Journal of
Proteomics & Bioinformatics

Nair et al., J Proteomics Bioinform 2016, 9:7
DOI: 10.4172/jpb.1000405

Research Article Open Access

Structural Investigation and *In-silico* Characterization of Plasmepsins from *Plasmodium falciparum*

Divya N. Nair, Vijay Singh, Deekshi Angira and Vijay Thiruvenkatam*

Department of Physics & Biological Engineering, Indian Institute of Technology Gandhinagar, India

Abstract

Malaria is the one most important parasitic disease of humans, which affects approximately one hundred countries and threatens half of the world's population. The *Plasmodium* aspartic protease called plasmepsins performs a vital role in providing nutrients to the malaria parasite, which make these proteins as an excellent drug target. In this study, we have carried out a comparative protein modeling, active site analysis and structural analysis of all ten plasmepsins from *Plasmodium falciparum*. In this report we have analyzed *in-silico* structure modeling and made efforts to characterize plasmepsins structure and further propose its functional information. The phylogenetic analysis and disulfide linkages indicate, plasmepsin I to IV and HAP have similar structure, function property. Whereas, plasmepsin IX to X and plasmepsin VI to VIII belong to a separate cluster. The integral membrane protein plasmepsin V has a functional characterization as compared to the others aspartic proteases from *Plasmodium falciparum*. The overall study summarizes the need of good model to understand the structure and function activity and to design potent small molecule inhibitors targeting all ten plasmepsins, specifically Plasmepsin V as important target.

Keywords: Plasmepsins; *Plasmodium falciparum; In-silico* analysis; Homology modeling

Introduction

Malaria is a life-threatening disease caused by *Plasmodium* parasites transmitted to humans through the infected *Anopheles* mosquitoes, specifically in the region of tropical and subtropical climate and is active during the dusk and dawn [1]. About 20 different *Anopheles* species are globally important around the world [2]. Transmission is more intense in places where the mosquito lifespan is longer, because the parasite has time to complete its development inside the mosquito and it prefers to bite humans rather than other animals [3,4]. Resistance to antimalarial medicines is a recurring problem. In recent years, parasite resistance to artemisinins has been detected in 5 countries. If resistance to artemisinins develops and spreads to other large geographical areas, the public health consequences could be dire [5,6]. World Health Organization (WHO) recommends the routine monitoring of antimalarial drug resistance and supports countries to strengthen their efforts in this important area of work [7]. Effective drug is a critical component of malaria control. Selection of *Plasmodium sp.* parasites resistant to multiple drugs calls for accelerated efforts to develop new anti-malarial drugs targeting novel essential parasite pathways [8]. In humans, the disease is the result of infection by *Plasmodium falciparum* (Pf), *Plasmodium malariae, Plasmodium ovalae or Plasmodium vivax. Plasmodium knowlesi,* majorly known as the fifth human malaria parasite. Of these species, *Plasmodium falciparum* is the most lethal and considered as important target for drug intervention [9-13].

During the intra-erythrocytic stage of infection, the malaria parasite *Plasmodium falciparum* digests most of the host cell hemoglobin. Hemoglobin (Hb) degradation is essential for the growth of the malarial parasites. The degradation process that occurs inside an acidic digestive vacuole is thought to involve the action of aspartic proteases of *Plasmodium*, termed plasmepsins (PMs) [14-16]. The plasmepsins perform a crucial role in the provision of nutrients for the red cell stages of the malaria parasite and thus make excellent drug targets [17]. Inhibition of aspartic proteases aid to kill parasites in human red blood cells in culture. Proteases are known to play an important role in numerous pathways and represent potent drug targets for several chronic infectious diseases. Hence, aspartic proteases are considered one among the important drug-target.

The *P. falciparum* genome comprises a group of 10 aspartyl proteases called plasmepsin, initially discovered by identifying the hemoglobin digestion pathway in malarial victims and have been strongly considered as potential anti-malaria drug targets. The plasmepsins I, II, IV and HAP are expressed in the erythrocytic stages of the life cycle of *P. falciparum* and are localized in the food vacuole [18]. Plasmepsin V is an integral membrane protein present in the endoplasmic reticulum of the parasite suggesting a role in protein processing within the parasite [19-21]. Plasmepsin V, IX and X were expressed concurrently with plasmepsin I to IV but are not transported to the food digestive vacuole. The remaining plasmepsins VI- VIII are expressed during the exo-erythrocytic cycle and their functions are unknown [22,23]. For several years, the structure-based drug design of antimalarial compounds targeting *P. falciparum* and the plasmepsin inhibitors have received much attention due to their potential therapeutic use [24,25]. However, our current study shows a wide range of *in-silico* modeling and the structure analysis of all ten plasmepsins. This may provide a good foundation for designing potential anti-malarial drugs targeting plasmepsins.

Also, in order to make the strongest and most effective drug for therapy of malaria infection it will be necessary to optimize the binding of compounds to the most critical enzyme in the parasite. Based on our study we have shown that plasmepsin V is a key enzyme in the

***Corresponding author:** Vijay Thiruvenkatam, Assistant Research Professor, Department of Biological Engineering Indian Institute of Technology Gandhinagar, Ahmedabad-382424, Gujarat, India, Tel: +91 78782 51653; E-mail: vijay@iitgn.ac.in

Received May 05, 2016; **Accepted** July 25, 2016; **Published** July 29, 2016

Citation: Nair DN, Singh V, Angira D, Thiruvenkatam V (2016) Structural Investigation and *In-silico* Characterization of Plasmepsins from *Plasmodium falciparum*. J Proteomics Bioinform 9: 181-195. doi:10.4172/jpb.1000405

Copyright: © 2016 Nair DN, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

parasite and its inhibitors will provide basic foundation for further development of malarial drug based on plasmepsin V.

Materials and Methodology

Sequence retrieval and alignment

The sequence of plasmepsin from *P. falciparum* was obtained from the protein sequence database of NCBI (GenBank Id: AAB41811). The genome ID and genome localization of each plasmepsin was retrieved from PlasmoDB [26]. PlasmoDB is a functional genomic database for malaria parasites. The multiple alignments were carried out using CLUSTAL X [27]. The identical and similar amino acids are shaded or colored.

Fold-recognition and domain analysis

The domain composition of plasmepsins was analyzed using the SMART-Simple Modular Architecture Research Tool [28] in combination with the *P. falciparum* (Pfam) database [29]. The classification of plasmepsins was done by using PRED-CLASS server [30].

In-silico physico chemical characterization

For physico-chemical characterization, theoretical isoelectric point (pI), molecular weight, extinction coefficient and instability index [31] were computed using the Expasy's ProtParam server [32].

Functional characterization

The eukaryotic and viral aspartyl protease active site was predicted using PROSITE [33,34]. The SOSUI server was employed to identify the nature and function of the protein. The transmembrane region of the plasmepsin was predicted using HMM-TM: Prediction of Transmembrane Alpha-Helical Proteins. NetNGlyc 1.0 Server predicted N-glycosylation sites using artificial neural networks [35,36]. The intrinsic protein disorder regions was predicted by FoldIndex© [37] and GLOBPLOT 2 [38]. FoldIndex is a graphic web server that discriminates between folded and intrinsically unfolded proteins which defines the mean net charge, $|\langle R \rangle|$, as the absolute value of the difference between the numbers of positively and negatively charged residues at pH 7.0, divided by the total residue number, and the mean hydrophobicity, $\langle H \rangle$, as the sum of all residue hydrophobicity, divided by the total number of residues, using the Kyte/Doolittle scale, rescaled to a range of 0–1. Subcellular localization of protein using amino acid composition was achieved by Phobius predictor [39]. Secondary structure elements prediction was performed using the PSIPRED protein sequence analysis work bench [40]. The presence of disulfide bridges was analyzed using the DiANNA web server [41] and DISULFIND [42].

Phylogenetic analysis

The sequences were aligned by Clustal X version 2.0 with default options and phylogenetic tree was constructed based on the bootstrap neighbor-joining method [43] using Molecular Evolutionary Genetics Analysis (MEGA) software version 4.1 [44]. The stability of internal nodes was assessed by bootstrap analysis with 10,000 replicates.

Homology modeling, refinement and validation

The three dimensional (3D) structure of plasmepsin V was performed based on the homology modeling in PHYRE2 Protein Fold Recognition Server [45]. The Swiss PDB viewer and PyMOL were used to visualize and refine the models. The quality and validation of the obtained model was performed using PROCHECK [46], ERRAT [47] and PROVE software's [48,49] from "SAVES: Meta Server Structure Analysis" under NIH MBI Laboratory Server. The model was also analyzed in SuperPose [50] and DALI server [51]. The obtained 3D-model was stereo-chemically evaluated on RAMPAGE server [52], which provides a score based on proline and glycine preferential positions according to a Ramachandran plot. The Prosa Web server was used to predict the Z-score of the modeled structure [53].

Docking studies using AutoDock/Vina

Molecular docking protocols are widely used for predicting the binding affinities of ligands. The PDBQT files for protein and ligand preparation and grid box generation were done using Graphical User Interface program AutoDock Tools (ADT). ADT assigned polar hydrogens, united atom Kollman charges, solvation parameters and fragmental volumes to the protein. The prepared files were saved in PDBQT format. AutoGrid was used for the preparation of the grid map using a grid box and the grid size was set to $28 \times 32 \times 24$ xyz points and grid center was designated at dimensions (x, y, and z): 4.055, 45.931.554 and 18.066. A scoring grid is calculated from the ligand structure to minimize the computation time. AutoDock/Vina was employed for docking using protein and ligand information along with grid box properties in the configuration file and in docking both the protein and ligands are considered as rigid. The results less than 1.0 Å in positional root-mean-square deviation (RMSD) was clustered together and represented by the result with the most favorable free energy of binding. The pose with lowest energy of binding or binding affinity was extracted and aligned with receptor structure for further analysis.

Results and Discussion

The ten plasmepsin sequences from *P. falciparum* were retrieved in FASTA format from the NCBI database and were analyzed using bioinformatics. The genome localization, gene ID, physiochemical features of each plasmepsin are represented in Table 1. In *P. falciparum* 3D7, the plasmepsin I-IV, VIII and IX genes are located in chromosome 14 whereas the plasmepsins V, VI, VII and X are localized on chromosome 13, 3, 10 and 8 respectively [54]. The protein sequence for plasmepsins vary in their length, which runs from 380 aa to 630 aa with variable molecular weight. The Isoelectric point or isoionic point (pI) is the pH at which the net charge on a protein is zero or neutral and hence, it does not show any mobility in an electric field. The computed pI of Plasmepsin I, II, IV and X (pI<7) showed acidic nature whereas the other plasmepsins (V, VI, VII, VIII, IX and HAP) having pI>7 had a basic nature. It is an important feature for any protein to know for experimental aspect of molecular biology, especially in 2D gel electrophoresis, isoelectric focusing etc. The high extinction coefficient of plasmepsin V indicates presence of high concentration of Cys, Trp and Tyr present in its sequence. The computed extinction coefficients help in the quantitative study of protein-protein and protein-ligand interactions in solution. The instability index provides an estimate of the stability of the protein and the protein whose instability index is smaller than 40 is predicted as stable [55,56]. The instability index of plasmepsin VI, IX and X was computed by the server to be above 40 which predicts that the protein may be unstable.

