Research Article

The P25 Ookinete Surface Proteins: Homology Modeling and Phylogenetic Relationships

Babita Sharma and Manoj Kumar Jaiswal

School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

Correspondence should be addressed to Babita Sharma; babita2005@gmail.com

Received 19 February 2013; Accepted 27 March 2013

Academic Editors: B. Oliva, J. Ruan, and D. Sahoo

Copyright © 2013 B. Sharma and M. K. Jaiswal. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sexual stages of *Plasmodium* such as zygote, ookinete, and young oocysts express 25 kDa surface protein P25, which along with P28 proteins protect the parasite from harmful environment inside mosquito midgut. Vaccines against these proteins induce antibodies in vertebrate host capable to inhibit parasite development in mosquito midgut and thus preventing the transmission of parasite from mosquito to other human host. Transmission-blocking vaccines help reduce malaria burden. The purpose of this study was *in silico* structural characterization of P25 family proteins and to predict their phylogenetic relationships with other proteins. Results indicate that members of P25 family have four EGF domains arranged in triangular fashion with major variations lying in the loop regions. All 22 cysteines are conserved forming 11 disulphide bonds. The C-loop of EGF domain IV in P25 proteins is smaller in comparison to P28 proteins. B loop of EGF domain II showed maximum RMSD variations followed by loops of EGF domain III. P25 proteins are tile-like triangular flat proteins that protect the parasite inside mosquito midgut. Obtained structures will help in understanding the biology of the parasite inside the mosquito midgut. These structures may also help in designing transmission-blocking vaccine against malaria in absence of experimentally determined structures.

1. Introduction

During sexual stages of the *Plasmodium* parasite, many surface proteins are synthesized *de novo*. P25 proteins are the major surface proteins of *Plasmodium* ookinetes, having molecular weight of 25 kDa. P25 proteins are present on ookinete surface of all known *Plasmodium* species [1–3]. Many proteins have been identified as promising vaccine candidates against malaria including two proteins of P25 family [4]. P25 proteins start expressing immediately after fertilization and continue to be expressed on zygote, ookinete, and young oocyst stages of *Plasmodium* [5]. These proteins are present in abundance and are evenly distributed over the entire ookinete surface [6–8]. Gene knockout experiments suggest that these proteins along with P28 proteins are essential for the survival of parasite inside mosquito midgut [9]. Structurally, all P25 proteins contain a signal sequence, four epidermal growth factor (EGF) domains, and a C-terminal glycosylphosphatidylinositol (GPI) anchor [10, 11]. EGF domains are known to be present especially in surface

proteins where they participate in recognition and adhesion-like processes [12], which indicates that they play important roles in host parasite interactions. P25 proteins show very few sequence polymorphisms presumably because these proteins are never exposed to the vertebrate immune system [13]. Transmission blocking vaccines are being developed using P25 proteins. Antisera induced against Pfs25 protein from *P. falciparum* belonging to this family significantly blocked malaria transmission in nonhuman primates [14]. Experiments show that transmission blocking antibodies arrest/slow down movement of ookinete inside the mosquito midgut [15], and ookinete delayed in forming oocyst is not able to survive in the harsh proteolytic environment inside mosquito midgut [16]. The levels of antibody are directly related to the transmission blocking activity [17]. Recombinant versions of transmission blocking antibodies also block transmission [18, 19]. Structure of only one P25 protein is available to date, that is, Pvs25 from *P. vivax* [20]. The 3D structure of a P25 transmission blocking vaccine candidate proteins will help in understanding the drug and disease

mechanism as they are important vaccine candidates against malaria. Structure of these proteins can help in studying interaction of these proteins with respective transmission blocking antibodies and the receptors molecules present in mosquito midgut, which may help in understanding the biology of the parasite inside mosquito midgut. Homology modeling of eight P25 proteins suggested that all the mature P25 proteins are triangular flat molecules, each having four EGF domains arranged in the form of a triangle. A comparative analysis of the structures revealed that all P25 proteins have a structural scaffold of 22 conserved cysteines which form 11 disulphide bonds in all the members of the family. These disulphide bonds act as a skeleton for all the members of the P25 family leading to similarities in the overall structures of the family members. Dissimilarities in structures are primarily present in the loop regions of the P25 proteins. These loop regions are responsible for variable molecular recognition of P25 proteins among different species, whereas the conserved regions are responsible for similarity in functions of the P25 proteins among different *Plasmodium* species.

2. Methods

2.1. Template Selection. The amino acid sequences of the selected eight P25 proteins as shown in Table 1 were downloaded from the nonredundant protein sequence database Swiss-Prot [21] release 50.4. In order to select suitable templates to model the selected eight P25 ookinete surface proteins, BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) [22] search was performed against the protein structure database PDB (<http://www.rcsb.org/pdb/Welcomedo>, 52103 structures) [23] taking an e-value of 0.001. This homology search resulted in only one significant hit, that is, Pvs25 (PDB id = 1Z27) protein from *Plasmodium vivax*. The sequence identity and sequence similarity between the P25 proteins and Pvs25 protein are shown in Table 1. Sequence alignment was done with the help of ClustalX and MUSCLE programs [24, 25]. All the P25 proteins contain 22 conserved cysteines in oxidized form resulting in formation of 11 disulphide bonds. Within an EGF domain, disulphide bonds are formed between 1-3, 2-4, and 5-6 cysteines. Out of the four EGF domains, EGF domain II, EGF domain III, and EGF domain IV contain six cysteines, whereas EGF domain I contains only four cysteines forming disulphide bonds between cysteines 1-3 and 2-4.

