Putative circumsporozoite protein (CSP) of *Plasmodium vivax* is considerably distinct from the well-known CSP and plays a role in the protein ubiquitination pathway

Manoswini Dash, Veena Pande, Abhinav Sinha

Abstract

Amidst technical challenges which limit successful culture and genetic manipulation of *P. vivax* parasites, we used a computational approach to identify a critical target with evolutionary significance. The putative circumsporozoite protein on chromosome 13 of *P. vivax* (*PvCSP*) is distinct from the well-known vaccine candidate *PfCSP*. The aim of this study was to understand the role of *PvCSP* and its relatedness to the well-known CSP. The study revealed *PvCSP* as a membrane bound E3 ubiquitin ligase involved in ubiquitination. It has a species-specific tetra-peptide unit which is differentially repeated in various *P. vivax* strains. The *PvCSP* is different from CSP in terms of stage-specific expression and function. Since E3 ubiquitin ligases are known antimalarial drug targets targeting the proteasome pathway, *PvCSP*, with evolutionary connotation and a key role in orchestrating protein degradation in *P. vivax*, can be explored as a potential drug target.

1. Introduction

In the era, when malaria has been successfully eliminated from many countries and some are on the verge of elimination, it is still a health concern for countries of the tropical and sub-tropical regions. Global strategies and intervention policies have brought down the cases of malaria and associated mortality by 41% and 62%, respectively, in the past fifteen years leading up to 2015 (World Health Organization (WHO), 2016). Nevertheless, high number of cases (219 million) and deaths (435,000) are still estimated in the annual World Malaria Report for 2018 (World Health Organization (WHO), 2018). Malaria biology is considered to be very intricate because of the parasite's complex life-cycle where it dwells between female *Anopheles* mosquito vectors and a gamut of hosts including rodents, aves, and primates. Two apicomplexan parasites namely *Plasmodium falciparum* and *Plasmodium vivax* are the major causative agents of human malaria, accountable for the global malaria burden. *P. falciparum* is the more lethal species while *P. vivax* believed to be non-lethal, was neglected for long (Baird, 2007; Bassat and Alonso, 2011; Tham et al., 2017). However, the assumption of the benign nature of *P. vivax* has progressively changed, with increasing cases of severity and death by *P. vivax* malaria reported globally (Aashish and Manigandan, 2015; Genton et al., 2008; Geleta and Ketema, 2016; Douglas et al., 2012; Tjitra et al., 2008). Besides the ‘not-so-fatal’ assumption, lack of accurate diagnosis, dormant liver stage formation, early transmission and technical difficulties for the continuous in vitro culture have also generated hindrances in *P. vivax* research (Price et al., 2007; Baird et al., 2012). Primaquine is the only licensed drug against dormant liver stage parasites, but it is contraindicated in G6PD<sup>1</sup> deficient malaria patients due to chances of hemolysis, thereby limiting its extensive use (Cappellini and Fiorelli, 2008). Tafenoquine is an alternative single dose medicine that was approved for the radical cure of *P. vivax* malaria by US Food and Drug Administration in 2018 and might offer a better and more compliant radical treatment. Moreover, the rate at which *P. falciparum* is developing resistance against the current first-line treatment (Artemisinin Combination Therapy), it will not be incorrect to assume that the similar situation can also arise in *P. vivax* though resistance in *P. vivax* against ACT<sup>2</sup> has not been reported yet (World Health Organization (WHO), 2018). It is, therefore, crucial to devise explicit prevention, treatment and control measures for *P. vivax*, along with *P. falciparum*, to effectively accomplish the goal of malaria elimination (Lover et al., 2018).

---

<sup>1</sup>Corresponding author.

E-mail address: abhinav@mrcindia.org (A. Sinha).

<sup>2</sup>Artemisinin Combination Therapy.

https://doi.org/10.1016/j.gene.2019.100024

Received 1 July 2019; Received in revised form 21 October 2019; Accepted 7 November 2019

Available online 12 November 2019

2590-1583/ © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).
There is no efficient immunization available against malaria till date, especially in *P. vivax* with very limited numbers of candidates studied in-depth (Tham et al., 2017). In *P. falciparum*, partial success has been achieved with the development of vaccine RTS, S targeting circumsporozoite protein (CSP), a sporozoite membrane protein and was launched in 2019 (Gordon et al., 1995). However, the *P. vivax* homologue, *Pv*CSP could not induce sterile protection despite significantly delaying the infection (Bennett et al., 2016; De Camargo et al., 2018). It is therefore important to identify novel and unique candidates to be used as a critical drug or vaccine target to combat the parasite.

This study focused on the putative circumsporozoite protein (puCSP; PVX_086150), which although annotated as circumsporozoite protein, lacks the functional thrombospondin domain. The gene is present on chromosome 13 and is hypothesized to be under selection pressure based on the hitchhiking model of molecular evolution (Gupta et al., 2010; Gupta et al., 2012). Since PVX_086150 holds evolutionary significance and shares annotation with well-known CSP (PVX_119355), it is assumed that the protein might have significant role in *P. vivax* biology. However, there is no experimental evidence available to date to confirm its function. Since half of the protein coding genes in the *P. vivax* genome are unannotated, lack of knowledge about these proteins could be one of the reasons why *P. vivax* biology is still not well understood. In the absence of direct experimental evidence, computational methods are the best measures to annotate the uncharacterized genes by utilizing the knowledge from orthologs and the resources that are generated from other studies and available in public repositories. We present a comprehensive and systematic computational study to annotate the PuVpuCSP to understand its role in *P. vivax* and detangle the in-silico framework.

2. Results

2.1. PuVpuCSP have transmembrane domain and RNA as repeat unit towards C-terminal

The PuVpuCSP is a 2.2 kb single exon gene, encodes a 739 amino acid long peptide and belongs to PA-TM-RING family. The PA-TM-RING family includes transmembrane E3 ubiquitin ligases and is characterized by an N-terminal transient signal peptide, a Protease Associated (PA) domain, a transmembrane domain and a C-terminal C\(_{3}\)H\(_{2}\)C\(_{3}\) type RING H2 finger domain (cd16454). PuVpuCSP carries five pass transmembrane domain (236–258 aa, 277–299 aa, 316–333 aa, 338–355 aa, and 362–379 aa), and a C-terminal C\(_{3}\)H\(_{2}\)C\(_{3}\) type RING zinc finger domain (689–733 aa), whereas N-terminal signal peptide and PA domain was not found (Fig. 1A). The 44 residues long RING zinc finger domain is the functional domain in PuVpuCSP, with characteristic C\(_{3}\)H\(_{2}\)C\(_{3}\) arrangement (Cys-X\(_{2}\)-Cys-X\(_{4}\)-Cys-X\(_{5}\)-His-X\(_{3}\)-His-X\(_{2}\)-Cys-X\(_{10}\)-Cys-X\(_{2}\)-Cys). PuVpuCSP also possesses a transcription termination factor-Rho (420–616 aa) (cd28310) in the central region.

