

# Regulatory Elements within the Prodomain of Falcipain-2, a Cysteine Protease of the Malaria Parasite *Plasmodium falciparum*

Kailash C. Pandey<sup>1</sup>, David T. Barkan<sup>2,3,4</sup>, Andrej Sali<sup>2,4</sup>, Philip J. Rosenthal<sup>1\*</sup>

**1** Department of Medicine, University of California San Francisco, San Francisco, California, United States of America, **2** Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, United States of America, **3** Graduate Group in Bioinformatics, University of California San Francisco, San Francisco, California, United States of America, **4** California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, United States of America

## Abstract

Falcipain-2, a papain family cysteine protease of the malaria parasite *Plasmodium falciparum*, plays a key role in parasite hydrolysis of hemoglobin and is a potential chemotherapeutic target. As with many proteases, falcipain-2 is synthesized as a zymogen, and the prodomain inhibits activity of the mature enzyme. To investigate the mechanism of regulation of falcipain-2 by its prodomain, we expressed constructs encoding different portions of the prodomain and tested their ability to inhibit recombinant mature falcipain-2. We identified a C-terminal segment (Leu<sup>155</sup>–Asp<sup>243</sup>) of the prodomain, including two motifs (ERFNIN and GNFD) that are conserved in cathepsin L sub-family papain family proteases, as the mediator of prodomain inhibitory activity. Circular dichroism analysis showed that the prodomain including the C-terminal segment, but not constructs lacking this segment, was rich in secondary structure, suggesting that the segment plays a crucial role in protein folding. The falcipain-2 prodomain also efficiently inhibited other papain family proteases, including cathepsin K, cathepsin L, cathepsin B, and cruzain, but it did not inhibit cathepsin C or tested proteases of other classes. A structural model of pro-falcipain-2 was constructed by homology modeling based on crystallographic structures of mature falcipain-2, procathepsin K, procathepsin L, and procaricain, offering insights into the nature of the interaction between the prodomain and mature domain of falcipain-2 as well as into the broad specificity of inhibitory activity of the falcipain-2 prodomain.

**Citation:** 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(5): e5694. doi:10.1371/journal.pone.0005694

**Editor:** Ashley M. Buckle, Monash University, Australia

**Received:** December 23, 2008; **Accepted:** March 24, 2009; **Published:** May 27, 2009

**Copyright:** © 2009 Pandey et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the National Institutes of Health (grants AI35800, AI35707, GM54762, and GM71790), the Medicines for Malaria Venture, the Sandler Family Supporting Foundation, and computer hardware gifts from Mike Homer, Ron Conway, IBM, Intel, HP, and NetApp. PJR is a Doris Duke Charitable Foundation Distinguished Clinical Scientist. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: prosenthal@medsfgh.ucsf.edu

## Introduction

Malaria is the most important parasitic disease in the world. *Plasmodium falciparum*, the most virulent human malaria parasite, is responsible for hundreds of millions of illnesses and about one million deaths each year (1). The control of malaria is hindered by increasing resistance to available drugs, making it important to develop new drugs to treat this disease. Among potential new targets for antimalarial chemotherapy are falcipain cysteine proteases (2). The best characterized of these proteases, falcipain-2 and falcipain-3, play key roles in the hydrolysis of hemoglobin by intraerythrocytic parasites (3–5). Inhibitors of falcipains demonstrate potent *in vitro* and *in vivo* antimalarial activity, and these proteases are the targets of efforts to develop novel cysteine protease inhibitors as new antimalarial drugs (2).

Falcipains are cathepsin L-like papain-family cysteine proteases (2). Features shared with other proteases of this sub-family include a catalytic domain of ~30 kDa with conserved active site amino acid residues and a prodomain with potent enzyme inhibitory activity (6). We have characterized a number of unusual features of falcipains. First, folding of the mature protease is mediated by a

short (14 amino acids for falcipain-2) N-terminal extension, rather than the enzyme prodomain (6, 7). Second, a 10 amino acid insertion near the C-terminus mediates interaction of the mature domain with its principal substrate, hemoglobin, and with the prodomain (8). Third, the prodomain does not have a typical signal sequence, but contains a membrane-spanning domain that predicts a type II integral membrane protein. Fourth, the falcipain prodomain is much larger than that of most other described papain-family proteases, with downstream sequence similar to papain and related enzymes, but unique upstream regions that mediate trafficking of falcipain-2 to the food vacuole, the site of hydrolysis of hemoglobin (9).

Considering its importance as a potential drug target, we were interested in evaluating the features of the falcipain-2 prodomain that mediate enzyme inhibition. We hypothesized that the inhibitory function is mediated by the downstream portion of the prodomain, which has amino acid sequence similar to that of other papain family proteases. In this region, cathepsin L-like papain family proteases, including falcipains, contain a number of conserved residues that appear to mediate interaction between the prodomain and mature protease (10), including six amino acids

(ERFNIN in papain) spanning nineteen residues (11, 12) and, further downstream, four conserved amino acids (GNFD in papain) spanning seven residues (13). Conservative substitutions at these motifs are common; the sequences are ERWNIN and ANFD in cathepsin L and DRWNIN and ANLD in cathepsin K. In cathepsin L, these residues appear to stabilize the prodomain structure through the formation of salt bridges (14). To determine the roles of these conserved amino acids and other portions of the falcipain-2 prodomain in enzyme inhibition, we expressed the prodomain and a series of truncated fragments, and evaluated their inhibitory activity (15). Our results define a 61 amino acid minimum inhibitory domain, which includes the ERFNIN and GNFD motifs, that strongly inhibits falcipain-2 and many other cysteine proteases. Modeling of the falcipain-2 prodomain suggests that the prodomain covers the enzyme active site, and thereby inhibits activity by preventing substrate access.

**Results**

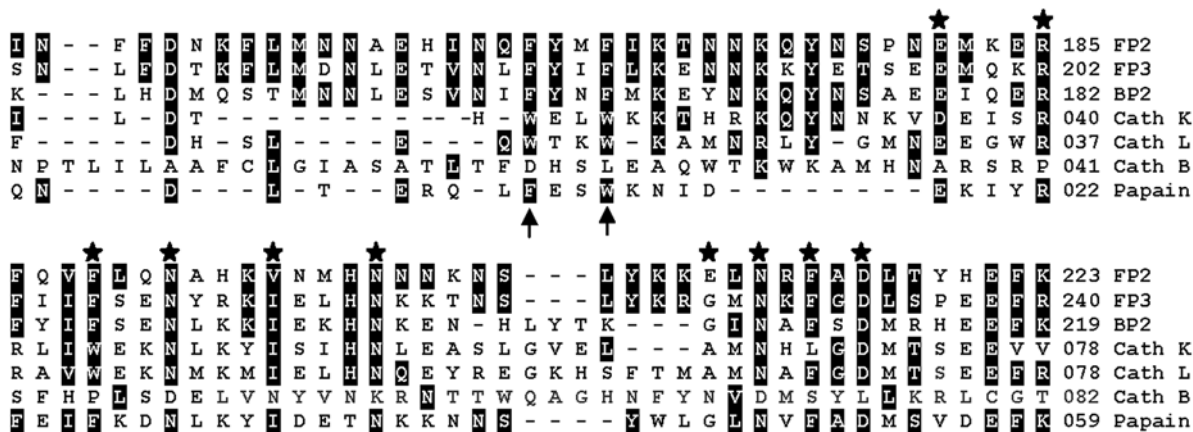
**Identification of the Inhibitory Domain of Falcipain-2**

