Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease

Stephanie X. Wang, Kailash C. Pandey, John R. Somoza, Puran S. Sijwali, Tanja Kortemme, Linda S. Brinen, Robert J. Fletterick, Philip J. Rosenthal, and James H. McKerrow

PNAS 2006;103;11503-11508; originally published online Jul 24, 2006;
doi:10.1073/pnas.0600489103

Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease

Stephanie X. Wang*, Kailash C. Pandey†, John R. Somoza‡, Puran S. Sijwali†, Tanja Kortemme§, Linda S. Brinen¶, Robert J. Fletterick‖, Philip J. Rosenthal†, and James H. McKerrow*§**

*Department of Pathology and the Sandler Center, Box 2550, Byers Hall N508, and †Department of Medicine, San Francisco General Hospital, Box 0811, University of California, San Francisco, CA 94143; §Department of Biopharmaceutical Sciences and California Institute for Quantitative Biomedical Research, and Departments of ¶Cellular and Molecular Pharmacology and ‖Biochemistry and Biophysics, University of California, San Francisco, CA 94143; and ‡Celera Genomics, 180 Kimball Way, South San Francisco, CA 94080

Edited by Robert M. Stroud, University of California, San Francisco, CA, and approved June 12, 2006 (received for review January 22, 2006)

Falcipain-2 (FP2), the major cysteine protease of the human malaria parasite *Plasmodium falciparum*, is a hemoglobinase and promising drug target. Here we report the crystal structure of FP2 in complex with a protease inhibitor, cystatin. The FP2 structure reveals two previously undescribed cysteine protease structural motifs, designated $FP2_{nose}$ and $FP2_{arm}$, in addition to details of the active site that will help focus inhibitor design. Unlike most cysteine proteases, FP2 does not require a prodomain but only the short $FP2_{nose}$ motif to correctly fold and gain catalytic activity. Our structure and mutagenesis data suggest a molecular basis for this unique mechanism by highlighting the functional role of two Tyr within $FP2_{nose}$ and a conserved Glu outside this motif. The $FP2_{arm}$ motif is required for hemoglobinase activity. The structure reveals topographic features and a negative charge cluster surrounding $FP2_{arm}$ that suggest it may serve as an exo-site for hemoglobin binding. Motifs similar to $FP2_{nose}$ and $FP2_{arm}$ are found only in related plasmodial proteases, suggesting that they confer malaria-specific functions.

cysteine protease | falcipain 2 | inhibitor | malaria | x-ray structure

An estimated 500 million people are infected with malaria worldwide, resulting in over a million deaths annually, mainly among children in subSaharan Africa (www.who.int/health_topics/malaria/en). There is an urgent need for new antimalarial therapy, because widespread drug resistance has limited the utility of most available treatments (1). *Plasmodium falciparum* is responsible for nearly all severe illness and death from malaria. During its life cycle within host red blood cells, the protozoan parasite relies on hemoglobin hydrolysis to supply amino acids for protein synthesis and to maintain osmotic stability (2, 3). The proteases involved in hemoglobin hydrolysis have been validated as promising drug targets (2). Both aspartyl proteases (plasmepsins) and cysteine proteases (falcipains) play a role. Inhibition of either protease class is lethal to the parasite, and inhibition of both is synergistic for parasite killing (4–6). Falcipain-2 (FP2), falcipain-2′ (FP2′), and falcipain-3 (FP3) are papain-family (C1) Clan CA cysteine proteases that cleave native or denatured human hemoglobin (7, 8). FP2, the most-abundant and best-studied enzyme among the falcipains, is a prime target for drug development. When the FP2 gene was deleted from *P. falciparum* by homologous recombination, trophozoites accumulated undegraded hemoglobin (9). Moreover, synthetic inhibitors targeting C1 proteases halted *P. falciparum* growth in cell culture and cured malaria in animal models (6, 10–12).

When compared with family C1 proteases from other organisms, mature falcipains contain two unique sequence motifs: an extension at the N terminus (17 aa for FP2) and an insertion between the catalytic His and Asn near the C terminus (14 aa for FP2). Although the N-terminal extension has been shown to mediate folding of FP2 in the absence of a prodomain (13, 14), the mechanism by which it achieves such a function remains largely unknown. The functional role of the C-terminal insertion has not been explored. Modeling analyses of FP2 and FP3 did not identify mechanistic or functional roles for these sequence motifs because of uncertainties in their conformations (15, 16). X-ray-derived structural data would therefore be key to elucidating the functions of the unique motifs and to guide rational drug design.

Here we report the structure of FP2 complexed with chicken egg white cystatin. The two "inserted" sequence motifs in FP2, previously undescribed for cysteine proteases, are found to correspond to unique surface structures. Structural, biochemical, and sequence data suggest that these unique structural elements participate in FP2 folding and host hemoglobin degradation. In addition, active-site properties and protein–protein interactions in the FP2–chicken egg white cystatin complex provide information for potential optimization of inhibitors.

Results and Discussion

The structure of FP2 has been a long- sought-after goal for supporting rational drug design against malaria. Previous attempts to crystallize FP2 with small-molecule inhibitors resulted only in poorly diffracting crystals that failed to yield a structure (ref. 17; L.S.B., J.R.S., S.X.W., and G. C. Cameron, unpublished data). To improve crystallization results, we tested macromolecular cysteine protease inhibitors with the expectation that protein–protein interactions over large surfaces would enhance crystal qualities and thus facilitate structural determination. Cystatins inhibit a wide range of papain-family cysteine proteases with high affinity, and the structures of multiple cystatins are available (18–20), making them ideal candidates for cocrystallization with FP2. Indeed, FP2 and chicken egg white cystatin (heretofore referred to as cystatin) formed a 1:1 complex that crystallized and diffracted well to produce an x-ray structure at 2.7 Å.