A scan of PuVpuCSP for repeats identified a 17 tandem repeat of R[D/C]NA (Arg-Asp/Cys-An-Ala) sequence towards the C-terminal (546–613 aa). Multiple sequence alignment of PuVpuCSP and its *Plasmodium* orthologs revealed that the repeat is very much unique to *P. vivax*, and is not present in other *Plasmodium* species except *P. knowlesi*, in which the repeating unit is present only once (Fig. 1B-i). Furthermore, comparison of the repeat region among *P. vivax* strains displays variation in the number of repeats (Fig. 1B-ii).

The topology, orientation and sub-cellular localization of a protein governs its cellular function. PuVpuCSP is located in association with plasma membrane (GO: 0005886) and contains transmembrane domains. Orientation of N- and C-terminals of PuVpuCSP with respect to the plasma membrane was predicted using topology prediction servers (Table 1). A total of 12 servers were used to predict the PuVpuCSP topology, of which 50% (6 out of 12) of servers (OCTOPUS, PHILLIUS, SPOCTOPUS, S-TMHMM, TOPCONS, and TOPPRED) predicted four transmembrane domains and both N- and C-terminals as cytoplasmic while rest of the servers predicted 5 transmembrane domains in PuVpuCSP except MEMSAT which predicted PuVpuCSP as single-pass transmembrane protein. Four prediction servers (MEMSAT, PHOBIUS, POLYPHOBUS, and SCAMPI) predicted N-terminal exposed to outside while the other two (HMMPOT and TMHMM) predicted C-terminal outside.

2.2. Orthologs of PuVpuCSP present across eukaryotes

DELTA-BLAST search against the Apicomplexan Refseq protein database using PuVpuCSP as a query retrieved 280 BLAST hits. Most of the retrieved hits were from *Plasmodium* species and are annotated as putative circumsporozoite protein or RING zinc finger protein or E3 ubiquitin ligase or conserved hypothetical protein with unknown function. The likelihood tree generated using the closest orthologs revealed *P. inui*, a simian parasite, as the closest to PuVpuCSP, followed by *P. knowlesi* among other *Plasmodium* species (Fig. 2).

The phylogenetic tree was constructed using orthologous proteins across species, sharing similar domain architecture (Supplementary Table S1) to PuVpuCSP to find out the relatedness and to ascertain the function. Orthologs of PuVpuCSP were found in other protozoans, algae, fungi, plants and higher organisms (Fig. 3). Other than *Plasmodium*, *Cryptosporidium* and *Toxoplasma* were found to be close homologs among Apicomplexans. In higher eukaryotes, the orthologs are characterized as an E3 ubiquitin ligases, while in lower eukaryotes, including the apicomplexans (around 70% of the total orthologs retrieved), majority are unannotated (Supplementary Table S1). The ortholog of PuVpuCSP is annotated as circumsporozoite protein only in *P. coatneyi*, *P. cynomolgi*, and *P. knowlesi* while in other apicomplexans, it is named as RING zinc finger protein based on its functional domain. The constructed phylogeny revealed conserved functional domains in diverse groups. However, the annotated proteins form a clade with substantial phylogenetic distance from PuVpuCSP and the proteins sharing a similar clade with PuVpuCSP are mostly hypothetical or putative.

PuVpuCSP is found to be a single copy gene from the self BLASTn search with an expectation value of \(> 1E-10\). However, PuVpuCSP also has paralogs in *P. vivax*, present on different chromosomes (PVX_111310, PVX_094410, PVX_113525, PVX_119830, and PVX_079770). From eggNOG search, a total of 6623 proteins in 237 species were retrieved as orthologs of PuVpuCSP across species, which were predicted to be involved in protein metabolic processes like protein ubiquitination and also possess metal and/or ion binding activity.

2.3. Moderate expression of PuVpuCSP across all parasitic stages

Examination of published transcriptome data from two distinct *P. vivax* isolates during the intra-erythrocytic developmental cycle (IDC) revealed moderate expression with minimum transcriptional changes of PuVpuCSP, with an average Log2 Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value of 5.5, across the blood stages (Zhu et al., 2016). However, Pv-CSP (PVX_119355) shows significant transcriptional changes though less expressed as compared to PuVpuCSP across different time points in the erythrocytic stage (Fig. 4A-i).

The transcriptome data generated by Westenberger and group revealed that PuVpuCSP (PVX_086150) was expressed with a fold change (FC) value of 2.8 and pANOVA < 0.05 (Westenberger et al., 2010). Expression of PuVpuCSP was marked in all stages (sporozoite, blood stage, gametocytes and oocyst) of parasite along with the sporozoite stage with average expression value of 116. However, the well-known CSP (PVX_119355) was expressed with the FC value of 144.14, with highest expression in the sporozoite stage (13,188.85). Comparison of the stage-specific expression of two CSPs showed that PuVpuCSP had moderate expression across all parasitic stages while the well-known CSP
Fig. 1. Schematic diagram of PpuCSP and multiple sequence alignment of PpuCSP orthologs. (A) Transmembrane domains (orange), transcription termination factor Rho (green) and RING zinc finger domain (blue) are marked according to their corresponding amino acid position in PpuCSP. The repeat region (marked with dotted diagonal lines) is predicted to lie within the transcription termination factor Rho. The above figure was prepared using DOG (Domain Graph) version 1.0 (Ren et al., 2009; Liu et al., 2015). (B-i) Multiple sequence alignment of PpuCSP and its Plasmodium orthologs showing the R[D/C]NA repeat region. (B-ii) Multiple sequence alignment of PpuCSP from different strains of P. vivax available, showing variation in the number of repeats. Only repeat region of the multiple sequence alignment has been shown in the above figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

| Transmembrane prediction servers and references | Orientation of N-terminal | Orientation of C-terminal | Number of predicted TM domains | The position of TM domains with respect to PpuCSP as position |
|-----------------------------------------------|--------------------------|--------------------------|-------------------------------|-------------------------------------------------------------|
| HMMPHOP                                      | Inside                    | Outside                   | 5                             | 239–256, 279–299, 316–333, 338–355, 362–379                   |
| MEMSAT                                       | Outside                   | Inside                    | 1                             | 237–258                                                     |
| OCTOPUS                                      | Inside                    | Inside                    | 4                             | 237–257, 274–294, 311–331, 345–375                          |
| PHILLUS                                      | Inside                    | Inside                    | 4                             | 237–257, 273–295, 314–333, 351–378                          |
| PHOBUS                                       | Outside                   | Inside                    | 5                             | 240–261, 273–294, 314–333, 340–357                          |
| POLYPHOBUS                                   | Outside                   | Inside                    | 5                             | 239–261, 273–295, 314–333, 338–357                          |
| SCAMPI                                      | Outside                   | Inside                    | 5                             | 238–258, 274–294, 314–334, 337–357                          |
| SPOCTOPUS                                   | Inside                    | Inside                    | 4                             | 237–257, 274–294, 311–331, 345–375                          |
| S-TMHHM                                      | Inside                    | Inside                    | 4                             | 237–257, 273–293, 313–334, 351–378                          |
| TMHHM                                       | Inside                    | Outside                   | 5                             | 236–258, 273–295, 316–333, 338–357                          |
| TOPCONS                                     | Inside                    | Inside                    | 4                             | 237–257, 274–294, 314–334, 351–371                          |

