J Mol Model (2009) 15:309–321
DOI 10.1007/s00894-008-0392-y

ORIGINAL PAPER

A very large C-loop in EGF domain IV is characteristic of the P28 family of ookinete surface proteins

Babita Sharma · Rahul Dev Ambedkar

Received: 31 August 2008 / Accepted: 7 October 2008 / Published online: 5 December 2008

Abstract The P28 family of proteins are 28 kDa proteins expressed on the surface of sexual stages—zygote, ookinete and young oocyst stages—of *Plasmodium* species when the parasite resides inside the mosquito midgut. Together with P25 proteins, P28 proteins protect the parasite from the harsh proteolytic environment prevailing inside the mosquito midgut. Vaccines against these proteins induce antibodies in vertebrate hosts that are capable of inhibiting parasite development in the mosquito midgut, thus preventing transmission of the parasite from the mosquito to another human host. These transmission-blocking vaccines are helpful in reducing the burden caused by malaria, which affects 300–600 million, and kills 1–3 million, people annually. The purpose of this study was to structurally characterise six members of the P28 family of ookinete surface proteins with the help of homology modelling, to compare these proteins in terms of transmission blocking and host parasite interactions, and to analyse phylogenetic relationships within the P28 family and with the P25 family. Our results indicate that all the members of the P28 family studied have four EGF domains arranged in triangular fashion with a very big C loop present in EGF domain IV, which could serve as a diagnostic feature of the P28 family as this loop is absent in the P25 family of ookinete surface proteins. The models of the P28 family of ookinete surface proteins obtained may help in understanding the biology of the parasite inside the mosquito midgut, and in designing transmission-blocking vaccines against malaria in the absence of experimentally determined structures of these important surface proteins.

Keywords Epidermal growth factor domain · Glycosylphosphatidylinositol anchor · Homology modelling · Phylogeny · Plasmodium 28 protein

B. Sharma (✉) · R. D. Ambedkar
School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India
e-mail: babita2005@gmail.com

Background

Malaria affects 300–500 million people and takes lives of 1–3 million people annually, with 40% of the world's population at risk [1–2]. Upon ingestion by a mosquito with the blood meal, gametocytes of *Plasmodium* undergo transformation, forming micro- and mega-gametes. These gametes fertilise to form zygotes, which develop into banana-shaped motile ookinetes. Ookinetes cross the mosquito midgut epithelium to form oocysts. Upon maturation each oocyst releases thousands of sporozoites, which migrate to the salivary glands of the mosquito host to infect another human host. P28 proteins are expressed on ookinete surfaces of all known *Plasmodium* species [3, 4]. These proteins are expressed after fertilisation, after P25 proteins, and continue to be expressed until the young oocyst stage [4]. P28 proteins are evenly and abundantly distributed on the ookinete surface [5–7] and are essential for parasite survival inside the mosquito midgut [8]. These proteins contain very few regions of genetic polymorphism [9] and hence are thought to be promising vaccine candidate proteins. Structurally, P28 proteins contain a secretary signal sequence followed by four EGF-like domains and a short hydrophobic glycosylphosphatidylinositol (GPI) anchor at the C-terminus [10]. Interaction of these proteins with single-chain immunotoxins results in

killing of the parasite, thus blocking transmission of the parasite to another human host [11]. The mechanism of this transmission blocking is unknown. It has been proposed that the transmission-blocking antibodies arrest/slow down the movement of the ookinete in the mosquito midgut [12]. Also, it has been shown that if an ookinete is delayed in forming an oocyst, it is not able to survive inside the mosquito midgut [13]. Recombinant versions of transmission-blocking antibodies also block transmission [14, 15]. P28 protein from *Plasmodium falciparum* has been shown to elicit transmission-blocking antibodies in mice [16], and, along with P28 from *Plasmodium vivax*, represents a promising transmission-blocking vaccine candidate.

The purpose of this study was to structurally characterise the P28 family of ookinete surface proteins with the help of homology modelling, and to compare members of this family in terms of structure, function and host parasite interaction. Our results suggest that mature P28 proteins are triangular molecules—a shape necessary to tile/tessellate the surface of the parasite. P28 proteins are probably thicker than P25 proteins. They all have four EGF domains, with a very large C-loop present in EGF-like domain IV. This loop, which may also harbour secondary structures, is one of the distinguishing features of the P28 family and is not present in the P25 family proteins that are expressed just before P28 proteins. We hypothesise that the large C-loop may act as a hook to help fit molecules onto the surface of the parasite. The structure of this loop varies greatly among the P28 family members studied.

Methods

Database search and sequence alignment

The single letter amino acid sequences for the six members of the P28 family were downloaded from either Swiss-Prot [18] database release 50.4 or from NCBI GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein) as indicated in Table 1. A PSI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) [19] search was performed using these P28 sequences as query against the PDB database (http://www.rcsb.org/pdb/Welcome.do) [20]. After removing redundant sequences, it was found that each search resulted in single significant hit, i.e. the Pvs25 protein from *P. vivax* [17], which was found to be the only suitable template for modelling of the P28 proteins. The ClustalX program (http://www.embl.de/~chenna/clustal/darwin/index.html) was used for sequence alignment [21]. The Bioedit program (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html) was used to represent conservation in sequence alignment.

Homology modelling

The three-dimensional (3D) structures of six members of the P28 family were modelled in a stepwise procedure, using MODELLER version 9v1. This software implements homology modelling of proteins by satisfaction of spatial restraints [22]; 100 models for each P28 protein were generated with the help of Modeller. Bond and angle values are taken from CHARMM-22 force field. 3D models were generated by molecular probability density function optimisation. The X-ray crystallographic structure of Pvs25 from *P. vivax* (PDB id 1Z27) [17], was used as the template for modelling P28 proteins. In all six P28 proteins, there are 20 conserved cysteines in oxidised form forming ten disulphide bonds. These disulphide bonds are formed between cysteines 1–3, 2–4 and 5–6 of EGF domains II and III. In the case of EGF domains I and IV, only four cysteines are conserved. The ten conserved disulphide bonds were assigned as constraints to the program during modelling. ModLoop (http://alto.compbio.ucsf.edu/modloop//modloop.html) [23, 24] was used to model the bigger and problematic loops carefully as these were showing maximum root mean square deviation (RMSD) variations. Special care was taken to model the C loop of EGF domain IV, as this was a very big loop of approximately 20 amino acids.

Model validation

Evaluation of the models of P28 proteins was carried out using the programs ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) [25, 26], PROCHECK (http://nihserver.mbi.ucla.edu/SAVS/) [27] and WHATIF (http://swift.cmbi.kun.nl/WIWWWI/) [28, 29]. Only those models that showed a satisfactory PROSA, PROCHECK and WHATIF profile were selected. A 3D structural superimposition to compare the structures was done using the program STAMP, which is a part of VMD version 1.8.4 [30]. Domain–domain interactions of the P28 models obtained were calculated with the help of PDBsum [31].

