Comparing sequence and structure of falcipains and human homologs at prodomain and catalytic active site for malarial peptide-based inhibitor design

Thommas M. Musyoka, Joyce N. Njuguna and Özlem Tastan Bishop*

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, 6140, South Africa

Corresponding author details:

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, P.O. Box 94, Grahamstown, 6140, South Africa

Tel: +27-466-038-072

E-mail address: O.TastanBishop@ru.ac.za

Abstract

Falcipains are major cysteine proteases of *Plasmodium falciparum* essential in hemoglobin digestion. Several inhibitors blocking their activity have been identified, yet none of them has been approved for malaria treatment. For selective therapeutic targeting of these plasmodial proteases, identification of sequence and structure differences with homologous human cathepsins is necessary. The protein substrate processing activity of these proteases is tightly controlled in space and time via a prodomain segment occluding the active site making it inaccessible. Here, we utilised *in silico* approaches to determine sequence and structure variations between the prodomain regions of plasmodial proteins and human cathepsins. Hot spot residues, key for maintaining structural integrity of the prodomains as well as conferring their inhibitory activity, were identified via residue interaction analysis. Information gathered was used to design short peptides able to mimic the prodomain activity on plasmodial proteases whilst showing selectivity on human cathepsins. Inhibitory potency was highly dependent on peptide amino acid composition and length. Our current results show that despite the conserved structural and catalytic mechanism of human cathepsins and plasmodial proteases, significant differences between the two groups exist and may be valuable in the development of novel antimalarial peptide inhibitors.

Keywords

Cysteine protease, Falcipain, Zymogen, Prodomain inhibitory segment, Homology modelling, Binding affinity

Abbreviations

Å	Angstrom
AAI	Amino acid interaction
BIC	Bayesian Information Criterion
BP-2	Berghepain 2
CP-2	Chabaupain 2
Cat-K	Cathepsin K
Cat-L	Cathepsin L
Cat-S	Cathepsin S
FPs	Falcipains
FP-2	Falcipain 2
FP-3	Falcipain 3
GRAVY	Grand average of hydropathy index
K_d	Dissociation constant
KP-2	Knowlesipain 2
KP-3	Knowlesipain 3
MAST	Motif Alignment Search Tool
MEGA	Molecular Evolutionary Genetic Analysis
MEME	Multiple Em for Motif Elucidation
Mr	Molecular weight
MSA	Multiple sequence alignment
NCBI	National Center for Biotechnology Information
NNI	Nearest-Neighbor-Interchange
PIC	Protein Interaction Calculator
pI	Isoelectric point
PlasmoDB	Plasmodium genome Database

PRODIGY	PROtein binDIng enerGY prediction
PROMALS3D	PROfile Multiple Alignment with predicted Local structures and 3D constraints
PSI-BLAST	Position-Specific Iterative Basic Local Alignment Search Tool
RBC	Red Blood Cell
VP-2	Vivapain 2
VP-3	Vivapain 3
YP-2	Yoelipain 2
Z-DOPE	Normalized Discrete Optimized Protein Energy
ΔG	Binding affinity
3D	Three dimensional

Running Title

Falcipains as malarial drug targets

Introduction

Malaria, caused by parasites from the genus *Plasmodium* and transmitted to human by a female anopheles mosquito bite, is still a devastating disease even though the global incidences have drastically dropped in recent years [1]. Parallel to evolving mosquito resistant to insecticides [1–4], continuously emerging resistant strains of parasite to current drugs [5–8] present an immense challenge for the eradication of malaria. A recent study promisingly showed that pre-existing resistance may not be a major problem for novel-target antimalarial candidates, and fast-killing compounds may result in a slower onset of clinical resistance [9]. Hence, the identification and development of alternative anti-malarial inhibitors with novel mode of action against new as well as known drug targets with certain key features are very important.

Proteases are considered as good parasitic drug targets and details are presented in a number of articles [10–16]. Cysteine proteases have a central role in *Plasmodium* parasites during hemoglobin degradation [17,18], tissue and cellular invasion [19], activation of pro-enzymes [20,21], immunoevasion and egression [11,21,22]. Red blood cell (RBC) invasion and rupturing processes as well as intermediate events involving hemoglobin metabolism are characterised by increased proteolytic activity. During the asexual intraerythrocytic stage, *Plasmodium* parasites degrade nearly 75% of host RBC hemoglobin [23,24] to acquire nutrients as they lack a *de novo* amino acid biosynthetic pathway. By this process, they can acquire all their amino acid requirements necessary for growth and multiplication with an exception of isoleucine which is exogenously imported as it is absent in human hemoglobin [10,25,26]. Hemoglobin degradation is an intricate and efficient multistage protein catabolic process occurring inside the acidic food vacuole [18,27].

This study focuses on a subgroup of papain-like Clan CA plasmodial cysteine proteases, namely Falcipains (FPs) of *P. falciparum* and their homologs. *P. falciparum* has four FPs; FP-1, FP-2, FP-2' and FP-3. FP-1 is the most conserved protease among the four proteases, and its role in parasite entry into RBCs is yet to be resolved. Although its inhibition using specific peptidyl epoxides blocked erythrocyte invasion by merozoites [28], FP-1 gene disruption in blood stage parasites does not affect their growth [29,30]. Despite its biological function remaining uncertain, FP-2' is biochemically similar to FP-2 and shares 99% sequence identity [22,31]. FP-2 (FP-2') and FP-3 share 68% sequence identity and are the major cysteine proteases involved in hemoglobin degradation in the parasite [32–35]. Expression of these proteins during the blood stage by plasmodia is strictly regulated in a site-specific and time-dependent manner [28,36,37]. These hemoglobinases have differential expression timing during the trophozoite stage: the early phase is characterised by FP-2 abundance while FP-3 is abundant at the late stages [17,22]. It was shown that targeted disruption of FP-2 gene in plasmodia results in accumulation of undigested hemoglobin in the food vacuole and its enlargement [17], therefore the protein can be considered as a promising drug target [38,39]. On the other hand, inhibiting individual proteases might not be essential due to redundancy in the hemoglobin digestion stage [10], hence any inhibitor design for FPs should consider blocking the activity of both FP-2 and FP-3. The importance of FP-2 as a drug target was also indicated in a recent study in which FP-2 polymorphisms were shown that are associated with artemisinin resistance [40].

Other *Plasmodium* species also express proteins highly homologous to FP-2 and FP-3 [41–44]. These include vivapains (vivapain 2 [VP-2] and vivapain 3 [VP-3]), knowlesipains (knowlesipain 2 [KP-2] and knowlesipain 3 [KP-3]), berghepain 2 [BP-2], chabaupain 2 [CP-2] and yoelipain 2 [YP-2] from *P. vivax, P. knowlesi, P. berghei, P. chabaudi* and *P. yoelii* respectively. All these proteins are related both in sequence and function to the papain-like

class of enzymes including human cathepsins. The plasmodial proteases have, however, unusual features compared to the human ones including, much longer prodomains and specific inserts in the catalytic domain - a “nose” (~ 17 amino acids) and an “arm” (~ 14 amino acids) [37,45,46]. In native environment, cysteine proteases are regulated either by their prodomain (zymogen form) or by other endogenous macromolecules like cystatins [47,48] and chagasin [49]. During erythrocyte entry, *P. falciparum* parasites secrete falstatin, a potent picomolar inhibitor of both FP-2 and FP-3 thus regulating the activity of these proteases on important surface proteins required for invasion [19,48]. In the zymogen form (Figure 1), a part of the prodomain flips over the active pocket and its subsites located on the catalytic domain [50], blocking its enzyme activity [51]. The acidic environment within a food vacuole (plasmodia) or lysosome (humans) triggers prodomain cleavage thus activating the catalytic domain [52,53].

The literature comprises a large number of inhibitor studies against FPs including peptide-based [31,54–56], non-peptidic [50,57–61] and peptidomimetic [58,62,63] studies. Hitherto, none of these inhibitors has been approved as an antimalarial drug as they have limited selectivity against host cathepsins, homologs to the parasites proteases. To overcome this, distinctive features between these two classes of proteins must be determined. Primarily, the current work utilises *in silico* approaches to characterize FP-2 and FP-3, their homologs from other *Plasmodium* species as well as human homologs (cathepsins) to identify sequence, physicochemical and structure differences that can be exploited for peptide-based antimalarial drug development. Although the two protein classes share high similarity, important differences that can be essential for inhibitor selectivity exist [50,64]. Our main aim in this study is to elucidate the inhibitory mechanism of plasmodial prodomain region responsible for endogenous regulation of the catalytic domain, information which may be useful in the design of novel inhibitors. For this purpose, using domain-domain interaction

approaches, specific hot spot residues critical for the mediation of the prodomain inhibitory effect were identified.

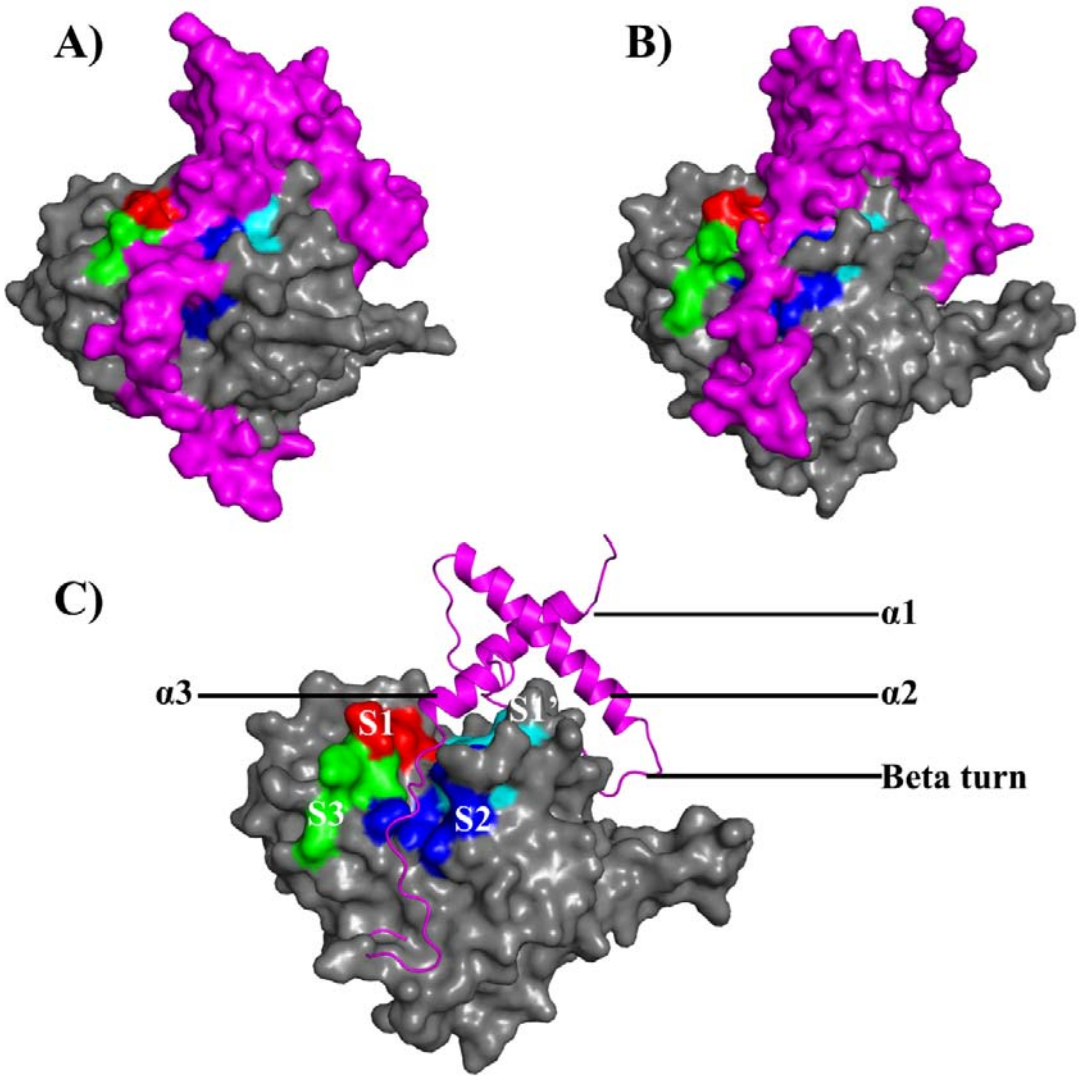


Figure 1. Clan CA cysteine protease zymogen prodomain-catalytic domain interaction modes. Surface representation of A) human Cat-K and B) FP-2. C) FP-2 prodomain structural elements (pink; in cartoon representation) interacting with the S1 (red), S2 (blue), S3 (green) and S1' (cyan) subsites of the catalytic domain.