General Structural Features of FP2. FP2 is a single polypeptide chain of 241 aa. Overall, it adopts a classic papain-like fold in which the protease is divided into L and R domains (Fig. 1*a*). The catalytic residues†† Cys-25, His-159, and Asn-175 are located at the junction between the two domains. When FP2 is compared with papain, human cathepsin K and L, and cruzain over the 114 most-conserved

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: FP2, falcipain-2; PDB, Protein Data Bank.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1YVB).

**To whom correspondence should be addressed. E-mail: jmck@cgl.ucsf.edu.

††The numbering of FP2 follows papain convention (21). Residues that appear to be "extra" by papain convention are labeled with residue numbers that specify the location of insertion, followed by small alphabet letters to specify the position of the insertion. For example, the first residue in the C-terminal motif that follows residue 169 is labeled as Glu-169a, whereas the Tyr residue at the end of the same motif is labeled as Tyr-169n. Residues in the N-terminal motif that precede residue 1 by papain numbering are numbered from 0 to −16. Residues in cystatin are numbered with superscripts in the format of I#, where I is to the chain identification, and # is the residue number.

BIOCHEMISTRY

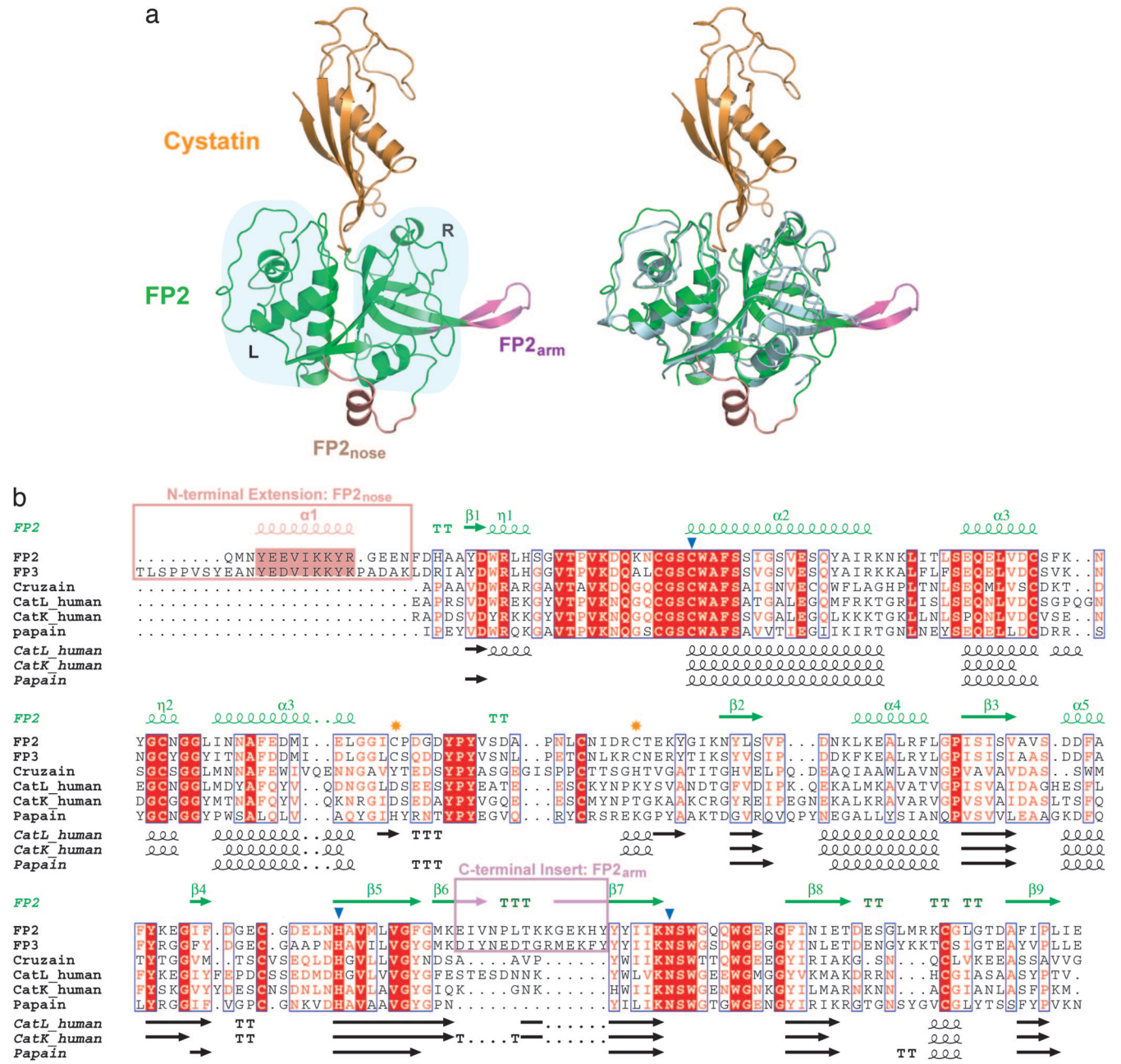

Fig. 1. Structure of FP2–cystatin. (*a*) Overall structure of FP2–cystatin is depicted in stereo in a conversational orientation for papain-like cysteine proteases: with the L domain (*Left*), the R domain (*Right*), and the active site on the top facing forward. Cystatin is colored in orange, and FP2 protease is colored in green with the N-terminal extension and C-terminal insertion highlighted in pink and salmon, respectively. The L and R domains are highlighted by shades of blue and yellow, respectively. The shaded area also corresponds to the size of human cathepsin H (superimposed on FP2 in light blue). (*b*) Amino acid sequence of the mature proteases FP2, FP3, cathepsin K, cathepsin L, papain, and cruzain are aligned by using the ClustalW program (29). The secondary structure of FP2 is depicted above its sequence and compared with those of cathepsins K and L. Sequence motifs resembling $FP2_{nose}$ and $FP2_{arm}$ are highlighted in salmon and pink boxes. The helical folding motifs in FP2 and FP3 are highlighted in a pink shade. The two cysteines that form a disulfide bond not found in papain are highlighted with orange stars.

amino acids, the largest structural deviation is only 1.1 Å, indicating little conformational change among these residues. The conserved residues come from both L and R domains and comprise the protease active site. In comparison to papain and the human cathepsins, FP2 has one additional disulfide bond that was required for its proteolytic activity (16). The most extraordinary features of FP2, however, are the two sequence motifs located at the N terminus and near the C terminus of the mature protein. Similar motifs are found in FP3 and related plasmodial proteases but not in papain, cruzain, or other papain (C1) cysteine proteases (Fig. 1*b*). In FP2, they add two unique structural elements: a "nose-like" projection connecting the L and R domains and an "arm-like" structure extending away from the protease surface. These two motifs are labeled as $FP2_{nose}$ and $FP2_{arm}$, respectively. The two orthogonal motifs have rendered the dimensions of FP2 (66.5 × 41.3 × 39.0 Å) significantly larger than those of papain (49.8 × 30.9 × 37.4 Å) when the molecules are oriented similarly. In addition to the apparent size increase, the structure of these unique motifs allows the postulation and testing of hypotheses regarding their biological functions.

