

the journal of biological chemistry

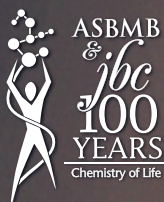
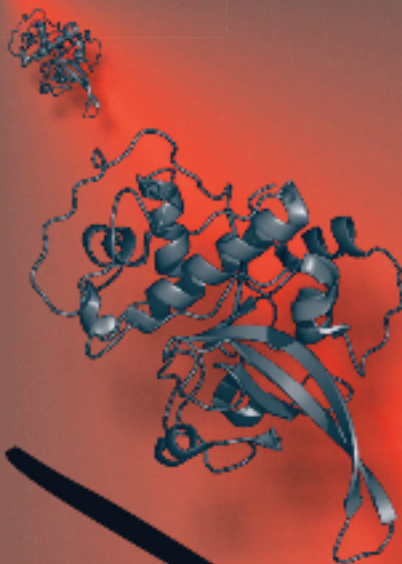
jbc

ASBMB

SEPTEMBER 1, 2006

VOLUME 281

NUMBER 35



AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY

Structural and Functional Characterization of Falcipain-2, a Hemoglobinase from the Malarial Parasite *Plasmodium falciparum**

Received for publication, April 19, 2006, and in revised form, June 13, 2006. Published, JBC Papers in Press, June 15, 2006, DOI 10.1074/jbc.M603776200

Tanis Hogg[‡], Krishna Nagarajan^{‡1}, Saskia Herzberg^{‡1}, Lili Chen[§], Xu Shen[§], Hualiang Jiang[§], Maria Wecke^{‡§}, Christoph Blohmke[‡], Rolf Hilgenfeld^{‡2}, and Christian L. Schmidt[‡]

From the [‡]Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany and [§]Drug Discovery and Design Centre, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences and Graduate School, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

Malaria is caused by protozoan erythrocytic parasites of the *Plasmodium* genus, with *Plasmodium falciparum* being the most dangerous and widespread disease-causing species. Falcipain-2 (FP-2) of *P. falciparum* is a papain-family (C1A) cysteine protease that plays an important role in the parasite life cycle by degrading erythrocyte proteins, most notably hemoglobin. Inhibition of FP-2 and its paralogues prevents parasite maturation, suggesting these proteins may be valuable targets for the design of novel antimalarial drugs, but lack of structural knowledge has impeded progress toward the rational discovery of potent, selective, and efficacious inhibitors. As a first step toward this goal, we present here the crystal structure of mature FP-2 at 3.1 Å resolution, revealing novel structural features of the FP-2 subfamily proteases including a dynamic β -hairpin hemoglobin binding motif, a flexible N-terminal α -helical extension, and a unique active-site cleft. We also demonstrate by biochemical methods that mature FP-2 can proteolytically process its own precursor in *trans* at neutral to weakly alkaline pH, that the binding of hemoglobin to FP-2 is strictly pH-dependent, and that FP-2 preferentially binds methemoglobin over hemoglobin. Because the specificity and proteolytic activity of FP-2 toward its multiple targets appears to be pH-dependent, we suggest that environmental pH may play an important role in orchestrating FP-2 function over the different life stages of the parasite. Moreover, it appears that selectivity of FP-2 for methemoglobin may represent an evolutionary adaptation to oxidative stress conditions within the host cell.

At the end of 2004, about 3.2 billion people in the world were estimated to be at risk of contracting malaria (1). The estimated worldwide number of annual clinical malaria episodes is 350–

500 million, mostly caused by the erythrocytic parasites *Plasmodium falciparum* and *Plasmodium vivax*, and the global annual death toll of the disease exceeds one million people (1). Efforts to reduce these numbers are being complicated by the rapid development of resistance by *Plasmodium* species against the currently available antimalarial drugs. Depending on the geographical region, up to 85% of clinical malaria cases cannot be cured by chloroquine, which was the standard drug only a few decades ago. Furthermore, more than 50% of the *Plasmodium* infections in sub-Saharan Africa and in some regions of Asia are insensitive to sulfamethoxine-pyrimethamine treatment, which otherwise shows a good efficacy against chloroquine-resistant *P. falciparum* strains. Of the malaria cases observed in Asia (Myanmar, Thailand, and Cambodia), 10–15% do not respond to treatment with mefloquine, which emerged as a successor to chloroquine in the 1980s. Resistance of *P. falciparum* to primaquine has also been reported (2), and widespread resistance of *P. vivax* against standard primaquine therapy is suspected (3). Even though quinine is the oldest antimalarial drug, it still shows relatively high efficacies in Asia and South America, *i.e.* 95 and 67%, respectively. However, severe side effects like hemolytic anemia, coma, and respiratory arrest unfortunately limit the wider application of this compound (4).

The only currently available antimalarial drugs without known resistance are artemisinin (from *Artemisia annua*) and its hemisynthetic derivatives (4). Recently, however, the occurrence of mutations that could lead to artemisinin resistance has been reported (5). Consequently, the WHO has called for a halt of artemisinin mono-therapies to curtail the emergence of resistance against this compound (WHO News Release 19-01-2006). Given the inexorable speed of the development of drug resistance and lack of an efficient vaccine, the search for effective, safe, and affordable drugs for malaria treatment is one of the most pressing health priorities worldwide (6).

Plasmodium infections are exclusively transmitted by anopheline mosquito bites. The sexual stage of the parasite life cycle is restricted to the insect host, whereas asexual reproduction and gametocyte development take place in the human host, where the parasite infects and multiplies within hepatocytes and erythrocytes. The primary infection of liver cells is not associated with the typical symptoms of malaria, yet the liver stages (schizonts and hypnozoites) are of particular importance

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. This paper is dedicated to the memory of Anastácia Joaquim, who died of misdiagnosed malaria on July 9, 2006.

The atomic coordinates and structure factors (code 2GHU) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed. Tel.: 49-451-500-4060; Fax: 49-451-500-4068; E-mail: hilgenfeld@biochem.uni-luebeck.de.

Structure and Function of Falcipain-2

since they are insensitive to many drugs (4) and are responsible for the well known relapses of plasmodial infections. Gametocyte reproduction and development occur within erythrocytes, and the massive, systemic rupture of infected red blood cells at the end of this phase causes the clinical symptoms of malaria. During intraerythrocytic replication, the parasites utilize cytosolic proteins of the host cell as a food source. To digest hemoglobin and other cytosolic proteins, *Plasmodia* have developed a specialized acidic (pH 5–6) organelle, termed the food vacuole, which contains a set of specific proteases. These include several aspartic proteases called plasmepsins (7, 8), which perform the initial cleavage of hemoglobin. The cysteine proteases falcipain-2 and falcipain-3 (9–12) as well as the very recently discovered falcipain-2B (13) (which appears to be identical to falcipain-2' (14)), further degrade the plasmepsin cleavage products. The short peptides produced by the falcipains are finally degraded by the metalloprotease falcilysin (15, 16).

During the late trophozoite and schizont stages, falcipain-2 is also involved in the degradation of erythrocyte-membrane skeletal proteins including ankyrin and the band 4.1 protein (17). This activity displays a pH optimum in the range of 7.0–7.5 and is thought to contribute to destabilization of the erythrocyte membrane, leading to host cell rupture and release of the mature merozoites. The autoproteolytic processing of its own precursor at neutral pH has been suggested as a third function of falcipain-2 (18, 19). Falcipain-2 is synthesized during the trophozoite stage as a membrane-bound proenzyme comprising 484 amino acid residues (18, 20). The proenzyme is transported to the food vacuole through the endoplasmic reticulum/Golgi system, and during this process the N-terminal 243 residues containing the membrane anchor are proteolytically removed. An autoproteolytic processing mechanism was suggested on the basis of inhibitor studies (19) and from the observation that the recombinant proenzyme undergoes spontaneous processing during *in vitro* refolding (18, 21). It remains unclear how falcipain-2 can perform the diverse activities of self-processing, hemoglobin degradation, and cytoskeletal degradation at different pH optima during the various stages of parasite development. Disruption of the falcipain-2 gene results in reduced hemoglobin degradation in the trophozoite stage and accumulation of undegraded hemoglobin within the parasite food vacuole but seems to be compensated for by overexpression of falcipain-2B and falcipain-3 in later stages of the *Plasmodium* life cycle (22). This suggests an overlapping function and the need for an effective inhibitor that simultaneously inhibits the falcipains. In contrast to falcipain-2, -2B, and -3, falcipain-1 does not digest hemoglobin but plays a role in host cell invasion (23) and may also be involved in oocyst production within the anopheles mosquito vector (24).

Based on their primary structures, the falcipains have been classified as members of the papain family of cysteine proteases (C1A) (12, 18, 25). Previous studies have demonstrated that inhibition of the plasmodial aspartic or cysteine proteases (26, 27) inhibits the development of the parasites *in vitro* and can cure parasitic infection in a mouse model (28). Moreover, the recent discovery of an essential ~14-residue hemoglobin binding motif near the C terminus of falcipain-2 and related plasmodial cysteine proteases (25) points toward the possibility of

designing peptidomimetic drugs that block falcipain activity by disrupting critical protein-protein interactions with its hemoglobin target. Hence, the falcipain-2 subfamily of plasmodial proteases is among the prime targets for discovery of novel anti-malarial drugs. Toward this goal, here we present the x-ray structure of falcipain-2, crystallized in the presence of the irreversible cysteine-protease inhibitor iodoacetamide, at 3.1 Å of resolution. The structure reveals unique features of the falcipain-2 subfamily proteases. The four copies of the enzyme in the crystallographic asymmetric unit provide individual snapshots of the dynamic and adaptable hemoglobin binding motif and the flexible N-terminal helical extension. A comparative structural analysis of the falcipain-2 active site with that of cruzain (cruzipain) from *Trypanosoma cruzi* suggests that existing vinyl sulfone inhibitors designed against cruzipain may be useful starting compounds for further chemical elaboration and rational design of novel antimalarials. We also demonstrate that mature, recombinant falcipain-2 can proteolytically process its own precursor *in trans* at a neutral to weakly alkaline pH and that the binding of hemoglobin to falcipain-2 is dependent on the pH of the medium and the oxidation state of the heme iron. Based on our results, we propose that methemoglobin is the preferred substrate for falcipain-2 within the acidic food vacuole and that the pH-modulated activity of falcipain-2 toward its various substrates may help to orchestrate critical events in the life cycle of the parasite.

EXPERIMENTAL PROCEDURES

Construction of Inactive Falcipain-2 Mutants—A pQE-30 plasmid (Qiagen) containing the DNA sequence coding for residues 211–484 of the falcipain-2 precursor with an N-terminal His tag (MRGSHHHHHHGSG) was kindly provided by Prof. T. Schirmeister (Würzburg). The expression plasmid for an inactive mutant form of the truncated falcipain-2 precursor containing an alanine residue instead of the active-site cysteine (pFPc285a) was constructed by PCR mutagenesis. The primers FPsen (TGGGCCTTTAGTAGTATAGGTTT) and FPantiA (GGCAGATCCACAATTTTTTGTATCCTTT) were used to amplify the entire plasmid, and the PCR product was purified by chloroform extraction and desalting with a Montage™ PCR filter device (Millipore). The plasmid was ligated using the Perfectly Blunt® Kit (Novagen) and transformed into NovaBlue competent cells (Novagen). Individual clones were selected, and pFPc285a was isolated and verified by DNA sequencing (MWG Biotech) followed by transformation into M15 (pRep4) competent cells (Qiagen) and test expression in 10-ml cultures.