To further identify a potential peptide segment, which could strongly bind to the plasmodial catalytic domains and mimic the native prodomain inhibitory effect, five short peptide sequences based on the identified hot spot residues were suggested. Flexible docking of these peptides against the catalytic domains identified a short 13-mer oligopeptide with preferential

binding towards plasmodial proteases. This oligopeptide could be a starting platform for the development and testing of novel peptide based antimalarial therapies against plasmodial cysteine proteases.

Material and methods

A workflow consisting of the different methods, tools and databases used in this study is shown in Figure 2. Unless otherwise indicated, amino acid numbering is based on individual protein full length as listed in Table S1.

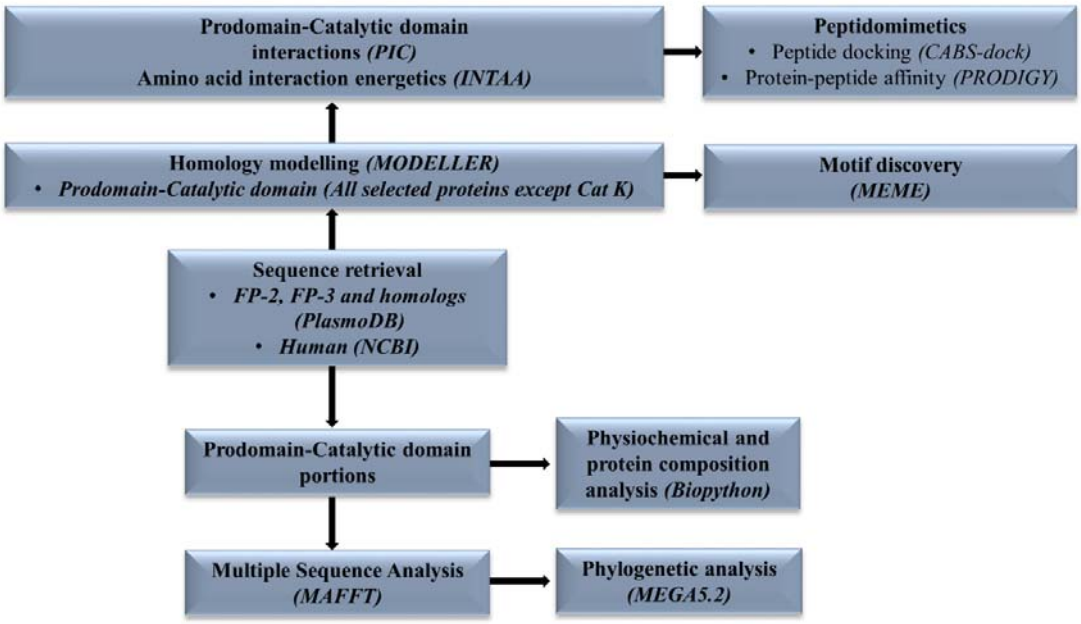


Figure 2. A graphical workflow of the methods and tools (in brackets) used in sequence and structural analysis of FP-2, FP-3 and their homologs.

Sequence retrieval and multiple sequence alignment

Using FP-2 (PF3D7_1115700) and FP-3 (PF3D7_1115400) as query sequences, seven plasmodial protein homologs together with three human homologs (Table 1) were retrieved from the PlasmoDB version 9.31 [65] and NCBI [66] databases respectively as described earlier [50]. A pronounced feature present in the cathepsin L (Cat-L) like plasmodial

proteases is the presence of an N-terminal signalling (non-structural) peptide sequence (~150 amino acids), which is responsible for targeting them into the food vacuole. For each of the plasmodial proteins, this segment was chopped off, and the remaining prodomain-catalytic portion saved into a Fasta file (Text S1). As guided by the partial zymogen complex crystal structure of Cat-K [PDB: 1BY8], ~ 21 amino acids (N-terminal) were also chopped off from the human cathepsin prodomain sequences. Together, these sequences were used in the rest of the study, and are referred as "partial zymogen" or "prodomain-catalytic domain" sequences interchangeably in the manuscript. Position details of the prodomain and catalytic portions per protein are listed in Table S1. To determine the conservation of the prodomain-catalytic portion, multiple sequence alignment (MSA) was performed using PROfile Multiple Alignment with predicted Local Structures and 3D constraints (PROMALS3D) web server [67] with default parameters except PSI-BLAST Expect value which was adjusted to 0.0001, and the alignment output visualised using JalView [68].

Phylogenetic inference

Using Molecular Evolutionary Genetic Analysis (MEGA) version 5.2 software [69], the evolutionary relationship of plasmodial proteases and human cathepsins was evaluated with the following preferences; Maximum Likelihood (statistical method) and Nearest-Neighbor-Interchange (NNI) as the tree inference option. A total of 48 amino acid substitution models were calculated for both complete (100%) and partial (95%) deletion and the best three models based on Bayesian Information Criterion (BIC) were selected (Table S2). For each selected model, the corresponding gamma (G) evolutionary distance correction value was selected to build different phylogenetic trees and comparison was made to determine robustness of dendrogram construction process. *Toxoplasma gondii* Cat-L [NCBI accession number: ABY58967.1] was included in the tree calculations as outgroup.

Physicochemical properties

Using an *ad hoc* Python and Biopython script, the amino acid composition and physicochemical properties, namely molecular weight (Mr), isoelectric point (pI), aromaticity, instability index, aliphatic index and grand average of hydropathy index (GRAVY) of the proteins were determined.

Motif analysis

Multiple Em for Motif Elicitation (MEME) standalone suite version 4.10.2 [70] was used to identify the composition and distribution of protein motifs within partial zymogen sequences. A Fasta file (Text S1) containing sequence information of the different proteins was parsed to MEME software with analysis preferences set as; -nostatus –time 18000 –maxsize 16000 –mod zoops –nmotifs X –minw 6 –maxw 50. The variable X (a whole number from 1) was varied until no more unique motifs were assessable as determined by motif alignment search tool (MAST) [71]. A heat map showing motif distribution was generated using an in house Python script. PyMOL was used to map the different motifs onto the protein structures (The PyMOL Molecular Graphics System, Version 1.6.0.0 Schrödinger, LLC).

Homology modelling and structure validation

MODELLER version 9.18 [72] was used to build homology models of the inhibitor complex of all proteins except for Cat-K which has already a crystal structure. Using a combination of templates, high quality prodomain-catalytic domain complexes of the plasmodial proteases as well as cathepsins (Cat-L and Cat-S) were calculated by MODELLER with refinement set to very slow. Table S3 shows the details of templates selected for each protein model. For the plasmodial proteases, the crystallographic structure of procathepsin L1 from *Fasciola hepatica* [PDB: 2O6X] was used as it had the highest similarity with most target sequences (30-38%) and high resolution of 1.40 Å. However, it lacked the arm (β-hairpin) region while

the nose residues were missing. To overcome these challenges, Cat-K [PDB: 1BY8] together with FP-2 [PDB: 2OUL] (for FP-2, VP-2, KP-2, BP-2 and YP-2) and FP-3 [3BWK] (for FP-3, VP-3, KP-3 and CP-2) were additionally used. For Cat-L and Cat-S, only two templates were used [PDB: 1BY8 and 2O6X]. For each protein, 100 models were calculated and ranked according to normalized discrete optimized protein energy (Z-DOPE) score [73]. The top three models per protein were further validated using ProSA [74], Verify3D [75], QMEAN [76] and PROCHECK [77] and the best quality model selected. Table 2 shows the quality scores obtained from different homology modelling assessment tools for the top model per protein. All validation methods gave consistently high quality scores for selected models and thus could be used for further experiments. QMEAN results showed that only small portions of the loop regions in Cat-L, Cat-S, and CP-2 were built with poor quality, while the majority of the prodomain-catalytic core region in all of the proteins was accurate (Figure 3 and Figure S1).

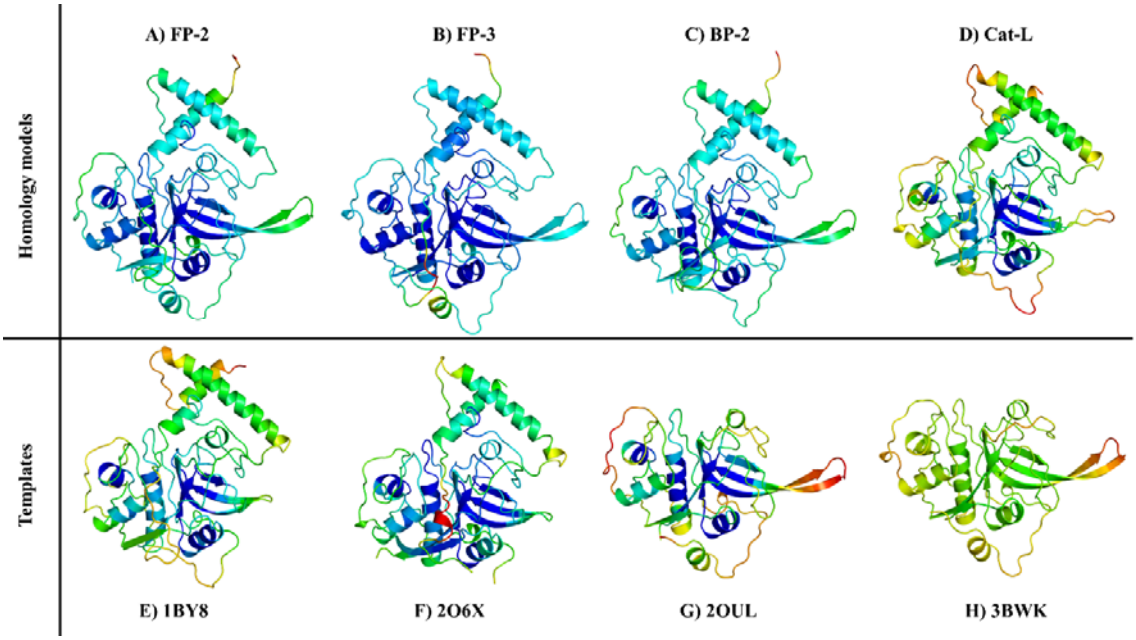


Figure 3. Homology models of different plasmodial proteases and human Cat-L together with the templates used in homology modelling. Colour code ranging from blue (accurate modelling) to red (poorly modelled regions).

As these loop regions were far from the catalytic pocket, the resulting models were considered acceptable for further analysis.

Prodomain-catalytic domain interaction studies and short inhibitor peptide design

To determine the prodomain inhibitory mechanism, residue interactions between prodomain and catalytic domain of plasmodial and human partial zymogen complexes were evaluated using the Protein Interaction Calculator (PIC) web server [78]. The interaction energy of identified residues was evaluated using the amino acid interaction (AAI) web server [79]. PyMOL was used to visualise the resulting interactions. For each protein, prodomain segment interacting with the catalytic domain's active pocket residues was identified and extracted into a Fasta file. From the interaction energies, residues within these inhibitory segments forming strong contacts with subsite residues were identified. Based on the identified hot spot residues, our next objective was to design short peptide(s) exhibiting the native prodomain effect whilst showing selectivity on human cathepsins. The conservation of prodomain inhibitory segments for all the proteins, and separately of only the plasmodial proteases, was determined using WebLogo server [80]. Peptides of varying lengths and composition based on amino acid conservation forming contacts with subsite residues were proposed. In order to evaluate the interaction of selected peptides on the catalytic domains, the prodomain segments of all proteins were chopped using PyMOL. Blind docking simulation runs of selected peptides were then performed on these sets of catalytic domains by CABS-dock protein-peptide docking tool [81] using the default parameters. To confirm the reliability of the results, docking experiments were repeated using catalytic domains of the same proteins that had been modelled and used in our previous studies [50]. Binding affinity (ΔG) and dissociation constant (K_d) for each protein-peptide complex was then evaluated using PROtein binDIng enerGY prediction (PRODIGY) web server [82].

Results and discussion

In this work, using combined *in silico* approaches the differences between falcipains and their plasmodial homologs as well as human cathepsins have been evaluated. Based on observed differences and interaction energy profiles between prodomain and catalytic domain subsite residues, short peptides that could mimic the native prodomain inhibitory mechanisms were proposed.