M. Dash, et al.  
Gene: X 4 (2019) 100024  
3
had significant expression in sporozoite stage. Change in expression was > 100 times more in the case of well-known CSP from the sporozoite to erythrocytic stage while in \textit{Pv}puCSP, change in expression was only about two-fold. Analysis of two other known stage-specific proteins (AMA1 and \textit{Pv}s25) also suggested that the \textit{Pv}puCSP is expressed almost at a similar rate in every stage and not confined to any particular parasite stage (Fig. 4A-ii; Supplementary Table S2). Analysis of hypnozoite transcriptome showed that \textit{Pv}puCSP is also getting expressed during the dormant stage with average Transcripts Per Kilobase Million (TPM) 24, following a similar pattern (Gural et al., 2018). The above array of expression of \textit{Pv}puCSP in almost all parasitic stages at a moderate rate further corroborates that it might be associated with some function crucial to the parasite across all stages. Analysis of transcriptome data from other \textit{Plasmodium} species didn’t provide any significant evidence about \textit{Pv}puCSP.

2.4. \textit{Pv}puCSP is a ubiquitin ligase, co-expressed with proteins involved in the protein ubiquitination pathway

A PPI network was developed using the DEGs\(^3\) to find out the proteins that are co-expressed with \textit{Pv}puCSP (Dharia et al., 2010). A total of 4326 genes were selected with FC ≥ 2 and p-value ≤ 0.05 from the transcriptome (Westenberger et al., 2010) (Supplementary Table S3). The PPI network was generated using the above DEGs based on the interaction data available in UniProt database. It resulted in 4091 nodes and 78,299 edges, where each node represented DEG and edges as their connected proteins (Supplementary Table S3). The data showed that \textit{Pv}puCSP has 37 degrees i.e. it interacts with 37 different proteins (Table 2). Nearly half of the proteins that showed interaction with \textit{Pv}puCSP, were ubiquitin-conjugated enzymes or ubiquitin-like proteins, which suggests that \textit{Pv}puCSP might play a role in the ubiquitination process. Other proteins showing interaction with \textit{Pv}puCSP were mostly hypothetical, while few of them were serine proteases and DNA-dependent RNA polymerases. Interestingly, more than half of the proteins interacting with \textit{Pv}puCSP were hub genes with the highest degree (Table 2). Five out of the top 10 hub genes are showing interaction with \textit{Pv}puCSP were mostly hypothetical, while few of them were serine proteases and DNA-dependent RNA polymerases. Interestingly, more than half of the proteins interacting with \textit{Pv}puCSP were hub genes with the highest degree (Table 2). Five out of the top 10 hub genes are showing interaction with \textit{Pv}puCSP, with a closeness centrality measure ranges from 0.5 to 1 in all the connected nodes. Interaction of \textit{Pv}puCSP with the hub genes showcases its importance (Fig. 4B). The Gene Ontology analysis of the above set of proteins revealed them as ubiquitin conjugative enzyme, ATP binding, and zinc ion binding and mostly shown to be involved in ubiquitin-mediated proteolysis (Supplementary Table S3).

\(^3\)Differentially expressed genes.
Fig. 3. Phylogenetic tree of PpucSP orthologs across species with similar domain architecture (Rho and RING-H2_PA-TM-RING domain-containing protein). The phylogeny was reconstructed using the maximum likelihood method. The tree was designed and color coded using a web-based tool, Interactive Tree Of Life (iTOL https://itol.embl.de/) (Letunic and Bork, 2016).

Fig. 4. Expression profiling and PPI network of PpucSP (A-i) The Log$_2$ FPKM value of PpucSP (PVX_086150) and well-known CSP (PVX_119355) was plotted at 6 time points during the intra-erythrocytic life cycle. The dots in the box plot show the exact Log$_2$ FPKM value of each protein at a different time point of erythrocytic life cycle. (A-ii) MOID normalized expression value of known stage-specific proteins (well-known CSP for sporozoite stage; AMA1 for erythrocytic stage and Pvs25 for gametocyte stage) were plotted for all parasitic stages (in human and mosquitoes). The trend lines show the exponential change in expression value of PpucSP and well-known CSP. (B) PPI network of PpucSP and interacting DEGS, where size and color represent the degree of the node (bigger size and dark color to the nodes with a higher degree). The expression data generated by Westenberger et al. was used for the above network.
Table 2
List of DEGs co-expressed with PuCSP (PVX_086150) obtained from the STRING enrichment analysis in Cytoscape, using expression data generated by Westenberger group.

