Structural and Functional Relationships in the Virulence-associated Cathepsin L Proteases of the Parasitic Liver Fluke, Fasciola hepatica*

Received for publication, October 15, 2007, and in revised form, November 28, 2007 Published, JBC Papers in Press, December 26, 2007, DOI 10.1074/jbc.M70521200

Colin M. Stack‡1, Conor R. Caffrey‡, Sheila M. Donnelly‡, Amritha Seshadri‡, Jonathan Lowther‡, Jose F. Tort‡, Peter R. Collins‡, Mark W. Robinson‡3, Weibo Xu‡, James H. McKerrow§, Charles S. Craik‡, Sebastian R. Geiger‡4, Rachel Marion‡5, Linda S. Brinen‡5, and John P. Dalton‡5

From the ‡Institute for the Biotechnology of Infectious Diseases, University of Technology Sydney, Sydney, New South Wales 2007, Australia, §Sandler Center for Basic Research in Parasitic Diseases, University of California, San Francisco, California 94158, the †Departamento de Genetica, Facultad de Medicina, Universidad del la Republica, General Flores 2125, CP 11800, Montevideo, Uruguay, the ‡Departments of Pharmaceutical Chemistry, Pharmacology, and Biochemistry and Biophysics, University of California, San Francisco, California 94158, and the **Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158

The helminth parasite Fasciola hepatica secretes cysteine proteases to facilitate tissue invasion, migration, and development within the mammalian host. The major proteases cathepsin L1 (FheCL1) and cathepsin L2 (FheCL2) were recombiantly produced and biochemically characterized. By using site-directed mutagenesis, we show that residues at position 67 and 205, which lie within the S2 pocket of the active site, are critical in determining the substrate and inhibitor specificity. FheCL1 exhibits a broader specificity and a higher substrate turnover rate compared with FheCL2. However, FheCL2 can efficiently cleave substrates with a Pro in the P2 position and degrade collagen within the triple helices at physiological pH, an activity that among cysteine proteases has only been reported for human cathepsin K. The 1.4-A three-dimensional structure of the FheCL1 was determined by x-ray crystallography, and the three-dimensional structure of FheCL2 was constructed via homology-based modeling. Analysis and comparison of these structures and our biochemical data with those of human cathepsins L and K provided an interpretation of the substrate-recognition mechanisms of these major parasite proteases. Furthermore, our studies suggest that a configuration involving residue 67 and the “gatekeeper” residues 157 and 158 situated at the entrance of the active site pocket create a topology that endows FheCL2 with its unusual collagenolytic activity. The emergence of a specialized collagenolytic function in Fasciola likely contributes to the success of this tissue-invasive parasite.

Clan CA papain-like cysteine peptidases, such as cathepsins B and L (1), are ubiquitous in helminth (worm) parasites of human and veterinary importance. These peptidases are involved in a variety of pathogen-specific functions, including penetration and migration through host tissues, catabolism of host proteins to peptides and amino acids, and modulation or suppression of host immune defenses by cleaving immunoglobulin or altering the activity of immune effector cells (2–4). The central role of Clan CA proteases in the survival of helminth parasites has positioned them as lead targets for the development of new chemotherapies and vaccines (5–7).

Fasciola hepatica is a helminth parasite that causes liver fluke disease (fasciolosis) in cattle and sheep worldwide. It is most prevalent in Europe with infection rates increasing because of the emergence of drug-resistant parasites and possibly as a result of climate change (8, 9). Human fasciolosis has recently emerged as a major zoonosis in rural areas of South America (particularly Bolivia, Peru, and Equador), Egypt, and Iran where organized farm management practices are poor. It is estimated that worldwide over 2.4 million people are infected with F. hepatica and about 180 million are at risk of infection (10, 11).

Secretion of cysteine proteases is associated with the virulence of F. hepatica and its capacity to infect a wide range of mammalian hosts (4, 6, 12–14). Cathepsin L1 (FheCL1) and cathepsin L2 (FheCL2) are the two major peptidases secreted by the infective larvae that traverse the host intestinal wall, by the migratory stages that penetrate the liver tissues, and by the mature adult parasites that reside in the bile ducts and feed on host blood, which they ingest through the punctured bile duct wall (4, 6, 15). Experiments using purifiednative enzymes demonstrated that FheCL1 and FheCL2 efficiently degrade host
hemoglobin, immunoglobulin, and interstitial matrix proteins such as fibronectin, laminin, and native collagen (6, 16, 17). Although FheCL1 and FheCL2 exhibited similar substrate specificity, FheCL2 showed a greater affinity for peptides containing Pro residues in the P2 position (18–20). We proposed that by producing proteases with overlapping specificity the parasite could digest these host macromolecules more efficiently, and therefore more effectively penetrate host organs (6, 16).

The *F. hepatica* cathepsin Ls belong to a lineage that eventually gave rise to the mammalian cathepsin Ls from which the mammalian cathepsin Ks diverged (2). Mammalian cathepsin L is ubiquitously expressed in tissues and performs a housekeeping function in protein turnover, but it also plays a part in more specialized functions such as antigen processing and presentation, hormone and protease activation, and extracellular matrix turnover (21). Cathepsin K, on the other hand, exhibits a more restricted expression profile being predominantly found in osteoclasts but also in multinucleated giant cells, macrophages, and lung epithelial cells (22, 23). A specific role for cathepsin K in bone resorption by osteoclasts has been related to the ability of the protease to cleave the covalently linked triple helices of native collagen, a unique property among the mammalian papain-like cysteine proteases (24). This unusual property was attributed to the presence of a tyrosine residue at position 67 within the S2 subsite of cathepsin K that interacts with proline in the P2 of substrates, including the Gly-Pro-Xaa repeat sequence (where Xaa is mainly proline or 4-trans-L-hydroxyproline) found in collagen. A parallel therefore exists between mammalian cathepsin K and the *F. hepatica* FheCL1, in which the active site Cys was replaced by a tyrosine residue to simulate those residues present in human cathepsins L and K. Although FheCL1 and FheCL2 exhibited similar substrate specificities, FheCL2 showed a greater affinity for peptides containing Pro residues in the P2 position (18–20). We proposed that by producing proteases with overlapping specificity the parasite could digest these host macromolecules more efficiently, and therefore more effectively penetrate host organs (6, 16).

To understand the role of the major secreted cathepsin L proteases of *F. hepatica* in the virulence of the parasite and its adaptation to various hosts, it is important to elucidate their biochemical properties and relate these to structure and function. Therefore, in this study, we have characterized the substrate specificity of active recombinant forms of FheCL1 and FheCL2. These properties were further explored by preparing variants of FheCL1 in which specific substitutions were made within the S2 subsite of cathepsin K that interacts with proline in the P2 of substrates, including the Gly-Pro-Xaa repeat sequence (where Xaa is mainly proline or 4-trans-L-hydroxyproline) found in collagen. A parallel therefore exists between mammalian cathepsin K and the *F. hepatica* FheCL2 as the latter can also cleave substrates with a P2 proline and possesses a tyrosine residue at the corresponding position 67.

**EXPERIMENTAL PROCEDURES**

Materials—Z-Phe-Arg-NHMec, Z-Leu-Arg-NHMec, Z-Pro-Arg-NHMec, Z-Val-Pro-Arg-NHMec, Z-Gly-Pro-Arg-NHMec, Z-Ala-Gly-Pro-Arg-NHMec, Z-Phe-Arg-NHMec, Z-Gly-Pro-Lys-NHMec, and Z-Phe-Ala-CHN₂ were obtained from Bachem (St. Helens, UK). Z-Leu-Arg-NHMec was purchased from Peptide Institute Inc. (Japan). E-64, DTT, and EDTA were obtained from Sigma. Cathepsin K inhibitor II was purchased from BD Biosciences. Prestained molecular weight markers and the AvrII and SnaBI restriction enzymes were obtained from New England Biolabs. Primers were obtained from Sigma-Genosys. The pPIC9K vector and *Pichia pastoris* strain GS115 were obtained from Invitrogen. Nickel-nitritoltriacetic acid-agarose and columns were obtained from Qiagen. Pre-cast 4–20% gradient SDS-polyacrylamide gels were purchased from Gradipore (Australia).