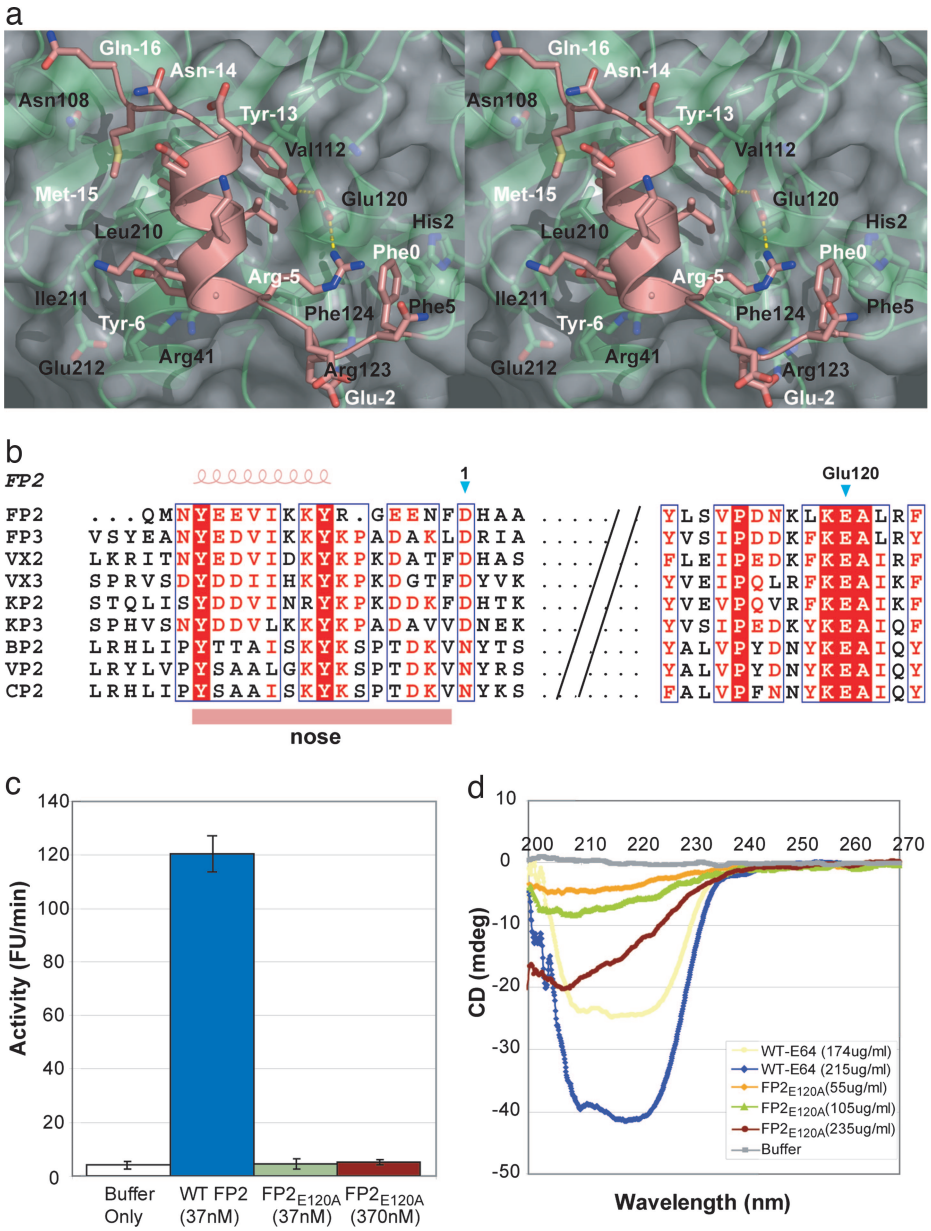

Fig. 2. Functional $FP2_{nose}$ in protein folding. (*a*) Interactions between $FP2_{nose}$ and the rest of the protease are depicted in stereoview, with $FP2_{nose}$ in pink and the protease core in green. All 17 amino acids in the nose region and residues from the protease core that make contact with the nose are shown in stick representation. Residues from the nose are marked in white labels with an N after their residue numbers. (*b*) Sequence alignment of plasmodial cysteine proteases from five different *Plasmodium* parasites (*Plasmodium vivax*, VX2 and VX3; *Plasmodium knowlesi*, KP2 and KP3; *Plasmodium berghei*, BP2; *Plasmodium vinckei*, VP2; and *Plasmodium chabaudi*, CP2) reveals a conserved 9-aa motif that may be important for protein folding as well as a KEA motif in the middle of the protease that interacts with the $FP2_{nose}$ motif. (*c*) The catalytic activity of $FP2_{E120A}$ is compared with wild-type FP2, using Z-Leu-Arg-AMC as a substrate as described (7). Results are presented as fluorescence unit (FU) per min. Error bars represent standard deviation from two assays, each performed in triplicate. (*d*) Data from CD analyses on $FP2_{E120A}$ mutant and wild-type FP2. Absorbance between 200 and 270 nm was depicted for samples at various concentrations; 1 μM of E64 was added to both samples.

Structural Requirements for $FP2_{nose}$-Mediated FP2 Folding. A surprising conclusion from biochemical analysis of recombinant falcipain expression was that active properly folded enzyme could be produced in the absence of the relatively long prodomain commonly found in the papain (Clan CA) family of proteases. Deletional