

# Insight into structural and biochemical determinants of substrate specificity of PFI1625c: Correlation analysis of protein-peptide molecular models



Kimjolly Lhouvum<sup>a</sup>, Vibin Ramakrishnan<sup>b</sup>, Vishal Trivedi<sup>a,\*</sup>

<sup>a</sup> Malaria Research Group, Department of Biotechnology, Indian Institute of Technology-Guwahati, Guwahati 781039, Assam, India

<sup>b</sup> Molecular Informatics & Design Laboratory, Department of Biotechnology, Indian Institute of Technology-Guwahati, Guwahati 781039, Assam, India

## ARTICLE INFO

### Article history:

Accepted 28 March 2013

Available online 8 April 2013

### Keywords:

Malaria  
Drugs  
Resistance  
Modeling  
Peptides  
Protein sorting  
Metalloprotease  
Mitochondria  
Apicoplast

## ABSTRACT

Bioinformatics and sequence comparison indicate PFI1625c as a putative metalloprotease present in plasmodium genome. The structure of PFI1625c consists of two domains with nearly identical folding topology. The active site of PFI1625c is located in a large central cavity between the two domains. Substrate binding regions of PFI1625c are lined by E-136, D-140 which provides negatively charged patches whereas F-53 facilitates binding of bulky hydrophobic residues of substrates. Probing PFI1625c active site with 199 different peptides from a combinatorial peptide library indicates preference of PFI1625c toward hydrophobic residue substituted peptides. Correlation analysis of each position of the peptide indicates that Ser 2 is the most crucial residue and no significant improvement was observed until it is mutated to a hydrophobic residue. The peptide P550 (LVIVAKRA) exhibits significantly better interaction within the active site than a template peptide (LSRVAKRA). The molecular dynamic's simulation studies confirms integrity of the complex, with all structures well within the qualitative limit of compactness and stability during the simulation time. There are structural and biochemical differences between PFI1625c with human metalloprotease and these are sufficient enough to allow us to exploit PFI1625c as drug targets. These computationally obtained insights provided clues about substrate selectivity in PFI1625c and it can be used to exploit PFI1625c as a target for future anti-malarial development.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

Malaria causes ~1 million deaths every year and with the development of drug-resistant parasite strains situation is quite alarming in tropical and sub-tropical area [1,2]. Malarial parasite depends extensively on proteolysis events for survival throughout its life cycle [3]. One of the most conspicuous events of the invasion is the removal or “shedding” of the fuzzy coat proteins covering the merozoite surface by serine protease SUB2 [4]. This process is also aided by rhomboid protease [5]. Serine protease gp76 degrades band 3 and glycophorin A to weaken the RBC cytoskeleton to facilitate a merozoite invasion [6]. Hemoglobin digestion, the most important metabolic pathway is completely controlled by co-ordinated action of different proteolytic events inside the acidic food vacuole of parasite [7–11].

Comparative Genome Analysis has identified 92 putative proteases in *P. falciparum* genome [12]. Potentially important proteases which remain uncharacterized consist of the calpain type

proteases, metacaspase, primary processing proteases or signal peptidase [12–14]. In *P. falciparum*, a calpain, yet unidentified, was believed to be essential in merozoite invasion, based on the observation that calpain inhibitors I and II strongly blocked invasion [15]. Parasite exports proteins to different organelles to carry out essential functions. Transport of proteins requires processing of signal sequence by signal peptidase for delivery of protein to reach their destinations [16]. Transport of proteins outside the parasite interacts with proteins of the erythrocyte membrane and remodels the host cell morphologically to induce disease associated pathology, responsible for deaths during malaria [17,18]. Proteins targeted to mitochondria and apicoplast require pre-processing of nuclear coded proteins by mitochondrial or apicoplast processing peptidases or signal peptidases [19]. These organelles are prokaryotic in origin and were endo-symbiotic within the parasite body to perform essential functions for parasite growth. A number of proteins transported into mitochondria and apicoplast play an important role in the invasion of merozoite to start the new infection cycle in the RBC [20]. Therefore current efforts have been made to unveil the signal processing machinery of the mitochondria/apicoplast for identification of a new drug target [21]. PFI1625c is present in the *P. falciparum* genome (<http://www.plasmodb.org>) and is

\* Corresponding author. Tel.: +91 361 2582217; fax: +91 361 258 2249.

E-mail addresses: [vtrivedi@iitg.ernet.in](mailto:vtrivedi@iitg.ernet.in), [Vishalash.1999@yahoo.com](mailto:Vishalash.1999@yahoo.com) (V. Trivedi).

proposed to play a role as a putative organelle processing peptidase. The constitutive expression of PFI1625c was found throughout the erythrocytic stages of the parasite life cycle [22]. In the current study an effort was made to perform an in-depth structural and functional characterization of PFI1625c. Multiple sequence alignment analysis and homology modeling provide evidence that PFI1625c belongs to metallo-protease M16 family. A combinatorial peptide library was prepared based on the signal peptide sequence LSRVAKRA present in yeast malate dehydrogenase (a template peptide used in the current study to model PFI1625c 3-D structure). Analysis of the molecular model of PFI1625c-peptide indicates a preference of PFI1625c toward hydrophobic groups substituted peptides. PFI1625c is evolutionarily very distant from human matrix processing peptidase. Hence our structural and peptide modeling studies characterize the substrate preference of PFI1625c and highlight its importance as an excellent drug target for anti-malarial drug development due to no potential cross reactivity with host proteases.

## 2. Materials and methods

### 2.1. Sequence analysis and alignment

Amino Acid Sequence of bc1 core protein of chicken, bovine, yeast mitochondrial processing peptidase and bacillus peptidase was retrieved from NCBI database and multiple sequence alignment was performed using Clustal W 2.0.11 [23]. A phylogenetic tree was generated by blasting PFI1625c into NCBI non-redundant protein database (nr) using the neighbor joining method of Clustal W2 phylogeny program [24].

### 2.2. Molecular modeling and structure validation

The amino acid sequence of PFI1625c was retrieved from the Plasmodium Database (<http://www.plasmodb.org>). PFI1625c 3D structure modeling was performed in the following sequential steps: template selection, sequence-template alignment, model building, refinement and validation. The suitable template was identified by searching PFI1625c into the protein data bank of NCBI using PSI-blast (<http://blast.ncbi.nlm.nih.gov/Blast>). The top hits were further analyzed for least number of gaps and highest sequence identity. Crystal structure of mitochondrial processing peptidase from yeast (PDB code 1HR6,  $\beta$ -subunit) was found to be suitable [25] with 38% identity over 435 residues. The 3D model of PFI1625c was generated using MODELLER 9v9 [26]. Auto model module of modeler 9v9 was used to generate 100 initial models with a ga431 score of 1 and these models were ranked based on their DOPE scores and mol pdf. Top ten models having the lowest DOPE scores and mol pdf were selected. The stereo-chemical quality of each model was confirmed by PROCHECK and ramchandran plot [27,28]. The statistics of non-bonded interactions between different atom types were analyzed by ERRAT program which gives a measure of the structural error at each residue in the protein [29]. The compatibility of the atomic model (3D) with its own amino acid sequence was determined by Verify\_3D [30]. The model with the least number of residues in the disallowed region was selected and energy minimized using Steepest Descent (SD) algorithm with GROMOS96 43a1 force fields in GROMACS 4.0.7 package (<http://www.gromacs.org/>). This process was repeated until most of the residues are not below 95% in ERRAT plot. The final model was checked by Verify-3D, Procheck and Ramchandran plot. The PROCHECK and ramchandran plot analysis indicate that final model is of good quality with more than 95% residues in the allowed region. The ERRAT plot gives most of the residues below 95%

with overall quality factor of 90.071. Overall no short contact was observed in the final model (Fig. S1).

