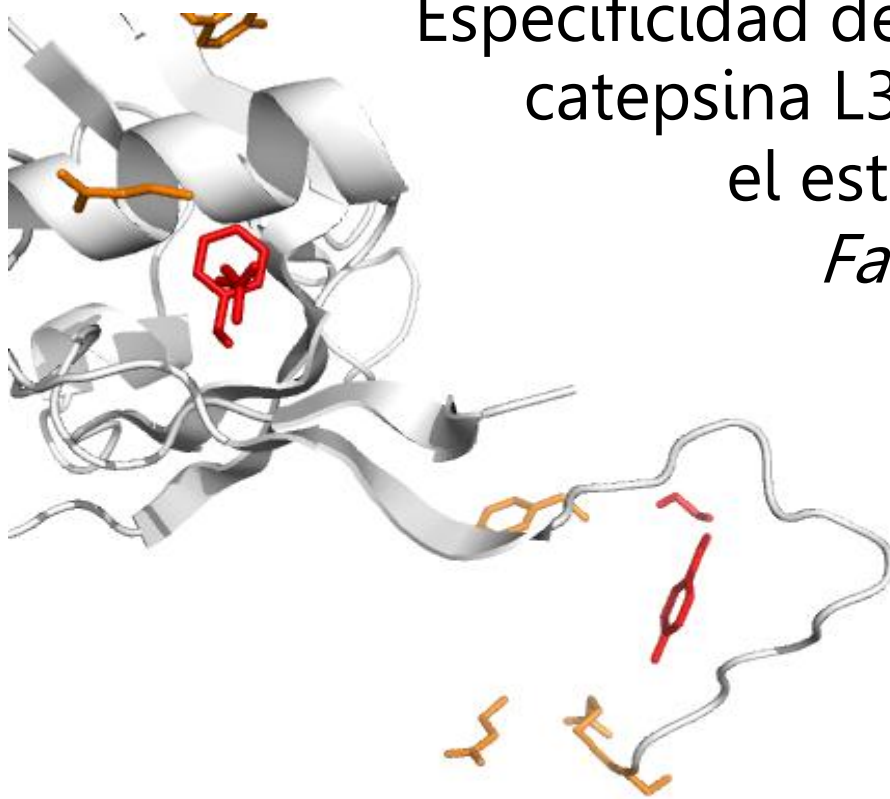




Tesis de Doctorado  
PEDECIBA  
**Biología Celular y Molecular**



Especificidad de sustrato de la  
catepsina L3 secretada por  
el estadio juvenil de  
*Fasciola hepatica*

**Lic. Ileana Corvo**

---

**Orientador: Dr. José Tort**

**Co-orientadores: Dra. Leda Roche y Dr. Carlos Carmona**

**Montevideo, Uruguay  
Agosto 2013**



*Durante todos estos años que hoy culminan con la presentación de esta tesis conté con el apoyo constante de mucha gente: orientadores, colegas, amigos y familiares. Por ello quisiera agradecer:*

***A mis orientadores,***

*por introducirme en el mundo de las catepsinas y guiarme durante la realización de este trabajo, por sus enseñanzas y sus fundamentales aportes. En particular a Pepe y a Leda, con quienes compartí el día a día en el trabajo.*

***A mis compañeros del Dpto. de Genética ,***

*que son un montón y no quisiera dejar de nombrar a alguno, por las horas de trabajo y experiencias que transitamos juntos. En particular a Mónica, por “heredarme” el sistema de expresión en Hansenula, a Nico y Gabriel por compartir las fasciolas, y en especial a Nati y Lu por las charlas, por su apoyo, por hacerme reír y estar ahí, en las buenas y en las malas.*

***A todos los compañeros del LOBBM,***

*por el tiempo compartido en el laboratorio, por estar siempre dispuestos a darme una mano o un reactivo.*

***A los compañeros de la UBP del Instituto de Higiene,***

*en especial a Martín que inició el estudio de las catepsinas de juveniles, a Tatiana, mi compañera de pasantía en San Francisco, y a Carlos por aceptar la co-orientación de esta tesis.*

***A los miembros del tribunal,***

*por el tiempo dedicado a leer y corregir este trabajo.*

***A mis compañeros de Celsius,***

*en particular a Paula, quien siempre me facilitó las cosas para que pudiera compatibilizar el desarrollo del doctorado con el trabajo en el Laboratorio.*

***A la gente que me abrió las puertas de sus laboratorios para realizar pasantías,***

*a los Dres. John Dalton y Mark Robinson del Infection, Immunity and Innovation (i3) Institute, University of Technology Sydney, a la Dra. Silvia Chifflet del Laboratorio de Bioquímica de la Facultad de Medicina, y a los Dres. Conor Caffrey, Anthony O’Donogue y James Mc. Kerrow del Center for Discovery and Innovation in Parasitic Diseases en UCSF.*

***A las instituciones que financiaron proyectos, cursos, congresos y pasantías,***

*CSIC, PEDECIBA, Wood Whelan Research Fellowship, AMSUD Pasteur*

***A toda mi familia,***

*a mamá, a papá, mis hermanos, mis suegros, cuñad@s y abuelos, que aunque algunos ya no estén, sé que estarían orgullosos de verme culminar esta etapa. En especial quiero agradecer a mi mamá, quien siempre me alentó a dedicarme a lo que me gusta y confió en que podía lograr las metas que me propusiese.*

***A mis amigos de la vida y de facultad,***

*Martín, Fabi, el Sapo, Andrés, Nati, Lu, Caro, Ana, Ale, Analía, Mica, Majo, Grace y Gaby por sacarme de la rutina, por las charlas, los viajes, las comidas y todos los momentos que han estado conmigo.*

***Finalmente, a Mauricio y Olivia,***

*a Mauricio por ser mi compañero de vida incondicional, por acompañarme en mis sueños, por confiar en mí y transitar juntos en los desafíos que plantea la vida. A Olivia, por haberse aferrado a mí, desde el principio y por siempre.*

## Contenido

---

AGRADECIMIENTOS.....	3
RESUMEN .....	7
ABREVIATURAS .....	8
I. INTRODUCCIÓN .....	9
1. <i>Fasciola hepatica</i> : modelo para el estudio de la interacción parásito-hospedador ..	10
1.1. Enfermedades parasitarias causadas por helmintos .....	10
1.2. Ciclo de vida de <i>Fasciola hepatica</i> .....	11
1.3. Estrategias de control de la fascioliasis.....	12
2. Efectores en la interfase parásito-hospedador .....	13
2.1. La familia de las cisteína proteasas.....	13
2.2. Las cisteína proteasas de <i>Fasciola hepatica</i> .....	16
2.2.1. Expresión estadio diferencial de catepsinas .....	16
2.2.2. Evidencias funcionales <i>in vitro</i> e <i>in vivo</i> .....	18
3. Especificidad de sustrato de las catepsinas L.....	20
3.1. Estructura molecular de las catepsinas.....	20
3.2. Sitio activo y mecanismo catalítico .....	22
3.3. Rol del bolsillo S <sub>2</sub> en la especificidad de sustrato .....	24
II. HIPÓTESIS Y OBJETIVOS .....	29
Hipótesis.....	30
Objetivo general .....	30
Objetivos específicos .....	30
III. MATERIALES Y MÉTODOS.....	32
1. Amplificación, clonado y secuenciación de catepsinas L de JRD.....	33
2. Expresión recombinante de catepsinas L activas.....	34
2.1. Transfección e integración de plásmido en <i>Hansenula polymorpha</i> .....	34
2.2. Selección de clones que expresan la enzima activa.....	35
2.3. Optimización de la producción de catepsinas recombinantes .....	35
2.4. Evaluación de la capacidad de autoactivación de pro <i>Fh</i> CL3 .....	36
2.5. Estudio del estado de glicosilación de <i>Fh</i> CL3 recombinante .....	36
3. Actividad enzimática con péptidos sintéticos fluorogénicos.....	36
3.1. Determinación de la cantidad de enzima activa.....	37
3.2. Estimación de los parámetros cinéticos .....	37
3.3. Ensayos de inhibición enzimática .....	37
4. Análisis de la digestión de sustratos proteicos .....	38
4.1. Identificación de péptidos de digestión del colágeno por HPLC-MS .....	38
5. Actividad con bibliotecas de péptidos sintéticos.....	39
5.1. Ensayos con la librería PS-SCL.....	39
5.2. Ensayos con la librería MSP-MS .....	39
6. Mutagénesis sitio dirigida .....	39

7. Modelado molecular de <i>FhCL3</i> .....	40
IV. RESULTADOS .....	41
1. Expresión de catepsinas L recombinantes activas en la levadura <i>Hansenula polymorpha</i> .....	42
Estrategia general .....	42
1.1. Identificación de catepsinas L del estadio juvenil.....	43
1.2. Expresión de <i>FhCL1</i> y <i>FhCL3</i> recombinantes .....	46
1.3. Activación de pro <i>FhCL3</i> a la enzima madura.....	47
1.4. Optimización de la expresión de catepsinas recombinantes .....	48
1.5. Expresión de las otras catepsinas de JRD .....	49
2. Caracterización enzimática de la catepsina L3 .....	50
Estrategia general .....	50
<b>Artículo:</b> La principal catepsina L secretada por el juvenil invasivo de <i>Fasciola hepatica</i> prefiere sustratos con prolina en el subsitio S <sub>2</sub> y puede cortar el colágeno .....	51
3. Estudio comparativo de la especificidad de sustrato de las catepsinas L del estadio juvenil y adulto.....	60
Estrategia general .....	60
3.1. Digestión de fibronectina y laminina .....	61
<b>3.2 Artículo:</b> La actividad colagenolítica de las principales catepsinas L involucradas en la virulencia del helminto patógeno <i>Fasciola hepatica</i> .....	62
4. Identificación de residuos relevantes en la especificidad de sustrato de las catepsinas de <i>Fasciola hepatica</i> .....	76
Estrategia general .....	76
<b>4.1 Artículo:</b> Disecando el sitio activo de la catepsina colagenolítica L3 del estadio invasivo de <i>Fasciola hepatica</i> .....	78
4.2. Contribución de otros residuos variables del subsitio S <sub>2</sub> a la especificidad de sustrato .....	91
V. DISCUSIÓN GENERAL .....	94
1. La familia de catepsinas L del estadio juvenil.....	95
2. Expresión de <i>FhCL3</i> recombinante activa .....	96
3. Especificidad de sustrato de las catepsinas de <i>F. hepatica</i> .....	98
3.1. Hidrólisis de sustratos peptídicos y librerías sintéticas .....	98
3.2. Hidrólisis de sustratos macromoleculares .....	101
4. Correlación estructura-función en el sitio activo de las catepsinas.....	104
VI. APLICACIONES Y PERSPECTIVAS .....	108
1. Aplicaciones en la prevención, tratamiento y diagnóstico de la fascioliasis .....	109
2. Aplicaciones biotecnológicas y biomédicas .....	110
VII. REFERENCIAS BIBLIOGRÁFICAS .....	112
VIII. ANEXOS .....	123

Las cisteína proteasas son enzimas proteolíticas presentes en todos los organismos. Las catepsinas L del trematode *Fasciola hepatica* constituyen una gran familia multigénica, con enzimas en su mayoría secretadas que están presentes en todos los estadios del parásito. *FhCL3* es la única catepsina tipo L que se ha encontrado hasta el momento en los productos de excreción/secreción del estadio juvenil de *F. hepatica*, y representa junto con las catepsinas B, mediadores moleculares clave que participan en la interfase parásito-hospedador. En este trabajo se puso a punto la expresión de *FhCL3* en la levadura *Hansenula polymorpha*, así como también la expresión de la enzima *FhCL1* secretada por los parásitos adultos, obteniéndose las enzimas recombinantes activas. Se realizó la caracterización de *FhCL3* mediante ensayos de activación de la proenzima, estabilidad y perfil de actividad a diferente pH, así como el estudio de su especificidad de sustrato con diferentes estrategias que incluyeron el análisis de la actividad hidrolítica sobre un panel de péptidos sintéticos fluorogénicos, inhibidores y dos bibliotecas peptídicas: PS-SCL y MSP-MS. *FhCL3* mostró acomodar preferentemente residuos de prolina y glicina en P<sub>2</sub> y P<sub>3</sub>, respectivamente. Se evaluó además la degradación de posibles sustratos naturales de la enzima en diferentes condiciones de pH y temperatura, observándose que *FhCL3* es eficiente en la degradación del colágeno, lo que apoya su rol en la migración de los parásitos juveniles a través del tubo digestivo y parénquima hepático. Demostramos que *FhCL3* tiene diferente especificidad que las enzimas secretadas por el parásito adulto, *FhCL1* y *FhCL2*, reafirmando la existencia de funciones estadio-específicas. Mediante mutagénesis dirigida se sustituyeron varios residuos del sitio activo de las catepsinas, evidenciando que el triptofano en la posición 67 del bolsillo S<sub>2</sub> de *FhCL3* es crucial para la actividad de la enzima. Nuestros resultados apoyan el papel activo de *FhCL3* durante el establecimiento de la infección, constituyéndose un blanco interesante para el desarrollo de posibles drogas o vacunas para el control de la fascioliasis.

---

<b>AMC</b>	Amino-metil cumarina
<b>ARNm</b>	ARN mensajero
<b>Boc-</b>	t-butyloxycarbonyl
<b>CB</b>	Catepsina B
<b>CL</b>	Catepsina L
<b>DTT</b>	Ditiotreitol
<b>E-64</b>	L-trans-epoxisuccinil-leucil-amido-(4-guanidinio) butano
<b>EDTA</b>	Ácido etilendiaminotetraacético
<b>ESTs</b>	Expressed Sequence Tags
<b>HPLC-MS</b>	cromatografía líquida acoplada a espectrometría de masas
<b>JRD</b>	juveniles recientemente desenquistados
$k_{cat}$	constante catalítica (número de recambio o <i>turnover</i> )
$k_{cat}/K_M$	constante de especificidad
<b>kD</b>	kiloDalton
$K_M$	constante de afinidad o de Michaelis-Menten
<b>MetOH</b>	Metanol
<b>MPM</b>	Marcador de peso molecular
<b>MSP-MS</b>	Multiplex Substrate Profiling – Mass Spectrometry
<b>Pb</b>	pares de bases
<b>PBS</b>	Tampón Fosfato Salino
<b>PES</b>	Productos de excreción/ secreción
<b>RT-PCR</b>	Transcripción reversa - Reacción en cadena de la polimerasa
<b>PM</b>	Peso Molecular
<b>PS-SCL</b>	Positional Scanning Synthetic Combinatorial Library
<b>Tos-</b>	tosil
<b>YEPD</b>	Yeast Extract, Peptone, D-glucose (Medio rico no selectivo)
<b>YMM</b>	Yeast Minimal Media (Medio selectivo)
<b>Z-</b>	benziloxi carbonilo

---



## **I. INTRODUCCIÓN**

## 1. *Fasciola hepatica*: modelo para el estudio de la interacción parásito-hospedador

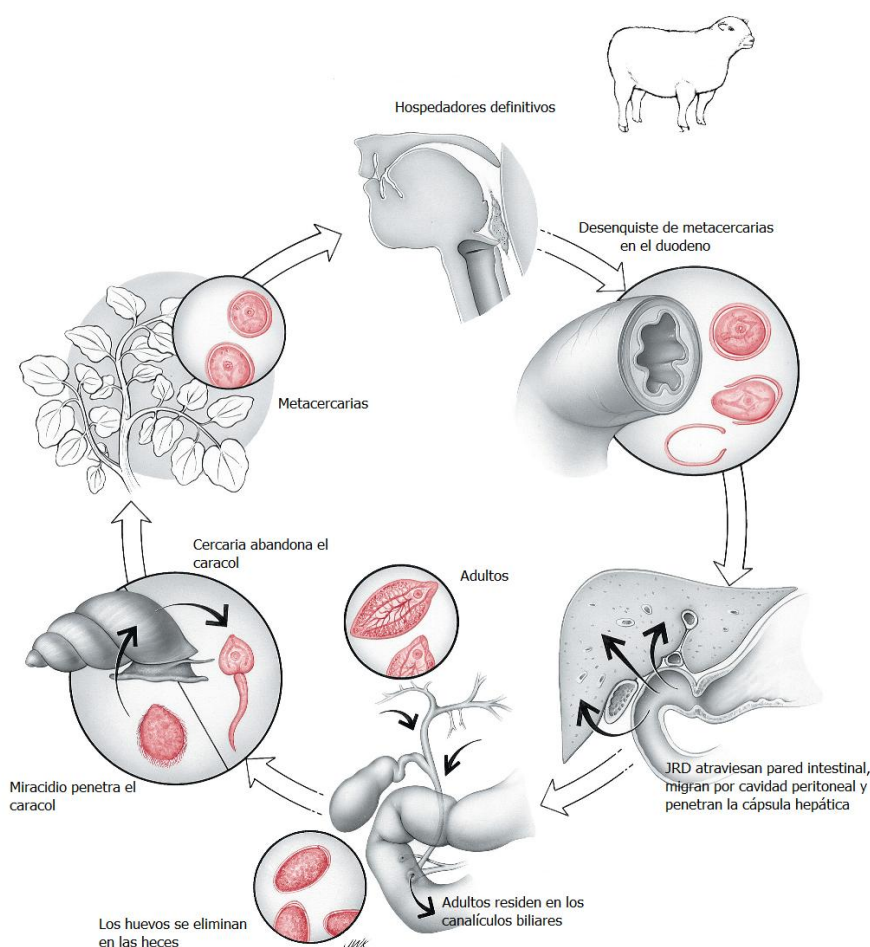
### 1.1. Enfermedades parasitarias causadas por helmintos

Las infecciones parasitarias producidas por helmintos son un grupo de enfermedades con un alto impacto global tanto en la salud humana como en el ámbito agroveterinario. Se estima que son responsables de más del 60% de las infecciones humanas y de más del 75% de las enfermedades infecciosas emergentes en humanos. En el ámbito productivo son responsables de importantes pérdidas por disminución de la productividad, la calidad y el gasto en atención sanitaria del ganado (Robinson & Dalton 2009). A pesar de su importancia, no se ha priorizado la investigación y búsqueda de nuevas estrategias para el control de estas enfermedades, lo que ha llevado a considerarlas como enfermedades “desatendidas” (Stein et al. 2007). Los helmintos parásitos incluyen los nematodos (gusanos redondos), los cestodos y los trematodos (gusanos planos). Las trematodiasis son infecciones parasitarias causadas por este último grupo de helmintos. Entre ellas las trematodiasis alimentarias, causadas por especies como *Paragonimus* spp, *Opisthorchis viverrini*, *Clonorchis sinensis*, *Fasciola gigantica* y *Fasciola hepatica* afectan a más de 100 millones de personas en el mundo, principalmente en regiones como el sudeste asiático, el altiplano boliviano y África sub-sahariana (Keiser & Utzinger 2009).

En particular *F. hepatica* y *F. gigantica*, son los agentes causantes de la fascioliasis en regiones subtropicales y tropicales, respectivamente. La fascioliasis es considerada una de las parasitosis más importantes entre los animales alimentados a pasturas a nivel mundial (ganado bovino y ovino), produciendo pérdidas por las mermas en la producción de carne, leche y lana, así como por los decomisos de hígados en los frigoríficos, interferencias en la fertilidad, aumento de la sensibilidad a adquirir nuevas infecciones y costos asociados a la aplicación de tratamientos. Así, las pérdidas económicas mundiales se estiman en 3 mil millones de dólares anuales (Mas-Coma, et al 2005). En nuestro país es una parasitosis importante con prevalencias que oscilan entre el 50-60% en el ganado bovino y ovino (Acosta 1991). Los datos de la Auditoría de Calidad de la Cadena Cárnica de 2007-2008 reportan que se pierden anualmente 6 millones de dólares solo a causa del decomiso de hígados (50% de hígados son decomisados), donde *F. hepatica* es uno de los principales agentes causales. La fascioliasis es también una enfermedad emergente en humanos. Es importante en países del altiplano sudamericano donde convive la población con el ganado como Bolivia, Perú y Ecuador, y también en Egipto, Irán y el sureste asiático. Asimismo, se registran entre 50-100 casos por año en países europeos como Portugal, Francia y España. La mayoría de los casos registrados en humanos se deben a infecciones con *F. hepatica*, aunque recientemente se ha reportado un aumento en las infecciones por *F. gigantica* en Vietnam (Mas-Coma et al. 2005). Se estima que hay entre 2.4 y 17 millones de personas infectadas con este parásito en el mundo y 91 millones en riesgo de infección (Keiser & Utzinger 2009).

## 1.2. Ciclo de vida de *Fasciola hepatica*

*Fasciola* presenta un ciclo de vida complejo que involucra diferentes estadios y hospedadores (Figura 1). El hospedador definitivo adquiere la infección al ingerir vegetación contaminada con metacercarias, la forma quística resistente del parásito, que al llegar al intestino se desenquista. Los juveniles recientemente desenquistados (JRD) atraviesan la pared intestinal, y migran por la cavidad peritoneal hasta el hígado. Luego de cruzar la cápsula hepática, migran por el parénquima como juveniles inmaduros durante 8-12 semanas hasta llegar a los canalículos biliares donde madura y se aloja el estadio adulto. En esta travesía los parásitos en desarrollo encuentran diferentes tejidos y macromoléculas, y deben enfrentarse a un microambiente fisiológico cambiante (ej. pH, disponibilidad de oxígeno) así como a la respuesta inmunológica del hospedador. La fase aguda de la enfermedad es debida a esta migración hacia los canalículos del hígado y se caracteriza por causar hemorragias, anemia, hepatomegalia y debilitamiento del animal afectado.



**Figura 1.** Ciclo de vida de *Fasciola hepatica* (tomado de <http://www.microbeworld.org>)

Una vez dentro del árbol biliar el parásito adulto se alimenta principalmente de sangre y alcanza la madurez sexual produciendo un gran número de huevos. La sangre es una fuente importante de carbohidratos para la producción de energía y de aminoácidos y ácidos grasos para la biosíntesis de moléculas del parásito y la producción de huevos. La superficie externa del parásito adulto es un tegumento sincitial multifuncional permeable a una gran variedad de pequeñas moléculas, participando en su absorción y distribución a los tejidos del parásito. Poseen un intestino ciego, encargado de digerir las macromoléculas obtenidas del hospedador y absorber los productos solubles, cuyo contenido se expulsa al medio cada 3-4 h conteniendo diferentes productos de excreción/secreción (PES). El ganado vacuno con fascioliasis puede mantener la infección por un término de 1-2 años, mientras el ganado ovino permanece infectado por períodos de hasta 20 años. La fase crónica se debe a la presencia del parásito adulto en los canalículos biliares, aunque muchas veces el animal convive con ambas fases debido a las reinfecciones recurrentes.

Los huevos son eliminados con las heces, y al eclosionar liberan una larva ciliada denominada miracidio, que infecta caracoles del género *Lymnea*. En este hospedador intermediario ocurren varios ciclos de reproducción asexual (esporocisto – redia - cercaria) que culminan con la liberación al medio de las formas de vida libre llamadas cercarias. Estas rápidamente se enquistan en la vegetación originando las metacercarias, que al ser ingeridas por el hospedador definitivo completan el ciclo del parásito (Andrews 1999).

### 1.3. Estrategias de control de la fascioliasis

El control de la fascioliasis se basa actualmente en el uso de drogas antihelmínticas, siendo los benzimidazoles, y en particular el triclabendazol la droga más utilizada a nivel mundial. Sin embargo, estos fármacos son costosos y pueden ocasionar problemas de tolerancia y hepato-toxicidad. Además, se ha reportado resistencia al triclabendazol en Irlanda, Inglaterra, Holanda, España y Australia, si bien existen otras drogas antihelmínticas que se han empleado con eficacia en estos casos (Brennan et al. 2007). Esto evidencia la necesidad de generar nuevas herramientas para la prevención y el tratamiento de la fascioliasis, apuntando tanto al descubrimiento de nuevos fármacos antihelmínticos así como al desarrollo de vacunas. En este sentido, se han empleado como antígenos en la formulación de vacunas algunas proteínas de *Fasciola*, nativas y/o recombinantes, como leucina aminopeptidasa, glutatión S-transferasa (GST), proteínas de unión a ácidos grasos y varias catepsinas, obteniéndose niveles de protección frente a la infección entre 65 y 89% (revisado en McManus & Dalton, 2006). Los niveles de protección más elevados han sido obtenidos en ensayos de vacunación con la leucina aminopeptidasa (Acosta et al. 2008) y/o con proteasas del tipo catepsinas, tanto con las enzimas nativas (Bentancor et al. 2002; Dalton et al. 1996; van Milligen et al. 2000; Mulcahy et al. 1998; Piacenza et al. 1999) como con las enzimas recombinantes (Golden et al.

2010; Maggioli et al. 2011; Reszka et al. 2005). A pesar de estas evidencias, ninguna vacuna se encuentra disponible actualmente en el mercado.

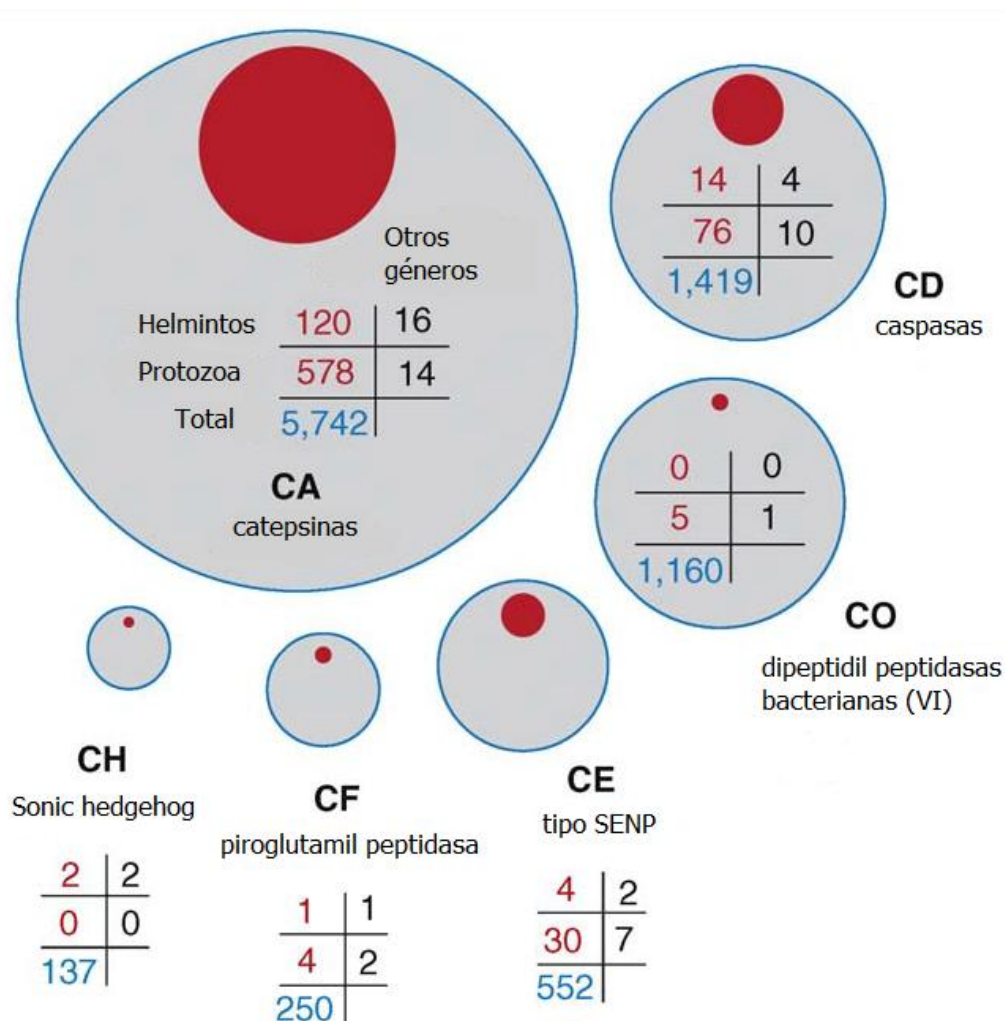
## 2. Efectores en la interfase parásito-hospedador

Las moléculas encontradas en la interacción del parásito con su hospedador son de gran interés ya que modifican o modulan el funcionamiento de las células del hospedador de manera de promover el desarrollo y sobrevivencia del parásito. Entre éstas, las proteasas producidas y secretadas por los parásitos son esenciales pues permiten atravesar barreras celulares y tisulares, degradar proteínas del hospedador para absorberlas como nutrientes y manipular su sistema inmune para evadir los mecanismos efectores encargados de erradicar la infección. Dada su abundancia y relevancia se utilizan como marcadores de inmunodiagnóstico (Atkinson, Babbitt, & Sajid 2009). Las proteasas son enzimas proteolíticas que catalizan la degradación de polipéptidos mediante la hidrólisis del enlace peptídico. De acuerdo al sitio donde cortan la cadena polipeptídica, las proteasas pueden ser endopeptidasas, si hidrolizan los enlaces peptídicos internos, o exopeptidasas cuando degradan a partir del extremo del sustrato. De acuerdo al residuo principal en su mecanismo de catálisis se clasifican en 5 clases principales: serina, cisteína, aspártico, treonina y metalo-proteasas. Hay también una sexta clase minoritaria, las glutámico proteasas. Para facilitar su estudio han sido agrupadas en familias de acuerdo a homología de secuencias y en clanes basados en homología estructural y catalítica.

### 2.1. La familia de las cisteína proteasas

En este trabajo nos centraremos en las cisteína o tiol proteasas, denominadas así por la presencia de un residuo conservado de cisteína en el sitio activo de la enzima, esencial para llevar a cabo la hidrólisis de sus sustratos. Las cisteína proteasas constituyen un grupo muy amplio de enzimas presentes en todos los organismos. Las cisteína proteasas más abundantes son las de la **superfamilia de la papaína**, que toma como referencia esta enzima aislada de la fruta papaya (*Carica papaya*). Esta superfamilia pertenece al **clan CA**, el cual contiene el mayor número de proteasas parasitarias descritas (Figura 2) e incluye además otras enzimas proteolíticas de plantas como la actinidina y bromelaína, y proteasas de animales denominadas catepsinas (del griego “digerir”). Las catepsinas típicas B, L, S, K, son enzimas lisosomales de mamíferos, aunque existen ortólogos en casi todos los organismos. Estas enzimas son endopeptidasas, aunque las catepsinas B también presentan actividad carboxipeptidasa, y la catepsina H de mamíferos presenta actividad aminopeptidasa. La excepción la constituye la catepsina C que posee únicamente actividad dipeptidil aminopeptidasa (Turk et al. 2000). Las diferencias de actividad se correlacionan con regiones específicas de similitud de secuencia alrededor del sitio catalítico que permiten dividir a las catepsinas animales en dos tipos: las

**catepsinas tipo L (CL)** y las **catepsinas tipo B (CB)** (Karrer, Peiffert, & Ditomas 1993). Ambos tipos de catepsinas se encuentran en toda la escala animal, pero están particularmente representadas en organismos parásitos. Las catepsinas constituyen el 80% de las proteasas de helmintos y 45% de las proteasas de protozoarios parásitos.



**Figura 2. Esquema de los 6 clanes de cisteína proteasas presentes en los parásitos: CA, CD, CE, CH, CF y CO.** El área de los círculos rojos es proporcional a la cantidad de secuencias de proteasas de helmintos y protozoarios parásitos depositadas en MEROPS, mientras los círculos grises son proporcionales a la cantidad total de proteasas en cada clan. La tabla indica el número de proteasas en cada clan de helmintos, protozoarios parásitos y otros géneros relacionados. Tomado de Atkinson et al., 2009.

La amplificación de estas enzimas puede haber ocurrido por duplicaciones génicas, como lo evidencia la presencia de genes organizados en tándem de isoformas similares en los tripanosomátidos (Campetella et al. 1992). La diversidad de catepsinas presentes en organismos parásitos sugiere que además de su actividad catabólica en el *turnover* celular y la

nutrición estas enzimas han adquirido nuevos roles (Atkinson et al. 2009). Consistentemente con esta diversidad, las proteasas parasitarias no solo se encuentran en compartimentos intracelulares como los lisosomas, sino que son secretadas al medio extracelular. Las enzimas parasitarias son sensibles al inhibidor genérico de cisteína proteasas E-64 al igual que las catepsinas de mamíferos, pero a diferencia de éstas que son activas a pH ácido, las parasitarias tienen actividad también a pH neutro y son estables en un amplio rango de pH (A. J. Dowd et al. 2000).

Las cisteína proteasas son capaces de degradar proteínas tanto intra como extracelulares, ya que algunas actúan dentro de las células mientras otras son secretadas al tubo digestivo del parásito y al medio externo. Lo que se conoce de la participación de las cisteína proteasas en la degradación de tejidos del hospedador proviene de evidencia indirecta de su expresión en aquellos estadios del parásito que penetran y migran a través de tejidos o la demostración de su capacidad de degradar proteínas de la matriz extracelular o las membranas basales *in vitro*. En estrecha relación encontramos el hecho de que las cisteína proteasas son también necesarias para que muchos protozoarios y helmintos puedan abandonar sus quistes o huevos y realizar el cambio de su cutícula. Como ejemplo podemos citar las cisteína proteasas involucradas en el desenquiste y migración de los trematodes parásitos *Paragonimus westermani*, *Schistosoma* y *Fasciola*, los cuales penetran por diferentes vías de entrada y se alojan en distintos órganos; *Paragonimus* y *Fasciola* penetran por el tracto gastrointestinal y se alojan en pulmón e hígado, respectivamente, *Schistosoma* penetra por la piel y se aloja en la red vascular (Dzik 2006; McGonigle et al. 2008; Na et al. 2006; Parker-manuel et al. 2011).

Se ha descrito que distintos tipos de proteasas actúan de forma coordinada para llevar a cabo la digestión de sus sustratos. La degradación de hemoglobina por parte del trematode hematófago *S. mansoni*, por ejemplo, ha sido extensamente estudiada y una catepsina tipo B denominada SmCB1 localizada en la gastrodermis y el intestino del parásito es la principal responsable de esta actividad. Otras cisteína proteasas de *S. mansoni* como la catepsina L1, la catepsina D y la catepsina C también contribuyen a la degradación de hemoglobina formando parte de una cascada proteolítica en la que también intervienen otras clases de proteasas, como las aspártico proteasas y las metaloproteasas (Delcroix et al. 2006).

Un papel importante juegan asimismo en la evasión de la respuesta inmunitaria del hospedador, ya sea degradando moléculas efectoras del sistema inmune o modulando las respuestas de células T. En algunos parásitos como *Haemonchus* y *Fasciola* las cisteína proteasas han demostrado su capacidad para degradar inmunoglobulinas *in vitro* (Berasain et al. 2000; Smith et al. 1993) Asimismo, las catepsinas han demostrado modular el sistema inmune suprimiendo las respuestas Th1, y promoviendo la expresión de efectores involucrados

en respuestas de tipo Th2 inmuno reguladoras las cuales resultan protectoras para el parásito (Donnelly et al. 2010; Prowse et al. 2002; Robinson et al. 2013).

## 2.2. Las cisteína proteasas de *Fasciola hepatica*

En 1989, Dalton & Heffernan (Dalton & Heffernan 1989) demostraron que los parásitos adultos de *F. hepatica* mantenidos *in vitro* secretaban cisteína proteasas al medio del cultivo. Mediante SDS-PAGE con gelatina demostraron que las cisteína proteasas eran los componentes secretados por el parásito más abundantes. Estudios más recientes del secretoma mediante electroforesis bidimensional y espectrometría de masas confirmaron que las moléculas principales secretadas son proteasas de tipo catepsinas aunque permitieron identificar además otros componentes menos representados tales como proteínas de unión a ácidos grasos (FABPs), enzimas reguladoras de los procesos de óxido reducción (tioredoxina, peroxirredoxina, GSTs), entre otras (Jefferies et al. 2001; Robinson et al. 2009)

Las catepsinas son sintetizadas dentro de las células epiteliales que recubren el tubo digestivo ciego de *Fasciola*, las cuales cumplen tanto funciones de absorción como de secreción. Previo a su secreción las catepsinas se hallan almacenadas como proenzimas en vesículas que se acumulan en la cara apical o luminal de las células epiteliales, pero son activadas al salir al intestino ligeramente ácido del parásito, previamente a su excreción al medio externo (Dalton, Skelly & Halton. 2004).

Estudios morfológicos y de la ultraestructura de *Fasciola* muestran alteraciones importantes en sitios de intercambio con el hospedador, tanto en la superficie del parásito como en la gastrodermis, a medida que los gusanos migran y se enfrentan a diferentes microentornos (Fairweather, Threadgold & Hanna 1999).

### 2.2.1. Expresión estadio diferencial de catepsinas

Las catepsinas L de *Fasciola hepatica* son la familia de cisteína proteasas más extensa conocida hasta el momento. El análisis de sus secuencias permite organizar esta familia en 5 grupos: *FhCL1*, *FhCL2*, *FhCL3*, *FhCL4* y *FhCL5*. De ellos, *FhCL1* se subdivide además en 4 sub-grupos denominados *FhCL1 A-D* (Irving et al. 2003; Morphew et al. 2011; Robinson, Tort, et al. 2008). La expresión de cisteína proteasas está finamente regulada durante el desarrollo de *Fasciola* en su hospedador definitivo y se correlaciona con las etapas de desenquiste, migración del parásito juvenil y maduración hacia la forma adulta (Robinson, Dalton & Donnelly, 2008)

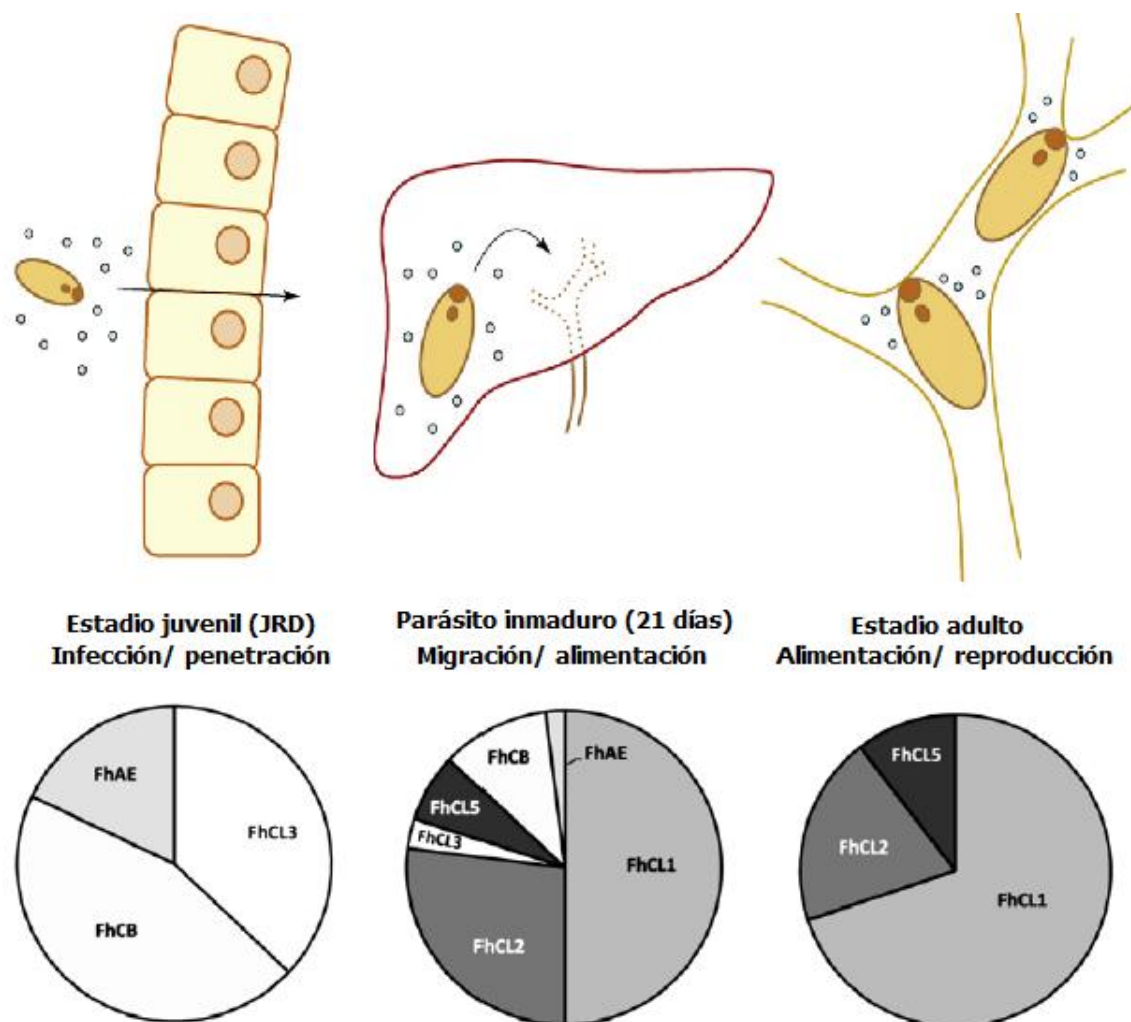
El análisis bioquímico de las proteasas secretadas por el parásito adulto mostró que los PES consistían principalmente de 2 catepsinas L, denominadas catepsina L1 y catepsina L2, con diferentes propiedades fisicoquímicas y especificidad de sustrato (Dowd et al. 1994; Smith et al. 1993). Una evidencia de la expresión estadio diferencial de proteasas surge de un



estudio de las proteínas expresadas por los JRD realizado en 1995, donde se halló una proteína de expresión predominante en ese estadio, la cual se demostró que era una cisteína proteasa diferente a las encontradas principalmente en adultos (Tkalcevic, Ashman & Meeusen 1995). La enzima fue posteriormente clonada y denominada *FhCL3* (Harmsen et al. 2004). La catepsina L3 tiene una identidad de secuencia de alrededor del 70% con las catepsinas *FhCL1* y *FhCL2* de adultos. Si bien la arquitectura del sitio catalítico se encuentra 100% conservada, se observaron diferencias con las CL de adultos en determinados residuos que han demostrado ser clave en la especificidad de sustrato de estas enzimas. Además, la catepsina L3 presenta sitios potenciales de *N*-glicosilación que están ausentes en *FhCL1* y *FhCL2*.