analysis of the N terminus of FP2 (13, 14) confirms that the $FP2_{nose}$ was necessary for proper folding. Structural analysis now reveals that $FP2_{nose}$, in fact, has a short but significant element of secondary structure, and that Glu 120 forms a buried hydrogen bond with tyrosine 13 and a salt bridge with arginine 5. This suggests that $FP2_{nose}$, although considerably shorter than the standard papain family prodomain, may still play a significant role as a template or chaperone for protease folding. In contrast to other members of the papain family, interactions of a residue in the protein core (Glu-120) with the $FP2_{nose}$ may provide the additional binding energy necessary to stabilize initial interactions for folding. Thus, FP2 and related hemoglobin-degrading proteases from malaria parasites may represent a papain family subset in which an alternative mode of protein folding has evolved.

$FP2_{nose}$ consists of 17 aa at the N terminus of the mature FP2. $FP2_{nose}$ allowed proper folding of FP2 without a prodomain. Serial truncation analyses further identified a 9-aa segment YEEVIKKYR as the functional motif within $FP2_{nose}$ (14). Multiple secondary structure prediction programs confirm that

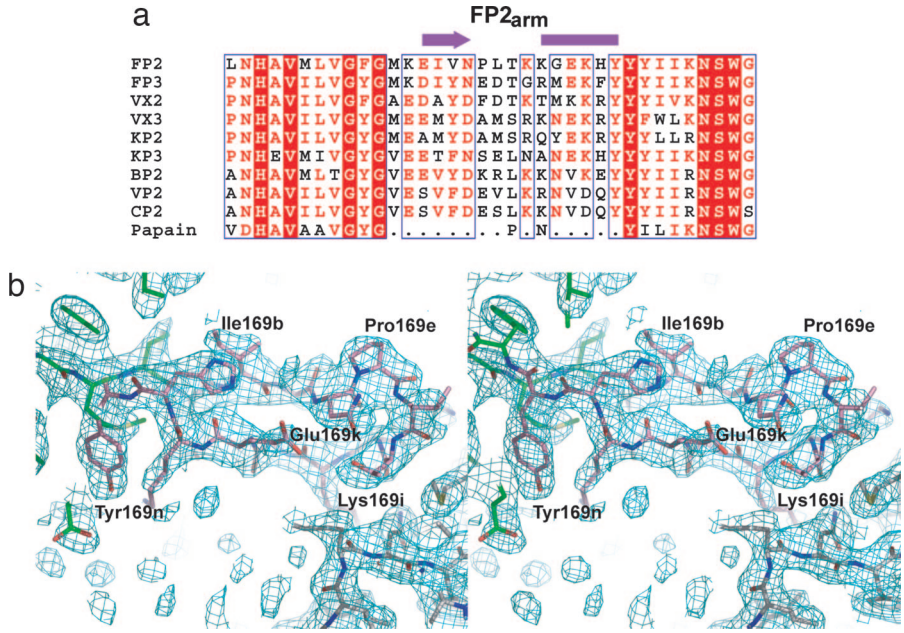

Fig. 3. Structure and conservation of FP2 arm. (*a*) Putative arm regions from FP2-like cysteine proteases from five *Plasmodium* parasites are aligned by the ClustalW program. (*b*) $3F_o-2F_c$ electron density map of $FP2_{arm}$ rendered at 1.2 σ is depicted in stereoview. Electron densities surrounding $FP2_{arm}$ are contributed by FP2 residues from a neighboring symmetry mate (with carbons colored in white).

