# Structural analysis of chorismate synthase from Plasmodium falciparum: A novel target for antimalaria drug discovery

**ARTICLE** *in* INTERNATIONAL JOURNAL OF BIOLOGICAL MACROMOLECULES · JULY 2011

Impact Factor: 2.86 · DOI: 10.1016/j.ijbiomac.2011.07.011 · Source: PubMed



**5 AUTHORS**, INCLUDING:

Satya Tapas
Centre for Cellular and Molecular Platforms
**9** PUBLICATIONS **29** CITATIONS
SEE PROFILE

Abhinav Kumar
University of Zurich
**1** PUBLICATION **11** CITATIONS
SEE PROFILE

Sonali Dhindwal
Virginia Commonwealth University
**11** PUBLICATIONS **42** CITATIONS
SEE PROFILE

Preeti Preeti
Institute for the Biology and Chemistry of P…
**7** PUBLICATIONS **22** CITATIONS
SEE PROFILE

International Journal of Biological Macromolecules 49 (2011) 767–777

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

# Structural analysis of chorismate synthase from *Plasmodium falciparum*: A novel target for antimalaria drug discovery

Satya Tapas [1], Abhinav Kumar [1], Sonali Dhindwal, Preeti, Pravindra Kumar*

*Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India*

ARTICLE INFO

*Article history:*
Received 10 May 2011
Received in revised form 12 July 2011
Accepted 13 July 2011
Available online 22 July 2011

*Keywords:*
Chorismate synthase
Shikimate pathway
*Plasmodium falciparum*
Homology modeling

ABSTRACT

The shikimate pathway in *Plasmodium falciparum* provides several targets for designing novel antiparasitic agents for the treatment of malaria. Chorismate synthase (CS) is a key enzyme in the shikimate pathway which catalyzes the seventh and final step of the pathway. *P. falciparum* chorismate synthase (*Pf*CS) is unique in terms of enzymatic behavior, cellular localization and in having two additional amino acid inserts compared to any other CS. The structure of *Pf*CS along with cofactor FMN was predicted by homology modeling using crystal structure of *Helicobacter pylori* chorismate synthase (*Hp*CS). The quality of the model was validated using structure analysis servers and molecular dynamics. Dimeric form of *Pf*CS was generated and the FMN binding mechanism involving movement of loop near active site has been proposed. Active site pocket has been identified and substrate 5-enolpyruvylshikimate 3-phosphate (EPSP) along with screened potent inhibitors has been docked. The study resulted in identification of putative inhibitors of *Pf*CS with binding efficiency in nanomolar range. The selected putative inhibitors could lead to the development of anti-malarial drugs.



## 1. Introduction

The phylum apicomplexa comprises of nearly 5000 species, most of which are parasites and are the causative agents of various pernicious diseases [1]. The most prominent apicomplexa is the *Plasmodium* species, which is the causative agent of malaria, a disease that is affecting nearly 300–500 million people every year worldwide. With the resurrection of drug resistant *Plasmodium falciparum*, the most fatal human malaria parasite, identification of new possible medicinal targets is an extreme priority. In such scenario, *P. falciparum* genome has facilitated the description of several metabolic pathways, particularly those differing from humans; thus providing new targets for drug development [2,3].

The shikimate pathway, in apicomplexan parasites provides a promising and exciting opportunity for reaching such objectives, as the pathway is only found in algae, plants, bacteria, protozoa and fungi, but is absent from mammals. The pathway is responsible for the formation of key aromatic amino acids involved in primary metabolism. In addition, the pathway in apicomplexa is important for the supply of folates, for which animals rely exclusively on an exogenous source [4]. Moreover, some of the enzymes of the pathway catalyze biochemically unique reactions in nature, making them excellent targets for new antiparasite drugs. Enzymes of this metabolic pathway have been studied extensively by various authors [5–7]. Chorismate synthase (EC 4.2.3.5) is the last enzyme of the pathway which catalyzes the conversion of the 5-enolpyruvylshikimate-3-phosphate (EPSP) to chorismate. The CS reaction comprises an *anti*-1,4-elimination of the 3-phosphate group and the C-6 *pro* R hydrogen with requirement of reduced FMN as a cofactor [8,9]. The catalysis action of CS does not involve any overall change in redox state, hence it is considered to be unique reaction in nature [9]. According to Bornemann et al., the reduced FMN donates an electron to EPSP to facilitate the loss of the phosphate and receive it back after the reaction [8].

According to the functionality, CS has been divided into two classes, monofunctional and bifunctional. Chorismate synthases from plants and eubacteria possess only *trans* elimination activity of substrate and are called monofunctional. However, CS from *Neurospora crassa* and *Saccharomyces cerevisiae* have an additional NADPH:FMN oxidoreductase activity, so-called bifunctional. Despite of resemblance with fungal CS, *Pf*CS is reported to be monofunctional [10].

In the present study, we performed *in silico* molecular modeling of three-dimensional structure of chorismate synthase enzyme from *P. falciparum*. We also performed an *in silico* structure-based inhibitors study using various substrate analogs and *in vitro* characterized inhibitors series. In absence of the crystal structure, we

* Corresponding author. Tel.: +91 1332 286312; fax: +91 1332 273560.
*E-mail addresses:* satyadbt@iitr.ernet.in (S. Tapas), abhi2pbt@iitr.ernet.in (A. Kumar), sonaldbt@iitr.ernet.in (S. Dhindwal), prityDBS@iitr.ernet.in ( Preeti), pravshai@gmail.com, kumarfbs@iitr.ernet.in (P. Kumar).
[1] These authors contributed equally to this work.

0141-8130/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijbiomac.2011.07.011

hope that the proposed 3D model will be helpful for providing novel target for structure based drug design against malaria.

## 2. Methodology

The computational analysis was done on Intel Core i3-2.13 GHz Processor running on Windows 2007 Home Basic. The AUTODOCK version 4 and Molecular Dynamics (MD) simulations (GROMACS) were performed on Red Hat Enterprise Linux 5 operation system (Red Hat Inc., Raleigh, NC) installed on a Dell Precision T5400 workstation [11,12]. The Molecular Docking was performed on GLIDE 5.5 program (Glide, version 5.5, Schrödinger, LLC, New York, NY, 2009) running on Windows 2003 on a HP xw8400 Workstation. All the graphical analysis and image production was done using PyMOL [13] and WinCoot software [14].

### 2.1. Multiple sequence alignment and phylogenetic tree construction

The CS sequences of *Plasmodium* species along with representative of bacteria, algae, fungi and plant were retrieved from NCBI database. A multiple sequence alignment for these sequences was generated using ClustalW with default parameters [15]. The graphical enhancement of the aligned sequences was performed using ESPript 2.2 server [16]. The phylogenetic tree was inferred using Phylogeny FR server [17].

### 2.2. Comparative molecular modeling, model optimization, evaluation and dimerization

Suitable template for modeling of *Pf*CS was searched using a PSI-BLAST against PDB database with default parameters [18]. From the best hits, crystal structure of chorismate synthase from *Helicobacter pylori* in complex with FMN (PDB 1UM0) was selected for model building. ClustalW program was used for multiple sequence alignment of query sequence with template sequence. Some manual corrections were done (as discussed in Section 3.2) in the alignment file for additional residues in *Pf*CS amino acid sequence. The sequence identity of modified *Pf*CS with the template was nearly 34%. Based on sequence alignment analysis, it was assumed that cofactor binding mode of *Pf*CS is similar to that of *Hp*CS. Therefore, FMN from the template was also incorporated in the modeled structure of *Pf*CS during model generation.

Program Modeller9v8 was employed to generate ten models of *Pf*CS [19]. Three sets of model having lowest DOPE scores were selected and stereo-chemical quality of each was evaluated by PROCHECK [20]. The model with least number of residues in the disallowed region was further refined for relieving steric clashes and improper contacts by energy minimization using Swiss PDB viewer 4.0.1 (http://spdbv.vital-it.ch/). Swiss PDB viewer implements GROMOS96 force field to compute energy and execute energy minimization. PROCHECK and ERRAT plot were further used to evaluate stereo-chemical quality of the model [21]. Loop refinement tool of the Modeller was used in an iterative fashion to refine the loop conformation of the model. Structural validation after each loop refinement step was done using ERRAT plot that gives a measure of structural error at each residue in the protein. This process was repeated iteratively until most of the amino acid residues were below 95% cutoff value in ERRAT plot. The refined model was further validated by VERIFY 3D of SAVES server (http://nihserver.mbi.ucla.edu/SAVES/) [22]. ProSA-Web server (https://prosa.services.came.sbg.ac.at/prosa.php) was also used to evaluate the generated 3D model of protein for potential error [23]. The dimer of *Pf*CS was generated by Modeller9v8 using the dimeric state of *Hp*CS as template. Furthermore, the ClusPro 2.0 Protein-Protein docking server (http://cluspro.bu.edu/login.php?redir=/) was used to generate *Pf*CS dimer to compare with dimeric model generated by Modeller9v8 [24].