Protein may have single or multiple functional regions called as domains, which perform specific biochemical functions [57]. In our study, we performed a conserved domain analysis using Pfam database and it was observed that all the plasmepsins contain Eukaryotic aspartic protease (ASP) domain. In addition to the ASP domain, plasmepsin V also contains Xylanase inhibitor N-terminal (Taxi_N) domain. The

Name of protein	Gene ID	Chromosome number	Genomic Localization	Protein Accession number	Aminoacid length	Molecular weight	Isoelectric Point	Ext. Coefficient	Instability index
Plasmepsin I	PF3D7_1407900	14	Pf3D7_14_v3: 288,297 - 289,655 (+)	P39898.2	452	51260.9	6.72	55030	28.47
Plasmepsin II	PF3D7_1408000	14	Pf3D7_14_v3: 293,471 - 294,832 (+)	P46925.1	453	51489.7	5.42	55155	37.4
HAP	PF3D7_1408100	14	Pf3D7_14_v3: 297,468 - 298,823 (+)	CAB40630.1	451	51693.2	8.04	55030	37.62
Plasmepsin IV	PF3D7_1407800	14	Pf3D7_14_v3: 283,086 - 284,435 (+)	AAW71463.1	448	50933.2	5.3	54000	30.06
Plasmepsin V	PF3D7_1323500	13	Pf3D7_13_v3: 975,403 - 977,175 (+)	AAW71468.1	590	68480.3	7.7	88225	37.05
Plasmepsin VI	PF3D7_0311700	3	Pf3D7_03_v3: 502,698 - 505,848 (+)	XP_001351190.1	432	49432.9	7.56	38320	43.68
Plasmepsin VII	PF3D7_1033800	10	Pf3D7_10_v3: 1,351,197 - 1,353,284 (+)	AAN35526.1	450	52328.1	8.26	67980	33.75
Plasmepsin VIII	PF3D7_1465700	14	Pf3D7_14_v3: 2,658,255 - 2,660,911 (-)	AAN37238.2	385	44254.2	9.1	48165	37.65
Plasmepsin IX	PF3D7_1430200	14	Pf3D7_14_v3: 1,188,349 - 1,191,466 (+)	AAN36894.1	627	74183	9.25	69860	41.52
Plasmepsin X	PF3D7_0808200	8	Pf3D7_08_v3: 416,344 - 418,065 (-)	XP_001349441.1	573	65114	5.35	47385	48

Table 1: The basic parameters of all plasmepsins from *P. falciparum* where gene id, chromosome number and gene localization were retrieved from PlasmoDB. Protparam database was used to find the molecular weight, isoelectric point, extinction coefficient and instability index. The protein accession number was taken from NCBI database.

Xylanase inhibitor N-terminal domain are mostly present in plants and have a major function to create the catalytic pocket necessary for cleaving xylanase [58,59]. The plant xylanase inhibitor proteins (XIPs) that inhibit fungal xylanase activity during rice blast fungal attack and are believed to act as a defensive barrier against fungal pathogens [60,61]. The presence of Taxi-N domain along with ASP domain in plasmepsin V indicates that it has an interesting characteristic which other plasmepsins do not have and thus it acquaints plasmepsin V as an important macro-molecule for further structure and function analysis. The domain organization, Bit score and the E-value of each domain search are shown in Table 2.

The multiple sequence alignment (MSA) of plasmepsins I, II, IV and HAP has more than 60% sequence similarity whereas other plasmepsins have diversity in their sequence. In MSA analysis, the active site region residues, Asp-Thr/Ser-Gly-Ser, in all the plasmepsins were found to be conserved except in HAP (Figure 1). In case of HAP the histidine residue is present instead of aspartic acid in its active site. The SCOPE database shows that plasmepsins are pepsin like proteins which falls in pepsin_retropepsin superfamily. It is evident from the results of multiple sequence alignments that there are two-conserved motifs and some short-length conserved regions were also found. The known conserved motifs [LIVMFGAC][LIVMTADN][LIVFSA]D[ST] G[STAV][STAPDENQ][^GQ][LIVMFSTNC][^EGK] [LIVMFGTA] of aspartic protease was subjected to search against plasmepsin for motif annotation using PROSITE search. These motifs or important conserved regions might help in inference of protein structure and sequence evolution history. Based on our analysis, all the plasmepsin sequence contains two distinct active site motif detected with 10-12 amino acid length were showed in Table 3 and Figure 1. The HAP protein contains only one motif which is present in the C-terminal region of the amino acid sequence. The malaria parasite *Plasmodium* uses plasmepsin to degrade hemoglobin in the red blood cells. It is experimentally proven that the plasmepsins I to V and HAP are capable of cleaving native hemoglobin as well as denatured globins with an optimized pH. But the activity of other plasmepsins has not yet been demonstrated. These results indicate that all the plasmepsins have a similar active site for aspartic protease activity.

The prediction of subcellular localization, disulfide linkage and glycosylation are shown in Figure 2. The online server Phobius predicted subcellular localizations which revealed that all plasmepsin I to V, IX and X are transmembrane proteins. Plasmepsin V is an integral membrane protein located in the endoplasmic reticulum of the parasite. Plasmepsin VI, VII and VIII contain no transmembrane region but they have a signal peptide region in their N-terminal. This signifies the fact that the plasmepsins which express in the erythrocytic cycle have transmembrane domain while others that express in exo-erythrocytic cycle have a signal peptide instead of transmembrane domain. Only plasmepsin V is predicted to have two transmembrane spanning regions which keep this protein separate from the other plasmepsins.

The disulfide bridge formation in a protein may play major role in the thermos-stability, functionality and structural stability of proteins. We calculated the cysteine residues and the disulfide bridges using various online tools like DIANNA server, DISULPHIDE and CYS_REC. The possible pairing and pattern with probability resulted in plasmepsin I- IV and VI- VIII having two disulfide bridges. Plasmepsin IX and X have four disulfide bridges were as plasmepsin V has five disulphide bridges. The disulphide architecture of plasmepsin IX is closely similar to plasmepsin V. The analysis also demonstrated that these proteins are glycosylated and most of the glycosylated proteins are known to be

Protein	Domain	Alignments (Start-End)	Bit Score	E-value
Plasmepsin I	Asp	138 - 448	246.9	2.8e-73
Plasmepsin II	Asp	139 - 449	250.8	1.8e-74
HAP	Asp	138 - 449	214.9	1.5e-63
Plasmepsin IV	Asp	137 - 445	249.6	4.3e-74
Plasmepsin V	TAXi_N, Asp	100 - 270 323 - 516	52.1 31.8	6.6e-14 8.3e-08
Plasmepsin VI	Asp	99 - 428	232.4	7.1e-69
Plasmepsin VII	Asp	91 - 443	169.8	8.1e-50
Plasmepsin VIII	Asp	61 - 384	221.0	2.1e-65
Plasmepsin IX	Asp	227 - 604	207.3	3.1e-61
Plasmepsin X	Asp	247 - 569	198.6	1.4e-5

Table 2: The domain sequence alignment of plasmepsins from *P. falciparum* done in the pfam database where the block diagram of green color ASP represents the Aspartic protease domain and red color block Taxi_N represents the Xylanase inhibitor N-terminal domain.

Protein Name	Motif I	Motif II	Residue number (Start-End)	Aspartyl proteases active site
Plasmepsin I	FIFDTGSANLWV	AIVDSGTSSITA	154 – 165, 334 – 345	157, 337
Plasmepsin II	FILDTGSANLWV	CIVDSGTSAITV	155 – 166, 335 – 346	158, 338
HAP	-	VILDSATSVITV	334 – 342	337
Plasmepsin IV	FIFDTGSANLWV	AVVDSGTSTITA	152 – 163, 332 – 343	155, 335
Plasmepsin V	LILDTGSSSLSF	MLVDSGSTFTHI	115 – 126, 362 – 373	118, 365
Plasmepsin VI	VVFDTGSSNLAI	AAIDTGSSLITG	115 – 126, 312 – 323	118, 315
Plasmepsin VII	VLFDTGSSQVWI	SIIDTGTYLIY	108 – 115, 321- 331	111, 324
Plasmepsin VIII	VLFDTGSTNLWI	AVIDTGTSSIAG	77 – 88, 267 – 278	80, 270
Plasmepsin IX	PIFDTGSTNIWI	LIFDSGTSFNSV	243- 251, 492 – 503	246, 495
Plasmepsin X	PIFDTGSTNVWV	VIFDTGTSYNTM	263- 271, 454 – 465	266, 457

Table 3: The conserved motif pattern of each plasmepsins and the residues number of aspartic acid responsible for the protease activity.

involved in various cellular biological functions like protein folding, cell-cell interactions, cell recognition and host-pathogen interactions.

The disorder predictors and unfoldability of protein have been proven to be useful in advancing our understanding of disordered regions with potential impact to improve the success rate of structural genomic efforts. The formation of protein crystals can be obstructed by the presence of highly flexible and disordered regions. The understanding of disordered regions as a result of structural bioinformatics efforts allowed us to extract and analyze patterns associated with these regions which can further help to overcome several potential bottlenecks for a successful structural genomics and X-ray crystallography. The unfoldability and disorder prediction of plasmepsins shows that plasmepsin IX has highly unfolded structure with the presence of six disordered regions [62]. The predicted secondary structure composition of plasmepsins was determined using the NPS@ server and Gor IV method which generates percentage of alpha helix, beta sheets and TM helix present. The results revealed that 40% beta sheets dominated among the secondary structure followed by 15–20% alpha helix. The disordered regions and the percentage of secondary structure of each plasmepsins are shown in Table 4.