2.2. Homology Modeling. In present study, eight sequences of P25 family were modeled with the help of Modeller version 8v2. Modeller implements homology modeling of proteins by satisfaction of spatial restraints [26]. The query P25 sequences and template Pvs25 sequence after alignment using MUSCLE program were provided to Modeller, and 100 models for each P25 protein were generated. Modeller derives restraints from known related structures automatically. These restraints include bond angles, distances, dihedral angles, pairs of dihedral angles, and some other spatial restraints. Bond and angle values are taken from CHARMM-22

TABLE 1: Members of ookinete surface antigen family P25.

Family	<i>Plasmodium</i> species	Swiss-Prot entries of P25 proteins	Sequence identity with Pvs25 protein	Sequence similarity with Pvs25 protein
P25	<i>P. falciparum</i>	P13829 (23-195)	44.63%	62.15%
	<i>P. berghei</i>	O00842 (23-192)	44.57%	61.71%
	<i>P. cynomolgi</i>	Q5EXQ8 (23-194)	81.71%	86.86%
	<i>P. gallinaceum</i>	P13401 (23-191)	47.43%	61.71%
	<i>P. inui</i>	Q5EXQ7 (23-194)	85.14%	91.43%
	<i>P. ovale</i>	Q969A0 (23-193)	49.14%	62.29%
	<i>P. reichenowi</i>	P19455 (23-193)	45.45%	60.80%
	<i>P. yoelii</i>	O00925 (23-192)	44.57%	61.71%

Starting and ending positions of ookinete surface proteins (without signal peptide and GPI anchor) are mentioned in parentheses with corresponding entries.

force field. All the 3D models were generated by molecular probability density function optimization. As template for homology modeling, X-ray crystallographic structure of Pvs25 from *Plasmodium vivax* (Protein database structure id of template that was used, <http://www.rcsb.org/pdb/explore/explore.do?structureId=1z27>) [20] was used. This template shows cysteine arrangement similar to P25 proteins. It is known that cysteines are the most conserved residues during evolution, and all P25 proteins contain four EGF domains with all the 22 cysteines in oxidized form; therefore, disulphide bonds were assigned as constraints to the program during modeling of the proteins. For each of the eight members, 100 models were generated with the help of Modeller program, and the best structures were selected on the basis of standard evaluation methods. In all the eight P25 proteins studied, loops were seen most variable, especially the B loop of the EGF domain II; therefore, the loop regions of P25 proteins were carefully modeled with the help of ModLoop server (<http://modbase.compbio.ucsf.edu/modloop/>) [27, 28].

2.3. Model Evaluation. For evaluation of models of the P25 proteins obtained by modeling with the help of Modeller, ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) [29, 30], PROCHECK (<http://nihserver.mbi.ucla.edu/SAVS/>) [31], and WHATIF (<http://swift.cmbi.ru.nl/servers/html/index.html>) [32] were used. Out of the total models generated, only those models were selected, which showed satisfactory ProSA-web, PROCHECK, and WHATIF profile, and then the best model was selected out of these. According to our observations during modeling, the loop regions of the P25 family members were showing maximum RMSD variations, and therefore these regions needed special attention. Taking this into account, the flexible large loops of P25 proteins were

FIGURE 1: Multiple sequence alignment of 8 sequences belonging to P25 family of ookinete surface antigens with Pvs25 protein which was used as a template to model P25 proteins. Negatively charged (D and E), positively charged (K and R), polar (N, Q, S, and T), aliphatic (I, L, M, and V), and aromatic (F and Y) residues are shown in red, blue, bluish-purple, green, and cyan colors, respectively. Histidine, glycine, and proline are represented in pink, buff, and gray colors, respectively. The 22 conserved cysteines and totally conserved residues in are shown the same background color, whereas semiconserved residues are shown in different background colors (as mentioned above) depending on the residue which is conserved.

modeled with the help of ModLoop server (<http://modbase.compbio.ucsf.edu/modloop/>) [33, 34]. 3D structural superimposition of the structures was done using the program STAMP which is a part of VMD (Visual Molecular Dynamics) version 1.8.4 [35]. Residue-residue interactions of the domains within the P25 proteins were calculated with the help of PDBsum (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/upload.html>) [36].

2.4. Phylogenetic Analysis. To analyze P25 protein sequences from evolutionary perspective, the phylogenetic tree was constructed using both Neighbor Joining (NJ) and Minimum Evolution (ME) method using MEGA4 (<http://www.megasoftware.net/>) software [37]. The settings of MEGA4 were as follows: 125 bootstrap replicates and 42535 seed; a close neighbor interchange was used as search option (level = 1) with an initial neighbor-joining tree, MaxTrees = 1. Gaps were considered as complete deletions. Evolution rates among sites were considered uniform.

3. Results and Discussion

3.1. The P25 Proteins. Two members of P25 family, Pfs25 from *P. falciparum* and Pvs25 from *P. vivax*, are already in clinical trial Phase II of vaccine development, which itself represents their importance as vaccine candidate proteins. The members of P25 family have a signal sequence which is followed by four tandem epidermal growth factor- (EGF-) like domains and a glycosylphosphatidylinositol (GPI) anchor, which anchors the protein to the parasite surface [4]. Single letter amino acid sequences of P25 proteins in FASTA format were downloaded from Swiss-Prot release 50.4 (<http://au.expasy.org/sprot/>) [21]. In order to find out the similarities between the sequences of P25 family, multiple sequence alignment was done with the help of MUSCLE and ClustalX programs. Figure 1 shows the multiple sequence alignment of the 8 members of P25 family with the template protein Pvs25 of the same family, whereas Figure 2 represents the web logo [38, 39] of the P25 family of ookinete surface proteins. Web logo of the family clearly indicates that all the twenty-two cysteines