Both plasmodial and human cathepsins have similar physicochemical properties

Protein function is largely governed by its structure, amino acid composition as well as its environment. Despite the low sequence identity between the two subclasses (cathepsins and plasmodial proteases), physicochemical analysis revealed that they have similar aromaticity and grand average hydropathy (GRAVY) values indicating that both groups of proteins are hydrophilic (Table 3). With an exception of CP-2, all the other proteins have an instability index score of ≤ 40 and thus can be considered as being stable in test-tube environment [83]. Interestingly, there is no significant difference between the aromaticity, GRAVY and instability index scores of partial zymogen complex and individual catalytic domains either. However, significant differences exist in the molecular weight and isoelectric point (pI). Plasmodial partial zymogens have higher molecular weight than that of human cathepsins, as they have longer sequences (two additional structural catalytic domain inserts and longer prodomains). A key factor that controls the functioning of cysteine proteases is pH of the milieu in which they are found. All the plasmodial prodomain-catalytic complexes and Cat-L have a slightly acidic pI of 5.66 ± 0.37 with their catalytic domains exhibiting lower pI. The other cathepsins have basic pI for both their partial zymogen complexes and catalytic domains. This difference in pI profiles might explain the localization aspects of these proteins

where the plasmodial proteases and Cat-L are found in acidic food vacuoles and lysosomes respectively while the remaining cathepsins are predominantly found in extracellular matrix.

Plasmodial clan CA proteases and human cathepsins exhibit separate evolutionary clustering

In addition to the previous findings for catalytic domain conservation discussed in detail in ref [7], current MSA identified two highly conserved ERFNIN and GNFD motifs, which are located in the α2-helix and the adjacent downstream loop region between β turn and α3-helix respectively (Figure 1 and Figure 4).

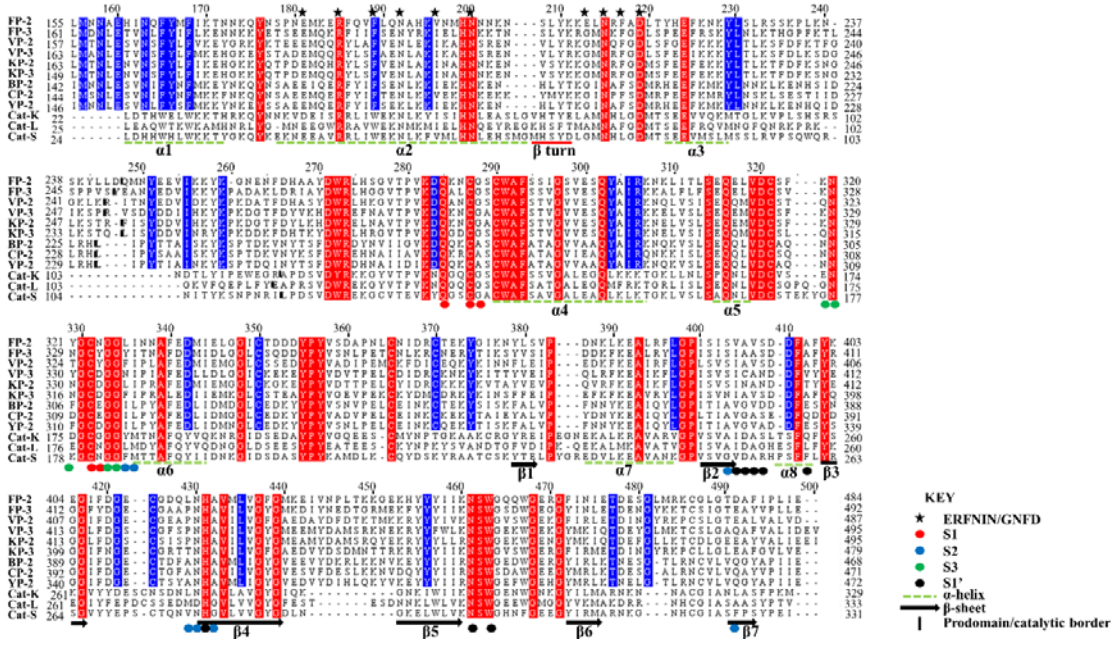
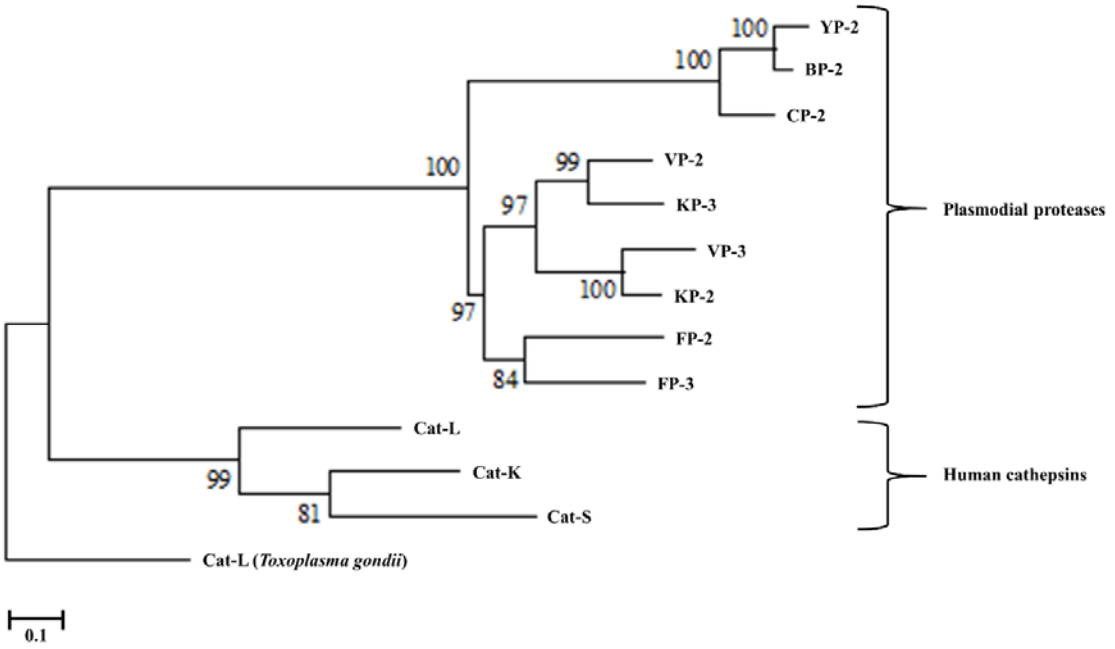


Figure 4. Structural-based multiple sequence alignment of FP-2, FP-3 and homologs prodomain-catalytic domains. Actual residue numbering per protein is given on the side, and the top numbering is based on partial zymogen alignment. The papain family characteristic prodomain ERFNIN and GNFD motif residues are indicated with an asterisk. Bold short lines depict prodomain-catalytic domain border. Dashed green lines indicate the position of α-helix and arrows β-sheet structural elements. Fully conserved residues in all the proteins are marked with red while residues only conserved in plasmodial proteases with blue. Position of subsite residues is shown with filled circles (Red=S1, Blue= S2, Green=S3 and S1'=black).

Despite the highly conserved nature of the ERFNIN motif across all the plasmodial proteins studied, FP-2 and CP-2 have Val residue in the place of Ile196 (numbering based on FP-2). In the human cathepsins, the motif's Phe190 (FP-2 numbering) is replaced by a Trp, a more

hydrophobic residue. Using site-directed mutagenesis, Kreusch *et al.*, identified two additional conserved Trp residues in human Cat-L (position 29 and 32 in Cat-L full length protein) which together with the highly conserved motifs (ERFNIN and GNFD) are important in the stability of the partial zymogen complex [84]. In plasmodial proteases, conservative substitution occurs on these two residues whereby they are replaced by less hydrophobic Phe residues (position 165 and 169 in FP-2). The contribution of these amino acid variations will be further discussed in the "*Prodomain regulatory effect mediated by α3 helix hydrophobic interactions with subsites S2 and S1' residues*" section. MSA result also revealed that cathepsins have a three amino acid insert in the α2 helix between the ERFNIN/GNFD motifs which is absent in the plasmodial proteases, and its importance is yet to be reported.

Phylogenetic analysis using partial zymogen sequences gave a distinct clustering between plasmodial proteins and human cathepsins forming two separate clades (Figure 5). There is no notable difference in tree topology in analysis performed using the catalytic domains only. This can be explained by the observed low sequence identity in both partial zymogen (Table 1) and catalytic domain sequences between the two groups of proteins [50]. The plasmodial proteases further clustered into two main subgroups based on the host. This is attributed to the previously reported sequence variations between the human and rodent plasmodial proteases [50]. FP-2 and FP-3 forms a separate sub-group from the other human plasmodial proteases possibly due to the high sequence similarity between the two proteins. The rate of mutation accumulation appears to vary between the two classes of proteins, being slowest in the human cathepsins. All human plasmodial proteases seem to evolve at the same rate as compared to the rodent orthologs which appear to show the highest substitution rate among all the proteins.



335

336 **Figure 5.** A phylogenetic tree of plasmodial and human FP-3, and FP-3 homologs
337 prodomain-catalytic protein sequences using MEGA5.2.2. The evolutionary history was
338 inferred by using the Maximum Likelihood method based on the Whelan and Goldman
339 model (WAG) model with a γ discrete distribution (+G) parameter of 2.4 and an evolutionary
340 invariable ([+I]) of 0.1. All positions with gaps were completely removed (100% deletion)
341 and bootstrap value set at 1,000. The scale bar represents number of amino acid substitutions
342 per site. *Toxoplasma gondii* CAT-L is used as the outgroup.
343
344 *Plasmodial proteases have unique motifs compared to human cathepsins*

345 Sequence motifs within proteins might be associated with a specific biological function. Thus
346 to better understand and characterise a group of proteins, identification of common and
347 distinguished motifs is of critical importance. A total of 13 unique motifs with varied
348 distributions were identified in the set of proteins studied (Figure 6A). These motifs were
349 then mapped onto the 3D structures of partial zymogen complexes (Figure 6B and 6C). Five
350 motifs (M1, M3, M5, M6 and M7) are present in both the plasmodial and human proteases.
351 Out of these five motifs, M1, M3, M5 and M7 are located at the catalytic domain of all
352 proteins while M6 is at α3-helix region of the prodomain (Figure 6B and 6C). Up to three
353 motifs; M2, M4 (located in α1-helix) and M8 (nose region) are only found within the
354 plasmodial proteases, except FP-2 lacks M8. A differential motif composition of the anterior

prodomain region (α1- α3 helix) of the two classes of proteins was observed with one long motif (M4) in plasmodial proteases while human cathepsins have two (M10 and M12).