A plasmid for the expression of mature, inactive falcipain-2 was constructed by amplifying the DNA sequence encoding amino acid residues 245–484 from pFPc285a using the primers PmutA (ATGAATTATGAAGAAGTTATAAAAAAATATAGA) and PmutAanti (ATTAGCTTATTCAATTAATGGAA-TGAA). The PCR product was purified, ligated into pETBlue-1 (Novagen), and transformed into Nova Blue competent cells as described above. Plasmids isolated from individual clones that contained the insert in the correct orientation as indicated by restriction mapping, termed pFPc285aM, were verified by DNA sequencing, transformed into RosettaBlue™ (DE3) cells (Novagen), and tested for expression in small cultures (10 ml).

Large Scale Expression, Purification, and Refolding of Wild-type and Mutant Falcipain-2—Bacteria containing pFpC285a or pFpC285aM were grown in 2–6-liter batches at 37 °C on YT or Luria-Bertani (LB) medium containing 100 µg/ml carbenicillin (plus 25 µg/ml kanamycin in the case of the M15 (pRep4)-derived strains). Recombinant gene expression was induced by the addition of 0.5 mM isopropyl 1-thio-β-D-galactopyranoside to the liquid cultures at an optical density of 0.6–0.7 (measured at 546 nm). The cultures were incubated for another 3 h at 37 °C and harvested by centrifugation. The cells were washed with 1/10th of the culture volume of cold 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, resuspended in 1/100th of the culture volume of the same buffer, and either stored at –20 °C or immediately broken by sonication. Recovery of recombinant wild-type falcipain-2 and the inactive mutants, which were all produced as inclusion bodies, was achieved by collection of the insoluble cell lysate fraction after centrifugation for 20 min at 20,000 × g. Inclusion bodies of the His-tagged falcipain-2 mutants were washed and dissolved, and the recombinant protein was purified by nickel nitrilotriacetic acid chromatography essentially as described by Sijwali *et al.* (11). Purified protein was refolded as described (11) with only minor modifications; protein concentration in the nickel nitrilotriacetic acid elution fractions was estimated using a theoretical extinction coefficient of 44,640 M⁻¹·cm⁻¹ at 279 nm (29), and the ratio of protein solution to refolding buffer was reduced to 1:50. The refolded protein was desalted on a HighTrapTM column (Amersham Biosciences) equilibrated with 20 mM Tris/HCl (pH 7.5), and eluted in the same buffer. The desalted protein was further purified on a Poros 20HQ 4.6/10 anion exchange column (PerSeptive Biosystems) equilibrated with 20 mM Tris/HCl, pH 7.5, and eluted with a gradient of 0–500 mM NaCl. For the purification of the wild-type protein, 10 mM iodoacetamide was included in the desalting and ion-exchange buffers.

Expression of mature, inactive falcipain-2 (FpC285aM) was the same as described above; however, since this protein could not be purified by metal affinity chromatography, the inclusion bodies were more intensely washed; twice with buffer A (2 M urea, 2.5% (w/v) Triton X-100, 20 mM Tris/HCl (pH 8.0)) and twice with buffer B (20% (w/v) sucrose, 20 mM Tris/HCl (pH 8.0)). After each centrifugation step, the inclusion bodies were resuspended by sonication. Inclusion bodies from a 6-liter culture were resuspended in 5 ml of buffer B plus 10 mM MgCl₂ and 125 units of benzonase (Merck). The suspension was stirred overnight at 4 °C followed by the addition of 25 ml of buffer B, centrifugation, and solubilization in 17 ml of denaturing buffer (8 M urea, 1 M imidazole, 20 mM Tris/HCl (pH 8.0)). Protein concentration was measured using a theoretical extinction coefficient of 40,605 M⁻¹·cm⁻¹ at 279 nm (29). This solution was either stored at –20 °C or directly subjected to refolding and ion-exchange purification.

General Activity Assay—The chromogenic substrate Z-Phe-Arg-pNA³ (Bachem AG) was used for routine activity assays.

The assay mixture contained protein sample in 50 µM Z-Phe-Arg-pNA, 10 mM dithiothreitol, and 100 mM sodium acetate (pH 5.5) in a final volume of 700 µl. The absorbance at 405 nm was measured against a blank consisting of protein-free assay mixture. The activity was calculated using an extinction coefficient of 10,500 M⁻¹·cm⁻¹. The pH dependence of the reaction was tested using BisTris-propane/HCl buffer in place of sodium acetate.

Hemoglobinase Assay—Hemoglobin-degrading activity was assayed by incubating human hemoglobin (Sigma, H7379) with varying concentrations of recombinant falcipain-2. Falcipain-2 was added to a mixture consisting of 0.2 mg/ml hemoglobin, 100 mM sodium acetate (pH 5.5), and 10 mM dithiothreitol. The assay mixture was incubated for 10–120 min at 37 °C, and 20 µl aliquots were extracted at fixed time points, mixed immediately with 10 µl of denaturing SDS-PAGE loading buffer, heated for 5 min at 95 °C, and analyzed by SDS-PAGE. To test the pH dependence of hemoglobin degradation, the sodium acetate buffer was replaced by 100 mM BisTris-propane/HCl at the indicated pH.

Self-processing Assay—The same procedure as described for the hemoglobinase assay was used to study the processing of inactive, truncated falcipain-2 precursor (FpC285a) by active, refolded falcipain-2. The assay mixture contained 0.2 mg/ml refolded FpC285a instead of hemoglobin.

Binding Studies—The binding of hemoglobin to mature, inactive falcipain-2 was studied by surface plasmon resonance using a Biacore[®] 3000 system at room temperature. Inactive falcipain-2 was immobilized to the carboxymethylated dextran matrix of a CM5 research-grade sensor chip (Biacore) by the standard primary-amine coupling reaction in HBS-EP buffer (10 mM HEPES/NaOH, pH 7.5, 150 mM NaCl, 3 mM EDTA). The flow cell was activated for 7 min with a 1:1 mixture of 0.2 M *N*-hydroxysuccinimide and 50 mM *N*-ethyl-*N'*-(3-diethylaminopropyl)-carbodiimide at a flow rate of 5 µl/min at 25 °C. Inactive, mature falcipain-2 (FpC285aM) at a concentration of 0.066 mg/ml in 10 mM sodium acetate (pH 3.9) was injected for 7 min at 5 µl/min, yielding a final immobilization level of 600 response units. The surface of the flow cell was subsequently blocked with a 7-min injection of 1 M ethanolamine (pH 8.5). Finally, 50 mM NaOH was injected for 10 s to remove any non-covalently bound protein. The control flow cell was activated and immediately blocked without immobilizing any protein to prevent nonspecific binding of the analyte to the sensor surface of the reference cell. Equilibration of the base line was carried out by a continuous flow of HBS-EP buffer (pH 6.0) through the chip for at least 3 h. Human hemoglobin (methemoglobin (Sigma H7379) or ferrous stabilized hemoglobin (Sigma H0267)) in 20 mM Na₂HPO₄/NaH₂PO₄, 150 mM NaCl, 3 mM EDTA buffer of the indicated pH was injected at a flow rate of 30 µl/min for 120 s at 25 °C. After each hemoglobin injection, the chip was regenerated by two injections of 50 mM NaOH at a flow rate of 50 µl/min for 6 s. Individual rate constants were determined using the Biacore software. Alternatively, the equilibrium dissociation constants were determined by plotting the intensity of the steady-state response (response units) against the hemoglobin concentration. The data were fitted to the

³ The abbreviations used are: Z-Phe-Arg-pNA, benzoyl-Phe-Arg-p-nitroanilide; NCS, non-crystallographic symmetry; r.m.s.d., root mean square deviation; Bis-Tris-propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; AMC, 7-amino-4-methylcoumarin.

Structure and Function of Falcipain-2

standard formula of the absorption isotherm using the program KaleidaGraph 3.6 (Synergy Software).

Crystallization and Diffraction Data Collection—Wild-type, mature falcipain-2 was crystallized in the presence of the irreversible inhibitor iodoacetamide. Crystals were obtained at room temperature using the hanging-drop vapor diffusion method by equilibrating a mixture containing 5 μ l of protein solution (2 mg/ml falcipain-2, 20 mM Tris-HCl (pH 7.5), 10 mM iodoacetamide, 150–200 mM NaCl) and an equivolume of precipitant solution (0.4–0.8 M $(\text{NH}_4)_2\text{SO}_4$, 100 mM sodium citrate, pH 4.5–5.0) over reservoirs containing 1 ml of precipitant. Crystals appeared after 3–4 days and reached maximum dimensions of $\sim 250 \times 50 \times 50 \mu\text{m}$ within 10 days. Crystals generally diffracted anisotropically to a minimum Bragg spacing (d_{min}) of ~ 10.0 – 6.0 \AA , possibly due to their very high solvent content (see below). Only a few of the >50 crystals screened diffracted to a resolution of better than 3.5 \AA , the best of which we report in this paper. Crystals were briefly soaked in a cryoprotectant consisting of precipitant solution with 25% (v/v) ethylene glycol and cooled to 100 K in a nitrogen-gas stream. The C-centered orthorhombic cell had dimensions $a = 145.83 \text{ \AA}$, $b = 168.28 \text{ \AA}$, and $c = 178.09 \text{ \AA}$. X-ray diffraction data were collected at 100 K and with an incident wavelength of 0.8045 \AA at the Joint University of Hamburg-IMB Jena-EMBL Beamline X13, DESY (Hamburg). Crystal indexing and integration of reflection intensities were performed with Mosflm 6.2.4 (30, 31), with subsequent data scaling and merging carried out with Scala 3.2.5 of the CCP4 software suite (32, 33).

Structure Determination and Refinement—Structure determination by molecular replacement phasing was implemented with Phaser 1.3 (34) using a “mixed-model” (35) derived from residues 4–218 of the crystal structure of KDEL-tailed cysteine endopeptidase (36), which displays 43% sequence identity with falcipain-2. The correct solution, consisting of four falcipain-2 monomers in the crystallographic asymmetric unit ($\sim 76\%$ solvent content) was found in space group $C222_1$ and yielded a log-likelihood gain (the difference between the log-likelihoods of the molecular replacement model and a random distribution of atoms) of 927 in the 40.0 – 3.5 \AA resolution range. An initial σ_A -weighted electron density map (37) calculated from the molecular replacement phases and measured amplitudes was subjected to prime-and-switch phasing with 4-fold non-crystallographic symmetry (NCS) averaging, as implemented in Resolve 2.05 (38), to reduce bias arising from the model-based phases. The NCS-averaged prime-and-switch electron density map was used to construct an initial model of falcipain-2 with the molecular graphics program Coot (39). Early rounds of crystallographic refinement were performed in CNS 1.1 (40) using anisotropic bulk solvent correction, torsion-angle simulated annealing, positional, and grouped B-factor protocols with strict 4-fold NCS constraints. The model was subjected to iterative rounds of building into σ_A -weighted $3F_o - 2F_c$ and $F_o - F_c$ maps, and crystallographic refinement with relaxed 4-fold NCS restraints was initiated at later stages once structural differences between the individual monomers were detected. Composite simulated-annealing omit maps were used regularly during the building process to verify and correct the model. The final falcipain-2 model, which exhibits good stereochemistry

(Table 1), consists of 962 amino acid residues and 40 water molecules. Final R and R_{free} values are 23.1 and 27.4% for data in the 40.0 – 3.1 \AA range (with 5% of the observed data reserved for calculation of R_{free}). Molecular figures were generated using the program PyMOL (41).