Falcipain-2 and homologs from related plasmodia have much larger prodomains than those of most papain-family proteases. The upstream portion of the falcipain-2 prodomain bears no obvious resemblance to sequences of non-plasmodial proteases, and mediates enzyme trafficking to the parasite food vacuole (9). In contrast, the downstream portion of the falcipain-2 prodomain is similar to that of papain, and in particular to the cathepsin L sub-family of papain-family proteases (Fig. 1). Sequence identity for this region between falcipain-2 and human cathepsin L is ~21%, and residues that have been identified as playing key roles in the functions of papain family prodomains are generally conserved in falcipain-2 and plasmodial homologs. The well characterized ERFNIN and GNFD domains (10), which contribute to proenzyme stability, are both fully conserved in falcipain-3, but falcipain-2 differs from the consensus sequence at one ERFNIN (I→V) and one GNFD (G→E) residue. Two highly conserved Trp residues (at positions 19 and 22, of procathepsin L), which also contribute to the stability of cathepsin L sub-family proteases (12), are each replaced by Phe in both falcipain-2 and falcipain-3 (Fig. 1; falcipain-2 positions 165 and 168).

We previously showed that the prodomain of falcipain-2 is a potent reversible inhibitor of the protease (6). To characterize the requirements for inhibition, we expressed a series of prodomain fragments in *E. coli* (Figure S1) and evaluated inhibition of falcipain-2 by each of the fragments (Fig. 2). All peptides were soluble in the buffers used for our experiments and stable under our experimental conditions. As we hypothesized, the large upstream portion of the prodomain, which includes a transmembrane domain flanked by cytosolic and luminal segments, and which mediates trafficking of falcipain-2 to the food vacuole (9), is not required for enzyme inhibition. Inhibitory potency was the same for a prodomain construct lacking only the upstream cytosolic and transmembrane domains (Tyr<sup>54</sup>-Asp<sup>243</sup>) and for constructs lacking the upstream 104 (Ser<sup>105</sup>-Asp<sup>243</sup>), 126 (Leu<sup>127</sup>-Asp<sup>243</sup>), or 154 (Leu<sup>155</sup>-Asp<sup>243</sup>) amino acids of the prodomain (Fig. 2). All of these constructs were very potent inhibitors of falcipain-2, with K<sub>i</sub> < 1 nM. Removal of the 27 C-terminal amino acids of the prodomain (Tyr<sup>54</sup>-Asp<sup>216</sup>) did not affect inhibitory potency, but removal of the 37 C-terminal amino acids (Tyr<sup>54</sup>-Leu<sup>206</sup>) led to an ~2000-fold loss of inhibitory potency, and removal of the 63 C-terminal amino acids (Tyr<sup>54</sup>-Asn<sup>180</sup>) led to a complete loss of inhibitory activity. A peptide spanning the ERFNIN and GNFD motifs (Tyr<sup>176</sup>-Asp<sup>216</sup>) demonstrated no inhibitory activity. These results allow identification of a minimum inhibitory domain for falcipain-2 (Leu<sup>155</sup>-Asp<sup>216</sup>), which includes two hydrophobic residues (Phe<sup>165</sup> and Phe<sup>168</sup> in falcipain-2; Phe<sup>182</sup> and Phe<sup>185</sup> in falcipain-3) and the ERFNIN and GNFD motifs, all of which are highly conserved among other cathepsin L sub-family proteases. We could not directly test the inhibitory activity of this minimum inhibitory peptide, as production of the recombinant peptide was unsuccessful.

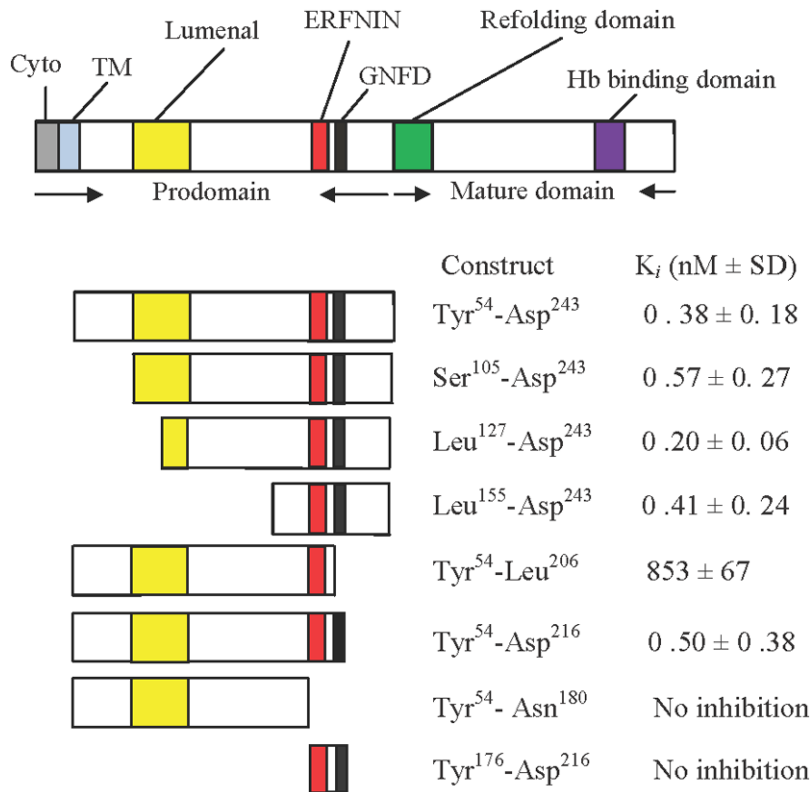
**Inhibitory Activity of the Falcipain-2 Prodomain Against Other Cysteine Proteases**

Cathepsin L sub-family protease prodomains generally inhibit only closely related proteases. For example, the prodomains of cathepsin L, cathepsin K, and cathepsin S are each potent inhibitors of all three proteases, but not of cathepsin B (10). In contrast, the falcipain-2 prodomain had a rather broad inhibitory specificity, with inhibition of the falcipain-2 homolog from



**Figure 1. Alignment of C-terminal amino acid residues of the prodomains of falcipain-2 and related cysteine proteases.** The sequences of falcipain-2 (FP2), falcipain-3 (FP3), berghhep-2 (BP2), human cathepsin K (Cath K), human cathepsin L (Cath L), human cathepsin B (Cath B), and papain were aligned using Expassy (European Bioinformatics Institute). Amino acids comprising the ERFNIN and GNFD motifs are labeled with stars, and conserved hydrophobic residues are indicated by arrows. Amino acids that are identical or similar to those of falcipain-2 are highlighted.