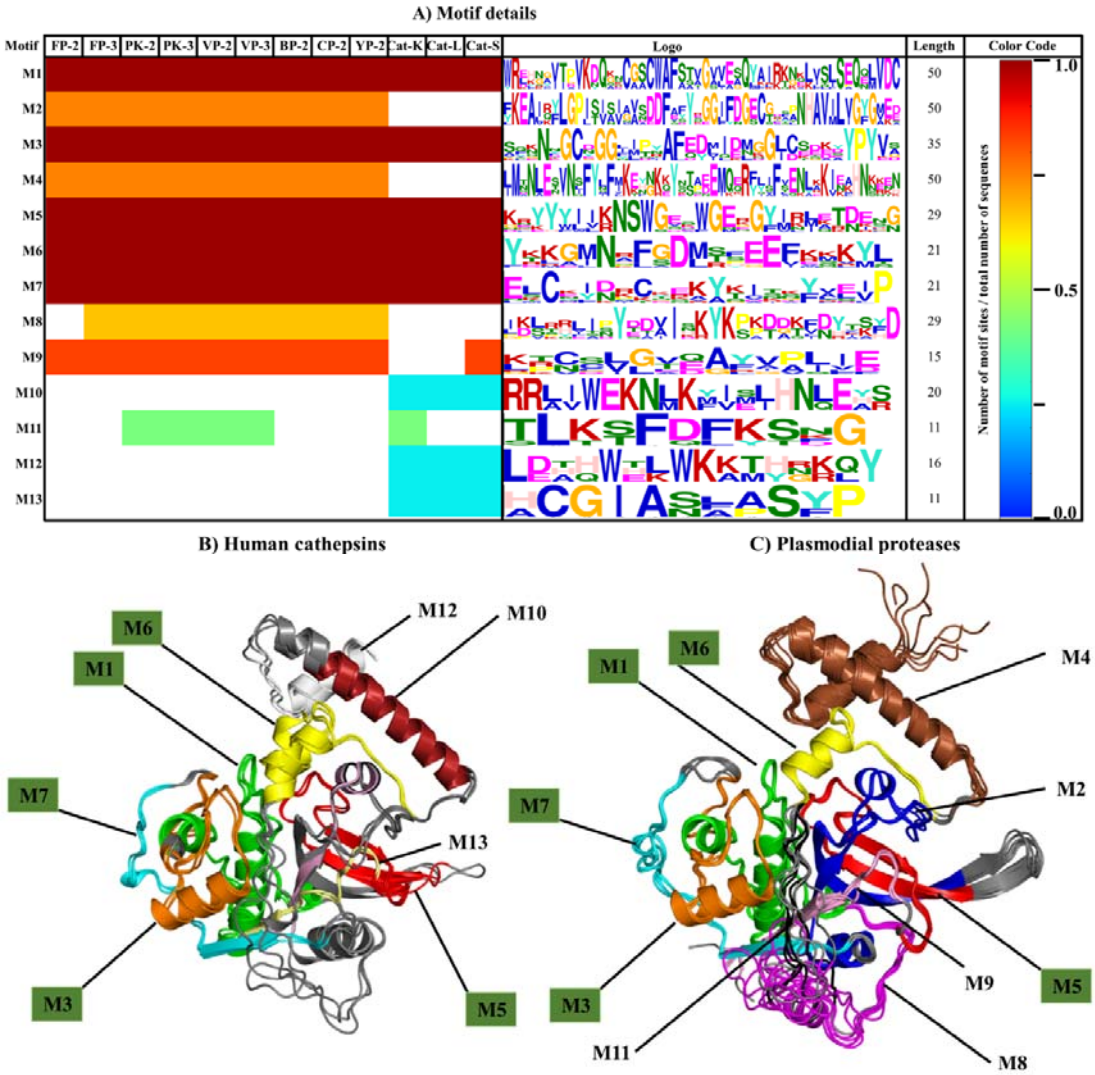


Figure 6. Motif analysis of plasmodial proteases and human cathepsins partial zymogen domains. A) A heat map showing the distribution, level of conservation and information of different motifs found in plasmodial and human proteases studied. A cartoon presentation showing the location of all motifs within the prodomain-catalytic structural fold. Labelled in green boxes are motifs present in both (B) human cathepsins and (C) plasmodial proteases.

PROSITE [85] and MyHits [86] webservers were used to search for the functional importance of identified motifs. M1 (PF00112.15) is the peptidase_C1 functional site and consists of PS00139 (QQnCGSCWAfST-cysteine protease active site), PS00008 (GVvesSQ-

N-myristoylation site), and PS00006 (casein kinase II phosphorylation site). M2 (PF00112) is a characteristic functional site of papain-like family cysteine proteases located at the C-terminus (α7-helix to β4), and forms part of the arm region of plasmodial proteases. M3 is located in α6-helix, and the adjacent loop regions of all the Clan CA group of enzymes have no function assigned to it. M4 (PF08246) is known as the cathepsin propeptide inhibitor domain (Inhibitor I29), and is located at α1 and α2 helixes of the N-terminus. The other motifs had no defined function assigned to them according to these webservers.

Prodomain regulatory effect mediated by α3 helix hydrophobic interactions with subsites S2 and S1' residues

Different non-canonical interactions were identified between the prodomain and catalytic domain of proteins. These included hydrophobic, cation-π, ionic, aromatic-aromatic and hydrogen bonds. In all partial zymogen complexes studied, no disulphide linkages between the two domains were observed. The main interactions exhibited are hydrophobic and hydrogen bonds, which participated either in anchoring and maintaining the folding integrity of the prodomain segment, or in mediating its inhibitory effect by interacting with subsite residues (Table S4). Our residue interaction results revealed that prodomain anchoring residues are located on the region between α1-helix and the β-turn which interacted with β3 and part of the arm region in the catalytic domain. Additionally, the C-terminus of the prodomain interacts with the N-terminus of the catalytic domain and part of β3 (Figure 7). A strong hydrogen and hydrophobic interaction network running from the N-terminal end to the GNFD motif prodomain residues, possibly for maintaining its structural fold, was identified in all proteins. In comparison with the human cathepsins, the plasmodial proteases had longer N-terminal prodomain regions with a series of highly conserved residues *viz.* Met156, Asn158, Glu160 and Asn163 (FP-2 numbering). These residues formed a hydrophobic interaction network with bonds of the order < -10.0 kJ/mol. Two additional aromatic-

aromatic interactions between Phe165-Phe168 and Phe/Tyr166-Phe189 (FP-2 numbering) in all the plasmodial proteases form strong bonds with energies less than -20.0 kJ/mol and -10.0 kJ/mol respectively. In human cathepsins, the first three Phe positions are substituted with Trp while the fourth position has charged residue substitution (Arg) resulting in weak interactions. A strong residue interaction network between the ERFNIN-GNFD motifs exists in all proteins, confirming the importance of these two motifs in the stability of the prodomain.

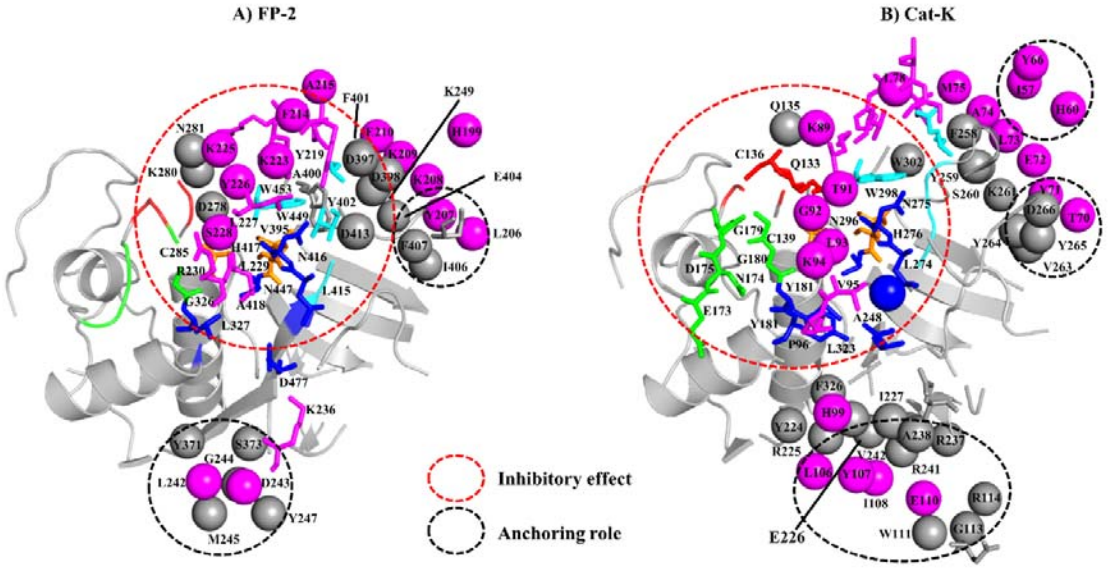


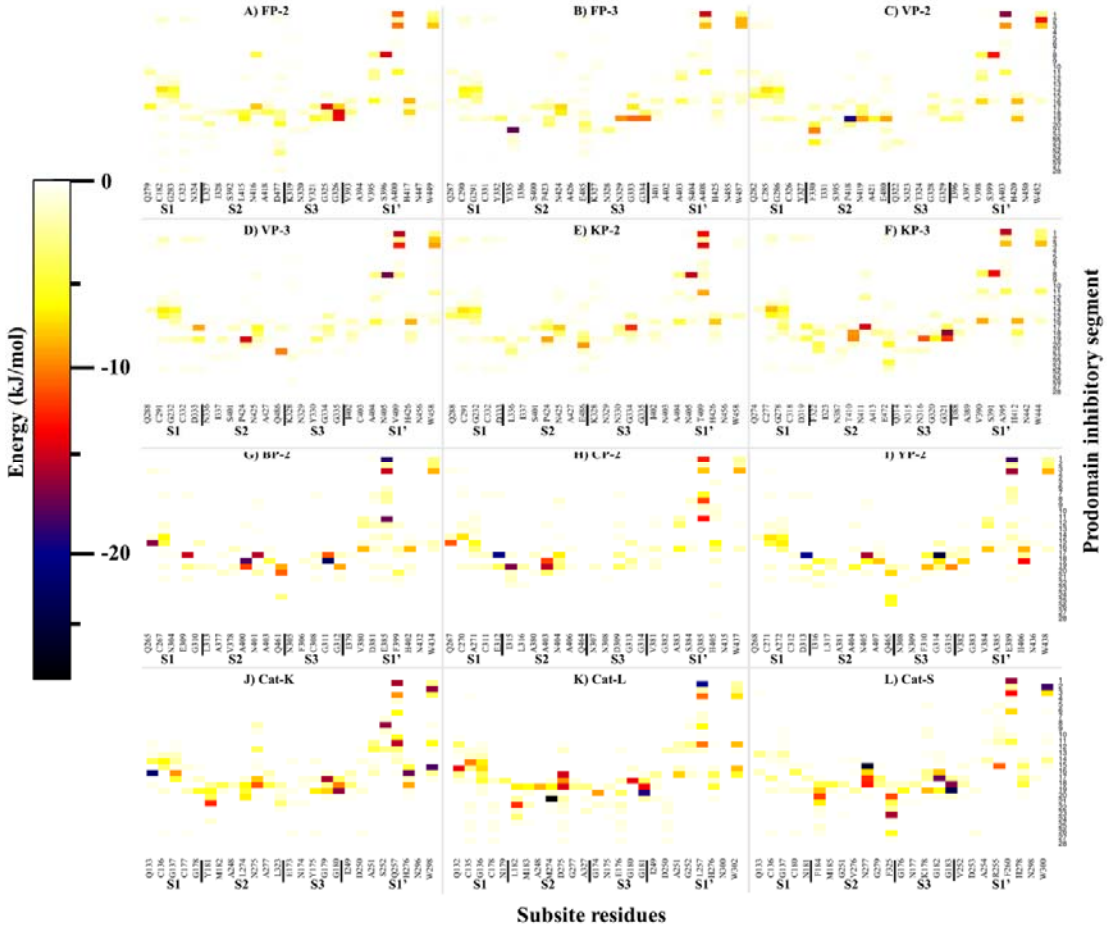
Figure 7. Prodomain-catalytic residue interaction network in (A) FP-2 and (B) Cat-K. For each protein full length residue numbering is used. Prodomain residues (magenta spheres) interacting with non-subsite residues (grey spheres) while sticks are prodomain residues interacting with subsite residues: S1 (red), S2 (blue), S3 (green) and S1' (cyan). Orange sticks are characteristic catalytic residues of the papain family of proteases (Cysteine, Histidine and Asparagine). Enclosed in red are prodomain-catalytic domain residues mediating inhibitory effect while those in black are involved in anchoring of the prodomain onto the catalytic domain.

A previous mutagenesis study on FP-2 identified two salt bridges (Arg185-Glu221 and Glu210-Lys403) that are important in the activation of the enzyme [87]. In our residue interaction analysis, Arg185 formed a stronger salt bridge with Asp216 (-21.2 kJ/mol) than with Glu220 (-9.5 kJ/mol). To validate these results, Asp216 and Glu220 were independently

413 mutated with an alanine residue and their interaction energy contribution with Arg185 was
414 determined. A complete loss of interaction for Glu70Ala mutation was observed (0.6 kJ/mol)
415 while Asp65Ala energy dropped by half to -12.5 kJ/mol, an indication that the ionic pair
416 between Arg31 and these two positions play a critical biological function. These two residues
417 are fully conserved in all of the proteins studied here. The second predicted salt bridge by
418 Glu210-Lys403 (FP-2 numbering) has high residue variation across the proteins. For the
419 charged Glu210 position in FP-2, all the other plasmodial proteases and Cat-S have a polar
420 residue (Gly) while the other cathepsins have a non-polar residue (Ala). The majority of the
421 residues in position Lys403 are mainly charged except KP-3, CP-2 and Cat-K in which have
422 a polar residue. The energetic contribution from the interactions forming this second salt
423 bridge were all weak (< -1.0 kJ/mol). However, PIC interaction results showed that position
424 209 in FP-2 consisted of highly conserved positively charged residue (mostly Lysine) across
425 the other plasmodial proteases which formed strong ionic contacts with Asp398 (fully
426 conserved in all plasmodial proteases) in the α8-helix region, an indication that the second
427 salt bridge was most likely formed by these residues. In addition, the mutagenesis study
428 identified aromatic-aromatic interactions in FP-2 between Phe214, Trp449 and Trp453 to be
429 also important in the activation. These residues were conserved in all proteins and formed
430 strong interactions, an indication that they are of functional importance as in FP-2.