RESULTS

Refolding and Purification of the Falcipain-2 Mutant Forms—The FPc285a (inactive mutant form of the truncated falcipain-2 precursor) and FPc285aM (inactive, mature falcipain-2) proteins were successfully refolded and purified (Fig. 1a). In the case of FPc285a, the recovery of soluble protein was significantly lower than for the active form. Moreover, handling of FPc285a was complicated by its relative instability as it precipitated gradually upon storage at 4°C and completely upon freezing. In contrast, recovery of FPc285aM was slightly better than for the wild-type protein, and it was stable for up to 2 weeks at 4°C . None of the mutant forms displayed any protease activity when assayed with Z-Phe-Arg-pNA as a substrate.

pH Dependence of Substrate Binding and Protease Activity—The pH-dependent proteolytic activity of recombinant, active falcipain-2 was measured with three different substrates: (i) human methemoglobin (Fig. 1c), (ii) the recombinant, inactive truncated falcipain-2 precursor (FPc285a) (Fig. 1d), and (iii) the chromogenic substrate Z-Phe-Arg-pNA (Fig. 1b). The hemoglobinase activity (Fig. 1c) displayed a clear pH optimum in the acidic range as has been previously described (18). Hemoglobin was completely digested within 30 min at pH 6.0, whereas only minimal or no proteolysis was observed at pH 7.0 and 7.5, respectively. The autoproteolysis assay shown in Fig. 1d demonstrates that mature, refolded wild-type falcipain-2 can process the inactive, truncated proenzyme *in trans*. In contrast to the hemoglobinase activity, *trans* processing exhibited detectable activity over a wide pH range and showed a maximum at pH 8.5. The peptidic substrate Z-Phe-Arg-pNA was cleaved with relatively high efficiency over the entire pH range studied (pH 6–8.5); however, the proteolytic activity decreased by $\sim 30\%$ above pH 7.5 (Fig. 1b).

To further investigate the basis for the different pH dependencies of the various catalytic activities of recombinant falcipain-2, we studied the binding of hemoglobin to mature, inactive falcipain-2 (FPc285aM) by surface plasmon resonance using a Biacore[®] 3000 system. These experiments showed that the binding of hemoglobin to falcipain-2 decreased with increasing pH (Fig. 2). The individual data collected at pH 7.0 and 8.0 could be fitted to a standard adsorption isotherm for the formation of a bimolecular complex assuming a single binding site (R values of 0.9967–0.9994). The data collected at pH 6.0 could be fitted significantly better assuming a second, low affinity binding site (R values 0.9995 and 0.9999). The dissociation constants of this second binding site were higher by about 35-fold (121 versus $3.3 \mu\text{M}$ for hemoglobin) to 90-fold (74 versus $0.8 \mu\text{M}$ for methemoglobin) than the constants for the high affinity binding site. The individual dissociation constants showed some variability depending on the protein preparation. The K_D at pH 6.0 for methemoglobin varied in a range from 0.8 to $2.3 \mu\text{M}$; however, the constants determined at higher pH values varied accordingly (e.g. the binding of methemoglobin at

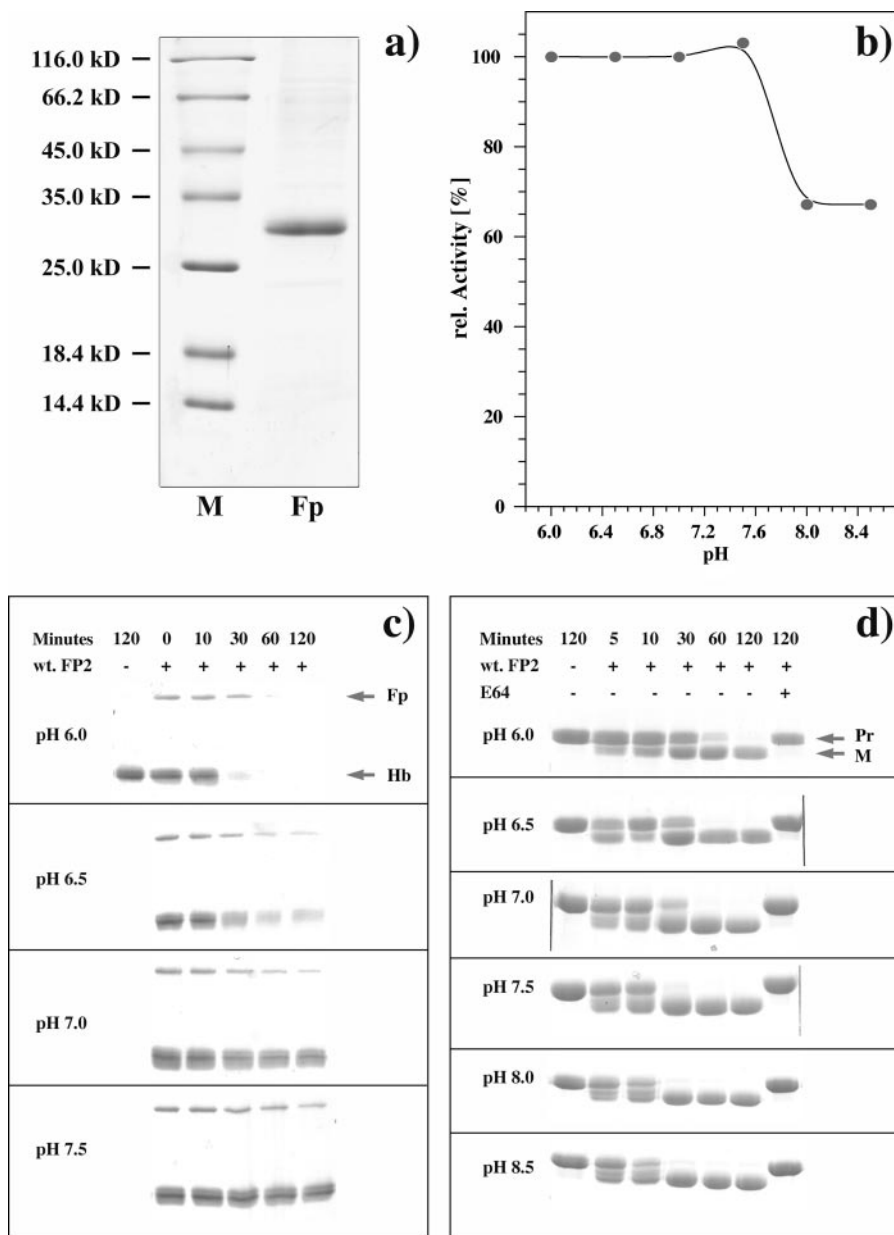


FIGURE 1. Catalytic activities of falcipain-2. *a*, SDS-polyacrylamide gel electrophoresis of the refolded and purified, inactive, mature falcipain-2 (FPc285aM). *M*, molecular weight marker proteins; *Fp*, 0.65 μ g of the purified falcipain-2 mutant. *b*, pH dependence of the hydrolysis of the chromogenic substrate Z-Phe-Arg-pNA. 100% activity corresponds to the hydrolysis of 101.6 μ mol of substrate/min/mg of falcipain-2. *c*, hemoglobinase activity. Each assay mixture contained 0.2 mg ml⁻¹ of human hemoglobin and 0.02 mg ml⁻¹ of active, refolded falcipain-2. The samples were incubated at 37 °C for the indicated times. *Fp*, falcipain-2; *Hb*, hemoglobin. *wt*, wild type. *d*, trans self-processing activity. The experimental setup was the same as for the hemoglobinase activity; however, the assay mixtures contained 0.2 mg ml⁻¹ refolded, truncated, and inactive falcipain-2 precursor (FPc285a) instead of hemoglobin. In a control experiment, FPc285a was incubated in the absence of the active falcipain-2 to verify the absence of protease activity in the FPc285a preparation. The cysteine protease inhibitor E64 was added to a final concentration of 10 μ M to one of the assay mixtures to confirm that the observed processing was performed by falcipain-2. *Pr*, truncated inactive falcipain-2 precursor; *M*, mature falcipain-2.

pH 7.0 was hardly measurable in the preparation that showed a K_D at pH 6.0 of 2.3 μ M). Thus, the observed reduction in hemoglobinase activity of falcipain-2 at higher pH values is due to decreased substrate binding under these conditions. Interestingly, very significant differences were observed between the binding of falcipain-2 to methemoglobin and hemoglobin, with methemoglobin exhibiting much smaller K_D values (Fig. 2).

β 4- β 5- β 6) that is extended at the central part of β 4 by the very short β 1 and is decorated by peripheral helices α 1, α 5, and α 6. Additional structural features that are unique to the falcipain-2 subfamily (with members such as the vivapain, vinckepain, and berghepain homologues (Fig. 4)) include a 17-residue N-terminal extension and a 14-residue β -hairpin protuberance extending from β 4 and β 5 that comprises the hemoglobin binding motif

Overall Structure—Mature falcipain-2 was crystallized at pH 4.5–5.0 in the presence of 0.4–0.8 M (NH₄)₂SO₄, 100 mM sodium citrate, and 10 mM irreversible inhibitor iodoacetamide. Its three-dimensional structure was determined by molecular replacement techniques using a modified set of coordinates of the KDEL-tailed cysteine endopeptidase, KDEL-CysEP, from *Ricinus communis* (castor bean) (36). The crystallographic asymmetric unit consists of four almost identical falcipain-2 monomers, with each exhibiting good stereochemistry, and 40 water molecules (Table 1). The residue numbering scheme chosen for the crystal structure reflects the sequence of mature falcipain-2 rather than its full-length precursor, *i.e.* residue 1 has been assigned to the Gln of the new N terminus liberated after self-processing (KYLLD ↓ QMNYF). Despite a limiting resolution of 3.1 Å, the 3F_o – 2F_c electron density maps were of good quality and allowed unambiguous tracing of the polypeptide chain for 958 of 964 amino acid residues. The Pro-189–Leu-190 segment of monomers A and C was modeled with zero occupancy, and the Glu-14–Glu-15 segment in monomer C was not modeled due to poor density. The overall fold of falcipain-2 is illustrated in Fig. 3, showing a core structure with a distorted ellipsoidal shape that is characteristic of the C1A proteases. The polypeptide chain of mature falcipain-2 folds into the prototypical structure of papain-like proteases, with two distinct domains separated by a long central substrate binding cleft containing the active site (42, 43) (Fig. 3). The left domain (L domain) is predominantly α -helical (α 2, α 3, α 4) with several long disulfide-stabilized segments lacking regular secondary structure, whereas the right domain (R domain) contains a large antiparallel β -sheet (β 2- β 7- β 3-