To the better understanding of the evolutionary relationships plasmepsins from *P. falciparum*, a phylogenetic analysis was performed by using MEGA 5.1 (Figure 3). The evolutionary history was evaluated based on the Neighbor-Joining method by bootstrap technique. This technique allows evaluating and judging the strength of branching pattern of tree. Also, the bootstrap value given at the node of each branch determines the probability of correct or incorrect relationship such that the value above 50% represents higher confidentiality of relationship and vice versa. The phylogenetic tree is separated into four distinct cluster groups with varied values where plasmepsin I, II, IV and HAP are present in single cluster. Cluster I contains plasmepsins, which are located in the acidic food vacuole and are active during the intra-erythrocytic phase of the life cycle. The sequence, structure and function of these proteins have more than 60% identity and fall together in a single cluster. The second cluster contains the plasmepsin IX and X, where these two proteins are known to express concurrently with plasmepsin I–IV, but their sequence identity is less than 30%. The third cluster contain plasmepsin VI, VII and VIII express within the sporogonic cycle in the mosquito and are functionally different from the other cluster of plasmepsins, Plasmepsin V forms a separate cluster indicating that this protein is functionally different from the

Citation: Nair DN, Singh V, Angira D, Thiruvenkatam V (2016) Structural Investigation and *In-silico* Characterization of Plasmepsins from *Plasmodium falciparum*. J Proteomics Bioinform 9: 181-195. doi:10.4172/jpb.1000405

```
gi|Plasmepsin_1   ALFYYLIDNVLFPKNKKINEIMNTSKHVIIGFSIENSHDRIMKTVKQHRLKNYIKESLKFFKTGLTQKPHLGNAGDSVTL
gi|Plasmepsin_2   GLFYYVYENVWLQRDNEMNEILKNSEHLTIGFKVENAHDRILKTIKTHKLKNYIKESVNFLNSGLTKTNYLGSSNDNIEL
gi|Plasmepsin_4   VFFFFLIGNFYSHR--KLYQVIKNTKHTTIGFKIDRPHDKVLSSVLKNKLSTYVKESFKFFKSGYAQKGYLGSENDSIEL
gi|HAP            GFSYYIFENFVFQKNRKINHIIKTSKYSTVGFNIENSYDRLMKTIKEHKLKNYIKESVKLFNKGLTKKSYLGSEFDNVEL
gi|Plasmepsin_9   NCLNFLEKDTMFLSPSHDKETFHINHMNKIKDEKYKQEYEEEKEIYDN-TNTSQEKNETNNE-QNLNINLINNDKVTLPL
gi|Plasmepsin_10  YEQDAQKGNKNFTNNENNSDNENNSDNENNSDNENNLDNENNLDNENNSDNSSIEKNFIALENKNATVEQTKENIFLVPL
gi|Plasmepsin_6   FCVLVHICPIHTLNIFKNDENEKGSLNIPLGKENNLFFNEIKLENRFKNNIKGYIQNVQKFHYLMEKNKPNVLSYIQEDL
gi|Plasmepsin_8   FCLFVITNLYNIIAVKAFKENLRVSKYYAGGK------HKLNLENKYIGISTIVLKG------------------------
gi|Plasmepsin_7   FILLLKQHIVILKNEEFTNPYSIRKKDIKAIVNVNNKLKSINIHKLDNINKKDLLGSYN--------ENYILIKLKKQDI
gi|Plasmepsin_5   LFCFVFVSIFFVSNVTIIKCNNVENKIDNVGKKIENVGKKIG-DMENKNDNVENKNDNVGNKNDNVKNASSDLYKYKLYG
                                  :*   :  .  ::.*.*:.: .    *          .  :. . *
gi|Plasmepsin_1   NDVANVMYYGEAQIGDNKQK-FAFIFDTGSANLWVPSAQCNTIGCKTKN-LYDSNKSKTYEKDGTKVE------------
gi|Plasmepsin_2   VDFQNIMFYGDAEVGDNQQP-FTFILDTGSANLWVPSVKCTTAGCLTKH-LYDSSKSRTYEKDGTKVE------------
gi|Plasmepsin_4   DDVANLMFYGEGQIGTNKQP-FMFIFDTGSANLWVPSVNCDSIGCSTKH-LYDASASKSYEKDGTKVE------------
gi|HAP            KDLANVLSFGEAKLGDNGQK-FNFLFHTASSNVWVPSIKCTSESCESKN-HYDSSKSKTYEKDDTPVK------------
gi|Plasmepsin_9   QQLEDSQYVGYIQIGTPPQT-IRPIFDTGSTNIWIVSTKCKDETCLKVH-RYNHKLSSSFKYYEPHTN------------
gi|Plasmepsin_10  KHLRDSQFVGELLVGTPPQT-VYPIFDTGSTNVWVVTTACEEESCKKVR-RYDPNKSKTFRRSFIEKN------------
gi|Plasmepsin_6   LNFHNSQFIADIGVGNPPQV-FKVVFDTGSSNLAIPSTKCIKGGCASHK-KFNPNKSRTFTKNLKNNQESVY--------
gi|Plasmepsin_8   -GYINRQFIGEINIGNPPQT-FKVLFDTGSTNLWIPSKNCFTRACYNKR-KYDHKISKNYKLVKKKDP------------
gi|Plasmepsin_7   FSKKLSTYYGEVQIGEQSENNMNVLFDTGSSQVWILNDTCKNSLCNNIHSKYKRTKSFVYKYDKKGLP--SV--------
gi|Plasmepsin_5   DIDEYAYYFLDIDIGKPSQR-ISLILDTGSSSLSFPCNGCKDCGIHMEK-PYNLNYSKTSSILYCNKSNCPYGLKCVGNK
                                 .        : .  :          : .                      *:.*:.
gi|Plasmepsin_1   -----MNYVSGTVSGFFSKDIVTIA-NLSFPYKFIEVTDTNG------FEPAYTLGQFDGIVGLGWKD----LSIGSVDP
gi|Plasmepsin_2   -----MNYVSGTVSGFFSKDLVTVG-NLSLPYKFIEVIDTNG------FEPTYTASTFDGILGLGWKD----LSIGSVDP
gi|Plasmepsin_4   -----ISYGSGTVRGYFSKDVISLG-DLSLPYKFIEVTDADD------LEPIYSGSEFDGILGLGWKD----LSIGSIDP
gi|HAP            -----LTSKAGTISGIFSKDLVTIG-KLSVPYKFIEMTEIVG------FEPFYSESDVDGVFGLGWKD----LSIGSIDP
gi|Plasmepsin_9   ---LDIMFGTGIIQGVIGVETFKIG-PFEIKNQSFGLVKREKASDNK--SNVFERINFEGIVGLAFPE----MLSTGKST
gi|Plasmepsin_10  ---LHIVFGSGSISGSVGTDTFMLG-KHLVRNQTFGLVESESNNNKNGGDNIFDYISFEGIVGLGFPG----MLSAGNIP
gi|Plasmepsin_6   ---TYIQYGTGTSILEQSYDDVYLK-GLKIKHQCIGLAIEES-------LHPFSDLPFDGIVGLGFSDPDFRSQNKYASP
gi|Plasmepsin_8   ---VEILFGTGEIHIAYVTDDIHLG-DIKVKNQEFGLASYIS-------DDPFSDMQFDGLFGLGISD------DKKKKQ
gi|Plasmepsin_7   ---IEIFYLSGKIVAFEGYDTIYLGKKLKIPHTNISFATKVD-------IPILEEFKWDGIIGLGFQNGD--SIKRGIKP
gi|Plasmepsin_5   CEYLQSYCEGSQIYGFYFSDIVTLP--SYNNKNKISFEKLMG--CHMHEESLFLHQQATGVLGFSLTK------PNGVPT
                                  :                  : :*  .                    :  :
gi|Plasmepsin_1   VVVELKNQNKIEQAVFTFYLPFDD-KHKGYLTIGGIEDRFYEG---QLTYEKLNHDL-----YWQVDLDLHFGNLT----
gi|Plasmepsin_2   IVVELKNQNKIENALFTFYLPVHD-KHTGFLTIGGIEERFYEG---PLTYEKLNHDL-----YWQITLDAHVGNIM----
gi|Plasmepsin_4   VVVELKKQNKIDNALFTFYLPVHD-KHVGYLTIGGIESDFYEG---PLTYEKLNHDL-----YWQIDLDIHFGKYV----
gi|HAP            YIVELKTQNKIEQAVYSIYLPPEN-KNKGYLTIGGIEERFFDG---PLNYEKLNHDL-----MWQVDLDVHFGNVS----
gi|Plasmepsin_9   LYENLMSSYKLQHNEFSIYIGKDS-KYS-ALIFGGVDKNFFEG---DIYMFPVVKEY-----YWEIHFDGLYIDHQKFCC
gi|Plasmepsin_10  FFDNLLKQNPNVDPQFSFYISPYD-GKS-TLIIGGISKSFYEG---DIYMLPVLKES-----YWEVKLDELYIGKERICC
gi|Plasmepsin_6   LIETIKKQNLLKRNIFSFYVPKKL-EKSGAITFGKANKKYTVEG-KSIEWFPVISLY-----YWEINLLDIQLSHKNLFL
gi|Plasmepsin_8   LIYDSIPKNILEKNMFAIYYPKNV-DDDGAITFGGYDKKFIREN-SSIEWFDVTSSK-----YWAIQMKGLKINDVFLDV
gi|Plasmepsin_7   FLDILKDDKILTNKNYKNQFGYYLSDKEGYITLGGIDNRLKNTPDEEIIWTPVSTEMG----YWTIQIMGIRKEYVNNHF
gi|Plasmepsin_5   FVDLLFKHTPSLKPIYSICVSEHG----GELIIGGYEPDYFLSNQKEKQKMDKSDNNSSNKGNVSIKLKNNDKNDDEENN
                                                                                  .*:.:    *
gi|Plasmepsin_1   ------------------------------------------------------------VEKATAIVDSGTSSITAPT-
gi|Plasmepsin_2   ------------------------------------------------------------LEKANCIVDSGTSAITVPT-
gi|Plasmepsin_4   ------------------------------------------------------------MQKANAVVDSGTSTITAPT-
gi|HAP            ------------------------------------------------------------SKKANVILDSATSVITVPT-
gi|Plasmepsin_9   GVNSIVYDLKKKDQENNKLFFTRKYFRKNKFKTHLRKYLLKKIKHQKKQKHSNHKKKKLNKKKNYLIFDSGTSFNSVPK-
gi|Plasmepsin_10  DE----------------------------------------------------------------ESYVIFDTGTSYNTMPS-
gi|Plasmepsin_6   CES---------------------------------------------------------------KKCRAAIDTGSSLITGPS-
gi|Plasmepsin_8   CSKN--------------------------------------------------------------HEGFCQAVIDTGTSSIAGPK-
gi|Plasmepsin_7   EENK-------------------------------------------------EEEEVIVKYEAFHDGGKNSIIDTGTYLIYAPK-
gi|Plasmepsin_5   SKDVIVSN--------------------NVEDIVWQAITRKYYYYIKIYGLDLYGTNIMDKKELDMLVDSGSTFTHIPEN
                        :                                                  :  .
gi|Plasmepsin_1   --EFLNKFFEGLDVVKIP----FLPLYITTCNNPKLPTLEFRS--ATNVYTLEPEYYLQQIFDFGI---SLCMVSIIPVD
gi|Plasmepsin_2   --DFLNKMLQNLDVIKVP----FLPFYVTLCNNSKLPTFEFTS--ENGKYTLEPEYYLQHIEDVGP---GLCMLNIIGLD
gi|Plasmepsin_4   --SFLNKFFRDMNVIKVP----FLPLYVTTCDNDDLPTLEFHS--RNNKYTLEPEFYMDPLSDIDP---ALCMLYILPVD
gi|HAP            --EFFNQFVESASVFKVP----FLSLYVTTCGNTKLPTLEYRS--PNKVYTLEPKQYLEPLENIFS---ALCMLNIVPID
gi|Plasmepsin_9   --DEIEYFFRVVPSKKCD----DSNIDQVVSS---YPNLTYVI--NKMPFTLTPSQYLVRKN-------DMCKPAFMEIE
gi|Plasmepsin_10  --SQMKTFLNLIHSTACT----EQNYKDILKS---YPIIKYVF--GELIIELHPEEYMILND-------DVCMPAYMQID
gi|Plasmepsin_6   --TFIQPLLEKINLERDC----SNKESLPIIS---FVLKNVEG--KEITLDFMPEDYIIEEGDTENNT-LECVIGIMPLD
gi|Plasmepsin_8   --EDLILLSRLLNPGKFC----QKRILLKNFS---FIFIDDKR--REREYELTPKDYIVNSFRIDPVLRSPCNFAFMPIN
gi|Plasmepsin_7   --NTMENYLKDLKINNCD----EKYNLPHLIFQIKSDEIKTIKGSAIIEIVLTPNDYVIEYVDKKNN-TKECILGIQPDE
gi|Plasmepsin_5   IYNQINYYLDILCIHDMTNIYEINKRLKLTNESLNKPLVYFEDFKTALKNIIQNENLCIKIVDGVQCWKSLENLPNLYIT
                                                     :*   *::   .::          :..  :
gi|Plasmepsin_1   LNKN----------------------------TFILG-DPFMRKYFTVFDYDN----HTVGFALAKKKL-----------
gi|Plasmepsin_2   FPVP----------------------------TFILG-DPFMRKYFTVFDYDN----HSVGIALAKKNL-----------
gi|Plasmepsin_4   IDDN----------------------------TFILG-DPFMRKYFTVFDYEK----ESVGFAVAKN-------------
gi|HAP            LEKN----------------------------TFVLG-DPFMRKYFTVYDYDN----HTVGFALAKNL------------
gi|Plasmepsin_9   VSSEY-------------------------GHAYILGNATFMRYYYTVYRRGNNNNSSYVGIAKAVHTEENEKYLSSLHN
gi|Plasmepsin_10  VPSER-------------------------NHAYLLGSLSFMRNFFTVFVRGTESRPSMVGVARAKSKN-----------
gi|Plasmepsin_6   VPPPR-------------------------GPIFIFG-NSFIRKYYTIFDNDH----KLIGLIEANHNF-----------
gi|Plasmepsin_8   ISSSN-------------------------GYLYILG-QIFLQKYYAIFEKDN----MKIGLAKSI--------------
gi|Plasmepsin_7   QSEEDN-----------------------VDGWTLG-QVFLKAYYTIFDKDN----LKIGFVRSKRNVTLR--------
gi|Plasmepsin_5   LSNNYKMIWKPSSYLYKKESFWCKGLEKQVNNKPILG-LTFFKNKQVIFDLQQ----NQIAFIESKCPSNLTSSRPRTFN
```