Phylogenetic analysis

To analyse P28 protein sequences from an evolutionary perspective we used MEGA4 (http://www.megasoftware.net) software [32]. A phylogenetic tree was constructed using both the neighbour joining method and the minimum evolution (ME) method. Trees were constructed with MEGA 4.0 software with the following settings: 125 bootstrap replicates and 42,535 seed; a close neighbour interchange was used as a search option (level = 1) with an initial neighbour-joining tree, MaxTrees = 1. Gaps were

Table 1 Members of the P28 family and the sequences used for modelling

Source organism	Swiss-Prot entry/ GenBank accession no. (amino acid range used)	Sequence identity with Pvs25 protein (%)	Critical sequence identity with Pvs25 protein (%)
Plasmodium berghei	AAG27292 (23–209)	33.7	41
Plasmodium falciparum	Q6LEB4 (23–213)	36.2	46
Plasmodium gallinaceum	Q05439 (22–217)	35.2	45
Plasmodium ovale	BAB43949 (23–210)	34.0	41
Plasmodium vinckei	AAG27290 (23–209)	37.2	48
Plasmodium vivax	O96556 (23–217)	37.2	48

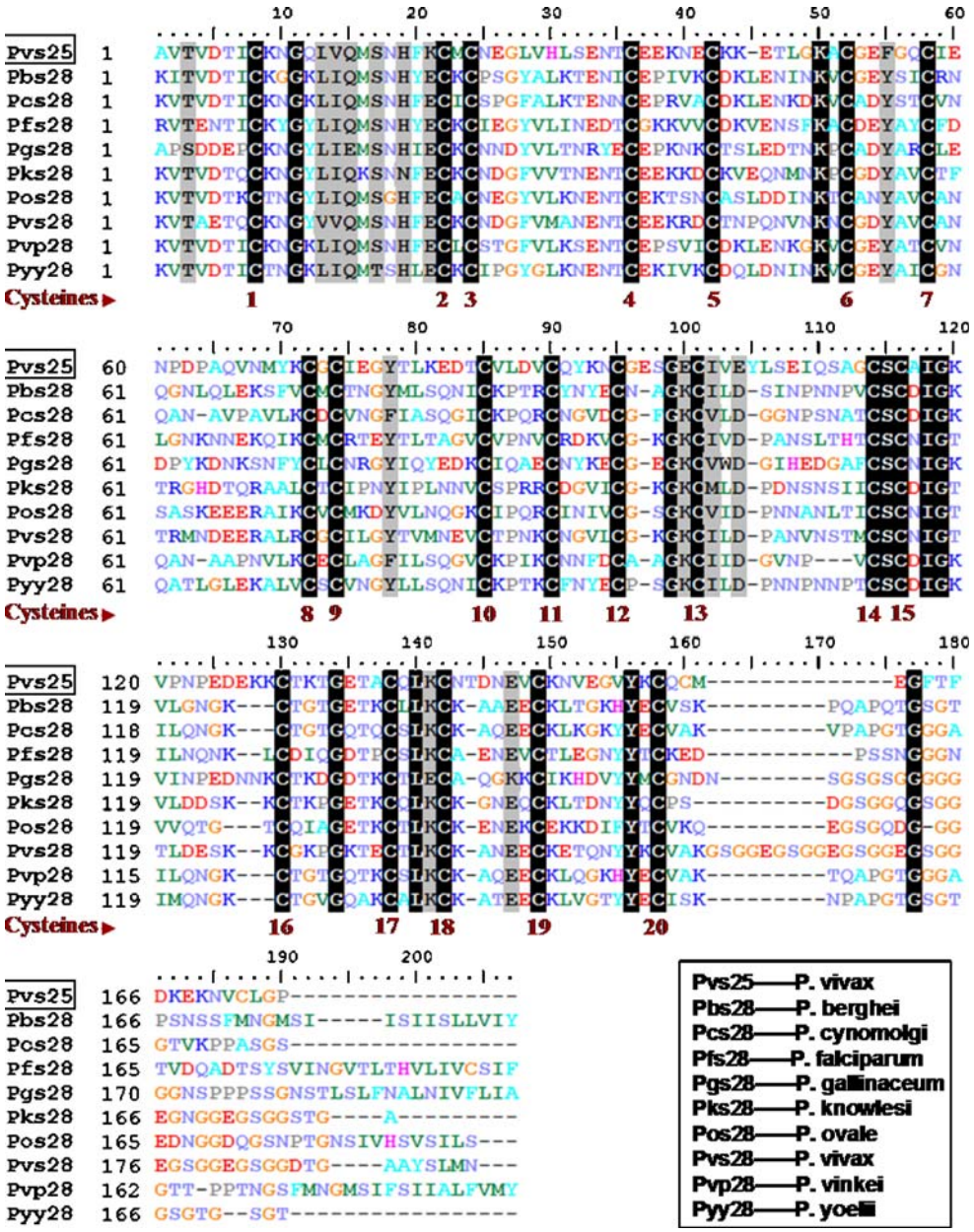

Fig. 1 Multiple sequence alignment with Pvs25 protein, which was used as a template to model P28 proteins, of eight sequences belonging to the P28 family of ookinete surface antigens. The 20 conserved cysteines and totally conserved residues are shown on a *black background* whereas semi-conserved residues are on a *grey background*. Amino acids: *Red* Negatively charged residues (D, E); *blue* positively charged residues (K, R); *bluish-purple* polar residues (N, Q, S, T); *green* aliphatic residues (I, L, M, V); *cyan* aromatic residues (F, Y); *pink* histidine; *buff* glycine; *grey* proline

considered as complete deletions. Evolution rates among sites were considered uniform.

Results and discussion

P28 proteins and sequence conservation

The P28 proteins present on the surface of the ookinetes of *Plasmodium* species are considered promising candidates for transmission-blocking malaria vaccines. Figure 1 shows a multiple sequence alignment of the nine members of the P28 family with Pvs25 protein, whereas Fig. 2 represents the WebLogo [33, 34] of this family. WebLogo clearly illustrates that all 20 cysteine residues are totally conserved in the P28 family. The structures of the models of P28 proteins were found to be similar to that of Pvs25 [17] owing to these 20 conserved cysteine residues (Fig. 1), with 34–38% sequence identity and 47–53% sequence similarity (Table 1) between the target P28 proteins and the template Pvs25 (PDB id 1Z27) protein. The lengths of the proteins used in this calculation are listed in Table 1.