431 A specific aim of this study was to determine the responsible residues that confer the
432 prodomain with its inhibitory function. From residue interaction results, only a small portion
433 of the prodomain (~22-mer) had significant contacts with individual protein subsite residues
434 and was responsible for the inhibitory effect (Figure 8). The main residues mediating the
435 inhibitory effect are located between the α3-helix and the inter-joining loop region, which
436 mostly interact with subsite S2 and S1' residues via hydrophobic interactions and hydrogen
437 bonds (Table S4). This correlates to our previous findings where residues forming these two

438 subsites were found to be critical in the inhibitory effect and selectivity using non-peptide
439 inhibitors [50].



440

441 **Figure 8.** A heatmap for residue interaction energies between prodomain inhibitory segment
442 and the catalytic subsite residues per protein. The inhibitory segment starts from the
443 conserved Asn residue in the GNFD motif (Figure 4).
444
445 A common interaction profile between the prodomain inhibitory segment and the catalytic
446 subsites of the different proteins is observed (Figure 8). For subsite S1, a limited residue
447 contact network was observed mainly with residues located at the α3-helix in all the
448 proteases. The C-terminal end of prodomain segment mainly exhibits contacts with S2 and S3
449 subsites, with human cathepsins and rodent plasmodial proteases forming stronger
450 interactions than the human plasmodial counterparts. In our previous study, high residue
451 variation across the proteins in S2 as well as S1'subsites was reported [50]. In all the proteins,

the first three prodomain inhibitory segment amino acids form strong hydrophobic contacts with residues at the opening of S1' subsite (Figure 8). Rodent plasmodial proteases have additional hydrogen bonding network, due to presence of charged residues at the fourth or fifth S1' position. From the interaction energy results, there are no observable contacts between residues Ala218 to Thr221 (FP-2 numbering) with any of the proteins' subsite residues. However, a strong hydrogen bonding network is formed between prodomain Ser231, Leu232, Arg233 with Leu429, Asn430 (S2), Val406, Ala412 (S1') and Gly334-335 (S3). Lys236 residue forms very strong ionic interactions with Asp491 (S2), a position mainly occupied by charged residues only in the human plasmodial proteases. The side-chain of Ser228 in FP-2 forms hydrogen bonding with thiol group of catalytic Cys285. A similar trend with other plasmodial proteases was observed (Table S4). From the interaction fingerprint, residues that are key in anchoring and maintaining the stability of the prodomain as well as mediating its catalytic domain regulatory effect were identified per protein.

Peptide inhibitory effect and selectivity dependent on composition and length

Despite their poor chemical properties, peptides remain a promising class of enzyme modulators as they are chemically diverse, highly specific and relatively safe [88,89]. Designing peptide based inhibitors requires prior understanding of how an enzyme recognizes its native peptide substrate then modifying the resulting interactions. Additionally, hot spot residues that regulate protein-protein/domain interactions may provide valuable insights. For FP-2, three peptide studies based on its prodomain-catalytic domain interaction network have already been performed. Rizzi *et al*., who designed peptidomimetics based on the interaction information between cystatin and FP-2 [90]. A major limitation of this study was that it was limited to FP-2 and the broad inhibitory potency of resulting cystatin mimics to other plasmodial proteases was necessary. Another study by Korde *et al*., using a synthetic 15-mer oligopeptide based on the N-terminal extension of FP-2 partial zymogen

(LMNNAEHINQFYMFI) showed that it could inhibit substrate processing activity of recombinant FP-2 *in vitro* [91]. However, our interaction fingerprint results showed that this terminal extension was not the native inhibitory segment and was not interacting with any of FP-2 catalytic domain subsite residues. Lastly, Pandey *et al.*, expressed the whole prodomain of FP-2 together with truncated segments and evaluated their inhibitory ability against a series of papain-family cysteine proteases. At the end, they determined that a FP-2 prodomain segment (Leu127-Asp243) which included the ERFNIN and GNFD motifs had a broad inhibitory activity against FP-3, BP-2, FP-2, Cat-L, Cat-B and cruzain [92]. Considering its length and molecular mass, the therapeutic potential of this peptide is uncertain.

In our study, peptides aimed at mimicking the inhibitory prodomain segment were designed and tested based on the identified prodomain-catalytic domain interaction fingerprint (Figure 8). Initially, a 22-mer peptide (peptide 1 = NRFGDLSFEEFKKKYLNLKLFD) based on the conservation of the prodomain segments responsible for the inhibitory mechanism for all the proteases was selected for docking against the catalytic domains of individual proteins using the CABS-dock webserver (Figure 9). CABS-dock performs blind docking simulations to identify the most probable binding site while maintaining the flexibility of the peptide ligand [81]. The ΔG of top protein-peptide complex model per protein was then determined using the PRODIGY server. A portion of this peptide interacted with active pocket residues of individual proteins and formed complexes exhibiting high binding affinities as that of a FP-2/Chagasin X-ray crystal complex [PDB: 2OUL] (Table 4). Despite the high predicted affinity scores with peptide 1, no differential binding was observed with the human cathepsins. As its N-terminus had highly conserved GNFD motif residues responsible for anchoring and maintaining the prodomain integrity, we chose to find out if a shorter peptide lacking these residues would bind differently. Thus, a different set of docking experiments with a peptide (peptide 2 = LTYHEFKNKYLSLRSSK) derived from the main inhibitory

segment of FP-2 was performed. Despite the variation in length, peptide 2 had similar results to peptide 1 and lacked differential binding affinity profile between the two protein classes.

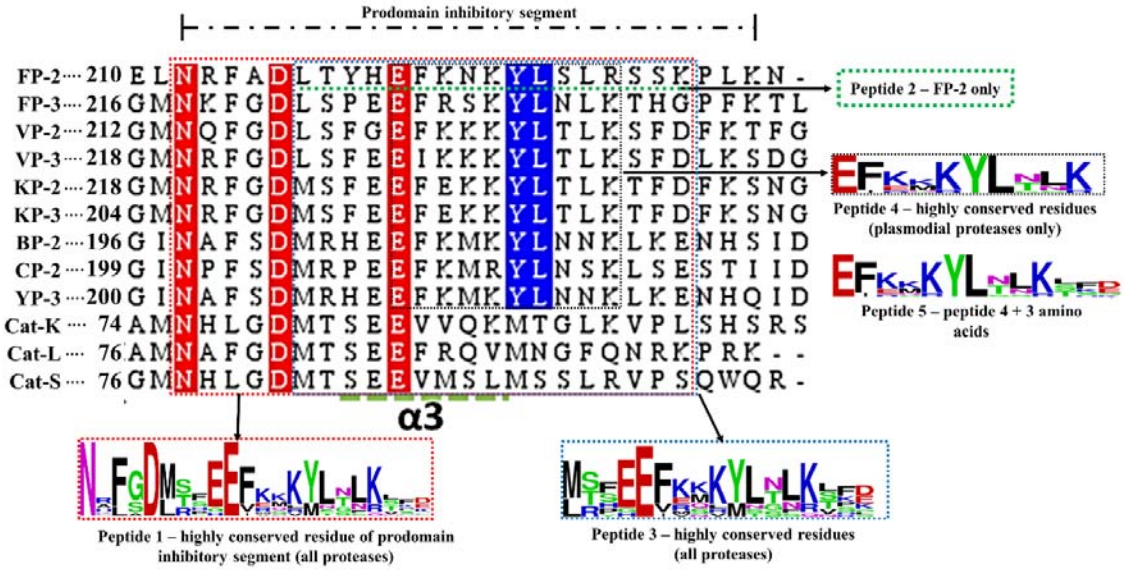
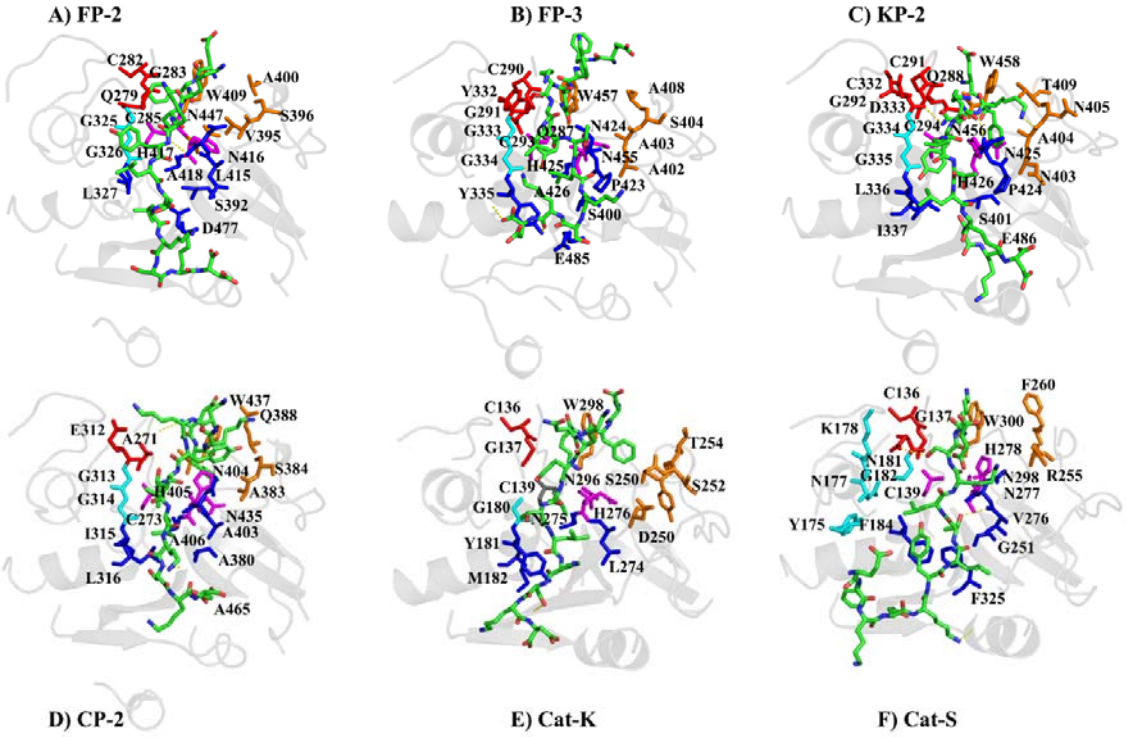


Figure 9. Sequence alignment of the prodomain inhibitory segment for the plasmodial and human cathepsin proteases studied. Marked sequence sections indicate the portions used to design different oligopeptides for docking studies and their conservation as determined by WebLogo server. Actual residue numbering per protein is given on the side.

A previous *in vitro* study by Pandey *et al.,* show that a FP-2 prodomain harbouring peptide 2 segment exhibited similar broad inhibitory activity on cruzain, Cat-B, Cat-L, FP-2, FP-3 and BP-2 [92]. However, from our energy interaction profiles, a large portion of the tested prodomain including the ERFNIN/GNFD motifs is mainly involved in anchoring it to the catalytic domain. Thus, the main inhibitory segment is much shorter and downstream of the GNFD motif. Peptide 2-YP-2 complex had the strongest binding association (-14.3 kcal/mol) while VP-2 had the lowest (-10.1 kcal/mol). With the already tested peptides exhibiting unselective high affinity binding on both human cathepsins and plasmodial proteases, additional docking experiments were performed with a different peptide derived from the most conserved residues in the same inhibitory segment as peptide2 from all proteases (peptide 3 = MTFEEFKQKYLTLKSKD). In some positions within the prodomain inhibitory

segments across the plasmodial proteases, high residue variations were observed and there was no consensus about which residue to include in the peptide. So the properties of the residues (polar, charged, non-polar, hydrophobic) occupying such positions were compared to determine if a common chemical property was preferred.. In addition, residues showing stronger interactions with the catalytic subsite residues were also taken into account. However, the ΔG between peptide 3 and plasmodial proteases was significantly lower in most plasmodial proteases than with the earlier peptides. Nevertheless, human cathepsins had similar binding affinity values with peptide 1 and 2. Guided by the residue interaction profile of prodomain residues with subsite residues (Figure 8 and S2), a fourth peptide (peptide 4 = EFKNKYLTLK) composed of the most conserved amino acids around α3-helix of the inhibitory segment of all plasmodial proteases was evaluated. A similar trend of non-selectivity was observed as with peptide 1, though with lower binding affinity. A fifth peptide, similar to peptide 4 except for its length, (EFKNKYLTLKSKD) was also evaluated. The residues in this peptide showed some conservation in the plasmodial proteases and had significant differences to the human cathepsins. Interestingly, it bound more strongly to all plasmodial proteases compared to the human cathepsins. A likely explanation of this differential binding affinity was that the peptide interacted with fewer residues on human cathepsins compared to the plasmodial proteins (Figure 10). In most of the plasmodial proteases, peptide 5 bound with almost same affinity as that of chagasin and FP-2 (-11.9 kcal/mol). From the prodomain-catalytic interaction analysis (Table 5 and Figure 10), the terminal end in peptide 5 interacts with last position of S2 which consists of a charged residue (only in human plasmodial proteases) forming a strong ionic interaction as well as other non-subsite residues thus forming a stronger complex. In most plasmodial proteases, peptide 5 formed multiple hydrogen bonds especially with S2 and S1' subsite residues. These two subsites residues have been found to be key in determining binding selectivity as they are the

546 main contributors to ligand binding [50]. Docking studies with previously modelled catalytic
547 domains gave results consistent with the current models.



