JBC Papers in Press. Published on December 26, 2007 as Manuscript M708521200
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M708521200

STRUCTURAL AND FUNCTIONAL RELATIONSHIPS IN THE VIRULENCE-ASSOCIATED CATHEPSIN L PROTEASES OF THE PARASITIC LIVER FLUKE, *FASCIOLA HEPATICA*.

Colin M. Stack[1,2], Conor R. Caffrey[3], Sheila M. Donnelly[1], Amritha Seshaadri[3,] Jonathan Lowther[1], Jose F. Tort[1,4], Peter R. Collins[1,5], Mark W. Robinson[1], Weibo Xu[1], James H. McKerrow[3], Charles S. Craik[6], Sebastian R. Geiger[7,8], Rachel Marion[7], Linda S. Brinen[7], John P. Dalton[1*]

[1]Institute for the Biotechnology of Infectious Diseases (IBID), University of Technology Sydney (UTS), Level 6, Building 4, Corner of Thomas & Harris Street, Ultimo, Sydney, NSW 2007, Australia. [2] presently at, Department of Medical Microbiology, University of Western Sydney (UWS), Narellan Road, Campbelltown, NSW, Australia.[3]Sandler Center for Basic Research in Parasitic Diseases, Byers Hall N508, 1700 4th Street, University of California San Francisco, San Francisco, CA 94158, USA. [4]Departamento de Genetica, Facultad de Medicina, UDELAR , Gral. Flores 2125, CP 11800, Montevideo, Uruguay [5] presently at, the School of Biotechnology, Dublin City University, Dublin 9, Republic of Ireland. [6]Departments of Pharmaceutical Chemistry, Pharmacology, and Biochemistry & Biophysics, University of California San Francisco, San Francisco, CA 94158, USA, [7]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA. [8]presently at, Gene Center Munich, Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universitat Munchen, Feodor-Lynen-Strasse 25, D-81377 Munich, Germany.

Running Title: *F. hepatica* Cathepsin L structure-activity relationships

Address correspondence to Prof. John P. Dalton, Institute for the Biotechnology of Infectious Diseases (IBID), University of Technology Sydney (UTS), Level 6, Building 4, Corner of Thomas & Harris Street, Ultimo, Sydney, NSW 2007, Australia. Ph: +61 2 9514 4142; Fax: +61 2 9514 4201; Email: john.dalton@uts.edu.au.

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 recombinantly produced and biochemically characterised. 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 to 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 Å three-dimensional structure of the FheCL1 has was determined by X-ray crystallography and the three-dimensional structure of FheCL2 constructed *via* homology-based modelling. 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 specialised 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; www.merops.sanger.ac.uk) 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 defences by cleaving immunoglobulin or altering the activity of immune effector cells (2, 3, 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, 6, 7).

Fasciola hepatica is a helminth parasite that causes liver fluke disease (fasciolosis) in cattle and sheep world-wide. It is most prevalent in Europe with infection rates increasing due to 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 organised 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, 13, 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 purified native enzymes demonstrated that FheCL1 and FheCL2 efficiently degrade host haemoglobin, immunoglobulin and interstitial matrix proteins such as fibronectin, laminin and native collagen (6, 16, 17). While FheCL1 and FheCL2 exhibited similar substrate specificities FheCL2 showed a greater affinity for peptides containing Pro residues in the P2 position (18, 19, 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 house-keeping function in protein turnover, but also plays a part in more specialised functions such as antigen processing and presentation, hormone and protease activation and extra-cellular 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 protease's ability to cleave the covalently-linked triple helices of native collagen, a unique property amongst 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* FheCL2 as the latter can also cleave substrates with a P2 proline and possesses a tyrosine residue at the corresponding position 67.

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. In the present study, therefore, we have characterised 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 the active site (positions 67 and 205) in order to simulate those residues present in human cathepsins L and K. In addition, the 1.4 Å three-dimensional structure of a variant FheCL1 zymogen, in which the active site Cys was replaced by a Gly (FheproCL1Gly^{25}), has been determined by X-ray crystallography. For FheCL2, the three-dimensional structure has been constructed *via* homology-based modelling. Analysis and comparison of these major parasite proteases

with the human cathepsins L and K provide a structural interpretation of the substrate-recognition mechanisms.

EXPERIMENTAL PROCEDURES

Materials

[1]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-CHN_2 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-Aldrich (Sydney, Australia). Cathepsin K Inhibitor II was purchased from Biosciences (CA, USA). Prestained molecular weight markers and the *Avr*II and *Sna*BI restriction enzymes were obtained from New England Biolabs (UK) Ltd. (Hitchin, UK). Primers were obtained from Sigma-Genosys (Pampisford, UK). The pPIC9K vector and *Pichia pastoris* strain GS115 were obtained from Invitrogen Corp. (San Diego, CA, USA). Ni-NTA agarose and columns were obtained from Qiagen (Crawley, UK). Collagen, calf skin, was purchased from Calbiochem (San Diego, CA). Pre-cast 4-20% gradient SDS-PAGE gels were purchased from Gradipore (Australia).

Expression and purification of recombinant cathepsin L zymogens in yeast

F. hepatica procathepsin L1 (FheCL1) and procathepsin L2 (FheCL2) were amplified by PCR from the pAAH5 *Saccharomyces cerevisiae* expression vector into which the full-length cDNA had been previously cloned in our laboratory (12, 25). FheCL1 variants (FheCL1 Leu67Tyr and FheCL1 Leu205Ala) were synthesized and incorporated a *Sna*BI restriction site at the 5' end of the gene and an *Avr*II restriction site and His_6-tag sequence at the 3' end (Geneart, Regensburg, Germany). The 980bp fragments were ligated into pCR-Script cloning vector (Stratagene, CA, USA) which were transformed into competent *E. coli* for amplification. Inserts were digested from plasmid preparations with *Avr*II and *Sna*B1 and inserted in-frame with the yeast alpha-factor at the *Avr*II/*Sna*B1 site of *Pichia pastoris* expression vector pPIC9K (Invitrogen). Plasmids were linearized with *Sac*I and then transformed into chemically competent GS115 cells (Invitrogen) as described previously (12). All inserts were sequenced to ensure congruence with original cDNAs.

Pichia pastoris yeast transformants were cultured in 500 ml BMGY broth, buffered to pH 8.0, in 5 L baffled flasks at 30°C until an OD_{600} of 2-6 was reached (12). Cells were harvested by centrifugation at 2000 x g for 5 min and protein expression induced by resuspending in 100 ml BMMY broth, buffered at pH 6.0 containing 1% methanol (Dowd *et al*, 1997). Recombinant proteins were purified from yeast medium by affinity chromatography using Ni-NTA-agarose (Qiagen) (12, 26). Purified recombinant zymogens were dialysed against phosphate buffered saline (PBS) and stored at –20°C. The 37 kDa zymogens were auto-catalytically activated and processed to 24.5 kDa mature enzymes by incubation for 2 hours at 37ºC in 0.1 M sodium citrate buffer pH 5.0, containing 2 mM DTT and 2.5 mM EDTA. The mixture was then dialysed 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 Leu67Tyr and FheCL1 Leu205Ala and FheCL2 were determined using a complete diverse positional-scanning synthetic combinatorial library (PS-SCL; 27). Screens were performed at 25ºC in 0.1 M sodium acetate, 0.1 M NaCl, 0.01 M DTT, 0.001 M EDTA, 0.01% Brij-35, 1% DMSO (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 cut-off at 435 nm. Screens were performed in duplicate and triplicate for wildtype and mutated enzymes, respectively.

Enzyme assays and kinetics with fluorogenic peptide substrates

Initial rates of hydrolysis of the fluorogenic dipeptide 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 Bio-Tek 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 (0.2 – 200 μM) and at fixed enzyme concentrations (0.5 – 5 nM). Assays were performed in PBS, pH 7.3, and 100 mM sodium acetate buffer, pH 5.5, each containing 2.5 mM DTT and 2.5 mM EDTA.