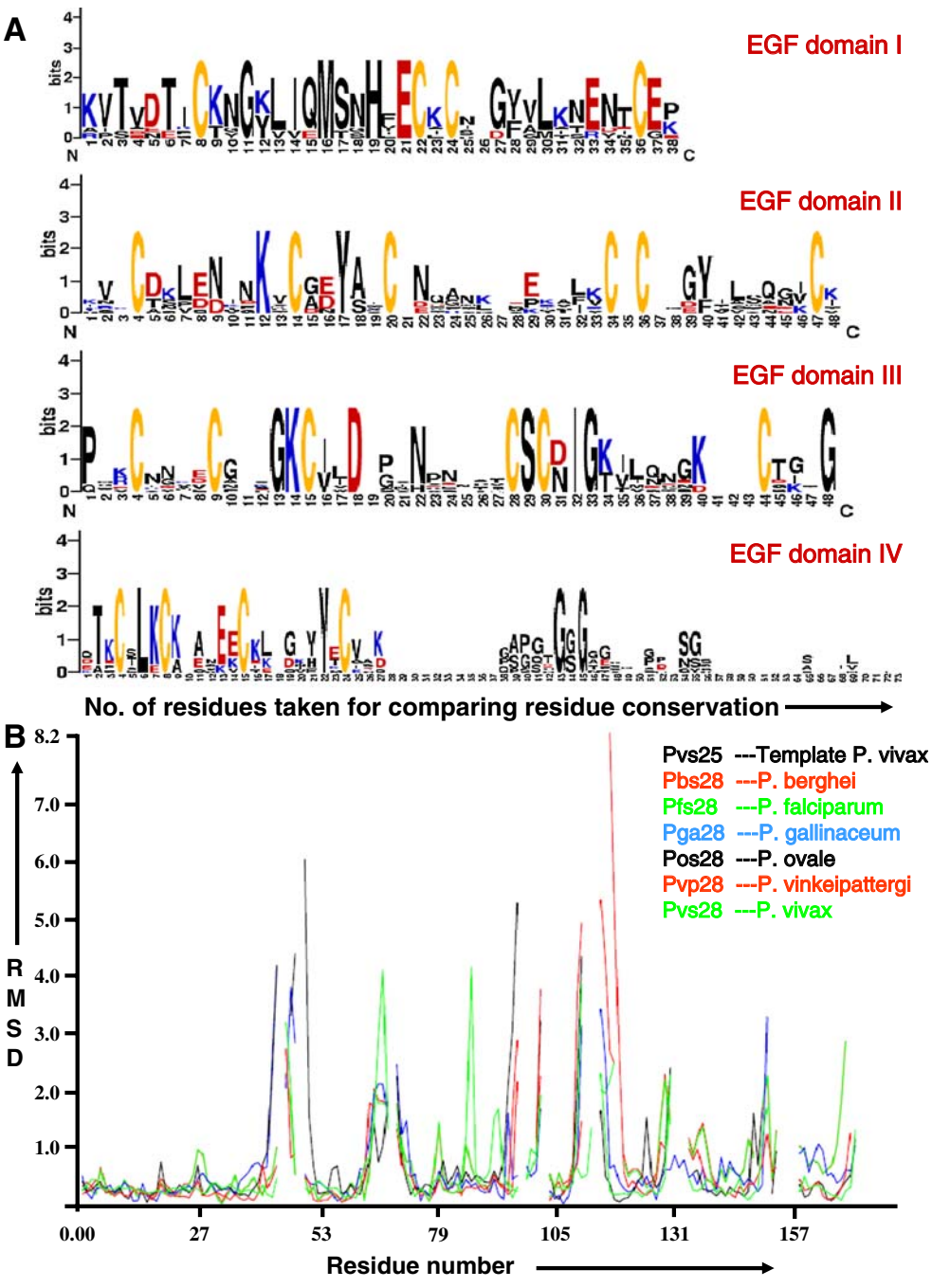

Fig. 2 **a** WebLogo [33, 34] representation of the six modelled sequences of the P28 family of ookinete surface antigens. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of the symbol within the stack indicates the relative frequency of each amino acid. *Red* Negatively charged residues (D, E); *blue* positively charged residues (K, R); *yellow* cysteine residues (C); *black* all other residues. Conservation of cysteine residues in all four EGF domains is clearly illustrated by the WebLogo representation. **b** Root mean square deviation (RMSD) of the six members of the P28 family compared to the template Pvs25 protein (PDB id 1z27). Residues 1–160 were taken for calculating RMSD

The 20 conserved cysteines provide a structural scaffold forming ten disulfide bonds in P28 proteins. Apart from the structural skeleton provided by the cysteines, P28 proteins have many other conserved and semi-conserved residues within the family [35]. These conserved interactions are important in making conserved contacts within the molecule among the four EGF domains of the protein as well as with other nearby molecules while forming a sheet on the surface of the parasite. Table 2 lists the conserved domain–domain interactions of the six members of P28 family studied as compared to the template protein Pvs25. Also, the EGF-domains have a high tolerance to polymorphism and mutations, i.e. are tolerant of insertions and deletions, and therefore a high level of similarity between the members of P28 family was expected, as was reflected in the models of P28 proteins we obtained. Table 2 clearly explains why the structure of P28 proteins are similar to that of Pvs25 protein.

Structural superimposition of P28 protein models with the Pvs25 template showed similarity overall except for EGF domain IV, where the presence of a very big C-loop differentiates P28 proteins from P25 proteins as shown in Fig. 3. Table 1 lists the overall sequence identities of P28 proteins with Pvs25 protein; sequence identity of Pvs25 protein is highest with Pvs28 protein from *P. vivax* and Pvp28 protein from *Plasmodium vinkei pattergi*. As well as the cysteines, many other residues are found to be conserved in P28 proteins. Residues Q^{15}, M^{16}, S^{17}, H^{19}, E^{21} of EGF domain I are conserved, whereas in EGF domain II, Y^{55} and Y/F^{79} are conserved. In EGF domain III, residues E/K^{99}, I^{102}, D^{104}, S^{115}, I^{118}, G^{119} and G^{134} are conserved, and in EGF domain IV, L^{140}, K^{141}, E^{147} and Y^{156} are conserved. High conservation of these residues across the P28 family is also reflected in the conserved domain–domain interactions found in the models generated for P28 proteins (Table 2). Many of these conserved interactions are thought to be involved in the interaction of one P28 molecule to another in order to form the ookinete surface protein sheet over the parasite surface.

The P28 model*s*

The models indicate that P28 proteins contain four EGF domains arranged in the form of a triangle, with EGF domain IV forming a very big C loop in comparison to the template protein Pvs25 as shown in Fig. 4. EGF domains are 30- to 40-residue domains containing two central beta strands followed by two smaller beta strands. A typical EGF domain contains three disulphide bonds formed with the help of six cysteines linked in a 1–3, 2–4 and 5–6 manner. This common linkage pattern leads to a general similarity in the structure of EGF domains. In the case of P28 proteins, only EGF domains II and III are typical EGF domains; EGF domains I and IV contain only four cysteines and form only two disulphide bonds instead of the usual three. EGF domain I of P28 proteins lacks the 1–3 disulphide bond seen in P25 proteins. The P28 EGF domain IV has only two disulphide bonds and the EGF domain is truncated after 160 amino acids. Thus, both EGF domains II and IV are incomplete in P28 proteins. Also, EGF domain IV in P28 proteins forms a very big C loop of approximately 20 residues.

The four EGFs of P28 proteins resemble each other except for the large loops like the C-loop of EGF domain IV as shown in Fig. 3. The number and spacing of the 20 P28 cysteines are conserved within the family. Models indicate that P28 family molecules may be thicker than P25 family molecules due to the presence of the unusually big C-loop in EGF domain IV. Also, we predict that this loop may provide the molecules with more flexibility to fit together to form a sheet over the parasite surface. The loop may also act as a "hook" to lock the proteins together while forming the protective sheet over the parasite surface, hence providing more protection to the parasite while it is in the mosquito midgut.

Validation of P28 protein models

Ramachandran plot statistics of the P28 protein models are shown in Table 3. No residue was found in the disallowed region in any of the six sequences of P28 family modelled, as illustrated in Fig. 5; except for Asn96 in Pbs28, Lys50 in Pos28 and Ala95 in Pvp28 protein. All three of these residues are present in the vicinity of a disulphide-bonded cysteine; any change in the conformation of these residues would lead to breakage of the nearby disulphide bond. As cysteines are the most conserved residues evolutionarily, for refinement of the model we chose to keep these three residues in the disallowed regions of the Ramachandran plot rather than breaking the disulphide bond. ProsaWeb analysis of the six members indicated that *z*-scores of all the P28 proteins lie within the standard defined values, and the energy plot indicated that the molecules have overall negative energy and therefore the models are energetically stable (Fig. 6).

P28 proteins contain very few polymorphisms as these proteins are expressed inside the mosquito host and are never exposed to the vertebrate immune system. This property of P28 proteins makes them suitable for use as vaccine candidates against malaria. A low degree of polymorphism is however found in P28 proteins, due to the presence of EGF domains. EGF domains tolerate polymorphism due to the extent of disulphide cross-linking within the EGF domain. P28 proteins have very few hydrophobic residues in their hydrophobic core and sequence changes are accommodated by solvent accessible atoms, leading to the same structural folds in the domain with variations lying in the loops of the EGF domains. This