YEEVIKKYR- and YEEVIKKYR-like sequences from other malarial proteases most likely adopt helical structures (details in Table 1, which is published as supporting information on the PNAS web site). As depicted in Fig. 2*a*, the helix in FP2 establishes extensive interactions with both L and R domains of the protease core. Specifically, Tyr-13 and Tyr-6 at the ends of the helix are tightly tucked into two hydrophobic pockets formed by numerous residues from the protease core. Tyr-13 forms an additional buried hydrogen bond with Glu-120. The two Tyr residues thus determine the orientation of the helix, and their contribution is further strengthened by a salt bridge between Arg-5 and Glu-120. Other less-conserved residues within the helical motif, such as the negatively charged Glu-2 and a hydrophobic Phe-0 at the end of $FP2_{nose}$, may also be important for interactions with the FP2 core. Overall, the strategically positioned $FP2_{nose}$ establishes strong interactions with both domains of FP2. We propose that these interactions confer local nucleation and stability, which lower the energy barrier for folding of FP2 into functional enzyme.

Sequence analyses support the current hypothesis of the function of $FP2_{nose}$. $FP2_{nose}$-like sequence motifs are found across a subclass of nine family C1 plasmodial cysteine proteases, which are also implicated in hemoglobin hydrolysis and likely share similar subcellular localization and chemical milieu (Fig. 2*b*). Indeed, the YEEVIKKYR-like sequences can be swapped between the proteases that harbor them without loss of folding (14). Within these sequences, the two Tyr and the conserved positive charge (Arg or Lys) are most highly conserved, consistent with their proposed role in folding. The proposed mechanism also explains why FP2 failed to gain catalytic activity once its N terminus was truncated beyond Tyr-13 (14), because the folding nucleus of $FP2_{nose}$ could no longer be sustained. The highly homologous nature of the YEEVIKKYR-like sequences suggests a shared folding mechanism for FP2 subfamily proteases involved in hemoglobin hydrolysis within the food vacuole of the trophozoite stage.

The YEEVIKKYR-like sequence alone, however, is not sufficient to fold and activate a homologous family C1 protease. For example, the distantly related plasmodial cysteine protease falcipain-1 (FP1) lacks the N-terminal nose-like motif, in particular the critical YEEVIKKYR-like sequence. A $FP2_{nose}$-FP1 chimera, created by adding $FP2_{nose}$ in front of the FP1 catalytic domain, did not exhibit enzymatic functional (P.S.S. and Y. C. Choe, data not shown). The failure of the chimera is likely attributable to the subtle, but key, differences between the catalytic domains of FP2 and FP1. Through BLAST search and sequence alignment, we identified Glu-120 as one of the key residues that might participate in the $FP2_{nose}$-mediated folding by interacting with $FP2_{nose}$. Glu-120 was found to be a part of a KEA motif in FP2 and related proteases harboring an $FP2_{nose}$-like folding motif, but not in FP1 or any other family C1 proteases (Fig. 2*b*). In the FP2 structure, Glu-120 forms a buried hydrogen bond with Tyr-13 and a salt bridge with Arg-5, both most likely critical in providing the binding energy necessary to stabilize the interactions between $FP2_{nose}$ and the protease core (Fig. 2*a*). Site-directed mutagenesis was carried out to convert Glu-120 to an Ala to test its hypothesized role in folding. The resulting $FP2_{E120A}$ mutant failed to gain catalytic activity after a standard refolding procedure (Fig. 2*c*). Because Glu-120 is a charged surface residue that does not participate in proteolysis, the drastic difference in catalytic activity would be unexpected unless the residue is important to other processes such as protein folding. CD analysis showed that the $FP2_{E120A}$ mutant, even at very high concentration, produced only weak CD signals that were distinctly different from those of wild-type FP2 (Fig. 2*d*). The lack of proper secondary structure in the $FP2_{E120A}$ mutant is consistent with a folding failure. It also explains the absence of catalytic activity. Our analyses thus have shown that the absence of the negative charge at Glu-120 indeed prevented FP2 folding into a functional enzyme. We therefore propose a molecular mechanism for the folding of FP2: the YEEVIKKYR sequence in $FP2_{nose}$ forms a stable helix that serves as an anchor. Subsequent folding and activation are further facilitated by residues from the main body of the protease, for example, Glu-120.

$FP2_{arm}$ Participates in Hemoglobin Recognition. At the distal end of the R domain and between the catalytic His-159 and Asn-175, 14 aa constitute $FP2_{arm}$, the second structural element unique to the FP2 subfamily plasmodial hemoglobinases (Figs. 1 and 3*a*). As its name suggests, $FP2_{arm}$ is a protruding "arm"-like structure of two extended β strands connected by an abrupt turn (Fig. 3*b*). Under