FIGURE 2: (a) Weblogo representation for the modeled 8 sequences of P25 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 symbols within the stack indicates the relative frequency of each amino acid. Negatively charged (D and E), positively charged (K and R), and cysteine (C) residues are shown in red, blue, and yellow colors, respectively, whereas the rest are in black color. Conservation of cysteine residues in all the four EGF domains is well represented by the logos above. (b) Represents the root mean square deviation of the eight members of P25 family when compared to template Pvs25 protein (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1z27>).

are totally conserved in the P25 family. Obtained models of the eight P25 proteins were found to be similar to the template sequence Pvs25 [20]. This similarity was as expected because of the presence of 22 conserved cysteine residues (Figure 1), 45–85% sequence identity, and 61–91% sequence similarity (Table 1) between the target P25 protein sequences and the template Pvs25 (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1z27>) protein sequence. The lengths of the proteins taken for modeling the P25 proteins are mentioned in Table 1. P25 proteins have many other conserved and semiconserved residues within the family apart

from the structural skeleton provided by the cysteines. These conserved interactions are important in making specific functional contacts within the molecule among the four EGF domains of the proteins as well as with other nearby molecules while forming sheet on the surface of the parasite *Plasmodium*. Table 2 represents the list of conserved domain-domain interactions of the eight members of P25 family. All the modeled members were compared with template protein Pvs25. EGF domains have high tolerance to polymorphism/mutations and are tolerant to insertions/deletions; therefore, a high level of similarity between the members of

FIGURE 3: Superimposed structures of the eight P25 proteins with Pvs25 (C-alpha atoms only) after structural alignment using program STAMP. Superimposition between theoretical 3D models of P25 proteins, and template Pvs25 shows similarity in the overall fold of the proteins. B loop of EGF domain II shows maximum conformational variations in P25 family. The 22 cysteines of the P25 family are conserved in all the members.

P25 family was expected. Our findings confirmed the same. Due to 3 complete EGF domains found in P25 domains, they show even lesser polymorphism when compared to P28 proteins. In other words P25 proteins are more resistant to mutations and deletions in comparison to P28 proteins. Table 2 explains why the structures of P25 proteins are similar to Pvs25 protein. P25 models when superimposed with Pvs25 template showed similarity in the overall fold as shown in Figure 3. Many other residues along with cysteines were found to be conserved in the eight P25 proteins modeled in this study. Residues V/I², T/S³, T⁶, K⁹, G¹¹, L¹³, Q¹⁵, M¹⁶, S¹⁷, H¹⁹, E²¹, and N/S³² of EGF domain I were conserved. In EGF domain II, K/E³⁷, K³⁹, K/E⁴², V/L⁴⁸, G⁵³, D/E⁵⁴, F/Y⁵⁵, and Y/F⁷⁸ were conserved. EGF domain III showed that residues S¹¹⁴, I¹¹⁷, G¹¹⁸, V¹²⁰, N/S¹²², D/N¹²⁵, T/S¹³⁰, G¹³³ and T¹³⁵ are conserved, and in EGF domain IV, the conserved residues were L¹³⁹, D/E¹⁴⁴, Q/E¹⁴⁶, K/R¹⁴⁹, Y/F¹⁵⁵, D/E¹⁶¹, G¹⁶², F/Y¹⁶³ and E¹⁶⁸. High degree of sequence conservation among the sequences of the P25 family was also reflected by the conserved domain-domain interactions found in the models obtained for the P25 proteins as shown in Table 2. Many of these conserved interactions are supposed to be involved in the interactions of P25 molecules to another P25/P28 molecule in order to form sheet over the parasite surface.

3.2. The P25 Models. The homology modeling of eight P25 proteins showed that all of them contained four EGF domains arranged in the form of a triangle as shown in Figure 4. EGF

domains are basically 30–40 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 the manner 1–3, 2–4, and 5–6. In P25 proteins, EGF domains II, III, and IV represented typical EGF domains with six cysteines and three disulphide bonds. However, EGF domain I contained only four cysteines, which formed only two disulphide bonds instead of three, which means that three EGF domains in P25 proteins are complete, whereas the IVth domain is incomplete. The four EGF domains of P25 proteins shared remarkable resemblance with each other. The number and spacing of the twenty-two cysteines were conserved within the P25 family of ookinete surface proteins.

3.3. Validation of P25 Protein Models. Ramachandran plot statistics of the P25 protein models are shown in Figure 5. No residue was found in the disallowed region of any of the eight sequences of P25 family modelled as shown in Figure 5. Figures 6 and 7 show ProsaWeb analysis of the P25 members, which indicate that Z-scores of all the P25 proteins lie within the standard defined values, and the energy plot showed that the P25 protein models have overall negative energy and therefore are stable energetically. P25 proteins contained very few polymorphisms as they only expressed inside the mosquito host and are never exposed to the vertebrate immune system. This property of the P25 proteins makes them suitable to be used as vaccine candidate proteins against malaria. Less polymorphism found in P25 proteins is due to the extent of disulphide cross-linking within the EGF domain. P25 proteins had very few hydrophobic residues in their hydrophobic core, and sequence changes were accommodated by the solvent accessible atoms leading to the same structural folds in the domain with variations lying in the loops of the EGF domains. The obtained models of P25 family showed that the loops represent maximum RMSDs as shown in Figures 2 and 3. Overall similarity among the P25 family members indicated that these proteins may have a similar major functions, whereas limited variations in the loop regions indicated towards different functional requirements of the loops in different species.