En estudios posteriores del transcriptoma y proteoma de *Fasciola* se encontró que en la metacercaria y en los JRD se expresan tanto catepsinas tipo L como tipo B, además de asparaginil endopeptidasas (legumainas), que son enzimas que hidrolizan el sitio C-terminal de residuos de asparagina y están involucradas en la activación de pro-enzimas a la forma madura (Figura 3). En el estadio adulto sin embargo sólo encontramos catepsinas L (Cancela et al. 2008; Hernández-González et al. 2010; Robinson, Tort, et al. 2008; Robinson et al. 2009). En parásitos juveniles se identificaron ARNm exclusivos de *FhCL3* y *FhCL4* (Cancela et al. 2008), sin embargo, solo *FhCL3* se encontró secretada en los PES, por lo que *FhCL4* o bien se secreta a niveles mucho más bajos o lleva a cabo funciones intracelulares específicamente (Hernández-González et al. 2010; Robinson et al. 2009) (Figura 3).

Las proteasas representan el 31% del total de proteínas secretadas por los JRD, y de éstas un 37% corresponde a catepsinas L (específicamente *FhCL3*), 45% a catepsinas B y un 18% a asparaginil endopeptidasas. A medida que los juveniles migran por la cavidad peritoneal y penetran la cápsula y parénquima hepático comienzan el proceso de maduración hacia la forma adulta. En estos parásitos, que aún se encuentran inmaduros y migrando a los 21 días post-infección, ya se detectan cambios en el repertorio de catepsinas que secretan, observándose una drástica disminución en los niveles de expresión de *FhCL3* y catepsinas B (ahora solo representan el 3% y 2% del total de catepsinas secretadas, respectivamente), y un aumento en la secreción de las catepsinas L1, L2 y L5, las únicas detectadas finalmente en el parásito adulto alojado en los canalículos biliares. En este estadio *FhCL1* es la enzima mayoritaria, constituyendo el 67% de las catepsinas secretadas, mientras *FhCL2* representa el 27% y *FhCL5* sólo un 6% (revisado en McVeigh, Maule, Dalton & Robinson, 2011) (Figura 3).



**Figura 3. Expresión diferencial de catepsinas durante el proceso de infección del hospedador definitivo.** Los gráficos de torta representan los niveles de expresión de las proteasas determinados mediante un análisis cuantitativo de espectrometría de masas. FhAE: *Fasciola hepatica* asparaginil endopeptidasas; FhCB: *Fasciola hepatica* catepsinas B; FhCL1-5: *Fasciola hepatica* catepsinas L 1-5 (adaptado de Robinson, Tort, et al. 2008; Robinson et al. 2009).

### 2.2.2. Evidencias funcionales *in vitro* e *in vivo*

Las catepsinas son las proteínas principales encontradas en la bilis de ovejas infectadas con *Fasciola* (Morphew et al. 2007). El perfil de catepsinas identificado por electroforesis bidimensional en parásitos adultos cultivados *in vitro* y el perfil obtenido al analizar la bilis de ovejas infectadas es muy similar. Si bien se identificó un mayor número de catepsinas *in vitro*, el 90% de las catepsinas encontradas *in vivo* también están representadas en la muestra *in vitro*. En ambas muestras las catepsinas pertenecían a los mismos 3 clanes hallados previamente en parásitos adultos, no detectándose en la bilis la presencia de catepsinas específicas identificadas en parásitos juveniles en cultivo. Las catepsinas exclusivas de las muestras *in vivo* se encuentran todas dentro del grupo de FhCL1, por lo que su expresión

diferencial reflejaría o bien una respuesta específica del parásito al entorno del hospedador (tal como la evasión de la respuesta inmunitaria o la adquisición de nutrientes), o el hecho que la cepa utilizada para los ensayos *in vitro* no es la misma que está infectando al ganado en estudio (Morphew et al. 2011). La notoria abundancia de estas enzimas en todos los estadios del parásito evidencia su importancia en el correcto desarrollo de su ciclo de vida, participando en procesos que van desde la infección del hospedador definitivo hasta la supervivencia del gusano adulto en los canalículos biliares.

Determinar el rol que cumplen las cisteína proteasas de *F. hepatica* en la interacción con su hospedero es uno de los principales desafíos que plantea el estudio de estas enzimas. Se han realizado numerosos esfuerzos por comprender el papel de las catepsinas en el proceso de invasión y establecimiento del parásito, y cómo éstas contribuyen al éxito de la infección. En este sentido, se ha demostrado que las catepsinas secretadas por el estadio adulto de *Fasciola* pueden degradar proteínas de la matriz extracelular y las membranas basales, digerir hemoglobina y cortar la región bisagra de las inmunoglobulinas lo cual sugiere su rol durante la migración de la larva inmadura a través del parénquima hepático, la obtención de nutrientes y la evasión de la respuesta inmunitaria (Berasain et al. 2000; Berasaín et al. 1997; Lowther et al. 2009; Smith et al. 1993). La principal fuente de nutrientes con que cuenta el parásito adulto de *Fasciola hepatica* en los canalículos biliares es la sangre. La enzima principal secretada por el adulto, *FhCL1*, es capaz de digerir la hemoglobina a péptidos pequeños de entre 4-14 aminoácidos cuando se encuentra a pH ácido ( $\leq 4.5$ ), los cuales se hidrolizarían posteriormente a aminoácidos libres dentro de las células del lumen intestinal para ser empleados en la síntesis de proteínas parasitarias (Lowther et al. 2009). Por otra parte, se ha observado que tanto *FhCL1* como *FhCL2* son capaces de digerir la región bisagra de las inmunoglobulinas G (IgG1 – 4), en presencia de agentes reductores como DTT (Berasain et al. 2000). Asimismo, ambas pueden hidrolizar moléculas de fibronectina, laminina, y colágeno soluble tipo IV que forman parte de la matriz extracelular y las membranas basales (Berasaín et al. 1997), pero sólo *FhCL2* es capaz de digerir el colágeno fibrilar tipo I (Stack et al. 2008).

En un estudio con un modelo murino de fascioliasis, donde se evaluó el efecto de la inhibición de las catepsinas con E-64 sobre los niveles de proteólisis en tejido hepático de ratones infectados con metacercarias, se observó un retraso en la progresión de la fascioliasis, una inhibición del 87% en la proteólisis a nivel hepático y una reducción en el número de huevos totales y embrionados producidos por el gusano (Alcalá-Canto et al. 2006). Un efecto similar en la disminución del daño hepático y la reducción en el crecimiento y fecundidad de *Fasciola* se observó al tratar ovejas infectadas con 50 mg/kg del inhibidor de cisteína proteasas Ep-475 (Alcala-Canto et al. 2007). Estos resultados apoyan el rol de las cisteína proteasas en la migración a través del tejido hepático y el desarrollo del gusano adulto, y sugieren que las catepsinas son un factor importante en los efectos patológicos observados *in vivo*. Posteriormente, un estudio del silenciamiento de las catepsinas B y L mediante ARN de

interferencia en JRD mostró que los parásitos interferidos tenían disminuida su capacidad de atravesar la pared intestinal, evidencia directa de la importancia de las cisteína proteasas en el proceso invasivo (McGonigle et al. 2008).

Estos resultados han sido interpretados como evidencia de que las enzimas secretadas por el parásito actúan de forma sinérgica. Por un lado, le permitirían atravesar las barreras físicas del hospedador, abriéndose paso por el parénquima hepático en su migración hacia los canalículos biliares, a la vez que modularían la respuesta inmune del hospedador evitando que las moléculas efectoras del sistema inmune logren erradicar la infección. Por otro lado, se plantea su participación en la adquisición de los nutrientes necesarios para cumplir con las funciones anabólicas y la alta producción de huevos. Es claro entonces el rol biológico esencial de estas enzimas para el parásito, haciendo de estas proteasas, y en particular las catepsinas L, buenos candidatos para la generación de vacunas o blancos de drogas antihelmínticas.

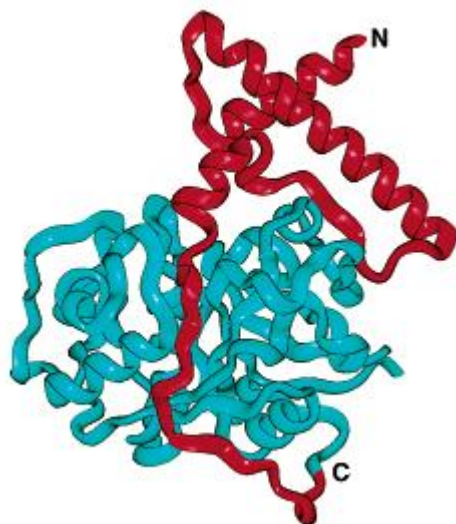
### **3. Especificidad de sustrato de las catepsinas L**

La habilidad de una enzima de discriminar entre varios sustratos potenciales es esencial para que pueda llevar a cabo su función correctamente. En un contexto biológico la selección del sustrato puede ser regulada de diferentes maneras, como la expresión o localización espacial diferencial de enzima y sustrato, las concentraciones de los mismos, o el requerimiento de cofactores u otros activadores enzimáticos. Sin embargo, el reconocimiento del sustrato en el sitio activo de la enzima es la interacción principal que finalmente determina se lleve a cabo o no la hidrólisis del mismo. Además, conocer la especificidad de sustrato de una enzima provee información importante para el desarrollo de sustratos e inhibidores selectivos que permitan estudiar la actividad de la enzima, a la vez que contribuye a la determinación de su rol biológico (Harris et al. 2000).

#### *3.1. Estructura molecular de las catepsinas*

Las catepsinas son proteínas generalmente globulares, pequeñas (25-35 KDa, la enzima madura), con dos dominios (izquierdo y derecho) que delimitan un surco en la superficie de la enzima donde se encuentra el sitio activo. Todas las proteasas de esta familia se producen como enzimas inactivas, denominadas proenzimas o zimógenos, conteniendo un péptido señal, un propéptido y un dominio catalítico, que representa la enzima madura activa proteolíticamente (Figura 4). El péptido señal consta por lo general de 10-20 aminoácidos y es responsable de la translocación de la enzima recién sintetizada al retículo endoplásmico (Lecaille, Kaleta & Brömme 2002). El propéptido, también denominado prosegmento, es de longitud variable y tiene una función tipo chaperona participando en el correcto plegamiento del dominio catalítico de la enzima y regulando su tráfico intracelular (Tao et al. 1994). Además, actúa como un inhibidor reversible de alta afinidad uniéndose en el surco del sitio activo de la

enzima en una orientación inversa al sustrato, impidiendo así la activación prematura de la proteasa hasta que alcanza su lugar de destino (Cappetta et al. 2002).



**Figura 4. Modelo esquemático de una catepsina (proenzima).**

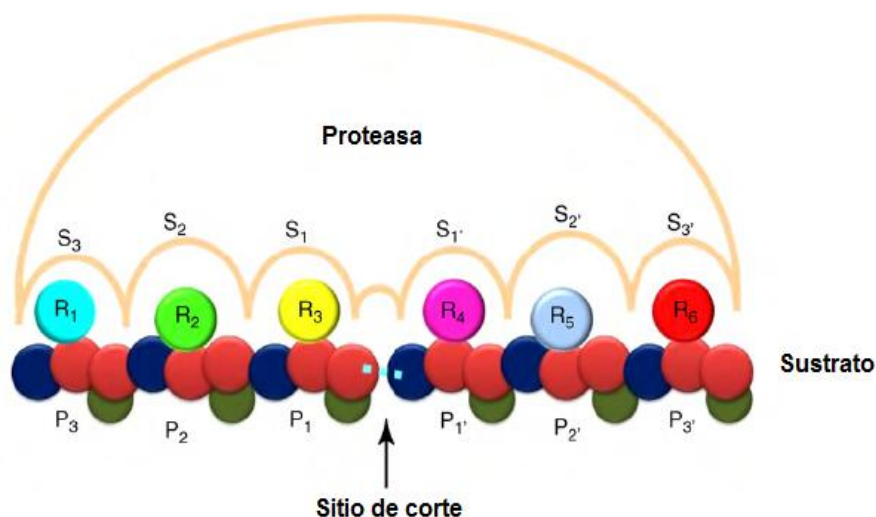
En rojo se observa el propéptido, ocupando el surco del sitio activo de la enzima. En azul se representa el dominio catalítico. N y C indican los extremos amino y carboxilo terminales respectivamente (adaptado de Lecaille, Kaleta, et al., 2002).

En la papaína, el propéptido tiene una longitud de 107 aminoácidos, y si bien esta región presenta mayor variación que el dominio catalítico, existe un motivo conservado (Gly-X-Asn-X-Phe-X-Asp<sup>-36</sup>) que es esencial para el correcto procesamiento de la enzima a su forma madura (Vernet et al. 1995). Durante este proceso, es necesaria la remoción del prosegmento para generar la forma proteolíticamente activa. La hidrólisis del propéptido ocurre a pH ácido de forma tanto intramolecular como intermolecular, observándose en muchos casos la autoactivación de la proenzima (Stack et al. 2008; Turk et al. 2000). El dominio catalítico de las cisteína proteasas tipo papaína contiene entre 220-260 aminoácidos, aunque algunas enzimas parasitarias poseen extensiones de sus dominios C-terminales (Lecaille, Kaleta, et al. 2002), u otro tipo de modificaciones. Las catepsinas tipo L y tipo B difieren en la secuencia y longitud de los propéptidos, así como en la presencia de un motivo conservado en el propéptido de las **catepsinas tipo L (ERFNIN)**, de acuerdo a la secuencia consenso conservada). Las catepsinas de tipo B por su parte, poseen una inserción de unos 20 aminoácidos denominada bucle ocluyente entre los residuos Tyr<sup>103</sup> y Cys<sup>128</sup> (numeración correspondiente a la catepsina B), donde dos residuos de histidina (110 & 111) son los responsables de la actividad dipeptidil carboxipeptidasa exclusiva de este grupo (Musil et al. 1991). El dominio catalítico es el que se encuentra más conservado, las cisteína proteasas poseen un mecanismo catalítico común que involucra un residuo de cisteína que forma una díada catalítica junto a un residuo de histidina. La cisteína del sitio activo se encuentra dentro de la secuencia peptídica altamente conservada **CGSCWAFS** (se resalta en negrita la cisteína del sitio activo), mientras la histidina<sup>159</sup> se ubica adyacente a residuos pequeños como glicina o alanina, seguida por residuos hidrofóbicos como valina, leucina o isoleucina, y una glicina. También se encuentran conservados los residuos de cisteína que participan en la formación de puentes disulfuro (Cys<sup>22</sup>-Cys<sup>63</sup>, Cys<sup>56</sup>-Cys<sup>95</sup> y Cys<sup>153</sup>-Cys<sup>200</sup>), y un residuo de prolina en la posición 2 de la enzima madura, que se

sugirió impide que el extremo N-terminal sea truncado e inactivado por la acción de aminopeptidasas (Lecaille, Kaleta, et al. 2002).

### 3.2. Sitio activo y mecanismo catalítico

En las cisteína proteasas el sitio activo se encuentra en una hendidura en forma de "V" localizada entre los 2 lóbulos de la enzima, en general es poco profundo y se halla expuesto al solvente, lo que hace que sus sustratos e inhibidores sean segmentos cortos de hebras- $\beta$  o bucles de proteínas que interaccionan con los residuos del sitio activo en una conformación extendida (Tyndall et al. 2005). La especificidad de las cisteína proteasas está determinada por las interacciones que se producen entre las cadenas laterales de diferentes aminoácidos del sitio activo de la enzima y el sustrato. Schechter & Berger describen esta interacción considerando el modelo donde un determinado subsitio del sitio activo de la enzima interacciona con un residuo aminoacídico específico del sustrato. Por convención, los subsitios de la enzima se denominan S (del inglés **S**ubsites) mientras los aminoácidos del sustrato con que interaccionan se denominan P (del inglés **P**eptide). De esta forma, los residuos aminoacídicos del lado amino terminal del sitio de corte del sustrato se denominan  $P_1$ ,  $P_2$ ,  $P_3$ , etc.; mientras los que se encuentran hacia el extremo carboxilo terminal se designan  $P_1'$ ,  $P_2'$ ,  $P_3'$ , etc. Los subsitios de la proteasa que interaccionan con dichos residuos en el sustrato se denominan  $S_3$ ,  $S_2$ ,  $S_1$ ,  $S_1'$ ,  $S_2'$ ,  $S_3'$  (Figura 5) (Schechter & Berger, 1967).

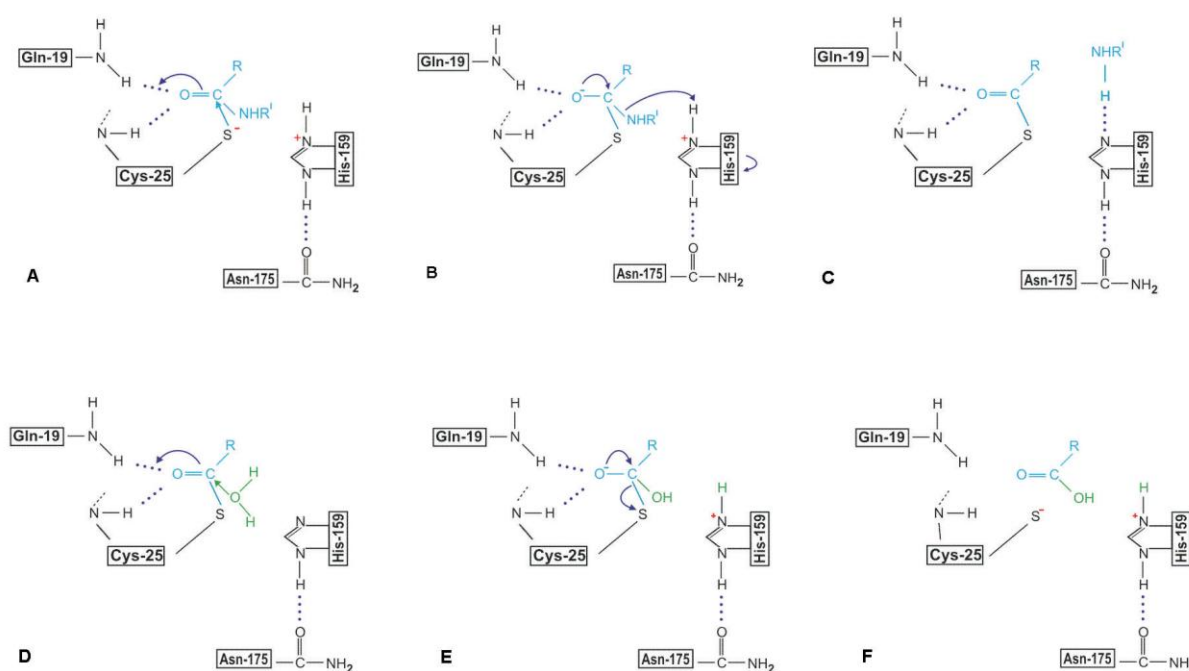


**Figura 5.** Nomenclatura de Schechter & Berger de los subsitios de las proteasas que interactúan con el sustrato, la flecha indica el enlace peptídico que es hidrolizado (modificado de Smooker, et al. 2010)

Mientras algunas proteasas muestran una especificidad restrictiva (ej. la tripsina hidroliza específicamente sustratos con lisina y arginina en  $P_1$ ), otras son capaces de hidrolizar una amplia gama de sustratos y muestran mayor flexibilidad con respecto a los aminoácidos

con los que interacciona cada subsitio. Las endoproteasas de la familia de la papaína tienen en general una especificidad amplia, y son capaces de hidrolizar la mayoría de los enlaces peptídicos que se encuentran accesibles en la estructura tridimensional de las proteínas. Se han descrito preferencias en los residuos que se encuentran a ambos lados del sitio de corte del sustrato, que modulan la eficiencia de las distintas enzimas de la familia, lo que podría haber permitido una adaptación a la hidrólisis de diferentes sustratos naturales a medida que evolucionaron las familias multigénicas en los distintos organismos.

El mecanismo catalítico de las cisteína proteasas involucra una díada catalítica cisteína - histidina, y un residuo de ácido aspártico que asegura la orientación correcta del anillo imidazol de la histidina. La activación de la cisteína involucra la transferencia de un protón entre el grupo tiol de la cisteína y la histidina, formando el par nucleofílico tiolato-imidazol. El primer paso en la catálisis es el ataque nucleofílico del anión tiolato al carbono carbonílico del enlace peptídico a ser hidrolizado formando un intermediario tetraédrico. Este intermediario alcóxido es estabilizado mediante la formación de enlace de hidrógeno con un residuo de glutamina en el denominado bolsillo oxianiónico. En el paso siguiente el grupo alcóxido revierte a carbonilo con la concomitante eliminación del grupo amina del extremo C-terminal del sustrato. Este paso involucra una reacción ácido base entre la amina saliente y el anillo imidazólico de la His el cual queda en su forma neutra no protonada y la amina es eliminada como R-NH<sub>2</sub>. Como consecuencia del segundo paso, la enzima resulta acilada unida a la porción N-terminal del



**Figura 6. Mecanismo catalítico de las cisteína proteasas** (la numeración de los residuos corresponde a la papaína). Tomado de Rzychon et al. 2004.

sustrato mediante un enlace tioéster. En la siguiente etapa la enzima acilada es hidrolizada por una molécula de agua, liberándose el extremo N-terminal del sustrato y reconstituyéndose la enzima activa (Figura 6) (Rzychon, et al. 2004).

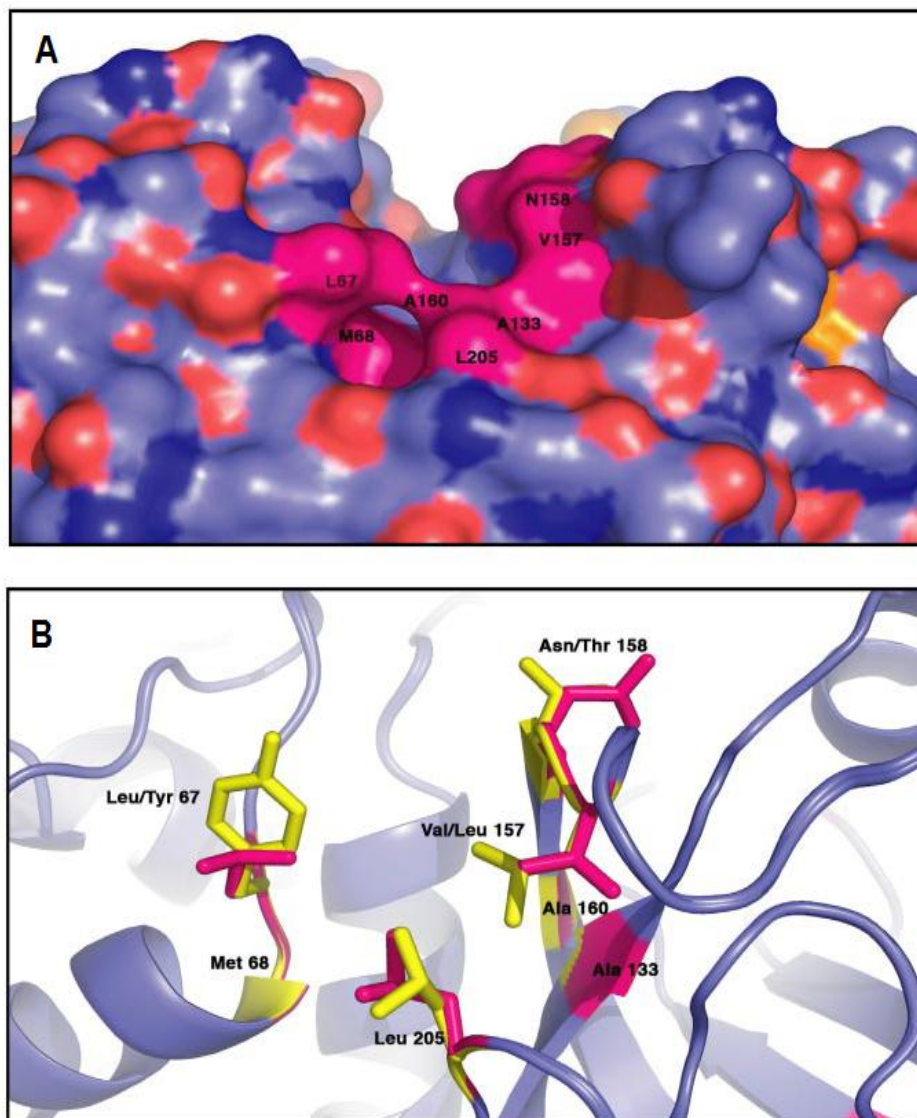
### 3.3. Rol del bolsillo $S_2$ en la especificidad de sustrato

La estructura de varias enzimas de esta familia (papaína, catepsinas humanas, cruzipapaína, *FhCL1*, entre otras) ha sido resuelta mediante cristalografía, ya sea de la enzima inactiva unida al propéptido o de la enzima unida a inhibidores, que al imitar la forma en que se une el sustrato, permite estudiar los sitios de unión enzima-sustrato (Choi et al. 1999; Gillmor, et al. 1997; Li et al. 2008; McGrath et al. 1997; Pauly et al. 2003; Stack et al. 2008). Los resultados muestran que los polipéptidos se unen en el sitio activo en una conformación extendida, donde las amidas de los residuos ubicados en la posición  $P_2$ ,  $P_1$  y  $P_1'$  forman una red de enlaces de hidrógeno con residuos conservados en la superficie subyacente de la enzima, mientras las cadenas laterales interaccionan con residuos más expuestos a ambos lados del surco del sitio activo. En el caso del residuo en  $P_3$  los datos indican que su interacción con los residuos correspondientes en el subsitio  $S_3$  ocurre solamente a nivel de las cadenas laterales. Las uniones de los residuos en el subsitio  $S_4$  y los subsitios  $S_2'$ - $S_4'$  son menos precisas. El bolsillo  $S_2$  es el que se encuentra mejor definido, observándose una buena superposición de las estructuras de los complejos enzima-inhibidor (análogos de sustrato), lo cual indica que en esta región el sustrato interacciona con la enzima en forma similar en todos los miembros de la familia. Este bolsillo involucra residuos a ambos lados del surco del sitio activo y tiene en general un carácter hidrofóbico rodeado de residuos alifáticos y metioninas (Turk et al. 1998). Los residuos principales que interaccionan con el sustrato en este subsitio son los ubicados en las posiciones 67, 68, 133, 157, 158, 160 y 205 (numeración de la papaína), de los cuales los residuos 67 y 157 se ubican en la entrada al bolsillo mientras el residuo 205 forma el piso de este (Figura 7). Algunos de estos residuos se encuentran conservados en la mayoría de las enzimas de la familia de la papaína (como Met68, Ala133, Ala160) mientras otros son variables y están involucrados en la especificidad de la interacción entre el residuo en  $P_2$  del sustrato y el subsitio  $S_2$ . En la tabla 1 se muestran los residuos del bolsillo  $S_2$  para algunas enzimas de la familia.

**Tabla 1.** Residuos del bolsillo  $S_2$  del sitio activo de algunas catepsinas

	67	68	133	157	158	160	205
Cat L humana	Leu	Met	Ala	Met	Asp	Gly	Ala
Cat K humana	Tyr	Met	Ala	Leu	Asn	Ala	Leu
<i>FhCL1</i>	Leu	Met	Ala	Val	Asn	Ala	Leu
<i>FhCL2</i>	Tyr	Met	Ala	Leu	Thr	Ala	Leu
<i>FhCL3</i>	Trp	Met	Ala	Val	Thr	Ala	Val
<i>FhCL5</i>	Leu	Met	Ala	Leu	Asn	Gly	Leu





**Figura 7. Representación del bolsillo S<sub>2</sub> de las catepsina L1 y L2 de *Fasciola hepatica*.** A) Representación de la superficie del sitio activo de FhproCL1, el bolsillo S<sub>2</sub> está resaltado en fucsia y los residuos clave implicados en el reconocimiento del sustrato se encuentran señalados indicando la posición de cada uno. B) Representación de la región del sitio activo de proFhCL1 y FhCL2, resaltándose el bolsillo S<sub>2</sub> de ambas enzimas, los residuos de FhCL1 se muestran en fucsia mientras los correspondientes residuos en FhCL2 se marcan en amarillo. Tomado de Stack et al. 2008.

Para estudiar la especificidad de sustrato de las catepsinas se utilizan tanto inhibidores análogos de sustratos, como oligopéptidos sintéticos con diferentes residuos en las posiciones P<sub>4</sub>-P<sub>1</sub> de unión a la enzima. La capacidad de escindir péptidos con arginina en P<sub>2</sub> se utiliza para discriminar entre la actividad de catepsina B y L, ya que las últimas no hidrolizan esta clase de péptidos. Más recientemente se han desarrollado otros métodos para evaluar la especificidad de sustrato como las librerías de péptidos combinadas (PS-SCL) donde pueden estudiarse a la vez todas las combinaciones posibles de los 20 aminoácidos en cada una de las 4 posiciones de interacción con la enzima (P<sub>1</sub>-P<sub>4</sub>) (Choe et al. 2006; Harris et al. 2000).

Estas bibliotecas se han utilizado para evaluar la especificidad de sustrato de un gran número de enzimas de la familia de la papaína, como las catepsinas humanas L, B, V, K, S, F, la bromelaína, la papaína, la cruzipaina, 2 catepsinas B de *S. mansoni*, las catepsinas L1, L2 y L5 de *F. hepatica*, entre otras (Choe et al. 2006; Norbury et al. 2012; Stack et al. 2008). Prácticamente todas las enzimas de la familia (exceptuando la bromelaína) muestran preferencia por aminoácidos hidrofóbicos en la posición P<sub>2</sub> del sustrato. Además, las catepsinas humanas y las parasitarias tienen una especificidad muy similar a la papaína, acomodando principalmente arginina y lisina en la posición P<sub>1</sub> y una amplia variedad de residuos en las posiciones P<sub>3</sub> y P<sub>4</sub> (Choe et al. 2006). Estos estudios también demostraron que las preferencias diferían principalmente en el aminoácido acomodado en S<sub>2</sub>, siendo los residuos aromáticos (fenilalanina, triptofano, tirosina) los elegidos por las catepsinas humanas L y V, mientras las catepsinas humanas S, K y F, y las catepsinas de *F. hepatica* acomodan mejor residuos alifáticos como leucina, valina e isoleucina. Algunas de estas enzimas tienen también la particularidad de hidrolizar péptidos con prolina en P<sub>2</sub> (catepsina K humana y catepsina L2 de *F. hepatica*) o ácido aspártico (catepsina L5 de *F. hepatica*) aunque con menor eficiencia que leucina (Choe et al. 2006; Norbury et al. 2012; Stack et al. 2008).

La mayoría de las catepsinas humanas (catepsinas L, S, K, F, W, V, O, B, C, H y X) son enzimas lisosomales que cumplen funciones en el recambio proteico y son capaces de hidrolizar una amplia gama de sustratos, mostrando poca especificidad (Rawlings et al. 2008). Sin embargo, otras como la catepsina K y la S están involucradas en funciones más específicas y sus sustratos naturales se hallan más acotados. La catepsina K se expresa fuertemente en osteoclastos, las células responsables de la reabsorción ósea (Drake et al. 1996). El 90% de la matriz del hueso está constituido por fibras de colágeno insoluble de tipo I, las cuales necesitan ser degradadas durante el proceso de reabsorción, mientras el colágeno tipo II está presente principalmente en la matriz fibrilar del cartílago articular. Todos los colágenos tienen en común una conformación de triple hélice dextrógira formada por 3 cadenas  $\alpha$ , siendo un homotrímero el colágeno tipo II y un heterotrímero el colágeno tipo I, formado por dos cadenas  $\alpha_1$  y una cadena  $\alpha_2$ . En ambos tipos de colágeno el largo y las propiedades biomecánicas de la molécula son similares. En los extremos de la triple hélice las cadenas  $\alpha$  adoptan una estructura no helicoidal denominada telopéptido. Para que la triple hélice se forme se requiere una glicina cada tres aminoácidos, por lo que la secuencia Gly-X-Y se encuentra repetida en los dominios del colágeno, siendo frecuentemente X e Y residuos de prolina e hidroxiprolina. El colágeno tipo II tiene un alto contenido de hidroxilisina, en comparación con el colágeno I (Gelse, 2003). Las metaloproteasas de la matriz (MMP) y la serina elastasa de neutrófilos eran las únicas proteasas de mamíferos típicas con capacidad de degradar las regiones de triple hélice del colágeno nativo tipo I, mientras las cisteína proteasas únicamente podían hidrolizar los extremos o telopéptidos que representan sólo el 4% de la molécula. Sin embargo, posteriormente se encontró que la catepsina K tiene la particularidad,

única dentro de las catepsinas humanas, de degradar el colágeno tipo I (Garnero et al. 1998) y tipo II (Kafienah et al. 1998) dentro de la región helicoidal, solubilizando así las fibras de colágeno. Se describió también que la actividad colagenolítica de la catepsina K es dependiente de la formación de complejos con las moléculas de condroitín sulfato residentes en el cartílago (Li et al. 2002) y el telopéptido N-terminal del colágeno tipo II aumenta la expresión de la enzima en los condrocitos articulares (Ruettinger et al. 2008). La actividad colagenolítica de la catepsina K se asoció a su habilidad para acomodar prolina en la posición  $P_2$  del sustrato, debido a que la alteración de la especificidad de sustrato de la enzima en el subsitio  $S_2$  elimina su capacidad de hidrolizar el colágeno (Lecaille, Choe, et al. 2002). Ciertos residuos dentro del bolsillo  $S_2$  demostraron ser claves para que la catepsina K acomode residuos alifáticos y prolina en la posición  $P_2$ . La sustitución de la tirosina en la posición 67 por una leucina y la leucina en posición 205 por una alanina (como en la catepsina L) alteran la especificidad del bolsillo  $S_2$  hacia las preferencias de la catepsina L, favoreciendo ahora la aceptación de residuos aromáticos como triptofano, tirosina y fenilalanina (Lecaille, Choe, et al. 2002). Los cambios recíprocos sobre la catepsina L humana (Leu67Tyr/ Ala205Leu) modifican el subsitio  $S_2$  de la enzima de manera que acomoda mejor residuos de leucina y prolina en  $P_2$ , como se determinó mediante hidrólisis de péptidos sintéticos fluorogénicos, sin embargo no adquiere la capacidad de degradar la triple hélice del colágeno nativo, como lo hace la catepsina K (Lecaille et al. 2007). Si bien la catepsina L Leu67Tyr/ Ala205Leu adquiere la capacidad de aceptar residuos de prolina en  $P_2$ , no consigue degradar colágeno dado que no tiene la habilidad de interactuar con el condroitín sulfato u otros glicosaminoglicanos la matriz extracelular del cartílago como la catepsina K (Li et al. 2004; Selent et al. 2007).

Las catepsinas de *F. hepatica* muestran una especificidad de sustrato solapada pero distintiva. Como mencionamos anteriormente, si bien tanto la catepsina L1, L5 como L2 muestran preferencia por la hidrólisis de residuos alifáticos como leucina y valina en  $P_2$ , sólo la catepsina L2 es capaz también de acomodar prolina en esa posición. Al igual que la catepsina K, *FhCL2* puede degradar el colágeno tipo I en la región helicoidal, mientras *FhCL1* sólo es capaz de cortarlo en los extremos telopeptídicos (Stack et al. 2008). Si bien la especificidad de estas enzimas sobre péptidos sintéticos no necesariamente se correlaciona con lo observado en los sustratos proteicos, en el caso de *FhCL1*, el análisis de los péptidos resultantes de la digestión de hemoglobina reveló que la mayoría contenían aminoácidos alifáticos en  $P_2$  en el orden Leu>Val>Ala. Esto concuerda con las preferencias observadas al utilizar péptidos sintéticos y al emplear librerías de péptidos, y junto al hecho de que estos aminoácidos constituyen el 42% de los presentes en la hemoglobina, sugieren que *FhCL1* evolucionó para adaptarse específicamente a cumplir un rol en la cascada proteolítica de degradación de hemoglobina para la adquisición de nutrientes (Lowther et al. 2009).

Se han reportado también algunos datos sobre la importancia de las interacciones en el bolsillo  $S_2$  en la especificidad de sustrato de las catepsinas de *Fasciola sp.* La sustitución de la

leucina 67 del subsitio  $S_2$  en la catepsina L5 por una tirosina (residuo presente en *FhCL2*) resultó en un aumento de la hidrólisis del péptido sintético Tos-Gly-Pro-Arg-AMC ( $k_{cat}/K_M$ ) con respecto a la enzima no modificada (Smooker et al. 2000). Sin embargo, la misma sustitución en la catepsina L1 de *Fasciola* (Leu67Tyr), no resultó en un aumento de la capacidad de la enzima de acomodar prolina en  $P_2$ . De acuerdo con el resultado anterior, esta variante de *FhCL1* tampoco fue capaz de hidrolizar el colágeno tipo I nativo, sugiriendo que en el caso de las enzimas de *F. hepatica* otras interacciones en el bolsillo  $S_2$  están también involucradas en la aceptación de prolina en  $P_2$ . El residuo en la posición 157, que es una leucina en *FhCL2* y 5 y una valina en *FhCL1*, se propuso como candidato para explicar las diferencias de comportamiento observadas al realizar la misma mutación en *FhCL1* y *FhCL5*. Este residuo se encuentra en la entrada del subsitio  $S_2$  opuesto al residuo 67, y se propone que juntos interaccionan con el anillo planar de la prolina para estabilizarlo dentro del bolsillo de *FhCL2*, donde el residuo 67 es tirosina y el 157 leucina. En *FhCL5* el cambio Leu67Tyr determina que la entrada al bolsillo  $S_2$  sea idéntica a *FhCL2*, mientras que la misma sustitución en *FhCL1* hace que Tyr67 quede opuesta a un residuo de valina en la posición 157, el cual al ser un átomo de carbono más corto que la leucina no se extendería lo suficiente dentro del bolsillo para interaccionar y estabilizar el anillo de prolina (Stack et al. 2008).

La principal catepsina secretada por el estadio juvenil de *F. hepatica*, *FhCL3*, presenta los residuos Trp67 y Thr158 en el subsitio  $S_2$ , lo que sugiere podría mostrar una especificidad de sustrato diferente a las proteasas presentes en el estadio adulto.

## **II. HIPÓTESIS Y OBJETIVOS**

## Hipótesis

La expresión estadio diferencial de las catepsinas de *Fasciola hepatica* nos lleva a planteamos la existencia de diferencias en la especificidad de sustrato de éstas. Puesto que *FhCL3* es expresada durante el desenquiste y migración del parásito juvenil, se presume su importancia durante el proceso de invasión del hospedero definitivo. El estudio de la especificidad de sustrato de *FhCL3* en comparación con catepsinas de otros estadios permitirá correlacionarla con su rol funcional en la infección parasitaria.

## Objetivo general

Abordar mediante diferentes estrategias el estudio de la especificidad de sustrato de la catepsina L3 del estadio juvenil de *F. hepatica*, buscando esclarecer su papel biológico en el ciclo de vida del parásito. Realizar ensayos comparativos con las enzimas secretadas por el parásito adulto, examinando las implicancias funcionales de la expresión estadio específica de las catepsinas.

## Objetivos específicos

### 1. Expresión de las catepsinas L de *F. hepatica* como proteínas recombinantes activas

- i. Amplificar las secuencias completas codificantes de las catepsinas de juveniles y clonarlas en el vector de expresión pHIX4-MFa-ScPAS3 de la levadura metilotrófica *Hansenula polymorpha*.
- ii. Sub-clonar la secuencia de la catepsina L1 de adultos en el vector de expresión pHIX4-MFa-ScPAS3.
- iii. Poner a punto las condiciones de cultivo e inducción para expresar las enzimas recombinantes en la levadura *H. polymorpha*.

### 2. Caracterización enzimática de la catepsina L3

- i. Estudiar la especificidad de sustrato frente a péptidos sintéticos fluorogénicos comúnmente hidrolizados por estas enzimas.
- ii. Determinar su estabilidad y actividad a diferente pH.
- iii. Evaluar su actividad sobre proteínas que se postulen como posibles blancos naturales.

### **3. Estudio comparativo de la especificidad de sustrato de las catepsinas L del estadio juvenil y adulto**

- i. Estudiar las propiedades cinéticas de las catepsinas de JRD sobre un panel de oligopéptidos sintéticos fluorogénicos, comparándolas con las catepsinas secretadas por el estadio adulto.
- ii. Comparar la cinética de inhibición de las catepsinas de juveniles y adultos por parte de inhibidores específicos.
- iii. Comparar la actividad de estas enzimas sobre proteínas que se postulan como posibles blancos naturales: proteínas de la matriz extracelular y membranas basales.
- iv. Identificar los péptidos generados y los sitios de corte sobre proteínas que resulten sustratos diferenciales de estas enzimas mediante análisis de los productos de degradación por cromatografía acoplada a espectrometría de masas.

### **4. Identificación de residuos relevantes en la especificidad de sustrato de las catepsinas de *Fasciola hepatica***

- i. Analizar la especificidad de sustrato de las enzimas recombinantes con bibliotecas de péptidos sintéticas.
- ii. Generar mutantes de sitio activo y expresarlos como proteína recombinante activa.
- iii. Evaluar el efecto de los cambios introducidos en su especificidad de sustrato mediante ensayos de actividad con bibliotecas de péptidos sintéticos.
- iv. Determinar la variación de los parámetros cinéticos mediante el uso de oligopéptidos sintéticos fluorogénicos.

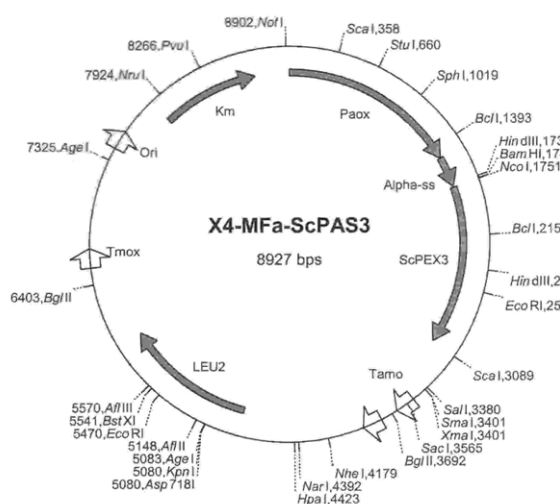
### **III. MATERIALES Y MÉTODOS**



La mayor parte de los Materiales y Métodos se encuentran descritos en los artículos publicados durante la realización de la tesis. En esta sección se brinda un resumen de las metodologías utilizadas, exponiéndose un poco más en detalle las correspondientes al primer objetivo del trabajo.