Table 2 Conserved domain–domain interactions of P28 proteins

<—— A T O M 1 ——>				EGF–domain involved	<—— A T O M 2 ——>				EGF–domain involved	H-bond distance
Atom no.	Atom name	Amino acid name	Residue no.		Atom no.	Atom name	Amino acid name	Residue no.		
Pbs28 protein and Pvs25 protein										
120	N	Ser	17	I <->	876	O	Cys	114	III	2.75
126	N	Asn	18	I <->	868	OG	Ser	113	III	2.76
136	ND1	His	19	I <->	868	OG	Ser	113	III	2.96
103	N	Gln	15	I <->	1,137	O	Tyr	150	IV	3.05
111	O	Gln	15	I <->	1,126	N	Tyr	150	IV	2.95
250	OE2	Glu	33	I <->	601	OH	Tyr	78	II	2.85
885	N	Ile	116	III <->	1,115	O	Lys	148	IV	2.81
691	N	Cys	90	III <->	672	O	Pro	87	II	2.83
Pfs28 protein and Pvs25 protein										
134	N	Ser	17	I <->	883	O	Cys	114	III	2.82
137	OG	Ser	17	I <->	883	O	Cys	114	III	2.58
140	N	Asn	18	I <->	875	OG	Ser	113	III	2.86
150	ND1	His	19	I <->	875	OG	Ser	113	III	2.70
117	N	Gln	15	I <->	1,164	O	Tyr	151	IV	3.07
125	O	Gln	15	I <->	1,153	N	Tyr	151	IV	2.93
270	OE2	Glu	33	I <->	630	OH	Tyr	78	II	2.78
892	N	Ile	116	III <->	1,140	O	Asn	149	IV	2.90
706	N	Cys	90	III <->	690	O	Pro	87	II	2.90
Pgs28 protein and Pvs25 protein										
123	N	Ser	17	I <->	908	O	Cys	114	III	2.81
126	OG	Ser	17	I <->	903	N	Cys	114	III	2.66
129	N	Asn	18	I <->	900	OG	Ser	113	III	2.81
139	ND1	His	19	I <->	900	OG	Ser	113	III	2.90
106	N	Glu	15	I <->	1,211	O	Tyr	153	IV	3.11
114	O	Glu	15	I <->	1,200	N	Tyr	153	IV	2.95
917	N	Ile	116	III <->	1,187	O	Val	151	IV	3.02
Pos28 protein and Pvs25 protein										
123	N	Ser	17	I <->	908	O	Cys	114	III	2.81
126	OG	Ser	17	I <->	903	N	Cys	114	III	2.66
129	N	Asn	18	I <->	900	OG	Ser	113	III	2.81
139	ND1	His	19	I <->	900	OG	Ser	113	III	2.90
106	N	Glu	15	I <->	1,211	O	Tyr	153	IV	3.11
114	O	Glu	15	I <->	1,200	N	Tyr	153	IV	2.95
917	N	Ile	116	III <->	1,187	O	Val	151	IV	3.02
721	N	Cys	90	III <->	706	O	Gln	87	II	2.94
Pvp28 protein and Pvs25 protein										
111	OE1	Gln	15	I <->	939	N	Ser	130	III	2.61
123	N	Ser	17	I <->	804	O	Cys	110	III	2.81
106	N	Gln	15	I <->	1,102	O	Ala	150	IV	3.14
114	O	Gln	15	I <->	1,098	N	Ala	150	IV	2.94
813	N	Ile	112	III <->	1,090	O	Cys	148	IV	2.85
654	N	Cys	89	III <->	636	O	Pro	86	II	2.89
Pvs28 protein and Pvs25 protein										
124	N	Ser	17	I <->	856	O	Cys	114	III	2.81
130	N	Asn	18	I <->	848	OG	Ser	113	III	2.90
140	ND1	His	19	I <->	848	OG	Ser	113	III	2.70
35	OE2	Glu	5	I <->	1,022	NZ	Lys	137	IV	2.95
107	N	Gln	15	I <->	1,142	O	Tyr	151	IV	3.08
115	O	Gln	15	I <->	1,131	N	Tyr	151	IV	2.93
254	OE2	Glu	33	I <->	602	OH	Tyr	78	II	2.62
865	N	Ile	116	III <->	1,118	O	Asn	149	IV	2.88
698	OD1	Asn	91	III <->	608	OG1	Thr	79	II	2.77
688	N	Cys	90	III <->	670	O	Pro	87	II	2.94

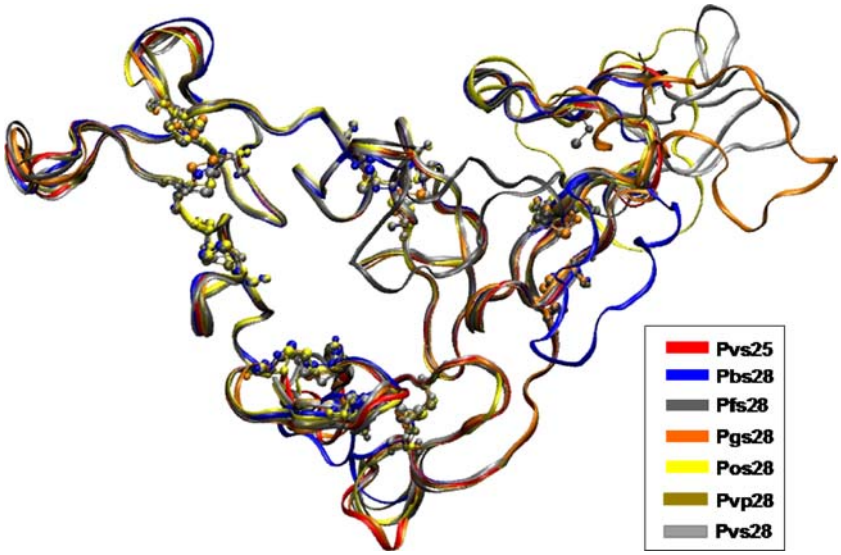

Fig. 3 Superimposed structures of theoretical 3D models of the six P28 proteins with Pvs25 (C-alpha atoms only) after structural alignment using the program STAMP [30], showing similarity of overall protein folding. In P28 proteins, EGF domain IV contains a extra-large (~20 residues) C-loop, which is not present in P25 proteins. Apart from this, the B loop of EGF domain II shows maximum conformational variations. The 20 conserved cysteines of the P28 family are shown in CPK representation

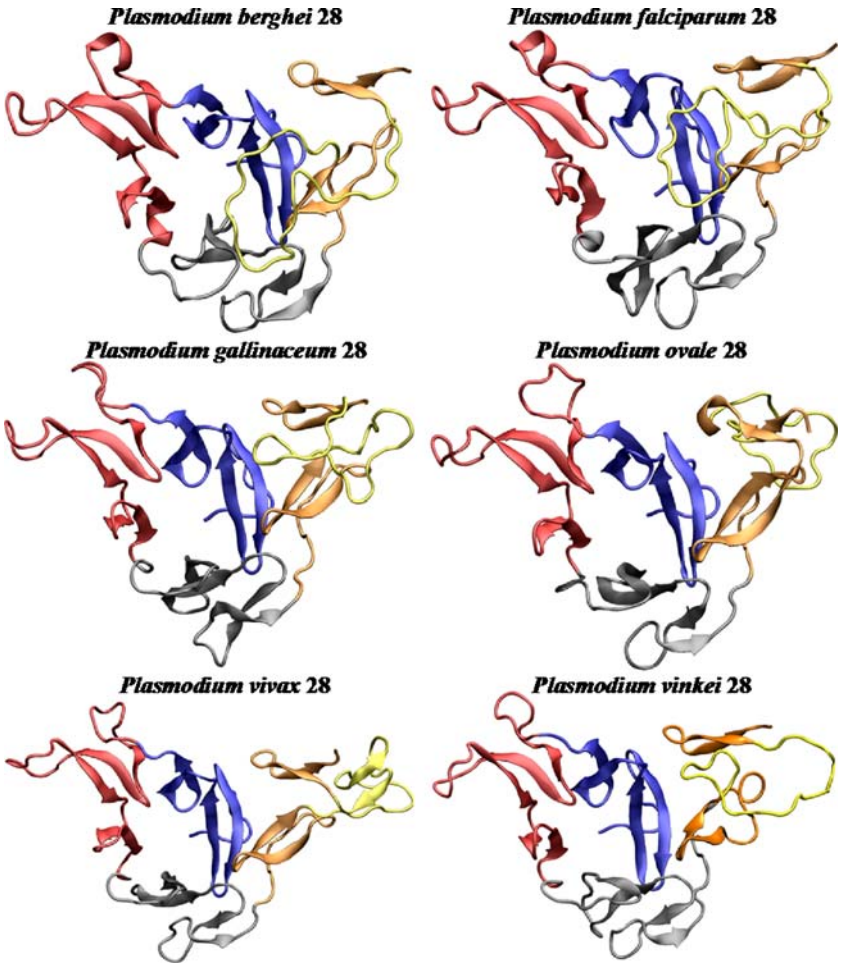

Fig. 4 Cartoon representations of six models of members of the P28 family of ookinete surface proteins. Models were generated assuming that the disulfide connectivity of the 20 cysteines in the P28 proteins is the same as that in the template Pvs25 (PDB code: 1Z27). All the P28 protein models show four EGF domains: *blue* EGF-domain I, *red* EGF-domain II, *grey* EGF-domain III, *orange* EGF-domain IV. The B-loop of EGF domain IV is shown in *yellow*