3.4. Electrostatic Representation. Electrostatic representation of the different faces of the P25 family members revealed that most of the charged residues were present on the surface of the proteins as shown in Figures 7(a)–7(e). All the P25 members studied carry an overall negative charge. The details of the charge, area, and volumes of the eight P25 proteins studied are given in Table 3. Though the P25 family members represented similar structures. Despite very similar structures P25 member molecules differ considerably when it comes to charge distribution. It can be clearly seen that negatively charged residues were differentially distributed on the surface of different P25 members (Figure 7). The central pores of all the eight P25 proteins were surrounded by negatively charged residues. Positively charged residues were few and mostly located towards the edges of the molecules. Pcs25 from *P. cynomolgi* was the maximum negatively

TABLE 2: Conserved domain-domain interactions in P25 proteins (hydrogen bonds only).

ATOM 1			EGF domain involved	ATOM 2			EGF domain involved	H-Bond distance		
Atom no.	Atom name	Res. name		Res. no.	Atom no.	Atom name			Res. name	Res. no.
(1) Pbs25 protein and Pvs25 protein										
119	N	SER	17	B	857	O	CYS	113	G	2.80
130	ND2	ASN	18	B	845	O	CYS	111	G	2.67
102	N	GLN	15	B	1166	O	TYR	153	O	3.08
110	O	GLN	15	B	1155	N	TYR	153	O	2.95
227	N	HIS	30	B	399	OE1	GLU	53	R	3.14
236	O	HIS	30	B	403	N	PHE	54	R	2.77
229	ND1	HIS	30	B	399	OE1	GLU	53	R	2.76
866	N	ILE	115	G	1142	O	ASP	151	O	2.83
(2) Pcs25 protein and Pvs25 protein										
116	N	SER	17	B	873	O	CYS	115	G	2.83
119	OG	SER	17	B	868	N	CYS	115	G	2.66
122	N	ASN	18	B	865	OG	SER	114	G	3.04
132	ND1	HIS	19	B	865	OG	SER	114	G	2.70
31	OD1	ASP	5	B	1056	NZ	LYS	140	O	2.70
99	N	GLN	15	B	1174	O	TYR	155	O	3.08
107	O	GLN	15	B	1163	N	TYR	155	O	2.96
104	OE1	GLN	15	B	1033	N	GLN	138	O	2.71
105	NE2	GLN	15	B	1026	O	ALA	136	O	2.76
225	O	HIS	30	B	405	N	PHE	54	R	2.75
218	ND1	HIS	30	B	401	OE1	GLU	53	R	2.66
879	N	ILE	117	G	1154	O	GLY	153	O	2.81
681	N	CYS	89	G	665	O	LEU	86	R	2.98
(3) Pfs25 protein and Pvs25 protein										
116	NE2	GLN	15	B	1005	O	LYS	134	G	2.66
127	N	SER	17	B	846	O	CYS	113	G	2.77
130	OG	SER	17	B	841	N	CYS	113	G	2.67
139	ND1	HIS	19	B	838	OG	SER	112	G	2.77
35	OD1	ASP	5	B	1032	NZ	LYS	138	O	3.00
110	N	GLN	15	B	1149	O	TYR	153	O	3.05
118	O	GLN	15	B	1138	N	TYR	153	O	2.96
115	OE1	GLN	15	B	1012	N	SER	136	O	2.85
254	OE2	GLU	33	B	582	OH	TYR	76	R	2.62
855	N	ILE	115	G	1129	O	GLY	151	O	2.62
(4) Pgs25 protein and Pvs25 protein										
127	N	SER	17	B	865	O	CYS	112	G	2.82
133	N	ASN	18	B	857	OG	SER	111	G	2.96
143	ND1	HIS	19	B	857	OG	SER	111	G	2.80
37	OD1	ASN	5	B	1052	NZ	LYS	137	O	2.70
110	N	GLN	15	B	1170	O	TYR	152	O	3.09
118	O	GLN	15	B	1159	N	TYR	152	O	2.93
235	O	LEU	30	B	406	N	PHE	54	R	2.76
255	OE2	GLU	33	B	581	OH	TYR	75	R	2.78
877	N	ILE	114	G	1150	O	GLY	150	O	2.75

TABLE 2: Continued.