doi:10.1371/journal.pone.0005694.g001



**Figure 2. Inhibitory activity of profalcipain-2 constructs.** The domains of falcipain-2 and the studied constructs are represented diagrammatically. Abbreviations: Cyto, cytosolic domain; TM, transmembrane domain; Hb, hemoglobin. The residues contained in each construct are shown, and the inhibitory capacity of mature falcipain-2 for each construct is indicated. The data provided are the  $K_i$  values for each polypeptide construct. Results are from two experiments, each performed in duplicate. doi:10.1371/journal.pone.0005694.g002

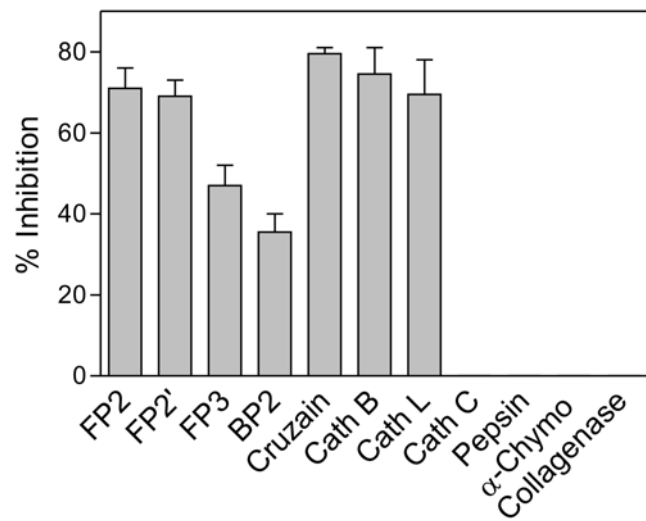
*Plasmodium berghei* (berghepain-2), the *Trypanosoma cruzi* protease cruzain, cathepsin L, and cathepsin B (Fig. 3). The only tested papain-family cysteine protease that was not inhibited was the dipeptidyl peptidase cathepsin C. The aspartic protease pepsin, serine protease  $\alpha$ -chymotrypsin, and metalloprotease collagenase were not inhibited by the falcipain-2 prodomain.

### Structural Explanation for Inhibitory Activity of Falcipain-2 Prodomain Fragments

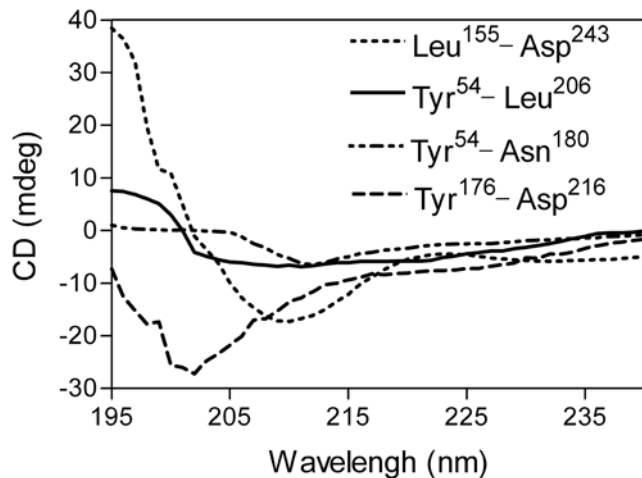
Structure-function studies identified a discrete portion of the falcipain-2 prodomain required for inhibition of the cognate mature protease. Prior work with other cathepsin L sub-family proteases suggests key roles for conserved hydrophobic amino acids as well as the ERFNIN and GNFD motifs in maintaining prodomain structure (10). We explored the roles of different domains in maintaining prodomain structure by circular dichroism analysis (Fig. 4). Secondary structure was seen in a fragment with potent inhibitory activity (Leu<sup>155</sup>-Asp<sup>243</sup>), but not in two larger constructs that lacked any sequence downstream of the ERFNIN and GNFD motifs (Tyr<sup>54</sup>-Leu<sup>206</sup>; Tyr<sup>54</sup>-Asn<sup>180</sup>) or in a peptide spanning the ERFNIN and GNFD motifs (Tyr<sup>176</sup>-Asp<sup>216</sup>). These results indicate that the ERFNIN and GNFD motifs and an upstream region including conserved Phe residues are required for proper folding or maintenance of secondary structure of the prodomain.

### Homology Modeling of Profalcipain-2

To explain the role of profalcipain-2 motifs in enzyme inhibition, we modeled the structure of the target falcipain-2



**Figure 3. Inhibition of different proteases by the prodomain of falcipain-2.** The inhibition of falcipain-2 (FP2), falcipain-2' (FP2'), falcipain-3 (FP3), berghepain-2 (BP2), cruzain, human cathepsin B (Cath B), human cathepsin L (Cath L), bovine cathepsin C (Cath C), pepsin,  $\alpha$ -chymotrypsin ( $\alpha$ -Chymo), and collagenase was measured as described in Experimental Procedures. In each case activity was measured with and without the prodomain and the percentage inhibition calculated. Error bars represent standard deviations from two experiments, each performed in duplicate. doi:10.1371/journal.pone.0005694.g003



**Figure 4. Circular dichroism analysis of prodomain constructs.** Different falcipain-2 prodomain constructs (200  $\mu\text{g/ml}$ ) were incubated in 20 mM sodium phosphate, pH 5.8, and absorbance between 195 and 240 nm was measured.  
doi:10.1371/journal.pone.0005694.g004

using the crystallographic structures of several papain-family cysteine proteases as templates. We used the software MODELLER-9v4 (16) to construct a homology model of profalcipain-2 (Fig. 5a), which aligned to mature falcipain-2, procathepsin L, procathepsin K, and procaricain at sequence identities of 100% (by definition, aligned with the sequence of the mature domain only), 30.6%, 30.9%, and 32.1% respectively (14, 20–24). The model was evaluated with DOPE (Discrete Optimized Protein Energy), a pairwise atomic distance statistical potential that assesses atomic distances in a model relative to those observed in many known protein structures (17). The DOPE Z-score of the model ( $-0.99$ ) is similar to the Z-scores of all templates (cathepsin L:  $-1.62$ ; mature falcipain-2:  $-1.13$ ; procathepsin K  $-0.95$ ; procaricain  $-1.25$ ); generally, a Z-score of  $-1$  or less indicates a relatively accurate model, with more than 80% of its C $\alpha$  atoms within 3.5 Å of their correct positions (17). Additionally, a separate assessment technique, TSVMOD, was applied. This method predicts the native overlap (defined as the fraction of  $\alpha$ -carbon atoms within 3.5 Å of the native structure) of a homology model in the absence of a solved structure using support vector machine learning (18, 19). The model's predicted native overlap (0.85) was similar to that of a model of mature falcipain-2 built using the mature sections of the above templates, indicating the falcipain-2 prodomain does not contribute significantly disproportionately to the overall model error. This assessment suggests that the fold of the profalcipain-2 model is correct despite the relatively low sequence identity between the falcipain-2 prodomain and the templates.