### 2.3. Generation of combinatorial peptide library

Signal peptide (LSRVAKRA) of malate dehydrogenase shows binding in PFI1625c in a similar conformation as present in the crystal structure of yeast mitochondrial processing protease (PDB Code 1HR9). The signal peptide of malate dehydrogenase (LSRVAKRA) was used as a template to prepare a peptide library to optimize the binding of peptide into the PFI1625c active site. A random single or double substitution into the template peptide sequence (LSRVAKRA) was done to prepare a combinatorial peptide library of 733 peptides representing different types of probable substitution (Table S1). The type of substitution was hydrophobic, neutral and long chain hydrophobic and charged residues. After critically analyzing a complete peptide library and excluding similar substitution with weak interactions, 199 peptides were used for molecular modeling, 3-D structure modeling of peptide sequence from the library was done with the help of EasyModeller 2.0 [31]. The structural quality of each modeled peptide was checked by Procheck and Ramchandran plot and most of the peptide models were of high quality.

### 2.4. Molecular Modeling of peptides from the combinatorial peptide library into PFI1625c active site

The individual peptide from the combinatorial peptide library was fitted into the PFI1625c 3-D structure using patchdock [32]. 'Fire-Dock' program is used initially to perform energy minimization of best 20 models to get stable PFI1625c-signal peptide complexes [33]. Atomic contact energy (ACE) and local proximity to the predicted binding sites in a PFI1625c active site of each conformation was analyzed. Control Docking of signal peptide (LSRVAKRA) in yeast mitochondrial processing peptidase (PDB Code 1HR9) using patchdock shows, peptide binds in the extended conformation as given in co-crystal structure [25]. The root mean square deviation between generated model and crystallographic complex (1HR9) is 1.52 Å. To validate the approach, docking was also performed by taking matrilysin and its peptide substrates from MEROPS database [34] and through literature search [35]. The docking experiments showed a strong correlation ( $r=0.941$ ) between the affinity of the peptide substrate ( $K_m$ ) to the atomic contact energy (Fig. S2). Correlation of experimentally determined biochemical data with docking score proves the reliability of using Patch-dock to generate PFI1625c-peptide molecular models and correlates the docking score with suitability of the bound peptide as potential substrate.

### 2.5. Molecular dynamics simulation

Molecular dynamics (MD) simulations and analysis were performed in an Intel Xeon Workstation using GROMACS program suite [36] with GROMOS96 43A1 force field. The dynamics of PFI1625C-peptide complex were studied at 300 K under NVT conditions with periodic boundary. Each of the ten complexes was energy minimized in vacuum using steepest descent algorithm in 500 steps. Solvent molecules (SPC water) were then added to the rectangular box with the distance between solute and the box set at 1 nm. The system was once again energy minimized in water before 1 ns production run with an integration step of 2 fs. Throughout the simulations, Bond lengths were constrained with Shake algorithm to a geometric accuracy of  $10^{-4}$ . Initial velocities were taken from a Maxwellian distribution at the chosen temperature of 300 K temperature bath with a coupling relaxation time of 0.1 ps. The non bonded list cutoff of 1.4 nm was used. The nonbonded interactions

(VdW and electrostatics) were shifted to zero between 0.8 nm to 1.1 nm cutoff.

### 3. Result

#### 3.1. PFI1625c is a metalloprotease

Multiple sequence alignment implicates that PFI1625c belongs to the metalloprotease family. The characteristic conserved sequences pattern present in PFI1625c matches with other known metalloprotease (Fig. 1). To further validate PFI1625c as metalloprotease, it was analyzed by ProtIdent web server [37] which is developed by fusing the functional domain and sequential evolution information. The first layer identifies the query protein as protease or nonprotease while the second layer identifies the protease class. Proldent server indicates that PFI1625c is a protease belonging to the metalloprotease family. A characteristic zinc binding motif and active site residues specific for a metalloprotease class are also present in PFI1625c. Phylogenetic analysis of PFI1625c against non-redundant database shows that PFI1625c forms a separate cluster with putative proteases of other apicomplexan species and it is evolving very distantly from human proteases (Fig. 2A).

#### 3.2. The overall structure of PFI1625c

The complete structure of PFI1625c can be divided into domains I and II. There is a sequence difference between two domains but the  $\alpha$ -helix and  $\beta$ -sheet folds in each domain to give similar folding topology of a non-superimposable image (Fig. 2B). A loop of 25 residues connects  $\alpha$ -helix 11 of a domain I and  $\beta$ -sheet 7 of domain II. The  $\alpha$ -helices and  $\beta$ -sheets are equally distributed in both the two domains.  $\beta$ -Sheets of each domain are projected into the interior core of the PFI1625c and consist of two hairpin structures and  $\Psi$ -loop motif. The  $\Psi$  loop structures from each domain are composed of two anti parallel  $\beta$  strands. The  $\alpha$ -helices are present toward the exterior portion of the PFI1625c to shield the interior  $\beta$  sheets (Fig. 2C). These arrangements give rise to a complete protease structure with an organized active site cavity and intact catalytic site present in other metalloproteases.

#### 3.3. Zinc binding site

A characteristic feature of metalloprotease is the presence of the metal binding motif [38]. Multiple Sequence alignment indicates that a conserved zinc binding motif (HXXEH) with H-46, E-49, H-50 residues are present in PFI1625c where zinc co-ordinates with E-126 (Fig. 3A). The structure of the protein implies that  $\alpha$ -helix 6 and 2 from a domain I come closer to each other with a unique folding pattern to align the zinc binding residues (H-46, E-49, H-50) to form a catalytic triad (Fig. 3A and B). The E-126, another zinc binding residue is also coming closer with the H-46 and H-50 to form the perfect geometry that fits the zinc metal.

#### 3.4. Substrate binding site

Substrate binding region of PFI1625c is constituted by E-136, D-140 and F-53 (Fig. 3B). The active site is present inside a well organized large cavity in the center of the structure. The active site cavity closer to domain I is negatively charged whereas rest cavity is of mixed environment (Fig. 3C). The domain I region of the active site is negatively charged to facilitate binding of positively charged residues of the substrate while the remaining portion of the cavity consists of mixed charges with a hydrophobic environment toward the interior to provide docking sites for hydrophobic residues of a peptide substrate. The cavity extends toward the domain II which may assist in binding the large substrates to allow

**Table 1**  
Analysis of top hit peptides from combinatorial peptide library.

| Peptide code | Amino acid sequence | Atomic contact energy (kcal/mol) |
|--------------|---------------------|----------------------------------|
| Template     | LSRVAKRA            | -69.85                           |
| P548         | LVDVAKRA            | -262.81                          |
| P550         | LVIVAKRA            | -262.47                          |
| P543         | LLIVAKRA            | -220.83                          |
| P536         | LIIVAKRA            | -144.42                          |
| P540         | LLKVAKRA            | -174.58                          |
| P33          | LSVVAKRA            | -114.1                           |
| P32          | LSLVAKRA            | -100.12                          |
| P31          | LSIVAKRA            | -98.26                           |
| P168         | ASIVAKRA            | -82.26                           |
| P545         | LLVVAKRA            | -73.05                           |
| P537         | LILVAKRA            | -57.47                           |
| P544         | LLLVAKRA            | -45.29                           |
| P546         | LVHVAKRA            | -33.53                           |
| P683         | LSIVHKRA            | -31.82                           |
| P170         | ASVVAKRA            | -23.61                           |
| P551         | LVLVAKRA            | -22.01                           |
| P538         | LIVVAKRA            | -13.35                           |

A random combinatorial peptide library was prepared from template peptide LSRVAKRA present in co-crystallized structure of mitochondrial processing peptidase (PDB code 1HR9). 3-D model of peptides were prepared as described in Section 2. The peptides from combinatorial peptide library were modeled into the PFI1625c active site using patch-dock and atomic contact energy was calculated. The peptides with hydrophobic substitutions fit well into the PFI1625c active site.