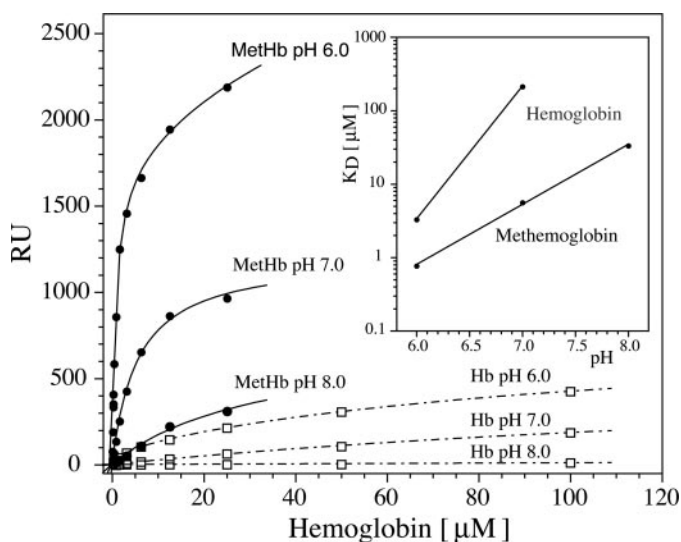


FIGURE 2. pH-dependent binding of human hemoglobin and methemoglobin to inactive falcipain-2 as measured by surface plasmon resonance. The intensity of the steady state response (response units) was plotted against hemoglobin concentration. The data were fitted to the standard formula of the absorption isotherm using the program KaleidaGraph 3.6 (Synergy Software). To fit the data collected at pH 6.0, the equation was modified to make allowance for two independent binding equilibria. The inset shows the pH dependence of the dissociation constants for the hemoglobin (Hb)-falcipain-2 and methemoglobin (MetHb)-falcipain-2 complexes. Only the constants for the high affinity binding site are plotted for pH 6.0. RU, response unit.

TABLE 1

Data collection and refinement statistics

Values in parentheses are for the highest resolution shell.

Data collection	
Wavelength (Å)	0.8045
Resolution (Å)	40.0-3.04 (3.2-3.04)
Space group	C222 ₁
Unit-cell parameters	
<i>a</i> (Å)	145.83
<i>b</i> (Å)	168.28
<i>c</i> (Å)	178.09
Overall reflections	262,326 (23,804)
Unique reflections	40,840 (4,411)
Multiplicity	6.4 (5.4)
Completeness (%)	95.8 (71.9)
R_{merge}^a (%)	15.2 (76.4)
$I/\sigma(I)$	11.5 (2.7)
Refinement	
Resolution (Å)	40.0-3.1 (3.29-3.1)
Number of reflections	39,960 (6,561)
Overall completeness	99.9 (100.0)
R_{cryst}^b	0.231 (0.397)
R_{free}^c	0.274 (0.405)
r.m.s.d. from ideal geometry	
Bonds (Å)	0.009
Angles (Å)	1.5
Protein atoms	7,614
Solvent atoms	40
Ramachandran plot	
Most favored ^c (%)	75.3
Additionally allowed ^c (%)	22.8
Generously allowed ^c (%)	1.8
Disallowed regions ^c (%)	0

^a $R_{\text{merge}} = \sum_{hkl} \sum_i |I(hkl)_i - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I(hkl)_i$, where $I(hkl)$ is the intensity of reflection hkl , and $\langle I(hkl) \rangle$ is the average intensity over all equivalent reflections.

^b $R_{\text{cryst}} = \sum_{hkl} |F_o(hkl) - F_c(hkl)| / \sum_{hkl} F_o(hkl)$. R_{free} was calculated for a test set of reflections (5%) omitted from the refinement.

^c As defined by Morris *et al.* (71).

(Figs. 3 and 4). Structural differences were clearly evident between the four monomers in both of these regions, specifically in $\alpha 1/\beta 1$ loop residues 13–16 and in residues 186–195 ($\beta 5$ -loop- $\beta 6$) of the

hemoglobin binding insert. These segments were, therefore, exclusively omitted from the 4-fold NCS restraints implemented during crystallographic refinement. Unless otherwise indicated, the following discussion will be based on the slightly better-defined monomer D of the crystallographic asymmetric unit.

To probe the entire Protein Data Bank for the closest known structural homolog of falcipain-2, a comprehensive VAST (Vector Alignment Search Tool) search (44) was carried out. Interestingly, this procedure identified cruzipain (also known as cruzain), the major cysteine protease of the blood parasite *T. cruzi* (the causative agent of Chagas disease) (45–47), as the most similar structural relative to falcipain-2, with a backbone root mean square deviation (r.m.s.d.) of 1.7 Å over 207 homologous residues. Using the same search algorithm, monoclinic papain (48) was found to exhibit a backbone r.m.s.d. of 1.8 Å over 204 residues. The overall sequence identity between falcipain-2 and cruzipain is ~39% compared with ~38% for papain.

The N-terminal Extension—The presence of a 17-residue N-terminal extension is one of the major distinctive structural features of the falcipain-2 subfamily when compared with other C1A proteases. Previous functional studies have shown that the N-terminal extension plays a crucial role in folding of the mature protein into its active conformation (20). The crystal structure of falcipain-2 reveals that the N-terminal extension comprises a short α -helix ($\alpha 1$, residues 4–11) followed by a 10-residue loop (Fig. 3). The $\alpha 1$ helix is stabilized by the highly conserved Tyr-4, Ile-8, and Tyr-11, which pack tightly against residues of the R domain elements $\beta 2$ (Leu-129), $\alpha 5$ (Glu-138, Leu-143), and the intervening $\beta 2/\alpha 5$ loop (Pro-132). The long loop (residues 12–21) of the N-terminal extension traverses the lower surface of the R domain before entering the domain core nearly 20 Å away at the base of the hemoglobin binding β -hairpin. The loop exhibits elevated conformational flexibility as evidenced by relatively high B factors and diffuse electron density in some of the monomers.

An Additional Cys-99—Cys-119 Disulfide Bond in Falcipain-2 and Related Homologues—Apart from the catalytic Cys-42 (corresponding to Cys-285 of the full-length precursor), there are a total of eight cysteine residues in falcipain-2 forming four disulfide bonds (Cys-39—Cys-80, Cys-73—Cys-114, Cys-99—Cys-119, and Cys-168—Cys-229). The disulfides Cys-39—Cys-80 and Cys-73—Cys-114 are well conserved among the papain-like enzyme structures. Like papain and the majority of C1A proteases, falcipain-2 also has the disulfide Cys-168—Cys-229 to fix the upper loops defining the S2 and S1' substrate binding sites, in contrast to some of the cathepsins which lack this stabilizing element. An additional Cys-99—Cys-119 disulfide bond in the falcipain family was predicted on the basis of homology modeling studies (49), and mutations of either Cys-99 or Cys-119 have been shown to abolish enzyme activity, probably by disrupting protein folding and leading to a catalytically inactive conformation (50). Here we confirm the presence of a Cys-99—Cys-119 disulfide, which appears to play an important role together with the Cys-73—Cys-114 disulfide in stabilizing the long $\alpha 4/\beta 2$ loop (~30 residues) that blankets the lateral surface of the L domain.

The Hemoglobin Binding Insert—A 14-residue insertion located in the C-terminal half of falcipain-2 has been recently

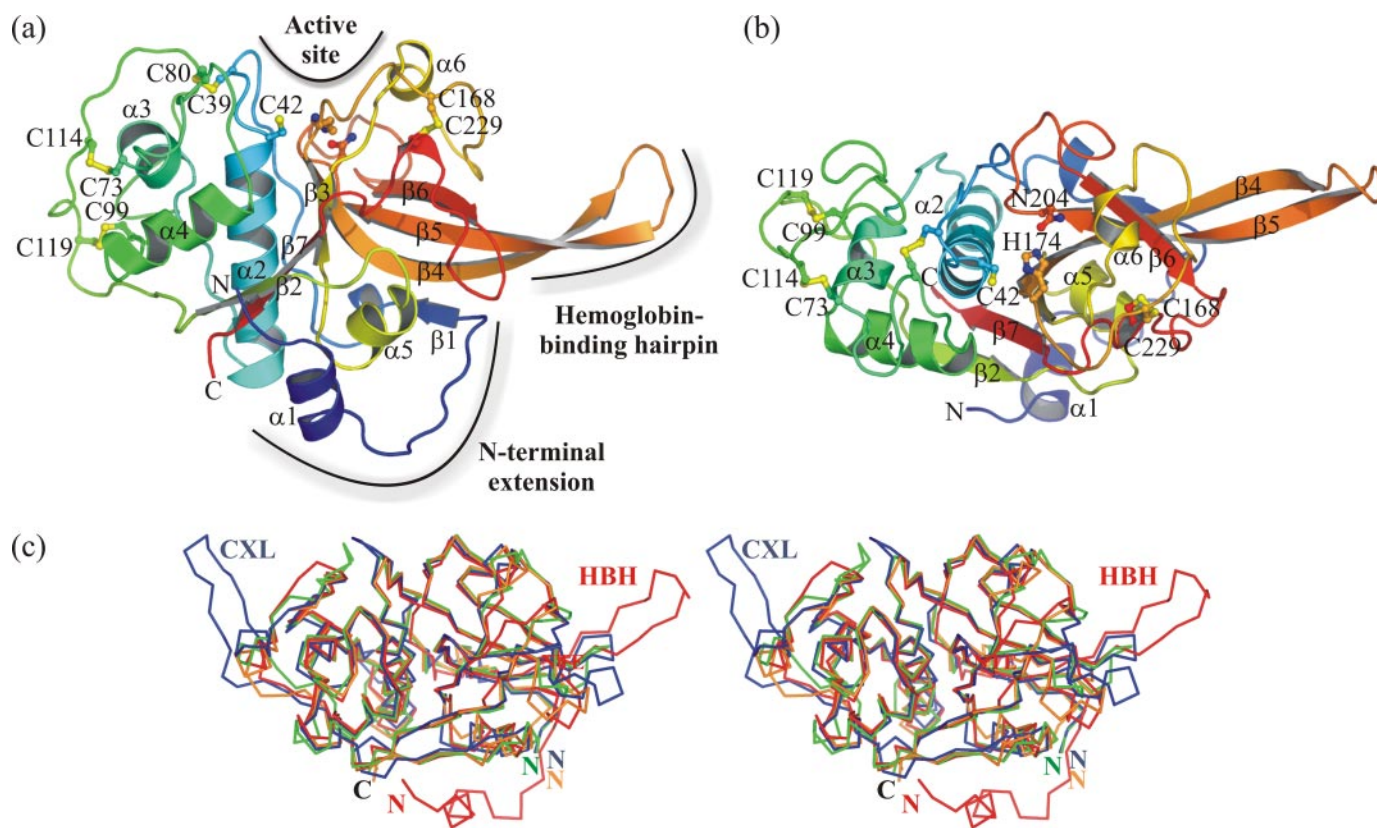


FIGURE 3. **Overall view of falcipain-2.** *a*, ribbon representation of falcipain-2 with the L domain on the left, the R domain on the right, and the active site cleft located on top of the molecule. The spectral color-coding is according to sequence, from dark blue (N terminus) to dark red (C terminus). Disulfide bonds and active site residues are shown in ball-and-stick models. Secondary structure elements, N and C termini, the active site cysteine (C42), and cysteine residues involved in disulfide bonds are labeled. *b*, an orthogonal view relative to (*a*) generated by a $\sim 90^\circ$ rotation toward the viewer. Active site residues Cys-42, His-174, and Asn-204 are labeled. *c*, $C\alpha$ superposition comparing falcipain-2 (red), human cathepsin X (blue; PDB-ID: 1EF7) (57), papain (orange; PDB-ID: 9PAP) (69), and cruzipain (green; PDB-ID: 1ME3) (47). N and C termini, the hemoglobin binding hairpin (HBH) of falcipain-2 and the 110–123 loop (CXL) of cathepsin X are labeled.