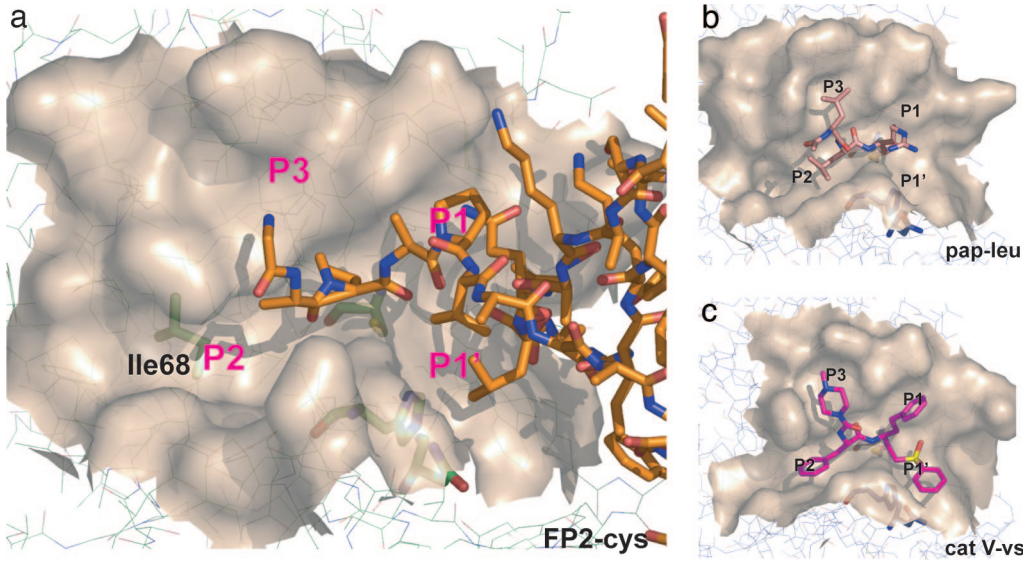

Fig. 4. Active sites of FP2 versus papain complexed with protein and cathepsin V inhibitors. (*a*) FP2–cystatin, (*b*) papain–leupeptin (PDB ID code 1POP), and (*c*) cathepsin V-vinyl sulfone (PDB ID code 1FH0) are depicted in the same orientation. The catalytic Cys, His, and Asn are depicted in stick mode in green. Substrate-binding pockets are labeled P3 to P1′. The labels are positioned precisely in the same spacing to reflect the size differences of the substrate-binding pockets. Ile-68 is also depicted in stick mode to illustrate a slight protrusion in the S2 pocket of FP2.

our crystallization conditions, $FP2_{arm}$ is stabilized by extensive crystal contacts with a neighboring symmetry mate in the same unit cell (Fig. 3*b*). A deep cleft is found between the base of $FP2_{arm}$ and the protease core, but the majority of $FP2_{arm}$ does not establish any contact with the protease core. As discussed below, mutagenesis studies suggest $FP2_{arm}$ may interact with the natural substrate of FP2, hemoglobin. However, because $FP2_{arm}$ is >25 Å away from the protease active site, $FP2_{arm}$ likely functions in hemoglobin hydrolysis through distal-site binding rather than direct participation in substrate cleavage. Surface charge distribution further suggests that $FP2_{arm}$ may achieve such a function through charge–charge interaction (*Supporting Text* and Fig. 5 *a* and *b*, which are published as supporting information on the PNAS web site). Finally, the structure also reveals that $FP2_{arm}$ is distantly located from the active site, so that it would not interfere with interactions between the protease and small ligands such as the inhibitors E64 and leupeptin.

The structure-based hypothesis of the function of $FP2_{arm}$ was tested by deleting 10 aa in $FP2_{arm}$ between Glu-169a and Gly-169j (22). The shortened FP2 showed wild-type-like catalytic activity against synthetic peptide substrates and the proteins casein and gelatin, suggesting that the catalytic function of the mutant is largely unaffected by the deletion. Most interestingly, however, this mutant specifically lost its catalytic activity against hemoglobin (22). Sequence analyses provide circumstantial support for the hypothesis that the loss of hemoglobin hydrolysis was due to loss of binding of FP2 to hemoglobin (22). $FP2_{arm}$-like motifs are found exclusively in the FP2 subfamily plasmodial hemoglobinases, although sequence homology in this motif is not as high as that of the $FP2_{nose}$-like motifs (Fig. 3*a*).

Although its importance is established by analysis of mutants, how $FP2_{arm}$ facilitates hemoglobin hydrolysis remains unknown. Surface charge analysis revealed that the region near the $FP2_{arm}$ motif was predominantly negatively charged, whereas positively charged patches are identified on human hemoglobin (Fig. 5 *a* and *b* and Table 2, which are published as supporting information on the PNAS web site). Best-ranked models of a FP2–hemoglobin complex generated by docking analysis consistently placed the complementarily charged surfaces near each other (experimental details, a FP2-hemoglobin model, and possible interactions are included in Fig. 6, which is published as supporting information on the PNAS web site, and Table 2). However, the FP2–hemoglobin model and proposed interactions between the two proteins remain speculative. Mutagenesis and structure analyses are necessary to uncover the interactions between FP2 and hemoglobin and whether $FP2_{arm}$ indeed directly participates in such interactions.