**Table 2**  
Comparison of peptide interaction within the active site of PFI1625c.

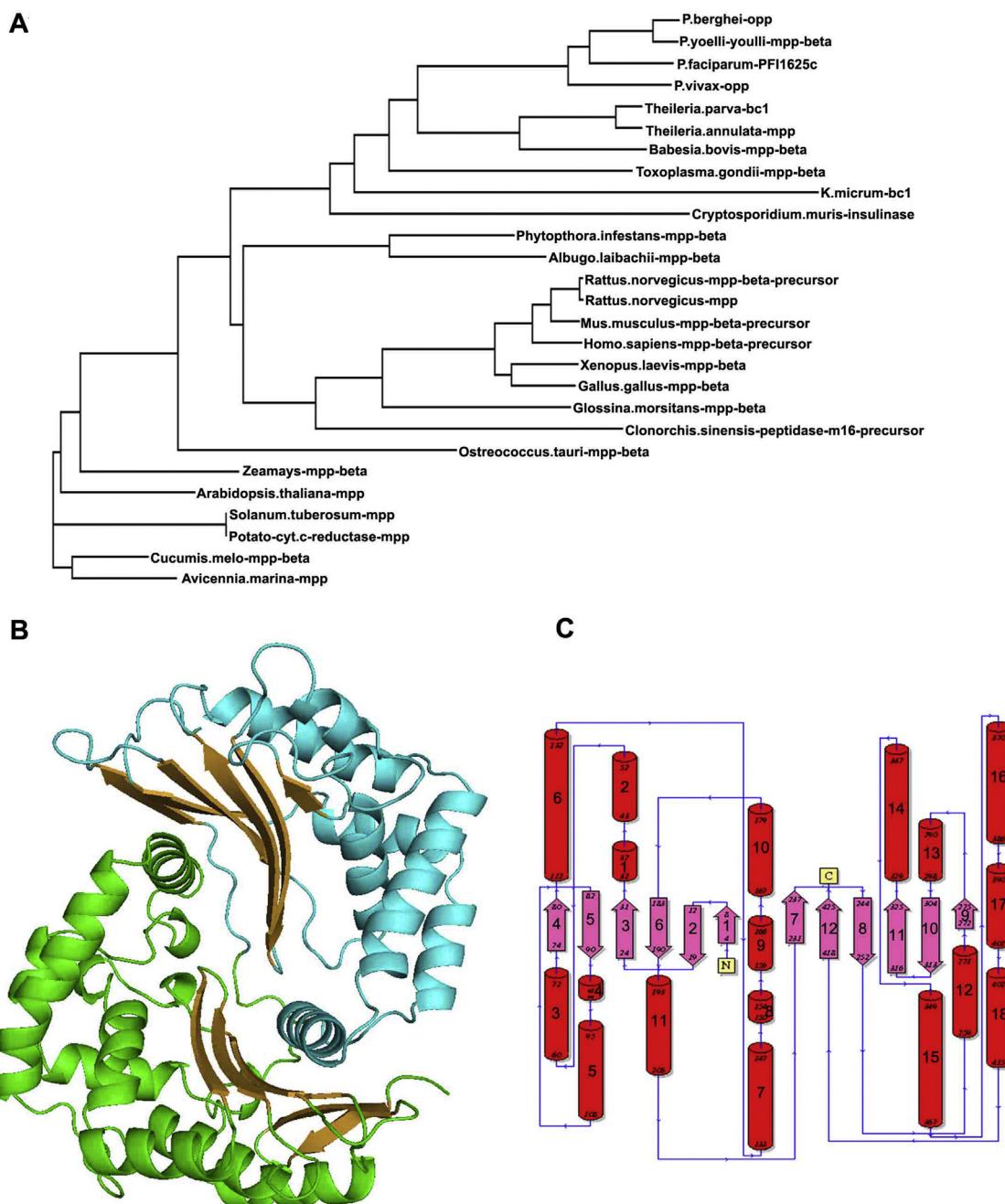
| Template peptide | Protein residue | Atom | Distance ( $\text{\AA}$ ) |
|------------------|-----------------|------|---------------------------|
| L1               | M264            | CB   | 2.04                      |
| L1               | M264            | CD2  | 2.46                      |
| S2               | Q267            | NE2  | 2.98                      |
| S2               | N310            | H    | 1.82                      |
| S2               | N310            | O    | 2.9                       |
| R3               | Q363            | NE2  | 4.88                      |
| V4               | R81             | N    | 4.27                      |
| A5               | R81             | NE   | 4.34                      |
| K6               | Y78             | OH   | 2.34                      |
| R7               | T79             | O    | 3.47                      |
| R7               | E136            | OE2  | 1.64                      |
| A8               | H46             | HD1  | 2.5                       |
| A8               | E49             | OE2  | 2.34                      |

| Peptide P550 | Protein residue | Atom | Distance ( $\text{\AA}$ ) |
|--------------|-----------------|------|---------------------------|
| L1           | Q267            | OE1  | 3.18                      |
| L1           | L364            | CD2  | 2.61                      |
| L1           | L364            | CD1  | 2.84                      |
| L1           | M264            | CE   | 2.01                      |
| V2           | C312            | CB   | 2.25                      |
| V2           | C312            | N    | 2.55                      |
| I3           | C312            | N    | 3.07                      |
| I3           | N310            | O    | 2.42                      |
| V4           | C312            | O    | 2.93                      |
| V4           | Y313            | CE1  | 3.34                      |
| A5           | R81             | NE   | 3.06                      |
| K6           | Y78             | CE1  | 3.71                      |
| R7           | E136            | OE2  | 1.29                      |
| A8           | E126            | OE2  | 2.15                      |
| A8           | H46             | NE2  | 2.19                      |

Interactions of the peptide with active site protein residues in PFI1625c-template peptide and PFI1625c-peptide P550 molecular model. The optimized peptide sequence P550 fits well into the PFI1625c active site as evidenced from a number of strong interactions as compared to template peptide.

access of the cleavage site by the catalytic residues present in the domain I (Fig. 3D). Substrate binding scaffolds comprised of 2  $\beta$ -sheets, contributed from both domains. The  $\beta$ -sheet from domain I has conserved HLNAY residues while the  $\beta$ -sheet from domain II has SFNTCY residues.