Rate constants for the inactivation of enzyme by Z-Phe-Ala-CHN_2 and Cathepsin K Inhibitor II were determined from progress curves in the presence of substrate (28, 29). When substrate and inhibitor bind to enzyme in rapid equilibrium and the substrate concentration does not change significantly during the course of the assay, the concentration of product, [P], at time *t* after the start of the reaction is given by the equation

$$[P] = \frac{v_0}{k_{obs}}\left[1 - \exp(-k_{obs}t)\right] + A_0 \qquad (1)$$

where v_0 is the initial rate of reaction, k_{obs} is the rate of inactivation and A_0 is the background fluorescence. k_{obs} is related to the inhibitor concentration by the equation

$$k_{obs} = \frac{k_{inact}[I]}{[I] + K_i\left(1 + \frac{[S]}{K_m}\right)} \qquad (2)$$

When $[I] << K_i$ plots of k_{obs} versus [I] were linear with slope equal to an apparent second-order rate constant $k_{obs}/[I]$. This value was then corrected for substrate concentration and the Michaelis constant to determine a true second-order rate constant k_{inact}/K_i.

The initial rate v_0 is related to inhibitor concentration by the equation

$$v_0 = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_{i(app)}}\right) + [S]} \qquad (3)$$

Since the inactivation was carried out with $[S]=K_m$, equation (3) reduces to

$$v_0 = \frac{V_{max}}{2 + \frac{[I]}{K_{i(app)}}} \qquad (4)$$

An apparent inhibition constant $K_{i(app)}$ for the formation of the initial reversible enzyme-inhibitor complex prior to inactivation was determined by plotting v_0 against [I] and fitting to equation (4).

Collagen digestion

Calf skin collagen type-1 was solublized in 0.2 M acetic acid at a concentration of 2 mg/ml and dialysed for two 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 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.

Production of inactive variant FheproCL1 Gly^{25}

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 Gly^{25} enzyme migrated as a single protein of 37 kDa on reducing 12% SDS-PAGE which represents the full zymogen containing a prosegment and mature enzyme domain (data not shown).

Data collection, structure solution and crystallographic refinement of FheproCL1 Gly^{25}

Initial crystallization screening experiments were performed at the Hauptman-Woodward Institute high-throughput crystallization lab. A total of 1536 conditions were tested using a nano-scale

microbatch-under-oil method, resulting in several preliminary hits that suggested a route to diffraction-quality crystals (30). Ultimately, high-quality crystals were grown in house *via* vapor diffusion in sitting drops. One µL of 10 mg/ml FheproCL1 Gly^{25} enzyme was mixed with one µL of the precipitating agent, 0.2 M NaSCN in 20% PEG 3350, and allowed to equilibrate at 23°C over a 100 µL reservoir of precipitating agent. Crystalline plates formed within two days, however, full-size growth to plates greater than 75 µm in thickness took nearly two months time.

Diffraction data were collected at the Advanced Light Source (ALS), beam line 8.3.1, using monochromatic (Si 111) radiation of 1.11588 Å (31). An ADSC Quantum 210 2x2 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 approximately 1.5 minutes in a cryo-protectant solution of crystal-growth solution plus 50% MPD. High- and low-resolution datasets 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 sigma 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 R_{factor} of 0.526. The initial molecular replacement solution was improved using ARP/wARP as implemented in the CCP4 program suite (34) resulting in a model that was better than 85% complete. Iterative rounds of visualization and manual model building and refinement were completed with QUANTA (Accelrys, San Diego, CA) and Refmac5 with anisotropic atomic displacement parameters (35), respectively. Water molecules were added automatically using ARPwaters in CCP4 (36) and were manually verified. In the final stages of refinement, XPLEO (37) was used to improve the fit of two areas of ambiguous density in the structure. Final visualization and manual adjustments to the structure as well as final assessment of water molecules were completed with COOT (38). Crystallographic parameters and statistics are summarized in Table 1 and final atomic coordinates have been deposited with the Protein Data Bank, accession ID 2O6X.

Homology-based molecular modelling

A model structure of the mature domain of FheCL2 was built using Modeller (release 8v1), a program for protein structure modelling (39, 40, 41). The 1.8 Å structure of human procathepsin L (PDB ID 1CS8), the 2.2 Å structure of human cathepsin K (PDB ID 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 ClustalX 1.81. Phylogenetic trees were generated from the alignment by the boot-strapped (1000-trial) neighbour-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 ID: 1CS8) and cathepsin K (PDB ID: 1ATK) are shown in Table 2 (see also Fig. 1 and 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 favours both aromatic and aliphatic P2 residues but will not accept proline, whereas cathepsin K prefers only

aliphatic resides and most particularly proline. Indeed, the acceptance of a P2 Pro residue confers cathepsin K with it 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 Leu67Tyr variant has a single amino acid change making the S2 subsite similar to FheCL2 and cathepsin K at positions 67 and 205. The FheCL1 Leu205Ala variant has a single amino acid that was designed to make the S2 subsite similar to human cathepsin L. The wildtype and variant *F. hepatica* cathepsin L peptidases were recombinantly-expressed in the methylotrophic yeast *Pichia pastoris*, purified and activated as described in materials and methods. 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 wildtype 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 ($pK_I = 3.87 \pm 0.07$ and pK_{II} 8.14 ± 0.08).

Substrate specificity profiling using a PS-SCL reveals unique and distinct activities of FheCL1 and FheCL2.

Wildtype 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 column), 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 Leu67Tyr and FheCL1 Leu205Ala, 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 favoured 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 to FheCL1 that did not accommodate this residue (Fig. 3, P2 column). 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 favoured 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 Leu205Ala) markedly altered the activity profile from wildtype enzyme (Fig. 4). This variant exhibited a broader substrate specificity by accepting Phe, Trp and Tyr at P2, residues that were not accepted by wildtype 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 Leu67Tyr variant did not show a significant change in the P2 preference to wildtype FheCL1; in particular, this substitution did not alter the activity of the enzyme towards Pro in the P2 position (Fig. 4). This was a

surprising result as we expected that the FheCL1 Leu67Tyr 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).

Wildtype 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 wildtype FheCL1, the variants FheCL1 Leu67Tyr, and FheCL1 Leu205Ala and wildtype FheCL2, we examined their hydrolytic activity against various fluorogenic di- and tri-peptides (Table 3). FheCL1 efficiently cleaved both Z-Phe-Arg-NHMec (k_{cat}/K_M = 1,021,092 $M^{-1}s^{-1}$) and Z-Leu-Arg-NHMec (k_{cat}/K_M = 8,395,402 $M^{-1}s^{-1}$); the enzyme cleaved the latter substrate over eight times more rapidly largely because its K_M for this substrate is much lower than for the former substrate. Although the substrates Z-Pro-Arg-NHMec (k_{cat}/K_M = 5,387 $M^{-1}s^{-1}$), Tos-Gly-Pro-Arg-NHMec (k_{cat}/K_M = 35,928 $M^{-1}s^{-1}$) and Boc-Ala-Gly-Pro-Arg-NHMec (k_{cat}/K_M = 46,021 $M^{-1}s^{-1}$) were cleaved relatively poorly the data, nevertheless, indicates that FheCL1 can accommodate proline residues in the P2 position.

In comparison to FheCL1, FheCL2 is much less efficient at cleaving substrates with Phe and Leu in the P2 position; the k_{cat}/K_M values for Z-Phe-Arg-NHMec and Z-Leu-Arg-NHMec with this enzyme are 24-fold 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). This kinetic data shows 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 Leu205Ala had a significant impact on the enzyme's substrate specificity 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-ArgNHMec and Boc-Ala-Gly-Pro-Arg-NHMec (Table 3). Substitution at position 67 to give the variant FheCL1 Leu67Tyr 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_{cat} and K_M values for each substrate. This substitution did not significantly alter the enzyme's specificity 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 wildtype and variant proteases with specific inhibitors

Peptidyl diazomethylketones 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 Leu205Ala, FheCL1 Leu67Tyr and FheCL2 by the cathepsin inhibitor Z-Phe-Ala-CHN_2 have been measured. Wildtype FheCL1 and FheCL2 had second-order rate constants of 20,838 $M^{-1}s^{-1}$ and 11,899 $M^{-1}s^{-1}$, respectively, showing that both enzymes were rapidly inactivated by Z-Phe-Ala-CHN_2 (Table 4). The two-fold greater rate of inactivation of FheCL1 compared to FheCL2 is further evidence that FheCL1 accommodates hydrophobic P2 residues more effectively than FheCL2.