548

549 **Figure 10**. Peptide 5-catalytic domain subsite residue interactions of various proteins
550 (Red=S1, Blue=S2, Cyan=S3, S1'=orange and Magenta=catalytic residues).
551
552 From the motif analysis (Figure 6), a large proportion of peptide 5 was represented in motif
553 M6. Despite the functional annotation of motif M4 indicating it as the cathepsin propeptide
554 inhibitor domain, majority of its residues were predominantly involved in anchoring the
555 prodomain. Taken together, our study is the first to identify the most key prodomain segment
556 involved in regulation of cysteine proteases, and to apply information based approaches to
557 propose a peptide with differential binding on both human and plasmodial proteases.

Conclusion

In the present study we aimed to characterise the differences between *P. falciparum* falcipains and their plasmodial and human homologs, especially where the prodomain interacts with the catalytic domain, in order to identify key residues which could be useful in antimalarial drug development approaches. This was done at both sequence and structure level. Through homology modelling, near native 3D partial zymogen complexes of both plasmodial and human proteases were obtained. This allowed structural characterization, thus deciphering how these segments confer their inhibitory mechanism endogenously. The main prodomain residues mediating the inhibitory effect were located in the α3-helix and the inter-joining loop region, and mostly interacted with subsite S2 and S1' residues. In our previous studies [50,57], we showed that residues forming these two subsites are critical in inhibitor design as they differ from human cathepsins. Hence, putting all the analysis together, with a continuous prodomain epitope mimicking strategy, a peptide which bound selectively more strongly on plasmodial proteases than the human ones was designed. The present approach offers a starting point which could lead to the establishment of novel antimalarial peptide drugs aimed at mimicking the natural plasmodial protease regulatory mechanism. Additional chemical modification either to obtain peptide derivatives with better physicochemical and pharmacokinetic properties as well as potency, bioavailability and stability might be necessary. Accessibility of parasite infected erythrocytes by macromolecules remains a major concern for the development of peptide based antimalarial inhibitors. A study by Farias *et al.*, using fluorescent peptides revealed that peptides with molecular weight up to 3146 Da can permeate into the blood stage parasites [93]. All the peptides determined had a mass of below 2753 Da, with peptide 5 having 1613 Da, an indication that it would readily be available inside the parasites. Korde *et al.*, demonstrated that a synthetic 15-mer oligopeptide of mass 1885 Da could localise into the intracellular compartments of trophozoites and schizoints

583 inhibiting FP-2 activity [91]. Additional modification of the peptide backbone as well as
584 amino acid side chains may also be performed yielding peptide based inhibitors.

Acknowledgement

This work is supported by the National Research Foundation (NRF) South Africa (Grant Numbers 93690 and 105267). T.M.M and J.N.N thank Rhodes University for the postgraduate financial support. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Author contributions

Ö.T.B conceived the project. T.M.M and J.N performed the experiments. All authors analysed the data. T.M.M and Ö.T.B wrote the article.

Disclosure statement

The authors declare no conflict of interest

References

1 WHO (2017) *World Malaria Report 2017*. Geneva.

2 Bass C & Jones CM (2016) Mosquitoes boost body armor to resist insecticide attack. *Proc. Natl. Acad. Sci.* **113**, 9145–9147.

3 Hemingway J, Ranson H, Magill A, Kolaczinski J, Fornadel C, Gimnig J, Coetzee M, Simard F, Roch DK, Hinzoumbe CK, Pickett J, Schellenberg D, Gething P, Hoppé M & Hamon N (2016) Averting a malaria disaster: will insecticide resistance derail malaria control? *Lancet* **387**, 1785–1788.

4 Alout H, Yameogo B, Djogbenou LS, Chandre F, Dabire RK, Corbel V & Cohuet A (2014) Interplay Between Plasmodium Infection and Resistance to Insecticides in Vector Mosquitoes. *J. Infect. Dis.* **210**, 1464–1470.

5 Tun KM, Imwong M, Lwin KM, Win AA, Hlaing TM, Hlaing T, Lin K, Kyaw MP, Plewes K, Faiz MA, Dhorda M, Cheah PY, Pukrittayakamee S, Ashley EA, Anderson TJC, Nair S, McDew-White M, Flegg JA, Grist EPM, Guerin P, Maude RJ, Smithuis F, Dondorp AM, Day NPJ, Nosten F, White NJ & Woodrow CJ (2015) Spread of artemisinin-resistant Plasmodium falciparum in Myanmar: a cross-sectional survey of the K13 molecular marker. *Lancet Infect. Dis.* **15**, 415–421.

6 Takala-Harrison S, Jacob CG, Arze C, Cummings MP, Silva JC, Dondorp AM, Fukuda MM, Hien TT, Mayxay M, Noedl H, Nosten F, Kyaw MP, Nhien NTT, Imwong M, Bethell D, Se Y, Lon C, Tyner SD, Saunders DL, Ariey F, Mercereau-Puijalon O, Menard D, Newton PN, Khanthavong M, Hongvanthong B, Starzengruber P, Fuehrer H-P, Swoboda P, Khan WA, Phyo AP, Nyunt MM, Nyunt MH, Brown TS, Adams M, Pepin CS, Bailey J, Tan JC, Ferdig MT, Clark TG, Miotto O, MacInnis B, Kwiatkowski

619 DP, White NJ, Ringwald P & Plowe C V. (2015) Independent Emergence of
620 Artemisinin Resistance Mutations Among Plasmodium falciparum in Southeast Asia. *J.*
621 *Infect. Dis.* **211**, 670–679.

622 7 Haldar K, Bhattacharjee S & Safeukui I (2018) Drug resistance in Plasmodium. *Nat. Rev.*
623 *Microbiol.* **16**, 156–170.

624 8 Fairhurst RM & Dondorp AM (2016) Artemisinin-Resistant Plasmodium falciparum
625 Malaria. In *Emerging infections 10* pp. 409–429. American Society of Microbiology.

626 9 Corey VC, Lukens AK, Istvan ES, Lee MCS, Franco V, Magistrado P, Coburn-Flynn O,
627 Sakata-Kato T, Fuchs O, Gnädig NF, Goldgof G, Linares M, Gomez-Lorenzo MG, De
628 Cózar C, Lafuente-Monasterio MJ, Prats S, Meister S, Tanaseichuk O, Wree M, Zhou Y,
629 Willis PA, Gamo F-J, Goldberg DE, Fidock DA, Wirth DF & Winzeler EA (2016) A
630 broad analysis of resistance development in the malaria parasite. *Nat. Commun.* **7**,
631 11901.

632 10 Deu E (2017) Proteases as antimalarial targets: strategies for genetic, chemical, and
633 therapeutic validation. *FEBS J.* **284**, 2604–2628.

634 11 Paul AS & Duraisingh MT (2018) Targeting Plasmodium Proteases to Block Malaria
635 Parasite Escape and Entry. *Trends Parasitol.* **34**, 95–97.

636 12 Nasamu AS, Glushakova S, Russo I, Vaupel B, Oksman A, Kim AS, Fremont DH, Tolia
637 N, Beck JR, Meyers MJ, Niles JC, Zimmerberg J & Goldberg DE (2017) Plasmepsins
638 IX and X are essential and druggable mediators of malaria parasite egress and invasion.
639 *Science* **358**, 518–522.

640 13 Alam A (2014) Serine Proteases of Malaria Parasite Plasmodium falciparum : Potential
641 as Antimalarial Drug Targets 1 . Global Malaria Burden and Need for. **2014**.

642 14 Gilson PR, Chisholm SA, Crabb BS & de Koning-Ward TF (2017) Host cell remodelling
643 in malaria parasites: a new pool of potential drug targets. *Int. J. Parasitol.* **47**, 119–127.

644 15 Qidwai T (2015) Hemoglobin Degrading Proteases of Plasmodium falciparum as
645 Antimalarial Drug Targets. *Curr. Drug Targets* **16**, 1133–1141.

646 16 Alam A (2017) Plasmodium Proteases as Therapeutic Targets Against Malaria. In
647 *Proteases in Human Diseases* pp. 69–90. Springer Singapore, Singapore.

648 17 Sijwali PS & Rosenthal PJ (2004) Gene disruption confirms a critical role for the cysteine
649 protease falcipain-2 in hemoglobin hydrolysis by Plasmodium falciparum. *Proc. Natl.*
650 *Acad. Sci. U. S. A.* **101**, 4384–4389.

651 18 Goldberg DE (1992) Plasmodial hemoglobin degradation: an ordered pathway in a
652 specialized organelle. *Infect. Agents Dis.* **1**, 207–211.

653 19 Pandey KC, Singh N, Arastu-Kapur S, Bogyo M & Rosenthal PJ (2006) Falstatin, a
654 cysteine protease inhibitor of Plasmodium falciparum, facilitates erythrocyte invasion.
655 *PLoS Pathog.* **2**, e117.

656 20 Drew ME, Banerjee R, Uffman EW, Gilbertson S, Rosenthal PJ & Goldberg DE (2008)
657 Plasmodium food vacuole plasmepsins are activated by falcipains. *J. Biol. Chem.* **283**,
658 12870–12876.

659 21 Dowse TJ, Koussis K, Blackman MJ & Soldati-Favre D (2008) Roles of proteases during
660 invasion and egress by Plasmodium and Toxoplasma. *Subcell. Biochem.* **47**, 121–139.

661 22 Rosenthal PJ (2011) Falcipains and other cysteine proteases of malaria parasites. *Adv.*
662 *Exp. Med. Biol.* **712**, 30–48.

663 23 Hanssen E, Knoechel C, Dearnley M, Dixon MWA, Le Gros M, Larabell C & Tilley L
664 (2012) Soft X-ray microscopy analysis of cell volume and hemoglobin content in

665 erythrocytes infected with asexual and sexual stages of Plasmodium falciparum. *J.*
666 *Struct. Biol.* **177**, 224–232.

667 24 Krugliak M, Zhang J & Ginsburg H (2002) Intraerythrocytic Plasmodium falciparum
668 utilizes only a fraction of the amino acids derived from the digestion of host cell cytosol
669 for the biosynthesis of its proteins. *Mol. Biochem. Parasitol.* **119**, 249–256.

670 25 Istvan ES, Dharia N V, Bopp SE, Gluzman I, Winzeler EA & Goldberg DE (2011)
671 Validation of isoleucine utilization targets in Plasmodium falciparum. *Proc. Natl. Acad.*
672 *Sci. U. S. A.* **108**, 1627–32.

673 26 Liu J, Istvan ES, Gluzman IY, Gross J & Goldberg DE (2006) Plasmodium falciparum
674 ensures its amino acid supply with multiple acquisition pathways and redundant
675 proteolytic enzyme systems. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8840–5.

676 27 Francis SE, Sullivan Jr DJ & Goldberg DE (1997) Hemoglobin metabolism in the malaria
677 parasite Plasmodium falciparum. *Annu. Rev. Microbiol.* **51**, 97–123.

678 28 Greenbaum DC, Baruch A, Grainger M, Bozdech Z, Medzihradszky KF, Engel J, DeRisi
679 J, Holder AA & Bogyo M (2002) A role for the protease falcipain 1 in host cell invasion
680 by the human malaria parasite. *Science* **298**, 2002–2006.

681 29 Eksi S, Czesny B, Greenbaum DC, Bogyo M & Williamson KC (2004) Targeted
682 disruption of Plasmodium falciparum cysteine protease, falcipain 1, reduces oocyst
683 production, not erythrocytic stage growth. *Mol. Microbiol.* **53**, 243–250.