**Fig. 1.** Characterization of different structural elements in PFI1625c. Multiple sequence alignments of PFI1625c from *P. falciparum* with bc1 core protein of *G. gallus* (Gg bc1), *B. taurus* (Bt bc1), metalloprotease of *S. cerevisiae* (Sc MPP) and protease from *B. halodurans* (Bh Pep). Similar and identical residues are indicated by standard annotation. The zinc binding region and substrate binding scaffold are enclosed within the rectangular box to highlight the presence of these regions in PFI1625c.

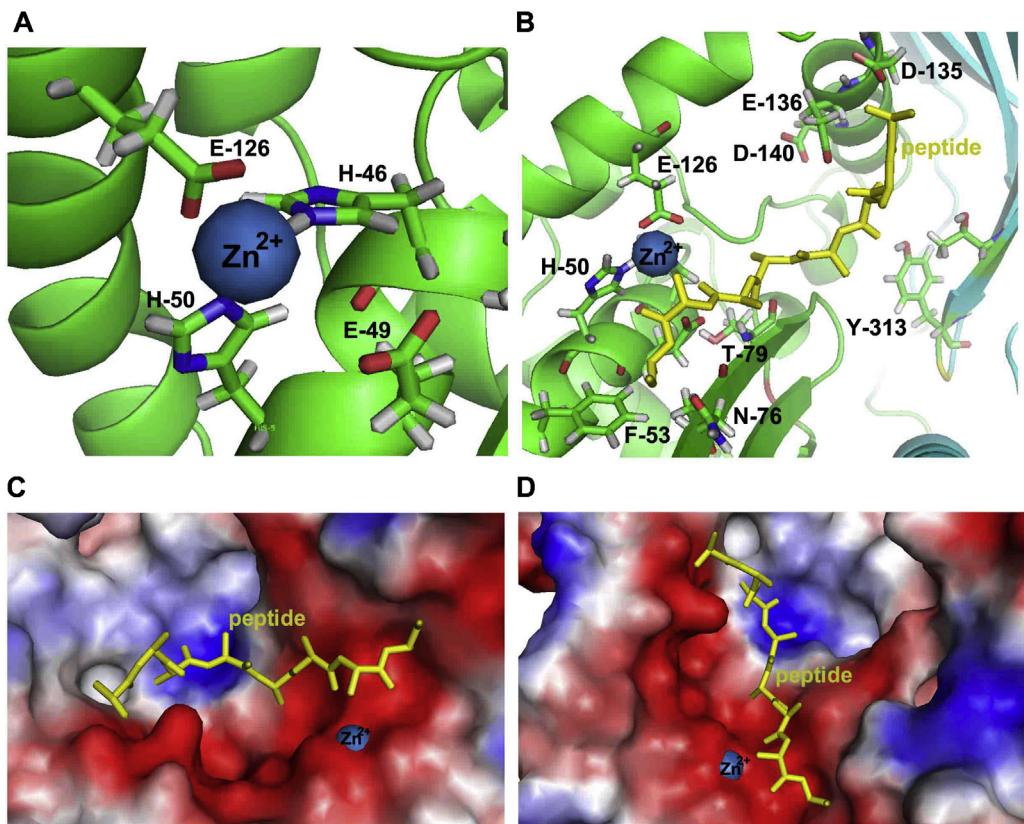


**Fig. 2.** Overall structure of PFI1625c. (A) Evolutionary relationship of PFI1625c with proteases present in other organism. Human metalloprotease is evolutionarily diverse from PFI1625c and form a distinct cluster. (B) 3-D structure of PFI1625c consists of two domains. Domain I is colored green and domain 2 is colored cyan. The β-sheets present in respective domains are colored in brown. (C) Topology diagram of PFI1625c to depict the organization of structural elements. The α-helices are colored red while β-sheets are colored pink and structural elements are connected by blue colored lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3.5. PFI1625c prefers hydrophobic residues in substrate

The interior of the substrate binding cavity is more hydrophobic in PFI1625c indicating that binding of hydrophobic residues in the cavity will be energetically more favorable than binding a hydrophilic residue. A template peptide (LSRVAKRA) binds PFI1625c with an ACE of -69.85 (Table 1). It is present in an extended conformation as predicted in the co-crystal structure of mitochondrial processing peptidase bound to signal peptide (PDB code 1HR9). The LSRVA occupy the hydrophobic region of the active site pocket while KRA occupy the more negatively charged region

(Fig. 4A). In the catalytic site, a template peptide makes polar contact, hydrogen bonding, van der waals with the residues surrounding substrate binding pocket (Fig. 4B). In the present binding mode residues of template peptide makes either weak interactions or no interaction at all (Table 1). A combinatorial peptide library of ~733 peptides based on the peptide sequence (LSR-VAKRA) was generated to represent all possible substitutions to map active site environments (Table S1). Individual peptides from the combinatorial peptide library (~199 peptides) were modeled into the binding pocket to study the peptide substrate interaction and binding within binding pocket. An in-depth interactions



**Fig. 3.** PFI1625c is a metalloprotease with a well defined active site. (A) Zinc binding motif of PFI1625c present in the domain I. The zinc ion is shown in blue and the residues interacting with  $Zn^{2+}$  to stabilize it within the catalytic site are shown with corresponding label. (B) Active site of PFI1625c with bound peptide substrate. The template peptide (LSRVAKRA) was taken from yeast mitochondrial processing peptidase (PDB ID 1HR9) and modeled into the active site. The residues interacting with peptide substrate are labeled and shown. PFI1625c surface representation in (C) horizontal and (D) vertical view with bound template peptide in yellow. Red represents –ve charge, blue represents +ve charge. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

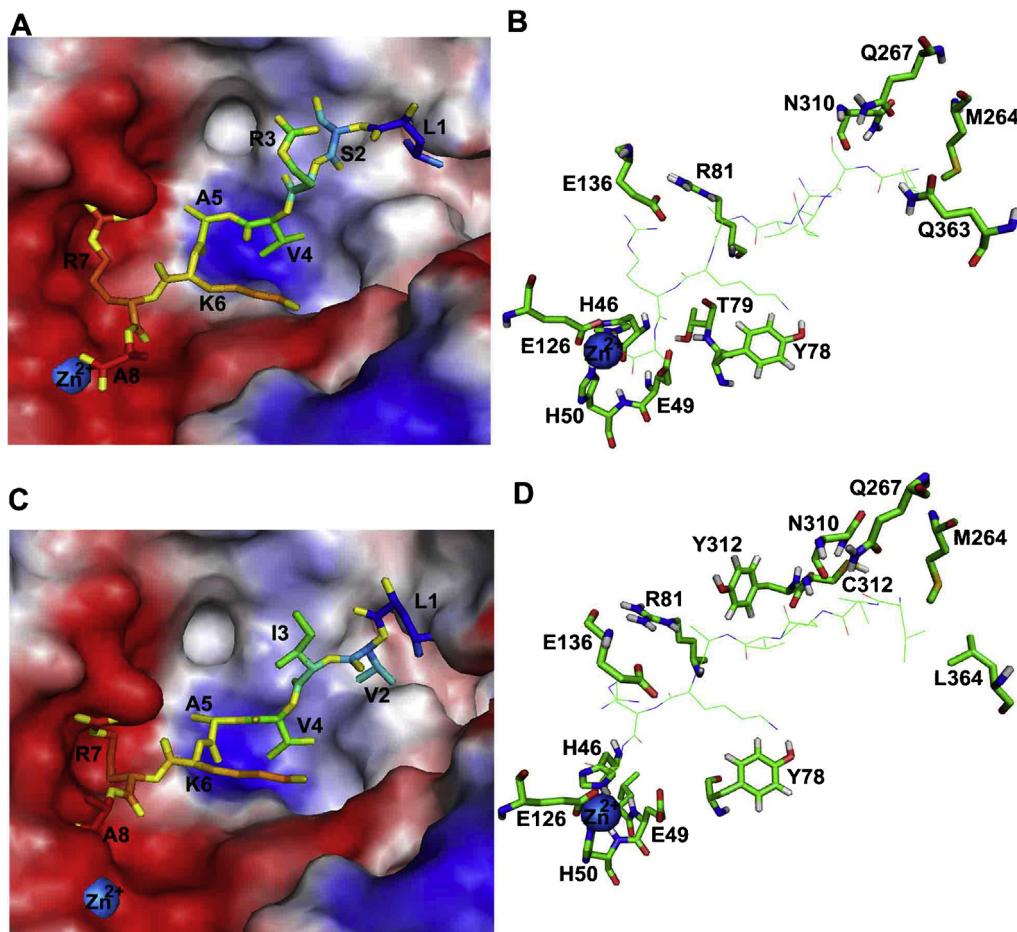
analysis indicates a number of potential peptide substrates with significantly high binding score and ACE values (Table 1).