The rate of inactivation of FheCL1 Leu205Ala was 24-fold greater than wildtype indicating that Ala at residue 205 in the S2 subsites binds a P2 Phe more effectively that 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 Leu205Ala 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 Leu67Tyr variant had a $k_{obs}/[I]$ value of 53,704 $M^{-1}s^{-1}$ (Table 4), which is only 2.5-fold higher than wildtype FheCL1 and 5-fold greater than wildtype FheCL2 and, therefore, this substitution has not such a major influence on the binding of Phe in the S2 pocket.

The inhibitor known as Cathepsin K Inhibitor II (Z-L-NHNHCONHNH-LF-Boc, CKII) is a potent time-dependent inhibitor of human cathepsin K; its selectivity for this enzyme is largely due to the effectiveness by which leucine occupies the S2 subsite (49). FheCL1 and FheCL2 were both potently inhibited by cathepsin K Inhibitor II with $k_{obs}/[I]$ values of 397,237 $M^{-1}s^{-1}$ and 269,447 $M^{-1}s^{-1}$, respectively, which are twenty times higher than that observed for the peptidyl diazomethylketone Z-Phe-Ala-CHN_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). This data is consistent with the kinetic data for hydrolysis of peptidyl fluorogenic substrates as both enzymes had highest k_{cat}/K_M values for Z-Leu-Arg-NHMec.

The rate of inactivation of FheCL1 Leu205Ala by cathepsin K Inhibitor II was 7-fold lower than wildtype. The $K_{i(app)}$ increased 11-fold demonstrating that this variant cannot accommodate leucine in the S2 subsite to the same extent as wildtype FheCL1. Since 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 Leu205Ala variant has S2 specificity more characteristic of human cathepsin L. The rate of inactivation of FheCL1Leu67Tyr by cathepsin K Inhibitor II was 3.5-fold lower than wildtype, although the $K_{i(app)}$ did not change significantly, against showing that the Tyr substitution at this position exerts a relatively lower affect on P2 binding.

Wildtype FheCL2 but not wildtype FheCL1 or its variants cleave native collagen type 1

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 β and γ chains while the α1 and α2 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 cleaves efficiently within the helical structures (Fig. 5, Panel A). Since low pH may cause some structural unravelling 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. 5, Panel B).

Like the wildtype FheCL1 enzyme, FheCL1 Leu205Ala and FheCL1 Leu67Tyr variants cleaved collagen but were unable to cleave within the tightly wound helices. While in all experiments FheCL1 Leu205Ala appeared to cleave collagen more efficiently than the wildtype enzyme, the pattern of digested fragments was similar (Fig. 5, Panel B). 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 Leu67Tyr variant to cleave within the helices of collagen is also consistent with the PS-PCL studies and substrate kinetics studies since this enzyme did not show any increase in preference for P2 Pro compared to wildtype FheCL1.

The FheproCL1Gly^{25}structure in comparison to other Clan CA cysteine proteases.

The experimentally determined structure of FheproGL1Gly25 is quite similar to that of previously described mammalian cathepsins. While the X-ray crystal structure of FheCL1 presented here is that of an inactive zymogen mutant in which active site Cys 25 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, FheproCL1Gly^{25}, is similar in tertiary structure to human cathepsin L1. Electron density is clear, connected and easily traceable for the entirety of the mainchain of the mature domain of FheproCL1Gly^{25}. 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

super-family of cysteine proteases (Fig. 6A). With the exception of the mutated catalytic cysteine at position 25, the expected catalytic machinery, highlighted in magenta in Fig. 6A, 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 β-sheet. Similar to other members of the papain super-family of enzymes there are three disulfide bonds in the mature domain of FheCL1 Gly25. These connect Cys 22–Cys 63, Cys 56–Cys 95 and Cys 153–Cys 200, respectively. Superimposition of the alpha carbons of the mature domain of FheproCL1 Gly^{25} on those of active papain (PDB entry ID: 9PAP) yields an r.m.s. deviation of 1.085 Å (the primary structure FheCL1 exhibits 32.8% identity and 62% similarity to papain). Superimposition of the main-chain atoms of the mature domain of FheproCL1 Gly^{25} on those of the mature domain of human cathepsin L yields an r.m.s. deviation of 0.780 Å, while superimposition of the two full-length molecules (PDB ID 1CJL) yields 1.115 Å (the primary structure FheCL1 exhibits 42.8% identity and 71% similarity to human cathepsin L). These structural comparisons are indicative of the high degree of overall fold and shape similarity that exists amongst the family of papain-like cysteine proteases (1, www.merops.sanger.ac.uk).

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 – though 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 beta-strand (50). One notable change in the structure as compared to the human zymogen structure is as follows. In FheproCL1Gly^{25}, a stretch of beta strand 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 beta strand only. The final visible residues of FheproCL1Gly^{25}, 89P through 96P are beta 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, 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 to 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 clefts 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 133, 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 133 and Ala 160 sit below them, deeper into the pocket, while Leu 205 lines the floor of the pocket (illustrated in Fig. 7B). Sequence alignment and homology based modelling of FheCL2 places the following residues within the S2 pocket: Tyr 67, Met 68, Ala 133, 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 'gate-keeping' 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. While tyrosine is much larger and bulkier than leucine, it is conformationally able to rotate somewhat freely based on the availability of an unrestrained torsion angle about the C_α-C_β 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 Leu67Tyr), a substitution of Tyr was made for Leu 67 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 ID 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 Leu205Ala), a substitution of Ala was made in the Leu 205 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 (wildtype and variants) and human cathepsin L and cathepsin K is shown in Fig. 8, and helps to summarise the findings of our substrate specificity analyses and the effect active site substitutions have on this. Firstly, it is clear that both FheCL1 and FheCL2 are similar to cathepsin K with regard to their preference for a P2 Leu over Phe. Secondly, both enzymes can accommodate Pro in the P2 position but this is more readily accepted by FheCL2 compared to FheCL1; neither enzyme, however, cleaves substrates with this residue in the P2 position as readily as human cathepsin K. Thirdly, substituting Tyr for Leu at residue 67 (variant FheCL1 Leu67Tyr) 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 3, and by PS-SCL as shown in Fig. 4. Lastly, Substitution of Leu205 with Ala (variant FheCL1 Leu205Ala) 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 cathespin L; thus, compared to wildtype FheCL1, FheCL1 Leu205Ala 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 diazomethylketone Z-Phe-Ala-CHN_2 more effectively than FheCL2 and therefore its inhibition by this reagent was two-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 Leu205Ala variant was 24-fold greater than for the wildtype 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 seven-fold less potent against the FheCL1 Leu205Ala variant which is interesting since it is 53-fold less effective against human cathepsin L, which possesses an Ala at position 205, compared to cathepsin K (49). Similar to our observations for substrate binding the replacement of Leu67 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. Due to 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 non-helical telomeric regions but does not possess intrahelical activity (24, 44). While 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 FheCL2's ability 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 Leu67 in FheCL1 with Tyr (FheCL1 Leu67Tyr) to make this enzyme's S2 subsite similar to FheCL2 and cathepsin K did not enhance it's ability to accept substrates with Pro residues in the P2 position, and nor did it confer the enzyme with collagenase-like activity suggesting that other S2 residue(s) besides that at position 67 is 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 wildtype 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 to the wildtype FheCL1, replacing the Leu with an Ala (FheCL1 Leu205Ala) not only enhanced the enzyme's ability to accept P2 Phe much more readily, but broadened the enzyme's overall substrate specificity 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 in order 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 Cδ1 and Cδ2 out of the pocket. While the Leu residue of FheCL1 is shorter than cruzain's Glu by approximately 1.5 Å, it is conformationally flexible. It can therefore position itself to make the most space possible available to an incoming P2 Phe. None-the-less, it clearly prefers the smaller Leu. The sidechain of the wildtype Leu, as determined in the X-ray crystal structure of FheproCL1Gly[25], extends 3.98Å 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 Leu205Ala, 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 Cß 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 'gate-keeper' 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 Tyr67 in determining the P2 Pro activity of cathepsin K and in its unique collagen-cleaving activity. In the present study, we found that replacing the Leu67 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 enzyme's ability to accept a P2 Pro. The three-dimensional