ATOM 1				EGF domain involved	ATOM 2				EGF domain involved	H-Bond distance
Atom no.	Atom name	Res. name	Res. no.		Atom no.	Atom name	Res. name	Res. no.		
(5) Pin25 protein and Pvs25 protein										
116	N	SER	17	B	847	O	CYS	115	G	2.85
132	ND1	HIS	19	B	839	OG	SER	114	G	2.70
31	OD1	ASP	5	B	1030	NZ	LYS	140	O	2.92
99	N	GLN	15	B	1150	O	TYR	155	O	3.09
107	O	GLN	15	B	1139	N	TYR	155	O	2.96
104	OE1	GLN	15	B	1007	N	GLN	138	O	2.75
105	NE2	GLN	15	B	1000	O	ALA	136	O	3.01
225	O	HIS	30	B	397	N	PHE	54	R	2.75
245	OE2	GLU	33	B	573	OH	TYR	77	R	2.67
853	N	ILE	117	G	1131	O	GLY	153	O	2.81
652	N	CYS	89	G	639	O	LEU	86	R	2.97
(6) Pos25 protein and Pvs25 protein										
120	N	SER	17	B	865	O	CYS	113	G	2.83
123	OG	SER	17	B	860	N	CYS	113	G	2.66
126	N	ASN	18	B	857	OG	SER	112	G	3.08
136	ND1	HIS	19	B	857	OG	SER	112	G	2.70
103	N	GLN	15	B	1172	O	PHE	153	O	3.12
111	O	GLN	15	B	1162	N	PHE	153	O	2.95
109	NE2	GLN	15	B	1022	O	GLU	134	O	3.01
234	O	HIS	30	B	413	N	TYR	54	R	2.79
874	N	ILE	115	G	1153	O	GLY	151	O	2.88
683	N	CYS	88	G	668	O	PRO	85	R	2.93
(7) Prs25 protein and Pvs25 protein										
125	N	SER	17	B	841	O	CYS	113	G	2.84
128	OG	SER	17	B	836	N	CYS	113	G	2.66
35	OD1	ASP	5	B	1024	NZ	LYS	138	O	2.86
108	N	GLN	15	B	1141	O	TYR	153	O	3.03
116	O	GLN	15	B	1130	N	TYR	153	O	2.91
113	OE1	GLN	15	B	1004	N	SER	136	O	2.73
252	OE2	GLU	33	B	576	OH	TYR	76	R	2.59
850	N	ILE	115	G	1121	O	GLY	151	O	2.77
(8) Pys25 protein and Pvs25 protein										
117	N	SER	17	B	851	O	CYS	113	G	2.81
123	N	ASN	18	B	843	OG	SER	112	G	2.94
127	OD1	ASN	18	B	843	OG	SER	112	G	3.32
128	ND2	ASN	18	B	839	O	CYS	111	G	2.66
133	ND1	HIS	19	B	843	OG	SER	112	G	2.69
100	N	GLN	15	B	1159	O	TYR	153	O	3.08
108	O	GLN	15	B	1148	N	TYR	153	O	2.95
105	OE1	GLN	15	B	1012	N	SER	136	O	2.68
106	NE2	GLN	15	B	1005	O	LEU	134	O	2.97
223	N	HIS	30	B	395	OE1	GLU	53	R	3.05
232	O	HIS	30	B	399	N	PHE	54	R	2.76
225	ND1	HIS	30	B	395	OE1	GLU	53	R	3.07
250	OD1	ASN	33	B	580	OH	TYR	76	R	2.58

FIGURE 4: Cartoon representation of 8 models of P25 family of ookinete surface proteins. Models were generated assuming that the disulfide connectivity of 22 cysteines in P25 proteins is the same as that in the template Pvs25 (PDB code: 1Z27). All the models of P25 proteins show four EGF domains as shown above. All the four EGF domains of each of the molecule are shown in different colors with EGF domain I in blue, EGF domain II in red, EGF domain III in gray, and EGF domain IV in orange color.

charged molecule with only three positively charged residues on dorsal surface (Figure 7(a)) and two on ventral surface (Figure 7(a)). P25 family members showed very different charge distribution when compared to each other as shown in Figures 7(a)–7(e). In order to know whether these proteins have a different amino acid composition in terms of charged residues, we carried out the amino acid composition analysis of positively and negatively residues of the eight P25 member proteins. Also a comparison was done with template protein Pvs25. It was found that the percentage of the positively charged amino acids (K, R and H) was almost constant among the P25 proteins and was similar to that of the template Pvs25 as shown in Figure 8, whereas in case of negatively charged residues (D and E), amino acid composition was found

variable Pcs25 showed maximum percentage of negatively charged residues in comparison to other members of P25 family. Overall, the P25 proteins are relatively more negatively charged and less positively charged. Pcs25 shows an overall negatively charged surface in GRASP surface presentation. The negatively charged residues form big patches on the surface of the molecules. There was a minimal hydrophobic core in P25 molecules with comparatively few residues buried, and most of them were solvent accessible. Charged residues were mainly situated on the surface of the molecule (relatively very high accessible molecular surface) thereby indicating significance of these residues in parasite's interaction to the mosquito midgut components. The conserved cysteines were present in the core region of the P25 molecules with least

FIGURE 5: Ramachandran plots for the models of P25 proteins using X-ray crystallographic structure of Pvs25 (PDB code 1Z27) as template. The number of residues in most favored region, allowed region, the generously allowed regions is shown below the Ramachandran plots. No residues were found in the disallowed regions of the Ramachandran plots of any of the six members of P25 family studied.

relative surface exposure which explains why cysteines were conserved in all the members of the P25 family. A similar cysteine arrangement and high sequence and structural identity between the members have been observed among

the eight members studied. Though members of P25 family have similar functions and all of them are supposed to have four EGF domains, they differed in the loop regions of the EGF domains, and therefore each of them was able to

(a)

FIGURE 6: Continued.

FIGURE 6: (a) Prosaweb analysis of theoretical 3D models of eight members of P28 family of ookinete surface proteins. The figure represents the Z-score plot of the eight P25 models obtained. Z-scores of all the eight models lie within the normal range of Z-scores as defined by Prosaweb. (b) Prosaweb analysis of theoretical 3D models of the eight members of P25 family of ookinete surface proteins. The figure represents the energy plots for the eight P25 models. Energy for all the models developed is close to zero or negative indicating the stability of the models energetically.

(a)

(b)

(c)

(d)

FIGURE 7: Continued.