Correlating the substitution at each position in a template peptide with other amino acid residues allow us to understand the catalytic preference of PFI1625c for substrate peptide. 17 peptides with similar group substitutions at position 1 or 2 did not give better interaction than template peptide except peptide P91 where Leu 1 is replaced by Ile and Ser 2 by Thr (Table S2). Substitutions at 1, 2 or 3 positions with acidic amino acid alone didn't show any improvement but substitution with acidic and hydrophobic amino acid has improved the peptide binding than a template peptide. 48 peptides with amino acid substitution at position 1 or 2 such as in peptide P152 where Leu 1 is substituted with Trp and Ser 2 is replaced with His provide better interaction with ACE of –114.38 (Table S2). Out of 21 peptides with hydrophobic substitution, 8 peptides gave significantly better interaction than a template peptide. The overall result showed that hydrophobic amino acid substituted peptides have the better interaction than a template peptide. The best interaction was found when the residue Ser 2 is replaced by Leu, Ile or Val and Arg 3 are replaced by Ile (Table 1). Correlation analysis of each position of the peptide indicates that Ser 2 is the most crucial residue and no significant improvement was observed until it is mutated to a hydrophobic residue. Peptide P550 (LVIVAKRA) fits well and shows the best interaction within the catalytic site of PFI1625c. In the catalytic site, P550 makes polar contact, hydrogen bonding, van der waals with the residues surrounding substrate binding pocket (Fig. 4C). In the present binding mode residues of P550 make several strong interactions as well as  $\pi$ - $\pi$  stacking interactions with the hydrophobic groups

present in binding pocket and hydrophobic groups substituted in P550 (Fig. 4D). In comparison to a template peptide, P550 makes several and very strong interactions within the binding pocket of PFI1625c (Table 2). Overall analyses of different peptide substrates indicate that peptide with hydrophobic substitution increases the interaction between peptide and active site residues especially residues present within a hydrophobic pocket of the active site.

### 3.6. Molecular dynamics analysis of PFI1625c-peptide complexes

Three programs *g\_gyrate*, *g\_hbond* and *g\_rms* in GROMACS have been used for analysis. “Compactness” of the protein throughout the simulation was monitored using *g\_gyrate*, which measures the radius of gyration. Radius of gyration values further underline the compactness of the complex with all structures with almost all complexes showing downward trend, with a negative slope (Fig. 5A). P31 and p543 were the most compact ones among the ensemble of structures sampled during 1 ns simulation whereas P168 was found to be comparatively loose. Root mean square deviation from starting structure was monitored using *g\_rms* and hydrogen bonding interaction between the peptide and the receptor was mapped with the help of *g\_bond* program. RMS deviation from the starting structure of the complex molecule were observed to be in the range of 0.2–0.3 nm with P168 being closest to the starting structure and P31 deviating maximum in one ns trajectory (Fig. 5B). Overall, simulation results underline integrity of the complex, with all structures well within the qualitative limit of compactness and stability during the simulation time.



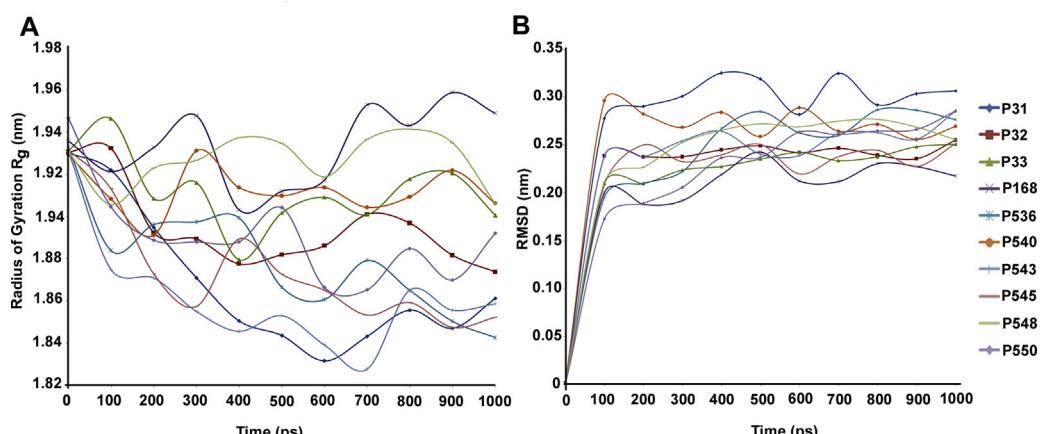
**Fig. 4.** Correlation analysis of PFI1625c-peptide molecular model. (A) PFI1625c-template peptide model and (C) PFI1625c-P550 peptide model binds peptide in an extended conformation to the large surface area within the active site of PFI1625c. P550 shows better fitting into the active site exploiting opposite charge pockets and hydrophobic patches. Red represents –ve charge, blue represents +ve charge. Interaction of (B) template peptide and (D) P550 with the residues present within the active site of PFI1625c.

Simulation results of top ten structures hence confirm the stability of interaction, though the simulation time may not be sufficient to create a large enough conformational ensemble to rank them accordingly.