GROMACS simulation suite version v.4.0.7 was used to perform Molecular Dynamics of predicted *Pf*CS model [12]. The energy minimization was performed by giving 1000 steps of Steepest Descent to eliminate bad atomic contacts, which converged in less than 800 steps only. Position restrained molecular dynamics was executed for 50 ps by partially restraining the atomic positions of the protein while letting the water molecules move into it during simulation. The equilibrated system was then subjected to non-restrained molecular dynamics simulation of 1000 ps (1 ns). The protein stability was assessed by determining the root mean square deviation (RMSD) between the structures generated before and after MD simulation.

### 2.3. Docking studies of PfCS

The docking studies of substrate and inhibitors were performed using GLIDE (*Grid-based Ligand Docking with Energetics*) version 5.5 software by Schrödinger running on Windows 2003 on a HP xw8400 workstation. GLIDE approximates a complete systematic search of the conformational and positional space of the docked ligand [25]. For docking, the protein was prepared by adding hydrogens, assigning bond orders and minimizing overall structure to RMSD of 0.30 Å using OPLS2001 force-field in Maestro's protein preparation wizard. All the substrates were prepared using Maestro's Ligprep module (Ligprep, Schrödinger Inc., New York, NY). Receptor grid for docking was generated by the centroids of the selected amino acid residues of the active site. GLIDE was further used for the docking of the substrates into active site using Extra Precision (XP). Best pose was selected on the basis of Glidescore, rank and by visually inspecting the molecule in PyMOL. The best

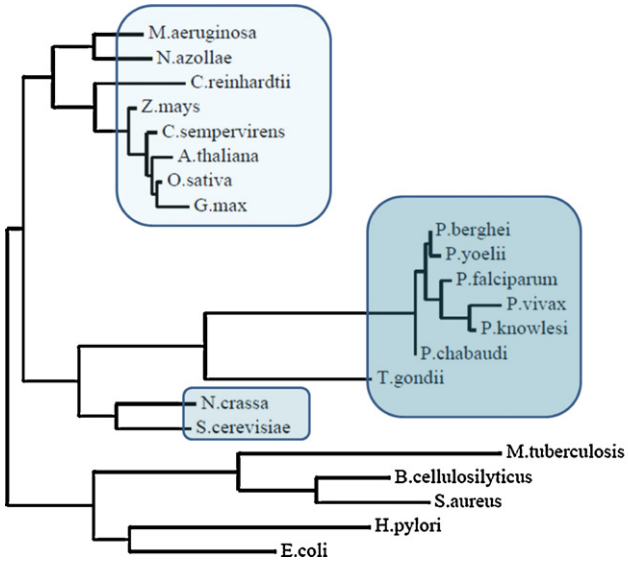


**Fig. 1.** Phylogeny of chorismate synthases. Phylogenetic reconstruction of chorismate synthase amino acid sequences from higher plant *Z. mays* (NP_001148583.1), *C. sempervirens* (CAA43034.1), *A. thaliana* (NP_001031158.1), *O. sativa* (BAD14928.1), *G. max* (ABA90483.1); apicomplexa *P. berghei* (XP_678920.1), *P. yoelii* (XP_724321.1), *P. falciparum* (AAB63293.1), *P. vivax* (AAL56611.1), *P. knowlesi* (XP_002261447), *P. chabaudi* (XP_743671.1), *T. gondii* (AAB52422.1); fungi *N. crassa* (AAC49056.1), *S. cerevisiae* (CAA42745.1); bacteria *M. aeruginosa* (YP_001656924.1), *N. azollae* (ADI63255.1), *M. tuberculosis* (NP_217056.1), *B. cellulosilyticus* (YP_004094866), *S. aureus* (NP_374580.1), *H. pylori* (AAB63293.1), *E. coli* (CAA68707.1); algae *C. reinhardtii* (XP_001695611.1). The apicomplexa CS follow a single branch and cluster together due to high sequence identity; however, they have nearest homologues to fungi CS. Though, apicomplexa CS physiologically similar to plant CS but make distant clusters in phylogenetic tree. The discussed separate clusters are highlighted in boxes. The phylogenetic tree was generated using Phylogeny FR server.

and the most energetically favorable conformation was selected for further analysis.

For further inhibitor analysis, AUTODOCK version 4 was employed to evaluate $K_i$ value. A Lamarckian Genetic Algorithm (LGA) in combination with a grid-based energy evaluation method was used for pre-calculating grid maps. Atomic potential grid map was calculated by AutoGrid 3 with a 0.375 Å spacing in a box of dimensions 30 Å × 30 Å × 30 Å. Box was centered at the active site of the model. AutoDock tools were used to calculate the atomic partial charges by the Gasteiger–Marsili method. Other docking parameters were set as default. For analysis, the configuration with lowest binding energy was selected as the representative.

## 3. Result and discussion

### 3.1. Sequence alignment and phylogenetic analysis

The sequence of *Pf*CS was obtained from NCBI with accession no AAB63293.1. The protein is expressed from AroC gene on chromosome no.6, which does not possess any intron [1]. The gene encodes the protein of 527 amino acids with predicted molecular weight of 59 kD and theoretical p*I* value 6.76. Sequence alignment reveals that *Pf*CS has maximum homology with fungal chorismate synthase (~33% identity). However, both the enzymes have different functional aspects. For its activity, CS requires the reduced FMN as cofactor. The category of CS that avails the reduced FMN exogenously is classified as monofunctional, generally found in bacteria, algae, plants and protozoa including *Plasmodium* species. On the other hand, CS from fungi viz. *N. crassa* and *S. cerevisiae* possess an additional NADPH:FMN oxidoreductase activity, and thus can be categorized as bifunctional enzymes. However, it is not possible to classify the enzyme on the basis of its sequence. The phylogenetic analysis of protein sequences of chorismate synthase revealed that apicomplexan including *Plasmodium* species and *Toxoplasma* species share maximum sequence identity so as to make a separate cluster in phylogenetic tree (Fig. 1). These CS sequences differ from other reported sequences in having a number of large internal insertions (Fig. 2). These insertions were thought to be signal sequence for apicoplast transport; as in plants, some cytosolic proteins can be targeted to plastids even without amino terminal leader sequence [26,27]. Thus, the shikimate pathway was proposed to be in apicoplast in apicomplexa. Eventually, Fitzpatrick et al. successfully confirmed that cytosol is the site of *Plasmodium* shikimate pathway [1].

Phylogenetic reconstruction using CS amino acid sequences from algae, higher plants, fungi, apicomplexan and prokaryotes indicates that the fungal and apicomplexan proteins have a common ancestor (Fig. 1). Although the sequence homology indicates that *Pf*CS is similar to fungal CS; however, biochemically it resembles with monofunctional CS from plant [1]. Plant CS is localized in plastid compared to *Pf*CS which is confirmed to be located in cytosol. Thus, *Pf*CS appears to possess unique properties that are distinct from the previously described CS.

The chorismate synthases exhibit three highly conserved signature motifs out of which two motifs possess conserved FMN binding residues [10]. *Pf*CS shares the same motif with some modifications. Hence previously described signature motif for CS