## 1. Amplificación, clonado y secuenciación de catepsinas L de JRD

La amplificación y clonado de las secuencias completas de las catepsinas de JRD se realizó mediante RT-PCR sobre ARN de JRD según se describe en el artículo correspondiente (Corvo et al. 2009). Varios clones fueron secuenciados en la Unidad de Secuenciación del Institut Pasteur de Montevideo para luego realizar un análisis filogenético utilizando el programa Clustal X (Thompson et al. 1997). Una secuencia representativa codificante de cada una de las variantes génicas mayores halladas en los JRD se sub-clonó de forma direccional en el vector de expresión pHIX4-MF $\alpha$ -ScPAS3 gentilmente cedido por R.J.S. Baerends y J.A.K.W. Kiel (Molecular Cell Biology Lab, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Holanda) (Figura 8). Este vector contiene la secuencia señal del  $\alpha$ -*mating factor* de *S. cerevisiae*, adyacente al sitio de clonado para permitir la exportación de la proteína sintetizada al medio de cultivo. La selección de transfectantes se hace a partir del gen *Leu2* que complementa la auxotrofia de la cepa de *Hansenula* empleada para el crecimiento en medios selectivos.



**Figura 8.** Esquema del vector pHIX4-MF $\alpha$ -ScPAS3 utilizado para la expresión de las enzimas recombinantes en la levadura *H. polymorpha* (el gen de catepsina se clona en lugar del gen ScPEX3, que codifica una proteína del peroxisoma de la levadura).

Brevemente, se extrajo ARN total de parásitos juveniles recientemente desenquistados utilizando el kit RNAqueus-Micro (Ambion), con parásitos cultivados y denquistados *in vitro* en el laboratorio por el Dr. Gabriel Rinaldi. Se sintetizó el ADNc utilizando el sistema SMART PCR cDNA (Clontech). Se amplificaron a partir del ADNc las secuencias codificantes de los genes de catepsinas mediante PCR utilizando oligonucleótidos que adicionan sitios de corte para *Sal*

I (polidT) y *Bam* HI (oligonucleótido degenerado que adiciona los 8 primeros aminoácidos del propéptido) (Corvo et al. 2009) y se clonaron en el vector pCR 4- TOPO. Para generar la mini-biblioteca se transformaron bacterias competentes *E. coli* XL1 mediante el protocolo convencional de *shock* térmico (Sambrook et al. 1989). Se extrajo ADN mediante lisis alcalina de varios clones para secuenciar y en base a esto se seleccionó uno de cada grupo de catepsinas para clonar en el vector pHIX4-MF $\alpha$ -ScPAS3 en marco de lectura con la secuencia de exportación ( $\alpha$ -*mating factor*) empleando las enzimas de restricción antes mencionadas. Tanto el vector como las construcciones con los genes de interés se propagaron inicialmente en bacterias *E. coli* XL1.

## 2. Expresión recombinante de catepsinas L activas

### 2.1. Transfección e integración de plásmido en *Hansenula polymorpha*

#### 2.1.1. Preparación de células electrocompetentes

Se tomó una colonia fresca de una placa en YEPD de *H. polymorpha* NCYC495 y se incubó en agitador termostatzado o.n. a 37 °C y 220 rpm. Se inocularon 2 mL del precultivo en 200 mL de medio YEPD y se incubó en iguales condiciones hasta alcanzar una DO<sub>600nm</sub> entre 1.2 - 1.5. El cultivo se centrifugó 10 min a 3000xg a TA, el *pellet* de células se resuspendió en 50 mL de TED (100 mM Tris, 50 mM EDTA, 25 mM DTT pH 8) y se incubó durante 15 min a 37 °C y 220 rpm. Se centrifugó en las mismas condiciones y se resuspendieron las células en 200 mL de STM (270 mM sacarosa, 10 mM Tris, 1 mM MgCl<sub>2</sub> pH 8) a 4°C. Se repitió la centrifugación y se resuspendió el *pellet* de células en 100 mL de STM a 4°C. Nuevamente se centrifugó, se resuspendió el *pellet* en 0.5 mL de STM frío y se hicieron alícuotas de 50  $\mu$ L que se almacenaron a -80 °C hasta su uso.

#### 2.1.2. Transfección de levaduras mediante electroporación

Se extrajo el ADN plasmídico de las bacterias mediante lisis alcalina siguiendo protocolos descritos (Sambrook et al. 1989) y se linearizó cortando con la enzima *Stu* I (Fermentas), para la incorporación del plásmido al genoma de la levadura una vez transfectadas. Se transfirieron a una cubeta de electroporación de 1 mm de espesor ~10-20 ng de plásmido adicionados a un tubo de células competentes y se aplicó un pulso eléctrico de 0.75 KV, 50  $\mu$ F y 129 Ohm en un electroporador BTX<sup>®</sup>-ECM630. Rápidamente se adicionaron 950  $\mu$ L de YEPD y se transfirieron a un microtubo de 1.5 mL para incubarse 1 h a 37°C y 220 rpm. Se centrifugó 3 min a 3000 x g y se lavó el *pellet* de células con 1 mL de medio YMM (tabla 2). Se realizó una segunda centrifugación, se resuspendieron las células en 200  $\mu$ L de medio YMM y se sembraron con rastrillo sobre placas de YMM - 1.5% agar. Las placas se incubaron en estufa a 37 °C durante 48-72 h en atmósfera húmeda.

### 2.1.3. Integración del plásmido al genoma de *H. polymorpha*

Se integró el plásmido al genoma según el protocolo estándar de Faber, Haima, Harder, Veenhuis, & Ab, 1994 con algunas modificaciones. Brevemente, se realizó un *pool* de 10 colonias de las placas de YMM - 1.5% agar en 20 mL de YEPD, se incubó 8-12 h a 37°C y 220 rpm. Se centrifugó el cultivo 10 min a 3000xg a TA y el *pellet* de células se resuspendió en 30 mL de medio YEPD fresco. Se repitió el cambio de medio y centrifugación 4 veces más, se realizaron diluciones  $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$  y  $10^{-5}$  del último cultivo en YMM y se rastrillaron 100  $\mu$ L de cada dilución en placas YMM – 1.5% agar. Las placas se incubaron en estufa a 37 °C durante 48-72 horas en atmósfera húmeda hasta la aparición de colonias.

### 2.2. Selección de clones que expresan la enzima activa

En las placas de transfección se observaron colonias de diferente tamaño, grandes, medianas y pequeñas. Se realizó un cribaje de actividad cisteína proteasa con sustratos fluorogénicos (ver punto 3) a partir de sobrenadantes de cultivos inducidos de cinco colonias de distintos tamaños. Para cada enzima, las dos colonias con mayor actividad fueron seleccionadas generándose *stocks* en glicerol 30% con un cultivo crecido o.n. en YEPD, almacenándose a – 80 °C.

### 2.3. Optimización de la producción de catepsinas recombinantes

La optimización de la expresión de las catepsinas se efectuó evaluando diferentes medios de inducción para promover la actividad del promotor AOX (tabla 2). Este promotor puede ser des-reprimido con glicerol e inducido con metanol, por lo que la combinación de ambos mecanismos fue utilizada para evaluar su efecto sobre los niveles de expresión de proteína recombinante. La producción de las enzimas recombinantes se evaluó mediante análisis en geles de poliacrilamida, *Western blot* con anticuerpos anti-*FhCL1* y midiendo actividad de hidrólisis de péptidos sintéticos fluorogénicos (ver punto 3). Se trabajó con geles de poliacrilamida al 12% o geles pre-armados NuPage Novex 4-12% Bis-Tris, coloreándose con Coomassie Blue o Flamingo (Tinción fluorescente Biorad). Para los ensayos de *western blot*, las proteínas separadas por SDS-PAGE se transfirieron a membrana de nitrocelulosa y se bloquearon en PBS 1% leche descremada según protocolos convencionales (Sambrook et al. 1989). Para la detección se utilizó suero anti-*FhCL1* producido en oveja 1:500 cedido gentilmente por el Dr. Carlos Carmona (Facultad de Ciencias/Instituto de Higiene, UdelaR) y un anticuerpo secundario anti-IgG oveja producido en cabra conjugado a fosfatasa alcalina (1:5.000) (Sigma).

#### 2.4. Evaluación de la capacidad de autoactivación de proFhCL3

La capacidad de autoprocesamiento de la proenzima se evaluó incubándola 3 h a 37 °C en soluciones tampón de diferente pH: acetato de sodio 0.1M pH 4 o 5, fosfato de sodio pH 6, 7 u 8, conteniendo 1mM DTT y 1mM EDTA. El procesamiento de la proenzima se evaluó mediante SDS-PAGE en geles NuPage Novex 4-12% Bis-Tris (Invitrogen) y tinción con Coomassie Blue. También se evaluó midiendo el aumento en la actividad de hidrólisis del péptido sintético Tos-GPR-AMC.

**Tabla 2.** Medios empleados para el crecimiento y la inducción de la expresión en *H. polymorpha*

Medios	Componentes
YEPD	1% Extracto de levadura
	1% Triptona
	1% Glucosa
YMM	0.67% Base nitrogenada de levaduras
	0.1M Buffer Fosfato de sodio pH6
	2% Glucosa
YEP 1% MetOH	1% Extracto de levadura
	1% Triptona
	1% Metanol
YMM 1% MetOH	0.67% Base nitrogenada de levaduras
	0.1M Buffer Fosfato de sodio pH6
	1% Metanol
YMM 0.5% Gly + 1% MetOH	0.67% Base nitrogenada de levaduras
	0.1M Buffer Fosfato de sodio pH6
	1% Metanol
	0.5% Glicerol

#### 2.5. Estudio del estado de glicosilación de FhCL3 recombinante

La presencia de glicosilación en la enzima recombinante FhCL3 se determinó utilizando un kit de deglicosilación enzimática (Enzymatic Protein Deglycosylation Kit, Sigma). La deglicosilación se llevó a cabo bajo condiciones desnaturizantes de acuerdo a las recomendaciones del fabricante, utilizando una mezcla de las enzimas de deglicosilación PNGasa F, O-glicosilasa y  $\alpha$ -2(3, 6, 8, 9) neuraminidasa. La reacción fue analizada en gel NuPage Novex 4-12% Bis-Tris (Invitrogen) y visualizada con tinción SYPRO Ruby (Molecular Probes) según las instrucciones del fabricante.

### 3. Actividad enzimática con péptidos sintéticos fluorogénicos

La actividad enzimática se midió en tampón fosfato de sodio 0.1 M pH 6, 1 mM DTT, 1 mM EDTA a TA y 20  $\mu$ M de sustrato. Las enzimas se emplearon a concentraciones finales en el rango nM. En un ensayo tipo, se preincubó la enzima 10 min en el tampón, luego se agregó el

sustrato y se midió la actividad en un espectrofluorímetro (*Fluostar*) con excitación a 345 nm y emisión a 440 nm durante 10-20 min. La actividad se midió como unidades de fluorescencia de AMC (UF) liberadas por unidad de tiempo (min o s).

### 3.1. Determinación de la cantidad de enzima activa

La concentración de enzima activa se midió con una titulación con el inhibidor irreversible de cisteína proteasas E-64. Para ello se incubó la enzima con distintas concentraciones de inhibidor (rango nM) durante 30 min, se midió la actividad enzimática residual y se graficó esta actividad (UF/min) en función de la concentración de inhibidor. Se extrapoló la concentración de inhibidor necesaria para que la actividad de la enzima sea cero y este valor se empleó como estimativo de la concentración de enzima activa.

### 3.2. Estimación de los parámetros cinéticos

Para estimar los parámetros cinéticos  $K_M$  y  $V_{max}$  se midió la velocidad de la reacción enzimática (UF/s) en estado estacionario (rango lineal de la curva) a diferentes concentraciones de sustrato (en el rango  $K_M/5 - K_M \times 5$ ). Se prepararon soluciones *stock* de 10 mM de cada sustrato en dimetilformamida al 70% y se almacenaron a  $-20^\circ\text{C}$ . Las diluciones empleadas en los ensayos cinéticos (1 – 200  $\mu\text{M}$ ) se prepararon en agua a partir del *stock*. La concentración de AMC liberado (en  $\mu\text{M}$ ) se calculó empleando una curva estándar de diferentes concentraciones de AMC (UF vs [AMC]), para transformar las unidades de velocidad de UF/s a  $\mu\text{M/s}$ . Para calcular  $K_M$  se realizó una regresión no lineal de la curva de Michaelis-Menten ( $V_{max} * [S] / (K_M + [S])$ ) empleando el programa OriginPro 6.1.  $k_{cat}$  se obtuvo mediante la fórmula  $k_{cat} = V_{max} / [E]$ .

### 3.3. Ensayos de inhibición enzimática

La inhibición enzimática se determinó a partir de la curva de progreso de la hidrólisis de un sustrato peptídico fluorogénico en presencia de diferentes concentraciones de inhibidor. Se utilizaron inhibidores competitivos análogos del sustrato en concentraciones nanomolares. Para ello, se prepararon diluciones seriadas de inhibidor en una placa de 96 pocillos y se incubaron con la enzima en tampón fosfato de sodio 0.1 M pH6, 1 mM DTT, 1 mM EDTA a TA durante 10 - 15 min para permitir la formación del complejo enzima-inhibidor. Luego se agregó el sustrato a una concentración final de 20  $\mu\text{M}$  y se midió la actividad en un espectrofluorímetro (*Fluostar*) con excitación a 345 nm y emisión a 440 nm durante 10-30 min. La actividad se expresó en unidades de fluorescencia de AMC liberadas por segundo (UF/s). En paralelo se midió la actividad de la enzima sin inhibidor incubada 15 min en el mismo tampón. La velocidad inicial de reacción ( $v$ ) se calculó como la pendiente de la curva UF vs tiempo (s) para las diferentes concentraciones de inhibidor. Se graficó la concentración de inhibidor [I] versus

$(v_0/v)-1$ , donde  $v_0$  es la velocidad en ausencia de inhibidor. La pendiente de esta curva es la constante de inhibición aparente  $K_{i(app)}$  en unidades de concentración de inhibidor (nM).  $K_i$  se calculó como  $K_{i(app)} / (1 + [S] / K_M)$  (Izquierdo-Martin & Stein 1992).

#### 4. Análisis de la digestión de sustratos proteicos

La digestión de proteínas de la matriz extracelular se realizó *in vitro* mediante incubación del sustrato con la enzima en un tampón adecuado a diferente pH (PBS pH 7.3, acetato de sodio 0.1M pH 5.5) con 1mM DTT y 1mM EDTA, a 28°C o 37°C durante diferentes tiempos. Las reacciones se detuvieron agregando 10  $\mu$ M de E-64 y los productos de degradación se separaron por SDS-PAGE en geles de poliacrilamida o geles NuPage Novex 4-12% Bis-Tris (Invitrogen). Los geles se tiñeron con Coomassie blue o tinción Flamingo (BioRad). Se utilizó fibronectina humana (Sigma), laminina I de ratón (Invitrogen), colágeno tipo I de piel de ternero (Calbiochem) o de cola de rata (gentilmente cedido por la Dra. Silvia Chifflet, Dpto. de Bioquímica, Facultad de Medicina, UdelaR), colágeno tipo II de septo nasal bovino (Sigma) e IgG1 humana (gentilmente cedida por la Dra. Patricia Berasain de la Unidad de Biología Parasitaria de la Facultad de Ciencias, UdelaR).

##### 4.1. Identificación de péptidos de digestión del colágeno por HPLC-MS

Las reacciones de digestión del colágeno tipo I fueron centrifugadas 15 min a 10.000 x g para remover precipitados y concentradas a un volumen de 15  $\mu$ l (~7 veces) utilizando un centricon Eppendorf de 5 kDa de corte. Los péptidos se analizaron por nano LC ESI-MS/MS usando un sistema Tempo nanoLC (Eksigent, USA) con una columna IntegraFrit empaquetada con una resina C18 (ProteoPrep II) acoplada a un espectrómetro de masas QSTAR Elite Q-TOF (AB Sciex). Se determinó la masa de los péptidos precursores ionizados y de los resultantes de la fragmentación de algunos iones precursores. La lista de masas generada por el programa MSX (Infochromics) se exportó al programa PEAKS Studio 5.0 (Bioinformatics Solutions Inc.). Se generó una base de datos conteniendo únicamente la secuencia del colágeno I bovino con la cual se realizó una búsqueda para asignar a las masas peptídicas la secuencia de aminoácidos correspondiente en el colágeno I. La modificación de cisteínas con propionamida se estableció como parámetro fijo y la oxidación de metioninas como parámetro variable. La tolerancia en la masa fue establecida en 100 ppm para los péptidos precursores y 0.2 Da para los fragmentados. Los péptidos que se identificaron con un *score* mayor al 60% fueron seleccionados y mapeados sobre la secuencia de aminoácidos de las cadenas  $\alpha$  1 y 2 del colágeno para determinar los sitios de corte de las enzimas. Estos ensayos se realizaron en el Instituto para la Biotecnología de Enfermedades Infecciosas de la Universidad Tecnológica de Sidney, Australia.

## 5. Actividad con bibliotecas de péptidos sintéticos

La determinación de la especificidad de sustrato de las enzimas recombinantes y las variantes generadas por mutagénesis sitio dirigida con las bibliotecas peptídicas se realizó en el Laboratorio de Patología del Centro para el Descubrimiento e Innovación en Enfermedades Parasitarias en la Universidad de San Francisco, Estados Unidos.

### 5.1. Ensayos con la librería PS-SCL

En una placa de 96 pocillos Microfluor de fondo plano se agregaron 0.5  $\mu$ l (12.5 nmoles) de cada una de las 20 sublibrerías. La reacción se inició con la adición de la enzima en 100  $\mu$ L de tampón fosfato de sodio 0.1M pH 6, 1 mM DTT, 1 mM EDTA, 0.01% PEG-6000 y 0.5% Me<sub>2</sub>SO (del sustrato) a 25 °C y se monitoreó con un espectrómetro de fluorescencia SpectraMax Gemini (Molecular Devices) con excitación a 380 nm, emisión a 460 nm y un corte de 435 nm. Cada enzima se ensayó por triplicado.

### 5.2. Ensayos con la librería MSP-MS

Comparamos la actividad enzimática de *FhCL1* y *FhCL3* con la biblioteca *MSP-MS*. Cada uno de los 3 *pools* (52, 52 y 20 tetradecapéptidos) de la librería se diluyó a 1  $\mu$ M en tampón fosfato de sodio 25 mM, pH 6, 1 mM DTT y 1 mM EDTA. Se agregó la enzima activa *FhCL1* o *FhCL3* en el mismo tampón de manera que la concentración final de cada enzima fue de 10 nM y se incubó a TA tomando alícuotas a los 5, 15, 60, 240 y 1200 min de incubación. La reacción se detuvo agregando ácido fórmico pH 3 (4% final), se eliminó el tampón por evaporación y se resuspendió en un volumen de ácido fórmico 0.1% igual al inicial. Como control se incluyó un tubo en iguales condiciones pero sin enzima. La digestión de los péptidos se analizó por LC-MS/MS inyectando 10  $\mu$ l de cada alícuota en una columna C18 (Michrom Bioresources) conectada a un espectrómetro de masas Thermo Finnigan LTQ. Se generó una lista de masas a partir de los datos crudos utilizando el programa PAVA (UCSF) para buscar la secuencia correspondiente con el programa Protein Prospector en una base de datos conteniendo los 124 tetradecapéptidos. Los péptidos generados por actividad de las enzimas se identificaron en comparación con los del ensayo control.

## 6. Mutagénesis sitio dirigida

Las sustituciones de aminoácidos en el sitio activo de *FhCL3* y *FhCL1* se realizaron por mutagénesis sitio dirigida con el kit QuikChange Site-Directed Mutagenesis (Stratagene). Se diseñaron diferentes pares de oligonucleótidos conteniendo las sustituciones de bases a introducir en el gen de la cathepsina según las instrucciones del fabricante (Stratagene) y fueron sintetizados por la compañía IDT (<http://www.idtdna.com/site>). Se realizó un PCR con los oligonucleótidos diseñados para introducir las mutaciones deseadas empleando como molde

ADN del plásmido pHIX4-M $\alpha$  conteniendo los genes de pro*FhCL1* o pro*FhCL3*. Las variantes dobles se realizaron usando como molde los plásmidos con las mutaciones simples. Los plásmidos con los genes mutados se propagaron en bacterias *E. coli* Top10, se preparó ADN mediante miniprep alcalina y se secuenció en la Unidad de Secuenciación del Institut Pasteur de Montevideo para confirmar la presencia de las sustituciones de bases. Con los plásmidos se transformaron levaduras competentes de *H. polymorpha* como se describió en la sección 2.1. y se obtuvieron las enzimas mutantes activas induciendo la expresión de las levaduras transfectadas en las mismas condiciones que *FhCL1* y *FhCL3*.

## 7. Modelado molecular de *FhCL3*

El modelado espacial de *FhCL3* se realizó mediante superposición con la estructura de *FhCL1* usando las coordenadas de pro*FhCL1* (2O6X) como molde principal con el programa Swiss-Model (Schwede et al. 2003). Para comparar con las otras cisteína proteasas se incluyó la porción madura de la enzima. El modelo de *FhCL2* se obtuvo del Swiss Model Repository (<http://swissmodel.expasy.org/repository/>). Los modelos se visualizaron y compararon por superposición con el programa DeepView Swiss-Pdb Viewer. Las otras estructuras empleadas en la comparación se obtuvieron de la base de datos Protein Data Bank (<http://www.pdb.org/>): catepsina L humana (1MHW), catepsina K humana (1ATK) y zingipaína del jengibre (1CQD). En algunos casos el inhibidor E64 se superpuso sobre las estructuras de las catepsinas para facilitar la visualización de los residuos que forman el sitio activo. Estos residuos se identificaron basándonos en la literatura y se confirmaron por alineamiento estructural con las catepsinas L y K humanas y la papaína (5PAD). Se utilizó la función *Mutate* del programa PDBViewer para generar *in silico* el cambio W67L y los rotámeros sobre el modelo de *FhCL3*. Estos se seleccionaron de acuerdo al *score* del rotámero y mediante inspección visual.



#### **IV. RESULTADOS**

## 1. Expresión de catepsinas L recombinantes activas en la levadura *Hansenula polymorpha*

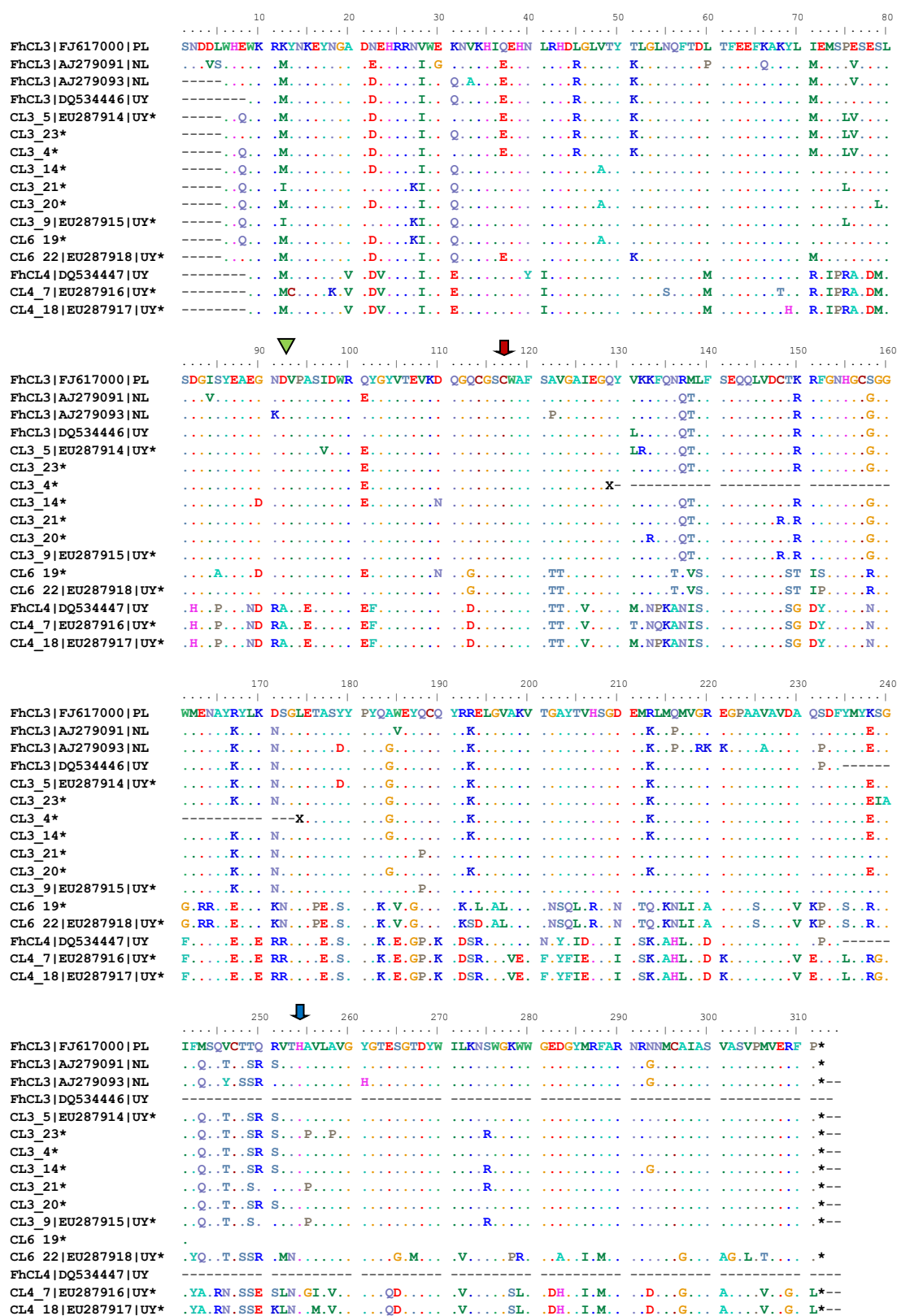
### *Estrategia general*

La purificación de las enzimas secretadas por los JRD se ve obstaculizada por la escasez de material biológico disponible y porque las catepsinas constituyen una familia multigénica de enzimas con características de peso molecular y punto isoeléctrico similares, y por lo tanto difíciles de separar por técnicas de purificación convencionales. Por consiguiente, nos propusimos el clonado y expresión recombinante de las catepsinas en un sistema adecuado. Las catepsinas maduras expresadas en sistemas bacterianos resultan inactivas por el plegamiento incorrecto del precursor (Smith & Gottesman 1989). Por este motivo se utilizan los sistemas eucariotas, donde existe una vía secretoria, para obtenerlas en forma activa a partir de ADNc de la proenzima (Cappetta et al. 2002). Nos planteamos la optimización de un sistema de expresión en la levadura metilotrófica *H. polymorpha*, dado que se encontraba disponible en el laboratorio y presenta ventajas sobre la expresión en *S. cerevisiae* y *Pichia pastoris*. Tanto en *Pichia* como en *Hansenula* la expresión de las proteínas recombinantes se induce utilizando metanol como fuente de carbono ya que está regulada por el promotor de la alcohol oxidasa 1 (AOX1). Los plásmidos que se utilizan para la expresión en *P. pastoris* están diseñados para integrarse mediante recombinación al gen de la AOX, mientras en *H. polymorpha* se integran al azar por recombinación no-homóloga en cualquier loci. Esto determina que el número de copias del gen a expresar no supere en general las 10 en *P. pastoris* mientras en *H. polymorpha* pueden integrarse hasta 150 copias. Además en *H. polymorpha* el promotor AOX puede ser inducido tanto mediante la adición de metanol al medio como por des-represión con glicerol, eliminando el peligro de la manipulación de metanol durante la fermentación cuando se trabaja a gran escala (<http://www.biotechresources.com/hansenula-yield.shtml>). *Hansenula* tiene la ventaja de ser muy eficiente en la secreción de proteínas al medio y muy pocas de sus proteínas endógenas son secretadas durante el cultivo. Si bien la catepsina L1 de *F. hepatica* se había expresado anteriormente por nuestro grupo en *S. cerevisiae* (Roche et al. 1997), su rendimiento era muy bajo por lo que decidimos sub-clonarla en el vector de *H. polymorpha* buscando aumentar los niveles de expresión.

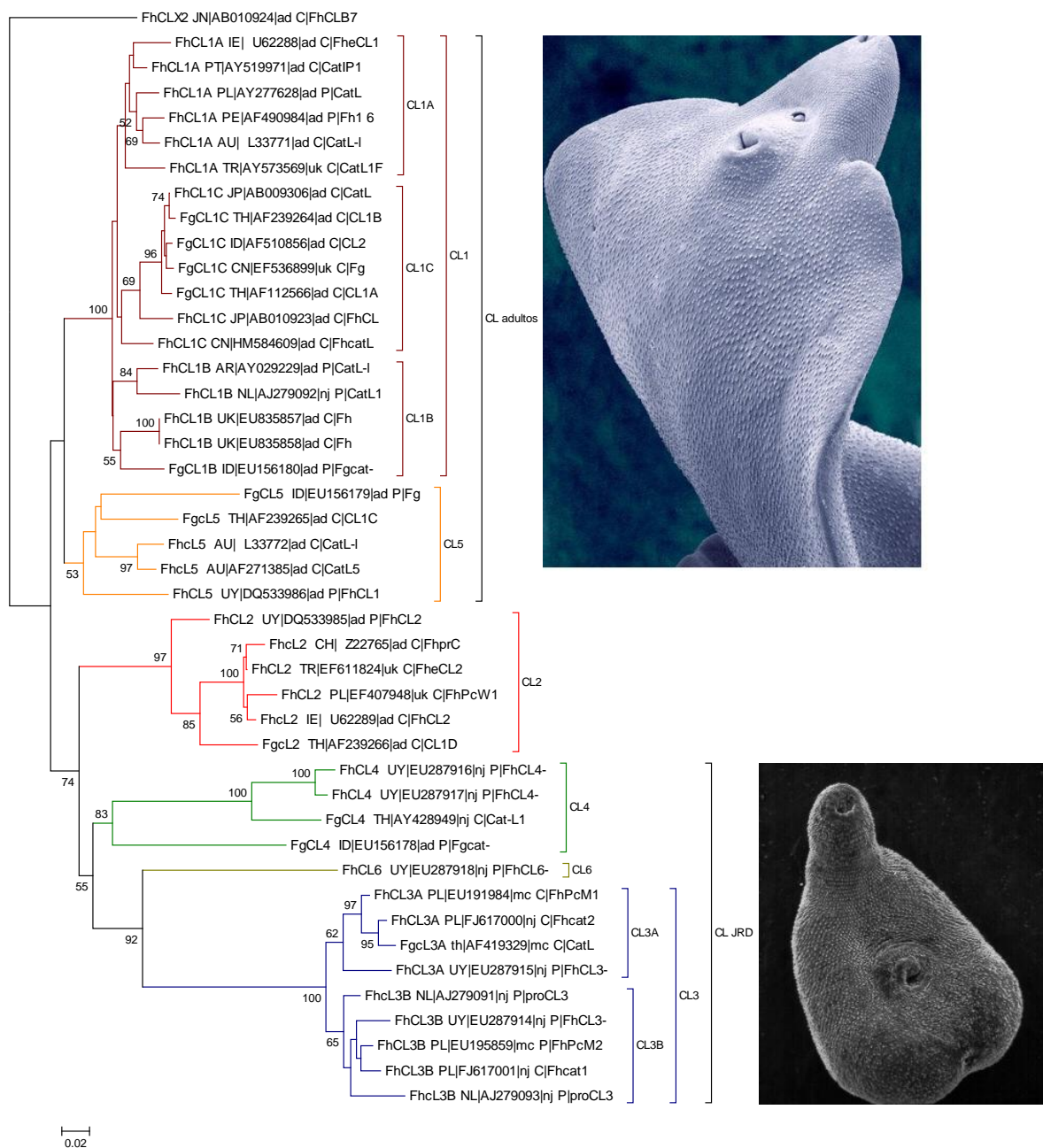
Como se describe a continuación, con este sistema se lograron buenos niveles de expresión de las enzimas recombinantes, que permitieron la caracterización de la enzima de juveniles *FhCL3*.

### 1.1. Identificación de catepsinas L del estadio juvenil

Las catepsinas L expresadas por los JRD se amplificaron por RT-PCR con oligonucleótidos diseñados en base a secuencias conocidas y se generó una minibiblioteca de ADNc para la propagación y secuenciación de los genes. El alineamiento y análisis filogenético de las secuencias obtenidas mostró que estas se distribuyen en 3 grupos: la mayoría se agrupa con las secuencias previamente reportadas para la catepsina L3 de *F. hepatica* y *F. gigantea*, mientras las restantes pertenecen al grupo de la catepsina L4 descrita por (Cancela et al. 2008) o forman un grupo separado de éstos, no descrito anteriormente, al cual denominamos *FhCL6* (Figuras 9 y 10). La secuencia aminoacídica de *FhCL6* tiene 75% de identidad con *FhCL3* y 64% de identidad con *FhCL4*, evidenciando la existencia de una familia multigénica propia del estadio juvenil. En el alineamiento se observa que la cisteína del sitio activo (flecha) se encuentra conservada en todas las catepsinas, la variación entre grupos se encuentra mayoritariamente en los últimos 20 aminoácidos del propéptido y en la enzima madura (flecha invertida) exceptuando los primeros 40 residuos del extremo N-terminal (Figura 9). La comparación de las secuencias de *FhCL3* obtenidas con las reportadas de parásitos aislados en Polonia (PL) y Holanda (NL) denota variabilidad intragrupo, independientemente del sitio de origen. Esto sugiere la presencia de más de un gen codificante de *FhCL3*, además de varios alelos. En el árbol filogenético puede verse cómo las secuencias correspondientes a *FhCL3* forman 2 grupos monofiléticos denominados *FhCL3A* y B (Figura 10). Para expresar las enzimas recombinantes se seleccionó una secuencia representativa de cada grupo (CL3\_14, CL4\_7 y CL6\_22, Figura 9).



**Figura 9. Alineamiento de las secuencias de catepsinas L de juveniles de *Fasciola hepatica*.** Se representa la identidad de las secuencias de la proenzima amplificadas durante este trabajo (\*), con respecto a una secuencia completa de FhCL3 (FJ617000) con diagrama de puntos utilizando el programa *Bioedit*. Se incluyen las secuencias parciales de FhCL3 y FhCL4 obtenidas previamente en nuestro laboratorio (Cancela *et al.* 2008; DQ534446 y DQ534447) y las secuencias de FhCL3 reportadas por Hamsen *et al.* 2004 (AJ279091 y AJ279093). Se indica el número de acceso de las secuencias que fueron depositadas en la base de datos GenBank: CL3 5 y 9, CL6 22, CL4 7 y 18, el sitio de inicio de la enzima madura (▽), la cisteína del sitio activo (⇓) y la histidina que forma la diáda catalítica (⇓).



**Figura 10. Árbol filogenético de las catepsinas L de los estadios adulto y juvenil de *Fasciola* sp.** Con secuencias representativas de todos los grupos de catepsinas L de *Fasciola* sp. depositadas en la base de datos Nucleotídica del *GenBank* se construye una filogenia con el método *neighbor-joining* con *bootstrapped* (1000-intentos) utilizando el programa MEGA5 (Tamura et al. 2011). Se indica la especie: Fh, *F. hepatica* o Fg, *F. gigantica*, el origen, AR: Argentina, AU: Australia, CH: Suiza, CN: China, IE: Irlanda, ID: Indonesia, TH: Tailandia, JP: Japón, TR: Turquía, NL: Holanda, PL: Polonia, UK: Inglaterra y UY: Uruguay; el número de acceso del *GenBank* y el nombre publicado originalmente en el artículo o secuencia depositada.

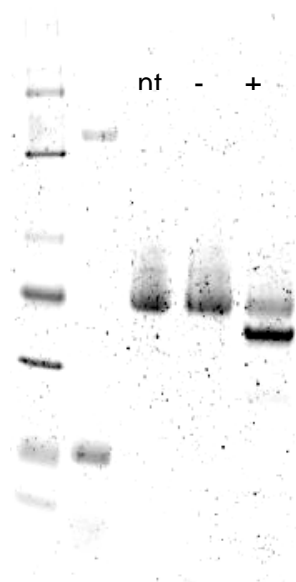
### 1.2. Expresión de *FhCL1* y *FhCL3* recombinantes

Para la expresión se utilizó el sistema de *H. polymorpha*, obteniéndose las enzimas recombinantes secretadas en el medio de cultivo. En el análisis de los sobrenadantes de cultivo de las levaduras crecidas en medio de inducción mediante SDS-PAGE se observa en el caso del clon que expresa pro*FhCL3* una banda de peso molecular aparente entre 38 kDa y 49 kDa, tamaño un poco mayor al esperado para la proenzima. El análisis mediante *Western blot* muestra que la banda es reconocida por sueros anti-*FhCL1*, y que migra más lentamente que la correspondiente a pro*FhCL1*. A su vez en el caso de *FhCL1* la enzima migra como varias bandas que por su tamaño parecen corresponder a formas oligoméricas de la proteína, como ya fuera descrito en trabajos previos (Dalton & Heffernan 1989) (Figura 11).



**Figura 11. Western blot con suero anti-*FhCL1*.** Carril 1: 10  $\mu$ l MPM (kDa) SeeBlue Plus2 (Invitrogen); carril 4: 20  $\mu$ l sdtte 10X *FhCL3* medio YEP 1%MetOH; carril 7: 20  $\mu$ l sdtte 10X *FhCL3* medio YMM 1%MetOH; carril 10: 2 $\mu$ l *FhCL1* sin activar a 1.2 mg/ml.

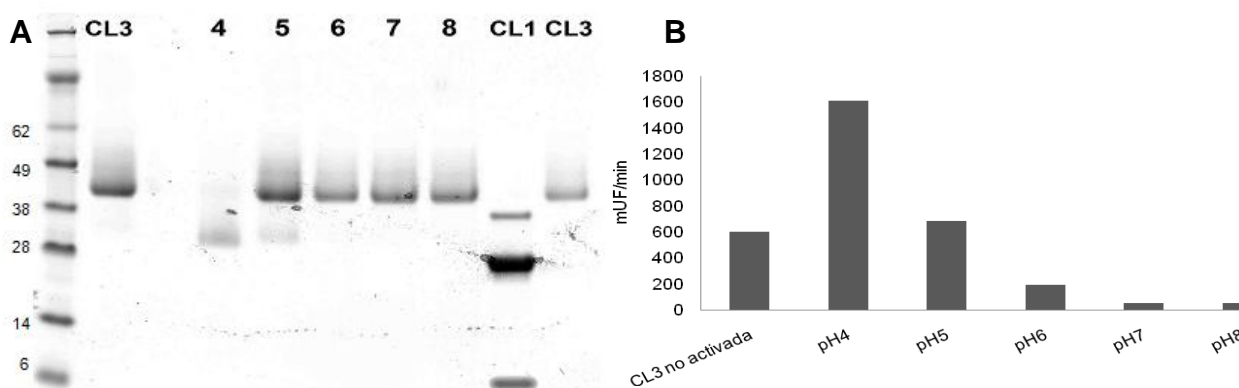
La identidad de *FhCL1* y *FhCL3* se confirmó por espectrometría de masas. La migración de la banda correspondiente a *FhCL3* se modifica al incubarla con enzimas deglicosilantes, indicando que la enzima recombinante está siendo glicosilada durante su pasaje por el sistema retículo endoplasmático y Golgi de la levadura (Figura 12). Esto explica el peso molecular mayor al esperado de acuerdo a su secuencia de aminoácidos (42 kDa vs 36 kDa). Asimismo, la enzima madura también tiene un peso molecular de aproximadamente 4 kDa más que el esperado (29 kDa vs 25 kDa) (Figura 13A).



**Figura 12. Ensayo de deglicosilación de FhCL3.** FhCL3 no tratada con deglicosilasas (nt), FhCL3 incubada en *buffer* de reacción con (+) o sin (-) enzimas de deglicosilación. Gel teñido con Sypro Ruby (Molecular Probes).

### 1.3. Activación de *proFhCL3* a la enzima madura

Durante la expresión de *FhCL3* se observó que una parte de la enzima secretada se procesa automáticamente a su forma madura, pues es posible detectar actividad directamente en el sobrenadante del cultivo, y se purifica una banda de peso molecular similar al estimado para la enzima madura (Corvo et al. 2009). La mayor parte se expresa en forma de proenzima, la cual es capaz de activarse autocatalíticamente a la forma madura, al incubarse durante 3 h a 37°C en tampón acetato pH ácido, mayormente a pH 4, y en menor medida a pH 5 (Figura 13).



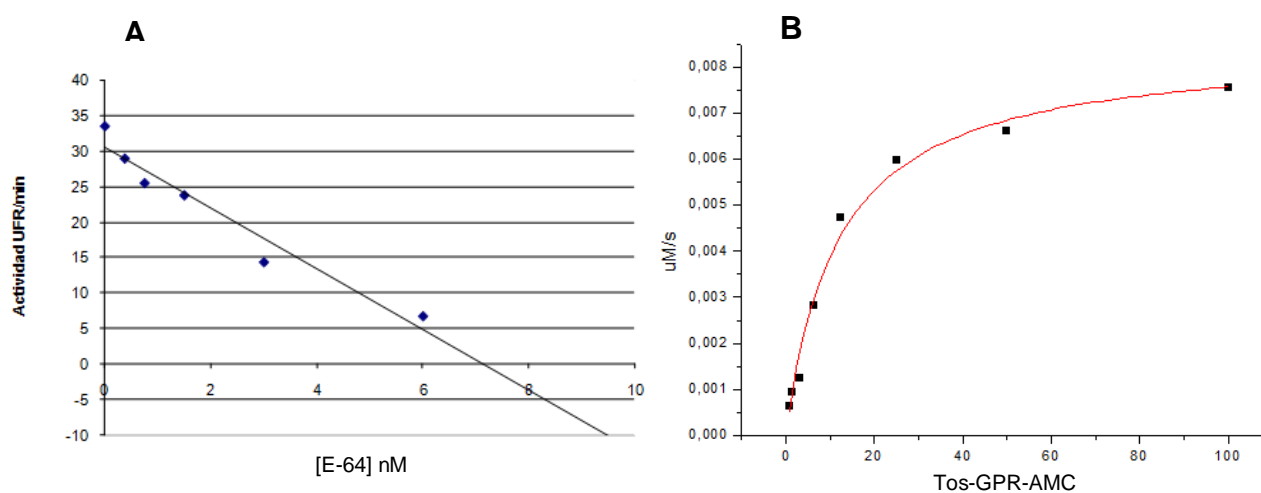
**Figura 13. Ensayo de activación de *proFhCL3* a pH 4, 5, 6, 7 y 8. A)** NuPage Novex 4-12% Bis-Tris teñido con Coomassie blue; CL3: enzima sin activar; CL1: enzima activada. MPM en kDa SeeBlue Plus2 (Invitrogen). **B)** Ensayo de actividad de FhCL3 sin activar e incubada 2h a 37°C en buffer de distinto pH (la tasa de hidrólisis del péptido Tos-GPR-AMC se expresa como mUF/min).