structure of FheCL1 was, therefore, analysed and compared to cathepsin K to explain these observations and 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 modelling 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, since 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 'gate-keeping' 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 ID 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 towards the top of the S2 opening. Based on spatial constraints, a proline would be accommodated in the S2 area of FheCL2, but this would 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 suggests its availability to further stabilize the presence of a P2 proline, perhaps with constructive aliphatic interactions.

By comparison, there are greater differences in the gate-keeping positions at the top entrance to the S2 pocket of FheCL1 compared to FheCL2 and cathepsin K (Fig. 7). Residue 67 is the smaller Leu in FheCL1 and it's terminal carbons, $C_{\delta}1$ and $C_{\delta}2$, extend only as far as the corresponding $C_{\delta}1$ and $C_{\delta}2$ 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 158, 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 approximately 1.5 Å less; this does not allow it 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 favour 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 Tyr67 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 Lecaille *et al*, (56) using human cathepsin L show that mutations that replace the Leu67 to Tyr67 in these enzymes are not sufficient alone to accommodate proline and, thus, 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 specialised properties, and hence, we have set the ground for future mutational studies. It is important to note that glycosaminoglycans such as chondroitin sulphate 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* is significant. By extension, the emergence of this enzyme group may have been essential to the parasites' adaptation to the wide variety of mammalian species it infects (13).

REFERENCES

1. Rawlings, N. D., Morton, F. R., and Barrett, A. J. (2006) *Nucleic Acids Res* **34**, D270-272
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 Mckerrow, 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) *Parasitic Flatworms: Molecular biology, Biochemistry, Immunology and Physiology*, Eds. A.G. Maule and N.J. Marks, pp1-36. Oxford, United Kingdom: CAB International.
5. Wasilewski, 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., Hoyle, D., Khaznadji, E., Moire, N., Brennan, G., Mousley, A., Kreshchenko, N., Maule, A., and Donnelly, S. (2003) *Int Parasitol* **33**, 1173-1181
7. Abdulla, M. H., Lim, K. C., Sajid, M., McKerrow, J. H., Caffrey, C. R. (2007) *PLoS Med.* 4(1):e14
8. Mitchell, G. B., Maris, L., and Bonniwell, M. A. (1998) *Vet Rec* **143**, 399
9. Borgsteed, F. H., Moll, L., Vellema, P., and Gaasenbeek, C. P. (2005) *Vet Rec* **156**, 350-351
10. Mas-Coma, 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**, S43-61
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**,1053-61
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 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., Ansorge, 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* **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 Biol Chem* **281**, 12824-12832
28. Morrison, J. F., and Walsh, C. T. (1988) *Adv Enzymol* **61**, 201-301
29. Tian, W. X., and Tsou, C. L. (1982) *Biochemistry* **21**, 1028-1032
30. Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K., and DeTitta, G. T. (2003) *J Struct Biol* **142**, 170-179
31. MacDowell, A. A., Celestre, R. S., Howells, M., McKinney, W., Krupnick, J., Cambie, D., Domning, E. E., Duarte, R. M., Kelez, N., Plate, D. W., 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) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48-57
33. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst D* **50**, 760-763.
34. Lamzin, V. S., Perrakis, A., and Wilson, K. S. (2001) *Crystallography of Biological Macromolecules*, Eds. Rossman, F, M.G., and Arnold, E., pp720-722. Dordrecht, Netherlands: Kluwer Academic Publishing
35. Murshudov, G. N., Vagin, A., and Dodson, E. (1997) *Acta Cryst D* **53**, 240-255
36. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) *Acta Cryst D* **53**, 448-455
37. van den Bedem, H., Lotan, I. J., Latombe, C. L., and Deacon, A. M. (2005) *Acta Cryst D* **61**, 2-13
38. Emsley, P., and Cowtan, K. (2004) *Acta Cryst D* **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 (http://www.pymol.org) (Delano Scientific, San Carlos, California)
43. Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001) *Bioinformatics* **17**, 1244-1245
44. Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S., 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, G. D., 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. Wang, D., Pechar, M., Li, W., Kopečková, P., Brömme, D., and Kopeček, J. (2002) *Biochemistry* **41**, 8849-8859.
50. Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J* **15**, 5492-5503
51. Gillmor, S. A., Craik, C. S., and Fletterick, R. J. (1997) *Protein Sci* **6**, 1603-1611

52. Brinen, L. S., Hansell, E., Cheng, J., Roush, W. R., McKerrow, J. H., and Fletterick, R. J. (2000) *Structure* **8**, 831-840
53. Zamyatin, H. (1972) *Prog Biophys Mol Biol* **24**, 107-123
54. Chothia, C. (1973) *Nature* **248**, 338-339
55. Brömme, D., Bonneau, P. R., Lachance, P., and Storer, A. C. (1994) *J Biol Chem* **269**, 30238-30242
56. Lecaille, F., Chowdhury, S., Purisima, E., Brömme, D., and Lalmanach, G. (2007) *Protein Sci* **16**, 662-670
57. Li, Z., Yasuda, Y., Li, W., Bogyo, M., Katz, N., Gordon, R. E., Fields, G. B., and Brömme, D. (1994) *J Biol Chem* **279**, 5470-5479

FOOTNOTES.

* **Acknowledgements**. John P. Dalton is a recipient of the NSW Government BioFirst Award in Biotechnology. Peter Collins was supported by a grant received from Enterprise Ireland and Department of Science, Education and Training (DEST), Australia. Dr. Colin Stack was funded by the Australian Research Council (ARC), Australia. Mark Robinson is holder of a Wain International Travel Scholarship, UK. Research at the Sandler Center is supported by National Institutes of Health Grant AI-053247 and The Sandler Family Supporting Foundation. The authors would like to 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 U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

[1]The abbreviations used are: BMGY, buffered glycerol-complex medium; BMMY, buffered minimal methanol medium; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; E-64, *trans*-Epoxysuccinyl- L-leucylamido(4-guanidino)butane;MYA, million years ago; MDP, 2-methyl-2,4-pentane diol; NaSCN, sodium thiocyanate; PEG, polyethylene glycol; PS-SCL, positional-scanning synthetic combinatorial library; Z-Phe-Arg-NHMec, benzyloxycarbonyl-L-phenylalanyl-L-arginine-4-methylcourmarinyl-7-amide; Z-Phe-Ala-CHN_2, benzyloxycarbonyl-L-phenylalanyl-L-alanine-diazomethylketone.

FIGURE LEGENDS

FIG. 1
Sequence alignment of the mature *Fasciola hepatica* cathepsin L1 (FheCL1) and cathepsin L2 (FheCL2) with human cathepsin L (hCatL), human cathepsin K (hCatK) and papain was performed with ClustalW (EBI, EMBL). Residues within the S2 subsite of the active site involved in determining substrate specificity are indicated with arrows and numbers (see also Table 2).