FP2–Cystatin and Other Cysteine Protease–Inhibitor Complexes. Outside of the two unique motifs discussed above, the rest of FP2 is structurally similar to homologous proteases in the C1 family. The FP2–cystatin complex shares many features with two known cysteine protease–cystatin structures: cathepsin H–stefin A and papain–stefin B (the stefins are a subclass within the cystatin superfamily; refs. 19 and 20). All three cystatins bind to target proteases in similar orientations at three discrete binding sites, resulting in extensive interactions with the target proteases. The buried surface area in the FP2–cystatin complex is ≈2,100 $Å^2$ (Fig. 4*a*). This large interaction surface, however, does not translate into tight binding. Leupeptin, which interacts with papain over only 700-$Å^2$ buried surface area, inhibits FP2 >30 times more effectively than does cystatin. Detailed analyses reveal that most of the binding energy for a small tight-binding inhibitor such as leupeptin derives from interactions at the P3 and P2 sites[‡‡] (Fig. 4 *b* and *c*), whereas cystatin largely interacts with FP2 at prime sites where substrate-binding pockets are relatively shallow and less defined (Fig. 4). Nevertheless, FP2 also has binding specificity at nonprimed sites. In fact, unlike most other cathepsin L-like cysteine proteases, FP2 strongly prefers a Leu at P2 over a Phe (7). Our structure shows that such a preference may be due to Ile-68 in FP2, because it creates a small protrusion in the otherwise flat base of the S2 pocket, which restricts aromatic side-chain binding (Fig. 4*a*). In addition, FP2 also interacts with cystatin at its nonprimed site, although the binding preference is not as pronounced as that found at the S2 pocket. It is nevertheless possible to predict that selective and potent synthetic inhibitors of FP2 could be designed by taking advantage of the cystatin-like prime-site interactions, the strong nonprime-site bind-

[‡‡]Substrate/inhibitor residues bound at the protease active site are labeled NH3. . . -P3-P2-P1-P1′-P2′-P3′. . . -COO- (23). P1–P1′ is the scissile bond. Residues N-terminal to the scissile bond form the nonprime site, and residues C-terminal to it constitute the prime site. The binding pockets within the protease that accommodate these residues are correspondingly labeled as . . . -S3-S2-S1-S1′-S2′-S3′-. . . .

BIOCHEMISTRY

ing found within active-site-based inhibitors, and the Leu specificity at P2. Rational designs of inhibitors against FP2 and FP3 are currently under way based on our structure through collaboration with GlaxoSmithKline (Tres Cantos, Spain).

Although the FP2–cystatin complex provides insights into protease–inhibitor interactions and serves a guide to inhibitor design, the most unusual and intriguing aspect of the FP2 structure remains $FP2_{nose}$ and $FP2_{arm}$. As noted above, a YEEVIKKYR-like motif, a KEA sequence, and a 14-aa insertion between the catalytic His and Asn are found in all FP2 subfamily proteases (Figs. 2*b* and 3*a*). Our structure and mutagenesis studies support the hypothesis that these functional motifs participate in protease folding, activation, and substrate binding.

Materials and Methods

Expressing, Solubilizing, Purifying, Refolding, and Assaying FP2. Expression, solubilization, refolding, and purification of FP2 were accomplished as described, with minor modifications (refs. 7 and 17; details in *Supporting Text*). All protein samples were concentrated by using a high-performance ultrafiltration cell (Model 2000, Amicon) with 10-kDa cutoff membrane at 4°C. Protein concentration was determined by Bradford assay (24). The kinetic activity of purified FP2 protein is assayed with Z-Leu-Arg-AMC, as described (22), with details included in *Supporting Text*.

Preparation of $FP2_{E120A}$. The $FP2_{E120A}$ mutant was generated by using an overlap-extension PCR (25) with primers that contain the point mutation (details in *Supporting Text*). Purified DNA fragment was first cloned in pGEM-T vector (Promega). The amplified DNA fragments were then released by BamHI and HindIII digestion, gel-purified, ligated into appropriately digested pQE-30 (Qiagen), and transformed into M15 (pREP4) *Escherichia coli* (Qiagen) for expression. Protein was purified and refolded as previously described for wild-type FP2. All sequences were confirmed by DNA sequencing.

Crystallization and Data Collection. Chicken egg white cystatin (Sigma) was added to activated FP2 until complete inhibition was achieved. The complex was then purified by ion-exchange chromatography on a Q-Sepharose column (Amersham Pharmacia). FP2–cystatin was crystallized in 10 mM Bis-Tris, pH 6.0, with 50 mM NaCl at 8–10 mg/ml. Thin rod-like crystals ($50 \times 40 \times 300$ μm) were grown with sitting drops (1 μl of protein + 1 μl of precipitant solution) in 0.1 M Hepes, pH 7.5/4.3 M NaCl at 4°C. Crystals were cryoprotected by soaking in mother liquors containing 5%, 10%, 15%, 20%, and 25% glycerol and subsequently frozen in liquid N_2 before data collection. Two data sets of FP2–cystatin were collected at beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA) at 100 K, using a CCD camera detector. The data sets were indexed, integrated, and merged by using DENZO/SCALEPACK (26) to yield a data set 96.5% complete at 2.7 Å (R_{merge} = 14.8%). The relatively high R_{merge} is due to high data redundancy and weak intensities of the reflections. Data processing and refinement details are tabulated in *Supporting Text*.

Structure Determination and Model Refinement. The FP2–cystatin crystals belong to the space group $P4_32_12$ (a = 96.04 Å, c = 124.16 Å). Each asymmetric unit contains one FP2–cystatin. The initial structure of FP2 was determined by molecular replacement with data up to 3.5 Å using a homology model based on human cathepsin K [Protein Data Bank (PDB) ID code 1ATK], using the rotational and translational functions from the Crystallography and NMR System software package (CNS 1.1; ref. 27). After a rigid body refinement, the first FP2 model gave an initial $R_{working}$ of 46.1%. A cystatin model (PDB ID code 1CEW) was then used in a second translational search to locate the bound cystatin. The FP2–cystatin model was refined by alternate cycles of energy minimization, simulated annealing, and group B factor refinement in the CNS suites. Model building and fitting were done by using QUANTA 2000 (Molecular Simulations, San Diego, CA). Complete data and refinement statistics are listed in Table 3, which is published as supporting information on the PNAS web site. All figures were generated with PyMOL (DeLano Scientific, South San Francisco, CA), and the surface electrostatic distribution was calculated by GRASP (28). The atomic coordinates and structure factors for the FP2–cystatin complex have been deposited in the PDB (ID code 1YVB).