**Expression and Purification of Recombinant Cathepsin L Zymogens in Yeast—F. hepatica** procathhepsin L1 (FheCL1) and procathhepsin L2 (FheCL2) were amplified by PCR from the pAAH5 *Saccharomyces cerevisiae* expression vector into which the full-length cDNA had been cloned previously in our laboratory (12, 25). FheCL1 variants (FheCL1 L67Y and FheCL1 L205A) were synthesized and incorporated an SnaBI restriction site at the 5’ end of the gene and an AvrII restriction site and Histag sequence at the 3’ end (Geneart, Regensburg, Germany). The 980-bp fragments were ligated into pCR-Script cloning vector (Stratagene), which were transformed into competent *Escherichia coli* for amplification. Inserts were digested from plasmid preparations with AvrII and SnaBI and inserted in-frame with the yeast α-factor at the AvrII/SnaBI site of *P. pastoris* expression vector pPIC9K (Invitrogen). Plasmids were linearized with Sall and then transformed into chemically competent GS115 cells (Invitrogen) as described previously (12). All inserts were sequenced to ensure congruence with original cDNAs.

*P. pastoris* yeast transformants were cultured in 500 ml of buffered glycerol complex medium broth, buffered to pH 8.0, in 5-liter baffled flasks at 30 °C until an A₅₀₀ of 2.6 was reached (12). Cells were harvested by centrifugation at 2000 × g for 5 min, and protein expression was induced by resuspending in 100 ml of buffered minimal methanol medium broth, buffered at pH 6.0 containing 1% methanol (20). Recombinant proteins were purified from yeast medium by affinity chromatography using nickel-nitritoltriacetic acid-agarose (Qiagen) (12, 26). Purified recombinant zymogens were dialyzed against phosphate-buffered saline (PBS) and stored at −20 °C. The 37-kDa zymogens were autocatalytically activated and processed to 24.5-kDa mature enzymes by incubation for 2 h at 37 °C in 0.1 M sodium citrate buffer, pH 5.0, containing 2 mM DTT and 2.5 mM EDTA. The mixture was then dialyzed against PBS, pH 7.3. The proportion of functionally active recombinant protein in these preparations was determined by titration against E-64.

**P1–P4 Specificity Using a Positional Scanning Synthetic Combinatorial Library**—The substrate specificities of FheCL1, FheCL1 L67Y, and FheCL1 L205A and FheCL2 were determined using a complete diverse positional scanning synthetic...
F. hepatica Cathepsin L Structure-Activity Relationships

combinatorial library (PS-SCL) (27). Screens were performed at 25 °C in 0.1 M sodium acetate, 0.1 M NaCl, 0.01 M DTT, 0.001 M EDTA, 0.01% Brij-35, 1% Me2SO (from the substrates), pH 5.5. Aliquots of 25 nmol in 1 μl from each of 20 sub-libraries of the P1, P2, P3, and P4 libraries were added to the wells of a 96-well Microfluor-1 U-bottom plate (Dynex Technologies). The final concentration of each compound of the 8000 compounds per well was 31.25 nM in a 100-μl final reaction volume. The assays were initiated by addition of preactivated enzyme, and the reaction was monitored with a SpectraMax Gemini fluorescence spectrometer ( Molecular Devices) with excitation at 380 nm, emission at 460 nm, and cutoff at 435 nm. Screens were performed in duplicate and triplicate for wild type and mutated enzymes, respectively.

Enzyme Assays and Kinetics with Fluorogenic Peptide Substrates—Initial rates of hydrolysis of the fluorogenic dipetidic substrates were measured by monitoring the release of the fluorogenic leaving group, NHMec, at an excitation wavelength of 380 nm and an emission wavelength of 460 nm using a BioTek KC4 microfluorometer. k_{cat} and K_m values were determined using nonlinear regression analysis. Initial rates were obtained at 37 °C over a range of substrate concentrations spanning K_m values (0.2–200 μM) and at fixed enzyme concentrations (0.5–5 μM). Assays were performed in PBS, pH 7.3, or PBS, pH 7.3. Reactions contained 10 μg of dialyzed collagen type 1, 1 mM DTT; and 2.5 mM EDTA and 5.47 μM activated peptidase in a final volume of 100 μl of one of the above buffers. Reactions were performed at 28 °C for 3 and 20 h or at 37 °C for 30 min. All reactions were stopped by the addition of 10 μM E-64. Collagen digests were analyzed by 4–20% gradient SDS-PAGE under reducing conditions and stained with Coomassie Brilliant Blue R-250.

Collagen Digestion—Calf skin collagen type 1 was solubilized in 0.2 M acetic acid at a concentration of 2 mg/ml and dialyzed for 2 days against 0.1 M sodium acetate, pH 4.0, 0.1 M sodium acetate, pH 5.5, or PBS, pH 7.3. Reactions contained 10 μg of dialyzed collagen type 1, 1 mM DTT, and 2.5 mM EDTA and 5.47 μM activated peptidase in a final volume of 100 μl of one of the above buffers. Reactions were performed at 28 °C for 3 and 20 h or at 37 °C for 30 min. All reactions were stopped by the addition of 10 μM E-64. Collagen digests were analyzed by 4–20% gradient SDS-PAGE, which represents the fullzymogen containing a prosegment and mature enzyme domain (data not shown).

Data Collection, Structure Solution, and Crystallographic Refinement of FheproCL1 Gly25—For the purpose of obtaining a high resolution three-dimensional structure of FheCL1, an inactive enzyme was produced by replacing the active site Cys residue at position 25 in the mature domain by a Gly (12, 26). This FheproCL1 Gly25 enzyme migrated as a single protein of 37 kDa on reducing 12% SDS-PAGE, which represents the fullzymogen containing a prosegment and mature enzyme domain (data not shown).

Production of Inactive Variant FheproCL1 Gly25—For the purpose of obtaining a high resolution three-dimensional structure of FheCL1, an inactive enzyme was produced by replacing the active site Cys residue at position 25 in the mature domain by a Gly (12, 26). This FheproCL1 Gly25 enzyme migrated as a single protein of 37 kDa on reducing 12% SDS-PAGE, which represents the fullzymogen containing a prosegment and mature enzyme domain (data not shown).

Enzyme Assays and Kinetics with Fluorogenic Peptide Substrates—Initial rates of hydrolysis of the fluorogenic dipetidic substrates were measured by monitoring the release of the fluorogenic leaving group, NHMec, at an excitation wavelength of 380 nm and an emission wavelength of 460 nm using a BioTek KC4 microfluorometer. k_{cat} and K_m values were determined using nonlinear regression analysis. Initial rates were obtained at 37 °C over a range of substrate concentrations spanning K_m values (0.2–200 μM) and at fixed enzyme concentrations (0.5–5 μM). Assays were performed in PBS, pH 7.3, or PBS, pH 7.3. Reactions contained 10 μg of dialyzed collagen type 1, 1 mM DTT; and 2.5 mM EDTA and 5.47 μM activated peptidase in a final volume of 100 μl of one of the above buffers. Reactions were performed at 28 °C for 3 and 20 h or at 37 °C for 30 min. All reactions were stopped by the addition of 10 μM E-64. Collagen digests were analyzed by 4–20% gradient SDS-PAGE, which represents the fullzymogen containing a prosegment and mature enzyme domain (data not shown).