684 30 Sijwali PS, Kato K, Seydel KB, Gut J, Lehman J, Klemba M, Goldberg DE, Miller LH &
685 Rosenthal PJ (2004) Plasmodium falciparum cysteine protease falcipain-1 is not
686 essential in erythrocytic stage malaria parasites. *Proc. Natl. Acad. Sci.* **101**, 8721–8726.

687 31 Singh N, Sijwali PS, Pandey KC & Rosenthal PJ (2006) Plasmodium falciparum:

688 biochemical characterization of the cysteine protease falcipain-2'. *Exp. Parasitol.* **112**,
689 187–92.

690 32 Pandey KC & Dixit R (2012) Structure-function of falcipains: malarial cysteine proteases.
691 *J. Trop. Med.* **2012**, 345195.

692 33 Hanspal M, Dua M, Takakuwa Y, Chishti AH & Mizuno A (2002) Plasmodium
693 falciparum cysteine protease falcipain-2 cleaves erythrocyte membrane skeletal proteins
694 at late stages of parasite development. *Blood* **100**, 1048–1054.

695 34 Dua M, Raphael P, Sijwali PS, Rosenthal PJ & Hanspal M (2001) Recombinant falcipain-
696 2 cleaves erythrocyte membrane ankyrin and protein 4.1. *Mol. Biochem. Parasitol.* **116**,
697 95–99.

698 35 Dhawan S, Dua M, Chishti AH & Hanspal M (2003) Ankyrin peptide blocks falcipain-2-
699 mediated malaria parasite release from red blood cells. *J. Biol. Chem.* **278**, 30180–6.

700 36 Sijwali PS, Koo J, Singh N & Rosenthal PJ (2006) Gene disruptions demonstrate
701 independent roles for the four falcipain cysteine proteases of Plasmodium falciparum.
702 *Mol. Biochem. Parasitol.* **150**, 96–106.

703 37 Sijwali PS, Shenai BR, Gut J, Singh A & Rosenthal PJ (2001) Expression and
704 characterization of the Plasmodium falciparum haemoglobinase falcipain-3. *Biochem. J.*
705 **360**, 481–489.

706 38 Teixeira C, Gomes JRB & Gomes P (2011) Falcipains, Plasmodium falciparum cysteine
707 proteases as key drug targets against malaria. *Curr. Med. Chem.* **18**, 1555–1572.

708 39 Marco M & Coteron JM (2012) Falcipain inhibition as a promising antimalarial target.
709 *Curr. Top. Med. Chem.* **12**, 408–444.

710 40 Siddiqui FA, Cabrera M, Wang M, Brashear A, Kemirembe K, Wang Z, Miao J,

711 Chookajorn T, Yang Z, Cao Y, Dong G, Rosenthal PJ & Cui L (2018) Plasmodium
712 falciparum Falcipain-2a Polymorphisms in Southeast Asia and Their Association With
713 Artemisinin Resistance. *J. Infect. Dis.*

714 41 Na BK, Shenai BR, Sijwali PS, Choe Y, Pandey KC, Singh A, Craik CS & Rosenthal PJ
715 (2004) Identification and biochemical characterization of vivapains, cysteine proteases
716 of the malaria parasite Plasmodium vivax. *Biochem. J.* **378**, 529–538.

717 42 Prasad R, Atul P, Soni A, Puri SK & Sijwali PS (2012) Expression, Characterization, and
718 Cellular Localization of Knowpains, Papain-Like Cysteine Proteases of the Plasmodium
719 knowlesi Malaria Parasite. *PLoS One* **7**, e51619.

720 43 Pei Y, Miller JL, Lindner SE, Vaughan AM, Torii M & Kappe SH (2013) Plasmodium
721 yoelii inhibitor of cysteine proteases is exported to exomembrane structures and
722 interacts with yoelipain-2 during asexual blood-stage development. *Cell. Microbiol.*

723 44 Caldeira RL, Gonçalves LMD, Martins TM, Silveira H, Novo C, Rosário V do &
724 Domingos A (2009) Plasmodium chabaudi: Expression of active recombinant
725 chabaupain-1 and localization studies in Anopheles sp. *Exp. Parasitol.* **122**, 97–105.

726 45 Rosenthal PJ & Nelson RG (1992) Isolation and characterization of a cysteine proteinase
727 gene of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **51**, 143–152.

728 46 Shenai BR, Sijwali PS, Singh A & Rosenthal PJ (2000) Characterization of native and
729 recombinant falcipain-2, a principal trophozoite cysteine protease and essential
730 hemoglobinase of Plasmodium falciparum. *J. Biol. Chem.* **275**, 29000–29010.

731 47 Tastan Bishop O & Kroon M (2011) Study of protein complexes via homology modeling,
732 applied to cysteine proteases and their protein inhibitors. *J. Mol. Model.* **17**, 3163–72.

733 48 Rennenberg A, Lehmann C, Heitmann A, Witt T, Hansen G, Nagarajan K, Deschermeier

C, Turk V, Hilgenfeld R & Heussler VT (2010) Exoerythrocytic Plasmodium parasites secrete a cysteine protease inhibitor involved in sporozoite invasion and capable of blocking cell death of host hepatocytes. *PLoS Pathog.* **6**, e1000825.

49 Monteiro AC, Abrahamson M, Lima AP, Vannier-Santos MA & Scharfstein J (2001) Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in Trypanosoma cruzi. *J. Cell Sci.* **114**, 3933–3942.

50 Musyoka TM, Kanzi AM, Lobb KA & Tastan Bishop Ö (2016) Analysis of non-peptidic compounds as potential malarial inhibitors against Plasmodial cysteine proteases via integrated virtual screening workflow. *J. Biomol. Struct. Dyn.* **34**, 2084–2101.

51 Sajid M & McKerrow JH (2002) Cysteine proteases of parasitic organisms. *Mol. Biochem. Parasitol.* **120**, 1–21.

52 Dahl EL & Rosenthal PJ (2005) Biosynthesis, localization, and processing of falcipain cysteine proteases of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **139**, 205–212.

53 Rozman J, Stojan J, Kuhelj R, Turk V & Turk B (1999) Autocatalytic processing of recombinant human procathepsin B is a bimolecular process. *FEBS Lett.* **459**, 358–62.

54 Cotrin SS, Gouvêa IE, Melo PMS, Bagnaresi P, Assis DM, Araújo MS, Juliano MA, Gazarini ML, Rosenthal PJ, Juliano L & Carmona AK (2013) Substrate specificity studies of the cysteine peptidases falcipain-2 and falcipain-3 from Plasmodium falciparum and demonstration of their kininogenase activity. *Mol. Biochem. Parasitol.* **187**, 111–116.

55 Chakka SK, Kalamuddin M, Sundararaman S, Wei L, Mundra S, Mahesh R, Malhotra P, Mohmmed A & Kotra LP (2015) Identification of novel class of falcipain-2 inhibitors as potential antimalarial agents. *Bioorg. Med. Chem.* **23**, 2221–2240.

757 56 Hernández González JE, Hernández Alvarez L, Pascutti PG & Valiente PA (2017)
758 Predicting binding modes of reversible peptide-based inhibitors of falcipain-2 consistent
759 with structure-activity relationships. *Proteins Struct. Funct. Bioinforma.* **85**, 1666–1683.

760 57 Musyoka TMTM, Kanzi AMAM, Lobb KAKA, Tastan Bishop Ö & Bishop ÖT (2016)
761 Structure Based Docking and Molecular Dynamic Studies of Plasmodial Cysteine
762 Proteases against a South African Natural Compound and its Analogs. *Nat. Publ. Gr.* **6**,
763 1–12.

764 58 Coteron JM, Catterick D, Castro J, Chaparro MJ, Diaz B, Fernandez E, Ferrer S, Gamo FJ,
765 Gordo M, Gut J, de las Heras L, Legac J, Marco M, Miguel J, Munoz V, Porras E, de la
766 Rosa JC, Ruiz JR, Sandoval E, Ventosa P, Rosenthal PJ & Fiandor JM (2010) Falcipain
767 inhibitors: optimization studies of the 2-pyrimidinecarbonitrile lead series. *J. Med.*
768 *Chem.* **53**, 6129–6152.

769 59 Domínguez JN, León C, Rodrigues J, Gamboa de Domínguez N, Gut J & Rosenthal PJ
770 (2005) Synthesis and evaluation of new antimalarial phenylurenyl chalcone derivatives.
771 *J. Med. Chem.* **48**, 3654–8.

772 60 Rudrapal M, Chetia D & Singh V (2017) Novel series of 1,2,4-trioxane derivatives as
773 antimalarial agents. *J. Enzyme Inhib. Med. Chem.* **32**, 1159–1173.

774 61 Himangini, Pathak DP, Sharma V & Kumar S (2018) Designing novel inhibitors against
775 falcipain-2 of Plasmodium falciparum. *Bioorg. Med. Chem. Lett.* **28**, 1566–1569.

776 62 Ehmke V, Kilchmann F, Heindl C, Cui K, Huang J, Schirmeister T & Diederich F (2011)
777 Peptidomimetic nitriles as selective inhibitors for the malarial cysteine protease
778 falcipain-2. *Medchemcomm* **2**, 800.

779 63 Mane UR, Gupta RC, Nadkarni SS, Giridhar RR, Naik PP & Yadav MR (2013) Falcipain

780 inhibitors as potential therapeutics for resistant strains of malaria: a patent review.
781 *Expert Opin. Ther. Pat.* **23**, 165–87.

782 64 Njuguna JN (2012) Structural Analysis of Prodomain Inhibition of Cysteine Proteases in
783 Plasmodium Species. Rhodes University.

784 65 Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A,
785 Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W,
786 Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ,
787 Treatman C & Wang H (2009) PlasmoDB: a functional genomic database for malaria
788 parasites. *Nucleic Acids Res.* **37**, D539-43.

789 66 Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J & Sayers EW (2009) GenBank.
790 *Nucleic Acids Res.* **37**, D26-31.

791 67 Pei J, Kim BH & Grishin N V (2008) PROMALS3D: a tool for multiple protein sequence
792 and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300.

793 68 Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview
794 Version 2--a multiple sequence alignment editor and analysis workbench.
795 *Bioinformatics* **25**, 1189–1191.

796 69 Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5:
797 molecular evolutionary genetics analysis using maximum likelihood, evolutionary
798 distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

799 70 Bailey TL, Williams N, Misleh C & Li WW (2006) MEME: discovering and analyzing
800 DNA and protein sequence motifs. *Nucleic Acids Res.* **34**, W369-73.

801 71 Bailey TL & Gribskov M (1998) Combining evidence using p-values: application to
802 sequence homology searches. *Bioinformatics* **14**, 48–54.

803 72 Webb B & Sali A (2016) Comparative Protein Structure Modeling Using MODELLER.
804 *Curr. Protoc. Bioinforma.* **54**, 5.6.1-5.6.37.

805 73 Shen M-Y & Sali A (2006) Statistical potential for assessment and prediction of protein
806 structures. *Protein Sci.* **15**, 2507–24.

807 74 Wiederstein M & Sippl MJ (2007) ProSA-web: interactive web service for the recognition
808 of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–
809 W410.

810 75 Eisenberg D, Lüthy R & Bowie JU (1997) VERIFY3D: Assessment of protein models
811 with three-dimensional profiles. In *Methods in enzymology* pp. 396–404.

812 76 Benkert P, Tosatto SCE & Schomburg D (2008) QMEAN: A comprehensive scoring
813 function for model quality assessment. *Proteins* **71**, 261–77.

814 77 Laskowski RA, MacArthur MW, Moss DS & Thornton JM (1993) PROCHECK: a
815 program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*
816 **26**, 283–291.

817 78 Tina KG, Bhadra R & Srinivasan N (2007) PIC: Protein Interactions Calculator. *Nucleic*
818 *Acids Res.* **35**, W473–W476.

819 79 Galgonek J, Vymětal J, Jakubec D & Vondrášek J (2017) Amino Acid Interaction
820 (INTAA) web server. *Nucleic Acids Res.* **29**, 2860–2874.

821 80 Crooks G, Hon G, Chandonia J & Brenner S (2004) WebLogo: a sequence logo generator.
822 *Genome Res* **14**, 1188–1190.

823 81 Kurcinski M, Jamroz M, Blaszczyk M, Kolinski A & Kmiecik S (2015) CABS-dock web
824 server for the flexible docking of peptides to proteins without prior knowledge of the
825 binding site. *Nucleic Acids Res.* **43**, W419-24.

826 82 Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM & Vangone A (2016) PRODIGY: a web
827 server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*
828 **32**, 3676–3678.