Figure 1: The multiple sequence alignment of the representative plasmepsin sequence performed with CLUSTAL X. Active site residues are indicated by an asterisk and the conserved residues which are common to all sequences are shadowed.

Citation: Nair DN, Singh V, Angira D, Thiruvenkatam V (2016) Structural Investigation and *In-silico* Characterization of Plasmepsins from *Plasmodium falciparum*. J Proteomics Bioinform 9: 181-195. doi:10.4172/jpb.1000405

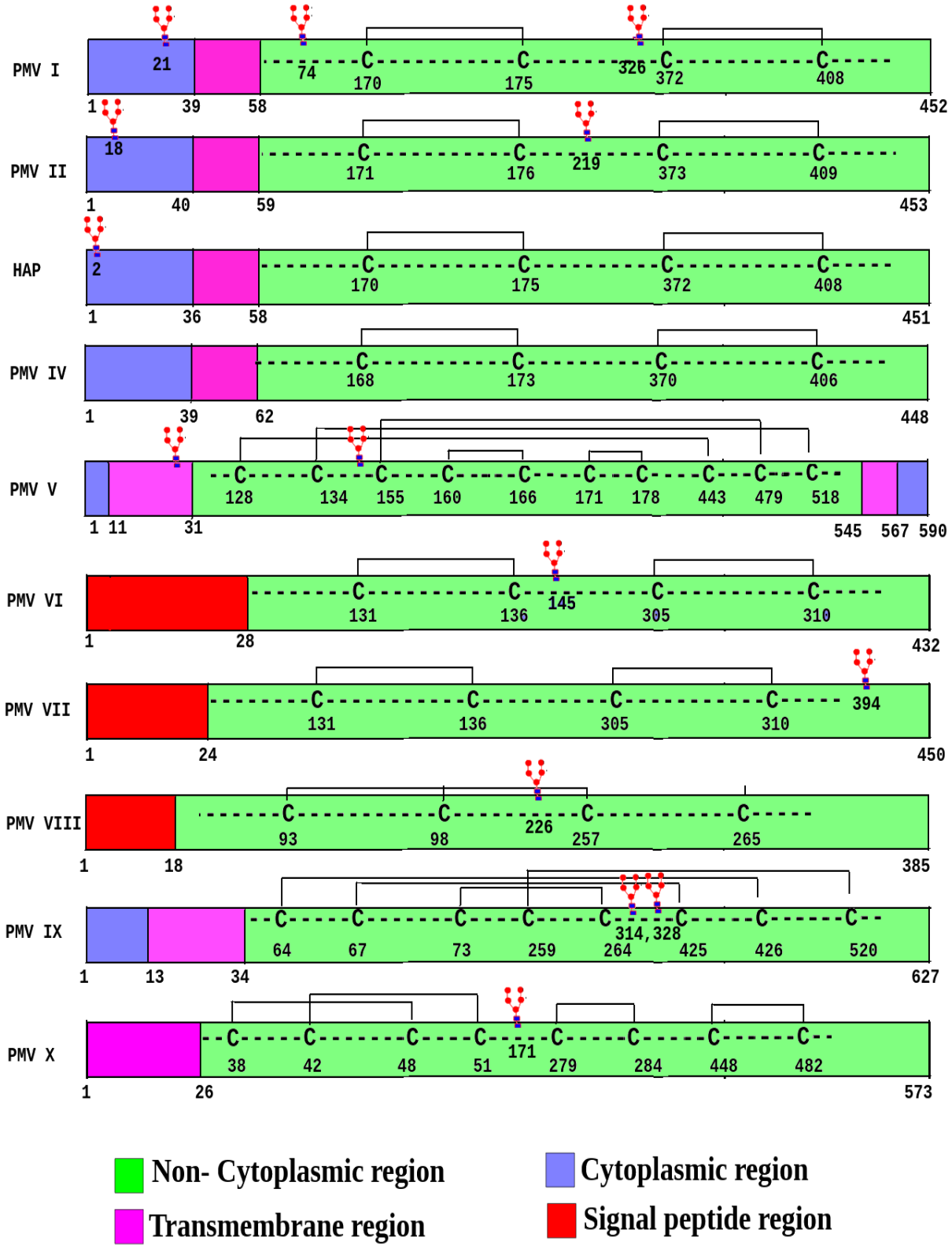


Figure 2: The sub-cellular localization, glycosylation and disulfide linkages are predicted using various tools. The sub-cellular localization (Red: Signal peptide region, Blue: Non-cytoplasmic region, Magenta: Transmembrane region and Green: Cytoplasmic region). Glycan chains are shown in the glycosylated regions and the "C" letter indicates cysteine residues and cysteine disulphide bridge are connected using lines.

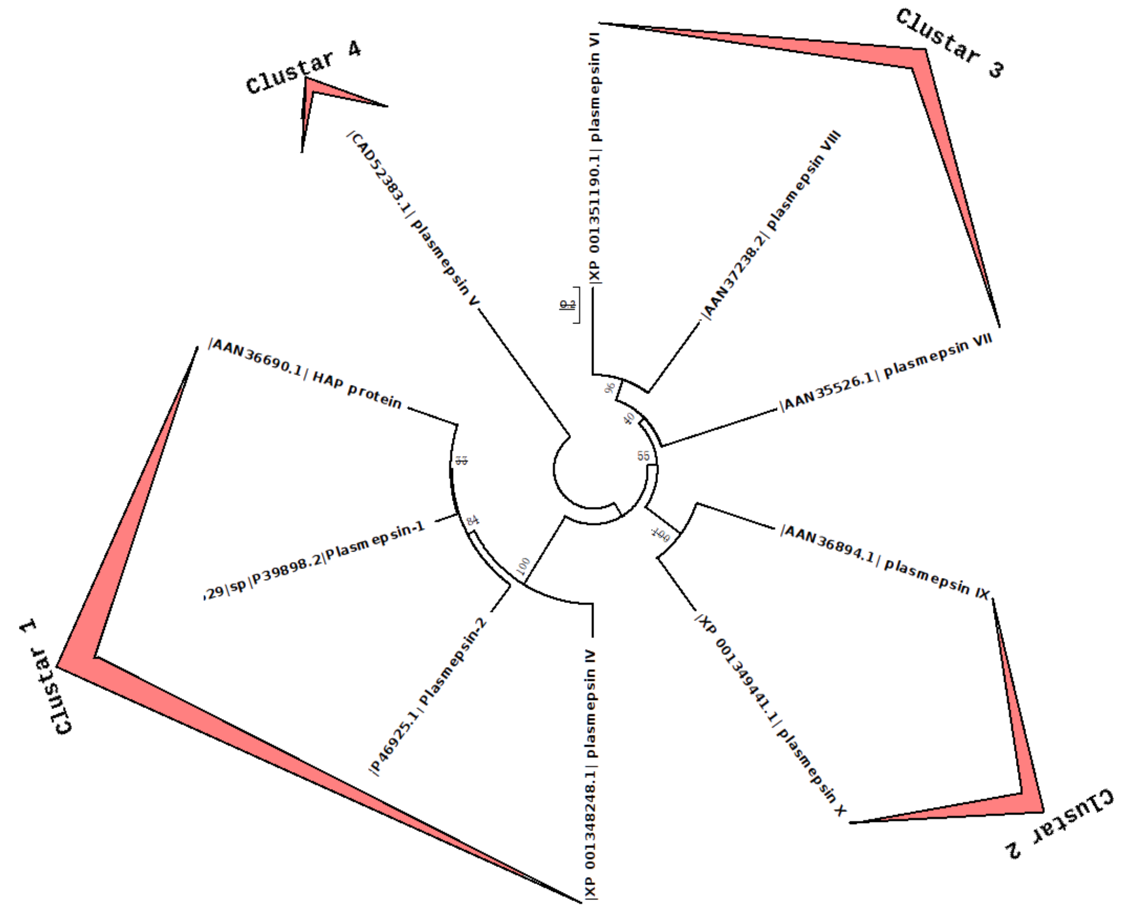


Figure 3: The phylogenetic analysis of all the ten plasmepsin from *P. falciparum* which makes four different clusters.