FIGURE 7: (a) Electrostatic molecular surface representation of the models of P25 proteins: front view. (b) Electrostatic molecular surface representation of the models of P25 proteins: opposite view. (c) Electrostatic molecular surface representation of the models of P25 proteins: edge I. (d) Electrostatic molecular surface representation of the models of P25 proteins: edge II. (e) Electrostatic molecular surface representation of the models of P25 proteins: edge III. Electrostatic molecular surface representation of the models of transmission blocking surface proteins of P25 family was prepared with the help of GRASP software. Surface potential was taken from -10kT (red) to $+10\text{kT}$ (blue). Figure (a) represents the front view of P28 proteins, (b) the opposite face after 180° rotation, and ((c)–(e)) the view of each edge of the molecule. ProsaWeb, PROCHECK, and WhatIF programs were used for statistical evaluation of the final models. As represented by the different views of the molecule, all the 8 members of P25 family have very different charge distribution. Most of them are negatively charged overall.

TABLE 3: Charge area and volume of P25 proteins.

Protein name	Total charge	Area (square Å)	Volume (cubic Å)
Pvs25	−10.0000	8678.15	21577.83
Pbs25	−10.0000	9034.09	20682.42
Pcs25	−20.0000	8761.99	20607.53
Pfs25	−12.0000	8828.38	20419.73
Pin25	−9.0000	8718.09	20415.64
Pgs25	−7.0000	8499.17	20355.93
Pos25	−12.0000	8828.38	20419.73
Prs25	−4.0000	8940.43	20510.40
Pys25	−12.0000	8775.96	20150.51

recognize different molecules present on the surface of the mosquito midgut. Maximum variations were seen in the B loop of EGF domain II followed by the loop regions of EGF domain III as shown in Figures 2 and 3. EGF domains I and IV showed maximum similarities with minimum RMSD variations throughout the P25 family members Figures 2 and 3. The charge distribution on the surface of the P25 molecules also indicated their species specific interaction with the molecules present in the mosquito midgut.

3.5. Phylogenetic Studies. In order to know the phylogenetic relationships of P25 proteins with other ookinete surface

proteins, we did a NCBI BLAST analysis of the P25 family members. Sequences showing similarities were downloaded and a phylogenetic tree was constituted. P25 proteins show significant sequence similarities to P28 family of ookinete surface proteins, which was also shown in our previous publication [40]. 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 were situated between rodent and avian P25 and P28 homologues, respectively, as shown in Figure 9. The position of *P. knowlesi* can be explained by the fact that it is now considered as the fifth human malaria parasite [41], whereas *P. reichenowi* causes *P. falciparum*-like cerebral malaria in chimpanzees, considered the closest evolutionary primate to humans. Also it has been observed that the Guanine + Cytosine (G+C) content of the genomic DNA of *P. falciparum* was very similar to both rodent and avian parasites, which suggested closer relationship between the two species. Observed similar structures of P28 proteins indicate convergent evolution of the family. Our studies on P28 proteins indicated that the rodent parasites are distant from the rest of the members of the family. The results were also in agreement with those reported by Waters et al. [42] based on phylogenetic analysis of asexually expressed SSU rRNA genes providing an experimental support to our study. Apart from the similarities with the P28 proteins, the P25 proteins are distantly related to Fibrillins and latent

FIGURE 8: (a) Amino acid composition analysis of the residues of the eight P25 members studied and their comparison with the template protein Pvs25. Groups taken for this analysis are given on the x-axis, whereas the percentage of the residues is given on the y-axis. Amino acid groups taken according to are aliphatic (G, A, V, L, and I), sulphur containing (C and M), aromatic (F, W, and Y), neutral (S, T, Q, and N), positively charged (K and R), negatively charged residues (D and E), and proline (P). P25 sequences represent almost similar amino acid composition among the members. ((b)-(c)) Amino acid composition analysis of negatively (D and E) and positively charged (K, R, and H) residues of the eight P25 members studied and their comparison with the template protein Pvs25. (b) represents the percentage amino acid composition comparison for individual positively and negatively charged amino acids of studied P25 proteins and Pvs25 protein whereas (c) represents group-wise amino acid composition comparison of P25 proteins, and template Pvs25. Groups taken for this analysis were according to Koolman et al., negatively charged/acidic (D and E), and positively charged/basic (K, R, and H). P25 sequences represent almost similar amino acid composition of positively charged residues in group-wise presentation (b). Percentage of individual amino acids differs significantly among the members.

transforming growth factor proteins. It is interesting that the *Plasmodium* P25 and P28 proteins do not share similarities with any other lower prokaryotic genera. All the proteins as shown in Figure 9 belonged to higher chordates. It is likely that the P25 proteins originated from some higher chordate protein.

4. Conclusion

In this study, we have proposed eight models for different members of P25 family of ookinete surface proteins of *Plasmodium* sp. All these models were developed by homology modeling looking at their high sequence identity with the

template Pvs25 sequence. All the models we obtained for the eight P25 proteins contained four EGF domains arranged in triangular-prism-shaped structures. When compared to each other and to that of the template protein Pvs25, the P25 proteins represented maximum variability in the B loop of EGF domain II as shown in Figures 2(b) and 3. P25 proteins along with P28 proteins are supposed to form sheets on the parasite surface, which provides protection to the parasite from the external harsh environment inside mosquito midgut. Parasites lacking both P25 and P28 proteins are not able to survive. P25 proteins interact with transmission blocking antibodies leading to transmission blocking of malaria, and therefore the obtained structures may help in vaccine development against

FIGURE 9: Phylogenetic relationship of P25 proteins among the studied *Plasmodium* species. This neighbor-joining (NJ) tree was constructed with the help of MEGA 4.0 software (<http://www.megasoftware.net/>) with following settings: Bootstrap (125 replicates, seed = 42535), search options Close-Neighbor-Interchange (CNI) level = 1, maximum trees = 1, and replicate trees = 125 which were used for bootstrap analysis.

malaria. As these therapeutic antibodies are becoming more important, such models are being used to predict protein-protein interaction complex structures of low resolution, which can then be used to guide experimental methods. Our study also indicates that the P25 proteins show close relationships to P28 proteins evolutionarily. It is expected that P25 and P28 proteins evolved due to duplication. P25 proteins do not show any similarity to prokaryotes; instead they show distant relations to higher chordate proteins, like Fibrillins and latent transforming growth factor proteins. As shown in Figure 9, almost all other sequences belong to higher chordates specially mammals. The models obtained in this study may serve many significant purposes. Such models can be used to predict the mechanism of transmission blocking as done in our previous two publications [43, 44]. A protein-protein docking study with proteins present in mosquito gut, like laminin and annexin, will help in understanding the host parasite interactions inside the mosquito midgut.