| PlasmoDB ID   | UniProt ID | Description of the protein | Degree | GO term  | GO description                           |
|---------------|------------|-----------------------------|--------|----------|------------------------------------------|
| PVX_084620    | ASK0L7     | Polyubiquitin 5             | 549    | GO:PM16005087 | Polyubiquitin 5, putative               |
| PVX_094805    | ASK7E2     | DNA repair protein RAD23    | 505    | GO:00020002 | Host cell plasma membrane               |
| PVX_122475    | ASJZ99     | Ubiquitin domain containing protein | 499 | GO:0008104 | Protein localization                    |
| PVX_091315    | ASK4E9     | Ubiquitin domain containing protein | 491 | GO:0006511 | Ubiquitin-dependent catabolic process |
| PVX_092345    | ASK506     | DNA-directed RNA polymerase, beta subunit | 464 | GO:GNF0218 | Merozoite development:                |
| PVX_084365    | ASK0G6     | DNA-directed RNA polymerase III subunit | 433 | GO:0043064 | DNA-directed RNA polymerase III subunit, putative |
| PVX_115255    | ASKBR4     | K02927 large subunit ribosomal protein L40e | 370 | GO:0003735 | Structural constituent of ribosome      |
| PVX_113480    | ASKL00     | Ubiquitin-conjugating enzyme E2 | 232 | GO:0004148 | Dihydrolipoyl dehydrogenase activity   |
| PVX_123140    | ASJJZQ3    | Ubiquitin-conjugating enzyme E2 | 232 | GO:0009308 | Amine metabolic process                |
| PVX_084235    | ASK0E0     | Ubiquitin-conjugating enzyme E2 4 | 232 | GO:0015935 | Ubiquitin-conjugating enzyme E2, putative |
| PVX_097985    | ASK986     | Ubiquitin-conjugating enzyme E2 | 232 | NA        | NA                                       |
| PVX_114795    | ASKJ40     | Ubiquitin-conjugating enzyme family protein | 232 | NA        | NA                                       |
| PVX_099465    | ASK6S4     | Ubiquitin conjugating enzyme E2 | 228 | GO:0003735 | Structural constituent of ribosome      |
| PVX_099185    | ASK6L9     | Ubiquitin conjugating enzyme E2 | 225 | GO:0009308 | Amine metabolic process                |
| PVX_083175    | ASK9R3     | Ubiquitin-conjugating enzyme E2 | 216 | NA        | NA                                       |
| PVX_085805    | ASK1A0     | Ubiquitin-conjugating enzyme E2 | 213 | GO:0015935 | Ubiquitin-conjugating enzyme E2, putative |
| PVX_095280    | ASK7N6     | RING-box protein HRT1        | 170    | GO:PM16005087 | NA                                       |
| PVX_087775    | ASK5A2     | Hypothetical protein         | 158    | NA        | NA                                       |
| PVX_089055    | ASK5J0     | Hypothetical protein         | 151    | GO:0006261 | DNA-dependent DNA replication           |
| PVX_114510    | ASK2B3     | Hypothetical protein         | 147    | GO:0004232 | Ubiquitin thiolesterase activity        |
| PVX_091605    | ASK6J7     | Hypothetical protein         | 146    | GO:0045047 | Protein targeting to ER                 |
| PVX_097850    | ASK6Y6     | Ubiquitin-like protein       | 142    | GO:CCYCL9 | NA                                       |
| PVX_097985    | ASK1B3     | Guanidine nucleotide exchange factor | 141 | GO:0004540 | Ribonuclease activity                  |
| PVX_088110    | ASKAC0     | Hypothetical protein         | 140    | NA        | NA                                       |
| PVX_092370    | ASK3I1     | Hypothetical protein         | 139    | NA        | NA                                       |
| PVX_114810    | ASK1H3     | Hypothetical protein         | 139    | NA        | NA                                       |
| PVX_115310    | ASK2S5     | S-phase kinase-associated protein 1A | 124 | GO:0044085 | Cellular component biogenesis           |
| PVX_101110    | ASK8U5     | Hypothetical protein         | 114    | NA        | NA                                       |
| PVX_119410    | ASKB74     | PVIII-420P                  | 93     | GO:PM15774020 | NA                                       |
| PVX_119315    | ASKB55     | PVIII-4110.P                | 62     | NA        | NA                                       |
| PVX_092460    | ASK529     | Subtilisin-like protease 2   | 61     | GO:0031124 | mRNA 3`-end processing                  |
| PVX_097935    | ASK0B3     | Subtilisin-like serine protease | 57  | GO:PM1502632 | NA                                       |
| PVX_097920    | ASKB0      | Subtilisin-like serine protease | 53  | GO:GNF0218 | Merozoite development                   |
| PVX_117040    | ASK372     | K19989 Derlin-2/3           | 51     | GO:CCYCL9 | NA                                       |
| PVX_111100    | ASKD2X     | Hypothetical protein         | 50     | GO:CCYCL01 | NA                                       |
| PVX_098150    | ASKLG9     | Circumsporozoite protein     | 37     | GO:0009308 | Amine metabolic process                |
| PVX_094820    | ASK7E5     | Hypothetical protein         | 16     | NA        | NA                                       |
| PVX_081725    | ASKA05     | Hypothetical protein         | 8      | NA        | NA                                       |

GO, Gene Ontology; NA, not available.

2.5. Tertiary PuCSP model possesses long IDR s and characteristic cross-brace zinc finger motif

The PDB search against PuCSP to obtain template(s) for homology modeling, retrieved hits with maximum 48% identity, 5% query coverage, and the hits were only confined to the C-terminal of the protein. Therefore, the templates were not found suitable for homology modeling and thus fold recognition method was implemented to generate a tertiary model. The validation study confirmed the model generated by I-TASSER to have better quality and was considered for further analysis (Table 3). The final model after loop refinement and energy minimization had the DOPE score changed from −3899.25415 to −61,670.992188.

The tertiary model of PuCSP shows that majority of the regions were alpha helices, followed by the coil and a beta sheet. The predicted transmembrane domains fold into alpha helices, which further supports that PuCSP is an integral membrane protein (Fig. 5B). The zinc RING finger domain (689–733 amino acids) presented with its characteristic alpha helix, small beta sheet, and variable loop length, binding with zinc metal ion with C6H23C4 arrangement (Fig. 5C). Some portion of

Table 3
Model validation report of PuCSP models generated by homology modeling and fold recognition methods.

| Servers     | Parameters checked | I-TASSER        | Orion   | SPARK-X   | Modeller |
|-------------|--------------------|-----------------|---------|-----------|----------|
| Verify 3D   | Averaged 3D-1D score > 0.2 | 65.22% | 48.57% | 59.13% | 42.22%   |
| PROCHECK    | Residues in most favoured region | 62.2% | 76.9% | 76.9% | 61.9%    |
|             | Additionally allowed region | 30.3% | 16.7% | 17% | 26.8%    |
|             | Generously allowed region | 5.2% | 2.3% | 3.5% | 6.6%     |
|             | Disallowed region | 2.3% | 2.2% | 2.6% | 4.7%     |
| ERRAT       | Overall quality factor | 84.583 | 21.216 | 11.44 | 22.8532  |
| ProSA       | Z score | −0.0369 | −0.02 | −0.07 | 1.5      |
| Modeller    | DOPE score | −58,795,4224 | −43,805,5391 | −46,093,4375 | −36,789,57422 |
|             | GA-341 score | 1 | 0.041819 | 0.961289 | 0.07887  |
PvpuCSP seemed to be inadequately modeled, which can be evidenced from the PvpuCSP model where a region was showing thread-like structure. These regions might be IDRs as PvpuCSP is predicted to have 57% of IDRs by DISOPRED3. Confirmation of the presence of transmembrane domains and RING domain towards the C terminal from the tertiary structure adds to the fact that PvpuCSP might be a membrane-bound E3 ubiquitin ligase.