Protein	Unfoldability*	Region of disorder segment	Percentage of Alpha helix	Percentage of Beta strand	Percentage of TM helix
Plasmepsin I	0.193	[78]-[118], [163]-[174], [287]-[298], [300]-[309]	21%	43%	6%
Plasmepsin II	0.162	[79]-[90], [93]-[103], [124]-[128]	21%	43%	6%
HAP	0.160	[74]-[115], [163]-[190], [197]-[201], [287]-[295], [304]-[308]	20%	44%	4%
Plasmepsin IV	0.174	[78]-[102], [390]-[395]	20%	40%	4%
Plasmepsin V	0.114	[47]-[93], [269]-[336],[462]-[469], [472]-[486], [507]-[513]	14%	32%	9%
Plasmepsin VI	0.169	[42]-[78], [129]-[170], [239]-[262], [353]-[362]	18%	40%	4%
Plasmepsin VII	0.098	[38]-[73], [119]-[129], [234]-[261], [271]-[329], [385]-[418]	13%	40%	4%
Plasmepsin VIII	0.143	[77]-[117], [119]-[126], [198]-[204]	10%	45%	4%
Plasmepsin IX	-0.018	[75]-[87], [89]-[225], [89]-[225], [427]-[508], [571]-[575], [595]-[627]	10%	38%	3%
Plasmepsin X	0.037	[49]-[228], [278]-[283], [289]-[293], [303]-[312]	17%	35%	-

Table 4: Identification of unfold ability, region of disorder segment in the sequence and the secondary structure predictions of each plasmepsin shown in table showing intrinsic protein disorder regions predicted by FoldIndex and GLOBPLOT 2 and unfold ability predicted by FoldIndex. In FoldIndex positive values represent proteins likely to be folded, and negative values represent those likely to be intrinsically unfolded. The percentage of alpha helix, beta sheet and Tm helix predicted using PSIPRED.

other plasmepsins. It is an integral membrane protein present in the endoplasmic reticulum of the parasite and helps to export hundreds of proteins from the parasite to the host cell. These studies thus indicate that each cluster family is separated based upon its occurrence and localization.

For several years, the structure-based drug design of anti-malarial compounds targeting plasmepsin and its inhibitors has received much attention due to their potential biomedical use. The X-ray crystallographic structure of plasmepsin I, II, IV and HAP are well known but other plasmepsin structures are not known. The PDB structures of plasmepsin I-IV and HAP structure were downloaded from the PDB database. These four structures were superposed and highlights the structures having similar secondary structure fold when compared to that of eukaryotic aspartic proteases. The single amino-acid protein chain of these plasmepsins folded into a topologically similar way consisting of two major beta-hairpin in their domain region along with two catalytic aspartic acid residues present in the N- and C-terminal of the beta-hairpin structure. The PDB structure of plasmepsin II (1SME) with its disulfide linkage and active site residues is shown in Figure 4a.

The unknown structures of other plasmepsins (Plasmepsin V, VI, VII, VIII, IX and X) were predicted using SWISS-MODEL and PHYRE structure prediction server. Plasmepsin V was modeled based on the X-ray crystallographic structure of plasmepsin V from *Plasmodium vivax* that served as the template 4ZL4 [63]. The missing regions in the crystal structure of plasmepsin V of *Plasmodium vivax* were also modeled in the built structure of plasmepsin V from *Plasmodium falciparum* (Figure 4d). The other plasmepsins VI-X were modeled against the PDB structure of cathepsin D (4OBZ), the most closely related human aspartic protease. All the single chain modeled structures except plasmepsin V were further superimposed over the crystal structure of cathepsin D from *Homo sapiens* [64] (Figure 4c). Plasmepsin V was superposed against its homolog from *Plasmodium vivax* (Figure 4d) and we also superposed the structure of all known plasmepsins, plasmepsin I-IV, together (Figure 4b). The structure of all the superposed model showed a similar topological fold which represent that the ten plasmepsins comes into the same fold of pepsin-like superfamily.

The quality and reliability of modeled structure was checked by several structural assessment methods like Z-score, RMSD and

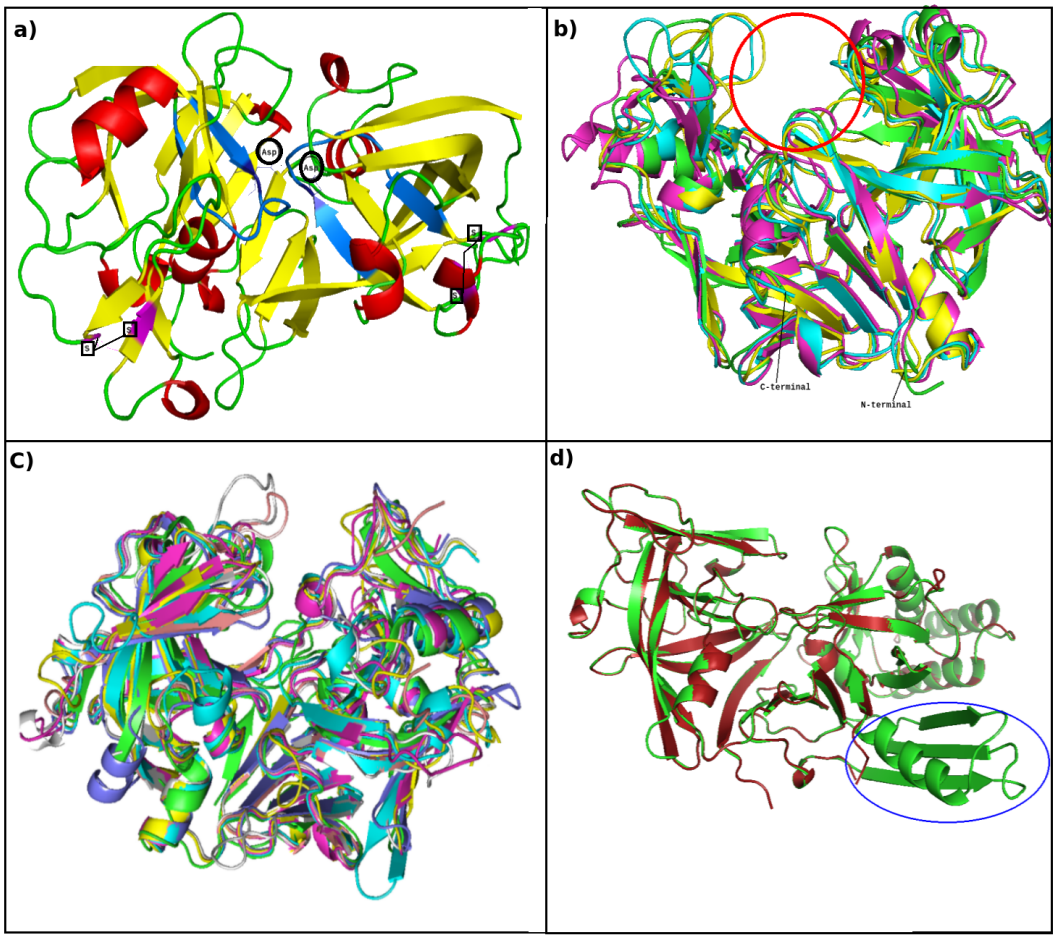


Figure 4: a) The crystal structure of plasmepsin, where the blue color represents it's active site region, active site residue ASP is marked and the magenta color represents disulfide [S] linkage region. b) The superposed structure of all the known plasmepsins where green, yellow, magenta and cyan represents plasmepsin I, II, IV and HAP respectively. The red color circle represents the cavity of the proteins. c) The superposed structure of HAP (green color) against the modeled structure were plasmepsin VI (cyan color), VII (yellow color), VIII (magenta color), IX (brown color), X (gray color). d) The superposed structure of plasmepsin V (green color) over the structure of *Plasmodium vivax* (red color) and the blue circle indicate the extra regions which not present in the structure of plasmepsin V from *Plasmodium vivax*.

Protein	Template	Sequence identity	RMSD Value	Z-score	Total number of buried outlier protein atoms	Overall quality factor	Averaged 3D-1D score ≥ 0.2	Number of residues in favored region	Nunmberof residues in allowed region	Number of residues in outlier region
Plasmepsin V	4ZL4	64%	0.4	38.0	101	75.01	80.0%	93.3%	8.2%	1.6%
Plasmepsin VI	4obz	39%	0.5	39.6	46	61.483	79%	91.94%	3.4%	3.4%
Plasmepsin VII	4obz	30%	0.5	52.1	50	60	65%	91.8%	5.6%	0.6%
Plasmepsin VIII	4obz	34%	0.8	43.3	41	55.625	70%	89.25%	5.38%	5.38%
Plasmepsin IX	4obz	35%	0.8	43.5	50	61.12	75%	91%	6.25%)	2.15%)
Plasmepsin X	4obz	35%	0.8	35.0	58	58.075	74.54%	90.1%	7.7%	3.3%

Table 5: The checking and validation of protein structures during and after the model refinement of plasmepsin V- X using various software's available in SAVES. The RMSD value of each model is predicted from DALI server, Z-score predicted using PROSA web server. The total number of outlier protein atoms are predicted from PROVE and averaged 3D-ID score calculated from VERIFY 3D using the modeled structures. The number of residues in the favored region, allowed region and outlier region is checked using Ramachandran plot.