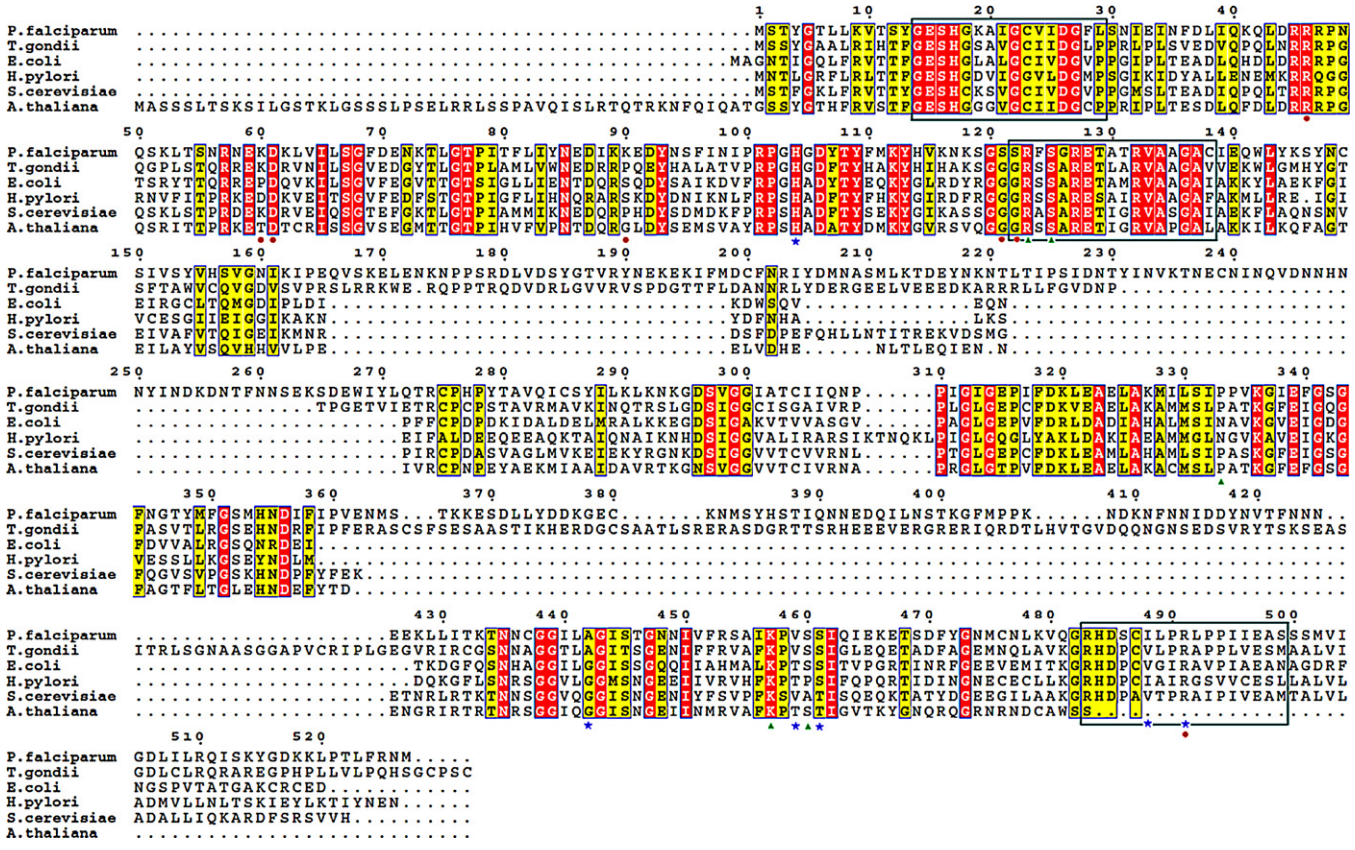


**Fig. 2.** Multiple sequence alignment of amino acid sequence of *Pf*CS with CS of *T. gondii* (AAB52422.1), *E. coli* (CAA68707.1), *H. pylori* (AAB63293.1), *S. cerevisiae* (CAA42745.1), *A. thaliana* (NP_564534.1). Identical and similar amino acids are highlighted with different background colors. The conserved signatory motif **G**-[DES]-**S**-**H**-[GC]-X2-[LIVM]-[GTIV]-X-[LIVT]-[LIV]-[DEST]-**G**-X-[PVL], [GES]-X2-**S**-[AG]-**R**-X-[ST]-X3-[VT]-X2-[GA]-[STAVY]-[LIVMFC] and **R**-[SH]-**D**-[PSV]-[CSAV]-X4-[SGAIVL]-X-[IVGSTAP]-[LIVM]-X-**E**-[STAH]-[LIVMAS] are highlighted in boxes. The triangle under the sequence indicates the active site residues interacting with FMN. Star indicates the residues involved in formation of FMN binding pocket, whereas circle indicates the residues interacting with EPSP. Figures are prepared in ESPript. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

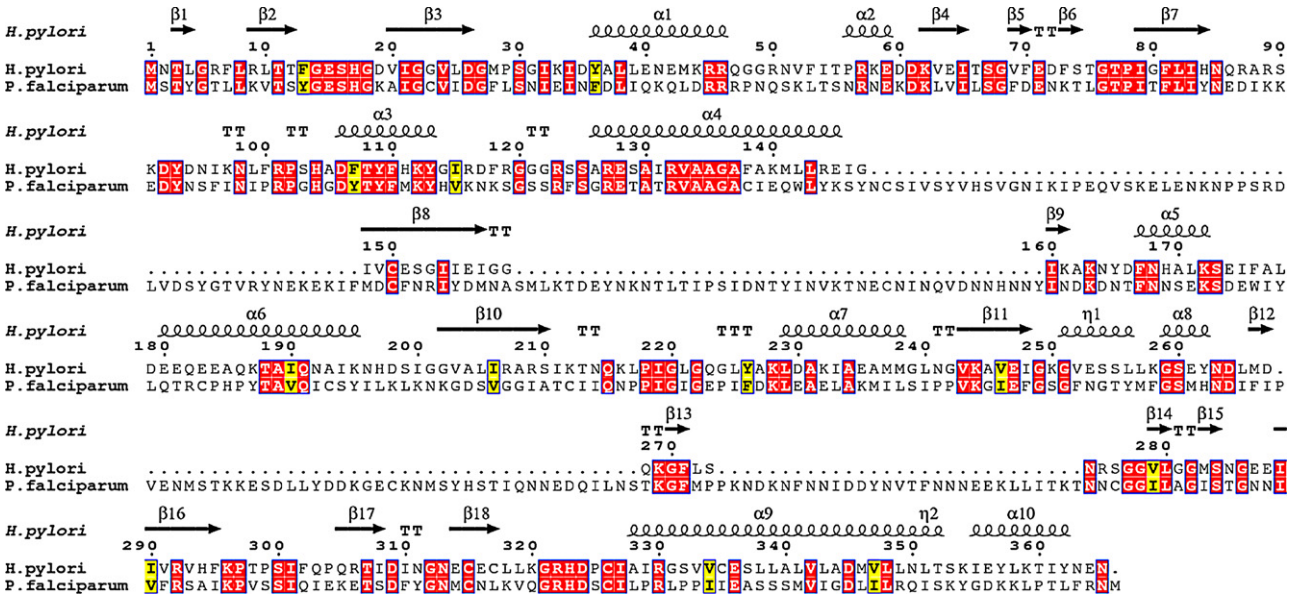


**Fig. 3.** Amino acid sequence alignment of *Pf*CS and *Hp*CS. The conserved residues are highlighted in different background. The secondary structural elements of *Hp*CS are shown above the aligned sequences. The internal amino acids inserts of *Pf*CS, Asn147 to Tyr251 and Pro360 to Thr433 do not show any homology with *Hp*CS except two conserved motifs [197-MDCFNRIYDMNA-210] and [399-TKGFMP-406]. The unmatched sequence has been carefully manipulated to generate a reliable model.

proposed by Ahn et al. may not be universal [10]. Here, we propose the modified signature sequence which may satisfy the conserved motif of apicomplexan CS. (i) Signature1: **G**-[DES]-**S**-**H**-[GC]-X2-[LIVM]-[GTIV]-X-[LIVT]-[LIV]-[DEST]-**G**-X-[PVL], (ii) Signature2: [GES]-X2-**S**-[AG]-**R**-X-[ST]-X3-[VT]-X2-[GA]-[STAVY]-[LIVMFC] and (iii) Signature3: **R**-[SH]-**D**-[PSV]-[CSAV]-X4-[SGAIVL]-X-[IVGSTAP]-[LIVM]-X-**E**-[STAH]-[LIVMAS] correspond to motifs 14-**G**E**SH**GKAIGCVID**G**FL-29; 122-SRF**S**G**R**ETATRVAAGAC-138 and 483-**R**H**D**SCILPRLPPII**E**AS-499 of *Pf*CS respectively, where X represents any amino acid and amino acids in bracket depicts the options at a given position. The bold faced residues indicate highly conserved residues at given position. The sample size for the above signature sequence was 32 out of which 10 samples were retrieved from apicomplexa.

### 3.2. 3D model generation

Homology search of *Pf*CS gave the hits of nineteen sequences against PDB database. The structure of bifunctional CS from *S. cerevisiae* (*Sc*CS) was the first hit with 33% sequence identity (PDB 1R52). It would be significant to mention here that homology model of *Pf*CS has been developed using *Sc*CS as template [28]. However, in the crystal structure of *Sc*CS electron density of 85 out of 376 amino acids (~23%) is missing most of which are highly conserved and are responsible for the formation of conserved domain near the active site of the protein [29]. Secondly, as discussed in Section 3.1, the *Pf*CS and *Sc*CS are functionally different proteins. Hence, a model based on *Sc*CS (1R52) as a template may not be reliable. With a temptation to generate a better and reliable model, we used structure of *Hp*CS complexed with FMN (PDB 1UM0) as template. Even though the identity of *Pf*CS was about 25% with 1UM0, the detailed knowledge of co-ordinates of complete amino acids was helpful to construct a reliable model. Sequence alignment with *Hp*CS gave indication of two different inserts in the sequence; insert one stretches from Asn148 to Tyr 251 and contains a conserved motif [198-MDCFNRIYDMNA-209] essential for a β-sheet in *Pf*CS core structure (Fig. 3). Insert two, stretches from Pro360 to Thr 433 which also contains a conserved domain [400-TKGFMP-405] responsible for the formation of a loop. These two inserts were cautiously studied and it was found that except the two conserved motifs in the insert the rest of the sequence may not have any significant role either in dimer formation or in substrate/cofactor recognition. Furthermore, sequence alignment of all *Plasmodium* CS shows very poor homology in these two insertions (figure not shown). As a result, in order to generate a reliable model the amino acids from the two inserts except conserved motif were removed from *Pf*CS model. After modification, the identity of modified *Pf*CS with the template was nearly 34%. The initial model of *Pf*CS was generated using Modeller 9v8. The model having lowest DOPE Score was further refined using loop refinement tool of Modeller and evaluated using PROCHECK and ERRAT plot after energy minimization. Ramachandran plot of the 3D model generated by PROCHECK showed 88% residues are present in the core region, 9.8% in allowed region, 1.9% in generously allowed region and 0.3% in disallowed region which includes only one residue. ProSA analysis showed that protein folding energy of the modeled structure is in good agreement favoring the validation of the model (Fig. 4A).