A detailed analysis of PvpuCSP with respect to existence of IDRs was done through IDR profiling. The per-residue disorder propensity analysis of PvpuCSP revealed enrichment of disorder-promoting residues (49.12%) followed by order-promoting and neutral residues, 36.53% and 14.61%, respectively (Supplementary Table S4). Composition profiling of PvpuCSP with a set of naturally abundant proteins (SwissProt 51) revealed that PvpuCSP is significantly enriched with order promoting residues Cys (C) and Asn (N), but depleted with Phe (F), Val (V) and Leu (L). The disorder promoting residues like Arg (R), Glu (E), Ser (S), Gly (G) and Ala (A) are enriched in PvpuCSP while Pro (P) and Gln (Q) are depleted (Fig. 6A). Depletion of amino acids with certain physico-chemical properties, like hydrophobicity, bulkiness, and propensity to form linker and beta structure, was observed while amino acids that are more frequent in coils were enriched.

Prediction of disordered regions using PONDR algorithms revealed a substantial portion of PvpuCSP as intrinsically disordered. PONDR VL-XT predicts 382 residues (51.69%) as disordered and 12 disordered regions with two longest regions of 90 (133–222) and 62 (550–612) residues at N and C-terminal, respectively. Consensus from all the IDR predictors used revealed that there are two major IDRs, one at the N-terminal (1–220) and the other towards C-terminal (420–680) (Fig. 6B).

Scanning of PvpuCSP for MoRFs within IDRs using ANCHOR tool exposed 8 MoRFs (1–51, 56–80, 108–205, 446–453, 459–471, 503–522, 599–633 and 640–655) (Fig. 6B).

The charge-hydropathy plot suggests PvpuCSP as an ordered protein as a whole (Fig. 6C). However, charge-hydropathy plot of the predicted disordered region reveals its intrinsic disordered nature. The CDF plot revealed PvpuCSP as a disordered protein (Fig. 6D).

3. Discussion

Research on P. vivax, one of the two major parasites that contribute to the global malaria burden has not received the same attention as P. falciparum. This might be due to the fatal nature of P. falciparum, which kills the host if not treated. On the contrary, as P. vivax is believed to be more ancient than P. falciparum it has probably learned to survive and grow inside the host, without being noticed by the immune system (Das, 2015). P. vivax maintains itself at low parasitemia, sometimes by forming an inactive form (hypnozoites) without killing the host. This nature of P. vivax, being hidden from the host immune system provides a survival advantage and might serve as a reservoir for malaria, thus negatively influencing the malaria elimination goal. Recent reports of severity by P. vivax from different parts of the globe have begun to change our outlook towards the infectious disease and research interest has also increasingly been focused on the hitherto neglected parasite. For the improvement of available treatment and prevention tools, a number of targets (defined and novel), need to be explored to identify suitable and critical leads. Even though many proteins have been identified and proved to have biological significance in the P. vivax life cycle from past genomic, transcriptomic and proteomic studies, most of them are hypothetical or putative. Therefore, annotating more and
more predicted genes to their biological processes and molecular functions is one of the essential pre-requisites for identification of critical targets before proceeding to expensive and tedious wet-lab experiments. Computational approaches to annotate the unknown genes using extensive real data available in public resources is a cost-effective and time-saving process.

Putative CSP was one of the genes present in a 200 kb genomic region, which was presumed to be under selection pressure (Gupta et al., 2012). On preliminary observation, it was believed that the identified putative CSP must be coding for the abundant sporozoite surface protein. However, on a closer look, it was found to be different from the well-known circumsporozoite protein in terms of its amino acid sequence and functional domain although the annotation ‘putative CSP’ was retained in the databases. Therefore, the present study was conceived to annotate the putative CSP through an integrated and systematic computational approach. It was also important to distinguish from or correlate the putative CSP and the well-known CSP in order to avoid the muddle in terms of sharing a similar name.

The only similarity observed between the well-known CSP and PvpuCSP was the presence of a central repeat region, though the repeat composition (amino acid) was different. Moreover, the tandem repeats of tetra-peptide R-[DC]-N-A were found to be exclusive to P. vivax and variation observed within the species in terms of a number of repeats make it a suitable marker to inspect variability in a population.

Although domain architecture of PvpuCSP is defined, its role in P. vivax is uncertain. The protein family with this architecture is not explored till date despite the fact that it is present in other clinically important Plasmodium species such as P. falciparum. The RING zinc finger (C3H2C3) domain present at the C-terminal of the protein, which is the functional domain in E3 ubiquitin ligase, provides insights that PvpuCSP might be involved in the ubiquitination process in P. vivax (Fig. 1). Ubiquitination is the process of targeted protein degradation through a chain reaction involving ubiquitin (Ub), enzymes to activate Ub (E1), Ub conjugating enzymes (E2) and ubiquitin ligases (E3) (Lorick et al., 1999). E3 ubiquitin ligase is a class of diverse proteins which plays an important role in protein ubiquitination by selecting appropriate protein for degradation. More specifically, it catalyzes the transfer of E2-ubiquitin complex or in some cases only modified Ub(s) to the lysine side chain of a specific protein to be degraded (Hershko, 1996; Ponts et al., 2008).

Accurate prediction of the topology of a transmembrane protein and its orientation are few of the big challenges since the experimental determination of membrane protein structure is hard to achieve with the existing crystallography method (Almeida et al., 2017). Membrane proteins are essential for signaling and small molecule transport and are also effective drug targets (Terstappen and Reggiani, 2001; Ubarretxena-Belandia and Stokes, 2010). There are several prediction servers available to assign the membrane topology. Most of them use physicochemical properties of the protein like hydropathicity of amino acids, (Kyte and Doolittle, 1982) positive-inside rule (von Heijne,
transmembrane protein from the lipid bilayer is extremely challenging (Lacapère et al., 2007). That is the reason why very less number of protein folds and tertiary structure are resolved till date although a vast number of protein sequences are available. Molecular modeling is another alternative to the protein structure determination when structure could not be determined by the above-mentioned tools. However, the lack of a suitable template with threshold homology to the protein of interest certainly restricts the prediction accuracy.

In the absence of appropriate template(s), there are many algorithms that have been proposed for accurately threading the target protein to the known folds and to obtain the best fit. The automated fold recognition servers used in this study to model PuPucSP have the basic fold recognition algorithm with added features to enhance its prediction accuracy. Since there is variation in approaches and algorithms used to predict the tertiary structure, all the models were validated to obtain the best fit model. A loop is a small and flexible form of a protein’s secondary structure which, helps to interconnect two secondary structures (alpha helices and beta sheets) and biologically has many vital functions because of its flexible nature, however it does not follow a regular and observable pattern. That is why; it is difficult to accurately model the loop region of a protein. Knowledge-based loop refinement increases the stability of the protein by changing the empirical distribution of amino acids.