Acknowledgments

Fellowship from Council of Scientific and Industrial Research (CSIR), India to Sharma B during Ph.D. is acknowledged. The authors thank Kushwaha, H. R., Yennamalli, R., and Patel, A., Centre for Computational Biology and Bioinformatics, Jawaharlal Nehru University, for their suggestions, discussions, and software support. Sharma B. thanks her advisors Professor P. K. Yadav, Professor K. C. Upadhaya, Professor R. Madhubala, Dr. S. Gaurinath, and Dr. A. Pareek for their help and guidance during the Ph.D. Sharma B. thanks her supervisor Dr. A. K. Saxena for allowing to design and implement the protocols of this study independently.

References

- [1] N. Kumar and R. Carter, "Biosynthesis of two stage-specific membrane proteins during transformation of *Plasmodium*

- gallinaceum* zygotes into ookinetes,” *Molecular and Biochemical Parasitology*, vol. 14, no. 2, pp. 127–139, 1985.
- [2] A. N. Vermeulen, T. Ponnudurai, and P. J. A. Beckers, “Sequential expression of antigens on sexual stages of *Plasmodium falciparum* accessible to transmission-blocking antibodies in the mosquito,” *Journal of Experimental Medicine*, vol. 162, no. 5, pp. 1460–1476, 1985.
 - [3] M. G. Paton, G. C. Barker, H. Matsuoka et al., “Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei*,” *Molecular and Biochemical Parasitology*, vol. 59, no. 2, pp. 263–275, 1993.
 - [4] G. A. Targett, “Malaria vaccines 1985–2005: a full circle?” *Trends in Parasitology*, vol. 21, no. 11, pp. 499–503, 2005.
 - [5] H. C. W. Fries, M. B. A. C. Lamers, J. van Deursen, T. Ponnudurai, and T. J. H. E. Meuwissen, “Biosynthesis of the 25-kDa protein in the macrogametes/zygotes of *Plasmodium falciparum*,” *Experimental Parasitology*, vol. 71, no. 2, pp. 229–235, 1990.
 - [6] L. A. Winger, N. J. Tirawanchai, J. Nicholas, H. E. Carter, J. E. Smith, and R. E. Sinden, “Ookinete antigens of *Plasmodium berghei*. Appearance on the zygote surface of an M_r 21 kD determinant identified by transmission-blocking monoclonal antibodies,” *Parasite Immunology*, vol. 10, no. 2, pp. 193–207, 1988.
 - [7] R. E. Sinden, L. Winger, and E. H. Carter, “Ookinete antigens of *Plasmodium berghei*: a light and electron-microscope immunogold study of expression of the 21 kDa determinant recognized by a transmission-blocking antibody,” *Proceedings of the Royal Society of London*, vol. 230, no. 1261, pp. 443–458, 1987.
 - [8] P. E. Duffy, P. Pimenta, and D. C. Kaslow, “Pgs28 belongs to a family of epidermal growth factor-like antigens that are targets of malaria transmission-blocking antibodies,” *Journal of Experimental Medicine*, vol. 177, no. 2, pp. 505–510, 1993.
 - [9] A. M. Tomas, G. Margos, G. Dimopoulos et al., “P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions,” *EMBO Journal*, vol. 20, no. 15, pp. 3975–3983, 2001.
 - [10] D. C. Kaslow, I. A. Quakyi, C. Syin et al., “A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains,” *Nature*, vol. 333, no. 6168, pp. 74–76, 1988.
 - [11] T. Tsuboi, D. C. Kaslow, M. M. G. Gozar, M. Tachibana, Y. M. Cao, and M. Torii, “Sequence polymorphism in two novel *Plasmodium vivax* ookinete surface proteins, Pvs25 and Pvs28, that are malaria transmission-blocking vaccine candidates,” *Molecular Medicine*, vol. 4, no. 12, pp. 772–782, 1998.
 - [12] E. Appella, I. T. Weber, and F. Blasi, “Structure and function of epidermal growth factor-like regions in proteins,” *FEBS Letters*, vol. 231, no. 1, pp. 1–4, 1988.
 - [13] R. Carter, P. M. Graves, I. A. Quakyi, and M. F. Good, “Restricted or absent immune responses in human populations to *Plasmodium falciparum* gamete antigens that are targets of malaria transmission-blocking antibodies,” *Journal of Experimental Medicine*, vol. 169, no. 1, pp. 135–147, 1989.
 - [14] C. Coban, M. T. Philipp, J. E. Purcell et al., “Induction of *Plasmodium falciparum* transmission-blocking antibodies in nonhuman Primates by a combination of DNA and protein immunizations,” *Infection and Immunity*, vol. 72, no. 1, pp. 253–259, 2004.
 - [15] D. C. Kaslow, “Transmission-blocking vaccine,” in *Malaria Vaccine Development: A Multi-Immune Response Approach*, S. L. Hoffman, Ed., pp. 181–227, American Society for Microbiology, Washington, DC, USA, 1996.
 - [16] Y. S. Han, J. Thompson, F. C. Kafatos, and C. Barillas-Mury, “Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes,” *EMBO Journal*, vol. 19, no. 22, pp. 6030–6040, 2000.
 - [17] E. M. Malkin, A. P. Durbin, D. J. Diemert et al., “Phase 1 vaccine trial of Pvs25H: a transmission blocking vaccine for *Plasmodium vivax* malaria,” *Vaccine*, vol. 23, no. 24, pp. 3131–3138, 2005.
 - [18] T. Tsuboi, Y. M. Cao, Y. Hitsumoto, T. Yanagi, H. Kanbara, and M. Torii, “Two antigens on zygotes and ookinetes of *Plasmodium yoelii* and *Plasmodium berghei* that are distinct targets of transmission-blocking immunity,” *Infection and Immunity*, vol. 65, no. 6, pp. 2260–2264, 1997.
 - [19] S. Yoshida, H. Matsuoka, E. Luo et al., “A single-chain antibody fragment specific for the *Plasmodium berghei* ookinete protein Pbs21 confers transmission blockade in the mosquito midgut,” *Molecular and Biochemical Parasitology*, vol. 104, no. 2, pp. 195–204, 1999.
 - [20] A. K. Saxena, K. Singh, H. P. Su et al., “The essential mosquito-stage P25 and P28 proteins from *Plasmodium* form tile-like triangular prisms,” *Nature Structural and Molecular Biology*, vol. 13, no. 1, pp. 90–91, 2006.
 - [21] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, “ExPASy: the proteomics server for in-depth protein knowledge and analysis,” *Nucleic Acids Research*, vol. 31, no. 13, pp. 3784–3788, 2003.
 - [22] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
 - [23] H. M. Berman, J. Westbrook, Z. Feng et al., “The protein data bank,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.
 - [24] R. Chenna, H. Sugawara, T. Koike et al., “Multiple sequence alignment with the Clustal series of programs,” *Nucleic Acids Research*, vol. 31, no. 13, pp. 3497–3500, 2003.
 - [25] R. C. Edgar, “MUSCLE: a multiple sequence alignment method with reduced time and space complexity,” *BMC Bioinformatics*, vol. 5, article 113, 2004.
 - [26] A. Sali and T. L. Blundell, “Comparative protein modelling by satisfaction of spatial restraints,” *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993.
 - [27] A. Fiser and A. Sali, “ModLoop: automated modeling of loops in protein structures,” *Bioinformatics*, vol. 19, no. 18, pp. 2500–2501, 2003.
 - [28] A. Fiser, R. K. G. Do, and A. Sali, “Modeling of loops in protein structures,” *Protein Science*, vol. 9, no. 9, pp. 1753–1773, 2000.
 - [29] M. Wiederstein and M. J. Sippl, “ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins,” *Nucleic Acids Research*, vol. 35, pp. W407–W410, 2007.
 - [30] M. J. Sippl, “Recognition of errors in three-dimensional structures of proteins,” *Proteins*, vol. 17, no. 4, pp. 355–362, 1993.
 - [31] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, “PROCHECK: a program to check the stereochemical quality of protein structures,” *Journal of Applied Crystallography*, vol. 26, pp. 283–291, 1993.
 - [32] G. Vriend, “WHAT IF: a molecular modeling and drug design program,” *Journal of Molecular Graphics*, vol. 8, no. 1, pp. 52–56, 1990.