La titulación de la enzima activa se realizó con el inhibidor irreversible E-64 obteniéndose en todas las preparaciones concentraciones en el entorno de 10-50  $\mu$ M de *FhCL3* activa (Figura 14 A). La concentración de enzima activa se empleó en el cálculo de los parámetros

cinéticos de las catepsinas frente a distintos sustratos como se describe en los artículos publicados durante la tesis (una curva obtenida para *FhCL3* con el sustrato Tos-GPR-AMC se presenta como ejemplo en la Figura 14 B).

#### 1.4. Optimización de la expresión de catepsinas recombinantes

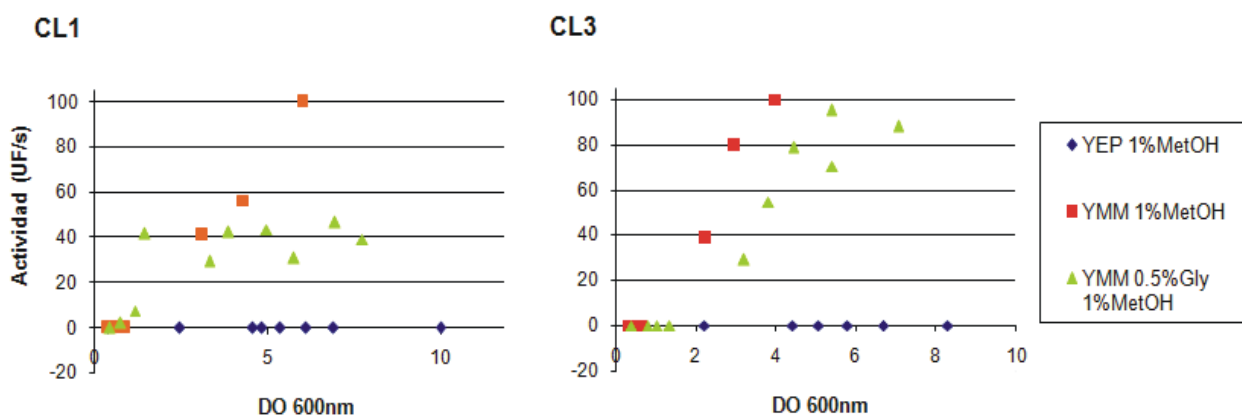
La optimización de la expresión de las enzimas recombinantes en la levadura *H. polymorpha* se realizó utilizando diferentes condiciones de inducción, variando tanto la composición del medio de inducción (Tabla 2) como la temperatura pre y post inducción (30 °C o 37 °C) y los tiempos de cultivo. Se siguió el crecimiento celular midiendo la densidad óptica del cultivo a 600 nm, evaluando en paralelo la cantidad de enzima secretada al medio mediante un ensayo de actividad enzimática. A modo de ejemplo, en la Figura 15 se compara la expresión de *FhCL1* y *FhCL3* en diferentes medios. En ambos casos, no se detecta actividad enzimática cuando la inducción se realiza en medio rico, pero si en YMM, siendo el medio conteniendo únicamente 1% MetOH el que mejor se expresan ambas catepsinas. Sin



**Figura 14. (A) Titulación de FhCL3 con el inhibidor E-64.** Se grafica la actividad enzimática en unidades relativas de fluorescencia (UFR) del producto (AMC) producido por minuto en función de la concentración de inhibidor. **(B) Cálculo de los parámetros cinéticos de FhCL3 para el sustrato Tos-GPR-AMC.** Se utilizó el programa OriginPro 6.1 para graficar velocidad inicial ( $\mu\text{M}$  AMC liberado por segundo) en función de la concentración de sustrato ( $\mu\text{M}$ ).

embargo, en el resultado del *Western blot* se observa que en ambos medios se está produciendo la enzima en cantidades similares (Figura 11). El análisis mediante SDS-PAGE del sobrenadante de los cultivos inducidos mostró que a partir de las 36 h de incubación comenzaban a aparecer otras proteínas en el medio, sin un enriquecimiento de la banda que correspondería a las catepsinas. Un aumento de la secreción de otras proteínas se observó también al realizar la inducción a 37 °C, por lo que la temperatura de 30 °C se definió como la más adecuada.





**Figura 15. Expresión de *FhCL1* y *FhCL3* en diferentes medios de inducción.** Se representa la actividad enzimática, medida directamente en el sobrenadante de los cultivos, en función del crecimiento celular (densidad óptica a 600nm). La actividad representa la hidrólisis del péptido sintético Z-VLK-AMC para *FhCL1* y Tos-GPR-AMC para *FhCL3*.

En base a estas observaciones se estableció el siguiente protocolo para producir las catepsinas recombinantes en la levadura *H. polymorpha*:

- Día 1: Inocular 500 mL de YEPD con 0,5mL de stock de glicerol de *H. polymorpha*. Utilizar un Erlenmeyer de 2 o 2.5L. Incubar a 30°C 200 rpm durante 24 h o hasta que el cultivo alcance una  $DO_{600nm}$  de 2-4.
- Día 2: Centrifugar las células 10 min a 3000 x g, descartar el sobrenadante y resuspender el *pellet* en 100 mL de YMM 1%MetOH. Utilizar un Erlenmeyer de 1 L.. Incubar a 30 °C 200 rpm durante 24 h.
- Día 3: agregar 1 mL de MetOH 10% e incubar durante 10-12 h más en las mismas condiciones. Detener el cultivo, centrifugar 20 minutos a 3000 x g 4 °C, separar el sobrenadante y almacenarlo a 4 °C.

### 1.5. Expresión de las otras catepsinas de JRD

Los genes de las catepsinas *FhCL4* y *FhCL6* fueron también clonados en el vector de expresión, pero no se detectó actividad cisteína proteasa ni se observó una banda del tamaño esperado en SDS-PAGE en los sobrenadantes de cultivo inducidos en las mismas condiciones que las enzimas *FhCL1* y *FhCL3*. Se procedió entonces a verificar la secuencia de los genes clonada en el vector de expresión, determinándose que ambos se encontraban correctamente en marco de lectura con la señal de exportación del  $\alpha$ -*mating* factor de la levadura y se correspondían con la secuencia esperada. Para descartar que la enzima se produjera pero no fuese secretada al medio se examinaron los extractos somáticos de levaduras inducidas mediante SDS-PAGE y *Western blot* con anticuerpos anti-*FhCL1*, no detectándose ninguna banda del tamaño esperado para las catepsinas.

## 2. Caracterización enzimática de la catepsina L3

### *Estrategia general*

En este trabajo iniciamos la caracterización de la proteína recombinante *FhCL3* expresada en su forma activa en la levadura *H. polymorpha*. Se estudio su actividad y estabilidad en presencia de agentes reductores, en un rango de diferentes pH.

Para abordar el estudio de la especificidad de sustrato de las cisteína proteasas se emplearon péptidos sintéticos conjugados a un fluoróforo. La habilidad de las enzimas para hidrolizar eficientemente cada péptido puede evaluarse estimando los parámetros cinéticos de la reacción de hidrólisis: la constante de afinidad ( $K_M$ ), la constante catalítica ( $k_{cat}$ ) y la constante de especificidad ( $k_{cat}/K_M$ ). Para las proteasas de la familia de la papaína se utilizan este tipo de sustratos para evaluar preferencias en los subsitios  $S_1$ - $S_4$  del sitio activo. Estos péptidos proveen una manera sencilla de discriminar por ejemplo las actividades de las catepsinas tipo B y tipo L, ya que las primeras hidrolizan péptidos con arginina en  $P_2$  (como Z-Arg-Arg-AMC) mientras las catepsinas L no son capaces de escindir esta clase de sustratos, siendo los péptidos con residuos hidrofóbicos, como fenilalanina o leucina en  $P_2$ , hidrolizados por ambos grupos de enzimas. A su vez, los péptidos con prolina en  $P_2$  permiten diferenciar la actividad de algunas catepsinas tipo L, como la catepsina K humana o *FhCL2* de adultos, dado que las demás catepsinas lisosomales humanas y *FhCL1* hidrolizan muy pobremente péptidos como tos-GPR-AMC (Dowd et al. 1994; Roche et al. 1997).

Calculamos los parámetros cinéticos de *FhCL3* frente a tres péptidos fluorogénicos con diferentes residuos en  $P_2$ , mostrando que la enzima tiene una especificidad diferente a las catepsinas producidas por el estadio adulto. Dado que *FhCL3* mostró una marcada especificidad hacia el péptido tos-GPR-AMC nos planteamos evaluar su capacidad de hidrolizar las moléculas de colágeno nativo, ya que está reportado que la capacidad de la catepsina K humana de degradar el colágeno se correlaciona con su actividad sobre este clase de péptidos (Lecaille, Choe, et al. 2002). En paralelo, evaluamos su capacidad de hidrolizar inmunoglobulina G (IgG), en comparación con *FhCL1* que es capaz de cortar la región bisagra de distintas subclases de IgGs (Berasain et al. 2000).

Hasta este momento, dentro de las catepsinas de *Fasciola* y humanos, solo *FhCL2* y la catepsina K humana estaban reportadas como capaces de hidrolizar péptidos con prolina en  $P_2$ , si bien *in vitro* ambas presentan mayor actividad sobre péptidos con  $P_2$ -leucina. El hecho de que *FhCL3* hidrolice con mayor especificidad tos-GPR-AMC resulta novedoso para una catepsina de la familia de la papaína. Una búsqueda de enzimas con actividad similar nos llevó a dos cisteína proteasas del jengibre, una de ellas ya cristalizada (Choi et al. 1999), que presentan la misma especificidad hacia prolina que *FhCL3*. Basándonos en la estructura cristalográfica de *FhCL1* (Stack et al. 2008) construimos un modelo de *FhCL3* para comparar el

sitio activo de las distintas catepsinas, buscando correlacionar la arquitectura del bolsillo S<sub>2</sub> con las diferentes especificidades observadas.

Los resultados de estos enfoques experimentales se presentaron en el siguiente artículo:

**Artículo:** *La principal catepsina L secretada por el juvenil invasivo de Fasciola hepatica prefiere sustratos con prolina en el subsitio S<sub>2</sub> y puede cortar el colágeno*

**Ileana Corvo**, Martín Cancela, Mónica Cappetta, Natalia Pi-Denis, José F. Tort, Leda Roche

*Molecular & Biochemical Parasitology, 2009 (167): 41-47.*



Contents lists available at ScienceDirect

## Molecular &amp; Biochemical Parasitology



## The major cathepsin L secreted by the invasive juvenile *Fasciola hepatica* prefers proline in the S2 subsite and can cleave collagen<sup>☆</sup>

Ileana Corvo<sup>a</sup>, Martín Cancela<sup>b</sup>, Mónica Cappetta<sup>a</sup>, Natalia Pi-Denis<sup>a</sup>, José F. Tort<sup>a</sup>, Leda Roche<sup>a,\*</sup>

<sup>a</sup> Dpto. de Genética, Facultad de Medicina, UDELAR, Montevideo, Uruguay

<sup>b</sup> Laboratorio de Biología Molecular de Cestodos, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

## ARTICLE INFO

## Article history:

Received 27 January 2009  
Received in revised form 8 April 2009  
Accepted 12 April 2009  
Available online 19 April 2009

## Keywords:

Trematode  
*Fasciola hepatica* NEJ  
Cathepsins  
Collagenase  
Invasion  
*Hansenula polymorpha*

## ABSTRACT

Secreted cysteine proteases are major players in host–parasite interactions; in *Fasciola hepatica*, a distinct group of cathepsins L was found to be predominantly expressed in the juvenile stages, but their enzymatic properties were unknown. Cathepsin L3 (FhCL3) is a main component of the juvenile secretory products and may participate in invasion. To characterize the biochemical properties, the proenzyme was expressed in the methylotrophic yeast *Hansenula polymorpha* and the mature enzyme was obtained from the culture medium. FhCL3 exhibited optimal activity and stability at neutral pH and a noticeable restricted substrate specificity with 70-fold preference for Tos-Gly-Pro-Arg-AMC over typical cathepsin substrates with hydrophobic or aliphatic residues in the S2 position. Accordingly, FhCL3 efficiently cleaved type I collagen over different pH and temperature conditions, but it did not cleave immunoglobulin. While most cathepsin cysteine proteinases are unable to digest collagen, mammalian cathepsin K, adult *F. hepatica* FhCL2 and the plant zingipain can also cleave collagen and substrates with Pro in P2 position, but only FhCL3 and zingipain hydrolyze these substrates with the highest efficiency. Molecular modeling and structural comparisons of the collagen cleaving cathepsins indicated that the strong substrate selectivity observed might be due to steric restrictions imposed by bulky aromatic residues at the S2–S3 subsites. The remarkable similarities of the active site clefts highlight the evolutive constraints acting on enzyme function. The presence of a collagen cleaving enzyme in *F. hepatica* juvenile stages is suggestive of a role in tissue invasion, an essential feature for the establishment of the parasites in their host.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

The identification of the major virulence factors is a primary goal in parasite research since they can eventually lead to their use in the development of specific control measures, either through drugs or immunogens. In parasitic trematodes secreted cysteine proteases have been recognized as major players in host–parasite interaction with proposed roles in host colonization, traversing tissue barriers, evading immune responses, and digesting nutrients [1,2]. Therefore these enzymes might be partly responsible for the pathogenic effects of the infection, which translate in the case of zoonotic species in health and sanitary issues and in relevant eco-

nomical losses. The infection with *Fasciola hepatica* or *F. gigantica* is responsible for fasciolosis, a debilitating liver infection affecting more than 700 million livestock animals worldwide, and resulting in severe economic losses conservatively estimated at US\$ 3.2 billion annually, not to mention the increase of the incidence of human fasciolosis in several regions that lead to its recognition as an emerging zoonotic disease by the World Health Organization [3].

It has long been established that *F. hepatica* has several cathepsin like cysteine proteases that prevail in secretions of different life stages of the fluke's cycle, while phylogenetic analysis demonstrated the presence of a complex multigenic family [2,4]. However, up to recently there was no clear evidence of differential expression of these enzymes. Proteomic analysis of adult secretions identified a restricted set of cathepsin L like enzymes that might start their expression during the stage of liver migration [5]. Using combined molecular, biochemical and proteomic approaches a subset of cysteine proteases including at least three cathepsin B like and two cathepsin L like enzymes was identified as predominantly expressed by the newly excysted juveniles (NEJ) [6]. The differential expression patterns observed might correlate with functional diversification within the cathepsin family. Due to their restrictive expression the NEJ enzymes might be the ones involved in the

**Abbreviations:** AMC, 7-amino-4-methyl coumarin; FhCL, *Fasciola hepatica* cathepsin L; Tos, Tosyl; Z, benzyloxycarbonyl.

<sup>☆</sup> Note: Nucleotide sequence data reported in this paper is available in the GenBank™ database under the accession number EU287914.

\* Corresponding author at: Dpto. Genética, Facultad de Medicina, Universidad de la República, UDELAR, Gral.Flores 2125, CP11800, Montevideo, Uruguay.  
Tel.: +598 2 9249562/9243414x3469; fax: +598 2 9249563.

E-mail address: lroche@fmed.edu.uy (L. Roche).

initial steps of mammalian host infection, the excystment of metacercariae and invasion through the intestinal wall as suggested by previous reports [7–9].

The NEJ cathepsins L are encoded by an early divergent clade of genes bearing substitutions in relevant positions of their active site, and may have consequently different enzymatic properties than the predominant counterparts in the adult flukes [5,6]. Secreted protein profiles and proteinase activities showed subtle changes *in vitro* during the first 48 h after excystment, with prevalence of cathepsin L-like activity. The predominant cathepsin found in the early NEJ excretion/secretion products (E/S) by proteomic analysis is proFhCL3, progressively changing to the mature enzyme during the first 48 h of NEJ development. Noticeably, at this stage the hydrolysis of an oligopeptide substrate with Pro in the P2 position is more efficient than the more typical cathepsin substrates bearing Phe or Leu in that position [6].

The property of hydrolyzing substrates with Pro in P2 is shared with mammalian cathepsin K, an enzyme involved in bone resorption by osteoclasts [10]. The vertebrate enzyme has the unusual ability of cleaving collagen at the Gly-Pro-Xaa repeat sequence (where Xaa is frequently hydroxyproline). A Tyr residue at position 67 within the S2 subsite of cathepsin K has been indicated as the principal responsible for this activity through stabilizing interactions with the Pro residue in the substrate [11]. Besides cathepsin K, only few other papain-like proteases have collagenolytic activity. Three ginger rhizome cysteine proteases (CP-II or zingipain, GP2 and GP3) have a marked preference for substrates with Pro in the P2 position, a fact that has been related to size exclusion due to the occurrence of a bulky tryptophan in the same position occupied by Tyr in cathepsin K [12,13]. The other C1A protease capable of collagen cleavage is FhCL2, secreted by the adult stages of *F. hepatica*. This enzyme has the same residues as mammalian cathepsin K at relevant positions in the S2 subsite, being a clear example on evolutive constrains [14]. While a collagenolytic activity in adult liver flukes might be relevant for diverse functions, the presence of a similar activity in early invasive stages is essential for the establishment of the parasite in the host.

As the initial stages of invasion are the main objectives of the protective immune responses, and the fluke pathology is mainly due to liver erosion during juvenile fluke migration, FhCL3 may be a relevant target for vaccine and chemotherapy design [15]. In order to further characterize this enzyme, we functionally cloned a cDNA encoding FhCL3, and characterized the recombinant enzyme. Tested with synthetic substrates the enzyme showed a marked preference for Pro residues in the selectivity S2 pocket of the active site. Accordingly, we demonstrated efficient collagenase activity, consistent with a role of FhCL3 in the tissue invasion process. Molecular modeling of the active site structure of cathepsins with collagenase activity compared to more typical C1A enzymes allowed the discussion of possible evolutive restrictions acting on enzyme function.

## 2. Materials and methods

### 2.1. Expression and purification of FhCL3 in the yeast *Hansenula polymorpha*

The FhCL3 cDNA fragment comprising most of the proenzyme coding sequence and the 3' UTR was amplified by PCR from NEJ cDNA using the forward primer 5'-GCGGATCCAATGATGTGTCATGGCAGAGTGGAAAGCG-3' (*Bam* HI site underlined; first 8 residues added to the propeptide in bold) and a *Sal* I adapter oligodT reverse primer (5'-GGCCACGCGTCGACAGTACTTTTTTTTTTTTTTTTIV-3', *Sal* I site underlined). The resulting fragment was cloned into pCRTOPO4 vector (Invitrogen). After total digestion with *Bam* HI and *Sal* I the fragment of the expected size was inserted in frame with the yeast

$\alpha$ -factor into the X4-Mf $\alpha$ -ScPas3 vector for expression and secretion in *H. polymorpha* [16]. FhCL1 was sub-cloned by amplification from a pre-existing adult cDNA clone using primers 5'-TAC GGA TCC AAT GAY GAT TTG TGG CAT-3' forward and 5'-CGA TTT CCG TGA TAA GCT TCT GTC GAC TCA-3' reverse (*Bam* HI and *Sal* I cloning sites underlined), and expressed in the same system. Vector and yeast strains were kindly provided by Dr. R.J.S. Baerends and Dr. J.A.K.W. Kiel, Molecular Cell Biology Lab, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands. Plasmid were linearized with *Stu* I to transform the yeast strain NCYC 495 (*leu* 1.1) by electroporation and the plasmid was integrated into the yeast genome as previously described [17]. Yeast transformants were cultured in 50 ml of yeast rich medium (YEPD) to an OD<sub>600</sub> of 5–6. Cells were harvested by centrifugation at 2000  $\times$  g for 10 min, and resuspended in buffered minimal media (0.67% yeast nitrogen base; 0.1 M phosphate buffer pH 6.0; 1% methanol) for induction at an initial OD<sub>600</sub> of 1. Cultures were maintained at 37 °C with orbital agitation until an OD<sub>600</sub> of 6–10 was reached (approximately 48 h incubation). After cell harvesting the supernatant was concentrated 20–30-fold in an Amicon concentrator with an ultrafiltration membrane of 10 kDa cut-off. Cathepsin enriched fractions were obtained by anion exchange chromatography (HiTrap QXL 5 ml GE Healthcare) from concentrated supernatants dialyzed against buffer Tris-HCl 0.1 M pH 7. Briefly, the column was equilibrated with Tris-HCl 0.1 M pH 7, and bound enzyme was eluted in a 0 to 0.5 M NaCl gradient. Elution fractions were assayed for FhCL3 activity using the fluorogenic peptide Tos-GPR-AMC and analyzed by SDS-PAGE electrophoresis. The FhCL3 and FhCL1 gel bands, migrating with an apparent molecular weight of 26–27 kDa were analyzed by tryptic digestion and matrix-assisted laser desorption ionization time of flight (MALDI-TOF) in a Voyager-DE-PRO mass spectrometer (Perspective Biosystems, Framingham, MA, Biochemical and Analytical Proteomic Unit, Institute Pasteur Montevideo) and verified to correspond to the mature FhCL1 and FhCL3, respectively. Purified enzyme was quantified by the BCA method [18].

### 2.2. Enzymatic assays

Enzyme concentration was determined by active-site titration with E-64c. Kinetic parameters were determined in reaction buffer containing 50 mM sodium phosphate, pH 6.5, 1 mM dithiothreitol, 1 mM EDTA at 24 °C, typically final enzyme concentrations were in the 10<sup>-9</sup> M range; and the substrate was added after 10 min of incubation of enzyme in reaction buffer. Proteolytic activity was monitored by the hydrolysis of 7-amino-4-methyl coumarin (AMC) from the synthetic peptide substrates Z-FR-AMC, Z-VLK-AMC, and Tos-GPR-AMC. Reaction rates with different substrate concentrations were measured as the slope of the progress curves obtained by continuous recording in a FluoStar spectrofluorimeter at 345 excitation and 440 emission wavelengths, using an AMC standard curve for product concentration calculation. Kinetic constants,  $k_{cat}$  and  $K_M$ , were estimated by non-linear regression analysis of the Michaelis-Menten plot. For pH activity profile, buffers pH 3.0–5.5 were prepared by mixing 0.5 M citric acid and 0.5 M sodium citrate; those at pH 6.0–9.0 were prepared by mixing 0.5 M Na<sub>2</sub>HPO<sub>4</sub> and 0.5 M NaH<sub>2</sub>PO<sub>4</sub>. The stability assay was performed by incubation of 0.1  $\mu$ M FhCL3 in buffer Tris-HCl 0.1 M pH 7 with 0.2 M NaCl in a water bath at 37 °C for different lengths of time. Residual activities were calculated as percentages of the activity of 0.1  $\mu$ M FhCL3 without incubation on duplicated experiments.

### 2.3. Digestion of type I collagen and immunoglobulin G1

Protein digestion was analyzed by incubating 10  $\mu$ g type 1 rat tail collagen (kindly provided by Dr. S. Chifflet, Dpto Bioquímica,

Facultad de Medicina, Uruguay, prepared according to [19] or 10  $\mu\text{g}$  human immunoglobulin G1 (kindly provided by Dr. P. Berasain, Unidad de Biología Parasitaria, Facultad de Ciencias, Uruguay, prepared as described in [20] with 1  $\mu\text{M}$  enzyme in buffer citrate pH 5.5 or phosphate pH 6.3 or 7, 1 mM DTT, 1 mM EDTA for different times at 28 or 37 °C. Fragments were separated by SDS-PAGE under reducing conditions and stained with Coomassie Brilliant Blue R-250.

#### 2.4. Molecular modeling

Homology molecular models of ProFhCL3 were generated with ModWeb [21] and SwissModel [22] using as principal template the coordinates of *F. hepatica* procathepsin L1 (2O6X, [14]). The models were visualized and compared by superposition using DeepView/Swiss PDBviewer version 3.7 [23], showing almost identical results. Only the mature portion of the enzyme was considered in the comparisons with other cysteine proteases. The model for FhCL2 was retrieved from the Swiss Model Repository (<http://swissmodel.expasy.org/repository/>) [24] and the remaining structures used in the comparison were retrieved from the Protein Data Bank (<http://www.pdb.org/>). They are human cathepsin L (1MHW), human cathepsin K (1B8, 1ATK), and ginger zingipain (1CQD). In the production of the figures shown, the E64 inhibitor complexed with human cathepsin K (1ATK) was superimposed in all the structures to facilitate the viewing of the active site cleft. The residues that make up the active site pockets were identified based on the literature and confirmed by structural alignment. The papain numbering is followed in the text and Table 2, while the actual numbering in each protein is indicated in Table 2 and Fig. 4.

### 3. Results

#### 3.1. Functional expression of *Fasciola hepatica* cathepsin L3

We had previously identified a NEJ cDNA fragment differentially expressed in juveniles corresponding to FhCL3 (GenBank accession DQ534446). Using primers derived from this sequence we obtained, cloned and sequenced a NEJ cDNA comprising most of the proenzyme coding sequence (GenBank accession EU287914). Sequence comparisons with public databases showed that this clone matched the original sequence (DQ534446) and also other juvenile expressed sequences (AJ279091, *F. hepatica* NEJ cDNA, 97% identity, EU191984, *F. hepatica* metacercariae cathepsin L, 94% identity, and the *F. gigantica* orthologue AF419329, 95% identity, among others). The homology to cathepsin L cDNA sequences from adult libraries was significantly lower in agreement with the preferential expression in metacercariae and NEJ.

The cDNA was subcloned in an expression vector and used to transform the methylotrophic yeast *Hansenula polymorpha*. The previously characterized FhCL1, which is the predominant enzyme in adult E/S products, was expressed in the same system and assayed in parallel for enzymatic properties. Both FhCL3 and FhCL1 were present in the culture media as active forms and migrated in SDS-PAGE with apparent molecular mass of 28 and 27 kDa, respectively, corresponding to the mature enzymes (Supplementary Fig. 1). The corresponding gel bands were confirmed as mature FhCL1 and FhCL3 by mass spectrometry. FhCL3 yield was 0.8–1 mg per liter of culture.

Functional expression of mature cathepsins L1 and L3 was evaluated by zymograms and synthetic fluorogenic substrates. Gelatinolytic activity was observed in supernatant from *H. polymorpha* expressing either FhCL1 or FhCL3 (data not shown). Proteolytic activity was dependent of reducing agent (1 mM DTT) and was completely abolished with the class inhibitor E-64 confirming the

**Table 1**

Kinetic parameters of FhCL3 over different peptide substrates.

Substrate	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{M}}$ ( $\mu\text{M}$ )	$k_{\text{cat}}/K_{\text{M}}$ ( $\text{mM}^{-1} \text{s}^{-1}$ )
Tos-GPR-AMC	$6.0 \pm 0.48$	$4.1 \pm 0.9$	1464.3
Z-VLK-AMC	$0.33 \pm 0.01$	$15.7 \pm 1.8$	20.76
Z-FR-AMC	$0.39 \pm 0.01$	$23.53 \pm 2.3$	16.78

Kinetic parameters for hydrolysis of synthetic peptidyl-AMC substrates by recombinant FhCL3 enzyme.

presence of an active cysteine proteinase expressed by *H. polymorpha*.

#### 3.2. Cathepsin L3 has preference for proline in the selective P2 position

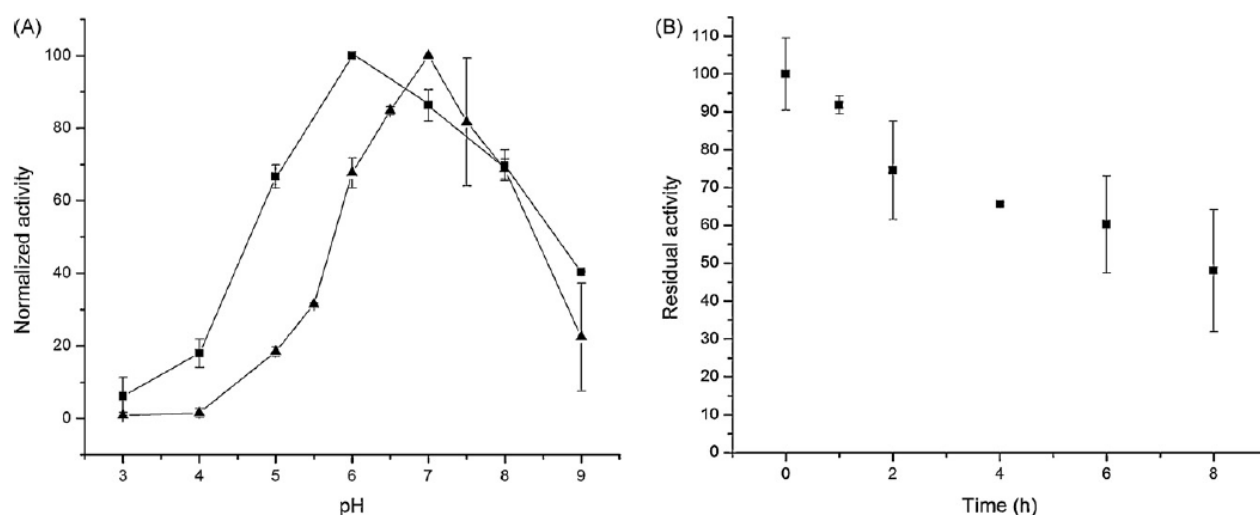
The activities of the enzymes towards different oligopeptide fluorogenic substrates were analyzed and data are shown in Table 1. FhCL1 showed kinetic parameters similar to those reported in previous studies [14,25], with a preference for Z-VLK-AMC over Z-FR-AMC and a 30 times lower activity towards Tos-GPR-AMC (data not shown). Noticeably, the FhCL3 enzyme exhibited a remarkable efficiency for cleavage of the substrate with Pro in the P2 position in comparison with the substrates with Leu or Phe in P2. The  $k_{\text{cat}}/K_{\text{M}}$  for Tos-GPR-AMC was around 70-fold higher over Z-VLK-AMC and Z-FR-AMC.

#### 3.3. Cathepsin L3 is stable and has optimum activity at pH7

The hydrolytic activity of FhCL1 towards Z-VLK-AMC and FhCL3 towards Tos-GPR-AMC were determined over a range of pH conditions, with both enzymes showing a wide pH activity profile consistently with previous reports of parasite cathepsins being active at various pH [26] (Fig. 1A). FhCL3 profile is switched to the neutral range with respect to FhCL1, with optimums at 7 and 6, respectively. FhCL3 has only 35% of the maximum activity at pH 5.5, whilst FhCL1 retain at least 70% of activity in agreement with most C1 family cathepsins. The stability of FhCL3 at neutral pH was performed by incubation at 37 °C and determination of residual activity after different lengths of time. Fifty percent of the activity was conserved after 8 h incubation (Fig. 1B) and 36% of it was retained after 24 h incubation.

#### 3.4. Cathepsin L3 has collagenolytic activity but is not efficient at immunoglobulin cleavage

Since the preference for Pro residues in the selectivity S2 pocket of the active site is shared with enzymes capable of cleaving collagen, we decided to test if the recombinant FhCL3 is efficient at processing this protein substrate, or other known liver fluke substrates as immunoglobulins. While FhCL1 cleaved IgG1 heavy chain as previously demonstrated, FhCL3 was not able to degrade this substrate in parallel experiments, even in conditions where its activity towards collagen was evident (Fig. 2). The collagenolytic activity was assayed by incubation of type I collagen with FhCL1 or FhCL3 at different conditions (Figs. 2 and 3). The conditions chosen were the physiological temperature and different pHs that may be faced by the parasite inside the host tissues, and also conditions that ensure the authentic activity without significant collagen denaturation, below the triple helix melting point (28 °C, pH 7). The resulting fragments were resolved by Coomassie Blue-stained SDS-PAGE. While FhCL1 displayed only a minimal collagenolytic activity, FhCL3 exhibited the ability to degrade type I collagen efficiently at pH 5.5, 6.3 and 7. Interestingly, a time course analysis of type I collagen degradation shows that FhCL3 reduce progressively the collagen bands, with an almost complete disappearance of the alpha



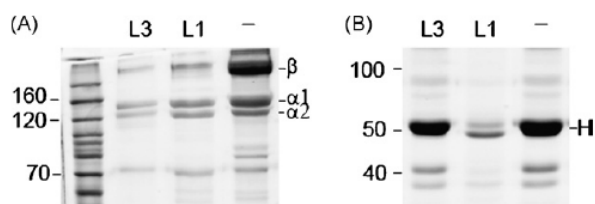
**Fig. 1.** pH and stability profile of FhCL3. (A) Activity profile of FhCL3 and FhCL1 at different pHs. The hydrolytic activity of FhCL1 over Z-VLK-AMC (■) and FhCL3 over Tos-GPR-AMC (▲) were assayed at different pH conditions (10 nM enzyme and 20  $\mu$ M substrate typical final concentrations). Activity was normalized considering the highest activity of each enzyme as 100. (B) FhCL3 stability assay. FhCL3 0.1  $\mu$ M was incubated at neutral pH for different lengths of time at 37 °C, residual activity towards Tos-GPR-AMC was plotted against time.

collagen bands after about 20 h incubation. In contrast extended incubation with FhCL1 produced only minor degradation (Fig. 3B).

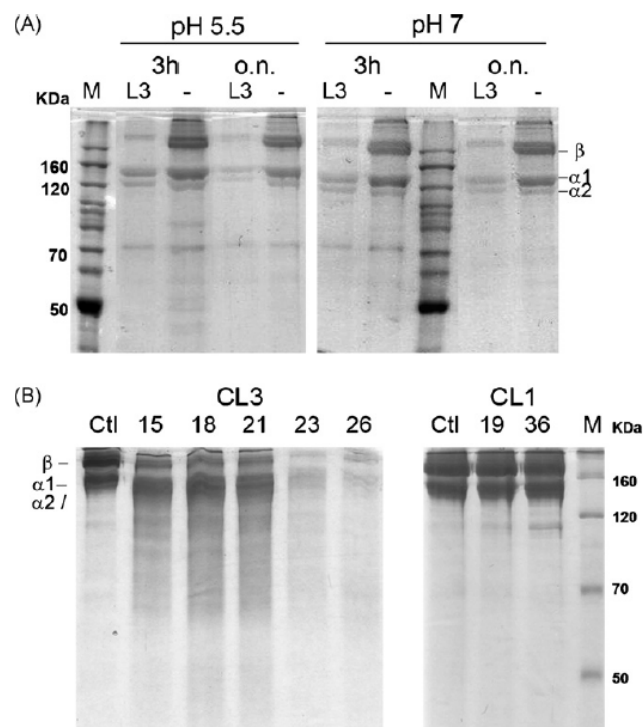
### 3.5. Molecular modeling suggest relevant residues for collagenolytic activity

In order to further analyze the structure–function relationships that might explain the peculiarities of FhCL3 activity, we generated a molecular model of the enzyme, and compared it with the adult liver fluke cathepsins FhCL1, FhCL2, human cathepsin L, and the collagenolytic papain-like cysteine proteases cathepsin K and ginger rhizome zingipain. The differences of the active site clefts of these enzymes are dependent of few residues forming the S2 and S3 subsites as previously established [5,27]. While the floor of the S2 pocket is the same in all enzymes (Met 68, Ala133, and Ala160 with the only exception of Gly160 in human cathepsin L), there is variation in the residues that form the walls and the bottom of the pocket (Table 2). The most relevant difference observed is in position 67 that functions as a gatekeeper separating the S2 and the S3 subsites. The residue in this position is Leu in human cathepsin L and FhCL1, Tyr in cathepsin K and FhCL2 and a bulkier Trp in ginger rhizome protease and FhCL3 (Fig. 4). The bulky aromatic residues Tyr and Trp can rotate about the C $\alpha$ –C $\beta$  bond, adopting diverse conformations. While some of these can result in a constrained S2 subsite, that would sterically exclude large residues, they can at the same time offer stabilizing ring to ring interactions with a proline in the pocket. On the opposite side of the cleft, on position 158 *Fasciola* cathepsins

FhCL2 and FhCL3 have a Thr residue, a feature not observed up to now in any other papain-like protease. The fairly reactive hydroxyl group of Thr is placed in a configuration where it is able to form hydrogen bonds with a variety of substrates at P2. Besides, there is always an aliphatic moiety at position 157, usually Val or Leu. Since the latter extends further into the pocket, it can provide the stabilizing aliphatic interactions that need to be established with Pro at P2.



**Fig. 2.** Collagen and IgG1 cleavage activity of adult and NEJ cathepsins L. 10  $\mu$ g of type I collagen (A) or IgG1 (B) were incubated with FhCL3, FhCL1 (1  $\mu$ M) or without enzyme (-) 4 h at 37 °C in buffer phosphate pH 7, 1 mM DTT, 1 mM EDTA and the reaction was analyzed by 10% SDS-PAGE. Molecular weight markers (kDa) are indicated on the left of each figure; collagen beta ( $\beta$ ) and alpha ( $\alpha$ ) chains and IgG1 heavy chain (H) are indicated on the right.



**Fig. 3.** Collagenolytic activity of FhCL3 and FhCL1 at different pH and temperature conditions. (A) Type I collagen was incubated with 1  $\mu$ M FhCL3 or without enzyme (-), at different pH conditions for 3 h or overnight (o.n.) at 37 °C and analyzed by 10% SDS-PAGE; (B) Type I collagen was incubated with FhCL3, FhCL1 (1  $\mu$ M) at 28 °C, pH7; aliquots were taken at different times (indicated in hours on each lane) and analyzed by 6% SDS-PAGE. Ctl indicates the control reactions without enzyme incubated for 36 h. Molecular weight markers (kDa) and collagen beta ( $\beta$ ) and alpha ( $\alpha$ ) chains are indicated at the sides of the figures.

**Table 2**  
S3 and S2 active site pocket residues.

Residue position <sup>a</sup>	S3 subsite					S2 subsite						
	60	61	64	66	66	67	68	133	157	158	160	205
Human L	Asn <sup>62</sup>	<i>Glu</i> <sup>63</sup>	<i>Asn</i> <sup>66</sup>	Gly <sup>67</sup>	Gly <sup>68</sup>	<i>Leu</i> <sup>69</sup>	Met <sup>70</sup>	Ala <sup>135</sup>	<i>Met</i> <sup>161</sup>	<i>Asp</i> <sup>162</sup>	<i>Gly</i> <sup>164</sup>	<i>Ala</i> <sup>214</sup>
Human K	Asn <sup>61</sup>	<i>Asp</i> <sup>62</sup>	<i>Gly</i> <sup>64</sup>	Gly <sup>65</sup>	Gly <sup>66</sup>	<i>Tyr</i> <sup>67</sup>	Met <sup>68</sup>	Ala <sup>134</sup>	<i>Leu</i> <sup>160</sup>	<i>Asn</i> <sup>161</sup>	Ala <sup>163</sup>	<i>Leu</i> <sup>209</sup>
Ginger	Asn <sup>62</sup>	<i>Asn</i> <sup>63</sup>	<i>Arg</i> <sup>66</sup>	Gly <sup>67</sup>	Gly <sup>68</sup>	<i>Trp</i> <sup>69</sup>	Met <sup>70</sup>	<i>Thr</i> <sup>134</sup>	<i>Ala</i> <sup>159</sup>	<i>Asn</i> <sup>160</sup>	Ala <sup>162</sup>	<i>Phe</i> <sup>211</sup>
FhCL1	Asn <sup>62</sup>	<i>Gly</i> <sup>63</sup>	<i>Gly</i> <sup>66</sup>	Gly <sup>67</sup>	Gly <sup>67</sup>	<i>Leu</i> <sup>69</sup>	Met <sup>70</sup>	Ala <sup>135</sup>	<i>Val</i> <sup>160</sup>	<i>Asn</i> <sup>161</sup>	Ala <sup>163</sup>	<i>Leu</i> <sup>209</sup>
FhCL2	Asn <sup>62</sup>	<i>His</i> <sup>63</sup>	<i>Gly</i> <sup>66</sup>	Gly <sup>67</sup>	Gly <sup>67</sup>	<i>Tyr</i> <sup>69</sup>	Met <sup>70</sup>	Ala <sup>135</sup>	<i>Leu</i> <sup>160</sup>	<i>Thr</i> <sup>161</sup>	Ala <sup>163</sup>	<i>Leu</i> <sup>209</sup>
FhCL3	Asn <sup>62</sup>	<i>His</i> <sup>63</sup>	<i>Gly</i> <sup>66</sup>	Gly <sup>67</sup>	Gly <sup>67</sup>	<i>Trp</i> <sup>69</sup>	Met <sup>70</sup>	Ala <sup>135</sup>	<i>Val</i> <sup>160</sup>	<i>Thr</i> <sup>161</sup>	Ala <sup>163</sup>	<i>Val</i> <sup>209</sup>

<sup>a</sup> Papain numbering. Variable positions are in bold italics. The actual position in each protein is indicated by superscript.

Finally, the bottom of the S2 subsite (position 205) that was shown previously to have a role in substrate specificity in adult liver fluke enzymes [14] does not seem to be involved in the ability to accept Pro, since the distance between the side chains at this position lie too far to establish significant interactions with the Pro ring.