The classical concept of structure-function paradigm, i.e. the activity of a protein is determined by its unique three-dimensional conformations, has been changing with the discovery of IDRs and their comparable abundance with structured proteins in nature. Nevertheless, the conformational plasticity is found to be fundamental for many crucial biological activities such as DNA binding, recognition, regulation and signaling (Iakoucheva et al., 2002; Uversky and Obradovic, 2008). Therefore, any change during the process of the protein formation and folding would eventually lead to many diseases states like neurodegeneration, cardiovascular disease, amyloidogenesis and many more (Cheng et al., 2006; Table of Contents 1, 2009; Kulkarni and Uversky, 2019; Du and Uversky, 2017). In the context of folds that proteins can acquire, they can be roughly classified into three categories: structured proteins, intrinsically disordered proteins and proteins with both structured and disordered regions. A large portion of the eukaryote proteome belongs to the third category i.e. hybrid proteins with both ordered and disordered regions (Dunker et al., 2013).

The tertiary model of PuPucSP shows that a significant portion of the protein forms coils as secondary structure. The above observation might be due to lack of a suitable template, however, presence of long IDRs couldn’t be denied as a probable reason. In PuPucSP, around 50% of the amino acids are disorder promoting, which adds first line evidence to its disordered nature. However, PuPucSP seemingly belongs to the category of hybrid proteins with a mix of well-defined domains of compact structure and disordered regions with structural flexibility. Furthermore, IDRs are mostly low complexity regions rich with different kind of repeats and lack of structural stability in the tandem repeats has been established from earlier studies. Therefore, co-localization of tetra-peptide tandem repeats and IDRs at the C-terminal of PuPucSP support of each other. It has also been reported that post-translational modifications are more frequent in IDRs (Xue et al., 2009). Therefore, the presence of glycosylation and phosphorylation sites profoundly in predicted IDRs of PuPucSP further confirms its intrinsic disordered nature (Supplementary Table S4).

Since a major portion of PuPucSP is predicted to have disordered regions, it might be correlated with the presence of zinc finger domains that might act as a DNA binding domains. It has been reported that proteins which act as chaperones for other proteins also carry the unfolded segments in order to bind with misfolded protein and RNA molecule (Tompa and Csermely, 2004). Besides, IDRs are more common in proteins that are de novo translated locally from their transcripts rather than in the proteins that are pre-translated before getting transferred (Lacapère et al., 2007). The enrichment of IDRs in
distally translated protein provides conformational flexibility to the proteins that lead to a larger surface area to interact with a diverse group of molecules (Van et al., 2014). Compositional bias of amino acid and multiple linear motifs are the common features of IDRs. Since PvpuCSP also possesses multiple stretches of a repeat of amino acid [R-D-N-A], it is suggested that it might undergo localized translation and the presence of zinc finger domain further supports the fact that it may bind with different biomolecules more specifically with the proteins to initiate ubiquitination. To summarize, both direct and indirect evidence supports the finding that PvpuCSP is a protein that carries N- and C-terminal IDRs and well defined transmembrane and RING zinc finger domains.

The PPI network built on DEGs in *P. vivax* revealed the interaction of putative CSP with other ubiquitin ligases and with DNA-directed RNA polymerase and DNA repair proteins which also suggests PvpuCSP as an integral part of ubiquitination along with DNA repair and signaling.

In conclusion, in the light of existing difficulties in maintaining long term *P. vivax* culture, the present study analyzed the PvpuCSP (PVX_086150) at every possible level starting from its primary sequence to the tertiary structure, using existing resources like orthologous sequences across species, high throughput transcriptome data, atomic coordinates of the template, etc. with necessary computational tools, to gather and correlate the information to assign putative function(s) to it. Comparison of both genes (PuCSP and PvpuCSP) revealed low sequence level similarity other than a similar repeat pattern. Moreover, presence of RING domain, which is entirely different from that present in PuCSP, leads to the conclusion that PvpuCSP is an integral membrane protein, with both terminals exposed on the cytosolic side of the parasite. However, the expression of PuCSP is not restricted to the sporozoite stage. The presence of repeat unit only in *P. vivax* and variation within *P. vivax*, make it a suitable marker to study the diversity of *P. vivax* in different population. However, the basis on which the protein was named as circumsporozoite protein in *P. vivax* and in few other *Plasmodium* species like *P. coatneyi*, *P. cynomolgi*, and *P. knowlesi* is not clear. This study proposes that PvpuCSP is a completely different protein from the well-known CSP irrespective of sharing a similar name. Since many of the ubiquitination related proteins are verified drug targets, and PvpuCSP is implicated in the ubiquitination pathway, it should be further explored as a novel antimalarial target. Taken together all the observations from this study, it can be concluded that PvpuCSP lacks direct evidence of its function at present, however, biochemical and functional studies to characterize PvpuCSP might project it as a critical target and open new paths towards vivax malaria treatment and control strategy.

4. Materials and methods

4.1. *PvpuCSP* characterisation and domain analysis

The protein sequence of PvpuCSP was retrieved from PlasmoDB (PVX_086150) and domain architecture was analyzed using multiple platforms viz. Conserved Domain Database CDD,\(^4\) SPARCLE,\(^5\) SMART\(^6\) (Letunic and Bork, 2018; Schultz et al., 1998). Proteins with similar domain architecture were searched in other organisms including apicomplexans through the CDART\(^7\) to understand its function based on homology (Supplementary Table S1) (Geer et al., 2002). Since presence of repeats in central region is a characteristic feature of well-known circumsporozoite protein (PVX_119355), PvpuCSP was scanned for tandem repeats using the Tandem Repeat Finder (Benson, 1999).

In order to confirm whether PvpuCSP is a single or multi-copy gene, BLASTn was performed against the *P. vivax* genome, with the presumption that the most similar hits with expect value [E] more than 1E−10 would be considered as single copy gene (Wu et al., 2006). Topology and orientation of PvpuCSP were predicted using transmembrane prediction servers (Table 1). Presence of signal peptide in PvpuCSP was also investigated using SignalP4.1 (Petersen et al., 2011) and signalHmmm (Burdzikiwicz et al., 2018), which is specifically designed to identify signal peptides from malaria parasite and related species.

A DELTA-BLAST\(^8\) search was performed using PvpuCSP as the query against Refseq_protein database to obtain homologous sequence carrying a similar domain as PvpuCSP in order to perform homology-based annotation (Boratyn et al., 2012). Retrieved hits with >35% identity with PvpuCSP were aligned using iterative refined method present in MAFFT (Katoh et al., 2005). The duplicate sequences were removed manually and poorly aligned regions were trimmed from the alignment using G-BLOCKS (Talavera and Castresana, 2007). The maximum likelihood phylogeny was reconstructed using the JTT matrix-based model applying 1000 bootstraps in MEGA7 and the tree with the highest log likelihood value was chosen for analysis. More information about orthologs and paralogs of PvpuCSP was also retrieved from EggNOG (Huerta-Cepas et al., 2016).