Sl. No	Protein Name	Binding energy	Residues in the hydrophobic cavity
1	Plasmepsin I	-7.3	Phe101, Gly115, Ile120, Ile30,Met13, Thr222, Val289, Ile287
2	Plasmepsin II	-7.1	Ile123, Tyr77, Ser79, Val78, Phe244, Phe294, Ile290, Ile212, Gly213, Tyr192, Ile123
3	HAP	-6.8	Leu73, Ile80, Met140, Trp34, Leu20, Phe15, Ala10, Trp94, Ser115, Tyr112
4	Plasmepsin IV	-7.2	Gly78, Typ77, Ser76, Ile294, Val292
5	Plasmepsin V	-5.6	Tyr341, Leu363, Asp365, Val486, Asn485
6	Plasmepsin VI	-6.9	Gly120, Tyr167, Thr169, Ala318, Arg396, Val334, Leu392
7	Plasmepsin VII	-7.8	Ile205, Trp202, Ile194, Ser181, Phe85, Thr327, Thr418, Ile322
8	Plasmepsin VIII	-7.0	Ile123, Ile359, Trp67, Leu170, Tyr239, Gly272, Val258, Tyr57, Lys53, Phe125
9	Plasmepsin IX	-7.1	Ile244, Asp246, Tyr409, Asp496, Ile577, Leu220, Tyr293, Phe333, Phe291
10	Plasmepsin X	-7.1	Ile264, Phe311, Arg244, Ile354

Table 6: The ligand binding energy and its interacting residues against the structure of plasmepsins after docking studies.

Ramachandran plot. The Z-Score and the RMSD value of the modeled 3D structure were predicted using PROSA web server and Dali server and the results are tabulated in Table 6. The Dali server gives result as a pair of pre-computed structural neighbor, which has been sorted by Z-score and the Z-score lower than 2 are considered spurious. The RMSD has often been used to measure the quality of reproduction of a known (i.e., crystallographic) binding pose by a computational method and the RMSD value of each model were less than 1 indicate the best models. The Ramachandran plot provides an easy way to view the distribution of torsion angles of a protein structure [65]. It also provides an overview of allowed and disallowed regions of torsion angle values, serving as an important indicator of the quality of protein 3D structures. Based on the Ramachandran plot, almost all the plasmepsin models are having 90% residues in the allowed region, whereas plasmepsin V has 93% residues in their allowed region. This indicates the degree of correctness of modeling of the plasmepsin proteins. The models were further checked by VERIFY3D, the analysis revealed the number of amino acids have scored ≥ 0.2 in the 3D/1D profile and the result showed in Table 5. All the predicted structure validation depicted that plasmepsin V has a better modeled structure as compared to the other plasmepsin structures. This summary may be useful for biologists seeking a good crystallographic structure and aiming towards exploiting the pre modeled structures for docking analysis.

All the plasmepsins of *P. falciparum* are aspartic proteases, which constitute one of the major protease subclasses. They further distinguish each subclass based on its structural homology. Most of the aspartic proteases, including the plasmepsins and cathepsin D, are members of the pepsin family. Most of the known aspartic proteases have a well- defined subsite pocket to inhibit its substrate pepstatin. These subsites are located on both side of the catalytic site, found only in eukaryotes. On the basis of X-ray crystallography, it is well known that pepstatin binds to plasmepsin I-IV in its enzyme binding subsites, which is localized in the extended beta-strand conformation. The structures also contain a beta-hairpin turn, which helps to interact with substrate and inhibitors by covering the binding cavity. The substrate binding cavity and its interacting residues are similar in plasmepsin I- IV and HAP protein. The co-crystal structure of plasmepsin II (PDB ID: 1W6I) is of 329 amino acid residue and its catalytic binding sites contain 1 non-polar residue (Val 78), 5 polar residues (Gly 36, Asn 76, Ser 79, Tyr 192 and Ser 218) and 2 negatively charged residues (Asp 34 and Asp 214) (Figure 5a). The ligand pepstatin formed five H-bond with the catalytic residues (Figure 5b). The other amino acid residues Ser 37, Tyr77, Gly 216, Thr 217, Phe 241, Leu 242 and Ile 290 were also involved in the hydrophobic interactions with pepstatin.

The modeled structures of all the plasmepsins were superposed over plasmepsin II, which was further, used for docking studies. The binding sites of each model were predicted based on the alignment of the binding residues of plasmepsin II. An automated docking tool, Autodock Vina that works by Lamarckian Genetic Algorithm, performed molecular docking studies [66]. Their docking score between ligand and model predicts the strength and the binding activity of a binding complex. The docking score of pepstatin with each plasmepsin model were calculated (Table 6). The docking score of pepstatin with plasmepsin I-VI and plasmepsin VII-X were below a score of -7 affinity, whereas HAP and plasmepsin VI have a binding score of -6.5 and -6.6 respectively. The least binding score was observed -5.1 with plasmepsin V indicating that plasmepsin V have a low binding affinity towards pepstatin. The least binding of plasmepsin V with pepstatin as compared to other plasmepsins might be because of low volume of the cavity due to the loop located at the cleft of the binding pocket. We further designed some small molecules similar to the pepstatin and used them as a ligand for docking studies. The shape of the binding pocket of plasmepsin V (Figure 6a) was observed to complement the shape or pose of the ligand

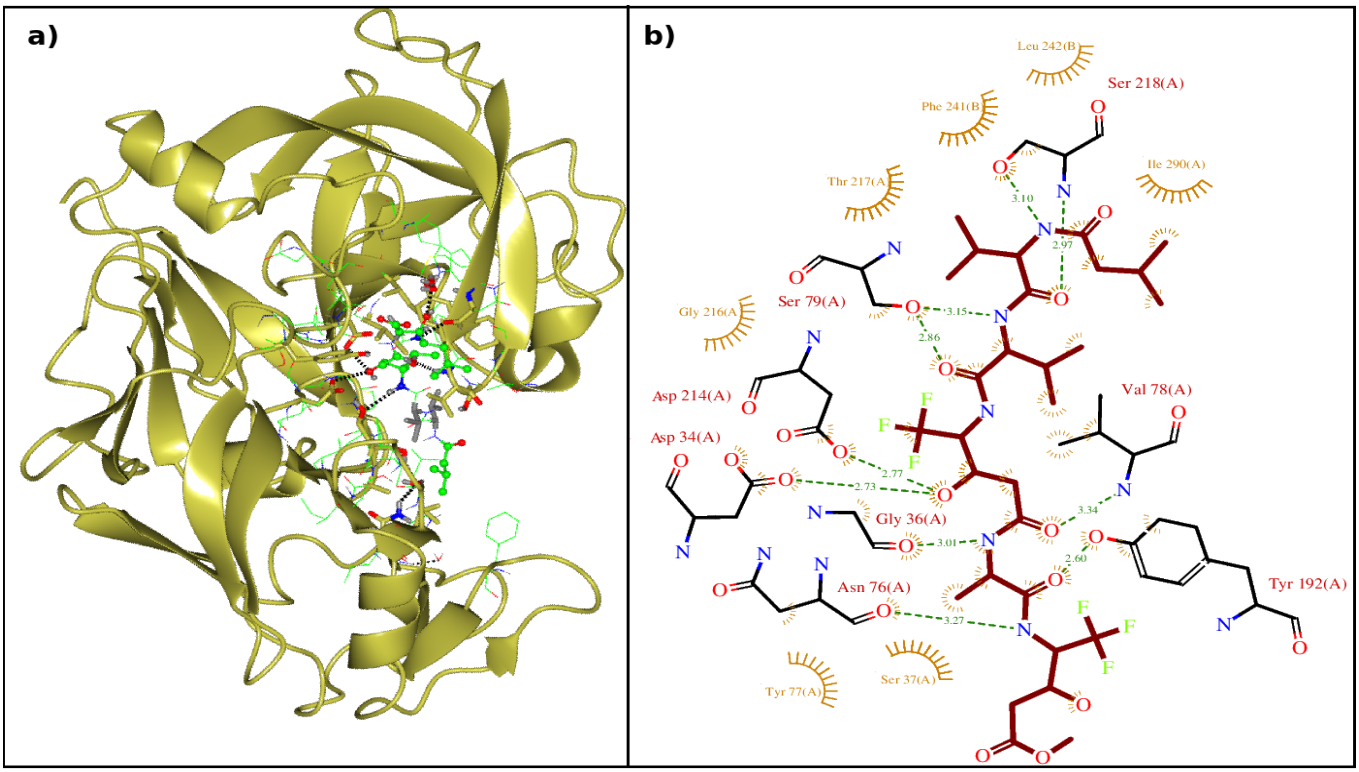


Figure 5: a) The crystallographic structure of plasmepsin II in complex with pepstatin where the blue dot represents the hydrogen bonding and the green ball and stick molecules indicate pepstatin. b) Intermolecular interactions in docked complex of plasmepsin 2 and pepstatin where green colored dotted line denotes hydrogen bond formation between the ligand and protein. (PDB ID: 1W6I).

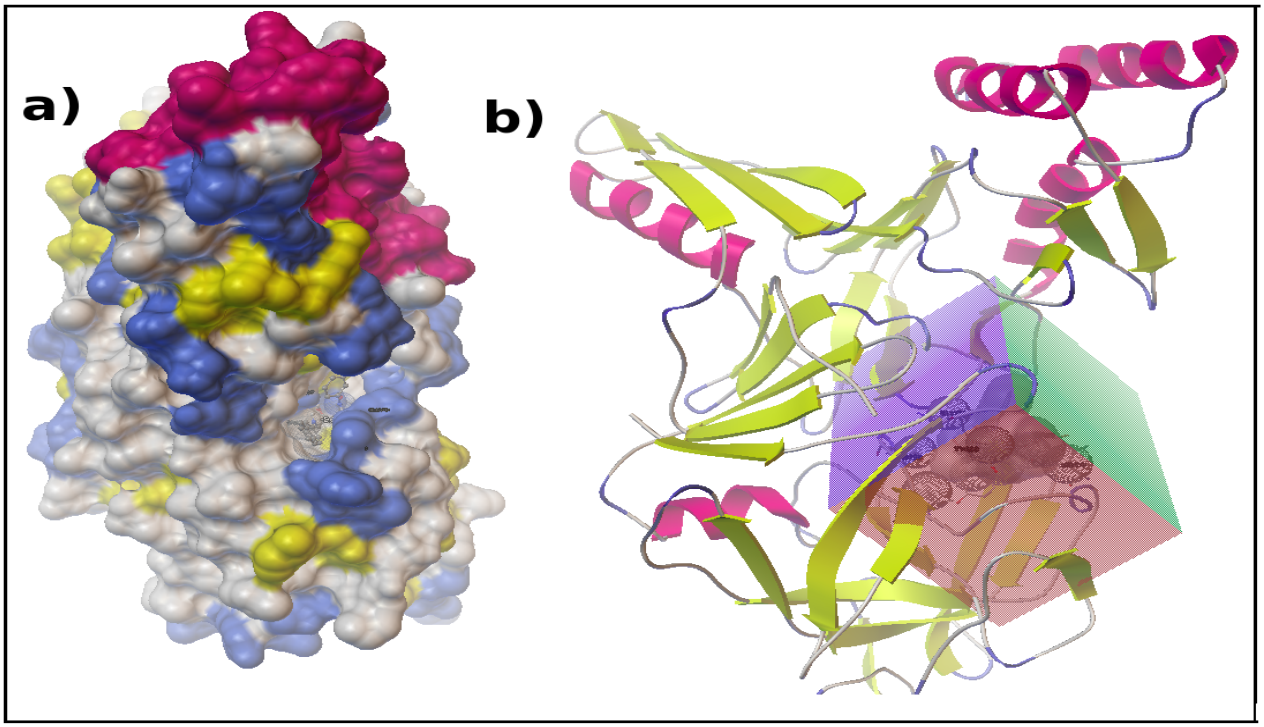


Figure 6: a) The modeled structure of plasmepsin V with its ligand M9 docked in its binding pocket. b) The grid defined for binding pocket of modeled plasmepsin V.