#### The Profalcipain-2 Model Suggests that the Conserved Residues Provide Stability to the Overall Fold

We examined the homology model for possible interactions involving residues in the conserved motifs. Several of these residues are highlighted in Fig. 5b. (i) The charged pair Arg<sup>185</sup> and Glu<sup>221</sup> appears to form a salt bridge. (ii) Glu<sup>210</sup> from the GNFD motif may form a separate salt bridge with Lys<sup>403</sup> in the mature domain. (iii) Phe<sup>214</sup> may participate in non-polar interactions, and possibly  $\pi$ -bond stacking, with two tryptophan residues on the mature domain, Trp<sup>449</sup> and Trp<sup>453</sup>. All of these interactions are

also present in at least one of the templates used to build the model, although none of them is conserved across all templates.

#### The Falcipain-2 Prodomain Appears to Block Substrates from Entering the Cathepsin B Active Site

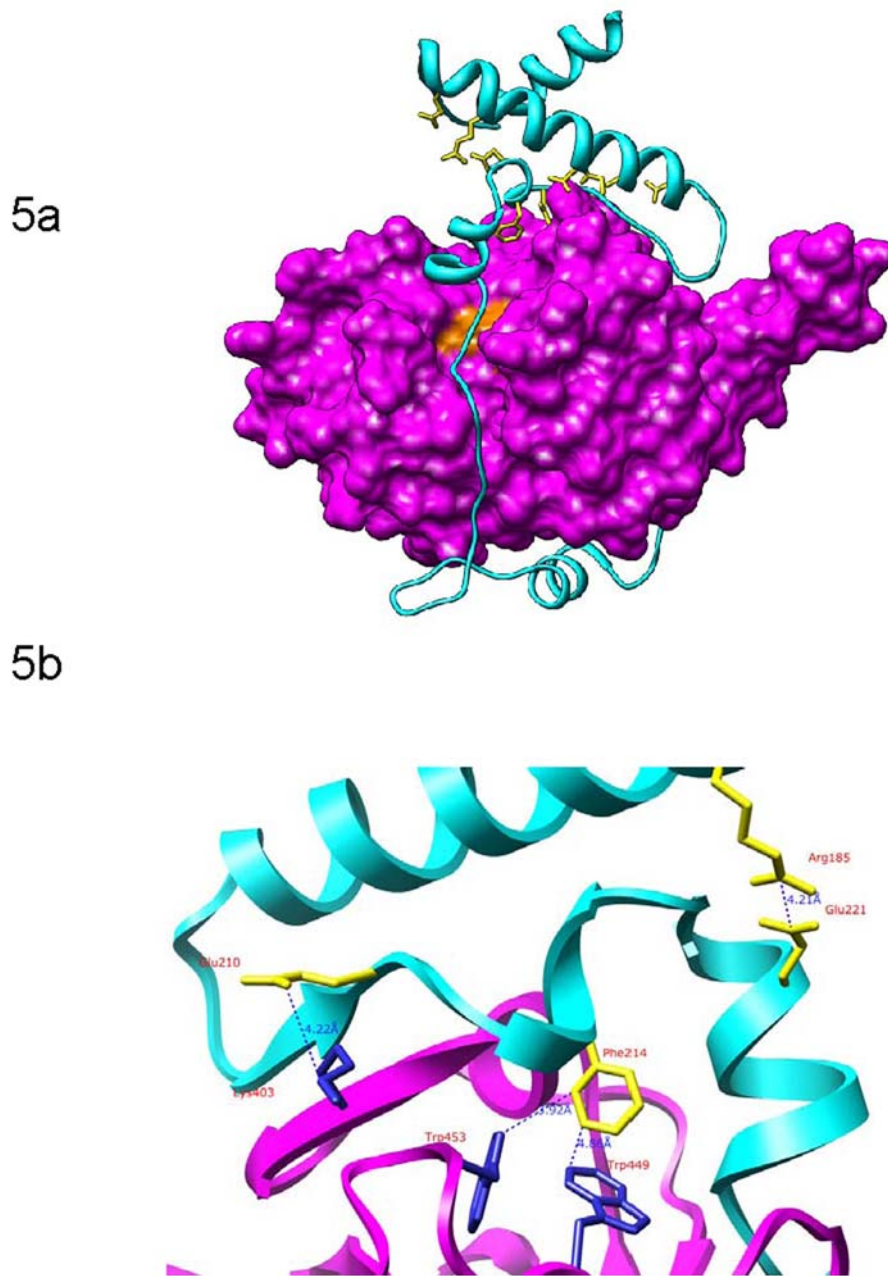
A separate homology model was constructed in which the falcipain-2 prodomain and cathepsin-B mature domain were modeled as a complex (Fig. 6a). The model was built based on an alignment of profalcipain-2 at 31.2% sequence identity with the crystallographic structure of procathepsin B (25, 26). The model received a DOPE Z score of  $-0.87$ , and a TSVMOD native overlap prediction of 0.82. These scores indicate that the overall fold is correct; poor scores would have suggested that there were significant errors in the modeled structure of the prodomain, and in that case the model would not have resembled the structures of the templates on which it was based. The model suggests that the prodomain of falcipain-2 binds mature cathepsin B in a manner similar to that observed in papain family zymogens, inhibiting catalytic activity by blocking substrate access to the active site. (Fig. 6b). While no structure has been solved for a propeptide in complex with an inhibited mature enzyme, it is likely that these propeptides bind to the enzymes in a conformation resembling the zymogen form (14, 25–27). This hypothesis is reflected in the model, which by construction is similar to its templates, and displays favorable stereochemistry and non-bonded atom distances as evaluated by MODELLER and DOPE.

#### Differences Between the Prodomains of Falcipain-2 and Cathepsin L

Cathepsin B activity is inhibited by the prodomain of falcipain-2 (Fig. 3) but not cathepsin L (10). To examine the structural basis of this selectivity, we compared the sequences and structures of these two proteins. Several differences were of note (Fig. 7). First, while the procathepsin L  $\alpha 1$  helix clashes with the occluding loop region of mature cathepsin B, thus preventing binding, the equivalent helix in falcipain-2 does not. Second, Phe<sup>186</sup> in profalcipain-2 participates in polar interactions with Phe<sup>165</sup> and Phe<sup>168</sup>; in procathepsin K and procathepsin L, Phe<sup>186</sup> is replaced by Arg. Third, a multiple sequence alignment reveals a conserved motif (LMNNAEHIN in falcipain-2) in the plasmodial proteases falcipain-2, falcipain-3, and berghepsin-2 that represents an insertion relative to the sequences of procathepsin K and procathepsin L (Fig. 1). Finally, an apparent salt bridge (interaction not shown) is formed between Glu<sup>210</sup> in the falcipain-2 prodomain and Lys<sup>184</sup> in mature cathepsin B; Glu<sup>210</sup> of falcipain-2 (which has replaced Gly in the GNFD motif) is replaced by Ala in cathepsin L and cathepsin K. Taken together, differences between modeled interactions for the cathepsin B mature domain with procathepsin L or profalcipain-2 appear to describe the structural basis for the observed selective inhibition of cathepsin B activity by profalcipain-2.