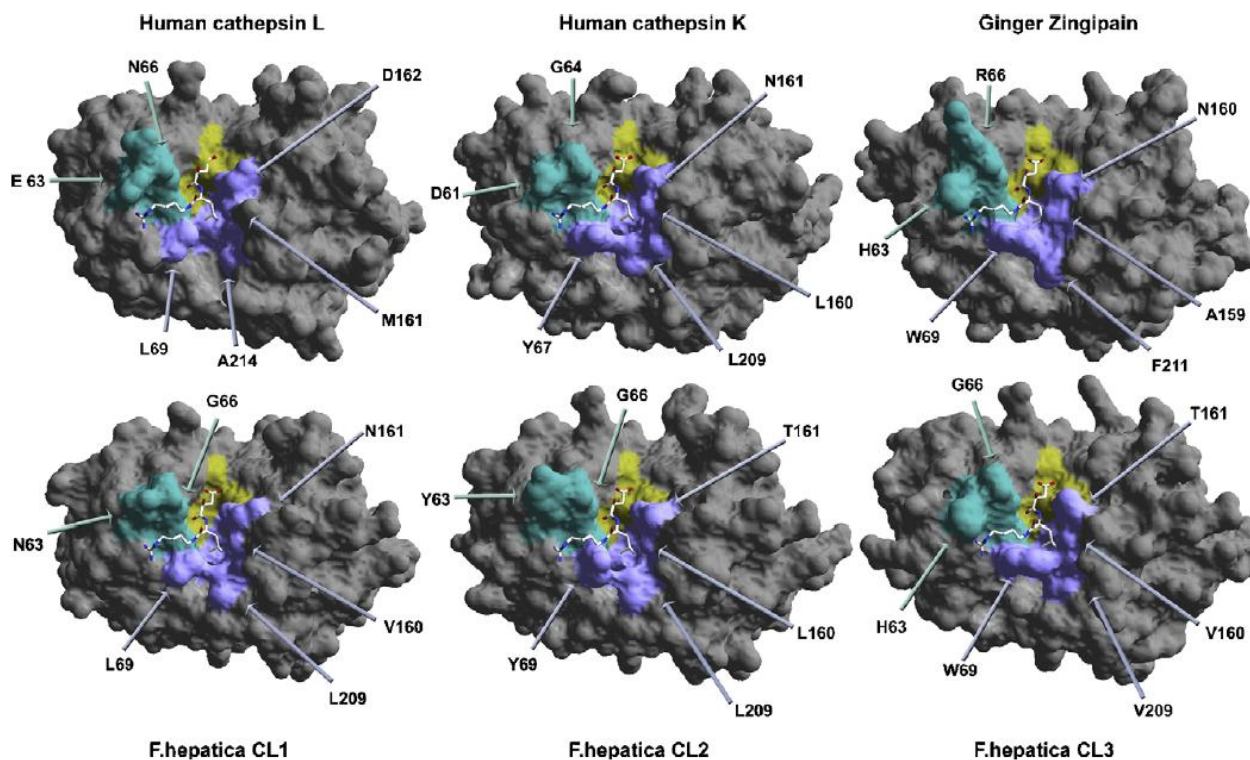
Interestingly, two of the five residues that form the S3 subsite showed remarkable variations that might be correlated with functional constraints: positions 61 at the bottom of the pocket and position 64 at the entrance of the S3 opposite to the gatekeeper residue 67 (Fig. 4). The presence of bulky Tyr or His aromatic residues in FhCL2 and FhCL3 restricts the possibilities of accommodating large moieties in S3, a shared feature with the plant collagenolytic protease.

Taken together these observations indicate that in general, similar to human cathepsin L, FhCL1 has wide pockets capable of accommodating large residues. In contrast, FhCL2 has a narrower and deeper S2 subsite due to the presence of a bulky aromatic Tyr on the right side and a protruding Thr on the opposite side

besides the catalytic His, and also a short S3 pocket. FhCL3 S2 is the most constrained due to a bulkier Trp at position 69, accompanied in the opposite side by a Thr moiety, although the rotation of the gatekeeper Trp can accommodate larger side chains. However, this rotation will result in a steric hindrance at the entrance of the S3 subsite, limiting the possible residues in this pocket only to very small ones. This FhCL3 topology is similar to the one found in the plant collagenolytic cysteine protease, while FhCL2 has an S2 pocket similar in shape to human cathepsin K and FhCL1 has the wider (and probably less restricted) active site, similar to that found in the mammalian cathepsin L (Fig. 4).

#### 4. Discussion

FhCL3 was previously demonstrated to be predominantly secreted by the invasive newly excysted juveniles [6]. In order to characterize its biochemical properties, we have cloned and



**Fig. 4.** Amino acid composition at the active site clefts of different cysteine proteases. Comparison of the crystallographic structure of human cathepsin L (1MHW), human cathepsin K (1ATK), and Ginger rhizome protease zingipain (1CQD), with *Fasciola hepatica* cathepsin L1 (206X) and molecular models of FhCL2 and FhCL3, shown in molecular surface representation. The residues lining the active site pockets are indicated by colors: conserved catalytic S1 site is indicated in yellow, S2 pocket in light blue and S3 in cyan. The active site residues that differ between the selected molecules are indicated (numbering corresponding to each molecule). A complete list of residues of the active site pockets is included in Table 2. The E64 inhibitor complexed with human cathepsin K (1ATK) was superimposed in all the structures to facilitate the viewing of the active site cleft. Structures were visualized with DeepView/Swiss PDBviewer version 3.7.



functionally expressed a cDNA encoding procathepsin L3 from *F. hepatica* NEJ in *Hansenula polymorpha*. A 28-kDa protein corresponding to the mature processed enzyme was purified from the supernatants of yeast cultures induced with methanol. The enzyme exhibited the typical characteristics of a C1A cysteine protease, being activated by reducing agents, and inhibited by E64. This is the first report of expression of a functional cysteine protease in *H. polymorpha*, which has been shown to be a suitable system for large scale production of recombinant proteins [16]. Previous reports of expression of FhCL1 in the more commonly used methylotrophic yeast *Pichia pastoris* obtained the proenzyme; and *in vitro* processing was required to produce the active form [28]. The mature form of FhCL1 was also secreted by *H. polymorpha* from a proFhCL1 clone, validating this methylotrophic yeast as an efficient system for the expression of active proteases.

Enzymatic assays for hydrolysis of fluorogenic peptide substrates showed that FhCL3 strongly preferred cleaving Tos-GPR-AMC over Z-VLK-AMC and Z-FR-AMC. Only a few cysteine proteinases are efficient in cleaving substrates with Pro in the P2 position namely mammalian cathepsin K, *F. hepatica* cathepsin L2 and ginger rhizome zingipains. While adult FhCL2 also is capable of cleaving substrates with Pro in P2, the activity towards Z-VLK-AMC is preferred, with a higher  $k_{cat}/K_M$  respect to Tos-GPR-AMC [14,29]. A similar situation is observed in mammalian cathepsins K, where Leu is the preferred residue at the selective P2 position, although Pro is very efficiently cleaved [30]. In contrast, FhCL3 is the first *Fasciola* cathepsin L shown to be 70-fold more efficient for the hydrolysis of a substrate with Pro in P2 with respect to the ones having Phe or Leu in that position. This peculiarity is shared with the ginger zingipain (CP-II), which is more than a hundred times more efficient over substrates bearing Pro than any other residue occupying the S2 pocket [12]. The maximal proteolytic activity and stability at a neutral pH may correspond to the secretion and function of FhCL3 in the environment encountered by the early NEJ stages at the host duodenum.

In order to advance in the knowledge of the possible biological role of cathepsin L3 we assayed the degrading activity over two biological relevant substrates previously assayed for FhCL1 and FhCL2: IgG1 and Type I collagen [14,20,31]. Cleavage of IgG subclasses by adult expressed FhCL1 and FhCL2 was shown to be specific and restricted, and may be involved in the mechanisms used by the parasite to evade the host immune response [20]. The absence of this activity in the juvenile enzyme is indicative of a restrained specificity of FhCL3. This is consistent with the expression of this enzyme while the parasite is in the host intestinal lumen, an environment where the interaction with immunoglobulins might be minimal. Moreover, while FhCL1 displayed only a minimal collagenolytic activity, FhCL3 exhibited the ability to cleave type I collagen in different pH conditions. The time course of collagen degradation, under incubation conditions that favor the native structure (28 °C, pH 7), exhibits an increase in the fading of alpha chains after about 24 h, which may be related to a cooperative effect on disruption of the quaternary structure.

Collagens, which are predominantly located in the extracellular matrix (ECM), are built up of three polypeptide chains containing the repeating triplet sequence Gly-X-Y, where Pro is commonly found in the X position and Hyp (4-hydroxyproline) in the Y position. Fibrillar collagens I and II with their triple helical structural arrangement are highly resistant to general proteolysis and require specific proteinases for their degradation. In mammals, the collagenase activity essential for tissue remodeling is performed mainly by specialized matrix metalloproteinases [32]. The degradation of collagen by papain-like enzymes is rare. Among mammalian papain-like enzymes, osteoclast secreted-cathepsin K, but not cathepsin L, possesses unique collagen I and collagen II degrading properties [10]. Besides CatK, the secreted *Fasciola* FhCL2 can cleave collagen effi-

ciently at pH 5.5 and to a lesser extent at pH 7, being proposed that this activity would be crucial for tissue traversing by the liver migrating flukes [14]. However, the expression of FhCL2 is not detected until later stages at the mRNA or protein levels, so the available evidence suggest that this enzyme is not present in the secretions of juveniles at the time they traverse the duodenum and peritoneal cavity. There is no clear understanding of the reason for a collagenolytic activity in the ginger rhizome, but it has two papain-like cysteine proteases (almost identical) that can effectively process this substrate [12,13]. The diverse enzymes might find slightly different environments that facilitate their action. In this sense, it has been reported that the ability of cathepsin K to degrade fibrillar collagen is dependent on the formation of a complex with bone and cartilage resident chondroitin-4-sulfate [33]. Previous data on FhCL2 and our data with FhCL3 indicate that the *Fasciola* enzymes do not require any macromolecular cofactors for collagen breakdown.

Degradation of collagen cross-linked triple helices by mammalian cathepsin K has been related to a more restricted S2 active site pocket and the exclusive preference for Pro at P2 among mammalian cathepsins [11]. More recent studies on cathepsin K have shown the relevance of the S3 subsite on substrate specificity [27]. For this reason we compared the substrate binding cleft of the *Fasciola* cathepsins in parallel to those present in the collagenolytic cysteine proteases and the reference human cathepsin L. The more obvious difference lies in the residue at the gatekeeper position 67 (papain numbering) where bulky aromatic residues in collagen cleaving cysteine proteases establish a steric hindrance for the entrance of large residues, and provide stabilizing interactions for a P2 proline. The relevance of other S2 positions for this activity is stressed by the results reported in site directed mutagenesis studies. The replacement in FhCL1 of the gatekeeper (67) Leu to Tyr does not enhance significantly the activity towards substrates with Pro [14], while the same substitution in the *F. hepatica* cathepsin L5 (a minor adult specific *Fasciola* cathepsin) resulted in an increment on the ability to cleave Tos-GPR-AMC [34], but also improved the efficiency toward Phe in P2. It has been proposed that the residue 157 on the opposite side of the pocket entrance is responsible for this difference. While a Leu 157 moiety found in FhCL2, FhCL5 and cathepsin K can contribute with stabilizing aliphatic interactions with Pro at P2, a shorter Val in FhCL1 (and its mutant) can not reach the Pro ring [14]. Surprisingly, in FhCL3 there is a Val in this position. It might be possible that the bulkier Trp at the gatekeeper position 67 might shift the proline ring closer to the opposite wall, favoring the interaction with the aliphatic Val, or else other residues in the vicinity contribute to Pro positioning. A remarkable difference found exclusively on *Fasciola* cathepsins L2 and L3 is the presence of a Thr in position 158. There is a positive correlation between the presence of this residue and the ability to cleave collagen and substrates with Pro in P2, but a direct proof of involvement must wait for mutagenesis studies, currently under way. Our homology modeling strongly suggests that a restricted S3 subsite might be a relevant characteristic of FhCL3 and the ginger rhizome enzyme. The presence of a large Tyr residue closing the bottom of the S3 pocket (position 61), adjacent to the gatekeeper Trp 67, would restrict this site to very small residues. Positional scanning using synthetic combinatorial libraries have been successfully used to analyze the specificity of diverse enzymes [14,27,35] and would be very useful to clarify the particular pocket preferences that confer FhCL3 its distinct properties.

The combination of preferences for small residues accepted at P3 and the stabilizing interactions with Pro at the S2 pocket are in line with the characteristic triplet repeat of collagens [32], and explains why FhCL3 can act as a "collagenase", a relevant acquisition for host colonization. Furthermore, the variations observed in the active sites of the *Fasciola* cathepsins L is indicative of a process

of functional divergence and specialization during the evolution of this gene family. These functions may involve digestion of different host proteins encountered at various stages. Notably, while FhCL1 resembles mammalian cathepsin L, FhCL2 has similarities with cathepsin K, and FhCL3 is structurally and functionally proximal to ginger rhizome zingipain. The ability to degrade substrates with Pro in the P2 position by cysteine cathepsins emerged independently in plants, mammals and trematode species, but interestingly converging into a similar active site architecture. This suggests strong evolutionary constraints acting on the enzyme active sites. In agreement with this observation, the residues at positions 61, 64, 67 were found to be under positive selection in the *Fasciola* cathepsin L family [4].

Functional studies are required to further understand the role of FhCL3 in the invasive process. Recently, new functional genomic tools have been developed in *Fasciola hepatica* showing that RNAi and reporter gene activity are able to be used in order to provide answers to relevant biological questions [36]. In this sense it was reported that RNAi with a FhCL1 gene fragment, comprising a region conserved across the cathepsin L family induced transient, abnormal locomotory phenotypes in *F. hepatica* NEJs, and significantly reduced penetration of the rat intestinal wall [37]. Specific silencing of the FhCL3 mRNA followed by assays of the invasive capability using *in vivo* and/or *ex vivo* models are under way to confirm the suspected role in invasion. Also, the assessment of the protective role of this molecule in vaccine trials in the bovine or ovine natural hosts may contribute to the design of control strategies for fasciolosis.

#### Acknowledgements

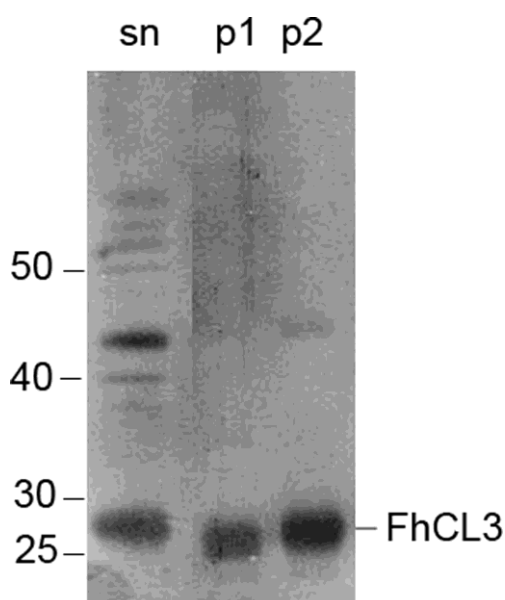
We would like to thank Dr. R.J.S. Baerends and Dr. J.A.K.W. Kiel, Molecular Cell Biology Lab, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands for kindly providing the yeast expression system. Dr. Silvia Chifflet and Dr. Patricia Berasain kindly provided the collagen and human IgG1 preparations used in this study. We are grateful to Dr. Lucía Piacenza, Dr. Gonzalo Peluffo and Laboratorios Celsius S.A. for helping us with enzyme purification. We also want to thank Dr. Carlos Carmona and Dr. Gabriel Rinaldi for helpful discussion and comments to the manuscript, and Lucía Pastro and Nicolás Dell'Oca for technical assistance. This work was partially supported by CSIC and PEDECIBA.

#### Appendix A. Supplementary data

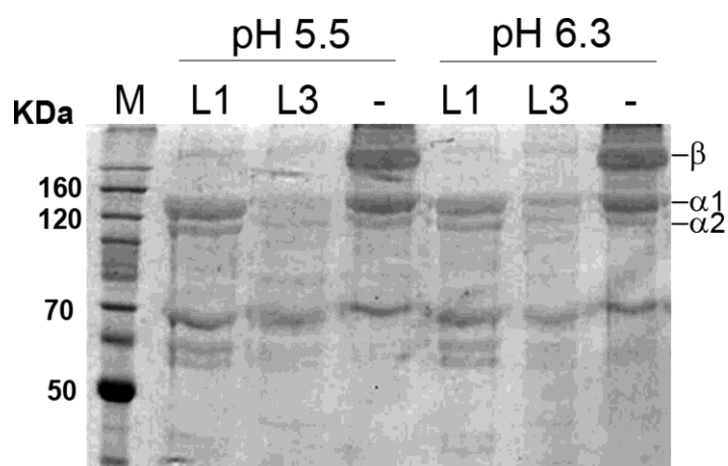
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2009.04.005.

#### References

- Dalton J, Caffrey C, Sajid M, et al. Proteases in trematode biology. In: Maule AG, Marks NJ, editors. Parasitic flatworms: molecular biology, biochemistry, immunology and physiology. CAB International; 2006.
- Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP. Proteinases and associated genes of parasitic helminths. *Adv Parasitol* 1999;43:161–266.
- Mas-Coma S, Bargues MD, Valero MA. Fascioliasis and other plant-borne trematode zoonoses. *Int J Parasitol* 2005;35(11/12):1255–78.
- Irving JA, Spithill TW, Pike RN, Whisstock JC, Smooker PM. The evolution of enzyme specificity in *Fasciola* spp. *J Mol Evol* 2003;57(1):1–15.
- Robinson MW, Tort JF, Lowther J, et al. Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Mol Cell Proteomics* 2008;7(6):1111–23.
- Cancela M, Acosta D, Rinaldi G, et al. A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie* 2008;90(10):1461–75.
- Carmona C, Dowd AJ, Smith AM, Dalton JP. Cathepsin L proteinase secreted by *Fasciola hepatica* in vitro prevents antibody-mediated eosinophil attachment to newly excysted juveniles. *Mol Biochem Parasitol* 1993;62(1):9–17.
- Tkalcevic J, Ashman K, Meeusen E. *Fasciola hepatica*: rapid identification of newly excysted juvenile proteins. *Biochem Biophys Res Commun* 1995;213(1):169–74.
- Tkalcevic J, Brandon MR, Meeusen EN. *Fasciola hepatica*: rapid switching of stage-specific antigen expression after infection. *Parasite Immunol* 1996;18(3):139–47.
- Lecaille F, Bromme D, Lalmanach G. Biochemical properties and regulation of cathepsin K activity. *Biochimie* 2008;90(2):208–26.
- Lecaille F, Chowdhury S, Purisima E, Bromme D, Lalmanach G. The S2 subsites of cathepsins K and L and their contribution to collagen degradation. *Protein Sci* 2007;16(4):662–70.
- Choi KH, Laursen RA, Allen KN. The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*. *Biochemistry* 1999;38(36):11624–33.
- Kim M, Hamilton SE, Guddat LW, Overall CM. Plant collagenase: unique collagenolytic activity of cysteine proteases from ginger. *Biochim Biophys Acta* 2007;1770(12):1627–35.
- Stack CM, Caffrey CR, Donnelly SM, et al. Structural and functional relationships in the Virulence-associated Cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. *J Biol Chem* 2008;283(15):9896–908.
- Reszka N, Cornelissen JB, Harmsen MM, et al. *Fasciola hepatica* procathepsin L3 protein expressed by a baculovirus recombinant can partly protect rats against fasciolosis. *Vaccine* 2005;23(23):2987–93.
- van Dijk R, Faber KN, Kiel JA, Veenhuis M, van der Klei I. The methylotrophic yeast *Hansenula polymorpha*: a versatile cell factory. *Enzyme Microb Technol* 2000;26(9/10):793–800.
- Faber KN, Haima P, Harder W, Veenhuis M, Ab G. Highly-efficient electrotransformation of the yeast *Hansenula polymorpha*. *Curr Genet* 1994;25(4):305–10.
- Smith PK, Krohn RI, Hermanson GT, et al. Measurement of protein using bicinchoninic acid. *Anal Biochem* 1985;150(1):76–85.
- Bornstein MB. Reconstituted rattail collagen used as substrate for tissue cultures on coverslips in Maximow slides and roller tubes. *Lab Invest* 1958;7(2):134–7.
- Berasain P, Carmona C, Frangione B, Dalton JP, Goni F. *Fasciola hepatica*: parasite-secreted proteinases degrade all human IgG subclasses: determination of the specific cleavage sites and identification of the immunoglobulin fragments produced. *Exp Parasitol* 2000;94(2):99–110.
- Pieper U, Eswar N, Davis FP, et al. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res* 2006;34(Database issue):D291–5.
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003;31(13):3381–5.
- Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;18(15):2714–23.
- Kopp J, Schwede T. The SWISS-MODEL repository: new features and functionalities. *Nucleic Acids Res* 2006;34(Database issue):D315–8.
- Roche L, Dowd AJ, Tort J, et al. Functional expression of *Fasciola hepatica* cathepsin L1 in *Saccharomyces cerevisiae*. *Eur J Biochem* 1997;245(2):373–80.
- Dalton JP, Neill SO, Stack C, et al. *Fasciola hepatica* cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *Int J Parasitol* 2003;33(11):1173–81.
- Choe Y, Leonetti F, Greenbaum DC, et al. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J Biol Chem* 2006;281(18):12824–32.
- Collins PR, Stack CM, O'Neill SM, et al. Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propetide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. *J Biol Chem* 2004;279(17):17038–46.
- Dowd AJ, Tort J, Roche L, Ryan T, Dalton JP. Isolation of a cDNA encoding *Fasciola hepatica* cathepsin L2 and functional expression in *Saccharomyces cerevisiae*. *Mol Biochem Parasitol* 1997;88(1–2):163–74.
- Lecaille F, Choe Y, Brandt W, Li Z, Craik CS, Bromme D. Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity. *Biochemistry* 2002;41(26):8447–54.
- Berasain P, Goni F, McGonigle S, et al. Proteinases secreted by *Fasciola hepatica* degrade extracellular matrix and basement membrane components. *J Parasitol* 1997;83(1):1–5.
- Prockop DJ, Kivirikko KI. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* 1995;64:403–34.
- Selent J, Kaleta J, Li Z, Lalmanach G, Bromme D. Selective inhibition of the collagenase activity of cathepsin K. *J Biol Chem* 2007;282(22):16492–501.
- Smooker PM, Whisstock JC, Irving JA, Siyaguna S, Spithill TW, Pike RN. A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*. *Protein Sci* 2000;9(12):2567–72.
- Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA, Craik CS. Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc Natl Acad Sci USA* 2000;97(14):7754–9.
- Rinaldi G, Morales ME, Cancela M, Castillo E, Brindley PJ, Tort JF. Development of functional genomic tools in trematodes: RNA interference and luciferase reporter gene activity in *Fasciola hepatica*. *PLoS Negl Trop Dis* 2008;2(7):e260.
- McGonigle L, Mousley A, Marks NJ, et al. The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *Int J Parasitol* 2007.



**Supplementary Fig. 1.** *FhCL3* expression and purification. Silver nitrate stained 12% SDS-PAGE showing 20 fold concentrated supernatant from culture media of yeast expressing *FhCL3* (sn) and *FhCL3* enriched fractions from anion exchange chromatography (p1, p2).



**Supplementary Fig. 2.** Collagenolytic activity of *FhCL3* and *FhCL1* at different pH and temperature conditions. Type I collagen was incubated with *FhCL3*, *FhCL1* (1  $\mu$ M), or without enzyme (-) overnight at 28 °C at different pH conditions. Reactions were analyzed by 10% acrylamide SDS-PAGE electrophoresis; molecular weight marker (kDa) is indicated on the left; collagen beta ( $\beta$ ) and alpha ( $\alpha$ ) chains are indicated on the right

**Fe de errata:** en la Tabla 2 los residuos en la posición 61 del subsitio  $S_3$  para Ginger, *FhCL1* y *FhCL2* son incorrectos. Los residuos correctos (como muestra la Fig. 4) son:

**Ginger:** His<sup>63</sup>  
***FhCL1*:** Asn<sup>63</sup>  
***FhCL2*:** Tyr<sup>63</sup>

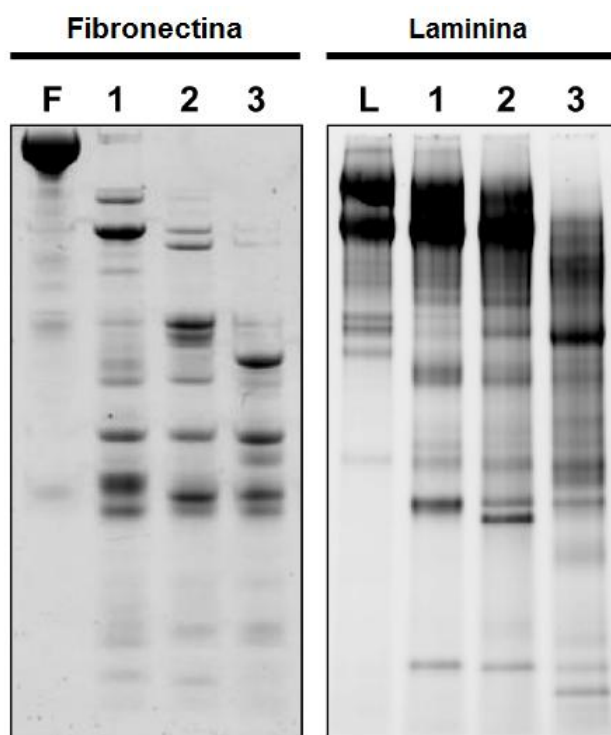
### 3. Estudio comparativo de la especificidad de sustrato de las catepsinas L del estadio juvenil y adulto

#### *Estrategia general*

Nos planteamos comparar la especificidad de sustrato de las tres catepsinas principales secretadas por *F. hepatica*: *FhCL1*, *FhCL2* y *FhCL3*. Para ello estudiamos la actividad sobre un panel de péptidos sintéticos e inhibidores. Analizamos su actividad sobre proteínas de la matriz extracelular y membranas basales con las cuales las catepsinas podrían entrar en contacto durante la migración del parásito en el hospedero mamífero, tales como distintos tipos de colágeno, fibronectina y laminina. Mientras que las tres enzimas degradan fibronectina y laminina de forma diferencial (sección 3.1), solo *FhCL2* y *FhCL3* fueron capaces de digerir el colágeno nativo. La caracterización de las diferencias en la cinética de digestión del colágeno por estas dos enzimas es el objeto de estudio del artículo presentado en la sección 3.2. En éste se usó la técnica de cromatografía acoplada a espectrometría de masas (HPLC-MS) para determinar los sitios de corte de *FhCL2* y *FhCL3* sobre las cadenas  $\alpha$  del colágeno I. Este abordaje ya había sido utilizado para obtener la secuencia de los péptidos resultantes de la hidrólisis de hemoglobina por *FhCL1* (Lowther et al. 2009). Se realizaron además simulaciones moleculares dinámicas sustituyendo los residuos del bolsillo  $S_2$  de la estructura de *FhCL1* por los presentes en las otras enzimas. Se analizó su capacidad de acomodar péptidos con diferentes combinaciones de aminoácidos en las posiciones  $P_2$  y  $P_3$  lo que permitió evaluar la contribución de los residuos del sitio activo de *FhCL2* y *FhCL3* en la interacción con el sustrato.

### 3.1. Digestión de fibronectina y laminina

En la Figura 16 se observa que las enzimas recombinantes *FhCL1* y *FhCL2* hidrolizan en forma similar ambos sustratos. *FhCL3* también es capaz de digerir la fibronectina, mostrando un patrón de bandas similar a las enzimas del adulto. En la degradación de laminina *FhCL3* es más efectiva, degradando extensamente las cadenas de la laminina.



**Figura 16. Digestión de fibronectina y laminina.** La digestión de cada proteína con las enzimas activas *FhCL1* (1), *FhCL2* (2) o *FhCL3* (3) se visualiza en geles NuPage Novex 4-12% Bis-Tris (Invitrogen) teñidos con Flamingo (BioRad). Los controles sin enzima se indican con F: fibronectina y L: laminina.

**3.2 Artículo:** *La actividad colagenolítica de las principales catepsinas L involucradas en la virulencia del helminto patógeno Fasciola hepatica*

Mark W. Robinson, **Ileana Corvo\***, Peter M. Jones, Anthony M. George, Matthew P. Padula, Joyce To, Martin Cancela, Gabriel Rinaldi, Jose F. Tort, Leda Roche, John P. Dalton

*PLoS Neglected Tropical Diseases*, 2011 5(4): e1012.

\* *Todos los experimentos presentados en este artículo, exceptuando las simulaciones moleculares dinámicas, fueron realizados por la tesista*

# Collagenolytic Activities of the Major Secreted Cathepsin L Peptidases Involved in the Virulence of the Helminth Pathogen, *Fasciola hepatica*

Mark W. Robinson<sup>1\*</sup>, Ileana Corvo<sup>2</sup>, Peter M. Jones<sup>1</sup>, Anthony M. George<sup>1</sup>, Matthew P. Padula<sup>1</sup>, Joyce To<sup>1</sup>, Martin Cancela<sup>2<sup>ra</sup></sup>, Gabriel Rinaldi<sup>2<sup>ra</sup>b</sup>, Jose F. Tort<sup>2</sup>, Leda Roche<sup>2</sup>, John P. Dalton<sup>3</sup>

**1** Infection, Immunity and Innovation (i3) Institute, University of Technology Sydney (UTS), Sydney, New South Wales, Australia, **2** Departamento de Genética, Facultad de Medicina, Universidad de la República (UDELAR), Montevideo, Uruguay, **3** Institute of Parasitology, McGill University, St. Anne de Bellevue, Quebec, Canada

## Abstract

**Background:** The temporal expression and secretion of distinct members of a family of virulence-associated cathepsin L cysteine peptidases (FhCL) correlates with the entry and migration of the helminth pathogen *Fasciola hepatica* in the host. Thus, infective larvae traversing the gut wall secrete cathepsin L3 (FhCL3), liver migrating juvenile parasites secrete both FhCL1 and FhCL2 while the mature bile duct parasites, which are obligate blood feeders, secrete predominantly FhCL1 but also FhCL2.

**Methodology/Principal Findings:** Here we show that FhCL1, FhCL2 and FhCL3 exhibit differences in their kinetic parameters towards a range of peptide substrates. Uniquely, FhCL2 and FhCL3 readily cleave substrates with Pro in the P2 position and peptide substrates mimicking the repeating Gly-Pro-Xaa motifs that occur within the primary sequence of collagen. FhCL1, FhCL2 and FhCL3 hydrolysed native type I and II collagen at neutral pH but while FhCL1 cleaved only non-collagenous (NC, non-Gly-X-Y) domains FhCL2 and FhCL3 exhibited collagenase activity by cleaving at multiple sites within the  $\alpha$ 1 and  $\alpha$ 2 triple helix regions (Col domains). Molecular simulations created for FhCL1, FhCL2 and FhCL3 complexed to various seven-residue peptides supports the idea that Trp67 and Tyr67 in the S2 subsite of the active sites of FhCL3 and FhCL2, respectively, are critical to conferring the unique collagenase-like activity to these enzymes by accommodating either Gly or Pro residues at P2 in the substrate. The data also suggests that FhCL3 accommodates hydroxyproline (Hyp)-Gly at P3-P2 better than FhCL2 explaining the observed greater ability of FhCL3 to digest type I and II collagens compared to FhCL2 and why these enzymes cleave at different positions within the Col domains.

**Conclusions/Significance:** These studies further our understanding of how this helminth parasite regulates peptidase expression to ensure infection, migration and establishment in host tissues.

**Citation:** Robinson MW, Corvo I, Jones PM, George AM, Padula MP, et al. (2011) Collagenolytic Activities of the Major Secreted Cathepsin L Peptidases Involved in the Virulence of the Helminth Pathogen, *Fasciola hepatica*. PLoS Negl Trop Dis 5(4): e1012. doi:10.1371/journal.pntd.0001012

**Editor:** Malcolm Jones, University of Queensland, Australia

**Received:** November 8, 2010; **Accepted:** December 21, 2010; **Published:** April 5, 2011

**Copyright:** © 2011 Robinson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** MWR is supported by a UTS Chancellor's Postdoctoral Fellowship. IC is supported by a Wood-Whelan Research Fellowship from the International Union of Biochemistry and Molecular Biology (IUBMB) and CSIC-UDELAR. PMJ is supported by Cure Cancer Australia and UTS fellowships. JPD is funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. The computational component of this research was undertaken at the NCI National Facility in Canberra, Australia, which is supported by the Australian Commonwealth Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

\* E-mail: mark.robinson@uts.edu.au

**ra** Current address: Laboratório de Cestodeos, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

**rb** Current address: Department of Microbiology, Immunology and Tropical Medicine, George Washington University Medical Center, Washington, DC, United States of America

## Introduction

Papain-like cysteine peptidases, including cathepsins B and L, are ubiquitously secreted extracorporeally by helminth parasites of human and veterinary importance where they perform many important roles that are critical to the development and survival of the parasite within the mammalian host [1]. These roles include penetration and migration through host tissues [2], catabolism of host proteins to peptides and amino acids [3,4], and modulation of the host immune response by cleaving immunoglobulin [5,6] or by altering the activity of immune effector cells [7]. Accordingly,

cathepsin peptidases are leading targets for novel anti-parasitic drugs and vaccines that block their function [8,9].

*Fasciola hepatica* is the causative agent of liver fluke disease (fasciolosis) of domestic animals in regions with temperate climates. Although traditionally regarded as a disease of livestock, fasciolosis is now recognised as an important emerging foodborne zoonotic disease in rural areas of South America (particularly Bolivia, Peru and Ecuador), Egypt and Iran [10]. It is estimated that over 2.4 million people are infected with *F. hepatica* worldwide and around 91 million are at risk of infection [11]. To infect their mammalian hosts, *F. hepatica* larvae, which are ingested with vegetation

### Author Summary

*Fasciola hepatica* is a helminth parasite that causes liver fluke disease (fasciolosis) in domestic animals (sheep and cattle) and humans worldwide. In order to infect their mammalian hosts, *F. hepatica* larvae must penetrate and traverse the intestinal wall of the duodenum, move through the peritoneum and penetrate the liver. After migrating through the liver, causing extensive tissue damage, the parasites move to their final niche in the bile ducts where they mature and feed on host haemoglobin to support the production of eggs. To achieve these tasks, *F. hepatica* secretes a number of distinct cathepsin L cysteine peptidases (FhCL). Thus, the infective larvae that penetrate the host gut secrete cathepsin L3 (FhCL3), the migrating liver-stage juvenile parasites secrete both FhCL1 and FhCL2 while mature bile duct parasites that feed on host blood secrete predominantly FhCL1 but also FhCL2. Here we show that the major cathepsin L peptidases secreted by *F. hepatica* (FhCL1, FhCL2 and FhCL3) display differential ability to degrade host collagen (an important component of host tissues) and investigate this phenomenon at the molecular level.

contaminated with dormant cysts (metacercariae), penetrate the intestinal wall, enter the liver capsule and migrate through the parenchyma before invading into the bile ducts [12]. To facilitate this tissue migration, *Fasciola* secrete various members of a multigenic family of cathepsin L peptidases that exhibit overlapping but complementary substrate specificities and together cleave host macromolecules very efficiently [13,14]. In fact, the ability of *Fasciola* to infect and adapt to a wide range of host species has been attributed to the effectiveness of this proteolytic machinery [14,15].

Phylogenetic analyses have shown that the *Fasciola* cathepsin L gene family expanded by a series of gene duplications followed by divergence which gave rise to three clades expressed by tissue-migrating and adult worms (Clades 1, 2, and 5) and two clades specific to the early infective juvenile stage (Clades 3 and 4) [13,14]. Consistent with these observations, our proteomics analysis identified representative enzymes from Clades 1, 2 and 5, but not from Clades 3 and 4, in the secretory products of adult *F. hepatica* [14]. More recently, we showed that the temporal expression and secretion of the specific cathepsin L clades correlated with the migration of the parasite through host tissues; members of cathepsin L clade 3 (FhCL3) are secreted by *Fasciola* infective larvae and effected penetration of the host intestinal wall while clades 1, 2 and 5 (FhCL1, FhCL2 and FhCL5) peptidases are secreted by the immature liver-stage flukes and adult worms and function in preparing a migratory path through the liver and in the acquisition of nutrient by degrading host blood and tissue components. While clade 4 (FhCL4) peptidases are expressed by infective larvae they do not seem to be secreted and, therefore, may play an intracellular house-keeping function [13,16]. Recent transcriptomic analyses of juvenile and adult stages have confirmed these observations [17,18].

The secreted *Fasciola* cathepsins are produced in specialised gastrodermal cells which line the parasites's gut and are packaged in secretory vesicles before being extruded into the gut lumen from where they are released into host tissues [19]. The peptidases can efficiently degrade a range of host macromolecules including haemoglobin, immunoglobulin and interstitial matrix proteins such as fibronectin and laminin [3,4,20–22]. Notably, however, studies in our laboratory using functionally-active recombinant enzymes have shown that FhCL2 and FhCL3 exhibit an unusual

ability to cleave native collagen [22,23]. This is of relevance because collagenase-like activity is restricted to very few enzymes (e.g. bacterial collagenases, matrix metalloproteinases and human cathepsin K) and, therefore, the evolution and maintenance of such an activity in *Fasciola* suggests that it is essential to the parasite's ability to degrade the connective tissue matrix of the organs through which it migrates.

The active site of papain-like cysteine peptidases is relatively short, and while consisting of four subsites (S2-S1-S1'-S2') with additional binding areas (S4-S3 and S3') the specificity of substrate binding is principally governed by the residues that make up the S2 subsite [24,25]. This S2 site forms a deep pocket capable of holding the P2 amino acid of the substrate and positioning the scissile bond into the S1 subsite for cleavage. In *Carica papaya* papain (PDB ID: 9PAP), the S2 subsite is composed of residues occupying positions 67, 68, 133, 157, 160 and 205. An analysis of these residues in the various cathepsin L clades clearly demonstrates divergence within the S2 subsite, in particular at the three positions that have the greatest influence on P2 binding i.e. at residues 67, 157 and 205 [14,15,23]. For FhCL2, the collagenolytic activity has been attributed to the presence of a particular residue, Tyr69, within the enzyme's S2 substrate binding site which is also found in human cathepsin K, the only mammalian cathepsin with the ability to cleave within the covalently-linked triple helices, of Col domains of native collagen [26,27]. The S2 Tyr69 is also suggested to allow both enzymes to cleave macromolecular and dipeptide substrates with a Pro residue in the P2 position [22,28]. More recently we showed that in FhCL3, this position is occupied by a larger Trp residue, a feature shared only with a ginger rhizome peptidase [29] which is also capable of cleaving collagen. Notably the S3 subsite of both FhCL3 and the plant enzyme are quite shallow, an observation that led us to advance the idea that the specificity of these enzymes might be restricted [23].

Our laboratory recently determined the three-dimensional structure of one of the major cathepsin L peptidases of adult *F. hepatica*, FhCL1 [22]. Similar to other cathepsins, the enzyme is composed of two domains (R and L) at the juncture of which is a cleft that forms the substrate-binding site and contains the enzyme catalytic machinery. Super-imposition of the alpha carbons of FhCL1 with cysteine peptidases from plants (e.g. papain, PDB ID 9PAP) and mammals (e.g. human cathepsin L PDB ID 1CJL) yields an r.m.s. deviation in the range 0.78 Å to 1.085 Å which is indicative of the very high conservation of the overall fold and shape that exists amongst all members of the papain family of cysteine peptidases [30]. Since the FhCL1 structure and fold can be described as practically identical to all other cathepsin L-like peptidases its scaffold can be exploited as a 'prototype' to investigate the role of critical amino acids within the S<sub>2</sub> subsite in substrate binding (Table 1), particularly those of FhCL2 and FhCL3, whose primary structures are 78% and 70% identical to FhCL1, respectively. In the present study, we investigated and compared the substrate specificity of active recombinant forms of FhCL1, FhCL2 and FhCL3 with specific emphasis on their ability to degrade native collagen. This activity was interpreted by obtaining the enzymatic kinetic parameters ( $K_m$ ,  $k_{cat}$ , and  $k_{cat}/K_m$ ) of these enzymes on a range of peptide substrates and binding kinetics for specific inhibitory compounds. Furthermore, using mass spectrometry we mapped the cleavage sites of native collagen I and derived peptides, adding evidence for differential activities between FhCL2 and FhCL3. Finally, using our FhCL1 crystal structure as a template we created molecular dynamics simulations to explain how the active sites of these enzymes accommodate collagen-like substrates and endow them with this unusual collagenolytic activity. Our study provides biochemical and



**Table 1.** Residues forming the S2 active site of human and *F. hepatica* cathepsin L peptidases.

	Residues					
	67	68	133	157	160	205
Human cathepsin L	Leu	Met	Ala	Met	Gly	Ala
<i>C. papaya</i> papain	Tyr	Pro	Val	Val	Ala	Ser
FhCL1	Leu	Met	Ala	Val	Ala	Leu
FhCL2	Tyr	Met	Ala	Leu	Ala	Leu
FhCL3	Trp	Met	Ala	Val	Ala	Val

Comparison of the residues from the S2 active site that contribute to differential substrate-binding in human cathepsin L, *C. papaya* papain and *F. hepatica* cathepsins FhCL1, FhCL2 and FhCL3.

doi:10.1371/journal.pntd.0001012.t001

structural insights into the molecular mechanism of tissue invasion by these important parasitic helminths.

## Materials and Methods

### Materials

Z-Phe-Arg-NHMec, Z-Leu-Arg-NHMec, Z-Val-Val-Arg-NHMec, Tos-Gly-Pro-Arg-NMec, Tos-Gly-Pro-Lys-NMec, Boc-Ala-Gly-Pro-Arg-NMec, Boc-Val-Leu-Lys-NMec, Boc-Val-Pro-Arg-NMec, Z-Phe-Ala-CHN2, Z-Gly-Pro-Gly-Gly-Pro-Ala and Z-Gly-Pro-Leu-Gly-Pro were obtained from Bachem (St. Helens, UK). Cathepsin K inhibitor II was purchased from BD Biosciences (Sydney, Australia). E-64, DTT, EDTA and bovine nasal septum collagen type II were obtained from Sigma-Aldrich (Sydney, Australia). Calf skin collagen type I was purchased from Calbiochem. *Pichia pastoris* strain X33 was obtained from Invitrogen (San Diego, CA, USA). Ni-NTA agarose and columns were obtained from Qiagen (Australia). Pre-cast NuPage 4–12% Bis-Tris gels and pre-stained molecular weight markers were purchased from Invitrogen (Australia).

### Expression and purification of recombinant *F. hepatica* cathepsins in yeast

Recombinant *F. hepatica* procathepsin L1, L2 and L3 (FhCL1, FhCL2 and FhCL3) were produced in yeast as previously described [22,23]. Briefly, *P. pastoris* (for FhCL1 and FhCL2 expression) and *P. angusta* (for FhCL3 expression) 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<sub>600</sub> of 2–6 was reached. Cells were harvested by centrifugation at 2000 × g for 5 min and protein expression induced by resuspending in 100 ml BMMY broth, buffered at pH 6.0 containing 1% methanol. Recombinant proteins were affinity purified from yeast using Ni-NTA-agarose. Recombinant propeptidases were dialysed against phosphate buffered saline (PBS) and stored at –20°C. The 37 kDa cathepsin L 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 dialysed against PBS, pH 7.3. The proportion of functionally active recombinant protein in these preparations was determined by titration against E-64.