4.2. Analysis of stage-specific *P. vivax* transcriptome data

The transcriptome data provides a better idea about the role of a protein based on its level of expression and the other protein along with which it gets expressed. The stage specific transcriptome data generated from previous studies were utilized to gather information about PvpuCSP expression and to distinguish it from well-known CSP by comparing their expression at different stages of parasite (Zhu et al., 2016; Westenberger et al., 2010; Gural et al., 2018; Bozdech et al., 2003; Bozdech et al., 2008; Hoo et al., 2016). The stage-specific transcriptome data of the intra-erythrocytic developmental cycle (IDC) of two distinct *P. vivax* isolates were screened (Zhu et al., 2016; Bozdech et al., 2003; Bozdech et al., 2008; Hoo et al., 2016). The dataset included the transcriptome data during the intra-erythrocytic cycle at 9-time points (TP) covering all erythrocytic stages (early ring, late ring, trophozoite, early schizont, late schizont) at 6-hour gaps. The transcriptome data of all stages of *P. vivax* in human and mosquito (sporozoite, erythrocytic stage at 8 different time points, gametes, zygotes, and ookinete) were also scanned in depth to check the stage of expression and interaction of PvpuCSP with other proteins (Westenberger et al., 2010). The known stage-specific proteins (PvAMA1 for blood stage and Pvs25 for gametocyte stage) were chosen as a control to check the change in expression. The recently published hypnozoite transcriptome data were also checked (Gural et al., 2018). Since limited expression data is available for *P. vivax*, the transcriptome of other Plasmodium species (*P. berghei*, *P. yoelii* and *P. falciparum*) were also explored using PvpuCSP orthologs (Bozdech et al., 2003; Hoo et al., 2016).

The transcriptome data covering all stages of the parasite in human and mosquito were further analyzed employing STRING\(^9\) database V10.5 (Szklarczyk et al., 2017) and Cytoscape (Shannon et al., 2003), to identify the other genes co-expressed with PvpuCSP and infer its biological function through protein-protein interaction network analysis. From the transcriptome data, the DEGs\(^10\) with a cutoff of Fold Change value (FC) ≥2 and p value ≤ 0.05 were filtered out. A PPI\(^11\) network was built on DEGs and PuCSP against a well-known CSP (Circumsporozoite protein) and domain architecture was analyzed using multiple platforms viz. Conserved Domain Database CDD,\(^4\) SPARCLE,\(^5\) SMART\(^6\) (Letunic and Bork, 2018; Schultz et al., 1998).

---

\(^4\) Conserved Domain Database.
\(^5\) Subfamily Protein Architecture Labelling Engine.
\(^6\) Simple Modular Architecture Research Tool.
\(^7\) Conserved Domain Architecture Retrieval Tool.
\(^8\) Domain Enhanced Lookup Time Accelerated BLAST.
\(^9\) Search tool for the retrieval of interacting genes.
\(^10\) Differentially Expressed Genes.
\(^11\) Protein-protein interaction.
network of the DEGs was built using Cytoscape, and functional enrichment was performed by mapping the DEGs to STRING app in-built in Cytoscape (Szklarczyk et al., 2017; Shannon et al., 2003) to obtain the degree and other centrality measures for each node.

In network analysis, each gene is referred as a node and its interaction with other proteins represented as edges. The nodes with the highest degrees (number of interacting proteins) were considered to be biologically significant and called as a hub gene (Barabási and Oltvai, 2004). The proteins with which PuPucSP showed interaction were clustered using MCODE plug-in (Sun et al., 2017). To get a more comprehensive idea about the molecular function of clustered proteins, comprehensive idea about the molecular function of PuPucSP, Gene Ontology enrichment analysis was performed using the proteins showing interaction with PuPucSP. All the genes in the network were uploaded to DAVID v6.8 online tool to perform gene enrichment analysis (Huang et al., 2009a; Huang et al., 2009b).

4.3. Molecular modeling and binding site prediction

PuPucSP was searched against PDB12 to screen out the suitable templates for homology modeling using DELTA-BLAST (Boratyn et al., 2012). Since the PDB hits obtained against the PuPucSP had significantly low sequence similarity and insufficient query coverage, PuPucSP was modeled via fold recognition method using automated modeling servers like Orion (Ghousam et al., 2016), SPARK-X (Yang et al., 2011) and I-TASSER server (Yang et al., 2014). Besides, the template information obtained from the fold recognition servers were selected as a template (PDB ID: SALU, 5WTJ, 2M6M, and 5XJY) based on its score and similarity with PuPucSP and the structure was predicted using the multi-template method in the modeler (Šali and Blundell, 1993). The functional domain [RING domain] of PuPucSP was modeled independently via homology modeling using the best match templates (PDB ID-2L0B and 2ECT) along with zinc ion using Modeller9.19 (Webb and Sali, 2017). The quality of the model was assessed based on DOPE13 score (Marchler-Bauer et al., 2013; Shen and Sali, 2006) and GA341 score (Melo et al., 2009). The model with the lowest DOPE score and GA341 score close to 1 considered to be a best-fit model for further analysis.

The best (with highest C score and Z score) out of all the models generated by I-TASSER, Orion, SPARK-X and Modeller were analyzed using the validation servers viz. using PROCHECK (Laskowski et al., 1993), ERRAT (Colovos and Yeates, 1993; MacArthur et al., 1994), Verify-3D (Eisenberg, 1997) and PROVE (Pontius et al., 1996) available in SAVES meta-server (Table 3), and the one with a good score and qualified the validation process was chosen for loop refinement (Mod-loop server) and energy minimization was done using YASARA energy minimization server (Krieger et al., 2009). PuPucSP was also scanned for IDRs14 using DISOPRED3 (Jones and Ward, 2003; Jones and Cozzetto, 2015).

4.4. IDR profiling of PuPucSP

Composition profiling analysis of PuPucSP with respect to the propensity and number of order-promoting (N, C, I, L, F, W, Y and V) and disorder-promoting (A, S, R, G, E, P, E and K) residues was carried out using Composition Profiler tool, which detects enriched and depleted amino acids in the query protein based on their many physico-chemical and structural properties (Vacek et al., 2007). A group of proteins with specific attributes that provides background amino acid distribution was compared against the query protein. Here, SwissProt 51 was used as a reference set because the database closely resembles the distribution of proteins in nature. The IDR profiling was done with 10,000 bootstrap iteration and Bonferroni correction to enhance the accuracy and reduce the error probability. The disordered regions were also predicted using PONDR-FIT, a variant of PONDR algorithm, and a meta-predictor that incorporates the output of other individual disorder predictors using artificial neural network (ANN) method (Xue et al., 2010). PONDR VL-XT, a variant of PONDR algorithm with a higher sensitivity to locate disordered stretches within a protein, was also utilized. ANCHOR algorithm was used to find out potential Molecular Recognition Features or MoRFs, which are small active sites present within the IDRs and play an important role to bind and interact with other substrates, in PuPucSP (Dosztányi et al., 2009). Stretches with < 10 residues long were filtered out and residues with ANCHOR score > 0.5 were considered to be MoRFs.