Diffraction data were collected at the Advanced Light Source, beam line 8.3.1, using monochromatic (Si-111) radiation of 1.11588 Å (31). An ADSC Quantum 210 2 × 2 CCD array detector was used with low temperature conditions of 100 K at the crystal position. Crystals of the single mutant protein were flash-cooled in liquid nitrogen after being soaked for ~1.5 min in a cryoprotectant solution of crystal growth solution plus 50% 2-methyl-2,4-pentane diol. High and low resolution data-sets were collected from the same crystal. Data processing was completed with MOSFLM (32) and SCALA. The structure was solved via molecular replacement using the MOLREP program of the CCP4 suite (33) with a polyserine search model derived from the 1.8 Å structure of 1CS8 (human procathepsin L). The topmost solution had an R-factor value of 0.535 and correlation coefficient of 0.288, each several σ levels above the next best solution, which had corresponding statistics of 0.604 and 0.083.
respectively. One unique solution was found with one molecule in the asymmetric unit and a starting resolution bin (1.48–1.40 Å).

Note the values given in parentheses under Data Collection are for highest resolution data.

| Crystallographic parameters: data collection and refinement statistics |
|-------------------------------------------------|
| **Data Collection**                               |
| Space group                                        |
| P2 2 2                                           |
| Unit cell parameters                              |
| a 57.17 Å                                        |
| b 105.78 Å                                       |
| c 49.11 Å                                        |
| Wavelength                                        |
| 1.1159 Å                                         |
| Temperature                                       |
| 100 K                                             |
| Resolution                                        |
| 1.4 Å                                             |
| Total no. of reflections                          |
| 333,582 (19,545)                                  |
| Total unique reflections                          |
| 59,595 (8465)                                     |
| Completeness                                      |
| 99.7 (98.7%)                                      |
| Redundancy                                        |
| 5.6 (2.3)                                         |
| Rmerge                                           |
| 0.080 (0.549)                                     |
| Rfree                                            |
| 0.029 (0.431)                                     |
| (Rmerge)                                         |
| 14.6 (1.8)                                        |

| **Refinement**                                   |
| Resolution range (Å)                            |
| 52.93–1.40 Å                                     |
| No. of reflections                               |
| 56,141                                           |
| Rfactor                                          |
| 0.128                                            |
| Rfree                                           |
| 0.165                                            |
| Free reflections                                 |
| 5.05%                                            |
| Average σ2 factor (Å²)                           |
| Protein                                          |
| 15.39                                            |
| Water                                            |
| 34.09                                            |
| Root mean square deviation from ideal bond lengths |
| 0.020 Å                                          |
| Bond angles                                      |
| 1.79°                                            |
| Ramachandran Plot                                |
| Residues in most favored regions                |
| 238 (90.5%)                                      |
| Residues in additional allowed regions          |
| 25 (9.5%)                                        |
| Residues in generously allowed regions          |
| 0 (0.0%)                                         |
| Residues in disallowed regions                  |
| 0 (0.0%)                                         |

Homology-based Molecular Modeling—A model structure of the mature domain of FheCL2 was built using Modeler (release 8, version 1), a program for protein structure modeling (39–41). The 1.8 Å structure of human procathepsin L (PDB code 1CS8), the 2.2 Å structure of human cathepsin K (PDB code 1ATK), and our 1.4 Å solved structure of FheproCL1 Gly-25 were used as three-dimensional templates of related fold. Generated models were visualized and compared with COOT (37) and with PyMOL (42).

Sequence Analysis—F. hepatica cathepsin L protein sequences were aligned using Clustal X 1.81. Phylogenetic trees were generated from the alignment by the bootstrapped (1000-trial) neighbor-joining method using MEGA (43).

RESULTS

Active Site Residues Involved in Substrate Specificity of FheCL1 and FheCL2—Residues that make up the S2 pocket of FheCL1 and FheCL2 were determined using the three-dimensional x-ray crystal structure of FheCL1 and homology-based model of FheCL2, respectively, (see below), and their comparison to the structure of human cathepsin L (PDB code 1CS8) and cathepsin K (PDB code 1ATK) is shown in Table 2 (see also Fig. 1 and Fig. 7A; papain numbering is used). Most variation between papain cysteine proteases occurs at residues 67 and 205, and studies with human cathepsin L and cathepsin K demonstrated that the difference in residues 67 (Leu and Tyr, respectively) and 205 (Ala and Leu, respectively) reflect the striking difference in the substrate specificity of these two enzymes; for example, human cathepsin L exhibits a broad specificity and favors both aromatic and aliphatic P2 residues but will not accept proline, whereas cathepsin K prefers only aliphatic residues and most particularly proline. Indeed, the acceptance of a P2 Pro residue confers cathepsin K with its unique ability to cleave native type I and type II collagens, proteins that contain repeated Gly-Pro-X motifs (44). Like human cathepsin L, FheCL1 possesses a Leu at position 67; however, unlike cathepsin L it possesses a Leu at position 205 rather than an Ala. The Leu at position 205 is similar to cathepsin K, and thus, the FheCL1 exhibits hybrid character in the S2 subsite. By contrast, FheCL2 possesses a Tyr at position 67 and Leu at position 205 and hence is identical to cathepsin K at both sites. It has been suggested by us (2) and others (44, 45) that the accommodation of Pro in the P2 position of peptide substrates by FheCL2 may be related to the presence of the Tyr67, analogous to the cathepsin K scenario.

To address the relationship between the residues presented at position 67 and 205 and the substrate specificity and function of FheCL1 and FheCL2, we prepared variants of FheCL1 as shown in Table 2. The FheCL1 L67Y variant has a single amino acid change making the S2 subsite similar to FheCL2 and cathepsin K at positions 67 and 205. The FheCL1 L205A variant has a single amino acid that was designed to make the S2 subsite similar to human cathepsin L. The wild type and variant F. hepatica cathepsin L peptides were recombinantly expressed in the methylo trophic yeast P. pastoris, purified, and activated as described under
Experimental Procedures. All enzymes were expressed as 37-kDa zymogens that autocatalytically processed at pH 4.5 to produce 24.5-kDa mature enzymes, which was confirmed by N-terminal sequencing (Fig. 2). Enzymatic assays showed that all substitutions made in the S2 subsite of the FheCL1 active site did not alter its pH profile for activity against the fluorogenic substrate Z-Phe-Arg-NHMec; both the wild type and variant free enzymes exhibited a Gaussian bell-shaped pH profile with an optimum for activity in the region pH 6.5 to 7.0 (pKᵢ = 3.87 ± 0.07 and pKᵢ = 8.14 ± 0.08).

Substrate Specificity Profiling Using a PS-SCL Reveals Unique and Distinct Activities of FheCL1 and FheCL2—Wild type FheCL1 and FheCL2 exhibited similar preferences for amino acids at P1. As expected for papain-like cysteine proteases, both enzymes had a clear preference for Arg at P1, but other residues accommodated in this position included Lys, Glu, Thr, and Met (Fig. 3, P1 panel), and these were all cleaved at similar relative rates to that observed for human cathepsin L and cathepsin K (44). Similar results were obtained for the variants FheCL1 L67Y and FheCL1 L205A, which were expected as the introduced substitutions do not affect the S1 active site pocket (not shown).