FIG. 2
Activation of purified recombinant FheproCL1, and FheproCL2. The 37 kDa zymogens were auto-catalytically activated and processed to 24.5 kDa mature enzymes by incubation for 2 hours at 37ºC in 0.1 M sodium citrate buffer pH 5.0, containing 2 mM DTT and 2.5 mM EDTA. Reaction samples were analysed by 4-20% SDS-PAGE; *lane 1 – 4*, activation reaction of FheCL1 at 0, 30, 60 and 120 mins; *lane 5 - 8*, activation reaction of FheCL2 at 0, 30, 60 and 120 mins. Similar results were obtained with variant peptidases (not shown). MW, molecular mass markers.

FIG. 3

Profiling of the P1-P4 substrate specificity of FheCL1 and FheCL2 using positional-scanning synthetic combinatorial libraries. The y axis represents activity against the substrates relative to the highest activity of the library whereas the x axis presents the amino acids as represented by the one-letter code (n = norleucine).

FIG. 4
Comparison of the P2 specificities of recombinant wildtype FheCL1, variant FheCL1 Leu205Ala (L205A), variant FheCL1 Leu67Tyr (L67Y) and wildtype FheCL2 using positional-scanning synthetic combinatorial libraries.

FIG. 5.
Comparison of the collagen-cleaving activities of wildtype FheCL1, variant FheCL1 Leu205Ala, variant FheCL1 Leu67Tyr and wildtype FheCL2. *Panel A*, type I collagen was incubated with FheCL1 and FheCL2 at pH 4.0 and 5.5 and at 28 ºC for 3 hours and the reaction analysed by 4-20% SDS-PAGE; *lane 1*, collagen alone; *lane 2* collagen plus FheCL1 pH 4.0; *lane 3*, collagen plus FheCL2 pH 4.0; *lane 4* collagen alone; *lane 5* collagen plus FheCL1, pH 5.5; *lane 6*, collagen plus FheCL2, pH 5.5. *Panel B*, type I collagen was incubated with active recombinant peptidase (5.47 μM) at 28 ºC at neutral pH (PBS, pH 7.3) for 20 hours and the reactions analysed as above; *lane 1*, collagen alone (i.e. no peptidase added); *lane 2*, collagen plus FheCL1; *lane 3*, collagen plus FheCL1 Leu209Ala; *lane 4*, collagen plus FheCL1 Leu67Tyr; *lane 5*, collagen plus FheCL2. Molecular mass standards are indicated on the right while collagen chains (α1, α2, β11, β12 and γ, see 44) are indicated on the right.

FIG. 6
A, the bi-lobed mature FheproCL1 Gly^{25} is shown as a cartoon. The predominantly helical domain is at right and the predominantly sheet domain is at left. The mutated active site residue Gly^{25}, lies in the cleft between the two domains and is indicated in red. B, the structure of full-length FheproCL1 Gly^{25} zymogen is shown, with the mature segment surface illustrated in blue and the prosegment as a cartoon. The extended C-terminal portion of the prosegment runs through the active site cleft. The catalytic machinery of the enzyme is highlighted in pink. Figures were created with PyMol (42).

FIG. 7
A, Surface representation of the active site region of FheproCL1 Gly^{25}. The S2 pocket of the enzyme is highlighted in pink and key residues implicated in substrate preference, including the 'gatekeeper' residues (V157 and N158) are noted. B, Representation of the active site region of FheproCL1 Gly^{25} and FheCL2. The S2 pocket of both enzymes are highlighted. 'Gatekeeper' residues of FheproCL1 Gly^{25} are further indicated in pink stick representation and those of the modelled FheCL2 in yellow. Figures created with PyMol (42).

FIG. 8
Comparison of the substrate specificity of human cathepsin L (CL), human cathepsin K (CK), recombinant FheCL1, FheCL1 Leu205Ala (L205A), FheCL1 Leu67Tyr (L67Y) and FheCL2. Data shown as relative k_{cat}/K_M for the hydrolysis of the substrates Z-Phe-Arg-NHMec, Z-Phe-Leu-NHMec and Tos-Gly-Pro-ArgNHMec. *Data for human CL and CK is taken from Lecaille *et al.* (44).

Table 1. Crystallographic parameters: data collection and refinement statistics

Data Collection		
Space Group	$P2_12_12$	
Unit cell parameters		
a (Å)	57.17	
b (Å)	105.78	
c (Å)	49.11	
Wavelength (Å)	1.1159	
Temperature (K)	100	
Resolution (Å)	1.4	
Total number reflections	333582 (19545)	
Total unique reflections	59595 (8465)	
Completeness (%)	99.7 (98.7)	
Redundancy	5.6 (2.3)	
R_{merge}	0.080 (0.549)	
$R_{p.i.m.}$	0.029 (0.431)	
$\langle I \rangle / \langle \sigma I \rangle$	14.6 (1.8)	
Refinement		
Resolution range (Å)	52.93 – 1.40	
Number of reflections	56141	
R_{factor}	0.128	
R_{free}	0.165	
Free reflections (%)	5.05	
Average B factor (Å^2)		
Protein	15.39	
water	34.09	
R.m.s deviation from ideal		
Bond lengths (Å)	0.020	
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%)	

Note: Values given in parenthesis in Data Collection section above are for highest resolution bin (1.48-1.40 Å)

Table 2. Residues contributing to substrate binding in the S2 subsite of human cathepsin L, human cathepsin K, *F. hepatica* cathepsin L1 (FheCL1), *F. hepatica* cathepsin L1 Leu67Ala variant (FheCL1 Leu67Ala), *F. hepatica* cathepsin L1 Leu205Tyr variant (FheCL1 Leu205Tyr), and *F. hepatica* cathepsin L2 (FheCL2)

	RESIDUE						
	67	**68**	**133**	**157**	**158**	**160**	**205**
Human cathepsin L	Leu	Met	Ala	Met	Asp	Gly	Ala
Human cathepsin K	Tyr	Met	Ala	Leu	Asn	Ala	Leu
FheCL1	Leu	Met	Ala	Val	Asn	Ala	Leu
FheCL1 Leu205Ala	Leu	Met	Ala	Val	Asn	Ala	Ala
FheCL1 Leu67Tyr	Tyr	Met	Ala	Val	Asn	Ala	Leu
FheCL2	Tyr	Met	Ala	Leu	Thr	Ala	Leu

Table 3: Kinetic parameters for hydrolysis of peptidyl-NHMec substrates by recombinant wildtype FheCL1, variants FheCL1 Leu205Ala, FheCL1 Leu67Tyr and wildtype FheCL2.

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
FheCL1	Z-FR-NMec	24.69 ± 1.30	24.18 ± 3.92	1,021,092
FheCL1 Leu205Ala	Z-FR-NMec	29.60 ± 1.05	19.21 ± 3.78	1,540,864
FheCL1 Leu67Tyr	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 Leu205Ala	Z-LR-NMec	9.15 ± 0.24	2.75 ± 0.61	3,327,273
FheCL1 Leu67Tyr	Z-LR-NMec	1.73 ± 0.38	0.38 ± 0.05	4,552,632
FheCL2	Z-LR-NMec	1.62 ± 0.08	1.39 ± 0.20	1,165,468
FheCL1	Z-PR-NMec	1.03 ± 0.04	191.21 ± 16.90	5,387
FheCL1 Leu205Ala	Z-PR-NMec	0.122 ± 0.005	48.41 ± 4.45	2,479
FheCL1 Leu67Tyr	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 Leu205Ala	Tos-GPR-NMec	0.113 ± 0.003	20.35 ± 1.00	5,405
FheCL1 Leu67Tyr	Tos-GPR-NMec	0.26 ± 0.01	6.96 ± 1.08	37,069
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 Leu205Ala	Boc-AGPR-NMec	0.20 ± 0.04	21.57 ± 8.85	9,179
FheCL1 Leu67Tyr	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

Table 4: Inactivation of recombinant wildtype FheCL1, variants FheCL1 Leu205Ala, FheCL1 Leu67Tyr and wildtype FheCL2 by the diazomethyl ketone inhibitor Z-Phe-Ala-CHN_2.