MD simulations were carried out using the predicted 3D structure of *Pf*CS to determine the stability of model in equilibration with solvent molecules, i.e. in the physiological state. The obtained MD trajectories during the simulation run of 1 ns were monitored and found to be stable (Fig. 4B). The graph revealed that RMSD values rise to 0.325 nm in the first 250 ps and then protein remained in the plateau state till the end of equilibration. An overall RMSD of 0.325 nm was obtained which indicates that the 3D modeled structure of *Pf*CS is good and has a stable conformation. The above result confirms that the 3D model constructed for *Pf*CS by comparative modeling is reliable for further detail structural analysis.

### 3.3. Predicted monomeric structure of PfCS

The predicted structure of the *Pf*CS monomer belongs to the α/β class as per the SCOP database [30]. Each monomer of *Pf*CS consists of 10α helices and 13β sheets (Fig. 5). The core of monomer has a unique "β-α-β sandwich fold" that appears as three layered structure in the predicted model. This unique fold consists of two antiparallel five stranded β-sheet layers (comprising of ↑β1 ↓β2 ↑β3 ↓β6 ↑β4 and ↑β8 ↓β7 ↑β9 ↓β12 ↑β10) and four α-helices (comprising of α1, α4, α9 and α7) which are sandwiched between these two β-sheet layers. These two layers of β-sheets may be

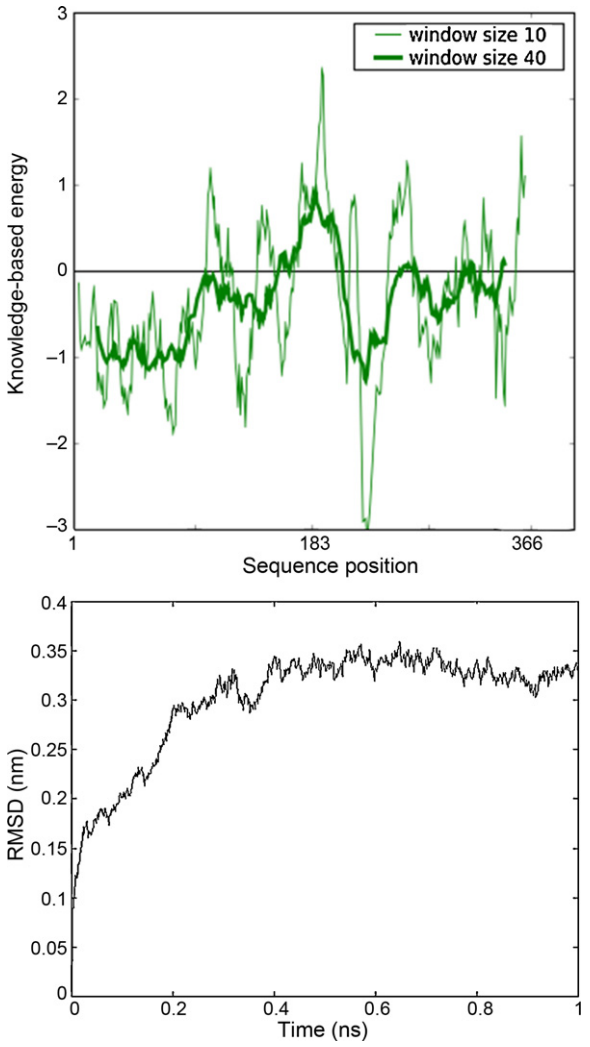


**Fig. 4.** (A) ProSA energy profile of *Pf*CS monomer. Thick line indicates average energy over each 40 residue fragment. The thin line indicates same with a smaller window size of 10 residues in the background of the plot. (B) The root mean square deviation (RMSD) of MD simulation *Pf*CS monomer relative to the structure of pre-MD simulation.

referred to as N-terminal and C-terminal β sheet layers, respectively. The predicted *Pf*CS structure has three β hairpins (β5a–β5b, β11a–β11b, β13a–β13b). The β13a–β13b hairpin has five residue turn whereas the other two hairpins have two residue turn. The β11a–β11b is a type II hairpin with a glycine at second position whereas β5a–β5b is a type I hairpin. The β5a–β5b hairpin is inserted between two adjacent antiparallel β-strands (β4 and β6). The α8 helix along with β11a–β11b hairpin is inserted between two adjacent antiparallel β strands (β10 and β12) and appears to be bulging out from the main body of predicted model. The unique β–α–β sandwich provides a scaffold for cofactor and substrate binding (further discussed below).

### 3.4. Analysis of dimeric structure

The residues responsible for dimerization as observed in solved crystal structure from various sources (1QXO, 1Q1L, 1UM0 and 1R53) are highly conserved in *Pf*CS [10,29,31,32]. This suggests that probably all CS exhibit at least a dimeric structure in solution.

Dimerization plays a functionally critical role, because it is required to form a complete FMN-binding pocket. The C terminal β-sheet layers from each subunit (↓β7 ↑β9 ↓β12 ↑β10 and ↓β10′ ↑β12′ ↓β9′ ↑β7′) contribute to the formation of eight stranded antiparallel β sheet layer which is a major feature of CS dimer (Fig. 6). The two sheets come together at β10 forming four inter-subunit hydrogen bonds at dimer interface. The residues responsible for dimerization are shown in Fig. 6. Important secondary structure elements involved in the stabilization of dimer are the helix α7, loop α8–β10 and helix α3 that interacts with α7′, β13b′ and loop α7′–β9′ of other subunit respectively. These interactions contribute to the dimer stabilization, presenting considerable amount of hydrophobic as well as hydrogen bonding interaction on the surface at the dimer interface. About 29 inter-subunit hydrogen bonds assist in the stabilization of the dimer.

### 3.5. Putative active site pocket and FMN binding mode

The FMN binding residues of CS active site pocket are highly conserved. Flexible loops of the enzyme contribute major role in creating the active site pocket to which cofactor gets bound through non-covalent bonds [10]. The residues of five major flexible loops that form FMN binding pocket are His104 (β6–α3), Arg123, Ser125 (α3–α4), Pro333 (α7–β10), Lys457, Val459, Ser460, Ser461 (β12–β13a), Ile488, Arg491 (β13b–α9). The isoalloxazine ring of FMN is buried into the hydrophobic pocket of active site, whereas the hydrophilic ribityl chain containing phosphate group is away from the molecular surface through a narrow cavity created in between loops (α7–β10) and (β6–α3) (Fig. 7). After the formation of stable dimer the surface of FMN binding pocket containing ribityl chain gets completely buried. Highly conserved Ser 125 in the active site pocket contributes to make stable interaction with isoalloxazine ring of FMN. However, Arg123, Ser460 interact with hydroxyl group and Lys457, Pro333 interact with phosphate group of ribityl chain of FMN. Arg491 which is a highly conserved amino acid in CS was not found to contribute any interaction with active site of generated model. The consecutive Pro333, Pro334 residues in the loop (α7–β10) makes a strict U turn and make convex surface to fit with bending ribityl chain of FMN. The helix α3 and loop in between β6 and α4 are expected to contribute major role in FMN binding mode.

In order to shed light on the role of loop movement in FMN binding, we have proposed a model assisted by following evidences. First, in the crystal structure of *Hp*CS and in *Pf*CS generated model, the cavity covering the ribityl chain of FMN has the width of 10.81 Å. This narrow cavity may not provide sufficient space for approximately 80° bent ribityl chain to align in proper orientation (Fig. 7B). So there might be movement of loop (loop containing α3) to broaden the active site pocket before binding of FMN. Secondly, in X ray diffraction pattern of *Hp*CS the electron density of 10 residues (87–96), which form similar loop near FMN binding pocket, is missing. Thus, giving indication to presence of highly flexible loop in this region (PDB 1UM0). Similar kind of deficiency (related to electron density) is also reported in the crystal structure of *Sc*CS (PDB 1R52). Thirdly, superimposition of FMN bound *Mycobacterium tuberculosis* CS (*Mt*CS) and its FMN free form (PDB 2O12 and 2O11) and those from *Hp*CS (PDB 1UM0 and 1UMF) clearly indicates movement of loop near FMN binding pocket making RMSD of 0.11 Å. Fourth, the α3 helix between flexible loops near FMN binding pocket forms various hydrogen bonds with residues from other subunit facilitating the dimerization and the compact packing of the FMN binding pocket.