#### Discussion

We evaluated features of the falcipain-2 prodomain that mediate enzyme inhibition. Our data show that only an  $\sim 11$  kDa C-terminal region of the prodomain is required for potent inhibition of the protease. The region includes two hydrophobic residues (both Phe in falcipain-2) and the ERFNIN and GNFD motifs, all of which are conserved among cathepsin L-like papain family proteases. The falcipain-2 prodomain also inhibited other papain family cysteine proteases, including similar cathepsin L sub-family proteases and the more distantly related

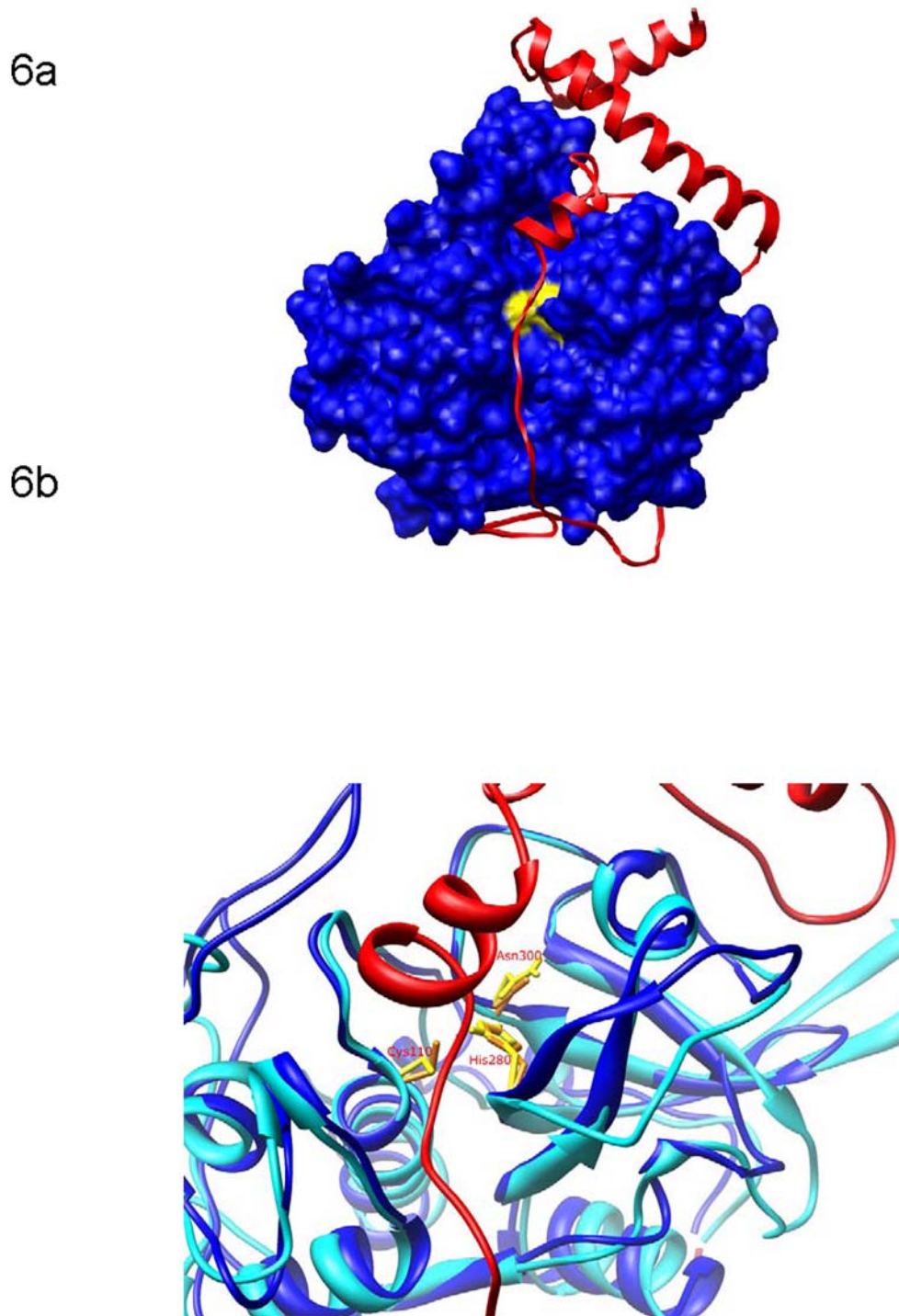


**Figure 5. Homology model of profalcipain-2. (a) Model created using MODELLER 9v4.** The 160 N-terminal residues of the prodomain are not included in the model. The prodomain (cyan) runs up the face of the mature enzyme (purple; catalytic triad residues in orange) before forming  $\alpha$ -helices containing the conserved ERFNIN and GNFD motifs (yellow). **(b) Close-up of several predicted interactions between the mature protease and the ERFNIN (R<sup>185</sup>) and GNFD (E<sup>210</sup>; F<sup>214</sup>) motifs.** Blue dashed lines indicate presumed stabilizing interactions (both electrostatic and hydrophobic) between residues. The structure has been rotated 180 degrees around the vertical axis from its representation in figure 5a. doi:10.1371/journal.pone.0005694.g005

cathepsin B. We explored the relevance of conserved falcipain-2 motifs by circular dichroism; the conserved residues were required to maintain the secondary structure of the prodomain. Thus, the first prerequisite for inhibitory activity was appropriate secondary structure. We also constructed a homology model of profalcipain-2 to help explain the observed experimental results. The model identified potential interactions between the inhibitory portion of the prodomain and mature falcipain-2 that appear to explain the inhibitory activity, and also the ability of the prodomain of falcipain-2, but not that of the related protease cathepsin L, to inhibit cathepsin B. Taken together, our results identify and

structurally characterize a minimum inhibitory domain of the falcipain-2 prodomain, offering a starting point for new considerations for the inhibition of key proteases of malaria parasites. Indeed, small molecules that inhibit falcipains via interactions independent of the active site might offer highly specific antimalarials without detrimental effects due to inhibition of host cysteine proteases.

Results of structure-function studies were straightforward. As expected, the upstream portion of the falcipain-2 prodomain, which mediates protein trafficking (9), was not required for inhibitory activity. Indeed, only a small portion of the prodomain