shown to fulfill an important function by capturing hemoglobin (25). This insert adopts an extended antiparallel β -hairpin conformation that causes the β_4/β_5 strands to protrude out from the surface of the R domain by ~ 20 Å (Figs. 3 and 5). The elevated conformational flexibility of the β -hairpin manifests itself in progressively diminishing electron density in a distal direction along the long axis of this segment in monomers A and C. This is not surprising since the β -hairpin of monomers A and C extends into bulk solvent and does not make contact with any adjacent monomers in the crystal (this crystal form has a solvent content of $\sim 76\%$). As a result, the distally positioned residues of the β -hairpin were difficult to model and refine in monomers A and C; hence, some of these residues exhibit rather strained main-chain torsion angles, and Pro-189 and Leu-190 at the tip of the hairpin in these monomers were modeled with zero occupancy. In monomers B and D, however, the β -hairpin is stabilized by intermolecular contacts, allowing for confident modeling and refinement of the entire insert including residues at the tip of the hairpin (Fig. 5).

The Active-site Cleft—The active site of falcipain-2 is located at the bottom of an elongated surface depression that has a slightly altered surface topology when compared with cruzipain and papain. The active-site Cys-42 exhibits strong residual $F_o - F_c$ electron density closely flanking the side-chain sulfur

atom, indicating the presence of an attached acetamide moiety, although the density was not strong enough to facilitate modeling of the inhibitor (Fig. 6*a*). Most striking was the presence of a large ellipsoidal $F_o - F_c$ electron density peak residing within the S2 subsite (which is formed by the side chains of Trp-43 (α_2), Leu-84 (α_3/α_4 loop), Ile-85 (N terminus of α_4), Ser-149 (C terminus of β_3), and Ala-175 (N terminus of β_5) as well as the main-chain carbonyl group of Gly-83 (α_3/α_4 loop)) (Fig. 6*a*). When we superimposed the structures of various C1A protease-inhibitor complexes with falcipain-2, we found the difference density to overlap perfectly with the side-chain atoms of residues in the inhibitor P2 position (Fig. 6*b*), suggesting the density may have arisen from S2-subsite binding of Tris (a component of the crystallization buffer) or ethylene glycol (the cryoprotectant) or possibly iodide (a product of iodoacetamide treatment). Computation of an anomalous-difference Fourier did not support the presence of an anomalous scatterer such as iodine, however, and the residual $F_o - F_c$ density lacks characteristic features that would allow an unambiguous assignment to any particular buffer molecule.

Comparison with the Falcipain-2-Cystatin Complex—One day after our experimental data were submitted to the Protein Data Bank (PDB), the 2.7-Å crystal structure of falcipain-2 in complex with egg white cystatin (a macromolecular cysteine

Structure and Function of Falcipain-2

4 α1 11 22 β1 310 27 42 α2

```

Falcipain-2  --QMNYEEVIIKKYR--GEENFDHAA----YDWRLHSSGVTPVKDQKNCGSCWAFSSIGSV 51
Falcipain-2B --QINYDAVIIKKYK--GNENFDHAA----YDWRLHSSGVTPVKDQKNCGSCWAFSSIGSV 51
Falcipain-3  --YEANYEDVIIKKYKP--ADAKLDRIA----YDWRLHGGVTPVKDQALCGSCWAFSSVGSV 53
Vivapain-2  --RITNYEDVIDKYKP--KDATFDHAS----YDWRLHKGVTPVKDQANCGSCWAFSTVGVV 53
Vivapain-3  --RVSDYDDIIHKYKP--KDGTFDYVK----HDWREFNAVTPVKDQKNCGACWAFSTVGVV 53
Vivapain-4  --YVSNYDDVLKKYKP--ADAVVDNEK----YDWREHNAVSEIKNQNLCGSCWAFGAVGAV 53
Falcipain-1  NHLKDNILISEFYTNGKRNEKDIFSKVPEILDYREKGIVHEPKDQGLCGSCWAFASVGNI 60
Vincapain-2  --YLVPYSAALGKYKS--PTDKVNYRS----FDWRDKDVIIDVKDQKCASCWAFSVAGVV 53
Chabapain-2  --HLIPYSAAISKYKS--PTDKVNYKS----FDWREHNAIIAVKDQKRCASCWAFATAGVI 53
Berghepain-2 --HLIPYTTAISKYKS--PTDKVNYTS----FDWRDYNVIIIGVKDQKCASCWAFATAGVI 53
KDEL-CysEP  ---TVPAS-----VDWRKKGAVTSVKDQGCGSCWAFSTIVAV 35
    
```

α2 58 67 α3 73 85 α4 95

```

Falcipain-2  ESQYAIRKNKLITLSEQELVDCSFK--NYGCNGGLINNAFEDMIELGGICPDGDYPVYVSDA 110
Falcipain-2B ESQYAIRKNKLITLSEQELVDCSFK--NYGCNGGLINNAFEDMIELGGICTDDDYPVYVSDA 110
Falcipain-3  ESQYAIRKKALFLFSEQELVDCSVK--NNGCYGGYITNAFDDMIDLGLCSQDDYPVYVSNL 112
Vivapain-2  ESQYAIRKNQLVSISEQQMVDCSTQ--NTGCYGGFIPLAFEDMIEMGLCSSEDYPVYVVADI 112
Vivapain-3  ESQYAIRKKELVSLSEQEMVDCSFK--NYGCDGGGNIPIAFEDLLDLGGICKEKEYPVYVDVT 112
Vivapain-4  ESQYAIRKNQHVLISEQELVDCSDK--NFGCFGGLASLAFDDMIDLGYLCSESDYPYVGFK 112
Falcipain-1  ESVFAKKNKNILSFSEQEVDCSKD--NFGCDGGHPFFYSFLYVLQN--ELCLGDEYKYKAKD 118
Vincapain-2  SAQYAIRQNKKISLSEQQLVDCAPN--NFGCEGGIIPYALEDLIDMGGLCEDKYPYVVANI 112
Chabapain-2  EAQYAIRQNKKISLSEQQLVDCSQS--NDGCEGGILPYAFEDLIDMGGLCEDKYPYVVADV 112
Berghepain-2 AAQYAIRKNQKVSLSEQQLVDCAQN--NFGCEGGILPYAFEDLIDMDGLCEDKYPYVSNV 112
KDEL-CysEP  EGINQIKTNKLVSLSEQELVDCDTDQNQGCNGGLMDYAFEFIKQRGGITTEANYPYEAYD 95
    
```

127 β2 130 136 α5 142 146 β3 149 154 α6 158 β4 164

```

Falcipain-2  PNLCNIDR--CTEKYGIKNYLSVPDNKLKEALRFLG--FISISVAVS--DDFAFYKEGIFDG 166
Falcipain-2B PNLCNIDR--CTEKYGIKNYLSVPDNKLKEALRFLG--FISISIAVS--DDFPFFYKEGIFDG 166
Falcipain-3  PETCNLKR--CNERYTIKSYVSIPDDKFKEALRYLG--FISISIAAS--DDFAFYRGGFYDG 168
Vivapain-2  PEMCKFDI--CEQYKINNFLEIPEDKFKEAIRFLG--ELSVSIAVS--DDFAFYRGGIFDG 168
Vivapain-3  PELCDIDR--CKNKYKITTYVEIPQLRFKEAIKFLG--FISVSICAN--DDFVYYEGGLFDG 168
Vivapain-4  PRKCEIKK--CKEKYTIKSYVKIPEEKYKEAIQFLG--ELTLGLTVN--DDFYDYKEGIFSS 168
Falcipain-1  DMFCLNYR--CKRKVSLSSIGAVKNQLILALNEVG--ELSVNVGVN--NDFVAYSEGVYNG 174
Vincapain-2  PELCEINK--CKEKYSIVEALVPYDNYKEAIQYLG--ELTIAVGVS--EDFYESGIFDG 168
Chabapain-2  PELCEINK--CKEKYTAIEALVPYDNYKEAIQYLG--ELTIAVGVS--EDFQDYDGGIFDG 168
Berghepain-2  PELCEINK--CTEKYSISKFALVPFNNYKEAIQYLG--ELTIAVGVD--DDFESYNGGIFDG 168
KDEL-CysEP  G--TCDVSKENAPAVSIDGHENVPENDENALLKAVANQEVSVAIDAGGSDFQFYSEGVFTG 154
    
```

175 β5 188 193 β6 203

```

Falcipain-2  ECGDQLNHAVMLVGFG--MKEI---VN----PLTKKGEKH-----YYYIIKNSWQQWG 211
Falcipain-2B ECGDELNHAVMLVGFG--MKEI---VN----PLTKKGEKH-----YYYIIKNSWQQWG 211
Falcipain-3  ECGAAPNHAVILVGYG--MKDI---YN----EDTGRMEKF-----YYYIIKNSWGSDWG 213
Vivapain-2  ECGEAPNHAVILVGFG--AEDA---YD----FDTKTMKKR-----YYYIVKNSWGVSWG 213
Vivapain-3  SCGFSPNHAVILVGYG--MEEM---YD----AMSRKNEKR-----YYFWLKNSWGEKWG 213
Vivapain-4  ECTEEPNHEVMIVGYG--VEEM---FN----SESNASEKH-----YYYIIKNSWGENWG 213
Falcipain-1  TCSEELNHSVLLVGYGQVEKTKLNYNNKIQTYNTKENSNQPDDNIIYWIKNSWSKKWG 234
Vincapain-2  ECEGVANHAVILVGYG--VESV---FD----EVLKKNVDQ-----YYYIIRNSWGSAWG 213
Chabapain-2  ECTGFANHAVILVGYG--VESV---FD----ESLKKNVDQ-----YYYIIRNSWSDAWG 213
Berghepain-2  ECTDFANHAVMLIGYG--VEEV---YD----KRLKKNVKE-----YYYIIRNSWGDAWG 213
KDEL-CysEP  SCGTELDHGVAIVGYG-----TTIDGTK-----YWTVVKNSWGPEWG 190
    
```