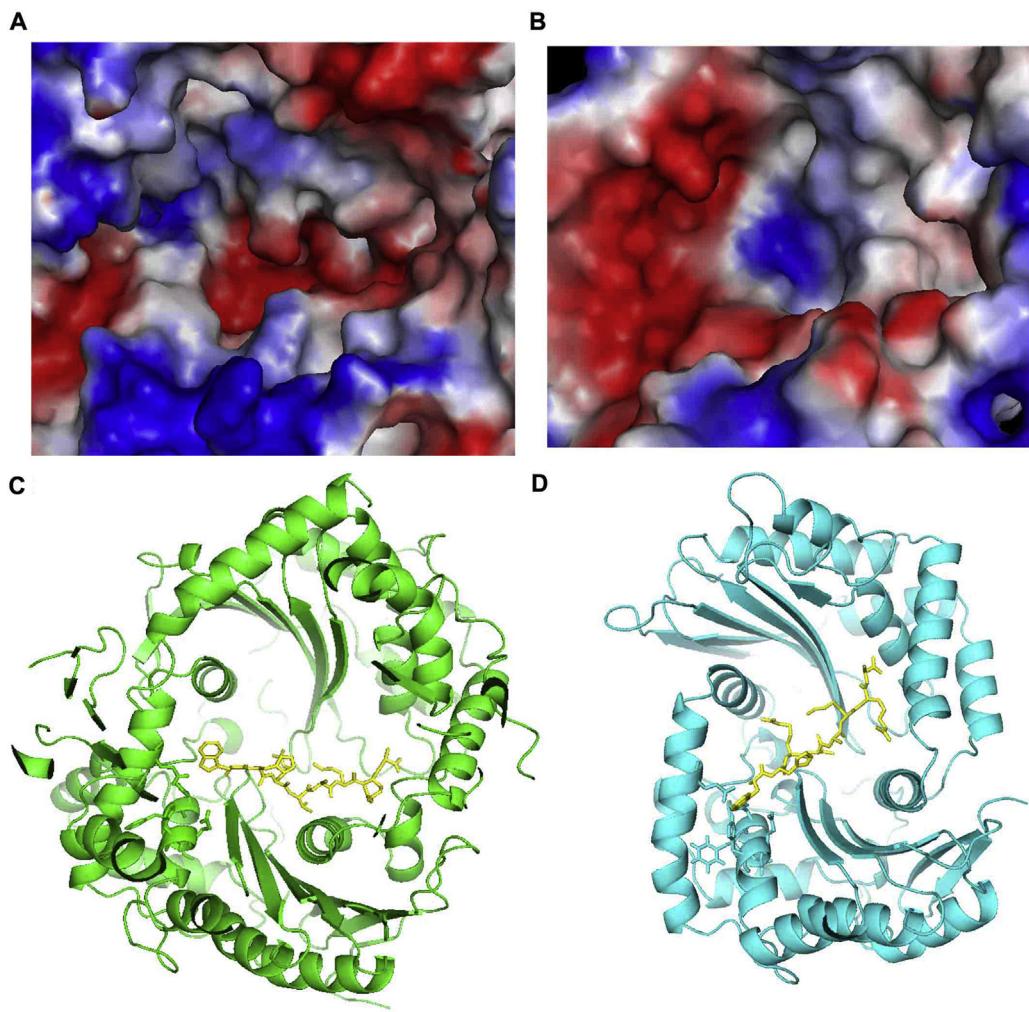
### 3.7. PFI1625c is a candidate drug target

Blasting PFI1625c against non-redundant database shows that PFI1625c forms a separate cluster and it is evolving very distant

from host proteases. The closest human metalloprotease (mpp  $\beta$  precursor) has diversified into a distinct group from the PFI1625c cluster (Fig. 2A). The closest available 3-D structure of human metalloprotease to PFI1625c is insulin degrading enzyme (IDE) and it is used to test the suitability of PFI1625c as a drug target. PFI1625c and IDE Exhibit 22% sequence identity with conserved catalytic residues but both are very distinct at structural level. The active site cavity of human IDE is buried inside a closed pocket and entry of a peptide is restricted and substrates need to enter through a



**Fig. 5.** Molecular dynamics simulation of top ten PFI1625c-peptide complexes. (A) Radius of Gyration and (B) RMSD of all complexes. MD analysis shows that the PFI1625c-peptide complexes are energetically stable under the simulation conditions.



**Fig. 6.** The surface structure of the active site and interaction of bound peptide within human insulin degrading enzyme (IDE) and modeled structure of PFI1625c. (A) and (B) Surface Structure of the peptide binding pocket of IDE and PFI1625c-peptide model. Blue is positive, red is negative and peptide (yellow) is present in an extended conformation. (C) and (D) Interaction of bound peptide with key residues present within the binding pocket of IDE and PFI1625c. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

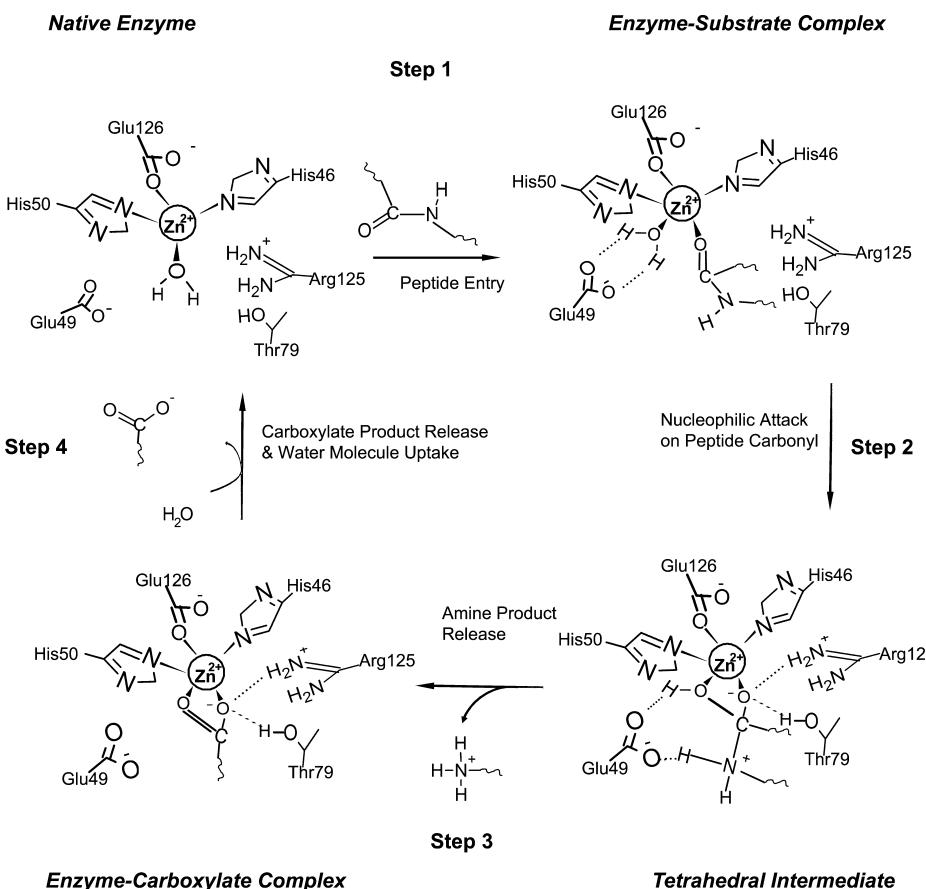
narrow tunnel (Fig. 6A) where as PFI1625c is present in an open groove with complete access of an incoming peptide (Fig. 6B). The active site of both proteases also exhibits significant differences in active site environment and charge distribution. The catalytic region of IDE is more positively charged or neutral except the zinc binding motif (Fig. 6C), PFI1625c is highly negatively charged (Fig. 6D). The validation of PFI1625c as a drug target is further affirmed by comparative studies with similar and characterized metalloprotease group in *P. falciparum* such as falcilysin, PfM18AAP, Pf-AM1 and PfLAP. Their functions in the parasite biology are characterized and their effect in the growth of the parasite is studied. Falcilysin has a role in parasite development during asexual blood stages [39]. Absence of PfM18AAP is shown to impart lethality in parasites using knockdown experiments. Inhibition of Pf-AM1 and PfLAP with specific inhibitors also has antimalarial effect [40,41]. All these comparative studies strongly support our vision toward PFI1625c as a drug target. Hence, PFI1625c exhibits a number of differences from host protease in terms of active site environment, substrate specificity and these properties are sufficient to exploit PFI1625c as drug targets.