### Enzyme assays and kinetics with fluorogenic peptide substrates

Initial rates of hydrolysis of the fluorogenic peptide substrates shown in table 2 were monitored by the release of the fluorogenic

**Table 2.** Kinetic parameters for hydrolysis of peptidyl-NHMec substrates by recombinant *F. hepatica* cathepsin Ls.

Enzyme	Substrate	K <sub>M</sub> μM ±		K <sub>cat</sub> s <sup>-1</sup> ±		K <sub>cat</sub> /K <sub>M</sub> Ms <sup>-1</sup>
FhCL1	Z-LR-NHMec	1.09	0.38	1.63	0.1	1492354.7
FhCL2	Z-LR-NHMec	2.13	0.29	0.95	0.035	444106.1
FhCL3	Z-LR-NHMec	48	5.7	1.2	0.05	25000
FhCL1	Z-FR-NHMec	1.9	0.57	0.12	0.008	64912.3
FhCL2	Z-FR-NHMec	7.8	1.5	0.09	0.008	11088
FhCL3	Z-FR-NHMec	20.05	2.9	0.02	0.001	999.3
FhCL1	Boc-VLK-NHMec	2.5	0.6	0.14	0.007	54266.7
FhCL2	Boc-VLK-NHMec	2.2	0.79	0.07	0.005	30712.5
FhCL3	Boc-VLK-NHMec	6.13	1.6	0.01	0.001	1747.8
FhCL1	Z-VVR-NHMec	2.84	0.7	0.021	0.001	7464.8
FhCL2	Z-VVR-NHMec	1.53	0.29	0.014	0.001	9237.5
FhCL3	Z-VVR-NHMec	11.7	2.1	0.043	0.020	3696.6
FhCL1	Tos-GPR-NHMec	27.2	4.9	0.02	0.001	671.6
FhCL2	Tos-GPR-NHMec	13.9	3.2	0.26	0.02	18559.5
FhCL3	Tos-GPR-NHMec	10.6	2.3	1.015	0.08	95774.7
FhCL1	Tos-GPK-NHMec	17.42	2.6	0.01	0.001	612.3
FhCL2	Tos-GPK-NHMec	12.9	3.4	0.18	0.013	13746.8
FhCL3	Tos-GPK-NHMec	9.8	3.7	0.15	0.015	36419.8
FhCL1	Boc-VPR-NHMec	22	2.5	0.01	0.0004	478.8
FhCL2	Boc-VPR-NHMec	11.39	2.1	0.23	0.01	20193.2
FhCL3	Boc-VPR-NHMec	21.1	3.2	0.033	0.002	1564
FhCL1	Boc-AGPR-NHMec	18.8	4	0.01	0.001	673.8
FhCL2	Boc-AGPR-NHMec	9.4	1.3	0.55	0.023	58027.1
FhCL3	Boc-AGPR-NHMec	9.6	1.6	0.58	0.025	60763.9

doi:10.1371/journal.pntd.0001012.t002

leaving group, NHMec, at an excitation wavelength of 380 nm and an emission wavelength of 460 nm using a Bio-Tek KC4 microfluorometer. *k<sub>cat</sub>* and *K<sub>m</sub>* values were determined using nonlinear regression analysis. Initial rates were obtained at 37°C over a range of substrate concentrations spanning *K<sub>m</sub>* values (0.2–200 μM) and at fixed enzyme concentrations (0.5–5 nM). Assays were performed in 100 mM sodium phosphate buffer (pH 6.0) containing 1 mM DTT and 1 mM EDTA. Rate constants for the inactivation of the *Fasciola* enzymes by Z-Phe-Ala-CHN2 and cathepsin K inhibitor II were determined from progress curves in the presence of substrate as previously described [22].

### Hydrolysis of native collagen and collagen-like peptides

Calf skin collagen type I and bovine nasal septum collagen type II (solubilised in 0.2 M acetic acid at a concentration of 2 mg/ml) were dialysed for two days against 0.1 M sodium acetate (pH 5.5) or PBS (pH 7.0). Digestion reactions contained 10 μg of dialysed collagen substrates, 1 mM DTT and 1 mM EDTA and 2 μM activated FhCL1, FhCL2 or FhCL3 in a final volume of 100 μl of one of the above buffers at 28°C. For collagen type I, reactions were performed for 3 h (pH 5.5) or 20 h (pH 7.0) whilst collagen type II was digested over 13–18 h. All reactions were stopped by the addition of 10 μM E-64. Digests were analyzed on reducing 4–12% NuPage Bis-Tris gels and visualised by staining with Flamingo fluorescent stain (Bio-Rad).

For digestion of collagen-like peptide substrates, 20 μg of Z-Gly-Pro-Leu-Gly-Pro and Z-Gly-Pro-Gly-Gly-Pro-Ala in DMSO were

incubated with FhCL2 or FhCL3 (15  $\mu$ M) in 100 mM sodium acetate buffer, pH 4.5, containing 1 mM EDTA and 2  $\mu$ M DTT for 30 min at 37°C. Digestion reactions were halted by the addition of 10  $\mu$ M E-64.

#### Analysis of collagen digests by mass spectrometry

Recombinant FhCL2 and FhCL3 were removed from collagen type I digests using Ni-NTA agarose. The reactions were then spun at 13,000 rpm for 15 min to remove particulates and were concentrated to a final volume of 15  $\mu$ l using a Concentrator 5301 (Eppendorf). Using an Eksigent AS-1 autosampler connected to a Tempo nanoLC system (Eksigent, USA), 10  $\mu$ l of the sample was loaded at 20  $\mu$ l/min with MS buffer A (2% acetonitrile+0.2% formic acid) onto a C8 trap column (Michrom, USA). After washing the trap for three minutes, the peptides were washed off the trap at 300 nL/min onto an IntegraFrit column (75  $\mu$ m $\times$ 100 mm) packed with ProteoPep II C18 resin (New Objective, Woburn, MA). Peptides were eluted from the column and into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer (AB Sciex) using the following program: 5–50% MS buffer B (98% acetonitrile+0.2% formic acid) over 15 minutes, 50–80% MS buffer B over 5 minutes, 80% MS buffer B for 2 minutes, 80–5% for 3 min. The eluting peptides were ionised with a 75  $\mu$ m ID emitter tip that tapered to 15  $\mu$ m (New Objective) at 2300 V. An Intelligent Data Acquisition (IDA) experiment was performed, with a mass range of 375–1500 Da continuously scanned for peptides of charge state 2+–5+ with an intensity of more than 30 counts/s. Selected peptides were fragmented and the product ion fragment masses measured over a mass range of 50–1500 Da. The mass of the precursor peptide was then excluded for 15 seconds. Peak list files generated by MSX (Infochromics) were exported to a local PEAKS Studio v5.0 (Bioinformatics Solutions Inc.) search engine for protein database searching. MS/MS data was used to search a custom-made database containing only bovine collagen sequences. The enzyme specificity was set to “no enzyme” and propionamide (acrylamide) modification of cysteines was used as a fixed parameter and oxidation of methionines was set as a variable protein modification. The mass tolerance was set at 100 ppm for precursor ions and 0.2 Da for fragment ions. Only 1 missed cleavage was allowed. Matched peptides achieving a score >60% were accepted during PEAKS searches [16]. The matching peptides were then mapped onto the primary amino acid sequence of bovine collagen to identify FhCL2 and FhCL3 cleavage sites and to plot P2 residue preference for each enzyme.

For collagen-like peptide substrates, digests were concentrated and analysed by MS/MS as described above with the following modifications. The mass range of 150–600 Da was scanned for peptides of charge state 2+ with an intensity of more than 100 counts/s. Selected peptides were fragmented and the product ion fragment masses measured over a mass range of 50–600 Da. The mass of the precursor peptide was then excluded for 120 seconds. An inclusion list describing all possible substrate ions that could be produced by enzymatic cleavage of the peptide substrates was generated and programmed into the Analyst acquisition software. The resulting data files were manually interrogated to determine the presence of peptide ions described in the inclusion list. The MS/MS spectra of those peptides were *de novo* sequenced for b and y ion fragments describing the peptide substrate’s sequence to a mass accuracy of approximately 0.2 Da.

#### Molecular dynamics (MD) simulations

For the MD simulations, starting coordinates for *F. hepatica* cathepsin L were taken from the 1.4 Å resolution crystal structure

of a FhCL1 mutant zymogen, in which the active site Cys was replaced by Gly ([22]; PDB 2O6X). The prosegment (residues 1–100) was removed and the active site Gly mutation reversed to the wild type Cys. Initial coordinates for a template peptide substrate (Ala-Leu-Ala-Leu-Pro) were derived from X-ray structures of inhibitors bound to human cathepsin K ([31]; PDB 1NLJ) and bovine cathepsin B ([32]; PDB 1SP4) after structural alignment with FhCL1. This initial peptide was altered to Ala-Leu-Arg-Asn-Ala using the mutate function in Swiss-PdbViewer ([33]; <http://spdbv.vital-it.ch/>) and then minimized while bound to the wild-type FhCL1 using the equilibration protocol described below. The equilibrated peptide was then extended by one Ala residue at its N- and C-termini using the coordinate generation function in the psfgen program [34], and then re-equilibrated. The resultant peptide (ligand A) was used to generate all other substrate starting coordinates by using the mutate function in Swiss-PdbViewer. Mutations to FhCL1 were similarly generated using Swiss-PdbViewer. Rotamers for mutated enzyme and substrate side-chains were chosen by visual inspection and using the rotamer score provided in Swiss-PdbViewer. The N-terminal residue of the substrate was acetylated and the C-terminus N-methylamidated. Each complex was optimally oriented to minimize cell volume [35] and solvated in a truncated octahedral periodic cell with a minimum of 20 Å between periodic images of the protein. The system was neutralized with sodium ions.

MD simulations were carried out with NAMD 2.6 [34] using the CHARMM27 force field with  $\phi/\psi$  cross-term map corrections [36]. Parameters for Hyp were from Veld and Stevens [37]. Water molecules were simulated with the TIP3P model [38]. Simulation conditions were maintained at 1.0 atm constant pressure by the Nosé-Hoover Langevin piston method [39,40] and at 310 K constant temperature by Langevin dynamics with a damping coefficient at 5 ps<sup>-1</sup>. The time step used for the simulations was 1.5 fs. A cutoff of 12 Å, with a switching function between 10 and 12 Å, was used for short-range non-bonded interactions. Long-range electrostatic interactions were computed using the particle mesh Ewald method [41] with a grid density of approximately 1/Å. A multiple time-stepping algorithm was used with interactions involving covalent bonds and short-range non-bonded interactions computed every time step, while long-range electrostatic forces were computed every two time steps. SHAKE [42] and SETTLE [43] were applied to constrain the lengths of all bonds that involve hydrogen.

The solvated starting structure was minimized using conjugate gradient minimization to a 0.5 kcal/(mol $\cdot$ Å) r.m.s. gradient with all enzyme heavy atoms fixed, with the exception of side-chain atoms of mutated residues, which were unrestrained. In addition, in this phase of the equilibration, ligand atoms were not fixed and harmonic positional constraints of 100 kcal/(mol $\cdot$ Å<sup>2</sup>) force constant were placed on the C $\alpha$  atoms of ligand residues 3–5 (P2, P1 and P1'). The unrestrained atoms were then further minimized during a 50 ps molecular dynamics run at 310 K. This starting model was then minimized with harmonic positional constraints on the NC $\alpha$ CO backbone of the protein and C $\alpha$  atoms of ligand residues 3–5. A 100 kcal/(mol $\cdot$ Å<sup>2</sup>) force constant was used to minimise the system to a 0.5 kcal/(mol $\cdot$ Å) r.m.s. gradient. The constraints were gradually removed by subsequent minimizations to a 0.1 kcal/(mol $\cdot$ Å) r.m.s. gradient, scaling the initial force constants by factors of 0.5, 0.15, 0.05, and 0. The unrestrained minimized structure was then heated from 50 K to 310 K in steps of 25 K using velocity reassignment during a 30 ps molecular dynamics run. The equilibrated system was then used for production runs with no restraints. All systems were run for 12 ns. All simulations remained stable to completion. For analysis,

the distance between the sulphur atom of the active Cys residue and the scissile backbone carbonyl carbon of the substrate (S-C distance) was recorded every 50 time-steps (0.075 ps); trajectory coordinates were recorded every 1000 time-steps (1.5 ps).

### Free energy of binding calculations

The free energy of binding of the peptide ligand to the peptidase contains an enthalpic and an entropic contribution. Free energy analysis of the production trajectories employed the single-trajectory MM/PBSA method combined with a determination of the change in configurational entropy using the harmonic approximation of normal-mode analysis [44,45]. Snapshots from the MD trajectory, stripped of water and counterions, were analysed. The enthalpy of binding is composed of the change in the molecular mechanics free energy upon complex formation, and the solvated free energy contribution. The molecular mechanics free energy difference was calculated using the SANDER module in AMBER 9 [46], with no cutoff for the non-bonded energies and the AMBER ff03 force field to describe the protein and peptide ligands [47]. Compatible parameters for Hyp were not available and binding energies for ligand F were not calculated. The AMBER PBSA module was used for the evaluation of the electrostatic free energy of solvation. A grid density of 3/Å was employed for the cubic lattice, the internal and external dielectric constants were set to 1 and 80, respectively, and 1000 linear iterations were performed. The non-polar solvation free energy was calculated from the solvent accessible surface area using the MSMS program [48], with a probe radius of 1.4 Å, the surface tension set to 0.00542 kcal/(mol·Å<sup>2</sup>), and the off-set to 0.92 kcal/mol·m.

The changes in configurational entropy upon ligand association were estimated by an all-atom normal-mode analysis performed with the AMBER NMODE module. Prior to the normal mode calculations, the complex, receptor, and ligand were subjected to minimization with a distance dependent dielectric constant  $\epsilon_r$  and convergence tolerance tighter than a root-mean-squared gradient of drms 10<sup>-4</sup> kcal/(mol·Å). Entropy and enthalpy calculations on all peptidase-ligand systems are performed separately and averaged over equally spaced snapshots, extracted over the final 4.005 ns of the production phase. The mean of the binding free enthalpies and entropies of all the snapshots were computed and then summed to obtain the binding free energy. For the enthalpy calculations, snapshots were taken every 10.5 ps (381 frames), for the entropy calculations, snapshots were taken every 190.5 ps (21 frames). VMD [49] was used to prepare the initial simulation system and analyse trajectories. Structural figures were prepared with PyMol [50]. Simulaid (<http://atlas.physbio.mssm.edu/~mezei/>) was used in the preparation of the truncated octahedral cell [35] and to convert the NAMD dcd format MD coordinate trajectories to AMBER format for the MM/PBSA analysis.

## Results and Discussion

### *F. hepatica* FhCL1, FhCL2 and FhCL3 exhibit distinct substrate specificities

Functionally active recombinant forms of the major cathepsin L peptidases of *F. hepatica*, FhCL1, FhCL2 and FhCL3, were expressed in yeast and isolated to homogeneity as previously described [22,23]. To compare their biochemical substrate specificity the kinetic parameters (*K<sub>m</sub>*, *k<sub>cat</sub>*, and *k<sub>cat</sub>/K<sub>m</sub>*) for each enzyme was determined against a range of small fluorogenic peptide (predominantly tripeptide) substrates (Table 2).

FhCL1 most efficiently cleaved substrates containing hydrophobic residues at the P2 position such as the dipeptides Z-Leu-

Arg-NHMec (*k<sub>cat</sub>/K<sub>m</sub>* 1,492,354 M<sup>-1</sup> s<sup>-1</sup>), Z-Phe-Arg-NHMec (*k<sub>cat</sub>/K<sub>m</sub>* 64,912 M<sup>-1</sup> s<sup>-1</sup>) and tripeptide Boc-Val-Leu-Lys-NHMec (*k<sub>cat</sub>/K<sub>m</sub>* 54,266 M<sup>-1</sup> s<sup>-1</sup>). In contrast, tripeptide substrates containing Pro at the P2 position, including Tos-Gly-Pro-Arg-NHMec (*k<sub>cat</sub>/K<sub>m</sub>* 671 M<sup>-1</sup> s<sup>-1</sup>), Boc-Ala-Gly-Pro-Arg-NHMec (*k<sub>cat</sub>/K<sub>m</sub>* 673 M<sup>-1</sup> s<sup>-1</sup>), Tos-Gly-Pro-Lys-NHMec (*k<sub>cat</sub>/K<sub>m</sub>* 612 M<sup>-1</sup> s<sup>-1</sup>) and Boc-Val-Pro-Arg (*k<sub>cat</sub>/K<sub>m</sub>* 478 M<sup>-1</sup> s<sup>-1</sup>), were cleaved relatively poorly (Table 2).

In comparison to FhCL1, substrates with Phe and Leu in the P2 position were much less effectively cleaved by FhCL2 and even less so by FhCL3. The *k<sub>cat</sub>/K<sub>m</sub>* values for FhCL2 and FhCL3 against Z-Phe-Arg-NHMec were 6- and 65-fold lower, respectively, than that observed for FhCL1. Similarly, the *k<sub>cat</sub>/K<sub>m</sub>* values for Z-Leu-Arg-NHMec were 3.5- and 66-fold lower than FhCL1 for FhCL2 and FhCL3 respectively. By contrast, FhCL2 and FhCL3 cleaved Pro-containing substrates much more readily than FhCL1 with *k<sub>cat</sub>/K<sub>m</sub>* values of 18,559 M<sup>-1</sup> s<sup>-1</sup> (28-fold greater, FhCL2) and 95,774 M<sup>-1</sup> s<sup>-1</sup> (142-fold increase, FhCL3) for Tos-Gly-Pro-Arg-NHMec; 58,027 M<sup>-1</sup> s<sup>-1</sup> (86-fold increase, FhCL2) and 60,763 M<sup>-1</sup> s<sup>-1</sup> (90-fold increase, FhCL3) for Boc-Ala-Gly-Pro-Arg-NHMec; 13,746 M<sup>-1</sup> s<sup>-1</sup> (22-fold increase, FhCL2) and 36,419 M<sup>-1</sup> s<sup>-1</sup> (60-fold increase, FhCL3) for Tos-Gly-Pro-Lys-NHMec and 21,193 M<sup>-1</sup> s<sup>-1</sup> (44-fold increase, FhCL2) and 1,564 M<sup>-1</sup> s<sup>-1</sup> (3-fold increase, FhCL3) for Boc-Val-Pro-Arg-NHMec (Table 2). Collectively, these data highlight significant differences in the substrate specificity of the three major *F. hepatica* cathepsin L peptidases. More specifically, the data demonstrates that FhCL3 prefers a bulky Pro residue in the P2 position of substrates over hydrophobic residues such as Leu or Phe, while FhCL2 can readily accept Pro despite preferring hydrophobic moieties at P2, and FhCL1 has an inverse preference to FhCL3.

### Kinetic analyses of recombinant *F. hepatica* peptidases with specific inhibitors

Peptidyl diazomethyl ketones are irreversible inhibitors of cysteine peptidases [51]. Changes in rates of inactivation by these inhibitors have highlighted different specificities at subsites of cysteine peptidases such as cathepsin L and cathepsin B [52]. In this study, we measured the rates of inactivation of FhCL1, FhCL2 and FhCL3 by the cathepsin inhibitor Z-Phe-Ala-CHN<sub>2</sub>. Both FhCL1 and FhCL2 were rapidly inactivated by Z-Phe-Ala-CHN<sub>2</sub> with the rate of inactivation of FhCL1 being almost 2-fold higher than that of FhCL2 (Table 3). This is in accordance with our previous data [22] and demonstrates that FhCL1 accommodates hydrophobic P2 residues more effectively than FhCL2. In contrast, the rate of inactivation of FhCL3 by Z-Phe-Ala-CHN<sub>2</sub> was 20-fold times lower, showing that Z-Phe-Ala-CHN<sub>2</sub> is a poor inhibitor of

**Table 3.** Inhibition parameters of Z-Phe-Ala-CHN<sub>2</sub> and cathepsin K inhibitor II against recombinant *F. hepatica* cathepsin Ls.

Enzyme	Inhibitor	K <sub>i(opp)</sub> nM	±	K <sub>i</sub> nM	±
FhCL1	Z-Phe-Ala-CHN <sub>2</sub>	31.4	0.61	1.62	0.03
FhCL2	Z-Phe-Ala-CHN <sub>2</sub>	55.2	4.3	5.32	0.42
FhCL3	Z-Phe-Ala-CHN <sub>2</sub>	969.3	67.8	336.5	23.2
FhCL1	Cathepsin K inhibitor II	12.2	0.75	0.63	0.03
FhCL2	Cathepsin K inhibitor II	4.8	1.0	0.46	0.09
FhCL3	Cathepsin K inhibitor II	26.1	0.3	9.05	0.15

doi:10.1371/journal.pntd.0001012.t003

FhCL3 (Table 3). This is in agreement with our kinetic substrate data using peptidyl fluorogenic substrates (Table 2) that revealed the poor capacity of FhCL3 to accommodate hydrophobic residues in the P2 position.

The inhibitor known as cathepsin K Inhibitor II (Z-LNHNHCONHNHLF-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 Leu occupies the S2 subsite [53]. FhCL1 and FhCL2 were both potently inhibited by cathepsin K inhibitor II with Ki values of 0.63 nM and 0.46 nM respectively. In contrast, CKII was 14-fold less effective against FhCL3 (Ki 336 nM) compared to FhCL1 and 20-fold less effective compared to FhCL2 (Table 3). The data are consistent with the kinetic data for hydrolysis of peptidyl fluorogenic substrates as both FhCL1 and FhCL2 had high  $k_{cat}/K_m$  values for Z-Leu-Arg-NHMec whereas that of FhCL3 against this substrate was much lower (Table 2).

### Recombinant FhCL2 and FhCL3, but not FhCL1, cleave native collagen at physiological pH

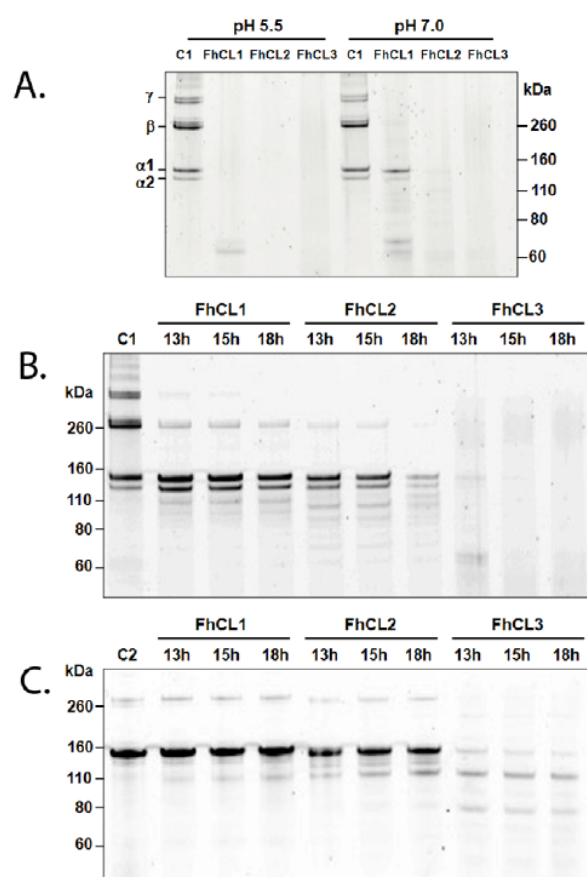
The  $\alpha$  chains of collagens are woven together to form triple helical, or Col, regions of collagen. These are flanked by non-collagenous, or non-helical, regions termed NC domains Type I and type II collagens are most abundant in nature and are the major components of vertebrate connective tissue. They share ~70% primary sequence identity and are composed largely of repeating Gly-X-Y motifs [27]. FhCL1, FhCL2 and FhCL3 effectively degraded type I collagen at pH 5.5 which induces a denaturation of the protein's helical Col structure. However, FhCL1 was much less able to degrade type I collagen at pH 7.0, where its native structure is preserved, and its activity was limited to the  $\beta$  and  $\gamma$  chains of the NC domains leaving the  $\alpha 1$  and  $\alpha 2$  chains of the Col domain intact (Fig. 1A). By contrast, both FhCL2 and FhCL3 degraded native collagen at pH 7.0 and cleaved efficiently within the Col helical structures as revealed by the breakdown of the  $\alpha 1$  and  $\alpha 2$  chains (Fig. 1A).

To determine the relative activity of FhCL1, FhCL2 and FhCL3 for collagen type I, digests were performed at pH 7.0 over a time course (up to 18 h) at 28°C (Fig. 1B). Only FhCL3 was capable of completely digesting collagen type I after 18 h incubation in these conditions. FhCL2 digested collagen  $\alpha$  chains to a lesser extent than FhCL3 while FhCL1 only digested the  $\beta 11$  and  $\beta 12$  dimers but not the collagen  $\alpha$  chains (Fig. 1B). Similarly, only FhCL3 was capable of degrading type II collagen whilst FhCL2 displayed much less activity against this substrate at pH 7.0 (Fig. 1C). FhCL1 was unable to cleave within the tightly wound type II collagen helices under these conditions (Fig. 1C).

### FhCL2 and FhCL3 cleave at different sites within the native collagen structure

To identify the cleavage sites for FhCL2 and FhCL3 within collagen type I  $\alpha 1$  and  $\alpha 2$  chains, the 18 h reaction aliquots (shown in Fig. 1B) were analysed by tandem mass spectrometry to determine the masses and sequence identities of the resulting hydrolytic products. Liberated peptides were mapped onto the primary amino acid sequence of bovine collagen to identify the cleavage sites of the *F. hepatica* peptidases (Fig. 2). FhCL2 cleaved collagen type I at 43 sites within the  $\alpha 1$  chain and 26 sites within the  $\alpha 2$  chain while FhCL3 cleaved at 24 sites within the  $\alpha 1$  chain and 24 sites within the  $\alpha 2$  chain. Strikingly, only three of these cleavage sites were shared between FhCL2 and FhCL3, all of which occurred in the  $\alpha 1$  chains (Fig. 2).

We examined the frequency of each amino acid in the P1, P2 and P3 position of the proteolytic cleavage sites identified in the



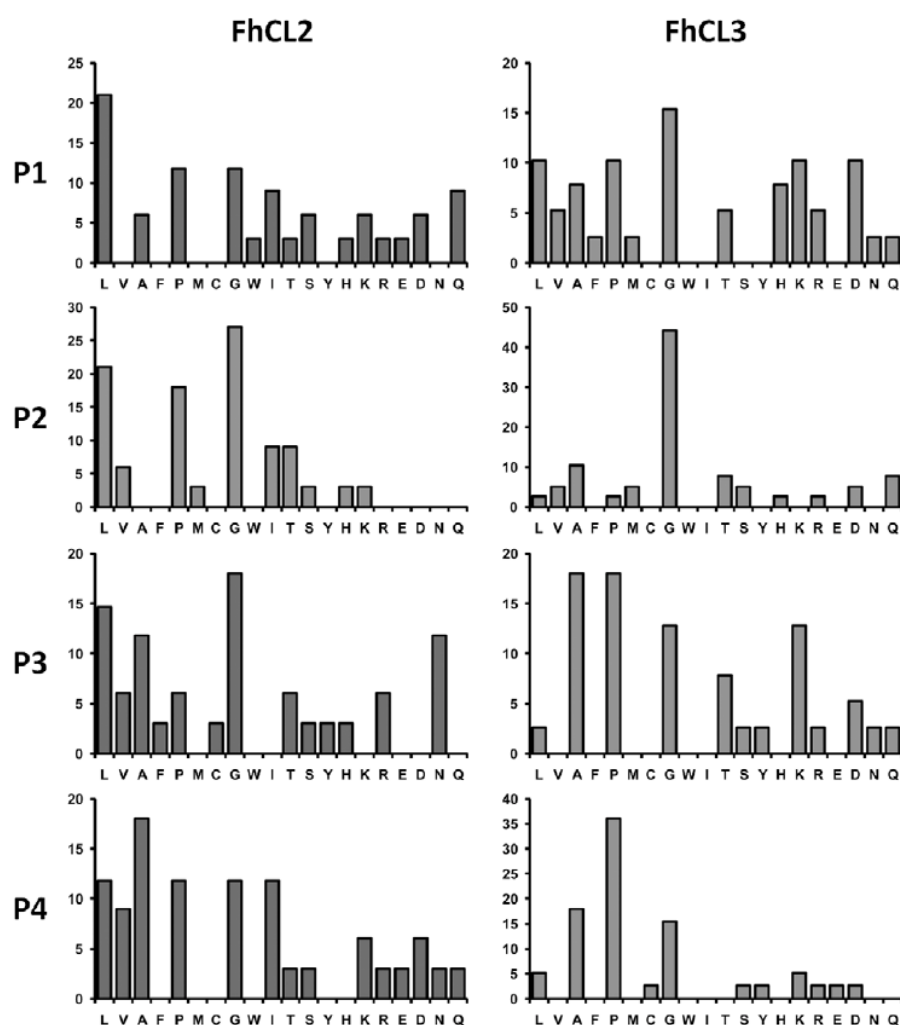
**Figure 1. Comparison of the collagen cleaving activities of recombinant *F. hepatica* FhCL1, FhCL2 and FhCL3.** (A) Type I collagen was incubated with 2  $\mu$ M FhCL1, FhCL2 and FhCL3 for either 3 h (pH 5.5) or 20 h (pH 7) at 28°C. Type I collagen (B) and type II collagen (C) were incubated with 2  $\mu$ M FhCL1, FhCL2 and FhCL3 at pH 7 for up to 18 h at 28°C. Reaction aliquots were analysed on 4–12% NuPage Bis-Tris gels (Invitrogen). C1, type I collagen alone; C2, type II collagen alone.

doi:10.1371/journal.pntd.0001012.g001

collagen digests described above to determine preferences for binding their respective active site S1, S2 and S3 subsites (Fig. 3). While substrate residues present at the P2 position from the scissile bond interact with the S2 subsite of the active site of papain-like cysteine peptidases are considered most critical in determining the efficiency by which the P1-P1' bond is cleaved [54], the binding of these residues are influenced by residues in the P3 positions. Consistent with our previous findings using positional scanning of synthetic combinatorial libraries the P1 position can be occupied by many different amino acids without a strong preference [22]. However, specificity is observed in the P2 position; Gly was most commonly found in the P2 position of the FhCL2 cleavages (27%), and this was followed by Leu (21%) and Pro (18%) (Fig. 3). By contrast, FhCL3 displayed a highly specific preference for Gly at the P2 position (44% of all cleavages) with a weak preference for all other amino acids including Leu and Pro in this position (3% for both residues, Fig. 3). The P3 and P4 positions were occupied by a wide range of amino acids.

To further investigate the cleavage of collagen by FhCL2 and FhCL3, the ability of both enzymes to cleave two small peptide substrates, Z-Gly-Pro-Leu-Gly-Pro and Z-Gly-Pro-Gly-Gly-Pro-Ala, mimicking the repeating Gly-X-Y motifs (where X is often





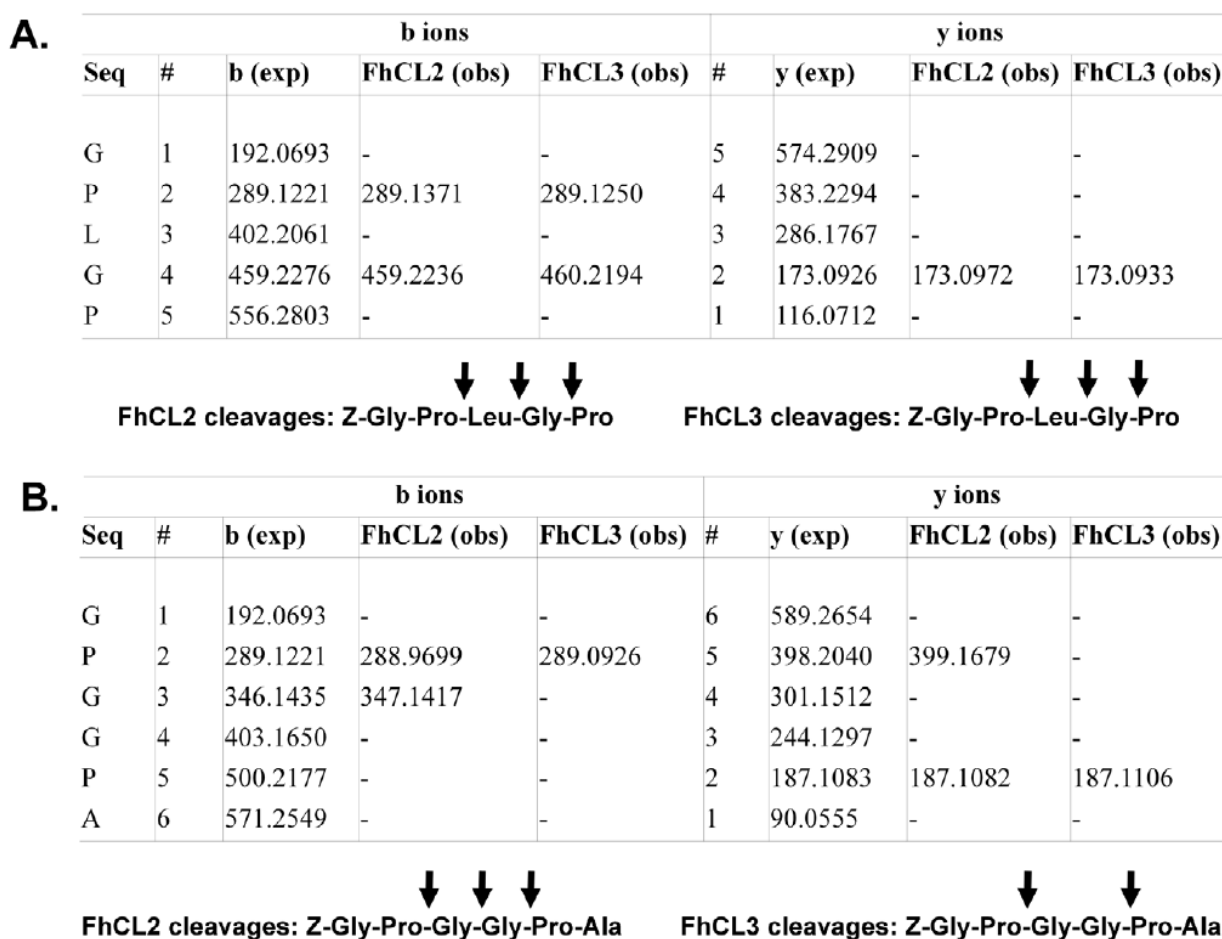
**Figure 3. Collagen P1-P4 residues.** Residues occupying the P1-P4 positions in peptides released from type I collagen following digestion by FhCL2 and FhCL3 as determined by LC-MS/MS analysis of digest samples (see Fig. 2). The frequency by which amino acids occur at the P1-P4 positions of collagen  $\alpha 1$  and  $\alpha 2$  chains (converted to a percentage of the total) are plotted for the 18 h reactions shown in Fig. 1B. doi:10.1371/journal.pntd.0001012.g003

Analysis of the cleavage sites within native type I collagen show that both FhCL2 and FhCL3 have a strong preference for Gly at P2, most particularly for the latter enzyme (Figs. 2 and 3). To investigate the molecular basis of this preference, simulations were performed using ligands with Gly at P2 (ligands D and E). The simulations with Gly at P2 generally showed markedly greater S-C distances than were observed in the complexes with Leu or Pro at P2 (Table 5). This supports the idea that the P2-S2 interaction has a strong influence on the S-C interaction. The binding affinity of collagen-like ligand D (which has Ala at P3, PAGP\*AGP) is substantially higher for FhCL2 compared wildtype FhCL1, but when in complex with FhCL3, ligand D essentially disengages. However, when ligand D is altered such that Leu occurs at P3 (ligand E, PLGP\*AGP), binding affinity to FhCL3 is restored but greatly reduced in the complex with FhCL2. These results further support the idea that the interaction of ligand residue P3 with the peptidase is a significant factor in ligand binding, and possibly of greater importance when Gly is at P2.

Collagens comprise polypeptide chains containing the repeating triplet sequence Gly-Pro-Y where 4-hydroxyproline (Hyp) com-

monly occupies the Y position [57]. Thus, simulations of complexes with a ligand containing Hyp at P3 and P1' and Gly at P2 (ligand F, PPGP\*PGP) were performed. For the FhCL2 variant, the ligand began to disengage from the peptidase whilst for FhCL3 the ligand remained closely bound. Although the binding affinity for the FhCL3-ligand F complex was not calculated, the average S-C distance was much lower than for the other complexes with Gly at P2 (Table 5). Moreover, the plot of the S-C distance frequencies showed high frequencies of very close approach for the FhCL3-ligand F complex (Fig. 7). These data suggest that FhCL3 is able to digest collagen with Hyp-Gly at P3-P2 whereas FhCL2 cannot. This may explain why we observed a greater ability of FhCL3 to digest type I and II collagen compared to FhCL2 (Fig. 1).

The FhCL3-ligand F simulation trajectories revealed that the side-chain of Trp 67 occupies the FhCL3 S2 subsite and sits against the peptide backbone of the ligand P2 Gly (Fig. 6A). This may stop solvent from entering the S2 subsite and interacting with the ligand P2 Gly, thus "sealing" the ligand in the enzyme's binding cleft. A similar disposition of the Trp 67 side-chain was



**Figure 4. LC-MS/MS data for the cleavage of small collagen-like peptide substrates by recombinant FhCL2 and FhCL3.** Substrates Z-Gly-Pro-Leu-Gly-Pro (A) and Z-Gly-Pro-Gly-Gly-Pro-Ala (B) were incubated with FhCL2 and FhCL3 for 15 min at pH 4.5 and the reactions were analysed by LC-MS/MS. The resulting peptide fragment masses (obs) that matched with y and b ions derived from theoretical fragmentation of the peptide substrates (exp) were used to map the FhCL2 and FhCL3 cleavage sites (arrows).  
doi:10.1371/journal.pntd.0001012.g004

also observed in the FhCL3-ligand E simulations. Tyr at position 67 in FhCL2 behaves in a similar manner to the Trp 67 of FhCL3 when binding ligands with Gly at P2. Thus, Tyr 67 occupied the S2 subsite cleft and contacted the peptide backbone of the ligand

P2 Gly in the FhCL2-ligand D complex (Fig. 6B). The position of the Tyr side-chain was further stabilised by a hydrogen bond between its hydroxyl oxygen and the backbone oxygen of residue 157.

**Table 4. *F. hepatica* cathepsin L variants used for the MD simulations.**

Subsite	S3	S2	S2	S1'	S2	
Residue	61	67	157	158	205	
Structure						Description
FhCL1	Asn	Leu	Val	Asn	Leu	Wild-type FhCL1
FhCL2	Tyr	Tyr	Leu	Thr	Leu	FhCL2 S1', S2, and S3 subsites
FhCL3	His	Trp	Val	Thr	Val	FhCL3 S1', S2, and S3 subsites
HK	Asp	Tyr	Leu	Asn	Leu	Human cathepsin K
GZ	His	Trp	Ala	Asn	Phe	Ginger rhizome peptidase

FhCL1 variants used for the MD simulations and their relationship to the substrate binding subsites of FhCL2 and FhCL3 variants. Substituted residues are in red and papain numbering is used for residue positions. For reference, the equivalent residues in the collagenolytic cathepsins human cathepsin K and ginger rhizome peptidase are also given.

doi:10.1371/journal.pntd.0001012.t004

**Table 5.** Computed free energies of binding and average nucleophilic sulphur-Scissile carbon distances<sup>a</sup>.

Ligand:		A AALR*NAA	B AAPR*NAA	C AGPR*NAA	D PAGP*AGP	E PLGP*AGP	F PPGP*PGP
FhCL1	$\Delta G$	-10.83 (1.18)	-9.10 (1.30)		-8.30 (0.96)		
	$\Delta d$	3.96 (0.24)	4.00 (0.25)		4.41 (0.30)		
FhCL2	$\Delta G$	-11.77 (0.75)	-6.30 (1.19)	-10.88 (0.96)	-10.83 (1.30)	-4.11 (1.02)	- <sup>b</sup>
	$\Delta d$	3.97 (0.24)	3.95 (0.24)	3.94 (0.23)	4.27 (0.34)	4.82 (0.51)	> 10
FhCL3	$\Delta G$	-10.40 (0.84)	-8.28 (0.82)	-12.71 (0.95)	-	-9.56 (1.20)	- <sup>b</sup>
	$\Delta d$	3.99 (0.26)	4.06 (0.25)	4.05 (0.25)	> 10	4.40 (0.28)	4.08 (0.5)

<sup>a</sup>Amino acid sequences of peptide ligands A–F are shown in single letter code with P2 in red and the position of the scissile bond shown by an asterisk; Hyp is indicated by P in ligand F. FhCL1 variants are shown in the leftmost column and are denoted as in Table 4. Mean energies ( $\Delta G$ ) are in kcal/mol, with corresponding standard error of the mean in parentheses. The average distances between the sulphur atom of the catalytic Cys and the backbone carbonyl carbon of scissile ligand residue 4 ( $\Delta d$ ) are in Ångstrom, with corresponding standard deviations in parentheses. All measures calculated over the final 4.005 ns of the simulation. Standard errors of the mean for all  $\Delta d$  measures are <0.01 Å;

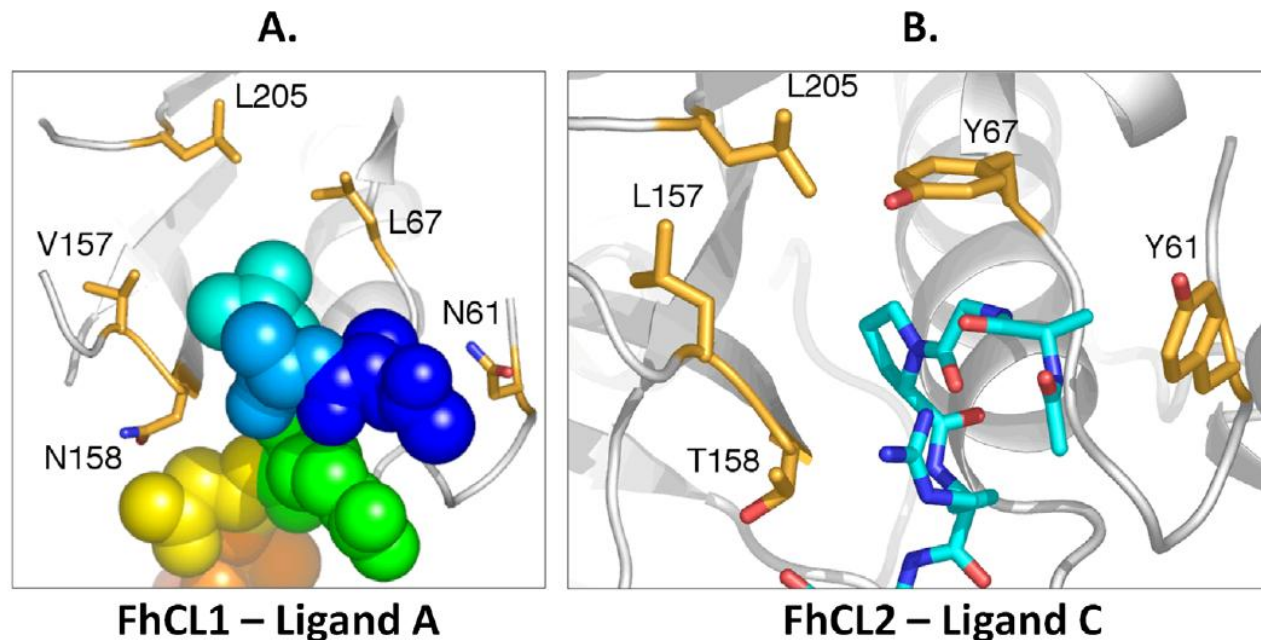
<sup>b</sup>Not calculated, see Methods.

doi:10.1371/journal.pntd.0001012.t005

*F. hepatica* has evolved a repertoire of cathepsin L peptidases as a result of gene duplication and diversification that exhibit subtle but distinct substrate specificities [1,8,14,15]. The expression of different members of this peptidase family is temporally regulated suggesting that they perform precise functions at different stages of the parasites' development [13]. This idea is supported by our previous data showing that the predominant enzyme, FhCL1, secreted by the mature adult parasites, which are obligate blood-feeders, is adapted to the degradation of host haemoglobin; the S2 subsite of the FhCL1 active site, which contributes mostly to substrate binding, readily accommodates P2 residues such as Leu,

Ala, Val and Phe that together represent >40% of the residues present in haemoglobin [3].