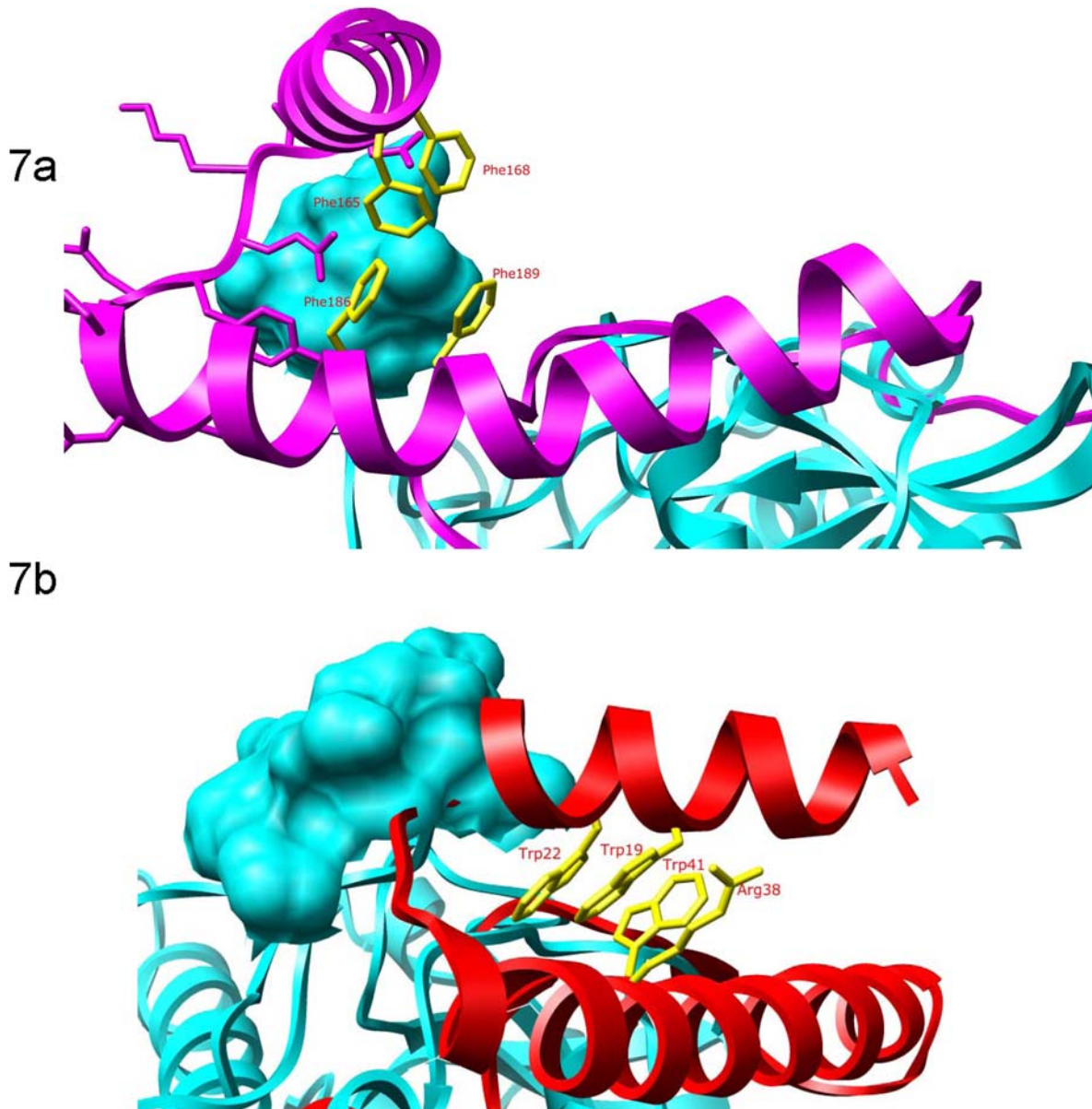


**Figure 6. Model rationalizing the inhibition of cathepsin B by the falcipain-2 prodomain. (a) Model of the falcipain-2 prodomain (red) and mature cathepsin B (blue; catalytic triad residues in yellow).** The prodomain binds to cathepsin B in a similar fashion as zymogens of other cysteine proteases, including procathepsin L and procathepsin B. **(b) Structural overlay of mature cathepsin B (blue) and falcipain-2 (cyan).** Catalytic triad residues are shown in the stick representation (yellow: cathepsin B; orange: falcipain-2). Cathepsin B amino acid numbering is used.

doi:10.1371/journal.pone.0005694.g006

(Leu<sup>155</sup>-Asp<sup>216</sup>) was required for sub-nanomolar inhibition of the mature enzyme. We did not demonstrate inhibition by the isolated Leu<sup>155</sup>-Asp<sup>216</sup> peptide, as production of this peptide proved difficult, but consideration of inhibition by a number of overlapping constructs (Fig. 2) clearly demonstrates that this peptide is

sufficient for inhibition of falcipain-2. Circular dichroism studies suggested that the limits of the minimum inhibitory domain are dictated by requirements for appropriate folding and maintenance of a secondary structure for the inhibitory portion of the prodomain.



**Figure 7. Modeled differences between falcipain-2 (a) and cathepsin L (b) prodomain binding to cathepsin B.** The model predicts a helix arrangement in the falcipain-2 prodomain (purple) that prevents steric clashes with the cathepsin B occluding loop (cyan). Phe<sup>186</sup> may mediate this arrangement; in cathepsin K and cathepsin L, Phe<sup>186</sup> is replaced by Arg. For cathepsin L (red), there is a large steric clash between the linker joining the two cathepsin L helices and the space-filled occluding loop (cyan).  
doi:10.1371/journal.pone.0005694.g007

Due to the conserved overall fold of cathepsin precursors (10), along with the high degree of structural similarity between these proteases and mature falcipain-2 (C $\alpha$  root-mean-square-deviation between falcipain-2 and cathepsin K is 0.92 Å; falcipain-2 and cathepsin L is 0.81 Å; and falcipain-2 and procaricain is 0.95 Å), profalcipain-2 is a good candidate for comparative modeling analysis. Our model has a good DOPE score, a pairwise atomic distance statistical potential that has been shown to perform well in evaluating errors in homology models (17). DOPE is particularly suited to determine the accuracy of the overall fold of a model. The DOPE score of the model of falcipain-2 was similar to those of mature falcipain-2 and procathepsin L, indicating that the overall fold of our homology model is accurate. A separate model assessment program, TSVMMod, gave essentially the same results.

In our model, residues in the ERFNIN and GNFD motifs were involved in several interactions important to the stability of the falcipain-2 prodomain fold (Fig. 5b). Two interactions, Arg<sup>185</sup>–Glu<sup>221</sup> and Phe<sup>214</sup>–Trp<sup>449</sup>/Trp<sup>453</sup>, appear to be conserved between falcipain-2 and cathepsin L, with equivalent residues present in procathepsin L (21). A third interaction, Glu<sup>210</sup>–Lys<sup>403</sup>, represents a unique charged pair interaction, as a Glu is found in falcipain-2, but not falcipain-3 or most related proteases, replacing the Gly in the GNFD motif. Side chain packing is the most difficult part of comparative modeling; however, in this case using the ERFNIN and GNFD motifs as well as the conserved Phe residues to guide the alignment resulted in conserved sequences across the downstream region of the prodomain (Fig. 1), increasing confidence in our predictions.

Many cathepsin L sub-family propeptides act in trans to inhibit related proteases (10). However, selectivity has been observed, and it has been demonstrated that the prodomains of cathepsin L and cathepsin K are unable to inhibit cathepsin B (25–27). Explanations for this observation include the following. First, cathepsin B lacks the ERFNIN motif, so that the protease lacks most of the  $\alpha 2$  helix found in cathepsin L sub-family proteases. Second, cathepsin B contains a large occluding loop insertion, conferring dipeptidase activity, but preventing propeptides containing the ERFNIN motif from binding due to a steric clash between the occluding loop and the prodomain residues connecting  $\alpha 1$  and  $\alpha 2$  (Fig. 7b). Interestingly, selectivity for prodomain inhibition was broader for falcipain-2, as the prodomain of falcipain-2 markedly inhibited cathepsin B (Fig. 3). Our homology model adds insight to this observation.