215 β7 220 228 310 230 235 β8 239

```

Falcipain-2  ERGFINIETDESGLMRKCGLGTDAFIPLIE----- 241
Falcipain-2B ERGFINIETDESGLMRKCGLGTDAFIPLIE----- 241
Falcipain-3  EGGYINLETDENGYKKTCSIGTEAYVPLLE----- 243
Vivapain-2  EKGFIRLETDINGYRKPCSLGTEALVALVD----- 243
Vivapain-3  EKGYMKIQTDEYGLMKTCSLGAQAFVALIDEV----- 245
Vivapain-4  EKGFMRIETDELGLQKTNNM--TEAYVPLLD----- 242
Falcipain-1  ENGFMRLSRNKNGDNVFCGIGEEVFYPIL----- 263
Vincapain-2  EDGYIRLKTNESGTLRNCVL--LQAFAPVIE----- 242
Chabapain-2  EEGYMRLKTDESGALRNCVL--VQAYVPIE----- 242
Berghepain-2  ERGYIRLKTNESGTLRNCVL--VQGYAPIE----- 242
KDEL-CysEP  EKGYIRMERGISDKEGLCGIAMEASYPIKKSSNNPSGIK 229
    
```

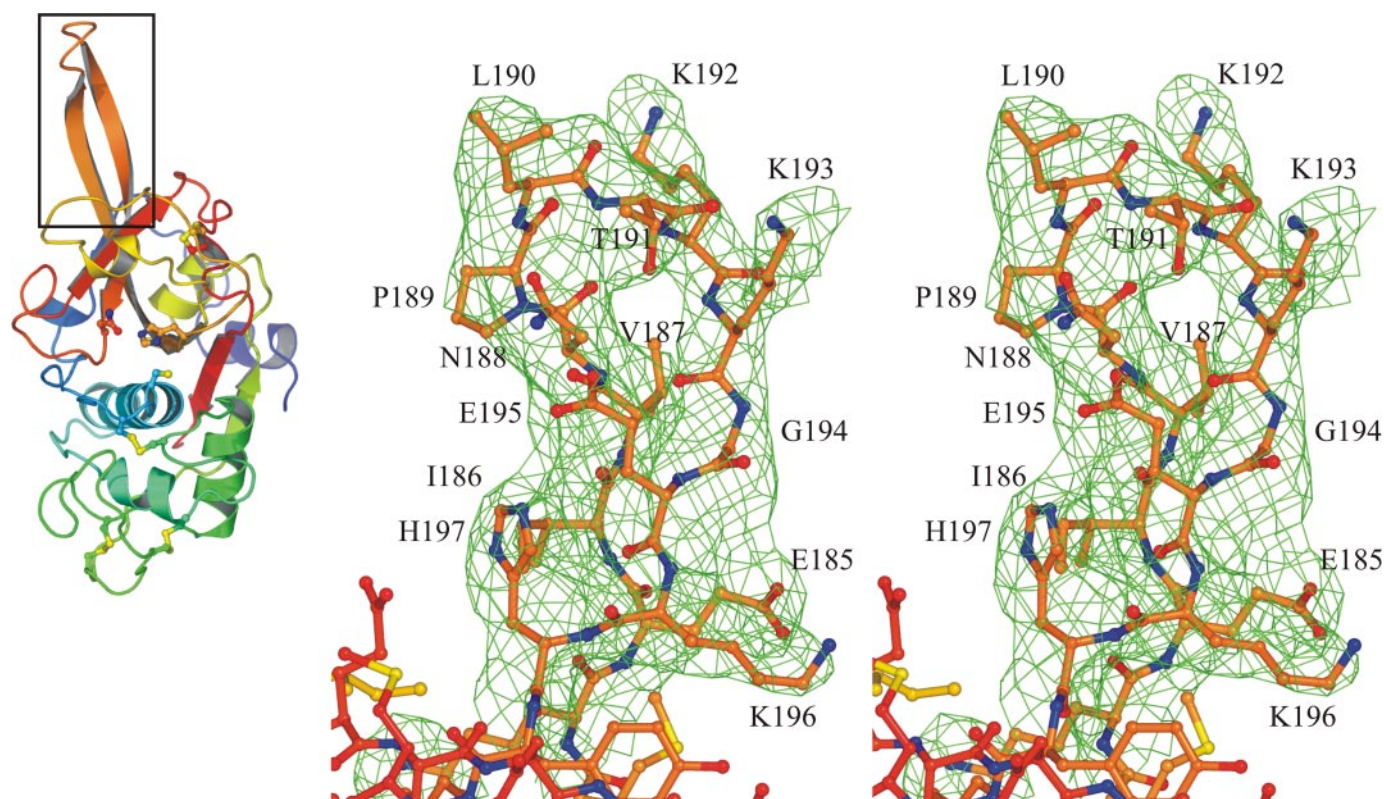


FIGURE 5. **Stereoscopic view of the falcipain-2 hemoglobin binding β -hairpin.** A ball-and-stick model for the hemoglobin-binding β -hairpin and neighboring residues of monomer D is shown (the boxed region of the *inset* highlights the area of detail). A simulated-annealing omit map ($F_o - F_c; 3\sigma$) calculated for residues 184–197 of the β -hairpin is superimposed, and residues of the β -hairpin are labeled.

protease inhibitor) appeared in the PDB (code 1YVB).⁴ Although the details of the structure have not yet been reported, the release of the coordinates has allowed us to make a brief comparison between the two structures. The overall $C\alpha$ r.m.s.d. between monomer D of our crystal structure and falcipain-2 in the cystatin complex is 0.63 Å, with the largest $C\alpha$ deviations between the two structures occurring in the hemoglobin binding insert (Leu-190 at the tip of the β -hairpin exhibits the most prominent $C\alpha$ r.m.s.d (3.0 Å)). Because the β -hairpin is remotely positioned from cystatin in the falcipain-2-cystatin complex, the structural displacement does not appear to be a direct consequence of cystatin binding.

DISCUSSION

Catalytic Activity of Falcipain-2—The pH profile of falcipain-2 activity shows pronounced differences depending on the offered substrates. The broad pH optima of Z-Phe-Arg-pNA hydrolysis and of self-processing indicates that the active site of the protein is in a catalytically active state at least in the range from pH 4.5 (as shown in Ref. 18) to 8.5. Even minor substrate

modifications can apparently result in significantly different pH-dependent activity profiles; Shenai and co-workers (18) observed no activity at pH 8.0 with Z-Phe-Arg-AMC as a substrate, whereas 60% of the maximal activity was still observed at this pH with Z-Phe-Arg-pNA (Fig. 1*b*). The exact reason for this discrepancy remains to be established. Apparently, pNA is a better leaving group than AMC at this elevated pH. The cleavagespecificity of falcipain-2 appears to change in a pH-/substrate-dependent manner from the rather unspecific cleavage of hemoglobin under acidic conditions to the highly specific cleavage of band 4.1 (17) and the falcipain-2 precursor at a neutral to weakly alkaline pH (Fig. 1*d*). The appearance of processing intermediates at pH 8.0 and above (Fig. 1*d*) in the case of the truncated falcipain precursor may indicate a stepwise removal of the pro-sequence.

Because the catalytic site of the enzyme itself seems to be active over a wide pH range, the different pH profiles of the substrates are likely to reflect pH-controlled binding of the respective substrates. This hypothesis was found to be true for hemoglobin binding (Fig. 2). The observation that the affinity of falcipain-2 for methemoglobin is significantly higher than for hemoglobin was completely unexpected, although it had been demonstrated

⁴ S. X. Wang, unpublished data.

FIGURE 4. **Sequence alignment between falcipain-2 and related plasmodial homologues.** The sequence of the KDEL-tailed cysteine endopeptidase (KDEL-CysEP) (36) has been included for comparison. Red shading highlights residues that are conserved in all sequences. Residues conserved in at least 70% of the sequences are shaded in green. Blue shading indicates conservative replacements. The alignment was calculated using the program ClustalX 1.83 (70) using the identity matrix with a gap opening penalty of 10 and a gap extension penalty of 0.1 for the pairwise comparison and the Blossum matrix with a gap opening penalty of 10 and a gap extension penalty of 0.05 for the multiple comparison. The threshold for delaying the most divergent sequences was set to zero. Residue-specific penalties and hydrophilic penalties were used. The hydrophilic residues were defined as GPSNDQEK. The gap-separation distance was set to 8, and the end-gap separation was turned off. The secondary-structure elements of falcipain-2 as determined for monomer D are indicated above the alignment. Where possible, numbers are given for the first and last amino acid residue of each secondary-structure element.