A P1-Arg fixed library was then used to explore P2–P4 specificities of FheCL1 and FheCL2. The enzymes show a distinct preference for hydrophobic amino acids in the P2; both favored Leu. Interestingly, however, the positional scanning method did not identify Phe as a suitable P2 residue even though our kinetic studies demonstrate that both FheCL1 and FheCL2, like other papain cysteine proteases, cleave fluorogenic substrates with a P2 Phe efficiently (see Table 3). The most striking observation was the distinct preference for Pro residues by FheCL2, particularly when compared with FheCL1 that did not accommodate this residue (Fig. 3, P2 panel). The unusual preference for a P2 Pro exhibited by FheCL2 is similar to that observed for human cathepsin K using the same methodology (44). However, whereas human cathepsin K favored equally Ile and Leu at P2 (44), both Fasciola cathepsins were more similar to human cathepsin L by preferring Leu over Ile.

The replacement of Leu for Ala at residue 205 (FheCL1 L205A) markedly altered the activity profile from wild type enzyme (Fig. 4). This variant exhibited a broader substrate specificity by accepting Phe, Trp, and Tyr at P2, residues that were not accepted by wild type FheCL1. The same residues are also accommodated by human cathepsin L (44), thus demonstrating that the replacement of Leu205 for Ala in FheCL1 generates an enzyme more similar to the human orthologue. By contrast, the FheCL1 L67Y variant did not show a significant change in the P2 preference to wild type FheCL1; in particular, this substitution did not alter the activity of the enzyme toward Pro in the P2 position (Fig. 4). This was a surprising result as we expected that the FheCL1 L67Y variant would behave similarly to FheCL2 and cathepsin K given that the residues at positions 67 and 205 were identical.

As anticipated, the P3 and P4 specificities for FheCL1 and FheCL2 were similar, and like human cathepsin L and cathepsin K, the Fasciola enzymes accepted a broad range of residues in these positions. The P3–P4 specificity of FheCL1 was unaffected by the P2 substitutions present in the variant proteases (not shown).
Wild type and Variant Protease Specificities against Fluorogenic Peptide Substrates Correlates with Residues at Position 67 and 205—To support and extend the data derived from the positional scanning libraries, and to determine substrate kinetic parameters ($K_m$, $k_{cat}$, and $k_{cat}/K_m$) for wild type FheCL1, the variants FheCL1 L67Y and FheCL1 L205A, and wild type FheCL2, we examined their hydrolytic activity against various fluorogenic di- and tripeptides (Table 3). FheCL1 efficiently cleaved both Z-Phe-Arg-NHMec ($k_{cat}/K_m = 1,021,092 \text{ M}^{-1} \text{s}^{-1}$) and Z-Leu-Arg-NHMec ($k_{cat}/K_m = 8,395,402 \text{ M}^{-1} \text{s}^{-1}$), the enzyme cleaved the latter substrate over eight times more rapidly because its $K_m$ value for this substrate is much lower than for the former substrate. Although the substrates Z-Pro-Arg-NHMec ($k_{cat}/K_m = 5,387 \text{ M}^{-1} \text{s}^{-1}$), Tos-Gly-Pro-Arg-NHMec ($k_{cat}/K_m = 35,928 \text{ M}^{-1} \text{s}^{-1}$), and Boc-Ala-Gly-Pro-Arg-NHMec ($k_{cat}/K_m = 46,021 \text{ M}^{-1} \text{s}^{-1}$) were cleaved relatively poorly, nevertheless, the data indicate that FheCL1 can accommodate proline residues in the P2 position.

In comparison with FheCL1, FheCL2 is much less efficient at cleaving substrates with Phe and Leu in the P2 position.

![FIGURE 3. Profiling of the P1–P4 substrate specificity of FheCL1 and FheCL2 using positional scanning synthetic combinatorial libraries.](image)

**TABLE 3**

| Enzyme   | Substrate          | $k_{cat}$ | $K_m$ | $k_{cat}/K_m$ |
|----------|--------------------|-----------|-------|---------------|
| FheCL1   | Z-FR-NMec          | 24.69 ± 1.30 | 24.18 ± 3.92 | 1,021,092 |
| FheCL1 L205A | Z-FR-NMec     | 29.60 ± 1.05 | 19.21 ± 3.78 | 1,540,864 |
| FheCL1 L67Y | Z-FR-NMec     | 3.58 ± 0.11  | 8.16 ± 0.86  | 438,725  |
| FheCL2   | Z-FR-NMec          | 1.70 ± 0.07  | 39.94 ± 5.12 | 42,564   |
| FheCL1   | Z-LR-NMec          | 36.52 ± 0.63 | 4.35 ± 0.21  | 8,395,402 |
| FheCL1 L205A | Z-LR-NMec     | 9.15 ± 0.24  | 2.75 ± 0.61  | 3,327,273 |
| FheCL1 L67Y | Z-LR-NMec     | 1.73 ± 0.38  | 0.38 ± 0.05  | 4,552,632 |
| FheCL2   | Z-LR-NMec          | 1.62 ± 0.08  | 0.19 ± 0.02  | 1,165,468 |
| FheCL1   | Z-PR-NMec          | 1.03 ± 0.04  | 191.21 ± 16.90 | 5,387       |
| FheCL1 L205A | Z-PR-NMec    | 0.122 ± 0.005| 48.41 ± 4.45 | 2,479        |
| FheCL1 L67Y | Z-PR-NMec    | 0.62 ± 0.04  | 136.98 ± 12.03 | 4,526        |
| FheCL2   | Z-PR-NMec          | 2.64 ± 0.13  | 84.03 ± 11.28 | 31,417      |
| FheCL1   | Tos-GPR-NMec       | 0.36 ± 0.03  | 10.02 ± 1.20  | 35,928      |
| FheCL1 L205A | Tos-GPR-NMec | 0.113 ± 0.003| 20.35 ± 1.00  | 5,405      |
| FheCL1 L67Y | Tos-GPR-NMec | 0.26 ± 0.01  | 6.96 ± 1.08   | 37,096      |
| FheCL2   | Tos-GPR-NMec       | 1.17 ± 0.08  | 15.33 ± 2.62  | 76,321      |
| FheCL1   | Boc-AGPR-NMec      | 0.48 ± 0.03  | 10.43 ± 2.63  | 46,021      |
| FheCL1 L205A | Boc-AGPR-NMec | 0.20 ± 0.04  | 21.57 ± 8.85  | 9,179       |
| FheCL1 L67Y | Boc-AGPR-NMec | 0.93 ± 0.03  | 11.13 ± 1.37  | 83,378      |
| FheCL2   | Boc-AGPR-NMec      | 2.48 ± 0.08  | 33.69 ± 0.99  | 73,731      |
the $k_{\text{cat}}/K_m$ values for Z-Phe-Arg-NHMec and Z-Leu-Arg-NHMec with this enzyme are 24- and 7-fold lower than for FheCL1, respectively. However, its ability to cleave Z-Pro-Arg-NHMec, Tos-Gly-Pro-Arg-NHMec, and Boc-Ala-Gly-Pro-Arg-NHMec is 6-, 2-, and 2-fold higher than that of FheCL1 (Table 3). The kinetic data show that the S2 subsite of FheCL2 is able to accommodate proline residues more readily than the S2 subsite of FheCL1 and is in agreement with the data obtained by PS-SCL (Figs. 3 and 4).

Substitution at the 205 position of FheCL1 to generate variant FheCL1 L205A had a significant impact on the substrate specificity of the enzyme by increasing its ability to cleave Z-Phe-Arg-NHMec, while reducing its effectiveness on Z-Leu-Arg-NHMec, Z-Pro-Arg-NHMec, Tos-Gly-Pro-Arg-NHMec, and Boc-Ala-Gly-Pro-Arg-NHMec (Table 3). Substitution at position 67 to give the variant FheCL1 L67Y reduced the efficiency of the enzyme for both Z-Phe-Arg-NHMec and Z-Leu-Arg-NHMec about 2-fold, which was reflected in a reduction of both $k_{\text{cat}}$ and $K_m$ values for each substrate. This substitution did not significantly alter the specificity of the enzyme for the substrate Z-Pro-Arg-NHMec or Tos-Gly-Pro-Arg-NHMec, although it almost doubled its efficiency on Boc-Ala-Gly-Pro-Arg-NHMec (Table 3).