Table 3 Ramachandran plot statistics of P28 proteins studied

Modelled member proteins of P28 family	Residues in most favoured regions [A,B,L]	Residues in additional allowed regions [a,b,l,p]	Residues in generously allowed regions [~a,~b,~l,~p]	Residues in disallowed regions [~a,~b,~l,~p]
P. berghei (Pbs28)	84.80%	14.6%	0.6%	0.0%
P. falciparum (Pfs28)	84.20%	15.2%	0.6%	0.0%
P. gallinaceum (Pga28)	82.40%	17.0%	0.6%	0.0%
P. ovale (Pos28)	81.20%	18.8%	0.0%	0.0%
P. vinckei (Pvp28)	84.20%	15.8%	0.0%	0.0%
P. vivax (Pvs28)	88.10%	11.9%	0.0%	0.0%

is very well reflected in the six models obtained of P28 family members. It can be seen in Figs. 2 and 3 that the loops represent maximum RMSDs. The overall similarity among P28 family members indicates a similar major function of these molecules, whereas variations in the loop regions within and between *Plasmodium* species suggest differing minor functional requirements of the loops in different species.

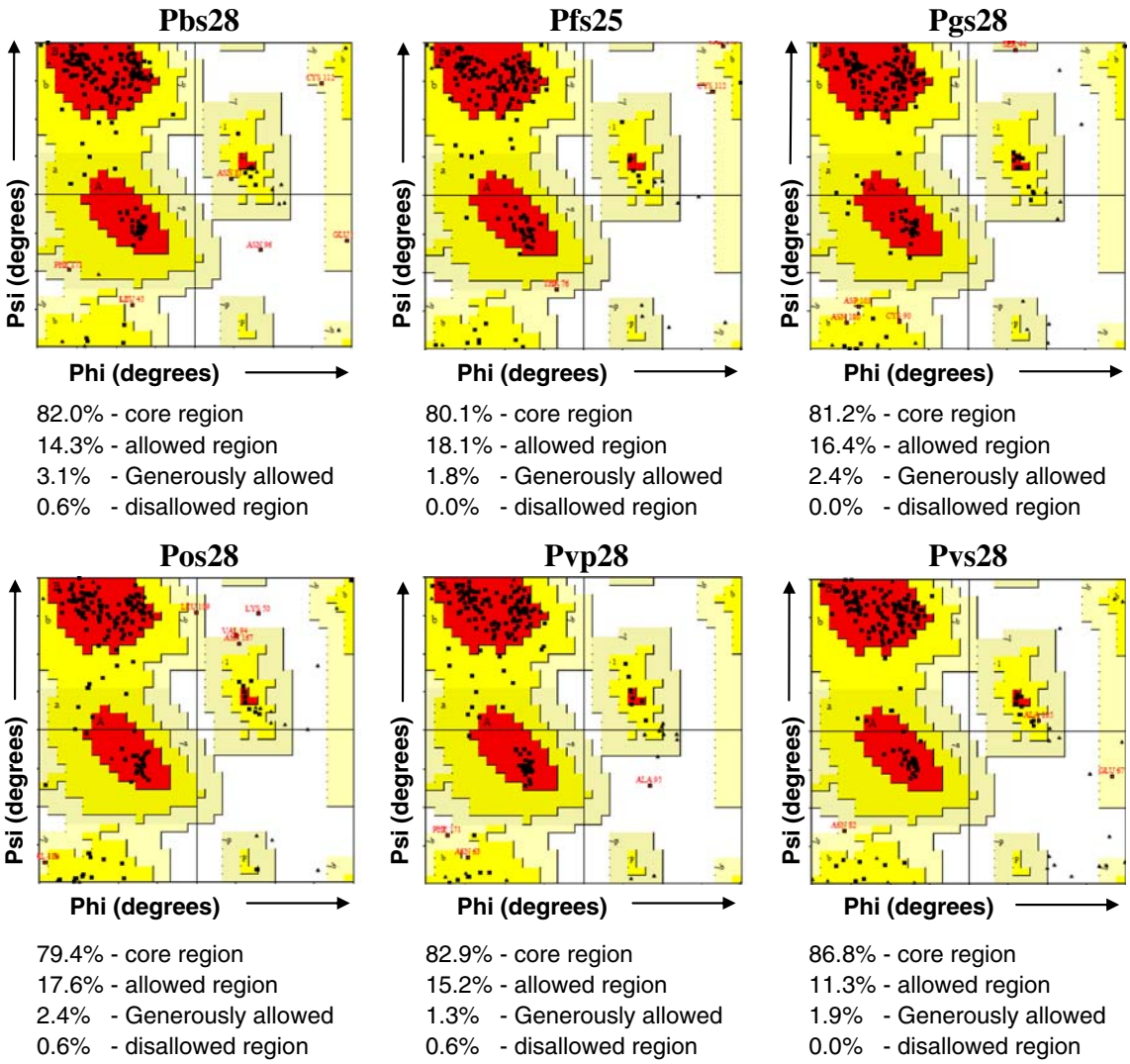

Fig. 5 Ramachandran plots for theoretical 3D models of P28 proteins using the X-ray crystallographic structure of native Pvs25 (PDB code 1Z27) as a template. The percentage of residues in the most favoured, allowed, and generously allowed regions are shown below the Ramachandran plots. No residues were found in the disallowed regions of the Ramachandran plots of any of the six members of P28 family studied