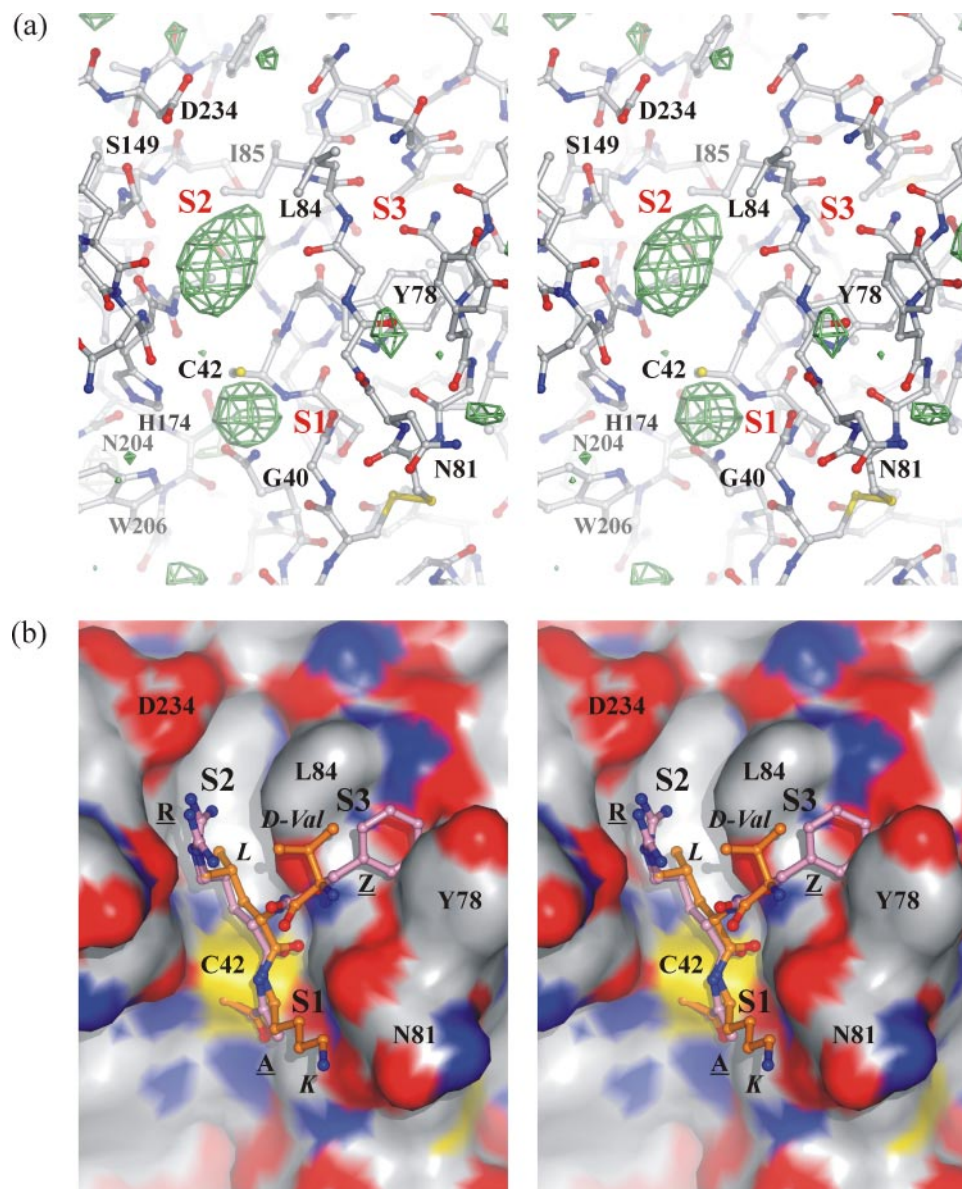


FIGURE 6. Stereoscopic representations of the falcipain-2 catalytic cleft. Residues are colored according to atom type (blue, nitrogen; gray, carbon; red, oxygen; yellow, sulfur). *a*, ball-and-stick model of the catalytic cleft overlaid with the final difference electron density map ($F_o - F_c$; 3σ , green). Prominent difference peaks are visible within the S2 pocket and next to the sulfur atom of the catalytic Cys42. Substrate specificity pockets S1, S2, and S3 are labeled in red; selected falcipain-2 residues are labeled in gray/black. *b*, surface representation of the catalytic cleft in an identical orientation as for panel *a*. Binding modes of benzoyl-arginine-alanine-fluoromethyl ketone in complex with cruzipain (61) are overlaid in pink ball-and-stick representations. Binding mode of D-valine-leucine-lysine-chloromethyl ketone in complex with the KDEL-tailed cysteine endopeptidase (36) is overlaid in orange ball-and-stick models. Substrate specificity pockets (S1, S2, and S3), inhibitor residues, and selected falcipain-2 residues are labeled. Labels for benzoyl-arginine-alanine-fluoromethyl ketone residues are underlined; those for D-valine-leucine-lysine-chloromethyl ketone are italicized.

that several factors contribute to the formation of methemoglobin during plasmodial infection, including the acidic pH of the plasmodial food vacuole, oxidative damage within infected erythrocytes (51, 52), and the reduced activity of NADH-methemoglobin reductase (53). A significantly increased methemoglobin content in the range of 20–42% has been detected in the plasmodial food vacuole compared with 0.6–1.0% in uninfected erythrocytes (51). Thus, the higher affinity of falcipain-2 for methemoglobin seems to reflect an adaptation to the conditions within *Plasmodia*-infected erythrocytes. This hypothesis is supported by further experimental data, where Sobolewski and co-workers (54) demon-

strated in *ex vivo* experiments using murine erythrocytes that the conversion of 95% of the hemoglobin into methemoglobin by the addition of NO had no detectable impact on the viability of *Plasmodium berghei*. The authors suggested that under these conditions, the parasites had either resorted to an alternative food source or were able to utilize methemoglobin. Akompong *et al.* (51) could even demonstrate that a reduction of the methemoglobin content by the addition of riboflavin resulted in a reduced size of the *P. falciparum* food vacuole and blocked parasite proliferation in erythrocyte cultures. Consequently, it seems that the formation of methemoglobin is not an unwanted side effect of a plasmodial infection but a prerequisite for the efficient utilization of hemoglobin as a food source by the parasite. Our data indicate that falcipain-2 is perfectly suited to perform this function.

Self-processing of Falcipain-2—Recombinant falcipain-2 was reported to be proteolytically processed during refolding (18). Our studies confirm these results, as N-terminal sequencing of the refolded protein exclusively revealed the sequence of the native mature protein (data not shown). Because we did not observe this auto-processing during the refolding of the inactive, truncated falcipain-2 precursor (FPC285a) (data not shown), we can now exclude the possibility that processing was an artifact due to trace amounts of a cysteine protease being present from the expression host. Based on inhibitor studies *in vivo*, Dahl and Rosenthal (19) previously concluded that falcipain-2 and -3 are autohydrolytically processed at a neutral pH before their secretion into the food vacuole. Our results de-

scribed in Fig. 1*d* demonstrate for the first time that falcipain-2 is capable of self-processing in *trans*, with an optimum at a weakly alkaline pH. This self-processing is fully sensitive to the cysteine-protease inhibitor E-64. Because the self-processing is still observable under acidic conditions, falcipain-2 may also contribute to the processing of other plasmodial proteins within the food vacuole. For instance, it has been shown that an as yet unidentified cysteine protease is involved in the processing of the plasmepsins (55, 56). However, the E-64 insensitivity of the plasmepsin processing (55, 56) would argue against a significant contribution of falcipain-2 to the processing of these particular proteins.

Structural Determinants of Protein Folding and Hemoglobin Recognition—Although the exact involvement of the 17-residue N-terminal segment in mediating correct folding of mature falcipain-2 is not clearly understood, previous mutational studies have pinpointed a critical role for the highly conserved Tyr-4 (20). The x-ray structure of falcipain-2 reveals this residue to be situated at the N terminus of helix $\alpha 1$, where it makes multiple side-chain interactions with $\beta 2$, $\alpha 5$, and the $\beta 2/\alpha 5$ bridging loop of the R domain core. This observation and the high sequence conservation of other residues mediating packing interactions between helix $\alpha 1$ and the bulk of the R domain would intuitively suggest that positioning of $\alpha 1$ in this region is critical for overall folding of the polypeptide and that the long $\alpha 1/\beta 1$ loop may act simply as a spacer to allow the N-terminal helix $\alpha 1$ to interact with the lower surface of the R domain. An essential role for the N-terminal extension in protein folding cannot be easily corroborated by a structural comparison of falcipain-2 with other C1A proteases lacking this structural element, because the underlying $\beta 2$ -loop- $\alpha 5$ structure does not exhibit any pronounced dissimilarities between the different proteins (Fig. 3c). The spatial proximity of the N-terminal $\alpha 1/\beta 1$ loop region to the hemoglobin binding β -hairpin does raise speculations of a possible function in hemoglobin capture. This is supported by a sequence comparison of the different plasmodial falcipain-2-subfamily proteases, which demonstrates that the $\alpha 1/\beta 1$ loop is restricted to a length of 10–11 residues (Fig. 4). Falcipain-1 is an exception in that it has a much larger $\alpha 1/\beta 1$ loop, but it has also been shown to be incapable of degrading hemoglobin (23). When taken together, the evidence alludes to a possible hemoglobin binding role for the N-terminal extension, particularly the $\alpha 1/\beta 1$ loop. Whether this structural element does indeed contribute to hemoglobin capture requires further examination.

The 14-residue hemoglobin binding β -hairpin (residues 183–196) exhibits a high degree of conformational flexibility when freely exposed to bulk solvent. The large positional deviations exhibited by the β -hairpin when comparing our structure and the structure of the falcipain-2-cystatin complex⁴ further underscore an inherent flexibility and plasticity of this element that may have functional implications in hemoglobin binding. A sequence comparison between the different falcipains and related homologues indicates a conservation of insert length (14 residues) but very low sequence identity (Fig. 4). Containing a much larger insertion in this region, falcipain-1 is once again an exception to this rule. A recent sequence analysis of 607 family C1A proteases has identified 40 members with insertions of longer than 12 residues at this location, with 18 of these members representing falcipains and falcipain homologues from other species (25). Other C1A proteases, such as cathepsin X (57), contain prominent insertions at other locations (Fig. 3c), although the biological role of these insertions remains less explored.

Toward Novel Antimalarial Drugs Targeting the Falcipains—Despite the absence of structural data concerning the falcipain-2-hemoglobin complex, an opportunity is now emerging to apply the current three-dimensional knowledge of the falcipain-2 hemoglobin-binding β -hairpin toward structure-guided design of inhibitors that interfere with falcipain-2-hemoglobin

complex formation. As an example, β -hairpin protein epitope mimetic inhibitors show great promise in the field of drug design, as exemplified by recent work demonstrating their utility in blocking complex formation between the p53 tumor-suppressor protein and its natural human inhibitor HDM2 (58). Indeed, preliminary work by Rosenthal and co-workers (25) has already shown that a noncyclic peptide comprising the hemoglobin binding motif effectively inhibits the *in vitro* degradation of hemoglobin by falcipain-2 in a dose-dependent manner. The feasibility of designing hemoglobin binding inhibitors that have broad-spectrum activity against members of the falcipain-2 subfamily awaits further research.

The active-site cleft of falcipain-2 subfamily proteases, as for other C1A proteases and cysteine proteases in general, represents a prime target for therapeutic intervention (59, 60). Peptides and irreversible peptidic inhibitors have traditionally been used to probe the substrate specificity of cysteine proteases and to generate information that can be translated into the design of non-peptidic inhibitors with drug-like properties. Structure-guided approaches have already proven indispensable in the design of potent and highly selective inhibitors against the closest known structural homolog of falcipain-2, cruzipain (cruzain), from *T. cruzi* (the causative agent of Chagas disease) (46, 47, 61, 62). An experimental cure for Chagas disease in mice using optimized vinyl-sulfone-derivatized pseudopeptides has been reported (63), and similar vinyl-sulfone inhibitors have shown marked antimalarial effects in mice (with up to a 40% cure rate after 4 days of oral administration) (64) and broad activity across multiple *P. falciparum* strains (65).

The S2 pocket is the major determinant of specificity for most cysteine proteases (66), and its predominantly hydrophobic nature in falcipain-2 agrees well with the observation that all characterized falcipain-2 cleavage motifs of its specific intracellular protein targets, including ankyrin (NVSAR ↓ FWLSD) (67), band 4.1 (SQEEIK ↓ KHHASI) (17), and the falcipain-2 precursor (KYLLD ↓ QMNYF) (18), contain a hydrophobic residue at the P2 position (marked in bold in the preceding sequences). Experiments with fluorogenic substrates have also shown that falcipain-2 has a strong preference for substrates with a hydrophobic residue, particularly Leu, at the P2 position (18). The crystal structure of falcipain-2 reveals that although the S2 subsite is generally hydrophobic, the electronegative side chain of Asp-234 is positioned very deep in the S2 pocket (Fig. 6). Previous x-ray structures of cruzipain and cathepsin B have uncovered a Glu side chain in the homologous position (Glu-205 in cruzipain; Glu-245 in cathepsin B) (45, 68). It has been demonstrated for cruzipain that the binding of substrates containing a basic side chain in the P2 position is mediated by salt-bridge formation between the P2 side chain and the carboxylate moiety of Glu-205 in the S2 pocket, thereby explaining the dual specificity of cruzipain and cathepsin B for substrates containing either a hydrophobic or basic residue at the P2 position (61). In falcipain-2 and falcipain-2B, however, the S2 pocket is slightly deeper due to the presence of the shorter Asp-234 side chain at this position, and it is therefore questionable whether an Arg or Lys in P2 could make a similar salt-bridge interaction (Fig. 6b). In designing selective inhibitors against falcipain-2 and falcipain-2B, it might be feasible to

exploit this feature of the S2 pocket by engineering a pseudopeptide inhibitor containing a longer basic side chain in the P2 position, such as homoarginine, which would presumably have the capacity to form a salt bridge with Asp-234. With the crystal structure at hand, synthetic chemistry can now be initiated to rationally explore the structural determinants of falcipain-2 specificity.