**Kinetic Analyses of Wild Type and Variant Proteases with Specific Inhibitors**—Peptidyl diazomethyl ketones are irreversible inhibitors of cysteine proteases (46). Changes in rates of inactivation by these inhibitors have highlighted different specificities at subsites of cysteine proteases such as cathepsin L and cathepsin B (47). In this study, rates of inactivation of FheCL1, FheCL1 L205A, FheCL1 L67Y, and FheCL2 by the cathepsin inhibitor Z-Phe-Ala-CHN$_2$ have been measured. Wild type FheCL1 and FheCL2 had second-order rate constants of 20,838 and 11,899 M$^{-1}$ s$^{-1}$, respectively, showing that both enzymes were rapidly inactivated by Z-Phe-Ala-CHN$_2$ (Table 4). The 2-fold greater rate of inactivation of FheCL1 compared with FheCL2 is further evidence that FheCL1 accommodates hydrophobic P2 residues more effectively than FheCL2.

The rate of inactivation of FheCL1 L205A was 24-fold greater than wild type indicating that Ala at residue 205 in the S2 subsites binds a P2 Phe more effectively than a Leu, which is consistent with our substrate kinetics studies (Table 3) and data derived from our tetrameric peptide library. These data highlighted further the major impact that Ala at position 205 has on binding P2 residues, and it is interesting to note that the second-order rate constant of 492,727 M$^{-1}$ s$^{-1}$ (Table 4) for the inactivation of FheCL1 L205A by Z-Phe-Ala-CHN$_2$ is similar to the value of 660,000 M$^{-1}$ s$^{-1}$ for the inactivation of mammalian cathepsin L by the same inhibitor (47, 48). By contrast, the FheCL1 L67Y variant had a $k_{\text{obs}}/I$ value of 53,704 M$^{-1}$ s$^{-1}$ (Table 4), which is only 2.5-fold higher than wild type FheCL1 and 5-fold greater than wild type FheCL2; therefore, this substitution has not been such a major influence on the binding of Phe in the S2 pocket.

The inhibitor known as cathepsin K Inhibitor II (Z-LNH-NHCONHNHFL-Boc, CKII) is a potent time-dependent inhibitor...
of human cathepsin K; its selectivity for this enzyme is largely because of the effectiveness by which leucine occupies the S2 subsite (49). FheCL1 and FheCL2 were both potently inhibited by cathepsin K inhibitor II with $k_{\text{obs}}/[I]$ values of 397,237 and 269,447 M$^{-1}$ s$^{-1}$, respectively, which are 20 times higher than that observed for the peptidyl diazomethyl ketone Z-Phe-Ala-CH$_2$(compare Tables 4 and 5). These values are similar to the value of 590,000 M$^{-1}$ s$^{-1}$ reported for the inactivation of cathepsin K by Wang et al. (49).

The rate of inactivation of FheCL1 L205A by cathepsin K inhibitor II was 7-fold lower than wild type. The $K_{\text{app}}$ increased 11-fold demonstrating that this variant cannot accommodate leucine in the S2 subsite to the same extent as wild type FheCL1. Because the rate of inactivation of human cathepsin L by cathepsin K Inhibitor II was 53-fold lower than that for human cathepsin K (Table 5), these data indicate that the FheCL1 L205A variant has S2 specificity more characteristic of human cathepsin L. The rate of inactivation of FheCL1 L67Y by cathepsin K inhibitor II was 3.5-fold lower than wild type, although the $K_{\text{app}}$ did not change significantly, against showing that the Tyr substitution at this position exerts a relatively lower effect on P2 binding.

Wild Type FheCL2 but Not Wild Type FheCL1 or Its Variants Cleaves Native Collagen Type I—FheCL1 and FheCL2 degraded type 1 collagen at pH 4.0 and 5.5 in reactions held at 28 °C, but the activity of FheCL1 was much less and was limited to the $\beta$ and $\gamma$ chains, whereas the $\alpha_1$ and $\alpha_2$ chains remained intact. Moreover, whereas FheCL1 produced clear degradation fragments, FheCL2 degraded the collagen completely, particularly at pH 4.0, indicating that only the latter degrades efficiently within the helical structures (Fig. 5A). Because low pH may cause some structural unraveling of the collagen, additional studies were performed at neutral pH. FheCL1 and FheCL2 both exhibit optimum activity against fluorogenic substrates in the neutral pH range. However, FheCL1 exhibited minimal activity against type 1 collagen in PBS, pH 7.3, whereas FheCL2 cleaved within all collagen chains (Fig. 5B).

Like the wild type FheCL1 enzyme, FheCL1 L205A and FheCL1 L67Y variants cleaved collagen but were unable to cleave within the tightly wound helices. Although in all experiments FheCL1 L205A appeared to cleave collagen more efficiently than the wild type enzyme, the pattern of digested fragments was similar (Fig. 5B). Nevertheless, the greater efficiency of cleavage of collagen is consistent with this variant’s enhanced activity against fluorogenic substrates (Table 3). The inability of the FheCL1 L67Y variant to cleave within the helices of collagen is also consistent with the PS-PCL studies and substrate kinetics studies, because this enzyme did not show any increase in preference for P2 Pro compared with wild type FheCL1.

The FheproCL1 Gly25 Structure in Comparison to Other Clan CA Cysteine Proteases—The experimentally determined structure of FheproGly125 is quite similar to that of previously described mammalian cathepsins. Although the x-ray crystal structure of FheCL1 presented here is that of an inactivezymogen mutant in which active site Cys has been mutated to glycine, the remainder of the active site machinery is intact, and the key specificity determinant, the S2 pocket, has not been altered.

The molecule, FheproGly125, is similar in tertiary structure to human cathepsin L1. Electron density is clear, connected, and easily traceable for the entirety of the main chain of the mature domain of FheproGly125. The mature domain is bi-lobed, with a substrate-binding cleft running between the two lobes of the enzyme, which is characteristic of the papain superfamily of cysteine proteases (Fig. 6A). With the exception of the mutated catalytic cysteine at position 25, the expected catalytic machinery, highlighted in pink in Fig. 6B, is present in the area of the substrate-binding cleft. The left-hand lobe of the mature domain (Fig. 6A) is predominantly helical in composition. The second domain contains several elements of $\beta$-sheet. Similar to other members of the papain superfamily of
shape similarity that exists among the family of papain-like cysteine proteases (1).

The prosegment of FheproCL1 Gly$^{25}$ folds in a manner very similar to human cathepsin L, as indicated by the value given for superimposition above, although it does show more divergence than is observed in the mature domain. In general, there is a globular region and an extended C-terminal portion, as illustrated in Fig. 6B, that connects the prosegment to the mature domain. As was described for human procathepsin L, the globular portion of the prosegment is fairly well structured and is comprised of distinct components of helix and β-sheet (50). One notable change in the structure as compared with the human zymogen structure is as follows. In FheproCL1 Gly$^{25}$, a stretch of β-sheet extends from 79P through 84P, which is then followed by a very short helical turn from 85P to 88P. In the human enzyme, this final helical turn is absent, and this segment as well as the remainder of the prodomain is made up of β-strand only. The final visible residues of FheproCL1 Gly$^{25}$, 89P through 96P are β-strand. It should be noted that the two species show the greatest structural divergence from residues 85P to 96P, with the chains carving a somewhat different path through three-dimensional space. The final four residues (97P, 98P, 99P, and 100P) of the prosegment of FheproCL1 Gly$^{25}$ are not visible in experimental electron density, which suggests that they are disordered and subject to motion within the crystal. Most of the prosegment of FheproCL1 Gly$^{25}$ sits adjacent to one side of the mature domain, in the region of a loop that extends from approximately residues 138–155 of the mature domain. The corresponding area of contact in the prosegment is residues 55P through 68P. The extended C-terminal tether of the prosegment that links the two domains lies across the active site cleft of the mature domain (Fig. 6B).