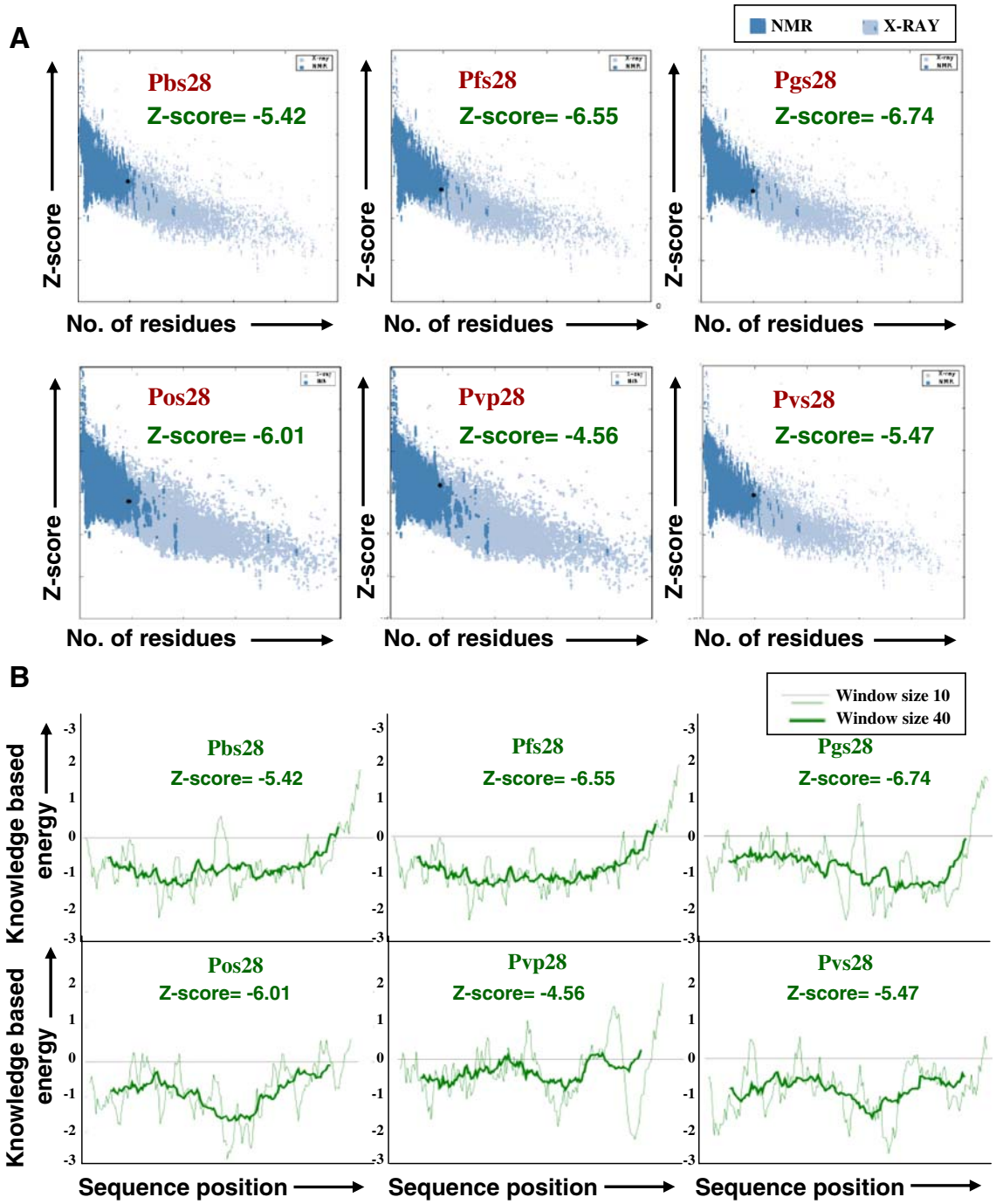

Fig. 6 ProsaWeb [25, 26] analysis of theoretical 3D models of six members of the P28 family of ookinete surface proteins. **a** *z*-Score plot of the P28 models obtained; the *z*-scores of all the six models lie within the normal range. **b** Energy plots for the six P28 models; the energy of residues for all the models developed is close to zero or negative indicating the stability of the models

Electrostatic representation

An electrostatic presentation of different faces of the P28 family members revealed that most of the charged residues are present on the surface of the proteins (Fig. 7). All the members studied carry an overall negative charge except for Pvp28 from *Plasmodium vinkei*, which carries an overall positive charge. The details of the charge, area and volume of the six P28 proteins studied are given in Table 4. Although the P28 family members present similar structures, the charge distributions on the molecules are quite different from each other. It can be clearly seen that a large negative patch is present on the ventral side (Fig. 7b) of both Pbs28 and Pfs28 proteins, which is formed by EGF

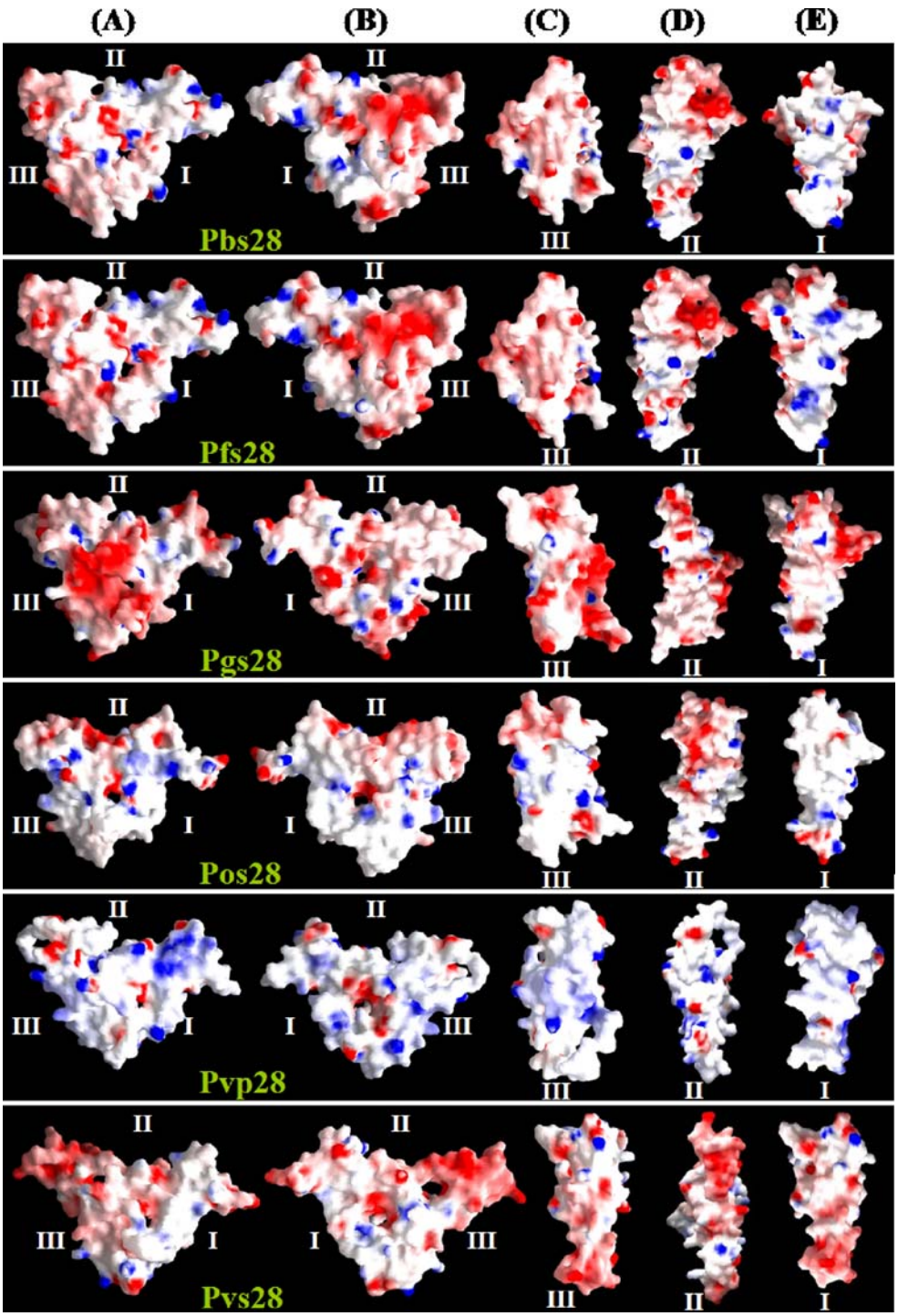

Fig. 7 Electrostatic molecular surface representation of the models of six transmission-blocking surface antigens of the P28 family prepared with the help of GRASP software (http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:GRASP). Surface potential was taken from −10 kT (red) to + 10 kT (blue). **a** Front view of P28 proteins, **b** the opposite face after 180 ° rotation, **c**–**e** view of each edge of the molecule. The programs PsoSA-web, PROCHECK and WHAT IF [25–29] programs were used for statistical evaluation of the final models. As can be seen in the different views of the molecules, all six members of the P28 family have very different charge distributions. Most of them are negatively charged overall

domain IV just below the big C-loop. A similar-sized negative patch is present on the dorsal surface of Pgs28 protein but this patch surrounds the central pore of the molecule, indicating the affinity of these areas to positively charged molecules. Members of the P28 family show a very different charge distribution when compared to each other, as illustrated in Fig. 7.