Enzyme	**k_{obs}/[I] ($M^{-1}s^{-1}$)**	**$K_{i(app)}$ (nM)**
FheCL1	20,838 ± 589	1,125 ± 213
FheCL1 Leu209Ala	492,727 ± 12592	15.4 ± 3.44
FheCL1 Leu67Tyr	53,704 ± 4331	93.6 ± 9.1
FheCL2	11,899 ± 477	1,670.7 ± 791

Table 5: Inhibition values for recombinant wildtype FheCL1, variants FheCL1 Leu205Ala, FheCL1 Leu67Tyr and wildtype FheCL2 Cathepsin K Inhibitor II.

Enzyme	**k_{obs}/[I] ($M^{-1}s^{-1}$)**	**$K_{i(app)}$ (nM)**
FheCL1	397,237 ± 55370	10.80 ± 0.48
FheCL1 Leu209Ala	58,212 ± 5306	116.33 ± 24.40
FheCL1 Leu67Tyr	113,182 ± 8220	13.27 ± 0.87
FheCL2	269,447 ± 4611	23.14 ± 3.59
cathepsin K [a]	590,000 ± 1200	6.0
cathepsin L [a]	11,000 ± 560	-

[a] taken from Wang *et al*, (49)

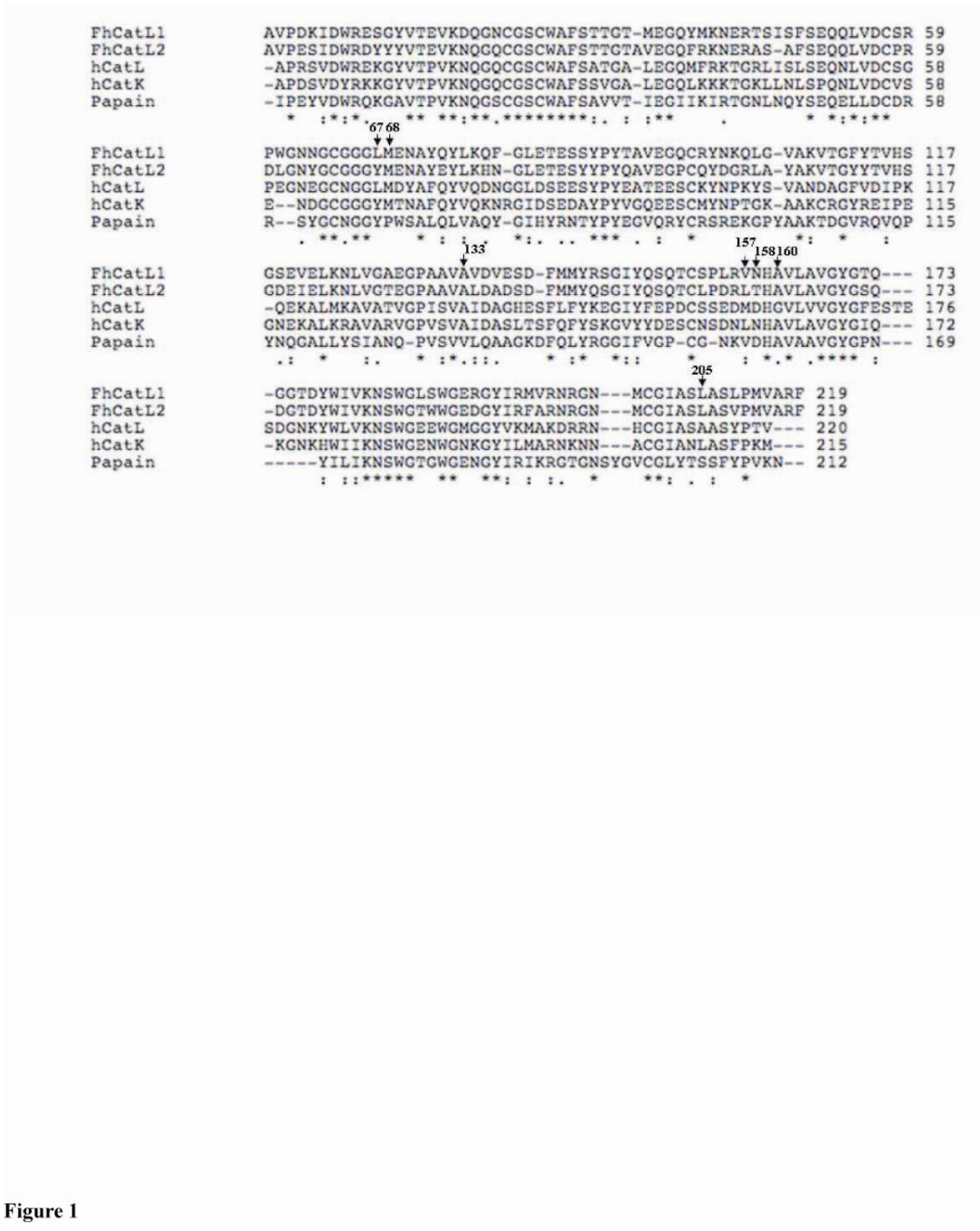