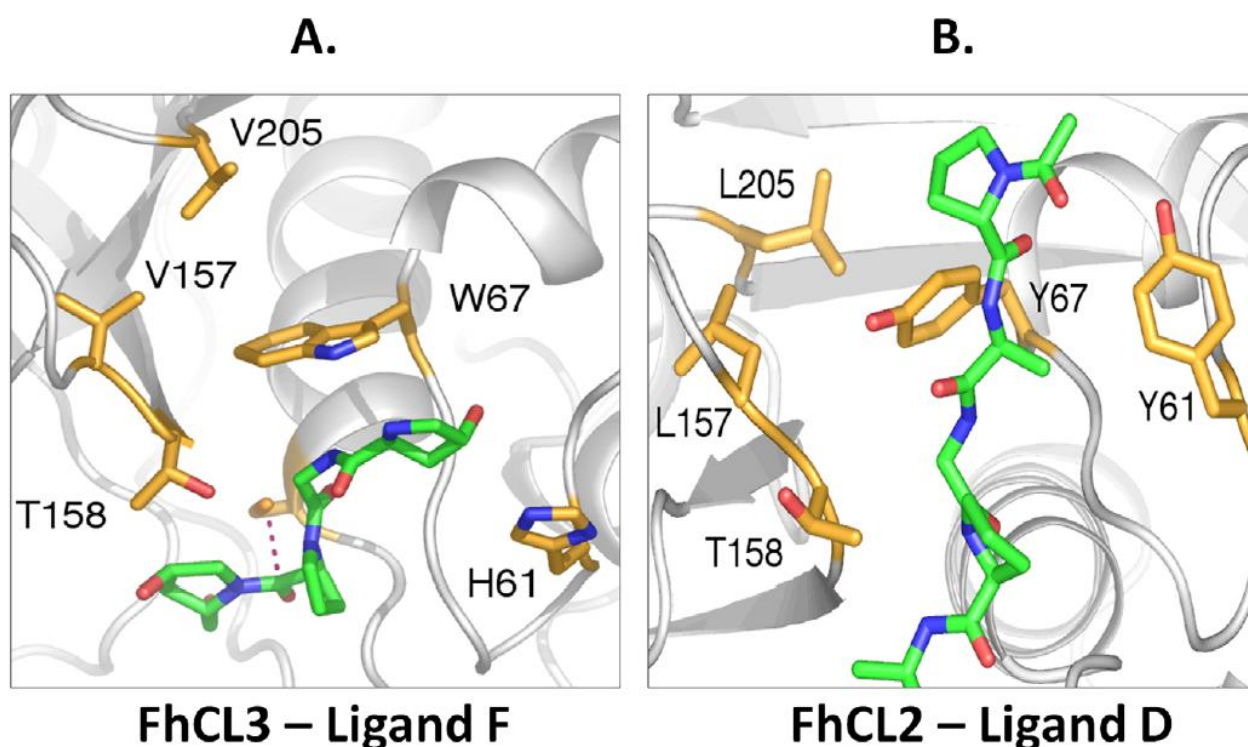
FhCL1 does not readily accept Pro into the S2 subsite as shown in this and other [22,28] studies and thus its activity against type I and II collagens observed here was restricted to the non-collagenous, NC, domains. By contrast, both FhCL2 and FhCL3 have evolved to accommodate Pro in the S2 subsite of their active sites; this property has been attributed to the presence of Tyr and Trp, respectively, at position 67 within the S2 subsite of these enzymes, a position that is occupied by Leu in FhCL1 [1,14,15,22,23]. In this study, our computational data show that



**Figure 5. Disposition of residues within ligand binding subsites S1', S2 and S3.** (A) Final trajectory frame from the MD simulation of Ligand A (AALR\*NAA) bound to wild type FhCL1, with Leu at P2. The ligand is shown in space filling representation and coloured in a spectrum pattern, with N-terminal residue 1 blue and C-terminal residue 7 red. The side-chain of P2 residue Leu (cyan) is bound in the S2 subsite. The side-chains of residues mutated in the MD simulations are shown in stick form with carbon atoms tan, oxygen red and nitrogen blue, with papain numbering. Secondary structural elements of the peptidase are shown in cartoon representation in grey. (B) Final trajectory frame from the MD simulation of Ligand C (AGPR\*NAA) bound to FhCL2, with Pro at P2. Interaction of Tyr 67 with Pro bound in subsite S2 is observed. The ligand is shown in stick form with carbon atoms cyan, oxygen red and nitrogen blue.

doi:10.1371/journal.pntd.0001012.g005





**Figure 6. Binding mode of Gly at P2 in the FhCL2 and FhCL3 S2 subsites.** The side-chains of residues mutated in the MD simulations are shown in stick form with carbon atoms tan, oxygen red and nitrogen blue, with papain numbering. The ligand is shown in stick form with carbon atoms green, oxygen red and nitrogen blue. Secondary structural elements of the peptidase are in shown in cartoon representation in grey. (A) Frame from the final 4 ns of the FhCL3-ligand F (PPGP\*PGP) simulation showing the interaction of the Trp (W) 67 side-chain within the S2 subsite cleft with the ligand P2 Gly (G). The N-terminal acyl group and Pro residue of ligand F are omitted for clarity. The S-C distance is indicated by a red dashed line. (B) Final frame from the FhCL2-ligand D (PAGP\*AGP) simulation showing the interaction of the Tyr (Y) 67 side-chain within the S2 subsite cleft with the ligand P2 Gly (G).

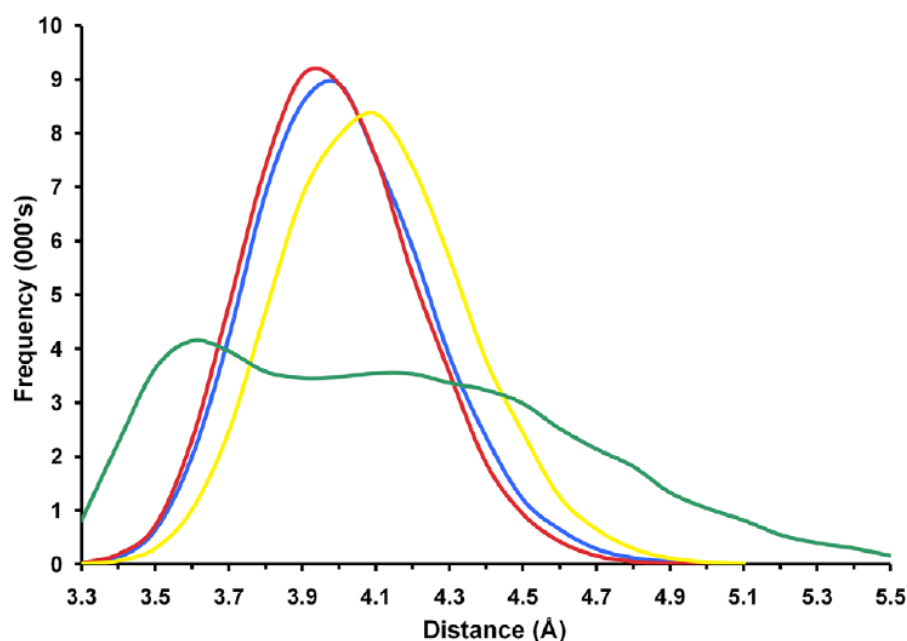
doi:10.1371/journal.pntd.0001012.g006

Tyr and Trp at position 67 have the ability to function in distinct ways to accommodate either Gly or Pro residues at P2 and explains why FhCL2 and FhCL3 have an ability to degrade the Gly-X-Y containing Col helices of collagen. The results are also in accordance with our previous suggestion that the interaction of substrate residue P3 with the peptidase is a significant factor in substrate binding, in particular when Gly or Pro is at P3 [23]. This is also the case when we compare FhCL3 specificity towards synthetic peptides; Pro is readily accepted in P2 only when Gly is at P3 (Tos-Gly-Pro-Arg-NHMec, Tos-Gly-Pro-Lys-NHMec and Boc-Ala-Gly-Pro-Arg-NHMec) but not when Val is at P3 (Boc-Val-Pro-Arg-NHMec) (Table 2). Therefore, the collagenolytic activity of FhCL2 and FhCL3 is not due simply to the P2 Pro-S2 subsite interaction, and with Pro at P2, Gly at residue P3 is critical. A comparison of the cleavage sites of FhCL2, FhCL3, human cathepsin K [56] and ginger rhizome GP2 [29] revealed that many of their cleavage sites within collagen where Pro is at P2, a Gly is present at P3. A P3 Gly, which lacks a side-chain, offers minimal steric interference with the large active site Tyr or Trp side-chain at position 67, and allows the Tyr or Trp ring to form a “lid” over the ligand’s P2 Pro ring, helping to sequester it in the S2 subsite (Fig. 6).

However, we observed that FhCL3 digested type I and II collagens more efficiently compared to FhCL2 and that these two enzymes cleave at mostly different sites (see Figs. 1 and 2). The computational data indicates that this may be, in part, because FhCL3 binds substrates containing a P3 and P1’ Hyp much

tighter than FhCL2. A difference between these two enzymes was also observed using a peptide substrate that mimics the Gly-X-Y repeat in the collagen Col domain, Z-Gly-Pro-Gly-Gly-Pro-Ala; FhCL2 cleaves at three sites with Gly or Pro in the P2 position, whereas FhCL3 was unable to cleave at one of these three sites despite having Pro occupying the P2 position and Gly at the P1 position. This result contrasts with our data using fluorogenic peptide substrates which showed that FhCL3 cleaved the tripeptides Tos-Gly-Pro-Arg-NHMec and Tos-Gly-Pro-Lys-NHMec with 5- and 3-fold better efficiency, respectively, than FhCL2. On the other hand, the two enzymes exhibited equal efficiency for the substrate Boc-Ala-Gly-Pro-Arg-NHMec. The influence of P4 and P’ regions of the peptides on substrate binding in these two enzymes need greater attention in future studies when suitable reagents become available. Notwithstanding, it is clear that the modification within the active site of FhCL2 and FhCL3 (Tyr or Trp at position 67) has subtly altered the substrate specificity of the two enzymes such that they exhibit different substrate profiles without compromising their unique ability to degrade host native collagen.

FhCL3 is expressed by the invasive stage of *F. hepatica* which must quickly penetrate the wall of the intestine to enter its host [1,13]. RNAi-mediated knockdown experiments have demonstrated that the secretion of this peptidase and a cathepsin B cysteine peptidase by these invasive parasites is critical to invasion of the intestinal tissue [2]. Once the intestine has been traversed expression of these enzymes is switched off and the parasite up-



**Figure 7. Frequency distributions of nucleophilic sulphur-scissile carbon distances.** Smoothed histogram plots of the frequencies of distances between the sulphur atom of the active site nucleophilic sulphur atom and the ligand P1 backbone carbonyl carbon atom. Measures taken over the final 4 ns of each simulation at 0.075 ps intervals (53334 measures per simulation); frequencies assigned to 0.1 Å bins between 3.3 Å and 5.5 Å. Simulation FhCL1-ligand A blue, FhCL2-ligand C red, FhCL3-ligand C yellow, FhCL3-ligand F green. doi:10.1371/journal.pntd.0001012.g007

regulates expression and secretion of FhCL1 and FhCL2 which are required to facilitate tunnelling through the liver mass and feeding on host tissue (the parasite undergoes rapid growth at this stage) [16]. The collagenolytic activity of FhCL3 and FhCL2 is important in degrading the extracellular matrix of the tissues through which this parasite moves. While collagenase activity has been demonstrated in ginger rhizome cysteine peptidases [29,58], only one other animal cysteine peptidase, human cathepsin K which functions in bone re-modelling [59], possesses collagenase activity. Accordingly, the evolution of this activity in *F. hepatica* must represent an important step in the development of a parasitic way of life.

### Supporting Information

**Figure S1 Residues contacting the ligand.** Final trajectory frame from the MD simulation of Ligand A (AALR\*NAA) bound to wild type FhCL1, with Leu at P2. The ligand is shown in stick

form with carbon atoms magenta, oxygen red and nitrogen blue. The side-chain of P2 residue Leu (cyan) is bound in the S2 subsite. Secondary structural elements of the protease in cartoon representation and the side-chains of residues that contact the ligand in the MD simulations shown in stick form. Sidechain atoms or backbone segments of the peptidase which contacted the ligand are coloured according to the relative frequency of contact over the final 4 ns with yellow (low) to green (high). (TIF)

### Author Contributions

Conceived and designed the experiments: MWR IC PMJ AMG MC GR JFT LR JPD. Performed the experiments: MWR IC PMJ MPP JT. Analyzed the data: MWR IC PMJ MPP JFT JPD. Contributed reagents/materials/analysis tools: MWR IC PMJ JPD. Wrote the paper: MWR PMJ JPD.

### References

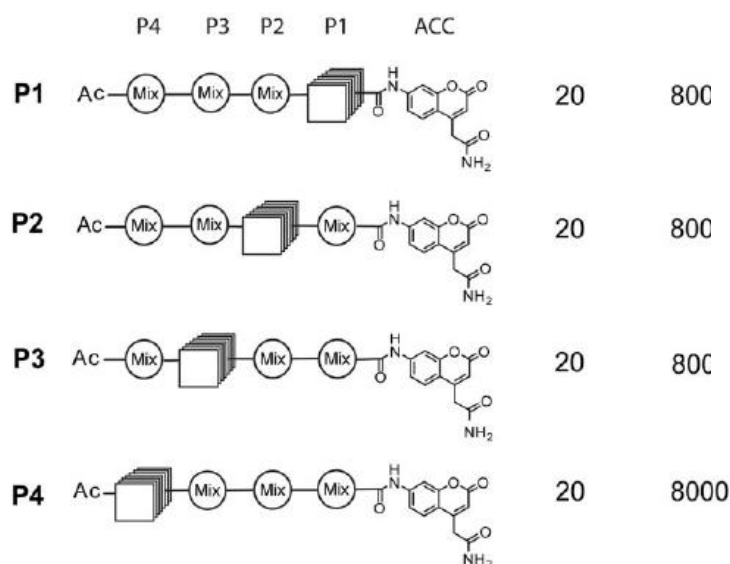
- Robinson MW, Dalton JP, Donnelly S (2008) Helminth pathogen cathepsin proteases: it's a family affair. *Trends Biochem Sci* 33: 601–608.
- McGonigle L, Mousley A, Marks NJ, Brennan GP, Dalton JP, et al. (2008) The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *Int J Parasitol* 38: 149–155.
- Lowther J, Robinson MW, Donnelly SM, Xu W, Stack CM, et al. (2009) The importance of pH in regulating the function of *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS Negl Trop Dis* 3: e369.
- Sripa J, Laha T, To J, Brindley PJ, Sripa B, et al. (2010) Secreted cysteine proteases of the carcinogenic liver fluke, *Opisthorchis viverrini*: regulation of cathepsin F activation by autocatalysis and *trans*-processing by cathepsin B. *Cell Microbiol* 12: 781–795.
- Smith AM, Dowd AJ, Heffernan M, Robertson CD, Dalton JP (1993) *Fasciola hepatica*: a secreted cathepsin L-like proteinase cleaves host immunoglobulin. *Int J Parasitol* 23: 977–983.
- Berasain P, Carmona C, Frangione B, Dalton JP, Goni F (2000) *Fasciola hepatica*: parasite-secreted proteinases degrade all human IgG subclasses: determination of the specific cleavage sites and identification of the immunoglobulin fragments produced. *Exp Parasitol* 94: 99–110.
- Donnelly S, O'Neill SM, Stack CM, Robinson MW, Tumbull L, et al. (2010) Helminth cysteine proteases inhibit TRIF-dependent activation of macrophages via degradation of TLR3. *J Biol Chem* 285: 3383–3392.
- Dalton JP, Neil SO, Stack C, Collins P, Walshe A, et al. (2003) *Fasciola hepatica* cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *Int J Parasitol* 33: 1173–1181.
- Abdulla MH, Lim KC, Sajid M, McKerrow JH, Caffrey CR (2007) Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. *PLoS Med* 4: e14.
- Robinson MW, Dalton JP (2009) Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiasis. *Phil Trans Royal Soc B* 364: 2763–2776.
- Keiser J, Utzinger J (2005) Emerging foodborne trematodiasis. *Emerg Infect Dis* 11: 1507–1514.
- Andrews SJ (1999) The life-cycle of *Fasciola hepatica*. In: Dalton JP, ed. *Fasciolosis* CABI: Oxford, United Kingdom. pp 1–29.

13. Cancela M, Acosta D, Rinaldi G, Silva E, Durán R, et al. (2008) A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie* 90: 1461–1475.
14. Robinson MW, Tort JF, Wong E, Donnelly SM, Lowther J, et al. (2008) Proteomic and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen, *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Mol Cell Proteomics* 7: 1111–1123.
15. Irving JA, Spithill TW, Pike RN, Whisstock JC, Smooker PM (2003) The evolution of enzyme specificity in *Fasciola* spp. *J Mol Evol* 57: 1–15.
16. Robinson MW, Menon R, Donnelly SM, Dalton JP, Ranganathan S (2009) An integrated transcriptomic and proteomic analysis of the secretome of the helminth pathogen, *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. *Mol Cell Proteomics* 8: 1891–1907.
17. Cancela M, Ruetalo N, Dell'Oca N, da Silva E, Smircich P, et al. (2010) Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*. *BMC Genomics* 11: 227.
18. Young ND, Hall RS, Jex AR, Cantacessi C, Gasser RB (2010) Elucidating the transcriptome of *Fasciola hepatica* – a key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnol Adv* 28: 222–231.
19. Collins PR, Stack CM, O'Neill SM, Doyle S, Ryan T, et al. (2004) Cathepsin L1, the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propeptide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. *J Biol Chem* 279: 17038–17046.
20. Na BK, Kim SH, Lee EG, Kim TS, Bae YA, et al. (2006) Critical roles for excretory-secretory cysteine proteases during tissue invasion of *Paragonimus westermani* newly excysted metacercariae. *Cell Microbiol* 8: 1034–1046.
21. Na BK, Kang JM, Sohn WM (2008) CsCF-6, a novel cathepsin F-like cysteine protease for nutrient uptake of *Clonorchis sinensis*. *Int J Parasitol* 38: 493–502.
22. Stack CM, Caffrey CR, Donnelly SM, Sessaadri A, Lowther J, et al. (2008) Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. *J Biol Chem* 283: 9896–9908.
23. Corvo I, Cancela M, Cappetta M, Pi-Denis N, Tort JF, et al. (2009) The major cathepsin L secreted by the invasive juvenile *Fasciola hepatica* prefers proline in the S2 subsite and can cleave collagen. *Mol Biochem Parasitol* 167: 41–47.
24. Turk D, Guncar G, Podobnik M, Turk B (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol Chem* 379: 137–147.
25. Turk D, Guncar G (2003) Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta Crystallogr D Biol Crystallogr* 59: 203–213.
26. Atley LM, Mort JS, Lalumiere M, Eyre DR (2000) Proteolysis of human bone collagen by cathepsin K: characterization of the cleavage sites generating the cross-linked N-telopeptide neopeptide. *Bone* 26: 241–247.
27. Gordon MK, Hahn RA (2010) Collagens. *Cell Tissue Res* 339: 247–257.
28. Smooker PM, Whisstock JC, Irving JA, Siyaguna S, Spithill TW, et al. (2000) A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*. *Protein Sci* 9: 2567–2572.
29. Kim M, Hamilton SE, Guddat LW, Overall CM (2007) Plant collagenase: Unique collagenolytic activity of cysteine proteases from ginger. *Biochim Biophys Acta* 1770: 1627–1635.
30. Rawlings ND, Morton FR, Barrett AJ (2006) MEROPS: The peptidase database. *Nucleic Acids Res* 34: D270–272.
31. Marquis RW, Ru Y, LoCastro SM, Zeng J, Yamashita DS, et al. (2001) Azepanone-Based Inhibitors of Human and Rat Cathepsin K. *J Med Chem* 44: 1380–1395.
32. Stern I, Schaschke N, Moroder L, Turk D (2004) Crystal structure of NS-134 in complex with bovine cathepsin B: a two-headed epoxysuccinyl inhibitor extends along the entire active-site cleft. *Biochem J* 381: 511–517.
33. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18: 2714–2723.
34. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.
35. Mezei M (1997) Optimal Position of the Solute for Simulations. *J Comp Chem* 18: 812–815.
36. MacKerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25: 1400–1415.
37. Veld PJ, Stevens MJ (2008) Simulation of the Mechanical Strength of a Single Collagen Molecule. *Biophys J* 95: 33–39.
38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.
39. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101: 4177–4189.
40. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation - the Langevin piston method. *J Chem Phys* 103: 4613–4621.
41. Darden T, York D, Pedersen L (1993) Particle mesh Ewald. An N log(N) method for Ewald sums in large systems. *J Chem Phys* 98: 10089–10092.
42. Ryckaert JP, Cicotti G, Berendsen HJC (1997) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comp Phys* 23: 327–341.
43. Miyamoto S, Kollman PA (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water molecules. *J Comput Chem* 13: 952–962.
44. Wang W, Donini O, Reyes CM, Kollman PA (2001) Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu Rev Biophys Biomol Struct* 30: 211–243.
45. Stoica I, Sadiq SK, Coveney PV (2008) Rapid and accurate prediction of binding free energies of saquinavir-bound HIV-1 proteases. *J Am Chem Soc* 130: 2639–2648.
46. Case DA, Cheatham T, Darden T, Gohlke H, Luo R, et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26: 1668–1688.
47. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed phase quantum mechanical calculations. *J Comput Chem* 24: 1999–2012.
48. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38: 305–320.
49. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–38.
50. DeLano WL (2002) The PyMOL Molecular Graphics System (Delano Scientific, San Carlos, California). Available: <http://www.pymol.org>. Accessed 2010.
51. Green GD, Shaw E (1981) Peptidyl diazomethyl ketones are specific inactivators of thiol proteinases. *J Biol Chem* 256: 1923–1928.
52. Kirschke H, Shaw E (1981) Rapid interaction of cathepsin L by Z-Phe-PheCHN12 and Z-Phe-AlaCHN2. *Biochem Biophys Res Commun* 101: 454–458.
53. Wang D, Pechar M, Li W, Kopecková P, Brömme D, et al. (2002) Inhibition of cathepsin K with lysosomotropic macromolecular inhibitors. *Biochemistry* 41: 8849–8859.
54. Schechter I, Berger A (1968) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27: 157–162.
55. Schwartz SD, Schramm VL (2009) Enzymatic transition states and dynamic motion in barrier crossing. *Nat Chem Biol* 5: 551–558.
56. Kafienah W, Bromme D, Buttle DJ, Croucher IJ, Hollander AP (1998) Human cathepsin K cleaves native type I and II collagens at the N-terminal end of the triple helix. *Biochem J* 331: 727–732.
57. Krane SM (2008) The importance of proline residues in the structure, stability and susceptibility to proteolytic degradation of collagens. *Amino Acids* 35: 703–710.
58. Choi KH, Laursen RA, Allen KN (1999) The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*. *Biochemistry* 38: 11624–11633.
59. Wilson S, Hashamiyan S, Clarke L, Saftig P, Mort J, et al. (2009) Glycosaminoglycan-mediated loss of cathepsin K collagenolytic activity in MPS I contributes to osteoclast and growth plate abnormalities. *Am J Pathol* 175: 2053–2062.

#### 4. Identificación de residuos relevantes en la especificidad de sustrato de las catepsinas de *Fasciola hepatica*

##### Estrategia general

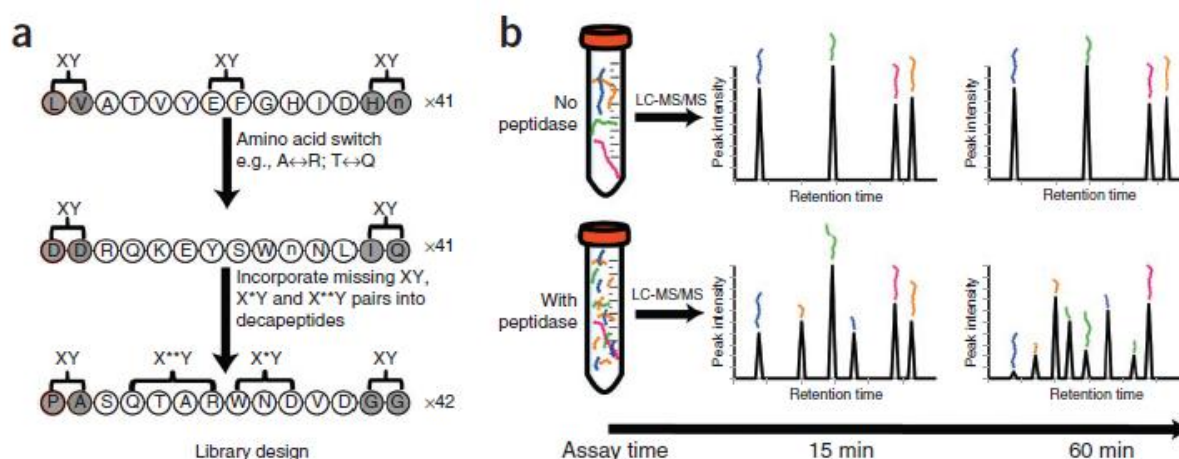
Para profundizar en el estudio de la especificidad de sustrato de *FhCL3* se emplearon dos tipos de librerías peptídicas, *Positional Scanning Synthetic Combinatorial Libraries (PS-SCL)* (Harris et al. 2000) y *Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS)* (O'Donoghue et al. 2012), de modo de ensayar todos los residuos en cada uno de los sitios de interacción con el sustrato. Las *PS-SCL* son librerías de péptidos sintéticos fluorogénicos diseñadas de manera que se mantiene fijo un único aminoácido en una de las posiciones adyacentes al sitio de corte ( $P_1$ ,  $P_2$ ,  $P_3$  o  $P_4$ ), mientras las demás posiciones contienen una mezcla equimolar de todos los aminoácidos. De esta forma, para cada una de las 4 posiciones del sustrato se tienen 20 sublibrerías con 8000 péptidos diferentes. Cada una de estas sublibrerías es testada de forma independiente pero en paralelo. De esta forma es posible determinar cuáles son los residuos preferidos por la enzima en cada subsitio utilizando todo el espectro de aminoácidos (Figura 17) (Choe et al. 2006).



**Figura 17. Diseño de la librería PS-SCL.** La librería utiliza péptidos conjugados al fluoróforo ACC (4-Aminocarbamoilmetilcumarina), no contiene cisteína pero sí norleucina. Consiste de 4 librerías  $P_1$ ,  $P_2$ ,  $P_3$  y  $P_4$  y cada una de ellas tiene 20 sublibrerías con 8000 compuestos diferentes. En total la librería cuenta con una diversidad de 160.000 tetrapéptidos (tomado de Choe et al., 2006).

Una estrategia diferente emplean las *MSP-MS*, las cuales consisten de 124 tetradecapéptidos lineales sintetizados de manera que todos los pares (XY) o casi pares ( $X^*Y$ ,  $X^{**}Y$ ) de aminoácidos estén representados al menos una vez en la librería. Estos péptidos largos son mezclados y ensayados en conjunto, y el método de detección de la hidrólisis no involucra la liberación de un fluoróforo sino que se analizan los productos de digestión por espectrometría de masas (Figura 18). Esto permite mapear todos los sitios de corte e identificar los residuos a ambos lados del sitio activo. Al ser péptidos más largos se asemejan

mejor a las regiones en forma de *loops* sensibles a hidrólisis que se hallan expuestas en la superficie de las proteínas (O'Donoghue et al. 2012).



**Figura 18. Esquema del ensayo de MSP-MS.** **a)** La librería de péptidos de 14 aminoácidos se diseña acomodando todos los pares (XY) y casi pares (X\*Y, X\*\*Y) de residuos en un decapeptido core (residuos no sombreados). X e Y corresponden a aminoácidos definidos, \* corresponde a un aminoácido al azar. Los extremos (residuos sombreados en gris) se generaron usando pares de aminoácidos (XY) seleccionados de 11 *pools* de aminoácidos. n, norleucina. **b)** Esquema del ensayo. Se incuban los péptidos con una proteasa y se toman alícuotas a distintos tiempos, deteniéndose la reacción con un inhibidor o ácido. Cada alícuota es inyectada en un sistema LC-MS/MS para detectar la hidrólisis tiempo-dependiente de los péptidos. Los sitios de corte son identificados en comparación con un control incubado sin proteasa (tomado de O'Donoghue et al., 2012).

Se obtuvieron resultados consistentes para las enzimas *FhCL1* y *FhCL3* en ambas aproximaciones, concordantes con lo reportado previamente para *FhCL1* (Stack et al. 2008), lo que reafirma que estas metodologías son adecuadas para evaluar especificidad enzimática. En base a estos resultados se emplearon las *PS-SCL* para estudiar cambios en la especificidad de *FhCL1* o *FhCL3* frente a sustituciones en residuos puntuales de los bolsillos  $S_2$  y  $S_3$ . En el artículo presentado en la sección 4.1. se realizan sustituciones en los residuos ubicados en las posiciones 61 y 67 de *FhCL3* por aquellos presentes en *FhCL1*, y los cambios recíprocos en *FhCL1* por los residuos de *FhCL3*, generando enzimas sustituidas en una o ambas posiciones. En la sección 4.2. se generan mutantes de *FhCL1* hacia *FhCL2*, estudiándose además de la posición 67 ya reportada (Stack et al. 2008), los residuos 157 y 158 del sitio activo. La evaluación de las enzimas con sustituciones se complementó con el uso de péptidos sintéticos fluorogénicos y sustratos macromoleculares, y se generaron modelos de *FhCL3* buscando comprender mejor la dinámica de las interacciones en el sitio activo.

**4.1 Artículo:** *Disecando el sitio activo de la catepsina colagenolítica L3 del estadio invasivo de Fasciola hepatica*

**Ileana Corvo**, Anthony J. O' Donoghue, Lucía Pastro, Natalia Pi-Denis, Alegra Eroy-Reveles, Leda Roche, James H. McKerrow, John P. Dalton, Charles S. Craik, Conor R. Caffrey & José F. Tort

*PLoS Neglected Tropical Diseases*, 2013. 7(7): e2269

# Dissecting the Active Site of the Collagenolytic Cathepsin L3 Protease of the Invasive Stage of *Fasciola hepatica*

Ileana Corvo<sup>1</sup>, Anthony J. O'Donoghue<sup>2</sup>, Lucía Pastro<sup>1,3</sup>, Natalia Pi-Denis<sup>1</sup>, Alegra Eroy-Reveles<sup>4</sup>, Leda Roche<sup>1</sup>, James H. McKerrow<sup>5</sup>, John P. Dalton<sup>6</sup>, Charles S. Craik<sup>2</sup>, Conor R. Caffrey<sup>5</sup>, José F. Tort<sup>1\*</sup>

**1** Departamento de Genética, Facultad de Medicina, Universidad de la República, UDELAR, Montevideo, Uruguay, **2** Department of Pharmaceutical Chemistry, Pharmacology, and Biochemistry and Biophysics, University of California, San Francisco, California, United States of America, **3** Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la República, UDELAR, Montevideo, Uruguay, **4** Department of Chemistry and Biochemistry, San Francisco State University, San Francisco, California, United States of America, **5** Center for Discovery and Innovation in Parasitic Diseases, Department of Pathology, University of California, San Francisco, California, United States of America, **6** Institute of Parasitology, McGill University, Quebec, Canada

## Abstract

**Background:** A family of secreted cathepsin L proteases with differential activities is essential for host colonization and survival in the parasitic flatworm *Fasciola hepatica*. While the blood feeding adult secretes predominantly FheCL1, an enzyme with a strong preference for Leu at the S<sub>2</sub> pocket of the active site, the infective stage produces FheCL3, a unique enzyme with collagenolytic activity that favours Pro at P<sub>2</sub>.

**Methodology/Principal Findings:** Using a novel unbiased multiplex substrate profiling and mass spectrometry methodology (MSP-MS), we compared the preferences of FheCL1 and FheCL3 along the complete active site cleft and confirm that while the S<sub>2</sub> imposes the greatest influence on substrate selectivity, preferences can be indicated on other active site subsites. Notably, we discovered that the activity of FheCL1 and FheCL3 enzymes is very different, sharing only 50% of the cleavage sites, supporting the idea of functional specialization. We generated variants of FheCL1 and FheCL3 with S<sub>2</sub> and S<sub>3</sub> residues by mutagenesis and evaluated their substrate specificity using positional scanning synthetic combinatorial libraries (PS-SCL). Besides the rare P<sub>2</sub> Pro preference, FheCL3 showed a distinctive specificity at the S<sub>3</sub> pocket, accommodating preferentially the small Gly residue. Both P<sub>2</sub> Pro and P<sub>3</sub> Gly preferences were strongly reduced when Trp67 of FheCL3 was replaced by Leu, rendering the enzyme incapable of digesting collagen. In contrast, the inverse Leu67Trp substitution in FheCL1 only slightly reduced its Leu preference and improved Pro acceptance in P<sub>2</sub>, but greatly increased accommodation of Gly at S<sub>3</sub>.

**Conclusions/Significance:** These data reveal the significance of S<sub>2</sub> and S<sub>3</sub> interactions in substrate binding emphasizing the role for residue 67 in modulating both sites, providing a plausible explanation for the FheCL3 collagenolytic activity essential to host invasion. The unique specificity of FheCL3 could be exploited in the design of specific inhibitors selectively directed to specific infective stage parasite proteinases.

**Citation:** Corvo I, O'Donoghue AJ, Pastro L, Pi-Denis N, Eroy-Reveles A, et al. (2013) Dissecting the Active Site of the Collagenolytic Cathepsin L3 Protease of the Invasive Stage of *Fasciola hepatica*. PLoS Negl Trop Dis 7(7): e2269. doi:10.1371/journal.pntd.0002269

**Editor:** Malcolm K. Jones, University of Queensland, Australia

**Received:** January 14, 2013; **Accepted:** March 23, 2013; **Published:** July 11, 2013

**Copyright:** © 2013 Corvo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported financially by CSIC-UDELAR, ANII and PEDECIBA. JPD is recipient of a Discovery Grant from the National Science and Engineering Research Council (NSERC), Canada. Support for AER and the SFSU mass spectrometry facility was provided by National Institutes of Health P20 MD000544 and National Science Foundation Grant CHE-1228656, respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

\* E-mail: jtort@fmed.edu.uy

## Introduction

The common liver fluke *F. hepatica*, together with *F. gigantica*, are the causative agents of fascioliasis, a zoonosis causing huge global losses in the agricultural sector by infecting more than 700 million ruminants worldwide. The disease is also recognized by the WHO as an important emerging neglected disease of humans, particularly in areas of South America, Asia, Iran and Egypt [1]. Infection with this parasite is acquired by the ingestion of plants contaminated with metacercariae, a resistant cystic form that emerges as a newly excysted juvenile (NEJ) in the duodenum,

and after traversing the gut wall migrates to the liver. The parasites spend 8–12 weeks feeding on, and severely damaging, the liver parenchyma before they move into the bile ducts and become obligate blood-feeders by sucking blood through punctures in the duct walls. As in other parasites the invasion and establishment is mediated by a delicate crosstalk between molecules generated by the parasite and the host, with proteolytic enzymes being major players in this interaction [2]. Tissue migration and feeding is facilitated by the abundant secretion of proteolytic enzymes, most particularly cathepsin L cysteine proteases [3,4].

### Author Summary

The flatworm *Fasciola hepatica* is responsible for fasciolosis, one of the most common parasitic diseases of livestock worldwide, with increased incidence of human cases. When contaminated plants are ingested, infective larvae are released and transverse the gut wall before migrating to the bile ducts within the liver. Migrating liver flukes erode host tissue while adults feed on blood and they mature and release thousands of eggs. Several developmentally-regulated cathepsin L like proteolytic enzymes (FheCLs) are essential to the migrating and feeding processes. Despite being similar in structure and sequence these enzymes show specialization attacking preferentially different substrates and taking part in the diverse process of invasion, immune evasion and feeding. Our analyses reveal unique differences in activity between the major infective juvenile (FheCL3) and adult (FheCL1) enzymes, and demonstrate that the juvenile enzyme has a particular active site that allows it to degrade collagen, the main component of connective tissues. We demonstrate that a single position on the active site, residue 67, is essential to this collagenolytic activity critical for parasite invasion.

*F. hepatica* possesses an expanded multigene family of cathepsin L-like proteases that includes at least 5 different Clan CA (papain-like) members that are developmentally regulated and play pivotal roles in parasite survival by facilitating migration, immune evasion and feeding [5,6]. Transcriptomic and proteomic studies have demonstrated that the infective NEJ express and secrete cathepsin L3 (FheCL3) indicating that this is critical to enabling the parasite penetrate the intestinal wall [7,8,9,10]. By contrast, the blood-feeding adult expresses predominantly cathepsinL1 (FheCL1), to a lesser extent, cathepsin L2 (FheCL2) and to a relatively minor extent FheCL5. FheCL1 can be involved in parasite feeding, since *in vitro* experiments showed it can digest hemoglobin; both FheCL1 and FheCL2 have been implicated in immune evasion based in their *in vitro* ability to cleave native immunoglobulins [11]. Correlating with the macromolecular substrates the parasite encounters at these different locations, the cathepsin L members exhibit distinct substrate specificities [4,11].

For papain-like proteases, the evidence points to the S<sub>2</sub> subsite as being most critical to defining substrate selectivity [12]. We have shown that the juvenile FheCL3 is unusual in having a particular preference for Pro residues in the P<sub>2</sub> position of peptide substrates. By stark contrast, FheCL1 has a marked preference for aliphatic and aromatic residues in the P<sub>2</sub> substrate position and does not readily accept Pro. FheCL2, on the other hand, exhibits an substrate preference in between these two enzymes by preferring P<sub>2</sub> aliphatic and aromatic residues but also accepting Pro, although much less efficiently than FheCL3. Most interestingly, we have previously demonstrated that the preference for P<sub>2</sub> Pro confers FheCL3 and FheCL2 with the rare ability to cleave native collagen [13,14]. Only two other cysteine proteases, mammalian cathepsin K, which is involved in bone resorption by osteoclasts [15], and the ginger rhizome cysteine proteases (CP-II or zingipain, GP2 and GP3) also exhibit this high affinity for Pro in P<sub>2</sub> and collagenolytic activity [16,17].

Comparison of crystallographic structures of several Clan CA cysteine proteases allowed the identification of residues that make up the active site cleft with the selective S<sub>2</sub> pocket being delimited by residues 67, 133, 157, 158 and 205 (papain numbering) [18,19,20,21,22,23,24]. While variations occur in

several of these positions within the *F. hepatica* cathepsin L family the residue at position 67 has been primarily implicated in P<sub>2</sub> Pro accommodation by stabilizing interactions with the planar ring of Pro in the peptide substrate [20,25]. In FheCL3 and zingipain this position is occupied by the large aromatic residue Trp while in FheCL2 and cathepsin K a Tyr is present. Structural comparisons and molecular dynamic simulations performed by us suggested that the substrate selectivity observed in FheCL3 might be due to steric restrictions imposed by the bulky aromatic residues not only at the S<sub>2</sub> subsite but also within the S<sub>3</sub> pocket [13,14]. The remarkable convergence between FheCL3 and zingipain is not only restricted to Trp67 but also the close-by position 61 at the bottom of the S<sub>3</sub> pocket is occupied by a large His residue. This suggested to us that together these two active site moieties could influence the capacity of the enzymes to best accommodate Pro over other aliphatic residues, and hence account for their collagenolytic activity.