829 83 Guruprasad K, Reddy BVB & Pandit MW (1990) Correlation between stability of a
830 protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of
831 a protein from its primary sequence. *"Protein Eng. Des. Sel.* **4**, 155–161.

832 84 Kreusch S, Fehn M, Maubach G, Nissler K, Rommerskirch W, Schilling K, Weber E,
833 Wenz I & Wiederanders B (2000) An evolutionarily conserved tripartite tryptophan
834 motif stabilizes the prodomains of cathepsin L-like cysteine proteases. *Eur. J. Biochem.*
835 **267**, 2965–2972.

836 85 De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E,
837 Bairoch A & Hulo N (2006) ScanProsite: detection of PROSITE signature matches and
838 ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**,
839 W362–W365.

840 86 Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV & Falquet L (2004)
841 MyHits: a new interactive resource for protein annotation and domain identification.
842 *Nucleic Acids Res.* **32**, W332-5.

843 87 Sundararaj S, Singh D, Saxena AK, Vashisht K, Sijwali PS, Dixit R & Pandey KC (2012)
844 The Ionic and Hydrophobic Interactions Are Required for the Auto Activation of
845 Cysteine Proteases of Plasmodium falciparum. *PLoS One* **7**, e47227.

846 88 Fosgerau K & Hoffmann T (2015) Peptide therapeutics: current status and future
847 directions. *Drug Discov. Today* **20**, 122–128.

848 89 Henninot A, Collins JC & Nuss JM (2018) The Current State of Peptide Drug Discovery:

Back to the Future? *J. Med. Chem.* **61**, 1382–1414.

90 Rizzi L, Sundararaman S, Cendic K, Vaiana N, Korde R, Sinha D, Mohmmed A, Malhotra P & Romeo S (2011) Design and synthesis of protein-protein interaction mimics as Plasmodium falciparum cysteine protease, falcipain-2 inhibitors. *Eur. J. Med. Chem.* **46**, 2083–90.

91 Korde R, Bhardwaj A, Singh R, Srivastava A, Chauhan VS, Bhatnagar RK & Malhotra P (2008) A Prodomain Peptide of *Plasmodium falciparum* Cysteine Protease (Falcipain-2) Inhibits Malaria Parasite Development. *J. Med. Chem.* **51**, 3116–3123.

92 Pandey KC, Barkan DT, Sali A & Rosenthal PJ (2009) Regulatory elements within the prodomain of Falcipain-2, a cysteine protease of the malaria parasite Plasmodium falciparum. *PLoS One* **4**, e5694.

93 Farias SL, Gazarini ML, Melo RL, Hirata IY, Juliano MA, Juliano L & Garcia CRS (2005) Cysteine-protease activity elicited by Ca2+ stimulus in Plasmodium. *Mol. Biochem. Parasitol.* **141**, 71–79.

865 Table 1. Details of all protein sequences retrieved from PlasmoDB and NCBI databases. Percentage
866 sequence identity (SI) is calculated based on the partial zymogen of query sequence (QS) and that of
867 corresponding homolog.

Source Organism	Host	Accession Number	Common name (abbreviation)	aa	% SI
P. falciparum (Pf)	Human	PF3D7_1115700	Falcipain-2 (FP-2)	484	QS
		PF3D7_1115400	Falcipain-3 (FP-3)	492	QS
P. vivax (Pv)		PVX_091415	Vivapain-2 (VP-2)	487	56[a]
		PVX_091410	Vivapain-3 (VP-3)	493	55[b]
P. knowlesi (Pk)	Human / Monkey	PKH_091250	Knowlesipain-2 (KP-2)	495	53[a]
		PVX-091260	Knowlesipain-3 (KP-3)	479	58[b]
P. yoelii (Py)	Rodents	PY00783	Yoelipain-2 (YP-2)	472	47[a]
P. chabaudi (Pc)		PCHAS_091190	Chabaupain-2 (CP-2)	471	48[b]
P. berghei (Pb)		PBANKA_09324	Berghepain-2 (BP-2)	470	50[a]
H. sapiens	-	NP_000387.1	Cathepsin-K (Cat-K)	329	34[a]
		AAA66974.1	Cathepsin-L (Cat-L)	333	33[a]
		AAC37592.1	Cathepsin-S (Cat-S)	331	32[a]

868 a=FP-2 homolog, [b]=FP-3 homolog

869 Table 2. Homology model quality validation scores of partial zymogen complexes using different
870 assessment tools.

					RAMACHADRAN (%)		
Protein	**Z-DOPE**	**Verify3D**	**ProSA**	**QMEAN**	**Favoured**	**Allowed**	**Disallowed**
FP-2	-1.05	78.48	-8.27	0.69	88.90	10.50	0.60
FP-3	-0.94	84.64	-7.84	0.67	89.70	10.30	0.00
VP-2	-0.64	81.27	-6.94	0.62	85.40	13.90	0.70
VP-3	-0.74	79.58	-7.31	0.70	88.60	11.40	0.00
KP-2	-0.63	85.89	-7.23	0.62	89.30	9.70	1.00
KP-3	-0.92	90.63	-7.88	0.63	86.10	13.90	0.00
BP-2	-0.62	86.85	-7.75	0.63	86.60	12.40	1.00
CP-2	-0.60	84.45	-7.02	0.63	83.90	13.80	2.30
YP-2	-0.42	75.54	-7.28	0.62	84.90	14.40	0.70
Cat-L	-1.47	87.38	-7.94	0.87	89.80	9.80	0.40
Cat-S	-1.61	85.39	-8.57	0.79	89.20	10.40	0.20
1BY8	*	85.16	-8.62	0.78	65.80	34.20	0.00
2O6X	*	94.77	-7.00	0.77	90.50	9.50	0.00
2OUL	*	98.13	-7.90	0.75	88.10	11.20	0.70
3BWK	*	93.51	-7.35	0.65	86.10	13.30	0.60

871 *=Template

872

873 Table 3. A summary of physicochemical properties of FP-2 and FP-3 and homologs partial zymogen
874 sequences. Included also are properties where catalytic domain significantly varied from partial
875 zymogen sequence.

Protein	Aromaticity	GRAVY	Instability index	Mwgt		pI	
				Complex	Catalytic	Complex	Catalytic
FP-2	0.12	-0.59	40.31	38021.06	27176.69	6.50	4.94
FP-3	0.14	-0.53	33.90	38029.91	27348.62	5.47	4.72
VP-2	0.14	-0.40	23.78	37941.11	27388.93	5.49	4.74
VP-3	0.14	-0.47	28.92	38405.76	28088.95	5.60	5.00
KP-2	0.14	-0.54	29.55	38583.83	27947.73	5.69	4.93
KP-3	0.14	-0.51	23.16	38284.39	27685.36	5.89	5.13
BP-2	0.14	-0.52	39.73	38014.94	27367.90	5.44	4.77
CP-2	0.13	-0.47	52.14	37883.92	27140.51	5.14	4.54
YP-2	0.14	-0.46	40.05	38120.25	27454.06	5.66	4.78
Cat-K	0.10	-0.62	33.12	34566.05	23495.47	8.83	8.92
Cat-L	0.13	-0.70	38.17	35074.20	24298.86	5.33	4.64
Cat-S	0.12	-0.56	37.20	34986.63	23963.97	8.44	7.64

876

877

878 Table 4. Amino acid sequences of proposed peptides, their predicted binding affinity values (ΔG -
879 kcalmol^{-1}) and dissociation constant (K_d) with individual catalytic domains of the different proteins
880 studied.

	Peptide									
	1		2		3		4		5	
Protein	ΔG	**K_D (M)**	ΔG	**K_D (M)**	ΔG	**K_D**	ΔG	**K_D**	ΔG	**K_D (M)**
FP-2	-11.8	2.2E-09	-11.2	8.4E-09	-8.1	1.2E-06	-9	2.7E-07	-11.4	4.4E-09
FP-3	-12.2	1.2E-09	-10.5	2.0E-08	-9.6	9.4E-08	-9.7	7.8E-08	-12.3	8.8E-10
VP-2	-11.0	8.3E-09	-10.1	3.8E-08	-12.3	8.8E-10	-9.0	2.7E-07	-10.9	1.1E-08
VP-3	-11.7	2.7E-09	-11.0	8.3E-09	-9.9	5.7E-08	-9.2	1.7E-07	-11.8	2.1E-09
KP-2	-12.7	4.7E-10	-11.8	2.1E-09	-6.7	1.3E-05	-8.1	1.2E-06	-10.6	1.8E-08
KP-3	-11.8	2.1E-09	-12.7	5.2E-10	-9.3	1.4E-07	-9.2	1.9E-07	-12.7	5.2E-10
BP-2	-11.5	3.7E-09	-12.2	1.2E-09	-10.9	9.9E-09	-7.9	1.7E-06	-11.7	2.7E-09
CP-2	-11.9	2.2E-09	-11.7	2.7E-09	-8.7	3.9E-07	-8.3	1.1E-07	-12.7	4.7E-10
YP-2	-11.7	2.7E-09	-14.3	5.5E-11	-12.4	7.9E-10	-10.4	2.5E-08	-9.3	1.5E-07
Cat-K	-13.9	6.3E-11	-11.5	3.5E-09	-11.9	2.0E-09	-11.4	4.2E-09	-8.3	1.1E-07
Cat-L	-12.2	1.0E-09	-11.8	5.1E-09	-12.3	9.7E-10	-10.2	3.5E-08	-8.7	9.2E-07
Cat-S	-10.4	2.4E-09	-11.5	3.5E-09	-12.4	7.7E-10	-9.9	5.2E-08	-8.6	4.8E-07

881 *FP-2 complex with chagasin: ΔG -11.9, κ_D=1.9e-09.
882 **Peptide 1**^{NRFGDLSFEEFKKKYLNLKLFD}, **2**^{LTYHEFKNKYLSLRSSK}, **3**^{MTFEEFKQKYLTLKSKD}, **4**^{EFKNKYLTLK}, **5**^{EFKNKYLTLKSKD}
883

884 Table 5. Peptide 5-catalytic domain residue interaction fingerprint. Bold are residues forming
885 hydrogen bonds with the peptide.

Protein	Subsite				Non-subsite residues
	S1	S2	S3	S1'	
FP-2	Q279,C282,G283	L327,**I328**,S392,L415,N416,A418,D477	G325,G326	V395,S396,A400,**H417**,A418,N447,W449	K280,W286,R470,C285
FP-3	Q287,C290,G291,Y332	**Y335**,S400,P423,A426,E485	G334,G335	A402,A403,S404,A408,**N424**,H425,N455,W457	L289,C293,W294,N338
VP-2	Q282,G286,Y327	F330,I331,P418,N419,A421	Q322,N323	I396,A397,V398,A403,H420,N451	D282,F331,P332,Y351,E367,F389
VP-3	G292,D333	N336,I337,S401,P424,N425	G334,G335	A404,V409,Y410,**H426**,W458	D287,K289,C294,D406,S423,S457,G459,K458,W462
KP-2	Q288,C291,G292,C332,D333	L336,I337,S401,P424,**N425**,A427,E486	G333,G334	N403,A404,N405,T409,H426,N456,W458	**K289**,N290,A293,C294,W295,E382,N405,D406,S457
KP-3	C277,G278,C318,D319	F322,N387,T410,N411,**E472**	Q314,N316,G320,G321	V390,S391,A395,**H412**,N442,W444	G275,S279,C280,P324,R325,E368,N387,D392,T409
BP-2	Q264,A268,C308,E309	I312,A377,**A400**,N401,A403	N304,N305,F306,G311	V378,G379,D381,H402,N432,W434	K266,A272,P314,Y334,E349,A367,I376
CP-2	Q267,C270,A271,E312	I315,L316,A380,A303,**N404**,A406	N307,N308,D309,G314	V381,G382,A383,S384,**H405**,N437	K265,W274,Q306,P337,K351
YP-2	C271,A272,D313	I316,L317,A381,A404,**N405**,A407	N308,N309,F310,G314	V382,G383,V384,A385,H406,N438	Q269,W275,F310,E353,I380
Cat-K	C136,G137	Y181,M182,A248,L274,N275,A277	G179	D250,A251,S252,H276	C139,W140,F186,Y224,F256,W302,E226
Cat-L	C135,N179	L182,A248,D275,G277	N175	I249,A251,L257,H276	P172,G224,A234,I263
Cat-S	C134,**N179**	F182,G249,V274,**N275**,F323	N175,K176,G180	R253,F258,W298	Y173,Y225,E227,P229,Y230

886

887