To determine whether these proteins have a different amino acid composition in terms of charged residues, we carried out an amino acid composition analysis of positively and negatively charged residues of the six P28 family proteins. A comparison was also done with template protein Pvs25. The percentage of positively charged amino acids (K, R and H) is almost constant in the P28 proteins and is similar to that of the template Pvs25, as shown in Fig. 8, whereas in the case of negatively charged residues (D and E), the amino acid composition was found to vary greatly and there were less negatively charged residues than in the

Table 4 Charge, area and volume of the proteins studied

Protein	Total charge	Area (square Å)	Volume (cubic Å)
Pbs28	−6.0000	9,914.90	22,214.45
Pfs28	−6.0000	9,721.52	22,200.61
Pga28	−10.0000	9,943.95	23,002.61
Pos28	−4.0000	9,418.26	21,335.01
Pvp28	4.0000	9,533.42	21,120.31
Pvs28	−5.0000	9,471.00	22,028.46

template protein Pvs25. Overall, the P28 proteins are relatively less negatively charged in comparison to the Pvs25 protein. Also, the negatively charged residues form big patches on the surface of the molecules. In the case of Pbs28 and Pfs28 proteins, negatively charged residues are concentrated near EGF domain IV on the ventral side of the molecule Fig. 7b (Pbs28 and Pfs28), and in the case of Pgs28, the negatively charged patch is situated near the central pore Fig. 7a (Pgs28 row) on the dorsal surface of the molecule. In the case of Pvs28, the extended EGF domain IV is more negatively charged.

There is a minimal hydrophobic core in P28 molecules with comparatively few residues buried, thus most residues are solvent-accessible. Charged residues are situated mainly on the surface of the molecule (relatively very easily accessible molecular surface) thereby indicating the significance of these residues in the interaction of the parasite and the mosquito midgut components. The conserved cysteines are present in the core region of the P28 molecules, with relatively less surface exposure, which explains why the cysteines are conserved in all members of the P28 family. A similar cysteine arrangement and high sequence and structural identity were observed among the six P28 family members studied. Although members of the P28 family have similar functions, and all of them are thought to have four EGF domains, they differ considerably in the loop regions and therefore each is able to recognise different molecules presented on the surface of the mosquito midgut. The differences in charge distribution on the surface of the P28 molecules also suggests that they have a species-specific interaction with molecules present in the mosquito midgut.

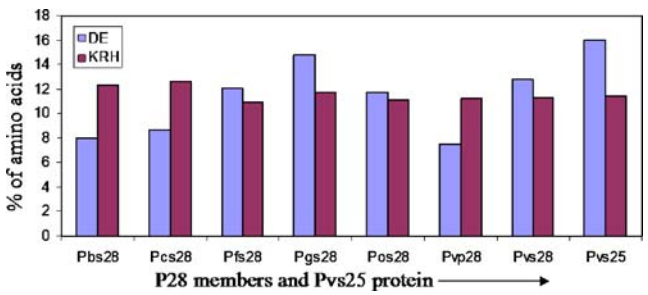

Fig. 8 Composition of charged amino acids in the six members of the P28 family and in the template protein Pvs25. *Blue bars* Negatively charged residues (D, E), *purple bars* positively charged residues (K, R, H). The percentage of positively charged residues is similar in all the P28 sequences, while the percentage of negatively charged residues varies within the P28 members and is less than that of Pvs25 protein in all the P28 proteins studied

Phylogenetic studies

Analysis of the phylogenetic tree for the genus *Plasmodium*, based on the amino acid sequences of ookinete surface proteins, suggests that all the human *Plasmodium* P25 and P28 proteins are situated between rodent and avian P25 and P28 homologues, respectively (Fig. 9). The position of *P. knowlesi* can be explained by the fact that it is now considered as the fifth human malaria parasite [36], whereas *P. reichenowi* causes a *P. falciparum*-like cerebral malaria in chimpanzees (the evolutionary closest primates to humans). Also, the G + C content of the genomic DNA of the parasite *P. falciparum* was found to be very similar to both the rodent and avian parasites [37], which suggests a closer relationship between these two species.

The similar structures observed in the P28 proteins are suggestive of convergent evolution of the family. Our studies on P28 proteins indicate that the rodent parasites are distant from other members of the family. These results are in agreement with those reported by Waters et al. [37] based on phylogenetic analysis of asexually expressed small subunit (SSU) rRNA genes. The P28 family was present in the common ancestor of all *Plasmodium* species. The evolutionary mechanisms responsible for the global divergence of these zygote/ookinete surface antigens between species, and the divergence between the P25 and P28 homologues within each species of *Plasmodium* are not yet known.

P28 proteins and repeats

In some P28 family members, the C loop contains multiple copies of repeats, e.g. a heptad repeat containing amino acids GSGGE/N is present in Pvs28 protein. In Pys28, five copies of the repeat sequence GTGS/T is present. The *P. yoelii* Pys28 contains five copies of a GTGS/T repeat, while the avian homologue, Pgs28, contains a series of Gly, Ser, and Pro residues in this same region. In contrast, the *P. berghei* P28 homologue Pbs21 has only one GTGS peptide sequence in the same loop, and the *P. falciparum* homologue Pfs28 has no repeat sequences. This region of the P28 homologues was found to be most divergent among members of the P28 family and is a characteristic difference between P28 proteins and P25 proteins. These repeated amino acid sequences are thought to be highly immunogenic. The importance of these regions for the parasite

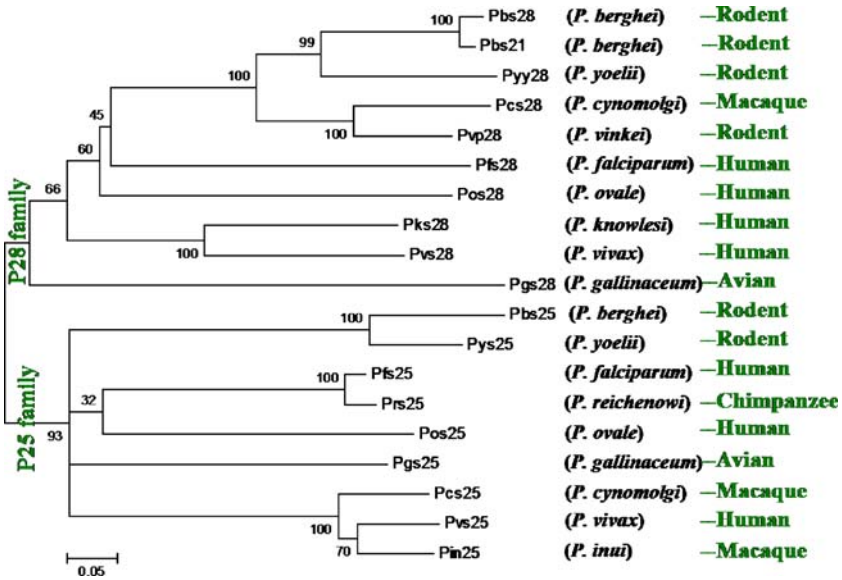

Fig. 9 Phylogenetic relationship of P25 and P28 proteins among the *Plasmodium species*. This neighbour-joining (NJ) tree was constructed with the help of MEGA 4.0 software (http://www.megasoftware.net) with the following settings: bootstrap (125 replicates, seed=42,535), search options close-neighbour-interchange (CNI) level = 1, maximum trees = 1, replicate trees = 125 were used for bootstrap analysis

remains to be determined. The antigenicity and immunogenicity of P28 proteins is sensitive to reducing agents, which break disulfide bonds. Single amino acid changes can lead to profound structural changes in P28 proteins.