#### 4. Discussion

Characterizing and validating proteins to examine their potential as a drug target is the need of the hour to develop novel

malarial chemotherapy. Protein-peptide docking is promising ultimate tool to decipher substrate or ligand specificity. Computational approaches to determine specificity in a peptide substrate toward HIV-1 protease indicate a correlation between substrate specificity and its binding energy with the protease [42]. In another study, protein-protein complexes are used to identify the determinants that are crucial for complex formation [43].

Proteases played a significant role in the malaria parasite life cycle and required for survival throughout its life cycle. Proteases present inside the food vacuole (FV) are involved in hemoglobin digestion to provide amino acid for protein synthesis [44]. Besides basic metabolic pathways, protease in the parasite is known to regulate invasion, egress and other molecular events [8,11,16,20,44]. A recent Bioinformatics analysis explored 92 proteases within plasmodium genome and a number of these uncharacterized proteases have potential to be utilized as an excellent drug target [12]. Potentially important proteases which remain uncharacterized consist of the calpain type proteases, metacaspase, primary processing proteases or signal peptidase. PFI1625c, an organeller processing peptidase is proposed to play an important role in protein targeting. Transport of proteins requires processing of signal sequence by signal peptidase for delivery of protein to reach their destinations [16]. PFI1625c has high similarity with other known metalloproteases with a conserved metal binding motif HXXEH [38]. A typical



**Fig. 7.** Proposed catalytic mechanism of PFI1625c mediated peptide cleavage. Based on the stereospecific positions of different residues within the PFI1625c active site, a schematic cleavage mechanism is proposed.

Zn<sup>2+</sup> dependent metalloprotease follows a mechanism involving different distinct reaction intermediate [45]. Structural determinants play a crucial role in substrate recognition and specificity [46].

Patch-dock was used as a tool to generate PFI1625c-peptide molecular models, as control docking of signal peptide (LSRVAKRA) in yeast mitochondrial processing peptidase (PDB Code 1HR9) gives complex (yeastMPP-peptide) close to known structure [25]. The root mean square deviation between generated model and crystallographic complex (1HR9) is 1.52 Å. To further validate the approach, docking was also performed by taking matrikyisin and its peptide substrates from MEROPS database [34] and through literature search [35]. The docking experiments showed a strong correlation ( $r=0.941$ ) between the affinity of the peptide substrate ( $K_m$ ) to the atomic contact energy (Fig. S2). Correlation analysis of PFI1625c-peptide complexes indicates a preference of PFI1625c toward hydrophobic residues in the substrate due to the local environment. Protein-peptide docking is a powerful tool to predict the binding affinities, specificity and such a study has immense potential in drug discovery by utilizing bioactive peptides [47,48].

Based on the stereospecific positions of different residues within the PFI1625c active site, a schematic cleavage mechanism is given in Fig. 7. Comparing PFI1625c with metalloprotease shows that E-136 and D-140 present near the zinc binding site to form the S<sub>2</sub> and S<sub>3</sub> sites. The negatively charged sites form a salt-bridge with an arginine present in an incoming peptide to facilitate strong binding (Fig. 3). The F-53 lies at the S<sub>1'</sub> site and is believed to interact with the aromatic residues found in the substrate. The residue E-49 which lies within hydrogen bonding distance to a water molecule coordinated to zinc, is predicted to polarize this water molecule,

thereby aligning it for nucleophilic attack on the carbonyl carbon of the peptide bond of the substrate (Fig. 7). The active site of PFI1625c or other metalloproteases differs from thermolysin in the absence of analogous residues Y-157 and H-231 which are proposed to stabilize the oxyanion of a tetrahedral intermediate. Solvent molecules or similar charged amino acids T-79 and R-125 could be involved in hydrogen bonding to stabilize oxyanion of a tetrahedral intermediate in PFI1625c (Fig. 7).

All peptides fitting well into the PFI1625c active site and stereochemistry indicate that most likely cleavage site is a peptide bond between A8 and positively charged R7, as positive charge can be stabilized by negatively charged patch within the catalytic site. Substitution of amino acids in the peptide occupying the hydrophobic cavity with non-polar residues increases the affinity of bound peptides within the PFI1625c active site due to the presence of a hydrophobic pocket.

## Acknowledgements

This work was partially supported by the Department of Biotechnology; Govt of India Grants (BT/PR13436/MED/12/450/2009) to V.T. KL acknowledges the financial support in the form of fellowship from Indian Institute of Technology-Guwahati, Assam, India.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2013.03.008>.