and the grid was made accordingly for the docking studies (Figure 6b). Compared to the peptide molecules the designed small molecules showed more binding affinity towards plasmepsin V (Table 7). Almost all the molecules except M2 produced a docking score less than -8 kcal/mol and M2 showed an affinity of -7.7 kcal/mol. The molecule M9 showed lowest binding score with two hydrogen bonding between GLY 367 and ASP 11. The other molecules M5 and M7 with model have hydrogen bond between GLY 367. The possible interacting residue of plasmepsin V model with each ligands were shown in Figure 7 and Table 7. These results might be helpful for structure based inhibitor designing for plasmepsin V. This study opens the wide area of research focusing on synthesis of these molecules for *in vivo* studies which may also lead to design new inhibitors for plasmepsin V.

The overall study of plasmepsins and the experimentally known results indicate that these proteins play an extensive role in the survival of parasite inside the host cell. Plasmepsin I–IV are situated in the acidic food vacuole and are active only during the intra-erythrocytic phase of the life cycle. Since, they are involved in the degradation of host hemoglobin; their inhibition is considered to be a good anti-malarial strategy. Plasmepsin VI-VIII is expressed within the sporogonic cycle in the mosquito, but not much research has been done on their functional part. Hence, these plasmepsins expressed in the intra-erythrocytic stage may not be suitable for designing drugs against malaria at this point of time. Similarly, plasmepsin IX and X are expressed concurrently with plasmepsin I–IV, but are not transported to the food vacuole. Their large molecular weight, disordered regions and their inefficiency to build a properly folded structure are major obstructions for designing good inhibitors against them. Plasmepsin V present in the endoplasmic reticulum of the parasite helps to export hundreds of proteins into the host cell to remodel its erythrocyte surface. The recent preliminary docking result in this article helps in better understanding of plasmepsins and also unfastens wide area for research focusing on structural characterization and design & synthesis of small molecule that are potent candidate for anti-malarial therapy with respect to Plasmepsin inhibition.

Conclusion

The present study is focused on the *in-silico* characterization and structural investigation of plasmepsins from *Plasmodium falciparum.* The domain prediction revealed that all the plasmepsins contain a eukaryotic aspartic protease domain and have conserved active site pattern in their N- and C-terminal. A physicochemical characterization was performed by computing theoretical isoelectric point (pI), molecular weight, extinction coefficient and instability index together with the prediction of disulfide linkages, motif profiles, sub-cellular localization and disordered regions using various servers. The fold of modeled structures was compared with the known three dimensional structure plasmepsins and the models were validated using Ramachandran's map, PROCHECK and WHAT IF. The docking studies of plasmepsin were done with its known inhibitor pepstatin. We also designed some small molecule inhibitors for plasmepsin V which may serve as a good foundation for designing new inhibitors for plasmepsins table 8.

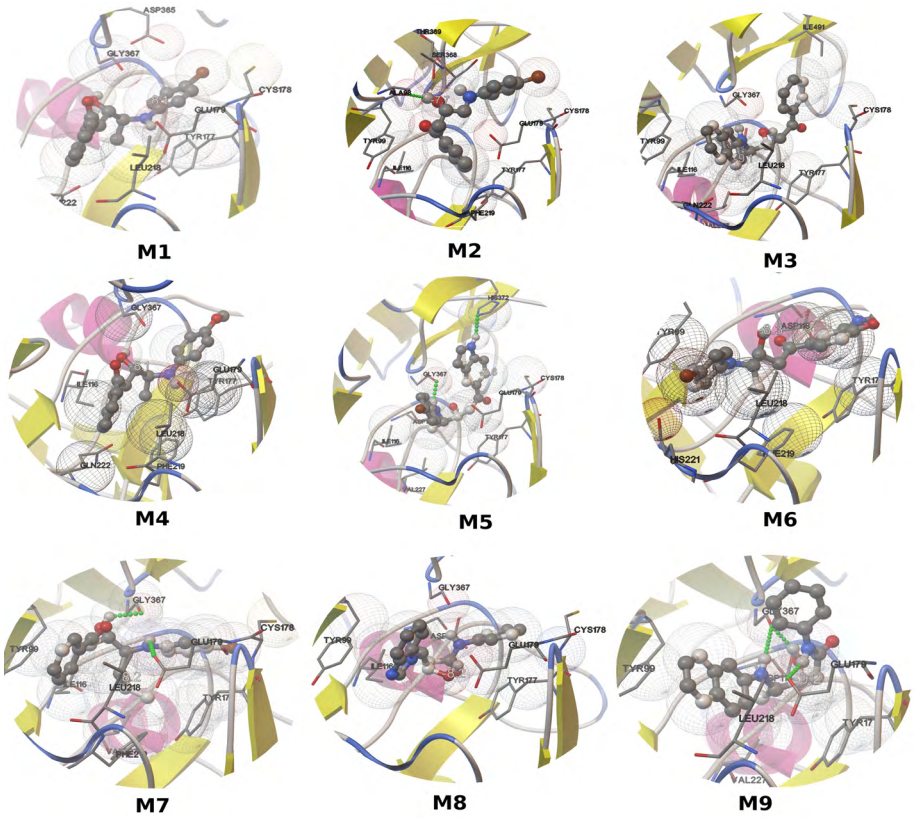
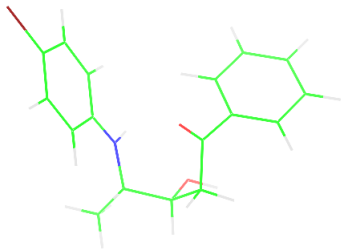
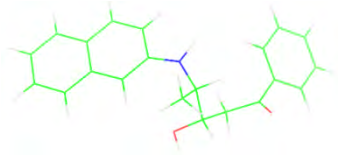
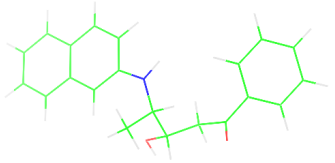
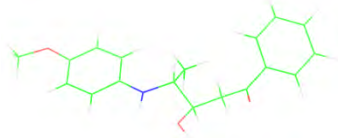
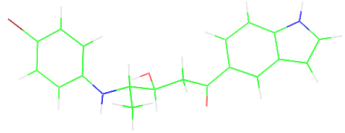
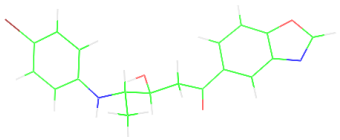
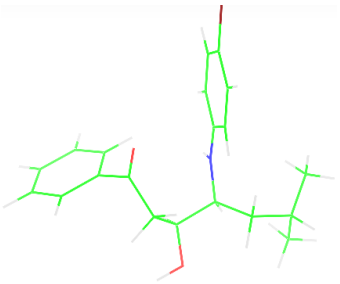
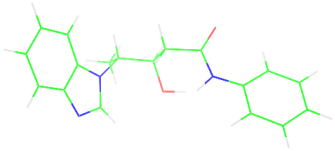


Figure 7: The best docking conformation of molecules (M1-M9) and the nearby residues of modeled plasmepsin V. The ball and stick structure represents ligand, stick represents the residues interacting with ligands and green dots represents hydrogen bonding.

Number	Molecule	Binding affinity (kcal/mol)	Dist from best mode rmsd l.b.	Dist from best mode rmsd u.b.	Interacting residues and nearby residues
M1		-8.1	0.938	5.987	ASP 365, GLY 367, CYS178, GLU 179, TYR177, LEU 218
M2		-7.7	4.057	7.547	THR369, SER 368, ALA98, TYR99, ILE 116, PHE 219, TYR177, GLU 179, CYS 178
M3		-9.1	3.713	6.837	ILE 491, GLY 397, TYR99, ILE 116, LEU218, CYS178, TYR 177
M4		-8.1	0.000	0.000	GLY367, ILE 116, GLU 179, TYR 177, LEU 218, CYS 178, TYR 177
M5		-8.6	3.402	5.198	GLY 367, HIS372, CYS 178, GLU 179, TYR 177, GLY 367, VAL 227
M6		-8.3	4.057	7.547	TYR 99, ASP118, TYR 177, LEU 218, HIS 221
M7		-8.2	2.335	5.571	TYR 99, GLY 367, CYS 178, GLU 179, TYR 177, LEU 218, ILE 116
M8		-8.2	0.000	0.000	GLY367, CYS 178, GLU 179, TYR 177, ILE 116, TYR 99

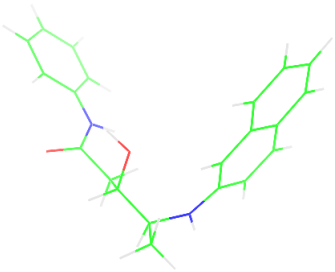
M9		-9.2	0.880	4.894	TYR 99, GLY 367, GLU 179, LEU 218, ASP 11, VAL 227

Table 7: The designed small molecule along with its binding energy towards plasmepsin V shown in table. The color code is based on the atoms presents in the molecule where green color represents carbon atom, gray color represents hydrogen atom, blue represents nitrogen and red color represents oxygen atoms.