- [33] R. W. W. Hooft, G. Vriend, C. Sander, and E. E. Abola, "Errors in protein structures," *Nature*, vol. 381, no. 6580, article 272, 1996.
- [34] A. L. Morris, M. W. MacArthur, E. G. Hutchinson, and J. M. Thornton, "Stereochemical quality of protein structure coordinates," *Proteins*, vol. 12, no. 4, pp. 345–364, 1992.
- [35] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996.
- [36] R. A. Laskowski, "PDBsum: summaries and analyses of PDB structures," *Nucleic Acids Research*, vol. 29, no. 1, pp. 221–222, 2001.
- [37] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.
- [38] G. E. Crooks, G. Hon, J. M. Chandonia, and S. E. Brenner, "WebLogo: a sequence logo generator," *Genome Research*, vol. 14, no. 6, pp. 1188–1190, 2004.
- [39] T. D. Schneider and R. M. Stephens, "Sequence logos: a new way to display consensus sequences," *Nucleic Acids Research*, vol. 18, no. 20, pp. 6097–6100, 1990.
- [40] B. Sharma and R. D. Ambedkar, "A very large C-loop in EGF domain IV is characteristic of the P28 family of ookinete surface proteins," *Journal of Molecular Modeling*, vol. 15, no. 3, pp. 309–321, 2009.
- [41] N. J. White, "*Plasmodium knowlesi*: the fifth human malaria parasite," *Clinical Infectious Diseases*, vol. 46, no. 2, pp. 172–173, 2008.
- [42] A. P. Waters, D. G. Higgins, and T. F. McCutchan, "*Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 8, pp. 3140–3144, 1991.
- [43] B. Sharma, "Structure and mechanism of a transmission blocking vaccine candidate protein Pfs25 from *P. falciparum*: a molecular modeling and docking study," *In Silico Biology*, vol. 8, no. 3-4, pp. 193–206, 2008.
- [44] B. Sharma and M. K. Jaiswal, "EGF domain II of protein Pb28 from *Plasmodium berghei* interacts with monoclonal transmission blocking antibody 13.1," *Journal of Molecular Modeling*, vol. 15, no. 4, pp. 369–382, 2009.