Conclusion

All *Plasmodium* species express members of the P28 family of cysteine-rich surface antigens on their zygote, ookinetes and young oocyst stages. Our study suggests that the hallmarks of P28 proteins are the four EGF domains (fourth EGF domain is truncated) arranged in the triangular-prism-shaped structures necessary to coat the surface of the parasite, thus protecting it from the harsh environment prevailing inside the mosquito midgut. The P25 family has 22 conserved cysteines and a complete EGF domain IV, whereas the P28 family has 20 conserved cysteines and an incomplete EGF domain IV. The P28 family also differs from the P25 family in the number of amino acid residues between the 3rd and 4th cysteines of EGF-like domains II and III. In the P28 family, EGF domain IV contains a very big C loop as observed in the models of P28 proteins. In some members, this loop contains multiple copies of repeats, which are thought to be highly immunogenic. P28 proteins may be more flexible and thicker in comparison to P25 proteins due to the presence of the large C-loop in EGF domain IV. The flexible nature of this loop may help to ensure a better fit of P28 proteins on the parasite surface, where it may also act as a hook. The shape of P28 proteins is expected to vary from species to species due to this C-loop and, unlike in P25 proteins, variation in this loop will lead to differential sheet formation. Our study provides a pointer to the experimental studies required to verify the results obtained here. Structures of P28 proteins may help in the design of transmission-blocking vaccines. Protein–protein interaction studies with these proteins with antibodies and proteins of the mosquito midgut may help our understanding of the biology of the parasite inside the mosquito host.

Acknowledgements B.S. thanks CSIR (Council of Scientific and Industrial Research) for providing a fellowship. I thank my supervisor Dr. A. K. Saxena for allowing me to design and implement the protocols for this study independently.

References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. Nature 434:214–217. doi:10.1038/nature03342
2. World Malaria Report Geneva. 2005 RBM/WHO/UNICEF, Geneva, Switzerland
3. Kumar N, Carter R (1985) Biosynthesis of two stage-specific membrane proteins during transformation of *Plasmodium gallinaceum* zygotes into ookinetes. Mol Biochem Parasitol 14:127–139. doi:10.1016/0166-6851(85)90032-5
4. Paton MG, Barker GC, Matsuoka H, Ramesar J, Janse CJ, Waters AP, Sinden RE (1993) Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei*. Mol Biochem Parasitol 59:263–275. doi:10.1016/0166-6851(93)90224-L
5. Winger LA, Tirawanchai N, Nicholas NJ, Carter HE, Smith JE, Sinden RE (1998) Ookinete antigens of *Plasmodium berghei*. Appearance on the zygote surface of an Mr 21 kD determinant identified by transmission-blocking monoclonal antibodies. Parasite Immunol 10:193–207. doi:10.1111/j.1365-3024.1988.tb00214.x
6. Sinden RE, Winger L, Carter EH, Hartley RH, Tirawanchai N, Davies CS, Moore J, Sluiters JF (1987) Ookinete antigens of *Plasmodium berghei*: a light and electron-microscope immuno-

gold study of expression of the 21 kDa determinant recognized by a transmission-blocking antibody. Proc R Soc Lond B Biol Sci 230:443–458. doi:10.1098/rspb.1987.0028
7. Duffy PE, Pimenta P, Kaslow D (1993) C: Pgs28 belongs to a family of epidermal growth factor-like antigens that are targets of malaria transmission-blocking antibodies. J Exp Med 177:505–510. doi:10.1084/jem.177.2.505
8. Tomas AM, Margos G, Dimopoulos G, van Lin LH, Koning-Ward TF, Sinha R, Lupetti P, Beetsma AL, Rodriguez MC, Karras M, Hager A, Mendoza J, Butcher GA, Kafatos F, Janse CJ, Waters AP, Sinden RE (2001) P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. EMBO J 20:3975–3983. doi:10.1093/emboj/20.15.3975
9. Tsuboi T, Kaslow DC, Gozar MM, Tachibana M, Cao YM, Torii M (1998) Sequence polymorphism in two novel *Plasmodium vivax* ookinete surface proteins, Pvs25 and Pvs28 that are malaria transmission-blocking vaccine candidates. Mol Med 4:772–782
10. Carter R, Graves PM, Quakyi IA, Good MF (1989) Restricted or absent immune responses in human populations to *Plasmodium falciparum* gamete antigens that are targets of malaria transmission blocking antibodies. J Exp Med 169:135–147. doi:10.1084/jem.169.1.135
11. Yoshida S, Ioka D, Matsuoka H, Endo H, Ishii A (2001) Bacteria expressing single-chain immunotoxin inhibit malaria parasite development in mosquitoes. Mol Biochem Parasitol 113:89–96. doi:10.1016/S0166-6851(00)00387-X
12. Kaslow DC (1996) Transmission-blocking vaccine. In: Hoffman SL (ed) Malaria vaccine development: a multi-immune response approach. American Society for Microbiology, Washington, DC, pp 181–227
13. Han YS, Thompson J, Kafatos FC, Barillas-Mury C (2000) Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. EMBO J 19:6030–6040. doi:10.1093/emboj/19.22.6030
14. Tsuboi T, Cao YM, Hitsumoto Y, Yanagi T, Kanbara H, Torii M (1997) Two antigens on zygotes and ookinetes of *Plasmodium yoelii* and *Plasmodium berghei* that are distinct targets of transmission blocking immunity. Infect Immun 65:2260–2264
15. Yoshida S, Matsuoka H, Luo E, Iwai K, Arai M, Sinden RE, Ishii A (1999) A single-chain antibody fragment specific for the *Plasmodium berghei* ookinete protein Pbs21 confers transmission blockade in the mosquito midgut. Mol Biochem Parasitol 104:195–204. doi:10.1016/S0166-6851(99)00158-9
16. Duffy PE, Kaslow DC (1997) A novel malaria protein, Pfs28, and Pfs25 are genetically linked and synergistic as falciparum malaria transmission-blocking vaccines. Infect Immun 65:1109–1113
17. Saxena AK, Singh K, Su HP, Klein MM, Stower AW, Saul AJ, Long C, Garboczi DN (2006) Plasmodium P25 and P28, essential proteins for survival of the malaria parasite in the mosquito, are tile-like triangular prism. Nat Struct Mol Biol 13:90–91. doi:10.1038/nsmb1024
18. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788. doi:10.1093/nar/gkg563
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 15:403–410
20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. doi:10.1093/nar/28.1.235
21. Ramu C, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31(13):3497–3500. doi:10.1093/nar/gkg546
22. Sali A, Blundell T (1993) L: Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815. doi:10.1006/jmbi.1993.1626
23. András F, Andrej S (2003) ModLoop: automated modeling of loops in protein structures. Bioinformatics 19:2500–2501. doi:10.1093/bioinformatics/btg362
24. Fiser A, Kinh Gian Do R, Andrej S (2000) Ali: Modeling of loops in protein structures. Protein Sci 9:1753–1773
25. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 10:1093. doi:10.1093/nar/10.3.1093
26. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17:355–362. doi:10.1002/prot.340170404
27. Laskowski RA, MacArthur MW, Moss DS, Thornton J (1993) M: PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291. doi:10.1107/S0021889892009944
28. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8:52–56. doi:10.1016/0263-7855(90)80070-V
29. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272–272. doi:10.1038/381272a0
30. Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. J Mol Graph 14:33–38. doi:10.1016/0263-7855(96)00018-5
31. Laskowaski RA (2001) PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res 29:221–222. doi:10.1093/nar/29.1.221
32. Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599. doi:10.1093/molbev/msm092
33. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) Weblogo: A sequence logo generator. Genome Res 14:1188–1190. doi:10.1101/gr.849004
34. Schneider TD, Stephens R (1990) M: Sequence Logos: a new way to display consensus sequences. Nucleic Acids Res 18:6097–6100. doi:10.1093/nar/18.20.6097
35. Tsuboi T, Kaslow DC, Gozar MM, Tachibana M, Cao YM, Torii M (1998) Sequence polymorphism in two novel *Plasmodium vivax* ookinete surface proteins, Pvs25 and Pvs28 that are malaria transmission-blocking vaccine candidates. Mol Med 4:772–782
36. White NJ (2008) *Plasmodium knowlesi*: the fifth human malaria parasite. Clin Infect Dis 46:172–173. doi:10.1086/524889
37. Waters AP, Higgins DG, McCutchan TF (1991) *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. Proc Natl Acad Sci USA 88:3140–3144. doi:10.1073/pnas.88.8.3140