CD to Evaluate the Folding of the $FP2_{E120A}$ Mutant. CD experiments were performed on a Jasco J-715 spectropolarimeter (Jasco, Easton, MD) with a peltier-effect temperature controller. CD signals were monitored between 190 and 300 nm, in 20 mM sodium phosphate, pH 5.8, at 20°C. Purified $FP2_{E120A}$ mutant and wild-type FP2, both in the presence of 1 μM E64, were concentrated by using a 10-kDa cutoff Amicon ultraconcentrator (Millipore) and transferred to the reaction buffer. All experiments were performed in a quartz cell of 1-cm path length from Hellma (Jamaica, NY). All solutions were stirred with a small magnetic stir bar while the temperature was continuously monitored by a thin probe inside the quartz cell.

We thank Dr. Brian Shoichet and Alan Graves for assistance with the CD experiments and Drs. Robert Fletterick, David Agard, Mohammed Sajid, Arthur Baca, Zachary Mackey, and Rajiv Bhatnagar for insightful discussions on the manuscript. We also thank Ryan Swenerton for critically reading our manuscript. P.J.R. was the principal investigator of the collaborating laboratory. This work was supported by the National Institutes of Health (Grants AI35707 and AI35800), the Sandler Family Supporting Foundation, and the Medicines for Malaria Venture. S.X.W. is a recipient of a postdoctoral fellowship from the American Heart Association. P.J.R. is a Doris Duke Charitable Foundation Distinguished Clinical Scientist.

1. White, N. J. (2004) *J. Clin. Invest.* **113,** 1084–1092.
2. Francis, S. E., Sullivan, D. J., Jr., & Goldberg, D. E. (1997) *Annu. Rev. Microbiol.* **51,** 97–123.
3. Lew, V. L., Tiffert, T. & Ginsburg, H. (2003) *Blood* **101,** 4189–4194.
4. Semenov, A., Olson, J. E. & Rosenthal, P. J. (1998) *Antimicrob. Agents Chemother.* **42,** 2254–2258.
5. Francis, S. E., Gluzman, I. Y., Oksman, A., Knickerbocker, A., Mueller, R., Bryant, M. L., Sherman, D. R., Russell, D. G. & Goldberg, D. E. (1994) *EMBO J.* **13,** 306–317.
6. Rosenthal, P. J., Wollish, W. S., Palmer, J. T. & Rasnick, D. (1991) *J. Clin. Invest.* **88,** 1467–1472.
7. Shenai, B. R., Sijwali, P. S., Singh, A. & Rosenthal, P. J. (2000) *J. Biol. Chem.* **275,** 29000–29010.
8. Sijwali, P. S., Shenai, B. R., Gut, J., Singh, A. & Rosenthal, P. J. (2001) *Biochem. J.* **360,** 481–489.
9. Sijwali, P. S. & Rosenthal, P. J. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 4384–4389.
10. Shenai, B. R., Lee, B. J., Alvarez-Hernandez, A., Chong, P. Y., Emal, C. D., Neitz, R. J., Roush, W. R. & Rosenthal, P. J. (2003) *Antimicrob. Agents Chemother.* **47,** 154–160.
11. Rosenthal, P. J., Lee, G. K. & Smith, R. E. (1993) *J. Clin. Invest.* **91,** 1052–1056.
12. Olson, J. E., Lee, G. K., Semenov, A. & Rosenthal, P. J. (1999) *Bioorg. Med. Chem.* **7,** 633–638.
13. Sijwali, P. S., Shenai, B. R. & Rosenthal, P. J. (2002) *J. Biol. Chem.* **277,** 14910–14915.
14. Pandey, K. C., Sijwali, P. S., Singh, A., Na, B. K. & Rosenthal, P. J. (2004) *J. Biol. Chem.* **279,** 3484–3491.
15. Sabnis, Y., Rosenthal, P. J., Desai, P. & Avery, M. A. (2002) *J. Biomol. Struct. Dyn.* **19,** 765–774.
16. Goh, L. L. & Sim, T. S. (2004) *Biochem. Biophys. Res. Commun.* **323,** 565–572.
17. Sijwali, P. S., Brinen, L. S. & Rosenthal, P. J. (2001) *Protein Expr. Purif.* **22,** 128–134.
18. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. & Turk, V. (1988) *EMBO J.* **7,** 2593–2599.
19. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990) *EMBO J.* **9,** 1939–1947.
20. Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M. & Turk, D. (2003) *J. Mol. Biol.* **326,** 875–885.
21. Kamphuis, I. G., Drenth, J. & Baker, E. N. (1985) *J. Mol. Biol.* **182,** 317–329.
22. Pandey, K. C., Wang, S. X., Sijwali, P. S., Lau, A. L., McKerrow, J. H. & Rosenthal, P. J. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 9138–9143.
23. Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27,** 157–162.
24. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254.
25. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989) *Gene* **77,** 51–59.
26. Otwinowski, Z. & Minor, W. (1997) *Processing of X-Ray Diffraction Data Collected in Oscillation Mode* (Academic, New York).
27. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
28. Nicholls, A., Sharp, K. A. & Honig, B. (1991) *Proteins* **11,** 281–296.
29. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.