## References

- [1] S.I. Hay, E.A. Okiro, P.W. Gething, A.P. Patil, A.J. Tatem, C.A. Guerra, et al., Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007, *PLOS Medicine* 7 (2010) e1000290.
- [2] S. Mok, M. Imwong, M.J. Mackinnon, J. Sim, R. Ramadoss, P. Yi, et al., Artemisinin resistance in *Plasmodium falciparum* is associated with an altered temporal pattern of transcription, *BMC Genomics* 12 (2011) 391.
- [3] H. Cai, R. Kuang, J. Gu, Y. Wang, Proteases in malaria parasites—a phylogenomic perspective, *Curr Genomics* 12 (2011) 417–427.
- [4] P.K. Harris, S. Yeoh, A.R. Dluzewski, R.A. O'Donnell, C. Withers-Martinez, F. Hackett, et al., Molecular identification of a malaria merozoite surface sheddase, *PLoS Pathogens* 1 (2005) 241–251.
- [5] S. Singh, M. Plassmeyer, D. Gaur, L.H. Miller, Mononeme: a new secretory organelle in *Plasmodium falciparum* merozoites identified by localization of rhomboid-1 protease, *Proceedings of the National Academy of Science of the United States of America* 104 (2007) 20043–20048.
- [6] E. Roggwiller, M.E. Betouille, T. Blisnick, C. Braun Breton, A role for erythrocyte band 3 degradation by the parasite gp76 serine protease in the formation of the parasitophorous vacuole during invasion of erythrocytes by *Plasmodium falciparum*, *Molecular and Biochemical Parasitology* 82 (1996) 13–24.
- [7] M. Allary, J. Schrevel, I. Florent, Properties, stage-dependent expression and localization of *Plasmodium falciparum* M1 family zinc-aminopeptidase, *Parasitology* 125 (2002) 1–10.
- [8] R. Banerjee, J. Liu, W. Beatty, L. Pelosof, M. Klemba, D.E. Goldberg, Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine, *Proceedings of the National Academy of Science of the United States of America* 99 (2002) 990–995.
- [9] K.K. Eggleson, K.L. Duffin, D.E. Goldberg, Identification and characterization of falcylisin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*, *Journal of Biological Chemistry* 274 (1999) 32411–32417.
- [10] M. Klemba, I. Gluzman, D.E. Goldberg, A *Plasmodium falciparum* dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation, *Journal of Biological Chemistry* 279 (2004) 43000–43007.
- [11] P.S. Sijwali, P.J. Rosenthal, Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by *Plasmodium falciparum*, *Proceedings of the National Academy of Science of the United States of America* 101 (2004) 4384–4389.
- [12] Y. Wu, X. Wang, X. Liu, Y. Wang, Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite, *Genome Research* 13 (2003) 601–616.
- [13] P. Prabhu, V. Patravale, Novel targets for malaria therapy, *Current Drug Targets* 12 (2011) 2129–2143.
- [14] B. Meslin, H. Zalila, N. Fasel, S. Picot, A.L. Bienvenu, Are protozoan metacaspases potential parasite killers? *Parasites & Vectors* 4 (2011) 26.
- [15] P. Olaya, M. Wasserman, Effect of calpain inhibitors on the invasion of human erythrocytes by the parasite *Plasmodium falciparum*, *Biochimica et Biophysica Acta* 1096 (1991) 217–221.
- [16] J.A. Boddey, A.N. Hodder, S. Gunther, P.R. Gilson, H. Patsiouras, E.A. Kapp, et al., An aspartyl protease directs malaria effector proteins to the host cell, *Nature* 463 (2010) 627–631.
- [17] F.S. Boretti, P.W. Buehler, F. D'Agnillo, K. Kluge, T. Glaus, O.I. Butt, et al., Sequestration of extracellular hemoglobin within a haptoglobin complex decreases its hypertensive and oxidative effects in dogs and guinea pigs, *Journal of Clinical Investigation* 119 (2009) 2271–2280.
- [18] A. Pamplona, T. Hanscheid, S. Epiphanio, M.M. Mota, A.M. Vigario, Cerebral malaria and the hemolysis/methemoglobin/heme hypothesis: shedding new light on an old disease, *International Journal of Biochemistry & Cell Biology* 41 (2009) 711–716.
- [19] R. Tuteja, A. Pradhan, S. Sharma, *Plasmodium falciparum* signal peptidase is regulated by phosphorylation and required for intra-erythrocytic growth, *Molecular and Biochemical Parasitology* 157 (2008) 137–147.
- [20] C.Y. He, M.K. Shaw, C.H. Fletcher, B. Striepen, L.G. Tilney, D.S. Roos, A plastid segregation defect in the protozoan parasite *Toxoplasma gondii*, *EMBO Journal* 20 (2001) 330–339.
- [21] G.G. van Dooren, V. Su, M.C. D'Ombrain, G.I. McFadden, Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme, *Journal of Biological Chemistry* 277 (2002) 23612–23619.
- [22] Z. Bozdech, J. Zhu, M.P. Joachimiak, F.E. Cohen, B. Pulliam, J.L. DeRisi, Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray, *Genome Biology* 4 (2003) R9.
- [23] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Research* 22 (1994) 4673–4680.
- [24] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution* 4 (1987) 406–425.
- [25] A.B. Taylor, B.S. Smith, S. Kitada, K. Kojima, H. Miyaura, Z. Otwowski, et al., Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences, *Structure* 9 (2001) 615–625.
- [26] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *Journal of Molecular Biology* 234 (1993) 779–815.
- [27] G.N. Ramachandran, C. Ramakrishnan, S. Sasisekharan, Stereochemistry of polypeptide chain configurations, *Journal of Molecular Biology* 7 (1963) 95–99.
- [28] R.A. Laskowski, D.S. Moss, J.M. Thornton, Main-chain bond lengths and bond angles in protein structures, *Journal of Molecular Biology* 231 (1993) 1049–1067.
- [29] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of non-bonded atomic interactions, *Protein Science* 2 (1993) 1511–1519.
- [30] R. Luthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, *Nature* 356 (1992) 83–85.
- [31] Kuntal, B.K. Aparoy, P. Reddanna, P. EasyModeller, A graphical interface to MODELLER, *BMC Research Notes* 3 (2010) 226.
- [32] Schneidman-Duhovny, D. Inbar, Y. Nussinov, R. Wolfson, H.J. PatchDock, SymmDock: servers for rigid and symmetric docking, *Nucleic Acids Research* 33 (2005) W363–W367.
- [33] N. Andrusier, R. Nussinov, H.J. Wolfson, FireDock: fast interaction refinement in molecular docking, *Proteins* 69 (2007) 139–159.
- [34] N.D. Rawlings, D.P. Tolle, A.J. Barrett, MEROPS: the peptidase database, *Nucleic Acids Research* 32 (2004) D160–D164.
- [35] H.I. Park, B.E. Turk, F.E. Gerkema, L.C. Cantley, Q.X. Sang, Peptide substrate specificities and protein cleavage sites of human endometase/matrilyn-2/matrix metalloproteinase-26, *Journal of Molecular Biology* 277 (2002) 35168–35175.
- [36] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, *Journal of Computational Chemistry* 26 (2005) 1701–1718.
- [37] K.C. Chou, H.B. Shen, ProtIdent: a web server for identifying proteases and their types by fusing functional domain and sequential evolution information, *Biochemical and Biophysical Research Communications* 376 (2008) 321–325.
- [38] A.B. Becker, R.A. Roth, Identification of glutamate-169 as the third zinc-binding residue in proteinase III, a member of the family of insulin-degrading enzymes, *Biochemistry Journal* 292 (Pt 1) (1993) 137–142.
- [39] M. Ponpuak, M. Klemba, M. Park, I.Y. Gluzman, G.K. Lamppa, D.E. Goldberg, A role for falcylisin in transit peptide degradation in the *Plasmodium falciparum* apicoplast, *Molecular Microbiology* 63 (2007) 314–334.
- [40] S. McGowan, C.J. Porter, J. Lowther, C.M. Stack, S.J. Golding, T.S. Skinner-Adams, et al., Structural basis for the inhibition of the essential *Plasmodium falciparum* M1 neutral aminopeptidase, *Proceedings of the National Academy of Science of the United States of America* 106 (2009) 2537–2542.
- [41] S. Maric, S.M. Donnelly, M.W. Robinson, T. Skinner-Adams, K.R. Trenholme, D.L. Gardiner, et al., The M17 leucine aminopeptidase of the malaria parasite *Plasmodium falciparum*: importance of active site metal ions in the binding of substrates and inhibitors, *Biochemistry* 48 (2009) 5435–5439.
- [42] S. Chaudhury, J.J. Gray, Identification of structural mechanisms of HIV-1 protease specificity using computational peptide docking: implications for drug resistance, *Structure* 17 (2009) 1636–1648.
- [43] I.H. Moal, R. Agius, P.A. Bates, Protein-protein binding affinity prediction on a diverse set of structures, *Bioinformatics* 27 (2011) 3002–3009.
- [44] I.Y. Gluzman, S.E. Francis, A. Okman, C.E. Smith, K.L. Duffin, D.E. Goldberg, Order and specificity of the *Plasmodium falciparum* hemoglobin degradation pathway, *Journal of Clinical Investigation* 93 (1994) 1602–1608.
- [45] V. Pelmenschikov, M.R. Blomberg, P.E. Siegbahn, A theoretical study of the mechanism for peptide hydrolysis by thermolysin, *Journal of Biological Inorganic Chemistry* 7 (2002) 284–298.
- [46] M. Yokoyama, T. Oka, H. Kojima, T. Nagano, T. Okabe, K. Katayama, et al., Structural basis for specific recognition of substrates by saposin virus protease, *Frontiers in Microbiology* 3 (2012) 312.
- [47] J. Audie, J. Swanson, Advances in the prediction of protein-peptide binding affinities: implications for peptide-based drug discovery, *Chemical Biology & Drug Design* 81 (2012) 50–60.
- [48] J. Audie, J. Swanson, Recent work in the development and application of protein-peptide docking, *Future Medicinal Chemistry* 4 (2012) 1619–1644.