**Significant Differences Exist in the Active Site Clefs of FheCL1 and FheCL2**—The composition of the active site cleft, particularly the deep and well defined S2 pocket, is a key determinant of the substrate specificity of the papain family of cysteine proteases (44). The ability to accept or exclude particular substrate moieties is highly dependent upon the size, shape, and volume of the available pocket, as well as the presence or absence of stabilizing interactions such as charge-charge pairs, hydrogen bonding, and hydrophobic interactions. The S2 pocket in FheproCL1 Gly$^{25}$ is lined with several residues that extend into the active site space. These include Leu$^{67}$, Met$^{68}$, Ala$^{131}$, Val$^{157}$, Ala$^{160}$, and Leu$^{205}$ (Table 2 and Fig. 7A). Leu$^{67}$ and Val$^{157}$ are situated at the entrance to the pocket and act as “gatekeepers.” Met$^{68}$, Ala$^{131}$, and Ala$^{160}$ sit below them, deeper into the pocket, whereas Leu$^{205}$ lines the floor of the pocket (illustrated in Fig. 7B). Sequence alignment and homology-based modeling of FheCL2 place the following residues within the S2 pocket: Tyr$^{67}$, Met$^{68}$, Ala$^{131}$, Leu$^{157}$, Ala$^{160}$, and Leu$^{205}$ (Table 3), all at locations corresponding to those observed in the structure of FheCL1 Gly$^{25}$. The differences between the S2 pockets in these similar enzymes include the presence of the dramatically larger Tyr in the “gatekeeping” position 67 at the opening of the S2 pocket, and the somewhat larger Leu$^{157}$ in the opposing position at the entrance to the pocket. Although tyrosine is much larger and bulkier than leucine, it is conformationally able to rotate somewhat freely based on the availability of an unre-
strained torsion angle about the Co–Cβ bond, and its presence at the top of the pocket does not necessary preclude the entry of P2 substrate residues (Fig. 7).

In one of the variants of FheCL1 constructed for this study (FheCL1 L67Y), a substitution of Tyr was made for Leu67 at the opening of the pocket, which renders the entrance to the S2 pocket somewhat more similar to that found in FheCatL2 and human cathepsin K (Table 2). Human cathepsin K is similar to the model constructed for FheCL2 sharing the presence of a larger Tyr residue at the entrance to the pocket. In this structure of human cathepsin K (PDB code 1ATK), the Tyr residue does not preclude access to the pocket and is positioned such that an inhibitor (E-64) is able to bind with a P2 Leu-like moiety just within the top of the S2 pocket. In the second FheCL1 variant constructed (FheCL1 L205A), a substitution of Ala was made in the Leu205 position at the base of the pocket that changed this site to be similar to that found in human cathepsin L (Table 2). As mentioned above the overall structure of the human enzyme is very similar to FheCL1 (50).

DISCUSSION

Substrate Specificity of FheCL1 and FheCL2—A comparison of the substrate specificity between the F. hepatica cathepsin L peptidases (wild type and variants) and human cathepsin L and cathepsin K is shown in Fig. 8, and helps to summarize the findings of our substrate specificity analyses and the effect active site substitutions have on this. First, it is clear that both FheCL1 and FheCL2 are similar to cathepsin K with regard to their preference for a P2 Leu over Phe. Second, both enzymes can accommodate Pro in the P2 position, but this is more readily accepted by FheCL2 compared with FheCL1; neither enzyme, however, cleaves substrates with this residue in the P2 position as readily as human cathepsin K. Third, substituting Tyr for Leu at residue 67 (variant FheCL1 L67Y) to make the S2 subsite of FheCL1 more like that of human cathepsin K did not significantly enhance its ability to cleave substrates with Pro in the P2 position; this was confirmed using three fluorogenic substrates as shown in Table, and by PS-SCL as shown in Fig. 4. Finally, substitution of Leu205 with Ala (variant FheCL1 L205A) increased the relative activity of the peptidase for substrates with Phe in the P2 position, but this increase was not sufficiently dramatic as to reverse its preference for Leu over Phe as observed for human cathepsin L; thus, compared with wild type FheCL1, FheCL1 L205A is more similar to human cathepsin L but is not identical in its substrate specificity.

Our results using inhibitors are consistent with our data derived from the PS-SCL and substrate specificity studies. FheCL1 accommodates hydrophobic P2 residues of diazomethyl ketone Z-Phe-Ala-CHN₂ more effectively than FheCL2, and therefore its inhibition by this reagent was 2-fold greater.
Replacement of the Leu205 by Ala, however, created an S2 pocket that accepted the P2 Phe more readily, and hence the inhibitory constant for Z-Phe-Ala-CHN$_2$ against the FheCL1 L205A variant was 24-fold greater than for the wild type enzyme. The cathepsin K inhibitor II, on the other hand, was 20 times more potent than Z-Phe-Ala-CHN$_2$ against both FheCL1 and FheCL2 and exhibited similar kinetics to that reported for human cathepsin K by Wang et al. (49). By contrast, it was 7-fold less potent against the FheCL1 L205A variant, which is interesting because it is 53-fold less effective against human cathepsin K, which possesses an Ala at position 205, compared with cathepsin K (49). Similar to our observations for substrate binding, the replacement of Leu$^{67}$ for Tyr did not have a dramatic effect on the binding of both inhibitors.

Differences in the substrate specificities of human cathepsin L and cathepsin K can be exquisitely demonstrated using collagen type 1 as a substrate. Because of its acceptance of a P2 proline, cathepsin K can completely degrade collagen by cleaving within the repeated Gly-Pro-Xaa motif in the helices of the tightly wound triple helical structure. Human cathepsin L, on the other hand, cleaves within the nonhelical telomeric regions but does not possess intrahelical activity (24, 44). Although both FheCL1 and FheCL2 could cleave native collagen, only FheCL2 cleaved this substrate within the helical structures. Most strikingly, FheCL2 cleaved native collagen even at neutral pH, which suggests that the enzyme could perform this function in vivo to facilitate parasite tissue migration. Collectively, these results support the idea that the ability of FheCL2 to accommodate proline in the P2 position of substrates confers the enzyme with collagenase-like activity, similar to that observed for cathepsin K. Although FheCL1 exhibited low activity against fluorogenic substrates with a P2 Pro, this was insufficient to endow this enzyme the ability to cleave the helices within native collagen. Replacement of Leu$^{67}$ in FheCL1 with Tyr (FheCL1 L67Y) to make the S2 subsite of this enzyme similar to FheCL2 and cathepsin K did not enhance its ability to accept substrates with Pro residues in the P2 position, nor did it confer the enzyme with collagenase-like activity suggesting that other S2 residue(s) besides that at position 67 are essential for this activity (see below).

The Amino Acid at Position 205 Lies at the Bottom of the S2 Pocket and Has a Major Impact on Substrate Specificity—Our PS-SCL, substrate binding, and inhibitor studies showed that the Leu at position 205 is indeed a determinant of substrate turnover and inhibitor specificity in FheCL1. The wild type enzyme is able to cleave substrates with Phe in the P2 position; however, substrates with the much smaller Leu moiety in the P2 position were cleaved more than 8-fold more rapidly. In comparison with the wild type FheCL1, replacing the Leu with an Ala (FheCL1 L205A) not only enhanced the ability of the enzyme to accept P2 Phe much more readily, but broadened the overall substrate specificity of the enzyme such that larger P2 residues such as Tyr and Trp were also accepted (see Table 3 and Fig. 4).