Charge-hydropathy (CH) plot and cumulative distribution fraction plot (CDF) are the two other binary classification measures followed to confirm the structural conformation of PuPucSP as a whole. Binary classification of a protein is to predict the nature of a protein as a whole by analyzing basic properties like charge, hydropathy, bulkiness, surface propensity of the amino acids. Relative high net charge with low mean hydropathy is the characteristic feature of a disordered protein or regions. In CH plot, the mean net charge and hydropathy value of a group of ordered and disordered proteins are used to generate a linear boundary line. Presence of query protein towards left and right of this boundary line predicts the query protein as ordered or disordered protein, respectively. In CDF plot, based on the distribution of a disorder score generated by VL-XT for each amino acid, the ordered state of protein is predicted. The proteins located towards the lower-right half of the CDF plot are assumed to be disordered whereas those located towards the upper-left half are compact proteins (Xue et al., 2009).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgene.2019.100024.

Funding

This work was supported by the Indian Council of Medical Research (ICMR-https://www.icmr.nic.in/), New Delhi, India in the form of Senior Research Fellowships to MD. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Author contributions

MD and AS designed the study. MD collected, analyzed and interpreted the data. MD and AS jointly wrote the manuscript. VP edited and reviewed the manuscript. AS conceptualize the idea and supervise the work. All authors read and approved the final manuscript.

Data availability statement

All data generated or analyzed during this study are included in this published article and its supplementary information files. The transcriptome data used in the study was taken from http://carrier.gnf.org/publications/Pv/ (Westenberger et al., 2010).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr. Neena Valecha, Regional Advisor-Malaria WHO-SEARO, for providing necessary research facilities and moral support during the study. We thank Dr. Sanju Kumari (ICMR PostDoctoral
Fellow) and Ms. Aparna Tiwari (CSIR Senior Research Fellow) for critically reviewing the manuscript. Thanks are also owing to all the lab members for their helpful comments and suggestions while preparing the manuscript. We extend our special thanks to Dr. Anubhav Srivastava (Research Fellow, D4 Research Team, Monash University, Australia) for his extensive review for language editing of the manuscript.

References

Aasih, A., Manigandan, G., 2015. Complicated vivax malaria, an often underestimated challenge. Trends Parasitol. 31 (11), 533–539. https://doi.org/10.1016/j.pt.2015.08.011.

Aashish, A., Manigandan, G., 2015. Complicated vivax malaria, an often underestimated challenge. Trends Parasitol. 31 (11), 533–539. https://doi.org/10.1016/j.pt.2015.08.011.

Lee, H.J., Georgiadou, A., Otto, T.D., et al., 2018. Transcriptomic studies of malaria: a comprehensive RNA-sequencing database. Nucleic Acids Res. 46 (D1), D493–D496. https://doi.org/10.1093/nar/gkx922.

Kulkarni, P., Uversky, V.N., 2019. Intrinsically Disordered Proteins in Chronic Diseases. pp. 1619–1623. https://doi.org/10.1101/gr.278202.

Baird, J.K., 2007. Neglect of Plasmodium vivax malaria. Trends Parasitol. 23 (11), 523–539. https://doi.org/10.1016/j.pt.2007.08.011.

Baird, K.J., Maguire, J.D., Price, R.N., 2012. Diagnosis and Treatment of Plasmodium vivax malaria. vol 80 Elsevierhttps://doi.org/10.1097/978-1-85979-100-9.

Barabási, Á.L., Oltvai, Z.N., 2004. Network biology: understanding the cell’s functional organization. Nat. Rev. Genet. 5 (2), 101–113. https://doi.org/10.1038/nrg1311.

Bassat, Q., Alonso, P.L., 2011. Defying malaria: fathoming severe Plasmodium vivax infection. Cell Host Microbe 23 (3), 395–406.e4. https://doi.org/10.1016/j.chom.2018.01.002.

Bassat, Q., Alonso, P.L., 2011. Defying malaria: fathoming severe Plasmodium vivax infection. Cell Host Microbe 23 (3), 395–406.e4. https://doi.org/10.1016/j.chom.2018.01.002.

Bozdech, Z., Lins, M.P., Ratner, B.J., Wong, E.D., Zhu, J., Delisi, J.L., 2003. The transcriptome of Plasmodium falciparum. PLoS Biol. 1 (1), 85–100. https://doi.org/10.1371/journal.pbio.0000005.

Bozdech, Z., Mok, S., Hu, G., et al., 2017. Membrane proteins structures: a review on computational modeling tools. Biochim. Biophys. Acta Biomembr. 1859 (10), 2021–2039. https://doi.org/10.1016/j.bbamem.2017.05.007.

Anfinsen, 1973. Principles that govern the folding of protein chains. Science (80- ) 181 (4093), 2021–2039. https://doi.org/10.1016/j.bbamem.2017.05.007.

Almeida-Rodríguez, G.J., Preto, A.J., Koolen, P.I., Bonvin, A.M.J.J., Moreira, I.S., 2017. Membrane proteins structures: a review on computational modeling tools. Biochim. Biophys. Acta Biomembr. 1859 (10), 2021–2039. https://doi.org/10.1016/j.bbamem.2017.05.007.

Dent, R.C., Almassy, C., Cutler, D., et al., 2015. Transcriptomic profiles of malaria parasites. Mol. Biochem. Parasitol. 200 (2), 173–174. https://doi.org/10.1016/j.molbiopar.2015.08.009.

Douglas, N., Kietzmann, A., Nicolai, J., 2007. Characterising the metapopulation dynamics of malaria parasites in the western Amazon region. Malar. J. 6 (1), 88. https://doi.org/10.1186/1475-2875-6-88.

Lee, P., Khiem, T., 2018. In silico identification of drug targets of Plasmodium vivax. J. Biosci. 43 (2), 245–258. https://doi.org/10.1007/s12038-018-0491-5.

Ghosh, A., Subramanian, S., 2011. Jennifer: a novel gene expression classifier for protein-protein interaction networks. Nucleic Acids Res. 39 (5), 1816–1824. https://doi.org/10.1093/nar/gkr1047.

Ghosh, A., Subramanian, S., 2011. Jennifer: a novel gene expression classifier for protein-protein interaction networks. Nucleic Acids Res. 39 (5), 1816–1824. https://doi.org/10.1093/nar/gkr1047.