Tools used	Website	Use
NCBI	http://www.ncbi.nlm.nih.gov/	Sequence retrieval.
PlasmoDB	http://plasmodb.org/plasmo/	Database of the *Plasmodium falciparum* genome sequencing consortium.
CLUSTAL X	Stand-alone software	It provides an integrated environment for performing multiple sequence and profile alignments and analysing the results,
SMART	http://smart.embl-heidelberg.de/	SMART used to explore domain architectures.
Pfam database	http://pfam.xfam.org/	Look at the domain organisation of a protein sequence.
PRED-CLASS server	http://athina.biol.uoa.gr/PRED-CLASS/	Classification of proteins into one of four possible classes like the membrane protein, globular protein, fibrous protein and the mixed (fibrous and globular) protein.
ProtParam	http://web.expasy.org/protparam/	It compute the various physical and chemical parameters of protein
PROSITE	http://prosite.expasy.org/	It describing protein domains, families and functional sites
HMM-TM:	http://aias.biol.uoa.gr/HMM-TM/	It predicts the transmembrane regions of alpha-helical membrane proteins, trained on crystallographically solved data.
NetNGlyc 1.0 Server	http://www.cbs.dtu.dk/services/NetNGlyc/	The *NetOglyc server* produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins.
FoldIndex	http://bioportal.weizmann.ac.il/fldbin/findex	Tool to predict whether a given protein sequence is intrinsically unfolded, which is based on the average residue hydrophobicity and net charge of the sequence.
PHYRE2 Protein Fold Recognition Server	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	Used for protein modeling.
MEGA software	http://www.megasoftware.net/ (Stand-alone software)	Used for conducting statistical analysis of molecular evolution and for constructing phylogenetic trees.
DISULFIND	http://disulfind.dsi.unifi.it/	Predicting the disulfide bonding state of cysteines and their disulfide connectivity starting from sequence alone
DiANNA web server	http://clavius.bc.edu/~clotelab/DiANNA/	Software for Cysteine state and Disulfide Bond partner prediction
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/	Used for protein secondary structural prediction
Phobius prediction	http://phobius.sbc.su.se/	A combined transmembrane topology and signal peptide predictor
GLOBPLOT 2	http://globplot.embl.de/	Exploring protein sequences for globularity and disorder
PyMOL	https://www.pymol.org/	Molecular visualization
Swiss PDB viewer	http://spdbv.vital-it.ch/	Molecular visualization
DALI server	ekhidna.biocenter.helsinki.fi/dali_server/start	Comparing protein structures in 3D
RAMPAGE server	RAMPAGE server	Used for Ramachandran Plot Analysis
AutoDock Vina	Stand-alone software	Automated docking software designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure.
PROCHECK	http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/	*It* checks the stereo chemical quality of a protein structure
PROVE software's	http://services.mbi.ucla.edu/	The volume-based structure validation procedures are implemented in the program PROVE
SuperPose	http://wishart.biology.ualberta.ca/SuperPose/	Used for the superposition of two or more structures using a modified quaternion approach
ERRAT	http://services.mbi.ucla.edu/ERRAT/	This method is for differentiating between correctly and incorrectly determined regions of protein structures based on characteristic atomic interaction is described.

Table 8: Lists out the different software's used for the protein modelling and bioinformatics web servers used for in-silico analysis.

Acknowledgments

The authors thank Department of Biotechnology (DBT) for RA fellowship to Divya Nair and IIT Gandhinagar for research facilities and infrastructure. We also thank Dr. Kirubakaran, IIT Gandhinagar for her inputs in writing this manuscript.

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.
2. Breman JG (2001) The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. Am J Trop Med Hyg 64: 1-11.
3. Tanner M, Greenwood B, Whitty CJ, Ansah EK, Price RN, et al. (2015) Malaria eradication and elimination: views on how to translate a vision into reality. BMC Med 13: 167.
4. Sachs J, Malaney P (2002) The economic and social burden of malaria. Nature 415: 680-685.
5. Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, et al. (2009) Artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med 361: 455-467.
6. Velarde-Rodríguez M, Van den Bergh R, Fergus C, Casellas A, Sanz S, et al. (2015) Origin of malaria cases: a 7-year audit of global trends in indigenous and imported cases in relation to malaria elimination. Glob Health Action 8: 29133.
7. http://www.who.int/mediacentre/factsheets/fs094/en/
8. Eastman RT, Fidock DA (2009) Artemisinin-based combination therapies: a vital tool in efforts to eliminate malaria. Nat Rev Microbiol 7: 864-874.
9. Baird JK (2005) Effectiveness of antimalarial drugs. N Engl J Med 352: 1565-1577.
10. Singh B, Daneshvar C (2013) Human infections and detection of Plasmodium knowlesi. Clin Microbiol Rev 26: 165-184.
11. Miller LH, Baruch DI, Marsh K, Doumbo OK (2002) The pathogenic basis of malaria. Nature 415: 673-679.
12. White NJ (1993) Malaria parasites go ape. Lancet 341: 793.
13. Mackintosh CL, Beeson JG, Marsh K (2004) Clinical features and pathogenesis of severe malaria. Trends Parasitol 20: 597-603.
14. Goldberg DE, Slater AF, Beavis R, Chait B, Cerami A, et al. (1991) Hemoglobin degradation in the human malaria pathogen Plasmodium falciparum: a catabolic pathway initiated by a specific aspartic protease. J Exp Med 173: 961-969.
15. Drew ME, Banerjee R, Uffman EW, Gilbertson S, Rosenthal PJ, et al. (2008) Plasmodium food vacuole plasmepsins are activated by falcipains. J Biol Chem 283: 12870-12876.
16. Coombs GH, Goldberg DE, Klemba M, Berry C, Kay J, et al. (2001) Aspartic proteases of Plasmodium falciparum and other parasitic protozoa as drug targets. Trends Parasitol 17: 532-537.
17. Banerjee R, Liu J, Beatty W, Pelosof L, Klemba M, et al. (2002) Four plasmepsins are active in the Plasmodium falciparum food vacuole, including a protease with an active-site histidine. Proc Natl Acad Sci U S A 99: 990-995.
18. Klemba M, Goldberg DE (2005) Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of Plasmodium falciparum. Mol Biochem Parasitol 143: 183-191.
19. Russo I, Babbitt S, Muralidharan V, Butler T, Oksman A, et al. (2010) Plasmepsin V licenses Plasmodium proteins for export into the host erythrocyte. Nature 463: 632-636.
20. Sleebs BE, Lopaticki S, Marapana DS, O'Neill MT, Rajasekaran P, et al. (2014) Inhibition of Plasmepsin V activity demonstrates its essential role in protein export, PfEMP1 display, and survival of malaria parasites. PLoS Biol 12: e1001897.
21. Li F, Bounkeua V, Pettersen K, Vinetz JM (2016) Plasmodium falciparum ookinete expression of plasmepsin VII and plasmepsin X. Malar J 15: 111.
22. Mastan BS, Kumari A, Gupta D, Mishra S, Kumar KA (2014) Gene disruption reveals a dispensable role for plasmepsin VII in the Plasmodium berghei life cycle. Mol Biochem Parasitol 195: 10-13.
23. Gil LA, Valiente PA, Pascutti PG, Pons T (2011) Computational perspectives into plasmepsins structure-function relationship: implications to inhibitors design. J Trop Med 6: 574- 583.
24. Marti M, Good RT, Rug M, Knuepfer E, Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 306: 1930-1933.
25. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, et al. (2009) PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res 37: D539-D543.
26. Thompson JD, Higgins DG, Gibson T (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.
27. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95: 5857-5864.
28. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, et al. (2014) The Pfam: protein families database. Nucleic Acids Res 42: 222 -230.
29. Guruprasad K, Reddy BV, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4: 155-161.
30. http://athina.biol.uoa.gr/PRED-CLASS/input.html
31. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, et al. (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform 3: 265-274.
32. http://us.expasy.org/tools/protparam.html
33. Bagos PG, Liakopoulos TD, Hamodrakas SJ (2006) Algorithms for incorporating prior topological information in HMMs: application to transmembrane proteins. BMC Bioinformatics 7: 189.
34. http://prosite.expasy.org/scanprosite/
35. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, et al. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21: 3435-3438.
36. http://www.cbs.dtu.dk/services/NetNGlyc/
37. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31: 3701-3708.
38. Ferrè F, Clote P (2005) DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res 33: W230-W232.
39. http://phobius.sbc.su.se/index.html
40. http://bioinf.cs.ucl.ac.uk/psipred/
41. Ceroni A, Passerini A, Vullo A, Frasconi P (2006) DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. Nucleic Acids Res 34: W177-W181.
42. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.
43. 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.
44. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845-858.
45. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. J App Cryst 26: 283-291.
46. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2: 1511-1519.
47. Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol 264: 121-136.
48. Maiti R, Van Domselaar GH, Zhang H, Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res 32: 590-594.
49. http://nihserver.mbi.ucla.edu/SAVES/
50. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W545-W549.
51. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, et al. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50: 437-450.

52. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35: W407-W410.

53. Hall N, Pain A, Berriman M, Churcher C, Harris B, et al. (2002) Sequence of Plasmodium falciparum chromosomes, 3-9 and 13. Nature 419: 527-531.

54. Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234: 364-368.

55. Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88: 1895-1898.

56. Klein JB, Thongboonkerd V (2004) Overview of proteomics. Contrib Nephrol 141: 1-10.

57. Zhang XC, Wang Z, Zhang X, Le MH, Sun J, et al. (2012) Evolutionary dynamics of protein domain architecture in plants. BMC Evol Biol 12: 6.

58. Fierens K, Brijs K, Courtin CM, Gebruers K, Goesaert H, et al. (2003) Molecular identification of wheat endoxylanase inhibitor TAXI-I, member of a new class of plant proteins. FEBS Lett 540: 259-263.

59. Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, et al. (2004) Structural basis for inhibition of Aspergillus niger xylanase by triticum aestivum xylanase inhibitor-I. J Biol Chem 279: 36022-36028.

60. Wu J, Wang Y, Kim ST, Kim SG, Kang KY (2013) Characterization of a newly identified rice chitinase-like protein (OsCLP) homologous to xylanase inhibitor. BMC Biotechnol 13: 4.

61. Oldfield CJ, Ulrich EL, Cheng Y, Dunker AK, Markley JL (2005) Addressing the intrinsic disorder bottleneck in structural proteomics. Proteins 59: 444-453.

62. Hodder AN, Sleebs BE, Czabotar PE, Gazdik M, Xu Y, et al. (2015) Structural basis for plasmepsin V inhibition that blocks export of malaria proteins to human erythrocytes. Nat Struct Mol Biol 22: 590-596.

63. Ersmark K, Feierberg I, Bjelic S, Hamelink E, Hackett F, et al. (2004) Potent inhibitors of the Plasmodium falciparum enzymes plasmepsin I and II devoid of cathepsin D inhibitory activity. J Med Chem 47: 110-122.

64. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95-99.

65. Trott O, Olson AJ (2010) AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31: 455-461.

66. Boddey JA, Moritz RL, Simpson RJ, Cowman AF (2009) Role of the Plasmodium export element in trafficking parasite proteins to the infected erythrocyte. Traffic 10: 285-299.