It has been observed in another papain family member, cruzain, from the protozoan parasite Trypanosoma cruzi, that the character (i.e. size, charge, and torsion-based flexibility) of the residue Glu$^{205}$ is crucial for determining which P2 residues can be accommodated in the S2 pocket (51). Examination of structures of cruzain bound to inhibitors containing Phe at P2 shows that this residue is flexible and can rotate in or out of the pocket depending upon the substrate residue entering the pocket (51). In several inhibitor-bound structures of cruzain where the P2 residue is a Phe, Glu$^{205}$ is swung out into the solvent to accommodate the size of the phenylalanine (52). In FheCL1, however, the Leu at position 205 is shorter by one carbon-carbon bond than Glu and therefore cannot rotate its C-$\beta$ and C-$\delta$ out of the pocket. Although the Leu residue of FheCL1 is shorter than cruzain’s Glu by ~1.5 Å, it is conformationally flexible. It can therefore position itself to make the most space possible available to an incoming P2 Phe. Nonetheless, it clearly prefers the smaller Leu. The side chain of the wild type Leu, as determined in the x-ray crystal structure of FheproCL1 Gly$^{25}$, extends 3.98 Å from the bottom of the pocket, filling a volume of 166.7 Å$^3$ (53) and exposing a surface area of 170 Å$^2$ (54). The observed broader specificity of FheCL1 L205A, i.e. the ability to minimally accommodate Tyr and Trp, can thus be understood by comparing the space available in the base of the S2 pocket. An Ala variant at 205 would extend only as far as the C-$\beta$ of Leu, or to a distance of 1.51 Å and a corresponding volume of 88.6 Å$^3$ and surface area of 115 Å$^2$, leaving considerable additional space for larger substrate peptide residues to be accommodated.

Ability to Cleave Proline at P2 Is Influenced by Residue 67 and the Surrounding Gatekeeper Residues—By engineering a variant cathepsin K with a Leu replacing the Tyr at position 67, Lecaille et al. (44) demonstrated an important role for Tyr$^{67}$ in determining the P2 Pro activity of cathepsin K and in its unique collagen cleaving activity. In this study, we found that replacing the Leu$^{67}$ of FheCL1 with a Tyr to engineer the S2 pocket of the active site of this enzyme to mimic FheCL2 and human cathepsin K did not significantly alter its S2 subsite specificity and, most particularly, did not enhance the ability of the enzyme to accept a P2 Pro. The three-dimensional structure of FheCL1 was therefore analyzed and compared with cathepsin K to explain these observations and to determine what additional factors within the S2 pocket of FheCL1 may influence the acceptance of a P2 Pro residue (Fig. 7).

The S2 pockets of human cathepsin K and FheCL2 are very similar, as evidenced by our modeling data and that presented in Table 2; however, there are some noteworthy differences within a 5 Å radius of this site. First, residue 133 is an alanine in the FheCL2 enzyme, but is a serine in human cathepsin K. The impact of this difference on substrate preference may be minimal, however, because this residue is more peripheral to the outer edge of the base of the S2 pocket (Fig. 7). More significant is residue 158, which is adjacent to the gatekeeping residue 157 and sits just above the upper lip of the pocket. Based on its position, this residue, which is Asn in cathepsin K and Thr in FheCL2, appears to have some secondary influence on accessibility of the opening of the pocket. In the published structure of human cathepsin K (PDB code 1ATK), Asn$^{158}$ is swung out of the way of the entrance to the pocket and does not preclude the entrance of any incoming P2 moiety. However, the Thr$^{158}$ of FheCL2, which is shorter by one carbon than asparagine, cannot move completely out of the way and would possibly have
either a carbon atom or an oxygen atom pointing in toward the top of the S2 opening. Based on spatial constraints, a proline would not be as readily accepted as in human cathepsin K. Our analysis suggests, however, that acceptance of proline in this pocket is not achieved by offering a topology of easy deep penetration access but rather by providing opportunities for stabilizing interactions with the 5-membered proline ring of the substrate at the entrance to the pocket, and that such stabilization involves interactions between the aromatic ring of Tyr^{67} and the P2 proline. The location and positioning of Leu^{157} in the structure of human cathepsin K and FheCL2 suggest its availability to further stabilize the presence of a P2 proline, perhaps with constructive aliphatic interactions.

By comparison, there are greater differences in the gatekeeping positions at the top entrance to the S2 pocket of FheCL1 compared with FheCL2 and cathepsin K (Fig. 7). Residue 67 is the smaller Leu in FheCL1, and its terminal carbons, C-81 and C-82, extend only as far as the corresponding C-81 and C-82 of Tyr, which stretches its terminal oxygen nearly 3.7 Å further from the protein main chain along the edge of the pocket. On the opposing side of the pocket entrance, residues 158 of FheCL1 is Asn^{159}, as in human cathepsin K, and is swung away from the entrance to the pocket, offering unimpeded access. However, residue 157 is a Val in FheCL1 and is one carbon shorter than the Leu found in FheCL2 and human cathepsin K, and accordingly its extension into the pocket is ~1.5 Å less; this does not allow it to extend far enough into the available space to participate in aliphatic interactions. Thus, the absence of both stabilizing Tyr and Leu residues would account for the reduced preference for P2 proline by FheCL1. On the other hand, the composition of the S2 pocket in FheCL1, being more open and accessible to deeper penetration, would more readily favor processing of longer amino acid moieties, such as Leu and Phe, as we have observed.

In summary, previous studies with human cathepsins K and L have shown that residues at positions 67 and 205 are essential in dictating substrate specificity (44, 55). Much attention has been given to the importance of a Tyr^{67} in conferring cathepsin K with the ability of accepting P2 Pro residues in the corresponding S2 subsite of the enzyme and the capacity to degrade native collagens. However, this study using FheCL1 and a recent study by Lecaill et al. (56) using human cathepsin L show that mutations that replace the Leu^{67} to Tyr^{67} in these enzymes are not sufficient alone to accommodate proline and thus to endow collagenolytic activity. Therefore, other residues at the opening of the active site pocket, namely the gatekeeper residues identified here that occupy sites 157 and 158, combine with Tyr to generate these specialized properties, and hence, we have set the groundwork for future mutational studies. It is important to note that glycosaminoglycans such as chondroitin sulfate are known to enhance the collagenolytic activity of human cathepsin K by binding to a site other than the active site (57). However, these do not influence the activity of FheCL2 (data not shown), which points to further intriguing differences between the parasite and mammalian enzymes.

Given that collagen is a major interstitial matrix protein that is highly resistant to proteolysis, our data showing that FheCL2 can degrade native collagen within the helical regions at physiological pH would suggest that this protease enabled Fasciola spp. to become proficient tissue-degrading pathogens. It is important to note that native collagenase-like activity is restricted to very few enzymes. These include the bacterial collagenases, matrix-metalloproteinases, and cathepsin K (24), and therefore the evolution and maintenance of such an activity in Fasciola are significant. By extension, the emergence of this enzyme group may have been essential to the adaptation of the parasites to the wide variety of mammalian species it infects (13).

Acknowledgments—We thank Dr. J. Holton for technical assistance with x-ray data collection and processing and Dr. T. Stout for assistance with structure refinement and analysis. Portions of this research were carried out at the Advanced Light Source, a national user facility supported by the Director, Office of Science, Office of Basic Energy Sciences, of the United States Department of Energy under Contract DE-AC02-05CH11231. Work at The Sandler Center was supported by National Institutes of Health Grant AI-053247.