In the model, the interaction of the helices equivalent to cathepsin L helices  $\alpha 1$  and  $\alpha 2$  is shifted (Fig. 7a). This shift is mediated by the presence of an additional aromatic residue in falcipain-2, Phe<sup>186</sup>. This residue is part of the hydrophobic core of aromatic residues that contributes to the helix interaction in cathepsin L and cathepsin K, normally mediated by two Trp residues on  $\alpha 1$  and the Phe residue in the ERFNIN motif on  $\alpha 2$ . In falcipain-2, Phe<sup>186</sup> provides additional stability, allowing  $\alpha 1$  to shift across  $\alpha 2$  and eliminating the steric overlap between the prodomain residues and the cathepsin B occluding loop. In procathepsin L and procathepsin K, which do not inhibit cathepsin B, Phe<sup>186</sup> is replaced by Arg<sup>38</sup> (procathepsin L) and Arg<sup>41</sup> (procathepsin K); arginine is a basic residue that interacts less favorably with the other hydrophobic residues. (Fig. 7b).

A recent study indicated that a synthetic fifteen residue peptide (Leu<sup>155</sup>-Ile<sup>169</sup>) from a region of the falcipain-2 prodomain immediately upstream of conserved Phe residues (Phe<sup>165</sup> and Phe<sup>168</sup>) inhibited falcipain-2 (28). The authors proposed that this segment plays an important role in inhibition of falcipain-2. However, inhibition by the peptide was at much lower (10,000 times less) potency than inhibition by our prodomain constructs, which acted at sub-nanomolar concentrations. In our model, the Leu<sup>155</sup>-Ile<sup>169</sup> residues form the  $\alpha 1$ -helix. As noted, these residues represent an insertion relative to cathepsin L and cathepsin K. The  $\alpha 1$  helix does not appear to actively inhibit falcipain-2, but rather appears to provide structural stability through an interaction with the  $\alpha 2$  helix. It is thus likely that the full prodomain inhibits falcipain-2 differently from the small peptide studied recently (28), as for this peptide to come within the proximity of the falcipain-2 active site would require replacement of the  $\alpha 3$  helix and a novel fold relative to other papain-family proteases.

Our work defines the minimum inhibitory region of the falcipain-2 prodomain. We show that several residues conserved across cathepsin L sub-family proteases are necessary for this inhibition, and present a structural model for the interaction of the falcipain-2 prodomain with both its own mature domain and that of other proteases. As natural inhibitors of parasite protease activity, propeptides present a promising basis for design of small molecules to treat malaria.

## Materials and Methods

### Reagents

Benzyloxycarbonyl-Leu-Arg-7-amino-4 methyl coumarin (Z-Leu-Arg-AMC) and Z-Phe-Arg-AMC were from Peptides International. Restriction endonucleases and polymerases were from New England Biolabs. Oligonucleotides were synthesized at the

Biomolecular Resource Center, University of California, San Francisco, and by Integrated DNA Technologies. The synthetic peptide was from AnaSpec. All other reagents were from Sigma-Aldrich or as mentioned in the text.

### PCR and Sequencing

All DNA fragments were amplified from the pTOP-FP2 plasmid, which encodes the falcipain-2 gene (6). The sequence of each construct was confirmed by DNA sequencing at the Biomolecular Resource Center, University of California, San Francisco. Portions of the falcipain-2 gene were amplified using primers specific for each construct (Table S1).

### Cloning, Expression, and Refolding of Different Prodomain Constructs

Amplified DNA fragments were digested with *Bam*HI and *Hind*III, ligated into digested plasmids (pRSET-B; Invitrogen) and used to transform AD (DE3) pLys *E. coli* (Invitrogen). Cells were induced with  $\beta$ -D-thio-galactopyranoside, and recombinant proteins were solubilized in 8 M urea, 20 mM Tris-Cl, pH 8.0 at room temperature for 60 min with gentle shaking. Insoluble material was separated by centrifugation at 27,000 g for 30 min at 4°C. For the purification of the recombinant protein, the supernatant was incubated with nickel-nitrilotriacetic acid resin (Ni-NTA; Qjagen) and purified under denaturing conditions, as previously described (6). Ni-NTA purified propeptides were bound to SP-sepharose columns (Amersham Bioscience) and eluted by a step-wise gradient of 0–1 M NaCl in 8 M urea, 20 mM Tris-Cl, pH 8.0. The denatured proteins were diluted 100-fold (final concentration 20  $\mu$ g/ml) in 100 mM Tris-Cl, 1 mM EDTA, 250 mM L-arginine pH 9.0, refolded at 10–12°C for 20 h, and concentrated using a 10 kDa cut-off membrane (Millipore) to 10 ml. Insoluble protein was removed using a 0.45  $\mu$ m syringe filter (Millipore).

### Inhibition of Falcipain-2 by the Prodomain

Inhibitor kinetics were calculated as previously described (15). In brief, different concentrations of prodomain constructs (2–50 nM) were pre-incubated with 2 nM falcipain-2 in 100 mM sodium acetate, 5 mM DTT, pH 5.5 for 10 min at room temperature. The substrate Z-Leu-Arg-AMC (10  $\mu$ M) was added, and fluorescence (excitation 355 nm; emission 460 nm) was continuously measured for 20 min at room temperature with a Labsystems Fluroskan Ascent spectrofluorometer. Enzyme concentration was determined by titration with the irreversible inhibitor morpholine urea-phenylalanine-homophenylalanine fluoromethyl ketone.  $K_i$  values were determined by nonlinear regression analysis using PRISM (GraphPad Software).

### Inhibition of other Proteases by the Falcipain-2 Prodomain

Substrates were Z-Leu-Arg-AMC (10  $\mu$ M) for falcipain-2, falcipain-3, and cruzain; Z-Phe-Arg-AMC (10  $\mu$ M) for cathepsin L and cathepsin K; Z-Arg-Arg-AMC (10  $\mu$ M) for cathepsin B; Pro-Arg-AMC (10  $\mu$ M) for cathepsin C; and FITC-casein (8  $\mu$ g/ $\mu$ l) for the other studied proteases. For each reaction, 1  $\mu$ g of purified falcipain-2 prodomain (or, for controls, no prodomain) and 2–10 nM of each enzyme were incubated for 10 min in 350  $\mu$ l of 100 mM sodium acetate, 5 mM DTT, pH 5.5 (for  $\alpha$ -chymotrypsin and collagenase 10 mM Tris, pH 7.5), substrate was added, and substrate hydrolysis was monitored as described above or, for FITC-casein, as previously described (8).



## Circular Dichroism

Experiments were performed on a Jasco J-175 spectropolarimeter. Signals were monitored between 195 and 300 nm in 20 mM sodium phosphate, pH 5.8 at 20°C. Purified proteins were concentrated (200 µg/ml) using a 10-kDa cutoff Amicon ultra-concentrator (Millipore) and transferred to the phosphate buffer. All experiments were performed in a quartz cell of 1 cm path length (Hellma).