Figure 1

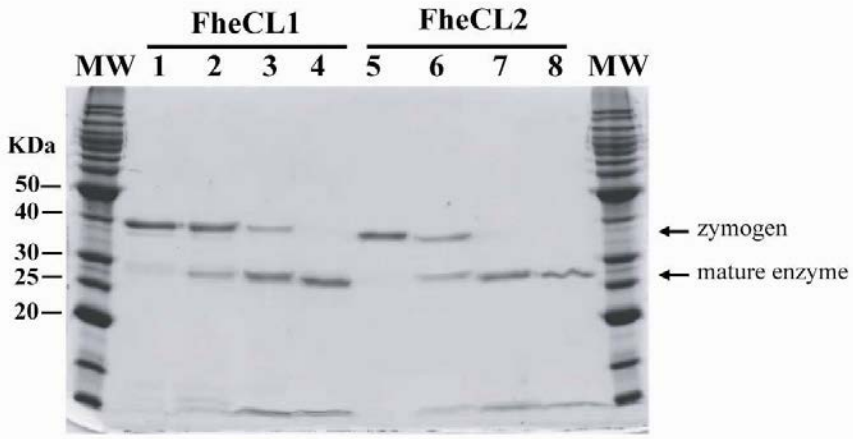

Figure 2

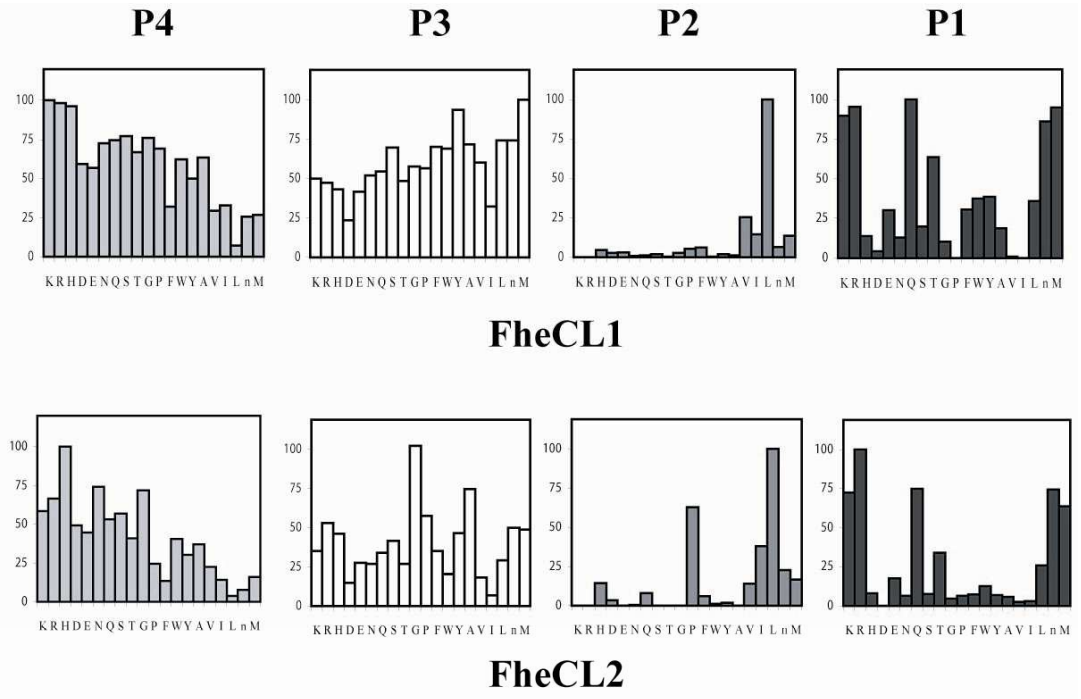

Figure 3

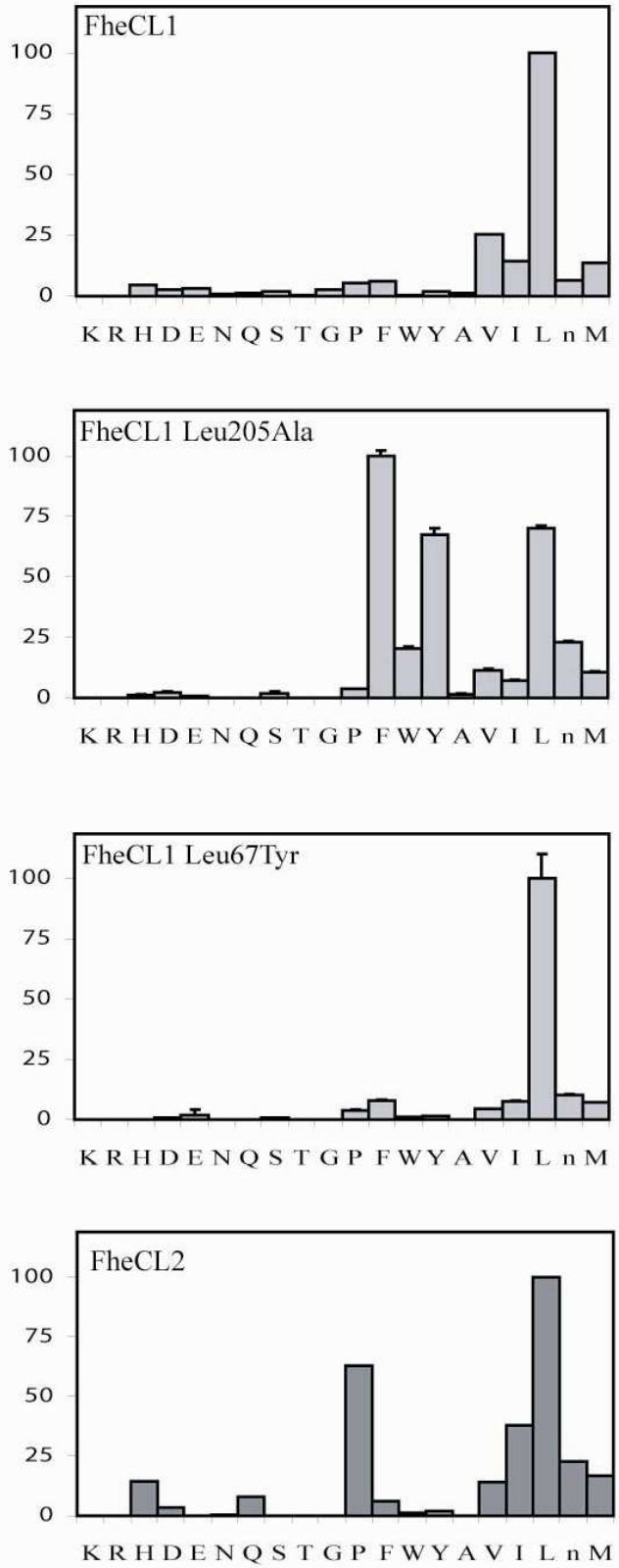

Figure 4

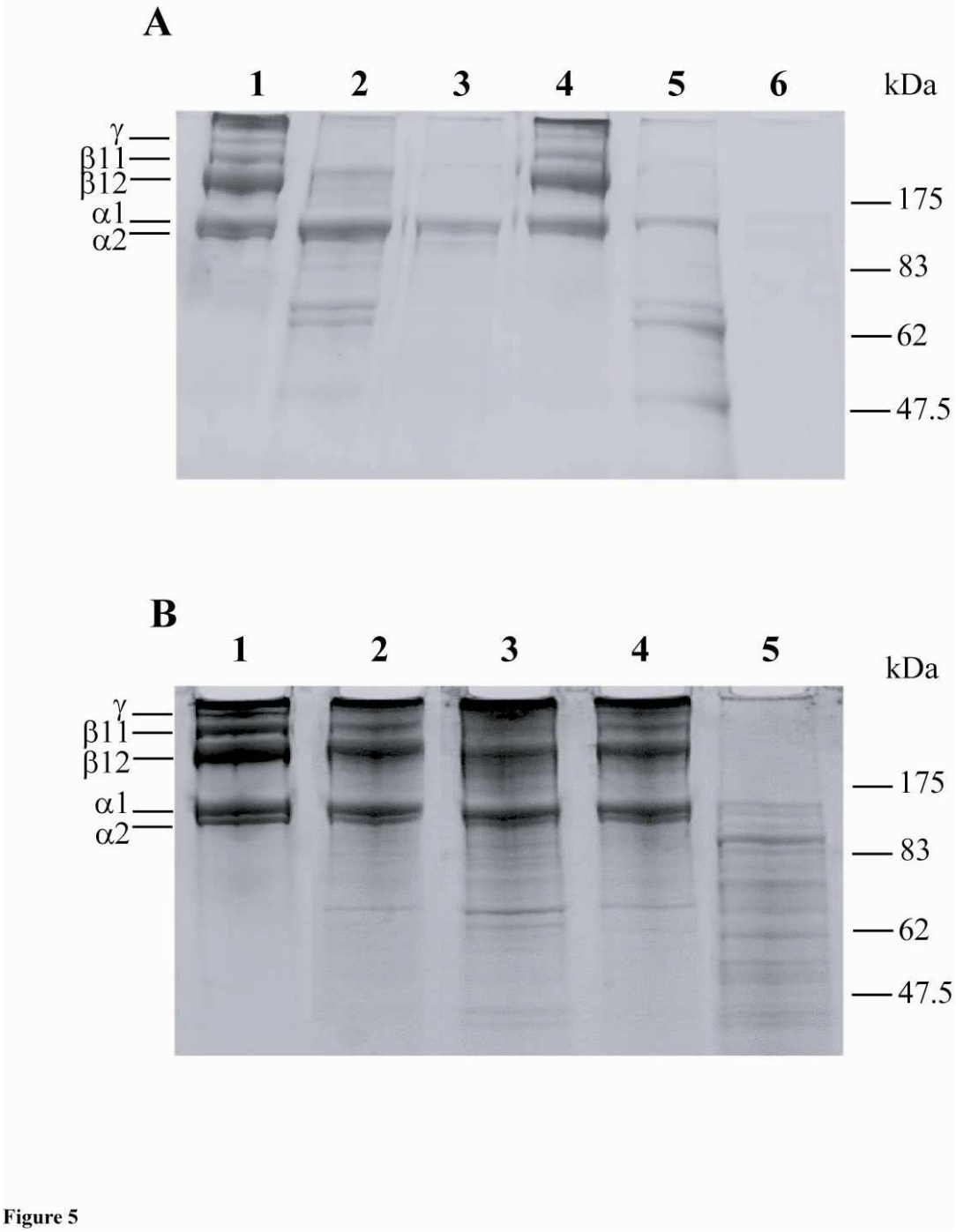

Figure 5

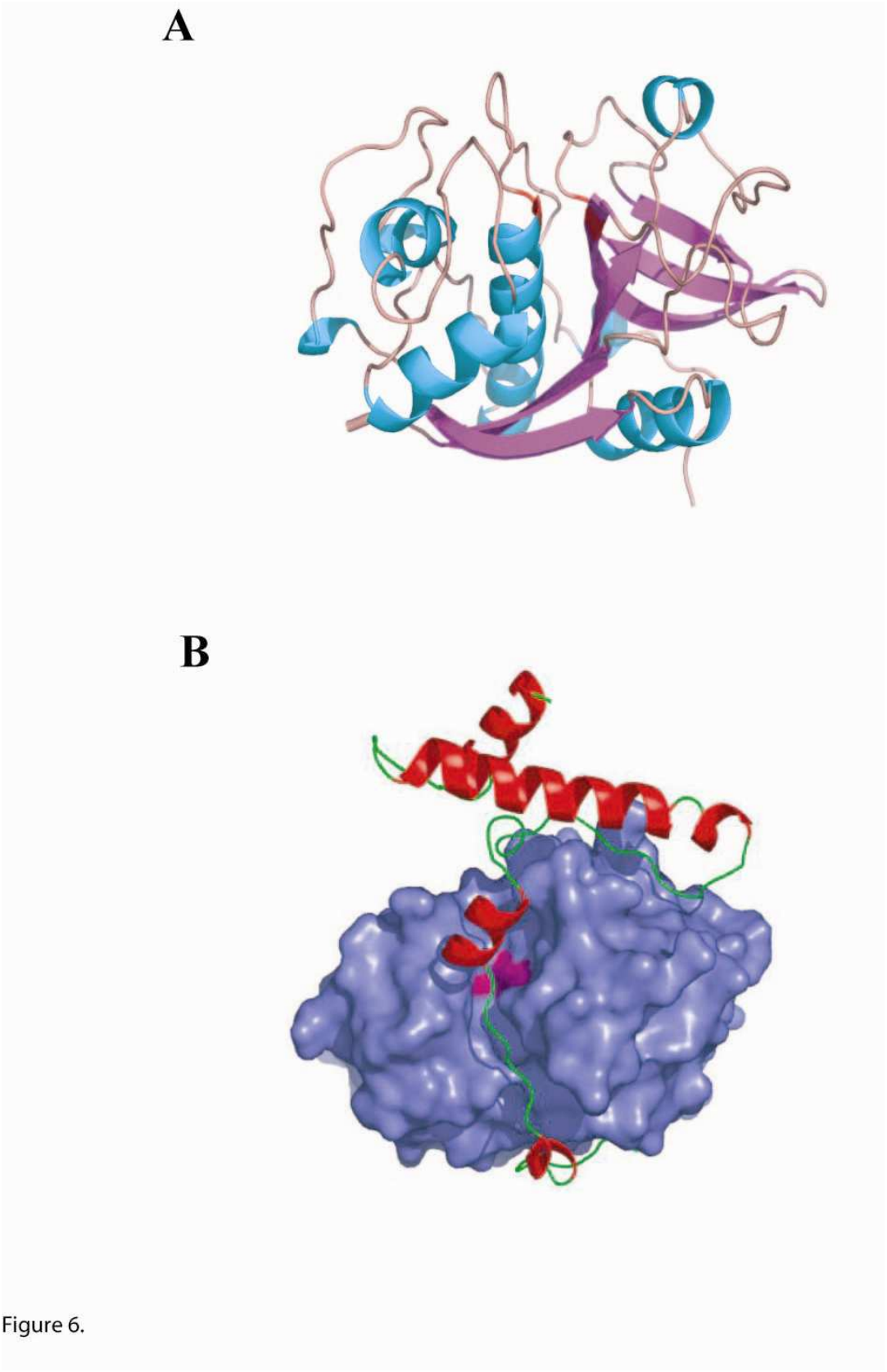

Figure 6.

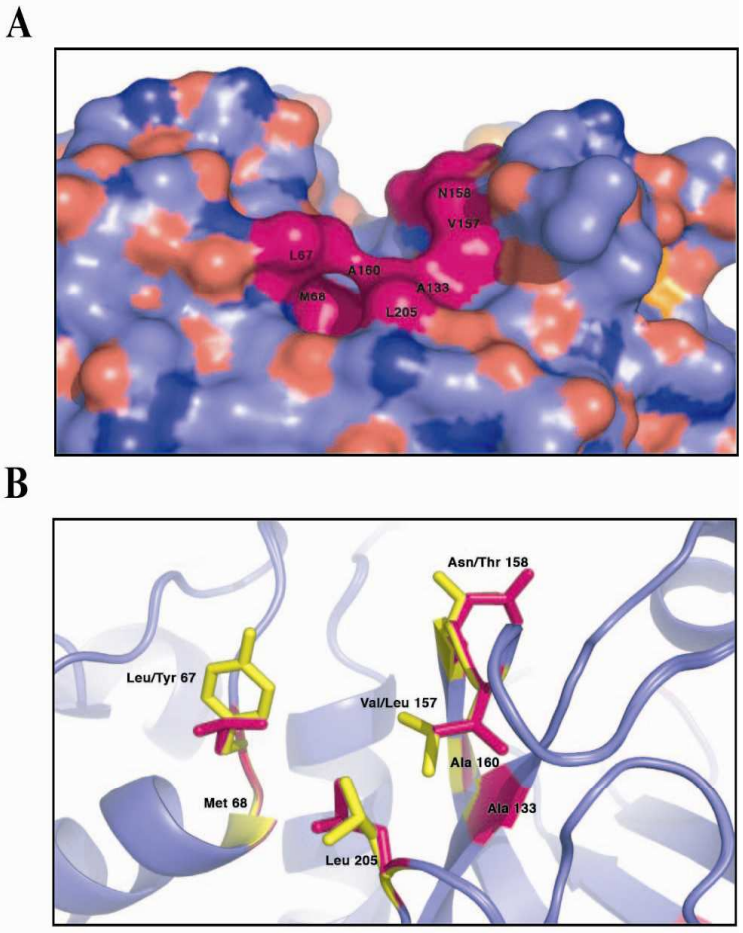

Figure 7

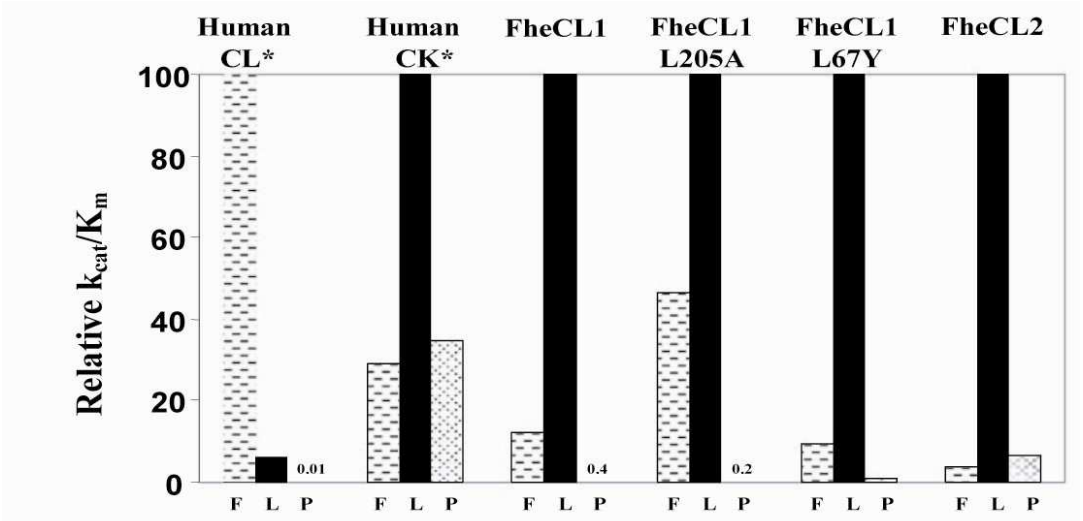

Figure 8