To account for these observations, we suggest that the binding of FMN to the *Pf*CS apoenzyme and its dimerization may occur in sequential steps and may involve shifting of loop (Fig. 8). Such conformational flexibility may result in accommodation of cofactor and substrate, facilitating the catalysis. During the initiation of FMN binding, the loop β6–α3 is lifted from original position thus creating an open conformation, hence creating a sufficient space

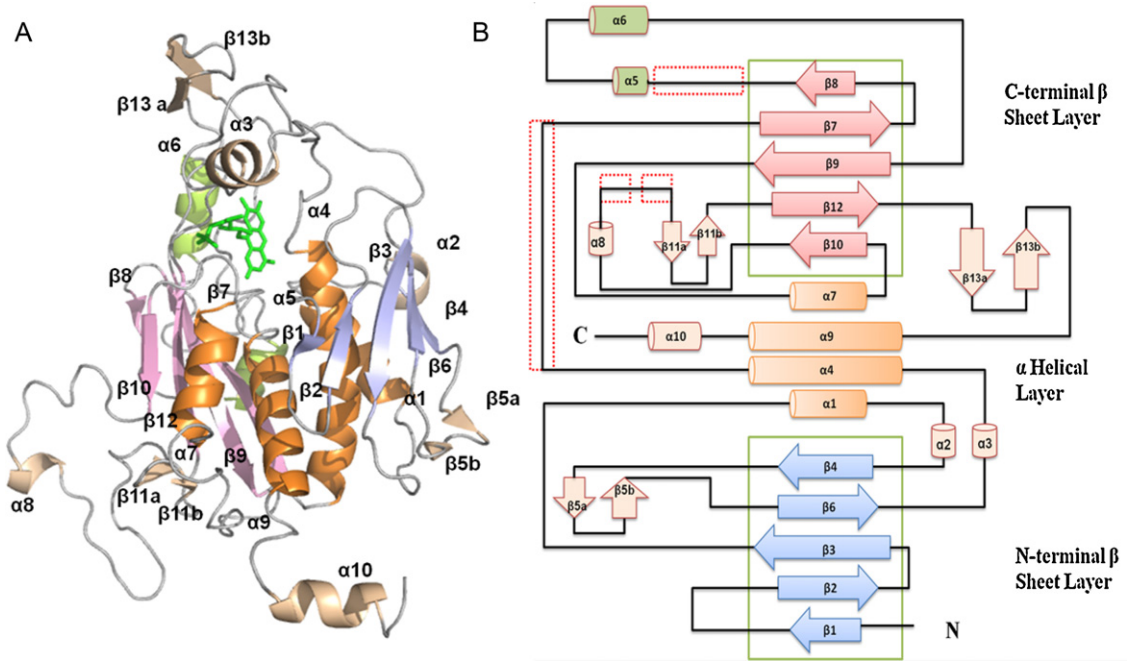


**Fig. 5.** Three dimensional structure of *Pf*CS monomer. (A) Ribbon diagram of *Pf*CS 3D model generated using Modeller9v8. (B) Topology diagram of *Pf*CS monomer. The secondary structure elements that belong to the three layered, β–α–β sandwich fold are colored in blue, light orange, and pink, respectively. The internal amino acid inserts have been highlighted in red dashed boxes on the loops. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

for the entry of FMN. Binding of FMN to its core pocket facilitates the formation of hydrogen bonds between Arg123, Ser460 and the hydroxyl group of ribityl chain. Successive to FMN binding, dimerization occurs with possible hydrogen interaction as discussed earlier (Section 3.3).

### 3.6. EPSP binding site

In order to study the binding mode of the substrate, EPSP was docked with the FMN bound *Pf*CS model. It is expected that EPSP will interact with a number of positively charged residues, as it is a negatively charged molecule. In the final docked structure, the interacting residues form an extremely basic environment favorable for EPSP binding. The residues involved in polar interactions with EPSP are Arg 46, Lys 60, Asp 61, Lys 90, Ser 121, Ser 122 and Arg 491(Fig. 9). The basic residues Arg 46, Arg 123, Arg 330 and Asp 61 are highly conserved in all CS and are involved in formation of substrate binding pocket. At position 121 and 122 in the CS of aligned species, glycine occurs as conserved residue, however in *Plasmodium* species the position is replaced by serine

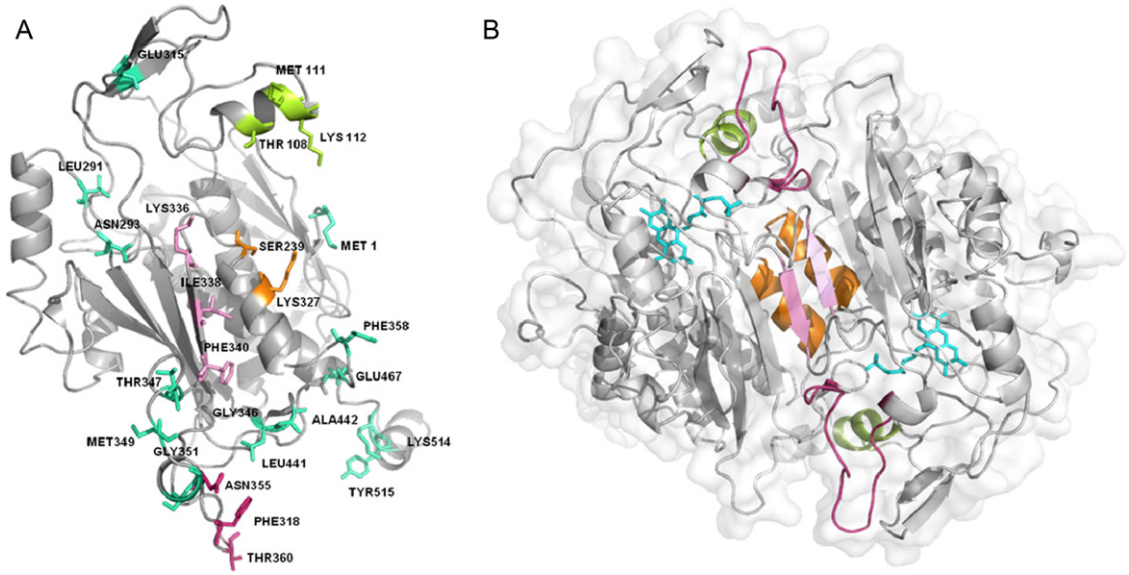


**Fig. 6.** Cartoon diagram of monomeric and dimeric form of *Pf*CS. (A) Interface residues involved in dimerization of *Pf*CS are indicated as colored sticks. (B) Ribbon diagram along with surface view of *Pf*CS dimeric state indicating important secondary elements in various colors involved in dimerization. Limon green and orange shows interface α-helix, light pink shows interface β sheet, warm pink shows interface loop and cyan shows other interface residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

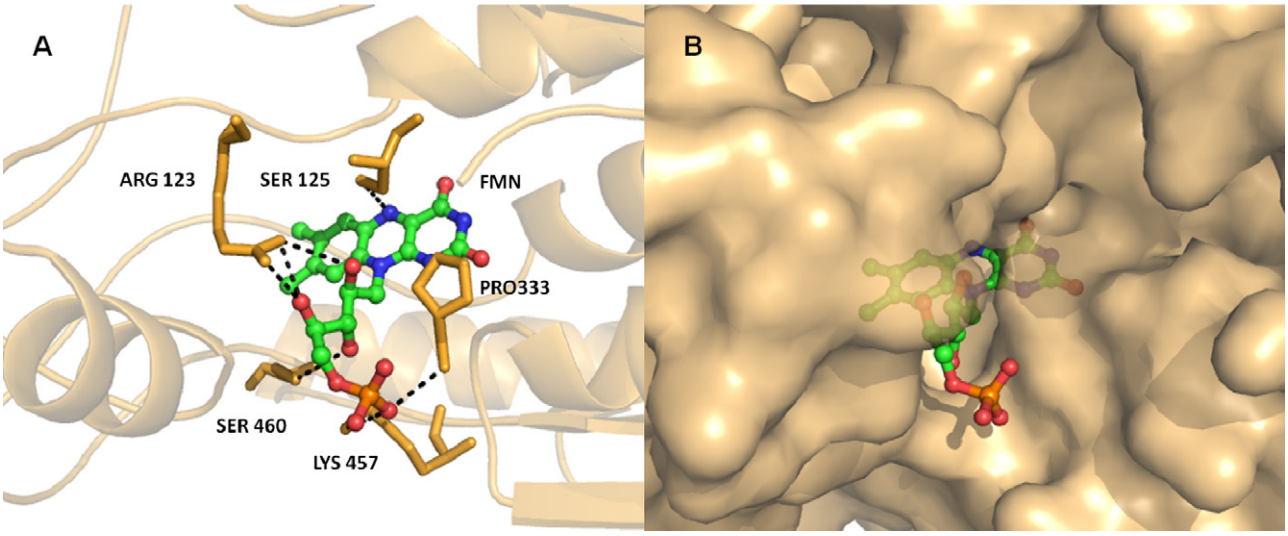


**Fig. 7.** Ribbon diagram of FMN binding pocket. (A) Residues showing polar interaction with FMN are shown in stick. Polar interactions are represented in dark dotted lines. (B) Surface view of FMN binding pocket. Isoalloxazine ring of FMN is buried in the monomeric pocket where as the ribityl chain comes out to the dimeric interface through a narrow cavity.