## Falcipain-2 Modeling

Falcipain-2 residues 161–484, encompassing the full mature domain and the C-terminal region of the prodomain, were aligned with procathepsin L, procathepsin K, and procaricain, at sequence identities of 20–25% in the prodomain region. 100 homology models were built based on the crystallographic structures of these proteins as templates (PDB codes were 1CS8, 1BY8, and 1PCI, respectively) and the crystallographic structure of mature falcipain-2 (1YVB), using the standard ‘automodel’ routine of MODELLER-9v4 (16). Models were evaluated with the Z-DOPE statistical potential (17) and the TSVMOD protocol for predicting absolute model error (18). The model receiving the best Z-DOPE score was subjected to loop refinement of residues 15–20 (sequence NKQYNS), restraining the first 14 residues to a helical conformation, using the ‘loop’ routine of MODELLER-9v4 (19).

## Cathepsin-B Modeling

The prodomain of falcipain-2 was modeled in complex with the crystallographic structure of mature cathepsin B. The same homology modeling and loop modeling procedures were performed as for falcipain-2, here based on the crystallographic

structures of the prodomain regions of procathepsin L, procathepsin K, and procaricain, and the solved structure of procathepsin B (PDB code 3PBH), as templates. Structural alignments of procathepsin L and cathepsin B in Fig. 7b were performed with the SALIGN command of MODELLER-9v4 (20).

## Supporting Information

**Figure S1** Expression of profalcipain-2 constructs. Different constructs were expressed in *E. coli*, and purified as described in Experimental Procedures. For each construct, 4 µg of protein was solubilized in SDS sample buffer, electrophoresed in a 12 % SDS-PAGE gel, and stained with Coomassie blue.

Found at: doi:10.1371/journal.pone.0005694.s001 (0.16 MB TIF)

**Table S1** Primers (Forward, F and Reverse, R) used to amplify DNA encoding the constructs shown are listed. Restriction endonuclease cleavage sites are in bold type.

Found at: doi:10.1371/journal.pone.0005694.s002 (0.04 MB DOC)

## Acknowledgments

We thank Katherine J. Walker, Ji H. Lee, and Jiri Gut for technical assistance and Puran S. Sijwali for useful advice.

## Author Contributions

Conceived and designed the experiments: KCP DTB AS PJR. Performed the experiments: DTB. Analyzed the data: KCP DTB AS PJR. Contributed reagents/materials/analysis tools: KCP PJR. Wrote the paper: KCP DTB PJR.

## References

- Greenwood BM, Bojang K, Whitty CJ, Targett GA (2005) Malaria. *Lancet* 365: 1487–1498.
- Rosenthal PJ (2004) Cysteine proteases of malaria parasites. *Int J Parasitol* 34: 1489–1499.
- Shenai BR, Sijwali PS, Singh A, Rosenthal PJ (2000) Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of *Plasmodium falciparum*. *J Biol Chem* 275: 29000–29010.
- Sijwali PS, Shenai BR, Gut J, Singh A, Rosenthal PJ (2001) Expression and characterization of the *Plasmodium falciparum* haemoglobinase falcipain-3. *Biochem J* 360: 481–489.
- Sijwali PS, Rosenthal PJ (2002) Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 101: 4384–4389.
- Sijwali PS, Shenai BR, Rosenthal PJ (2002) Folding of the *Plasmodium falciparum* cysteine protease falcipain-2 is mediated by a chaperone-like peptide and not the prodomain. *J Biol Chem* 277: 14910–14915.
- Pandey KC, Sijwali PS, Singh A, Na BK, Rosenthal PJ, et al. (2004) Independent intramolecular mediators of folding, activity, and inhibition for the *Plasmodium falciparum* cysteine protease falcipain-2. *J Biol Chem* 279: 3484–3491.
- Pandey KC, Wang SX, Sijwali PS, Lau AL, McKerrow JH, et al. (2005) The *Plasmodium falciparum* cysteine protease falcipain-2 captures its substrate, hemoglobin, via a unique motif. *Proc Natl Acad Sci USA* 102: 9138–9143.
- Subramanian S, Sijwali PS, Rosenthal PJ (2007) Falcipain cysteine proteases require bipartite motifs for trafficking to the *Plasmodium falciparum* food vacuole. *J Biol Chem* 282: 24961–24969.
- Wiederanders B, Kaulmann G, Schilling K (2003) Functions of propeptide parts in cysteine proteases. *Curr Protein Pept Sci* 4: 309–326.
- Karrer KM, Peiffer SL, DiTomas ME (1993) Two distinct gene subfamilies within the family of cysteine protease genes. *Proc Natl Acad Sci U SA* 90: 3063–3067.
- Kreusch S, Fehn M, Maubach G, Nissler K, Rommerskirch W, et al. (2000) An evolutionarily conserved tripartite tryptophan motif stabilizes the prodomains of cathepsin L-like cysteine proteases. *Eur J Biochem* 267: 2965–2972.
- Vernet T, Berti PJ, de Montigny C, Musil R, Tessier DC, et al. (1995) Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Proregion participates in the regulation of intramolecular processing. *J Biol Chem* 270: 10838–10846.
- Coulombe R, Grochulski P, Sivaraman J, Menard R, Mort JS et al; M (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J* 15: 5492–5503.
- Pandey KC, Singh N, Arastu-Kapur S, Bogyo M, Rosenthal PJ (2006) Falcipain, a cysteine protease inhibitor of *Plasmodium falciparum*, facilitates erythrocyte invasion. *PLoS Pathog* 2: e117.
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*; Chapter 2, Unit 2.9.
- Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15: 2507–2524.
- Eramian D, Eswar N, Shen MY, Sali A (2008) How well can the accuracy of comparative protein structure models be predicted? *Protein Sci* 17: 1881–93.
- Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9: 1753–1773.
- Marti-Renom MA, Madhusudhan MS, Sali A (2004) Alignment of protein sequences by their profiles. *Protein Sci* 13: 1071–1087.
- Guay J, Falgueyret JP, Ducret A, Percival MD, Mancini JA, et al. (2000) Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. *Eur J Biochem* 267: 6311–6318.
- Wang SX, Pandey KC, Somoza JR, Sijwali PS, Kortemme T, et al. (2006) Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease. *Proc Natl Acad Sci USA* 103: 11503–11508.
- LaLonde JM, Zhao B, Janson CA, D’Alessio KJ, McQueney MS, et al. (1999) The crystal structure of human procathepsin K. *Biochemistry* 38: 862–869.
- Groves MR, Taylor MA, Scott M, Cummings NJ, Pickersgill RW, et al. (1996) The prosequence of procaricain forms an alpha-helical domain that prevents access to the substrate-binding cleft. *Structure* 4: 1193–1203.
- Cyglar M, Sivaraman J, Grochulski P, Coulombe R, Storer AC, et al. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure* 4: 405–416.
- Turk D, Podobnik M, Kuhelj R, Dolinar M, Turk V, et al. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Angstroms resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett* 384: 211–214.
- Groves MR, Coulombe R, Jenkins J, Cyglar M (1998) Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes. *Proteins* 32: 504–514.
- Korde R, Bhardwaj A, Singh R, Srivastava A, Chauhan VS, et al. (2008) *J Med Chem* 51: 3116–3123.