Acknowledgments—We thank Prof. T. Schirmeister (University of Würzburg) for supplying the initial falcipain-2 expression clone and Dr. A. Petersen (Research Center Borstel) for N-terminal sequencing. We thank S. Siedler, D. Mutschall, and S. Zoske for excellent technical assistance. R. H. thanks the Fonds der Chemischen Industrie for continuous support.

REFERENCES

- WHO (2005) *World Malaria Report*, WHO, Geneva
- Baird, J. K., Wiady, I., Sutanihardja, A., Suradi, Purnomo, Basri, H., Sekartuti, Ayomi, E., Fryauff, D. J., and Hoffman, S. L. (2002) *Am. J. Trop. Med. Hyg.* **66**, 659–660
- Baird, J. K., and Hoffman, S. L. (2004) *Clin. Infect. Dis.* **39**, 1336–1345
- Baird, J. K. (2005) *N. Engl. J. Med.* **352**, 1565–1577
- Jambou, R., Legrand, E., Niang, M., Khim, N., Lim, P., Volney, B., Ekala, M. T., Bouchier, C., Esterre, P., Fandeur, T., and Mercereau-Puijalon, O. (2005) *Lancet* **366**, 1960–1963
- Robert, A., Benoit-Vical, F., Dechy-Cabaret, O., and Meunier, B. (2001) *Pure Appl. Chem.* **73**, 1173–1188
- Kesavulu, M. M., Prakasha Gowda, A. S., Ramya, T. N., Surolia, N., and Suguna, K. (2005) *J. Pept. Res.* **66**, 211–219
- Silva, A. M., Lee, A. Y., Gulnik, S. V., Majer, P., Collins, J., Bhat, T. N., Collins, P. J., Cachau, R. E., Luker, K. E., Gluzman, I. Y., Francis, S. E., Oksman, A., Goldberg, D. E., and Erickson, J. W. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10034–10039
- Rosenthal, P. J., Kim, K., McKerrow, J. H., and Leech, J. H. (1987) *J. Exp. Med.* **166**, 816–821
- Rosenthal, P. J., McKerrow, J. H., Aikawa, M., Nagasawa, H., and Leech, J. H. (1988) *J. Clin. Invest.* **82**, 1560–1566
- Sijwali, P. S., Brinen, L. S., and Rosenthal, P. J. (2001) *Protein Expression Purif.* **22**, 128–134
- Sijwali, P. S., Shenai, B. R., Gut, J., Singh, A., and Rosenthal, P. J. (2001) *Biochem. J.* **360**, 481–489
- Goh, L. L., and Sim, T. S. (2005) *Biochem. Biophys. Res. Commun.* **335**, 762–770
- Singh, N., Sijwali, P. S., Pandey, K. C., and Rosenthal, P. J. (2006) *Exp. Parasitol.* **112**, 187–192
- Murata, C. E., and Goldberg, D. E. (2003) *J. Biol. Chem.* **278**, 38022–38028
- Murata, C. E., and Goldberg, D. E. (2003) *Mol. Biochem. Parasitol.* **129**, 123–126
- Hanspal, M., Dua, M., Takakuwa, Y., Chishti, A. H., and Mizuno, A. (2002) *Blood* **100**, 1048–1054
- Shenai, B. R., Sijwali, P. S., Singh, A., and Rosenthal, P. J. (2000) *J. Biol. Chem.* **275**, 29000–29010
- Dahl, E. L., and Rosenthal, P. J. (2005) *Mol. Biochem. Parasitol.* **139**, 205–212
- Pandey, K. C., Sijwali, P. S., Singh, A., Na, B. K., and Rosenthal, P. J. (2004) *J. Biol. Chem.* **279**, 3484–3491
- Sijwali, P. S., Shenai, B. R., and Rosenthal, P. J. (2002) *J. Biol. Chem.* **277**, 14910–14915
- Sijwali, P. S., and Rosenthal, P. J. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4384–4389
- Greenbaum, D. C., Baruch, A., Grainger, M., Bozdech, Z., Medzihradsky, K. F., Engel, J., DeRisi, J., Holder, A. A., and Bogoy, M. (2002) *Science* **298**, 2002–2006
- Eksi, S., Czesny, B., Greenbaum, B. D., Bogoy, M., and Williamson, K. C. (2004) *Mol. Microbiol.* **53**, 243–250
- Pandey, K. C., Wang, S. X., Sijwali, P. S., Lau, A. L., McKerrow, J. H., and Rosenthal, P. J. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9138–9143
- Rosenthal, P. J., Wollish, W. S., Palmer, J. T., and Rasnick, D. (1991) *J. Clin. Invest.* **88**, 1467–1472
- Gelhaus, C., Vicik, R., Hilgenfeld, R., Schmidt, C. L., Leippe, M., and Schirmeister, T. (2004) *Biol. Chem.* **385**, 435–438
- Semenov, A., Olsen, J. E., and Rosenthal, P. J. (1998) *Antimicrob. Agents Chemother.* **42**, 2254–2258
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) in *The Proteomics Protocols Handbook* (Walker, J. M., ed) pp. 571–607, Humana Press, Inc., Totowa, NJ
- Steller, I., Bolotovskiy, R., and Rossmann, M. G. (1997) *J. Appl. Crystallogr.* **30**, 1036–1040
- Leslie, A. G. W. (2006) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 48–57
- Evans, P. (2006) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82
- Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. Sect. D* **50**, 760–763
- Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) *Acta Crystallogr. Sect. D* **60**, 432–438
- Schwarzenbacher, R., Godzik, A., Grzechnik, S. K., and Jaroszewski, L. (2004) *Acta Crystallogr. Sect. D* **60**, 1229–1236
- Than, M. E., Helm, M., Simpson, D. J., Lottspeich, F., Huber, R., and Gietl, C. (2004) *J. Mol. Biol.* **336**, 1103–1116
- Read, R. J. (1986) *Acta Crystallogr. A* **42**, 140–149
- Terwilliger, T. C. (2004) *Acta Crystallogr. Sect. D* **60**, 2144–2149
- Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D* **60**, 2126–2132
- Brünger, 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D* **54**, 905–921
- DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA
- Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (1998) *Handbook of Proteolytic Enzymes*, Academic Press, San Diego, CA
- Barrett, A. J., and Rawlings, N. D. (2001) *Biol. Chem.* **382**, 727–733
- Gibrat, J.-F., Madej, T., and Bryant, S. H. (1996) *Curr. Opin. Struct. Biol.* **6**, 377–385
- McGrath, M. E., Eakin, A. E., Engel, J. C., McKerrow, J. H., Craik, C. S., and Fletterick, R. J. (1995) *J. Mol. Biol.* **247**, 251–259
- Brinen, L. S., Hansell, E., Cheng, J., Roush, W. R., McKerrow, J. H., and Fletterick, R. J. (2000) *Structure* **8**, 831–840
- Huang, L., Brinen, L. S., and Ellman, J. A. (2003) *Bioorg. Med. Chem.* **11**, 21–29
- Pickersgill, R. W., Harris, G. W., and Garman, E. (1992) *Acta Crystallogr. Sect. B* **48**, 59–67
- Sabnis, Y., Rosenthal, P. J., Desai, P., and Avery, M. A. (2004) *J. Biomol. Struct. Dyn.* **19**, 765–774
- Goh, L. L., and Sim, T. S. (2004) *Biochem. Biophys. Res. Commun.* **323**, 565–572
- Akompong, T., Ghori, N., and Haldar, K. (2000) *Antimicrob. Agents Chemother.* **44**, 88–96
- Müller, S. (2004) *Mol. Microbiol.* **53**, 1291–1305
- Stocker, R., Hunt, N. H., Buffinton, G. D., Weidemann, M. J., Lewis-Hughes, P. H., and Clark, I. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 548–551
- Sobolewski, P., Gramaglia, I., Frangos, J. A., Intaglietta, M., and van der Heyde, H. (2005) *Infect. Immun.* **73**, 6704–6710
- Francis, S. E., Banerjee, R., and Goldberg, D. E. (1997) *J. Biol. Chem.* **272**, 14961–14968
- Banerjee, R., Francis, S. E., and Goldberg, D. E. (2003) *Mol. Biochem. Parasitol.* **129**, 157–165
- Gunčar, G., Klemenčič, L., Turk, B., Turk, V., Karaoglanovic-Carmona, A., Juliano, L., and Turk, D. (2000) *Structure* **8**, 305–313
- Fasan, R., Dias, R. L., Moehle, K., Zerbe, O., Obrecht, D., Mittl, P. R. E., Grütter, M. G., and Robinson, J. A. (2006) *ChemBiochem* **7**, 515–526
- Lecaille, F., Kaleta, J., and Brömme, D. (2002) *Chem. Rev.* **102**, 4459–4488

60. Leung-Toung, R., Zhao, Y., Li, W., Tam, T. F., Karimian, K., and Spino, M. (2006) *Curr. Med. Chem.* **13**, 547–581
61. Gillmor, S. A., Craik, C. S., and Fletterick, R. J. (1997) *Protein Sci.* **6**, 1603–1611
62. Choe, Y., Brinen, L. S., Price, M. S., Engel, J. C., Lange, M., Grisostomi, C., Weston, S. G., Pallai, P. V., Cheng, H., Hardy, L. W., Hartsough, D. S., McMakin, M., Tilton, R. F., Baldino, C. M., and Craik, C. S. (2005) *Bioorg. Med. Chem.* **13**, 2141–2156
63. Engel, J. C., Doyle, P. S., Hsieh, I., and McKerrow, J. H. (1998) *J. Exp. Med.* **188**, 725–734
64. Olsen, J. E., Lee, G. K., Semenov, A., and Rosenthal, P. J. (1999) *Bioorg. Med. Chem.* **7**, 633–638
65. Singh, A., and Rosenthal, P. J. (2001) *Antimicrob. Agents Chemother.* **45**, 949–951
66. Bode, W., and Huber, R. (2000) *Biochim. Biophys. Acta* **1343**, 160–186
67. Dhawan, S., Dua, M., Chishti, A. H., and Hanspal, M. (2003) *J. Biol. Chem.* **278**, 30180–30186
68. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatar, T., Katunuma, N., and Bode, W. (1991) *EMBO J.* **10**, 2321–2330
69. Kamphuis, I. G., Kalk, K. H., Swarte, M. B., and Drenth, J. (1984) *J. Mol. Biol.* **179**, 233–256
70. Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., and Higgins, D. (1997) *Nucleic Acids Res.* **25**, 4876–4882
71. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) *Proteins* **12**, 345–364