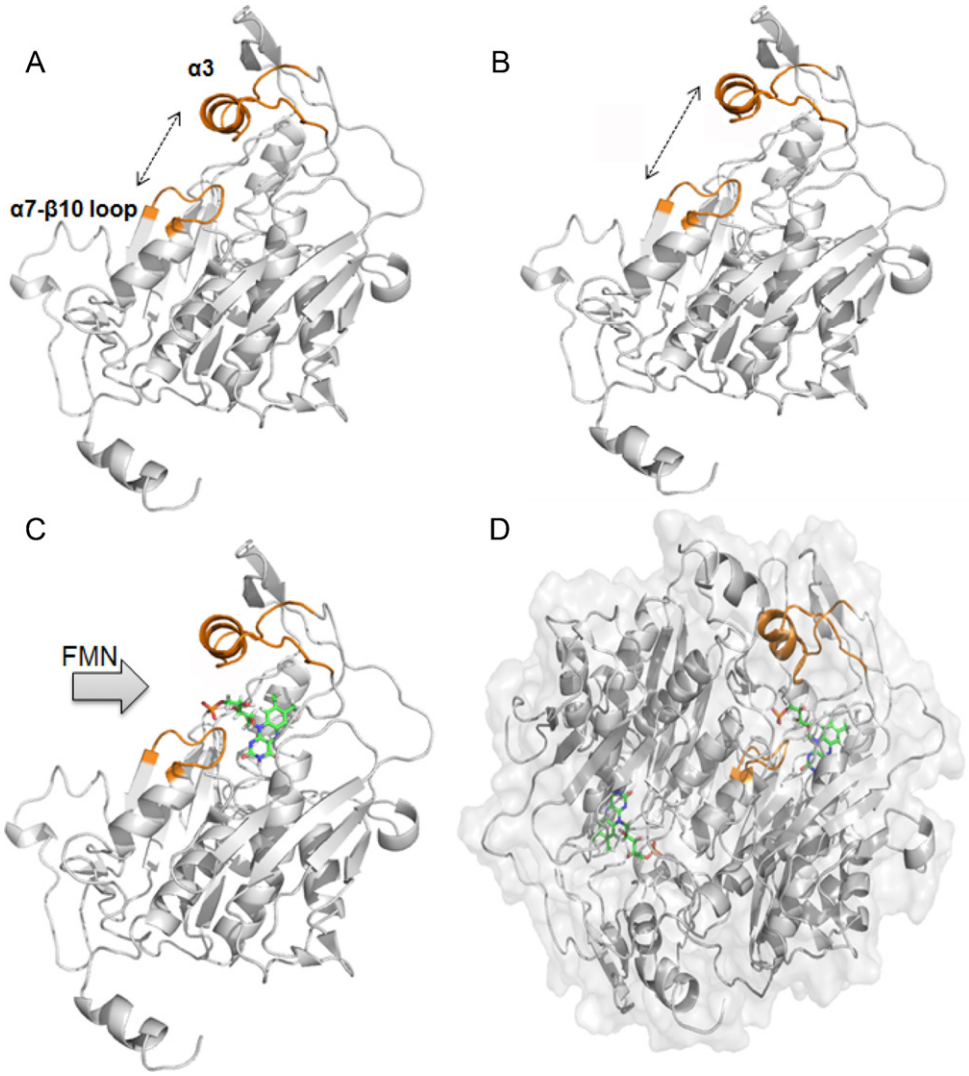


**Fig. 8.** Proposed sequential binding mode of FMN through loop movement. (A) Closed conformation of *Pf*CS prior to binding of FMN. (B) Open conformation caused by the opening of loops highlighted in orange color. (C) Entry of FMN in the open conformation of *Pf*CS. (D) Dimer formation followed by closed packing of FMN cavity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

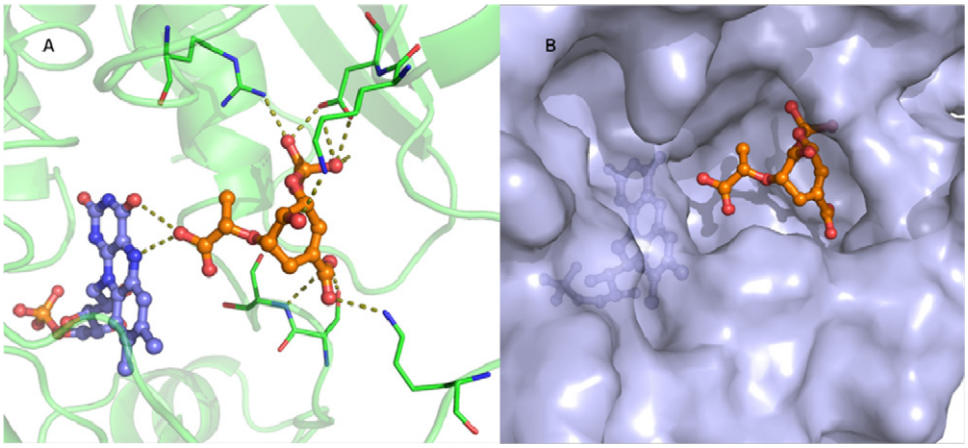


**Fig. 9.** Ribbon diagram of EPSP binding pocket. (A) Residues showing polar interaction with EPSP are shown as stick. Polar interactions are shown as dotted lines. (B) Surface view of EPSP binding pocket.

(Fig. 2). The O-7 group of EPSP shows polar interaction with the O-4 of FMN.

The positively charged residues are positioned at the three corners of EPSP site corresponding to the negatively charged EPSP. It is expected that modeled *Pf*CS exhibits the same mechanism of catalysis as per its monofunctional counterpart *Hp*CS.

## 4. Study of putative inhibitors of *Pf*CS

CS is one of the few untouched targets for developing anti-malaria drugs. Generation of a reliable model has opened the possibilities of computer-assisted inhibitor design against *Pf*CS. To achieve this objective within a reasonable time scale, we have used GLIDE docking software version 5.5 [33]. During the study of