REFERENCES

1. Rawlings, N. D., Morton, F. R., and Barrett, A. J. (2006) Nucleic Acids Res. 34, D270–D272
2. Tort, J., Brindley, P. J., Knox, D., Wolfe, K. H., and Dalton, J. P. (1999) Adv. Parasitol. 43, 161–266
3. Sajid, M., and Mckerrrow, J. H. (2002) Mol. Biochem. Parasitol. 120, 1–21
4. Dalton, J. P., Caffrey, C. R., Sajid, M., Stack, C., Donnelly, S., Loukas, A., Don, T., McKerrow, J., Halton, D. W., and Brindley, P. J. (2006) in Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology, and Physiology (Maule, A. G., and Marks, N. J., eds) pp. 1–36, CAB International, Wallingford, Oxon, UK
5. Wasiilewski, M. M., Lim, K. C., Philips, J., and McKerrow, J. H. (1996) Mol. Biochem. Parasitol. 81, 179–189
6. Dalton, J. P., O’Neill, S. M., Stack, C., Collins, P., Walsh, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyde, D., Khazadji, E., Moire, N., Brennan, G., Mousley, A., Kreshchenko, N., Maule, A. G., and Donnelly, S. (2003) Int. Parasitol. 33, 1173–1181
7. Abdulla, M. H., Lim, K. C., Sajid, M., McKerrow, J. H., and Caffrey, C. R. (2007) Plos Med. 4, e14
8. Mitchell, G. B., Maris, L., and Bonnivell, M. A. (1998) Vet. Rec. 143, 399
9. Bergstoe, F. H., Moll, L., Vellema, P., and Gaasenbeek, C. P. (2005) Vet. Rec. 156, 350–351
10. Mas-Cona, S., Bargues, M. D., and Valero, M. A. (2005) Int. J. Parasitol. 35, 1255–1278
11. MacManus, D. P., and Dalton, J. D. (2006) Parasitology 133, S43–S61
12. Collins, P. R., Stack, C. M., O’Neill, S. M., Doyle, S., Ryan, T., Brennan, G. P., Mousley, A., Stewart, M., Maule, A. G., Dalton, J. P., and Donnelly, S. (2004) J. Biol. Chem. 279, 17038–17046
13. Irving, J. A., Spithill, T. W., Pike, R. N., Whisstock, J. C., and Smooker, P. M. (2003) J. Mol. Evol. 57, 1–15
14. Beckham, S. A., Law, R. H., Smooker, P. M., Quinsey, N. S., Caffrey, C. R., McKerrow, J. H., Pike, R. N., and Spithill, T. W. (2006) Biol. Chem. 387, 1053–1061
15. Hanna, R. E., and Trudgett, A. G. (1983) Parasite Immunol. 5, 409–425
16. Berasain, P., Goni, F., McGonigle, S., Dowd, A., Dalton, J. P., Frangione, B., and Carmona, C. (1997) J. Parasitol. 83, 1–5
17. Berasain, P., Carmona, C., Frangione, B., Dalton, J. P., and Goni, F. (2000) Exp. Parasitol. 94, 99–110
18. Dowd, A. J., Smith, A. M., McGonigle, S., and Dalton, J. P. (1994) Eur. J. Biochem. 223, 91–98
19. Dowd, A. J., McGonigle, S., and Dalton, J. P. (1995) Eur. J. Biochem. 232, 241–260
20. Dowd, A. J., Tort, J., Roche, L., Ryan, T., and Dalton, J. P. (1997) Mol.
F. hepatica Cathepsin L Structure-Activity Relationships

Biochem. Parasitol. 88, 163–174
21. Ishidoh, K., and Kominami, E. (1998) Biol. Chem. 379, 131–135
22. Drake, F. H., Dodds, R. A., James, I. E., Connor, J. R., Debouck, C., Richardson, S., Lee-Rykaczewski, E., Coleman, L., Rieman, D., Barthlow, R., Hastings, G., and Gowen, M. (1996) J. Biol. Chem. 271, 12511–12516
23. Buhling, F., Reisenauer, A., Gerber, A., Kruger, S., Weber, E., Brömme, D., Roessner, A., Ansgorre, S., Welte, T., and Rocken, C. (2001) J. Pathol. 195, 375–382
24. Atley, L. M., Mort, J. S., Lalumiere, M., and Eyre, D. R. (2000) Bone (Elmsford) 26, 241–247
25. Roche, L., Dowd, A. J., Tort, J., McGonigle, S., MacSweeney, A., Curley, G. P., Ryan, T., and Dalton, J. P. (1997) Eur. J. Biochem. 232, 241–246
26. Stack, C. M., Donnelly, S., Lowther, J., Xu, W., Collins, P. R., Brinen, L. S., and Dalton, J. P. (2007) J. Biol. Chem. 282, 16532–16543
27. Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Brömme, D., Ellman, J. A., and Craik, C. S. (2006) J. Struct. Biol. 154, 418–429
28. Morrison, J. F., and Walsh, C. T. (1988) Adv. Enzymol. 61, 201–301
29. Tian, W. X., and Tsou, C. L. (1982) J. Biol. Chem. 257, 241–247
30. Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K., and DeTitta, G. T. (1982) Biochemistry 21, 1028–1032
31. McEwen, K., Earnest, T. N., Dickert, J., Meigs, G., Ralston, C., Holton, J. M., Cambie, D., Domning, E. E., Duarte, R. M., Kelez, N., Plate, D. W., Cork, C. W., DeTitta, G. T., Cork, C. W., Earnest, T. N., Dickert, J., Meigs, G., Ralston, C., Holton, J. M., Alber, T., Berger, J. M., Agard, D. A., and Padmore, H. A. (2004) J. Synchrotron. Radiat. 11, 447–455
32. Leslie, A. G. (2006) Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 48–57
33. Collaborative Computational Project, No. 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
34. Lamzin, V. S., Perrakis, A., and Wilson, K. S. (2001) in Crystallography of Biological Macromolecules (Rossman, F. M. G., and Arnold, E., eds) pp. 720–722, Kluwer Academic Publishing, Dordrecht, Netherlands
35. Murshudov, G. N., Vagin, A., and Dodson, E. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255
36. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 448–455
37. van den Bedem, H., Lotan, I. J., Latombe, C. L., and Deacon, A. M. (2005) Acta Crystallogr. Sect. D Biol. Crystallogr. 61, 2–13
38. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132
39. Marti-Renom, M. A., Stuart, A., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Annu. Rev. Biophys. Biomol. Struct. 29, 291–325
40. Sali, A., and Blundell, T. L. (1993) J. Mol. Biol. 234, 779–815
41. Fiser, A., Do, R. K., and Sali, A. (2000) Protein Sci. 9, 1753–1773
42. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, Version 1.0, Delano Scientific, San Carlos, CA
43. Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001) Bioinformatics (Oxf.) 17, 1244–1245
44. Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S., and Brömme, D. (2002) Biochemistry 41, 8447–8454
45. Smooker, P. M., Whisstock, J. C., Irving, J. A., Siyaguna, S., Spithill, T. W., and Pike, R. N. (2000) Protein Sci. 9, 2567–2572
46. Green, D. G., and Shaw, E. (1981) J. Biol. Chem. 256, 1923–1928
47. Kirschke, H., and Shaw, E. (1981) Biochem. Biophys. Res. Commun. 101, 454–480
48. Kirschke, H., Wikstrom, P., and Shaw, E. (1988) FEBS Lett. 228, 128–130
49. Lecaille, F., Chowdhury, S., Purisima, E., Brömme, D., and Lalmich, S. (2007) Protein Sci. 16, 662–670