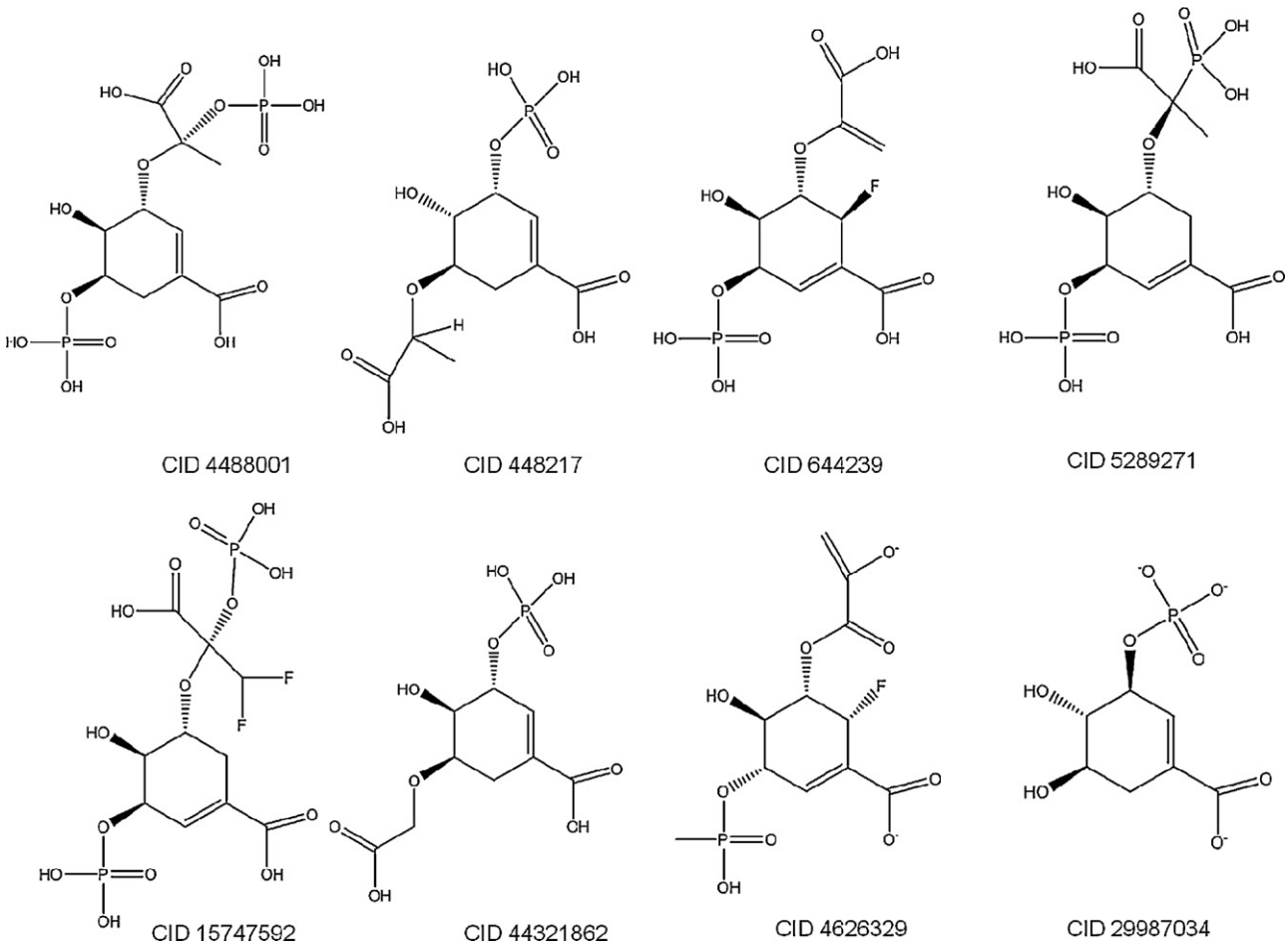


**Fig. 10.** 2D structure of EPSP analogs with corresponding PubChem Database ID.

**Table 1**
The table shows top screened EPSP analogs obtained from NCBI PubChem database along with GLIDE score and GLIDE energy using GLIDE version 5.5 software. The comparative analysis of *Pf*CS docking result was performed with *Helicobacter pylori* CS (*Hp*CS), *Streptococcus pneumoniae* (*Sp*CS). The theoretical $K_i$ value obtained using AUTODOCK software version 4.

| PubChem ID | Docking result with *Hp*CS, *Sp*CS and *Pf*CS | | | | | | | Interacting residues with *Pf*CS |
|---|---|---|---|---|---|---|---|---|
| | *Hp*CS | | *Sp*CS | | *Pf*CS | | | |
| | GLIDE score | GLIDE energy | GLIDE score | GLIDE energy | GLIDE score | GLIDE energy | AUTODOCK $K_i$ | |
| CID 448001 | −5.170 | −40.765 | −11.417 | −47.982 | −6.951 | −41.744 | 109.20 nM | Arg46, Glu86, Lys90, Arg491 |
| CID 448217 | −4.23 | −37.79 | −13.572 | −53.303 | −7.502 | −32.178 | 192.02 nM | Arg47, Lys60, Glu86, Lys90, Arg491 |
| CID 644239 | −4.561 | −31.403 | −12.668 | −46.192 | −7.472 | −28.747 | 815.02 nM | Arg47, Lys60, Glu86, Lys90, Arg491 |
| CID 5289271 | −4.585 | −34.278 | −9.682 | −45.522 | −7.579 | −36.780 | 526.36 nM | Lys60, Glu86, Lys90, Arg483, Arg491 |
| CID 15747592 | −5.168 | −36.051 | −9.303 | −47.691 | −7.133 | −33.213 | 610.03 nM | Arg47, Lys60, Glu86, Lys90, Arg483, Arg491 |
| CID 44321862 | −4.382 | −31.002 | −14.722 | −48.389 | −7.509 | −29.617 | 1.79 μM | Lys60, Glu86, Lys90, Arg491 |
| CID 4626329 | −5.084 | −35.483 | −11.558 | −49.125 | −7.533 | −38.574 | 2.33 μM | Arg46, Arg47, Lys60, Glu86, Lys90, Ser121, Gly126 |
| CID 29987034 | −4.596 | −24.19 | −12.704 | −42.465 | −8.369 | −29.493 | 1.79 μM | Arg46, Arg47, Asp61, Asn85, Glu86, Arg491 |

**Table 2**
The table shows selected benzofuran-3[2H]-one analogs prioritized with high docked Glide score using GLIDE version 5.5 software. The comparative analysis of docking result of *Pf*CS was performed with *Helicobacter pylori* CS (*Hp*CS) and *Streptococcus pneumoniae* (*Sp*CS). The theoretical $K_i$ value obtained using AUTODOCK software version 4. The experimental $IC_{50}$ value with *Sp*CS was obtained from the Binding Database (www.bindingdb.org).

| PubChem ID | Docking result with *Hp*CS, *Sp*CS and *Pf*CS | | | | | | | Experimental | Interacting Residues with *Pf*CS |
|---|---|---|---|---|---|---|---|---|---|
| | *Hp*CS | | *Sp*CS | | *Pf*CS | | | $IC_{50}$ (nM) with *Sp*CS | |
| | GLIDE score | GLIDE energy | GLIDE score | GLIDE energy | GLIDE score | GLIDE energy | AUTODOCK $K_i$ | | |
| ChEBI 323325 | −3.318 | −37.664 | −4.756 | −39. 983 | −4.510 | −40.009 | 2.66 μM | >50,000.0 | Gln50, Thr54, Glu59, Ser486 |
| ChEBI 323327 | −4.850 | −32.366 | −3.032 | −37.659 | −4.370 | −37.007 | 5.01 μM | 3400.0 | Lys52, Gln50, Arg483 |
| ChEBI 323293 | −3.712 | −47.538 | −5.906 | −42.194 | −4.161 | −36.357 | 32.74 μM | 5000.0 | Arg47, Gln50, Glu86, Arg491 |
| ChEBI 323801 | −4.145 | −43.090 | −5.211 | −34.759 | −3.251 | −34.716 | 7.91 μM | 8500.0 | Lys52, Gln50 |
| ChEBI 323292 | −3.506 | −36.805 | −4.917 | −37.472 | −3.149 | −32.218 | 6.53 μM | 25,900.0 | Gln50, Glu59 |
| ChEBI 323282 | −3.095 | −40.933 | −4.247 | −36.127 | −2.733 | −34.469 | 9.56 μM | 5500.0 | Lys52, Gln50 |

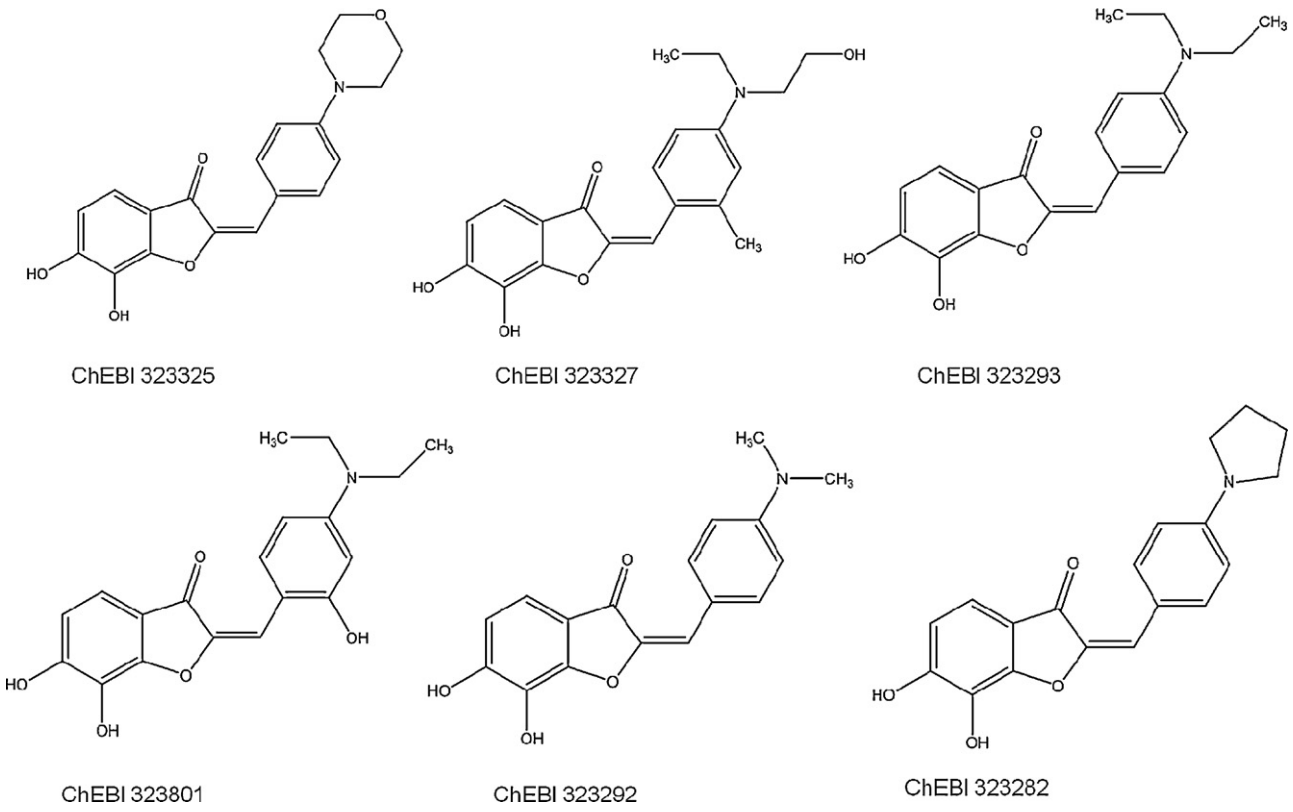


**Fig. 11.** 2D structure of series of selected benzofuran-3[2H]-one analogs reported by Thomas et al. as potent inhibitor of *Sp*CS.

putative inhibitors using GLIDE, the protein–ligand interacting site was restricted to the binding site of the EPSP (as discussed in Section 3.6). To screen the ligand(s) that bound to the active site of *Pf*CS, we used EPSP analogs available in PubChem database as well as the previously reported inhibitors of *Streptococcus pneumoniae* CS (*Sp*CS) [34].

EPSP analogs were retrieved from PubChem database and docking procedure yielded about eight best binding analogs (Fig. 10). The selected molecules were further assessed for their theoretical $K_i$ values using AUTODOCK software version 4. The predicted binding efficiency of retrieved analogs was estimated based on the values of dissociation constant $K_i$, docking score, docking energy and extent of interactions with active site residues. The comparative study of eight different analogs with highest predicted binding efficiency into active site of *Pf*CS, *Hp*CS and *Sc*CS was made and corresponding docking results along with interacting residues of *Pf*CS are shown in Table 1.

Thomas et al. has described a series of analogs of benzofuran-3[2H]-one as inhibitors of *Sp*CS with varying degree of potency [34]. We retrieved the submitted structures of these analogs from the Binding Database (www.bindingdb.org) and studied their binding capability in active site pockets of *Sp*CS, *Hp*CS and *Pf*CS (Fig. 11). Six of these inhibitors with highest predicted binding efficiency along with corresponding docking parameters and interacting residues are shown in Table 2. In our study, 6,7-dihydroxy-2-[1-(4-morpholin-4-yl-phenyl)-meth-(E)-ylidene]-benzofuran-3-one (ChEBI ID: 323325), with reported experimental $IC_{50}$ value 50 μM with *Sp*CS, was found to be most potent molecule binding into the active site of *Pf*CS ($K_i$ = 2.66 μM). In the docked structure of this inhibitor with *Pf*CS, the 6,7 hydroxyl groups of catechol is found to interact with Thr54 and Gln59. It also interacts with Gln50 and Ser486 using its furan ring. The catechol and the furan ring are found to be present in the same binding pocket as that of EPSP (figure not shown).

## 5. Conclusion

In present study, the molecular model of *Pf*CS was generated using the crystal structure of *Hp*CS as template. The modeled structure of *Pf*CS monomer shows a three layered "β–α–β sandwich fold" which is a characteristic feature of CS family. We have also proposed the modified signature sequence that satisfies the conserved motif of apicomplexa CS. The model also facilitated the recognition of conserved residues in active site pocket. As all the CS mainly exists in dimeric form, the dimeric *Pf*CS was generated to recognize all the interface residues involved in dimerization. The reliable model also permitted the study of binding efficiency of substrate analogs as well as previously reported inhibitors of *Sp*CS. The study yielded a small highly focused virtual subset of EPSP analogs, with speculated inhibitory capacity against *Pf*CS and having binding efficiency in nanomolar range. This study can thus help to direct the focus of synthetic chemists, designing the next generation of antimalarial agents.

## Acknowledgments

This work was supported by financial aid from the Department of Science and Technology, New Delhi, India. We thank Dr. Shailly Tomar for helpful discussions. We also acknowledge Macromolecular Crystallographic Unit (MCU) for providing computational facilities. Satya thanks AICTE, Abhinav thanks DBT, Sonali thanks MHRD and Preeti thanks CSIR for financial support.

## References

[1] T. Fitzpatrick, S. Ricken, M. Lanzer, N. Amrhein, P. Macheroux, B. Kappes, Mol. Microbiol. 40 (2001) 65–75.
[2] R.G. Ridley, Nature 415 (2002) 686–693.
[3] G.A McConkey, J.W. Pinney, D.R. Westhead, K. Plueckhahn, T.B. Fitzpatrick, P. Macheroux, B. Kappes, Trends Parasitol. 20 (2004) 60–65.

[4] F. Roberts, C.W. Roberts, J.J. Johnson, D.E. Kyle, T. Krell, J.R. Coggins, G.H. Coombs, W.K. Milhous, S. Tzipori, D.J. Ferguson, D. Chakrabarti, R. McLeod, Nature 393 (1998) 801–805.
[5] M.V. Dias, J.C. Borges, F. Ely, J.H. Pereira, F. Canduri, C.H. Ramos, J. Frazzon, M.S. Palma, L.A. Basso, D.S. Santos, W.F. de Azevedo Jr., J. Struct. Biol. 154 (2006) 130–143.
[6] L.A. Basso, L.H. da Silva, A.G. Fett-Neto, W.F. de Azevedo Jr., I.S. Moreira, M.S. Palma, J.B. Calixto, S. Astolfi Filho, R.R. dos Santos, M.B. Soares, D.S. Santos, Mem. Inst. Oswaldo Cruz 100 (2005) 475–506.
[7] J.H. Pereira, F. Canduri, J.S. de Oliveira, N.J. da Silveira, L.A. Basso, M.S. Palma, W.F. de Azevedo Jr., D.S. Santos, Biochem. Biophys. Res. Commun. 312 (2003) 608–614.
[8] S. Bornemann, M.K. Ramjee, S. Balasubramanian, C. Abell, J.R. Coggins, D.J. Lowe, R.N. Thorneley, J. Biol. Chem. 270 (1995) 22811–22815.
[9] P. Macheroux, S. Bornemann, S. Ghisla, R.N. Thorneley, J. Biol. Chem. 271 (1996) 25850–25858.
[10] H.J. Ahn, H.J. Yoon, B. Lee 2nd, W.S. Suh, J. Mol. Biol. 336 (2004) 903–915.
[11] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, J. Comput. Chem. 30 (2009) 2785–2791.
[12] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theor. Comput. 4 (2008) 435–447.
[13] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA, 2002, http://www.pymol.org.
[14] P. Emsley, K. Cowtan, Acta Crystallogr. D: Biol. Crystallogr. 60 (2004) 2126–2132.
[15] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.
[16] P. Gouet, E. Courcelle, D.I. Stuart, F. Métoz, Bioinformatics 15 (1999) 305–308.
[17] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.M. Claverie, O. Gascuel, Nucleic Acids Res. 36 (2008) 465–469.
[18] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403–410.
[19] A. Sali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779–815.
[20] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, J. Appl. Cryst. 26 (1993) 283–291.
[21] C. Colovos, T.O. Yeates, Protein Sci. 2 (1993) 1511–1519.
[22] R. Lüthy, J.U. Bowie, D. Eisenberg, Nature 356 (1992) 83–85.
[23] M. Wiederstein, M.J. Sippl, Nucleic Acids Res. 35 (2007) 407–410.
[24] S.R. Comeau, D.W. Gatchell, S. Vajda, C.J. Camacho, Nucleic Acids Res. 32 (2004) 96–99.
[25] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, J. Med. Chem. 47 (2004) 1739–1749.
[26] P.J. Keeling, J.D. Palmer, R.G. Donald, D.S. Roos, R.F. Waller, G.I. McFadden, Nature 397 (1999) 219–220.
[27] G.I. McFadden, M.E. Reith, J. Munholland, N. Lang-Unnasch, Nature 381 (1996) 482.
[28] N. Arora, U.V.N.M. Chari, A.K. Banerjee, U.S.N. Murty, Internet J. Genom. Proteom. 3 (2007).
[29] S. Quevillon-Cheruel, N. Leulliot, P. Meyer, M. Graille, M. Bremang, K. Blondeau, I. Sorel, A. Poupon, J. Janin, H. van Tilbeurgh, J. Biol. Chem. 279 (2004) 619–625.
[30] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, J. Mol. Biol. 247 (1995) 536–540.
[31] J. Maclean, S. Ali, Structure 11 (2003) 1499–1511.
[32] C.M. Viola, V. Saridakis, D. Christendat, Proteins 54 (2004) 166–169.
[33] J.B. Cross, D.C. Thompson, B.K. Rai, J.C. Baber, K.Y. Fan, Y. Hu, C. Humblet, J. Chem. Inf. Model. 49 (2009) 1455–1474.
[34] M.G. Thomas, C. Lawson, N.M. Allanson, B.W. Leslie, J.R. Bottomley, A. McBride, O.A. Olusanya, Bioorg. Med. Chem. Lett. 13 (2003) 423–426.