To get a clear picture of the substrate specificity of the major proteases of *F. hepatica*, we used a recently developed method involving multiplex substrate profiling and mass spectrometry (MSP-MS), that provides for unbiased subsite profiling of proteases across the entire active site [26]. In addition, the P<sub>1</sub>–P<sub>4</sub> subsite specificities were determined by Positional Scanning-Synthetic Combinatorial Libraries of fluorogenic tetrapeptides (PS-SCL), a well-established technology to study protease substrate specificity [20,27,28,29]. To test the relevance of active site positions 61 and 67 in selectivity we prepared recombinant variants of FheCL3 with the specific alterations in the S<sub>2</sub> and S<sub>3</sub> subsites, mimicking those present in FheCL1, and the reciprocal variants of FheCL1 in an attempt to confer this protease with collagenolytic activity. All the approaches highlight the unusual and marked preference of FheCL3 for P<sub>2</sub> Pro, and additionally reveal that the P<sub>3</sub> pocket has a less marked but distinctive preference for Gly. The mutational analysis emphasizes the dual role of residue 67 in modulating interactions with both P<sub>2</sub> and P<sub>3</sub> substrate residues and its crucial importance in juvenile FheCL3 specificity and activity. Our findings provide structural insights into the molecular determinants of active site preferences of two proteases that are vital for parasite development, which might in turn prove useful in the design of strategies to control parasite infection.

### Methods

#### Generation of the FheCL1 and FheCL3 active site pocket variants

Six FheCL3 and FheCL1 variants bearing substitutions at the S<sub>2</sub> and S<sub>3</sub> active site pockets were constructed by site-specific mutagenesis using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) as indicated in Table S1. Briefly, different pairs of complementary oligonucleotides containing the base pair substitutions to be introduced in the cathepsin gene sequences were generated and used in an outside PCR reaction employing as templates clones of FheCL1 or FheCL3 in the X4-Mfz-ScPas3 expression plasmid (kindly provided by Dr. R.J.S. Baerends and Dr. J.A.K.W. Kiel, Molecular Cell Biology Lab, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands). Double variants were obtained by using plasmids bearing the single mutations as templates. The amplified modified plasmids were propagated in bacteria, sequenced to confirm the presence of the desired mutations, and then electroporated in the *Hansenula polymorpha* yeast strain for production as previously described [30].



### Production of FheCL1, FheCL3 and the enzyme variants in yeast

FheCL1 and FheCL3 recombinant proenzymes were produced in the yeast *Hansenula polymorpha* as previously described [13,14]. Briefly, yeast transformants were cultured in 500 ml BMGY broth at 37°C to an OD<sub>600</sub> of 2–6, harvested by centrifugation at 2000 g for 10 min and induced by resuspending in 50 ml of buffered minimal media (0.67% yeast nitrogen base; 0.1M phosphate buffer pH 6.0; 1% methanol) for 36 hs at 30°C. Recombinant propeptidases were secreted to the culture media, and recovered by 20–30 fold concentration of culture supernatants by ultrafiltration with a 10 kDa cut-off membrane. The proenzymes were autocatalytically activated to the mature form by incubation for 2 h at 37°C in 0.1 M sodium citrate buffer (pH 5.0) with 2 mM DTT and 2.5 mM EDTA, dialyzed against PBS pH 7.3 and stored at –20°C. The protein concentration was assessed by the BCA method [31]. The proportion of functionally active recombinant enzyme was determined by titration against E-64c. The enzymes variants were obtained with the same protocol used for production of FheCL1 and FheCL3.

### Multiplex substrate profiling by mass spectrometry (MSP-MS)

The enzymatic activity of FheCL1 and FheCL3 were compared by MSP-MS, a procedure designed for unbiased profiling of protease activity [26]. A highly diversified peptide library consisting of 124 synthetic tetradecapeptides containing all possible amino acid pairs and near neighbor pairs, was used to test enzymatic activity. All peptides had unmodified termini and consist of natural amino acids except Met that was substituted by norleucine and Cys omitted because of potential disulfide bond formation. The library was distributed into three pools consisting of 52, 52 and 20 peptides and diluted to 1 μM in 25 mM sodium phosphate, pH 6.0, 1 mM DTT, 1 mM EDTA. An equal volume of FheCL1 or FheCL3 in the same buffer was added to the peptide pools such that the final concentration of each enzyme was 10 nM. An enzyme-free assay was set up as a control. Assays were incubated at room temperature and aliquots were removed after 5, 15, 60, 240 and 1200 minutes. All reactions were acid quenched to pH 3.0 or less with formic acid (4% final), evaporated to dryness and reconstituted to the original volume in 0.1% formic acid. Ten μl of each time point were injected onto a 150×0.3 mm Magic C18AQ column (Michrom Bioresources) connected to a Thermo Finnigan LTQ ion trap mass spectrometer equipped with a standard electrospray ionization source. Peak lists were generated from the raw files using PAVA software (UCSF) and searched against a database consisting of all 124 peptides using Protein Prospector. Newly formed cathepsinL1 or L3 cleavage products were identified by comparison with control assays.

### P<sub>1</sub>–P<sub>4</sub> specificity testing using a PS-SCL

The substrate specificities of FheCL1, FheCL3 and all the variants were determined using a PS-SCL [26]. Assays were performed in 0.1 M sodium phosphate buffer pH 6.0, 1 mM DTT, 1 mM EDTA, 0.01% PEG-6000 and 0.5% Me<sub>2</sub>SO (from the substrates) at 25°C. Aliquots of 12.5 nmol in 0.5 μl from each of the 20 sub-libraries of the P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub> and P<sub>4</sub> libraries were added to the wells of a 96-well Microfluor-1 flat-bottom plates. The final concentration of each compound of the 8,000 compounds per well was 15.62 nM in a 100 μl final reaction volume. The assays were performed in triplicate, the reaction was initiated by addition of the enzyme diluted in the above buffer and monitored with a SpectraMax Gemini fluorescence

spectrometer (Molecular Devices) with excitation at 380 nm, emission at 460 nm and cutoff at 435 nm.

### Enzymatic assays using synthetic fluorogenic peptides

Kinetic parameters were determined in a reaction buffer containing 0.1 M sodium phosphate buffer, pH 6.0, 1 mM DTT and 1 mM EDTA at 25°C; typically final enzyme concentrations were in the 10<sup>–9</sup>M range, and the substrate was added after 10 min. of incubation of enzyme in reaction buffer. Enzyme concentration was determined by active-site titration with E-64c. Enzyme activity was monitored by the hydrolysis of 7-amino-4-methyl coumarin (AMC) from the synthetic peptide substrates Z-VLK-AMC and Tos-GPR-AMC. Reaction rates with different substrate concentrations (5–100 μM) were measured in duplicate as the slope of the progress curves obtained by continuous recording in a FluoStar spectrofluorimeter at 345 excitation and 440 emission wavelengths, using an AMC standard curve for product concentration calculation. Kinetic constants,  $k_{cat}$  and  $K_M$ , were estimated by non-linear regression analysis of the Michaelis-Menten plot using the OriginPro 6.1 software.

### Digestion of type I collagen

Protein digestion was analyzed by incubating 10 μg of type I collagen from rat tail (Sigma) with 5 μM enzyme in PBS pH 7.3, 1 mM DTT and 1 mM EDTA for different times at 28°C. Digestion reactions were stopped by adding 10 μM of E64c to the reaction mixture. Fragments were separated by SDS-PAGE gels (8% acrylamide) under reducing conditions and stained with Coomassie Brilliant BlueR-250.

### Homology modeling

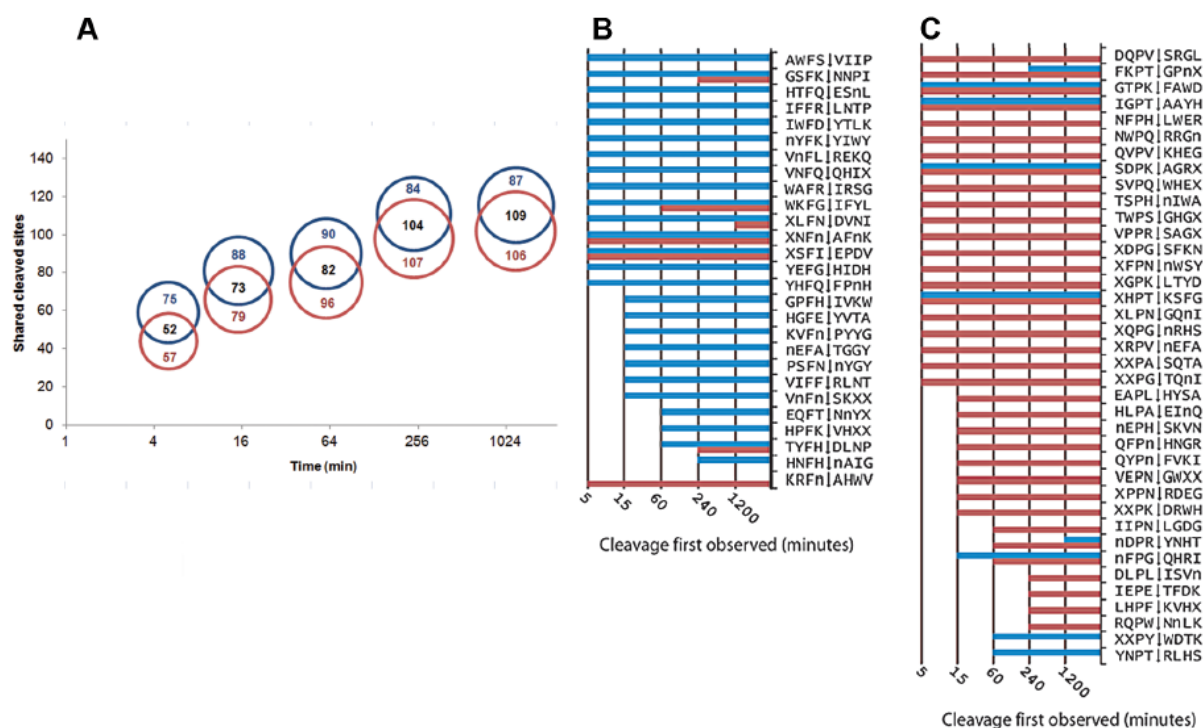
Homology models of FheCL3 were generated with SwissModel [32] using as principal template the crystal structure of FheCL1 (206×). Template and models were superimposed for visualization with Swiss PDBViewer version 4.1. (<http://www.expasy.org/spdbv/>) [33] Active site residues were identified based on the literature and confirmed by structural alignment with human cathepsin L (1MHW), human cathepsin K (1ATK) and papain (5PAD). The FheCL3 rotamers and the W67L mutant were generated with the mutate function in the PDBViewer, and selected based on rotamer score and visual inspection.

## Results/Discussion

### FheCL1 and FheCL3 multiplex substrate profiling by mass spectrometry (MSP-MS)

MSP-MS is a novel method designed to profile protease activity, based on the cleavage of a library of 124 tetradecapetides, providing theoretically unbiased information on preferences at both sides of the cleavage point [26]. The extended nature of the tetradecapeptides allow a much more natural interaction across the protease active site providing a detailed picture of the contribution of the S and S' subsites that accommodate the substrate. The characteristics of the S' sites are generally poorly known mainly because most substrates used for enzymatic profiling place a fluorophore or chromophore in the P<sub>1</sub>' position, a moiety very unlike any amino acid that the enzyme can normally accept in that pocket.

The FheCL1 or FheCL3 were added separately to the library and all the cleaved peptides were identified at time intervals by mass spectrometry. While both enzymes cleaved at more than 170 sites after one hour incubation, FheCL1 had produced approximately 75% within five minutes of the reaction and >95% by 15 minutes. Notably, compared to FheCL1, FheCL3 produced relatively fewer



**Figure 1. Differential cleavage of FheCL1 and FheCL3 using a multiplex combinatorial library (MSP-MS).** (A). Proportional Venn diagrams showing the number of shared and exclusive cleavage sites for FheCL1 (blue) and FheCL3 (red) obtained by MSP-MS at different incubation times (5, 15, 60, 120 and 2400 min, (time axis scale is log<sub>2</sub>)). (B) A list of all substrates containing Phe in the S<sub>2</sub> position cleaved by either FheCL1 (blue) and/or FheCL3 (red) The time that cleavage was first observed is illustrated by the corresponding bar charts (C) List of all substrates cleavage points containing Pro in the S<sub>2</sub> position cleaved by either FheCL1 (blue) and/or FheCL3 (red) at different times as indicated. doi:10.1371/journal.pntd.0002269.g001

cleavages at early time-points, while minor cleavages still occur for up to 20 hours of reaction, indicating differences in the ability to accommodate substrates (Figure 1 A and Figure S1). Significantly, only approximately half of the cleavage sites identified at any time were cleaved by both enzymes, leaving many that were exclusive for either FhCL1 or FhCL3 (Figure 1 B, C). A good example of this differential cleavage is offered by Peptide#38 where FheCL1 cleaves once between T<sup>Δ</sup>F (EAWMT<sup>Δ</sup>FIVPPRSAG) but FheCL3 cleaves twice between W<sup>Δ</sup>M and R<sup>Δ</sup>S (EAW<sup>Δ</sup>MTFIVPPR<sup>Δ</sup>SAG) and never cleaves between T<sup>Δ</sup>F even after 1200 minutes incubation (data not shown).

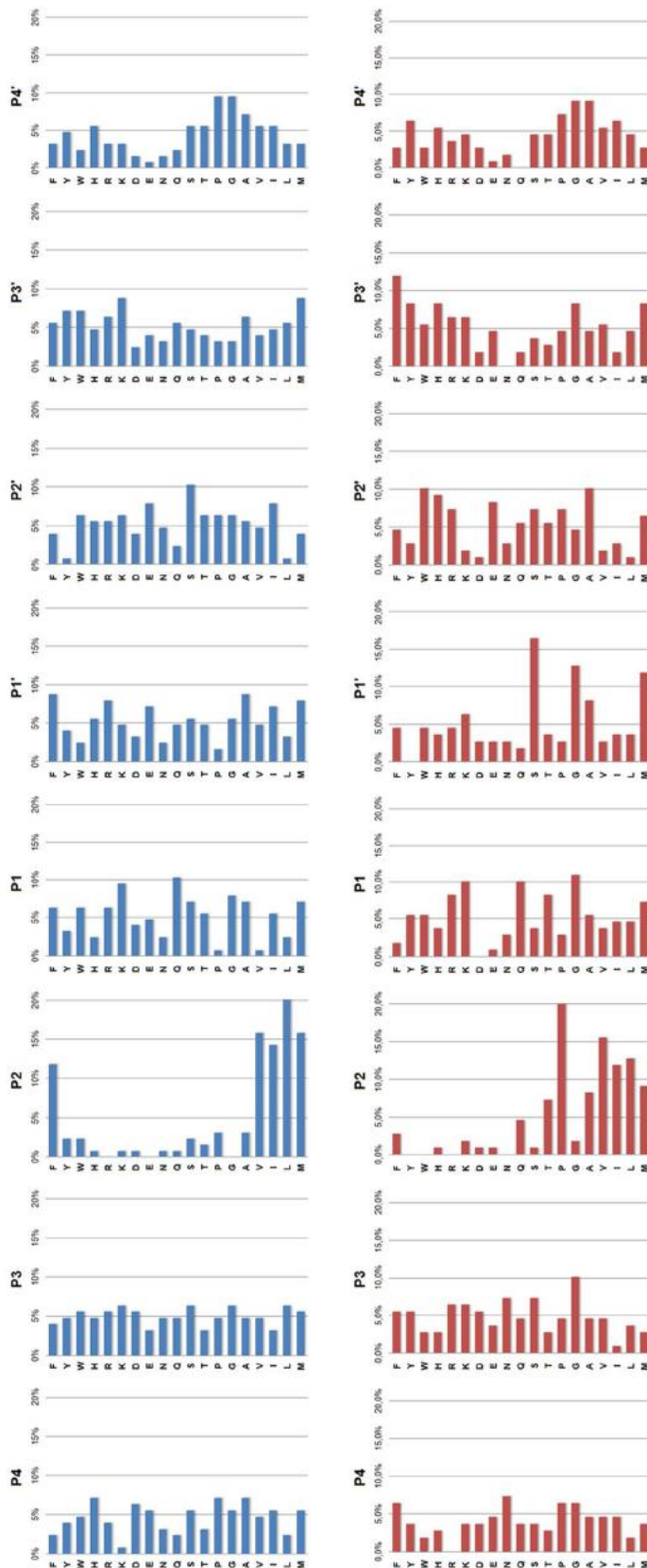
The positional analysis indicate that the substrate specificity in both FheCL1 and FheCL3 is dominated by the amino acid at P<sub>2</sub> consistent to what is known about clan CA cysteine proteases [12] (Figure 2). The substrate signature at this position showed that besides aliphatic residues that can be accommodated by both enzymes, FheCL1 can readily accept Phe at P<sub>2</sub> but has very low tolerance for Pro, while FheCL3 is the opposite (Figure 1B–C). In fact the preferred amino acids at this position are Leu and Pro for FheCL1 and FheCL3, respectively, confirming our previous studies using short fluorogenic peptides [13,14]. The profile also shows that both enzymes share a strong selection against charged P<sub>2</sub> residues (Figure 2). Also on the non-prime side, the juvenile enzyme, has a slight preference for Gly in the S<sub>3</sub> pocket (Figure 2). This S<sub>3</sub> preference is more noticeable at early digestion times (5 min. reaction), while other residues can be progressively accommodated at this site as the length of the incubation increases (Figure S1).

On the prime side of FheCL3, substrate preference is dominated by the P<sub>1</sub>' site and shows a preference for Ser, Gly and to a lesser extent Met (norleucine) and Ala (Figure 2). Previous reports using internally quenched penta or heptapeptide substrates investigated the prime side preferences for papain and mammalian cathepsins B, L, S and K and showed that a broad range of amino acids were accommodated in these subsites [34,35,36]. However, while subtle differences were noted between the enzymes none of them can be considered as major contributions to specificity, except for a slight preference of hydrophobic moieties in papain P<sub>3</sub>' [34], and a general avoidance of Pro at P<sub>1</sub>' [36]. Our data confirmed the avoidance of Pro, and highlights the preference of FheCL3 for Gly and Ser, a feature that might be relevant for the enzyme's ability to degrade collagen helical domains.

#### FheCL1 and FheCL3 active site preferences based on PS-SCL

Whereas the MSP-MS assay offers a more "natural" way of determining substrate specificity because the longer linear peptides are more like the loop regions found in protein substrates, Positional Scanning- Synthetic Combinatorial Libraries (PS-SCL) offer increased diversity for the study of P<sub>4</sub> to P<sub>1</sub> interactions since they comprise a collection of all possible fluorogenic tetrapeptides. This methodology has been widely used in the characterization of cysteine proteases [27,28], and profiles of adult liver fluke proteases are known [20,29]. The PS-SCL profile for the recombinant FheCL1 used in this work is practically identical with that reported by Stack et al. [20], independently supporting the accuracy of this tool in assessing enzyme specificity (Fig. S1).

Active Site of *Fasciola* Collagenolytic Cathepsin L



**Figure 2. Profiling of the P<sub>4</sub>-P<sub>4</sub>' substrate specificity of FheCL1 and FheCL3 using a multiplex combinatorial library (MSP-MS).** Frequency of amino acids found at positions P<sub>4</sub>-P<sub>4</sub>' of cleavage sites after 5 min incubation with FheCL1 (top) or FheCL3 (bottom). Results are expressed as percentage per site. The amino acid frequency at each position within the tetradecapeptide library ranges from 4.2% to 6.8%. Met is substituted by norleucine in the library. doi:10.1371/journal.pntd.0002269.g002

Furthermore, despite the differences in the methodological approaches, the PS-SCL results are generally consistent with the MSP-MS observations.

FheCL1 displays a typical papain-like cysteine protease profile with S<sub>2</sub> predominance, *i.e.* marked preferences for aliphatic residues, particularly Leu, at this position. Some minor selectivity can be found for the S<sub>1</sub> interactions, where the basic residues Arg and Lys, together with Gln, Thr and Met are preferred. In contrast, the S<sub>3</sub> and S<sub>4</sub> pockets show a broad specificity completing a picture similar to that found by the MSP-MS analysis (Figure 3).

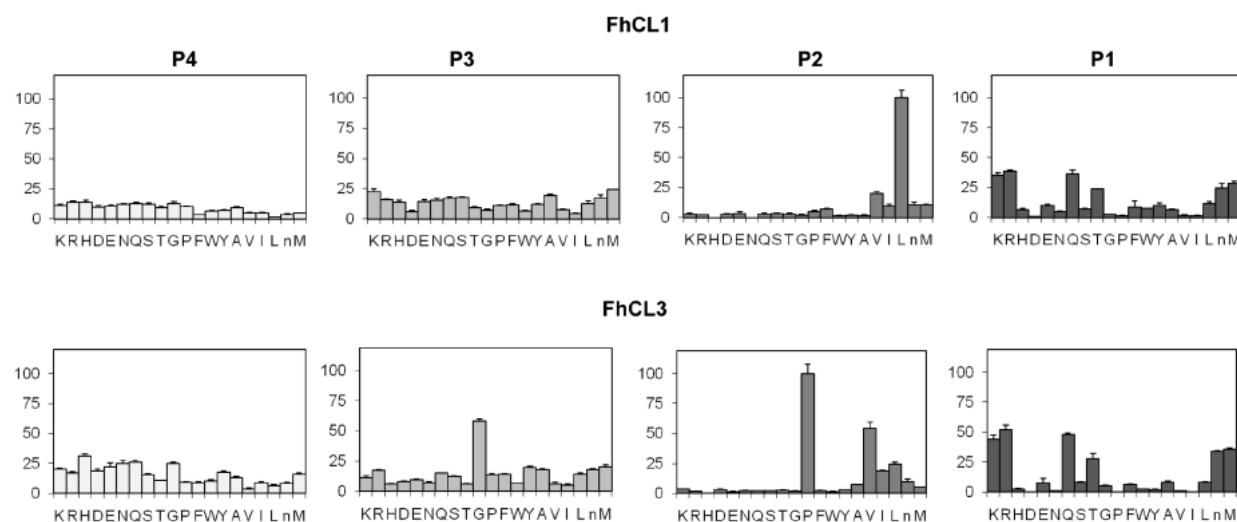
The most obvious difference between FhCL1 and FheCL3 are the very distinct profiles observed for the P<sub>2</sub> and P<sub>3</sub> residues. The FheCL3 S<sub>2</sub> pocket can accommodate Pro very readily, accepting it twice better than Val and four times better than Leu. In addition, unlike most known cysteine proteinases, the S<sub>3</sub> pocket of FheCL3 demonstrates selectivity, specifically for Gly (Figure 3). Consistent with the PS-SCL data the MSP-MS results at 5 min of digestion shows FheCL3 has a preference for Gly in P<sub>3</sub> (Figure 2), and as the reaction proceed other amino acids are also accommodated in S<sub>3</sub> as indicated by an increased frequency in later times. This effect is expected since in the MSP-MS assay all peptides are mixed and assayed together, consequently the preferential cleavages would be observed early in the reaction. Selectivity at S<sub>3</sub> is relatively rare, although PS-SCL studies have shown that the plant enzymes papain and bromelain have a noticeable preference for Pro at P<sub>3</sub> [27].

Our previous studies showed that FheCL2 also has a slightly increased preference for Gly at P<sub>3</sub>, and an augmented acceptance of Pro at P<sub>2</sub> although maintaining Leu as the preferred residue in this position [20]. Therefore, FheCL2 active site appears to have intermediate characteristics between FhCL1 and FheCL3, both at S<sub>2</sub> and S<sub>3</sub> subsites (compare Figure 3 with that of Stack et al. [20] <http://www.jbc.org/content/283/15/9896.full.pdf+html>).

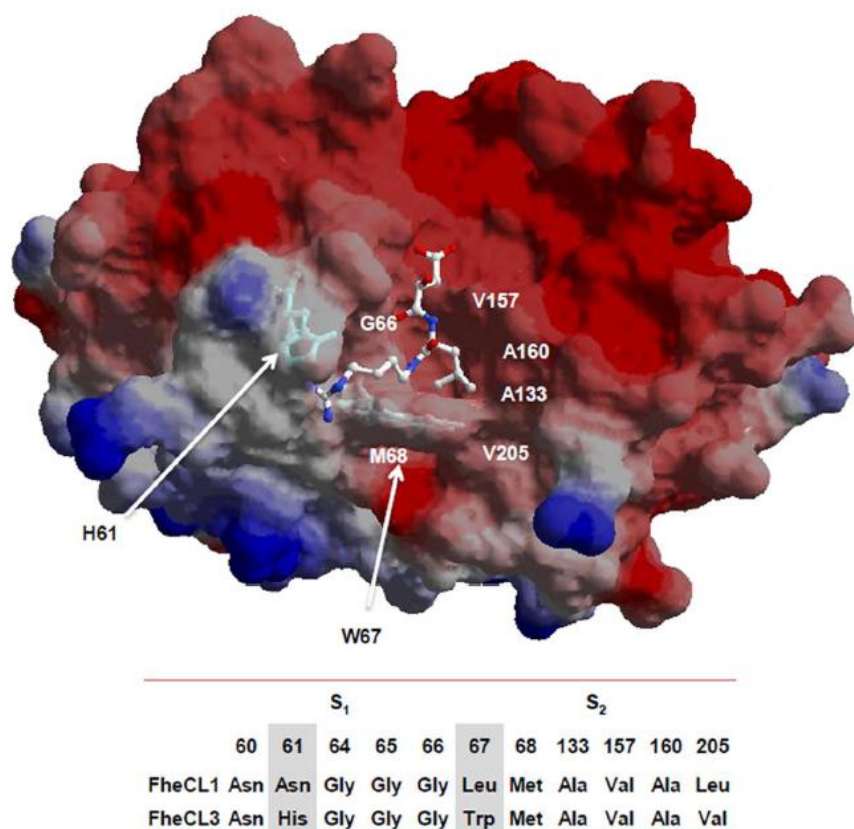
The PS-SCL profile of FheCL5, an enzyme secreted in very low abundance by adult *F. hepatica*, has also been reported and is more similar to FheCL1 with strong Leu preference at P<sub>2</sub>, although it has the unique ability to accept Asp [29]. These results support the idea of functional divergence and specialization of the different members of the liver fluke cathepsin L family occurred following several gene duplications as proposed by phylogenetic analysis [5,6].

### FheCL3 variants in S<sub>2</sub> and S<sub>3</sub> active site pockets

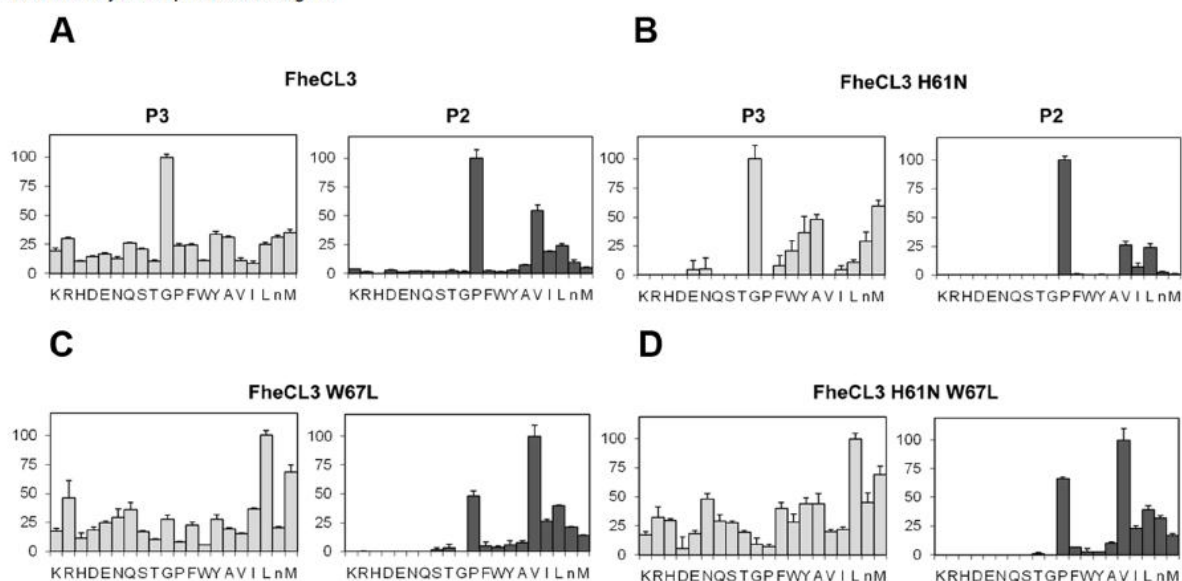
Since non-prime side differences in specificity between *Fasciola* cathepsin Ls are mainly restricted to S<sub>2</sub> and S<sub>3</sub>, we investigated the contribution of the variable residues lining those sites by mutation analysis. These pockets differ only at three positions: 61, 67 and 205 located at the bottom of the S<sub>3</sub> pocket, at the hinge of subsites S<sub>2</sub> and S<sub>3</sub>, and at the bottom of the S<sub>2</sub> subsite, respectively (Figure 4). The first two variations involve amino acids with different properties, while the third involves a substitution between similar aliphatic moieties. Based on these observations, we changed residues 61 and 67 of FheCL3, for those present in FheCL1, generating the variants FheCL3 H61N, FheCL3 W67L and a double mutant bearing both substitutions. Their preferences at P<sub>2</sub> and P<sub>3</sub> were assessed with the PS-SCL approach. FheCL3 H61N showed only a subtle change in enzyme specificity, decreasing Gly preference in relation to the other amino acids as has been predicted (Figure 5 B). The FheCL3 W67L variant resulted in a marked reduction in the preference for Pro at P<sub>2</sub> compared to FheCL3, while simultaneously increasing the aliphatic residues preferences and making Val the most favorably accommodated residue. Importantly, we found that the FheCL3 W67L variant also altered P<sub>3</sub> specificities, changing the Gly preference to an increased preference for Leu (Figure 5 C). The double mutant enzyme, FheCL3 H61N/W67L, presented S<sub>2</sub> and



**Figure 3. Profiling of the P<sub>1</sub>-P<sub>4</sub> substrate specificity of liver fluke cathepsin Ls using PS-SCL libraries.** The activity against the substrates is represented relative to the highest activity of the library with each enzyme (namely Leu and Pro at P<sub>2</sub> for FheCL1 and FheCL3, respectively, n= norleucine). The error bars display the standard deviation from triplicate experiments. doi:10.1371/journal.pntd.0002269.g003



**Figure 4. Residues contributing to substrate binding.** Model of mature FheCL3 showing the S<sub>2</sub> and S<sub>3</sub> subsites of the active site: the His61 and Trp67 residues that were mutated in the present study are highlighted as sticks. The E64 inhibitor complexed with human cathepsin K (1ATK) was superimposed to facilitate viewing the active site cleft. The image was generated with SPDBviewer [33]. The residues constituting the active site S<sub>2</sub> and S<sub>3</sub> subsites of FheCL1 and FheCL3 are indicated below the molecular representation.  
doi:10.1371/journal.pntd.0002269.g004



**Figure 5. PS-SCL Profiling of the P<sub>2</sub> and P<sub>3</sub> substrate specificity of FheCL3 to FheCL1 enzyme variants.** The activity against the substrates is represented in the y axis relative to the highest activity of the library, whereas the x axis shows the different amino acids using the one-letter code (n=norleucine). Error bars display the standard deviation from triplicate experiments. (A) FheCL3, (B) FheCL3 H61N (C) FheCL3 W67L (D) FheCL3 H61N W67L.  
doi:10.1371/journal.pntd.0002269.g005

**Table 1.** Kinetic parameters of FheCL1, FheCL3 and the modified enzymes over two different substrates.

Enzyme	Z-VLK-AMC			Tos-GPR-AMC			VLK/GPR $k_{cat}/K_M$ ratio
	$k_{cat}$ ( $s^{-1}$ )	$K_M$ ( $\mu M$ )	$k_{cat}/K_M$ ( $M s^{-1}$ )	$k_{cat}$ ( $s^{-1}$ )	$K_M$ ( $\mu M$ )	$k_{cat}/K_M$ ( $M s^{-1}$ )	
FheCL1	2.32±0.120	3.94±0.93	588889	0.0350±0.0060	15.10±6.70	2355	250:1
FheCL1 N61H	5.44±0.280	8.94±1.39	608391	0.0380±0.0040	11.76±4.30	3227	188:1
FheCL1 L67W	0.78±0.060	10.58±2.40	73378	0.0340±0.0050	10.90±4.60	3148	23:1
FheCL1 N61H L67W	1.30±0.130	12.01±2.90	108243	0.0310±0.0001	3.20±0.08	9615	11:1
FheCL3 W67L	0.03±0.003	44.80±10.60	590	0.0053±0.0005	5.20±1.80	1017	1:2
FheCL3 H61N W67L	0.03±0.004	28.70±8.60	1037	0.0072±0.0005	6.70±1.90	1072	1:1
FheCL3*	0.33±0.010	15.70±1.80	20760	6.0000±0.4800	4.10±0.90	1464300	1:70

\*FheCL3 kinetic parameters were taken from Corvo *et al*, 2009 [14].  
doi:10.1371/journal.pntd.0002269.t001

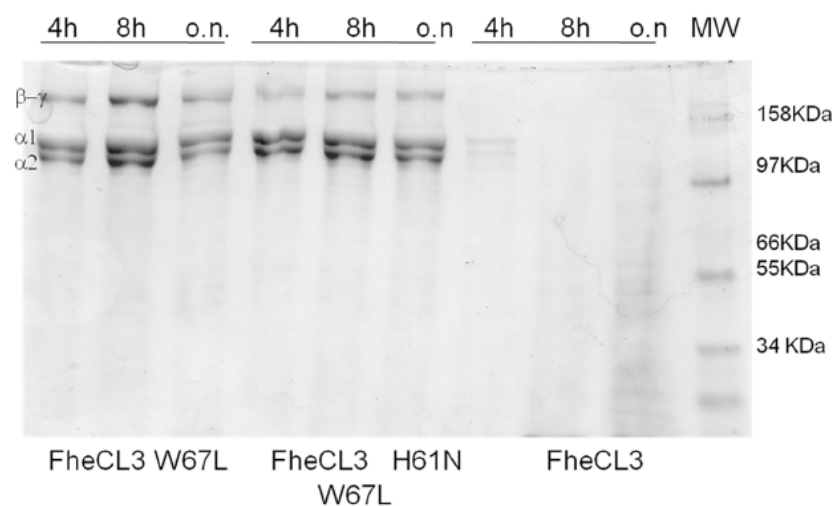
Consequently, despite finding the expected variations in Z-VLK-AMC and Tos-GPR-AMC activity in FhCL1mutants, these changes are not enough to absorb the more than 200-fold difference in specificity that FheCL1 has for these two types of substrates and the enzymes still prefer substrates with P<sub>2</sub> Leu (Table 1). Previously, Stack *et al*. [20] found that the L67Y change in FheCL1 did not significantly modify the activity towards Tos-GPR-AMC which is consistent with our studies. However, a 13-fold increase on the activity towards this substrate was observed when a similar L67Y change was introduced into FheCL5 [38]. FheCL5 active site is more restricted at both the S<sub>2</sub> and S<sub>3</sub> pockets than FheCL1 due to the presence of the bulkier Leu157 and Tyr61 residues respectively. The L67Y change would impose a further restriction in the active site such that small residues at P<sub>3</sub> and P<sub>2</sub> would be favored. Consequently, the improved acceptance of Tos-GPR-AMC could be explained by the presence of the adjacent Gly and Pro positioned at P<sub>3</sub> and P<sub>2</sub> respectively, rather than by the modest rise in activity towards Pro at P<sub>2</sub> as originally proposed [38]. The same rationale explains the recent observation that a FheCL5 L67F mutation increased

activity towards Tos-GPR-AMC, and the inverse FheCL2 Y69L variant reduced P<sub>2</sub> Pro acceptance [29].

#### Functional collagenolytic assay of FhCL3 mutants

Given the unusual characteristic of FheCL3 to efficiently degrade native type I collagen, we assessed the efficacy of the parent FheCL3 and its variants to hydrolyze type I collagen *in vitro*. Unlike wild type FheCL3, both FheCL3W67L and FheCL3 H63N/W67L variants were unable to cleave collagen at neutral pH and 28°C, conditions that preserve its native structure (Figure 7). The reduced activity of FheCL3 mutants indicate that Trp67 might be crucial for the enzyme activity that might be centered in cleaving substrates enriched in small amino acids (Gly, Pro) like collagen.

Our findings agree with previous observations that the substitutions Y67L and L205A in human cathepsin K (for residues present in human cathepsin L), abolish its collagenolytic activity [37]. This human cathepsin K variant acquires the S<sub>2</sub> preferences of human cathepsin L, and the reciprocal replacements to human cathepsin L conferred it with a specificity similar to cathepsin K



**Figure 7. Collagen cleaving activities of recombinant *F. hepatica* FheCL3 and the FheCL3 enzyme variants W67L and H61N/W67L.** Type I collagen was incubated with 5  $\mu M$  of each enzyme for different lengths of time in PBS pH 7.3 at 28°C and separated by SDS-PAGE.  
doi:10.1371/journal.pntd.0002269.g007

[25]. We have also prepared a double variant of FheCL1 at the same positions, i.e. FheCL1 L67Y/L205A but this did not exhibit collagenolytic activity (data not shown). This lack of correlation with human cathepsin L and K mutants behavior is surprising, although differences at other positions within the active sites exist between the mammalian and fluke enzymes that must also be important in determining collagenolytic ability. These differences in turn might prove useful in the design of specific inhibitors or drugs for the parasite enzymes over host homologues.

#### Homology modeling of FhCL3 active site

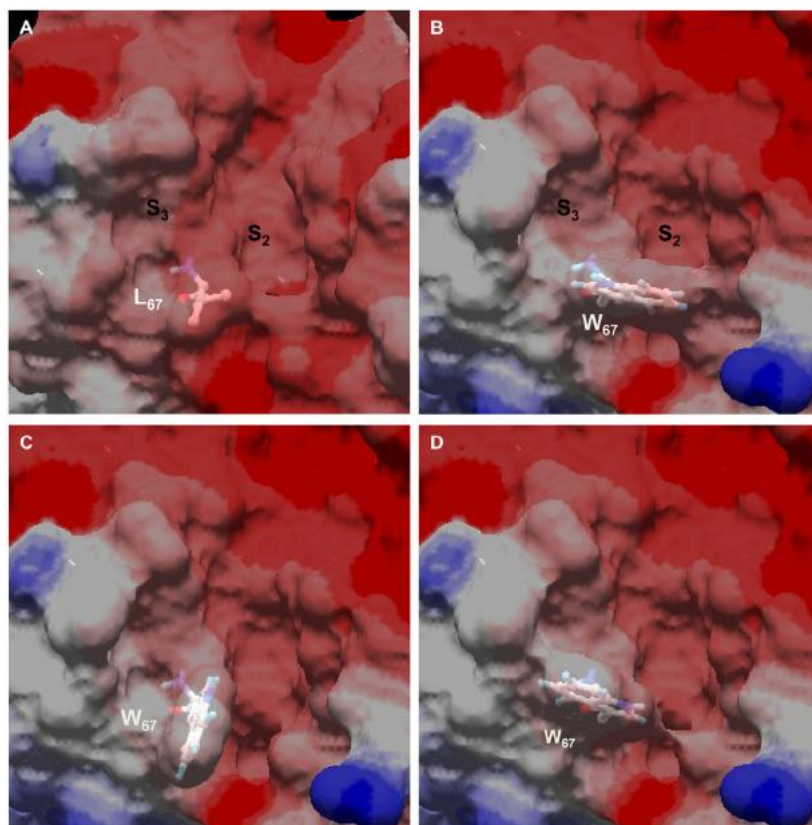
Our analysis of active site variants highlights the role of residue 67 which is determining by its gate-keeper position not only the conformation of the  $S_2$  subsite, but also of the  $S_3$  pocket. Using molecular modeling we analyzed the possible conformations of Trp67 in the active site of FheCL3 as compared to FheCL1 (Figure 8). The most stable rotamer protrudes and partially occludes the  $S_2$  subsite (Figure 8 B). An alternative conformer places the indole ring towards the  $S_3$  subsite reducing this site volume (Figure 8 C), while a third low energy rotamer is coaxial with the active site cleft leaving two more open but narrow active site pockets (Figure 8 D). The rotation of this residue might be fundamental to accommodating the distinct substrates of FheCL3, defining the nature of the amino acids that can be accepted in these subsites. The planar ring of Pro occupying the  $P_2$  subsite can be stabilized by stacking interactions with the aromatic heterocycle of Trp. Furthermore, aliphatic moieties can also be accommodated at this site due to the hydrophobic nature of FheCL3  $S_2$  pocket.

However, at the same time than stabilizing some interactions the bulky Trp can be imposing steric hindrances in the neighbor subsite thus favoring small residues.

Based on this observation we reanalyzed the MSP-MS data looking at the amino acid pairs present at  $S_3$ - $S_2$ . We noticed that FhCL3 can accommodate different residues at  $P_3$  if  $P_2$  is occupied by Pro, and that tiny Gly is slightly preferred at early times combined either with Pro or aliphatic moieties. In fact if small residues are present in  $P_3$ , other residues can be placed in  $P_2$  excepting aromatic ones, which are disfavored in any combination by FheCL3 (data not shown). These combined preferences for Pro and to a lesser extent for Gly residues by FheCL3, can explain why native collagen, that is rich in these amino acids is an appropriate substrate for this enzyme.

#### Conclusions

We have characterized the FheCL3 cysteine protease of the infective larval stage of *F. hepatica* that exhibits a particular collagenolytic activity and analyzed the differential contribution of active site residues involved. Our results highlight that a Trp residue strategically located at the gatekeeper position between the  $S_2$  and  $S_3$  active site pockets is vital to this activity and contributes to narrow and constrained pockets that can best accommodate small residues, particularly, Pro at  $P_2$  and Gly at  $P_3$ . These peculiarities are not shared by other known cysteine proteases, suggesting that the enzyme may be a good target for the development of small molecule inhibitors for parasite control. Furthermore, our mutation analyses reveal the under-appreciated



**Figure 8. Different rotamers of Trp67 have different effects on active site pocket sizes.** Comparison of FheCL1 (A) and FheCL3 active site pockets showing the three favoured rotamers (B, C and D) of Trp67 in FheCL3. The active site pockets  $S_2$  and  $S_3$  are indicated. doi:10.1371/journal.pntd.0002269.g008

significance of interactions at P<sub>3</sub> that together with those at P<sub>2</sub> contribute to modulating cysteine protease specificity. Novel extended peptide libraries provide first glimpses of other interactions particularly at the prime side of the active site cleft, showing noticeable differences whose contributions to specificity and selectivity need to be assessed in future studies.

## Supporting Information

**Figure S1 Time course of peptide degradation by FheCL1 and FheCL3 analyzed by MSP-MS.** Top panel: total amount of cuts obtained by the different enzymes at 5 min, 15 min, 1 h, 4 h and 20 h incubation. Bottom panels: amino acids found at positions P<sub>3</sub> and P<sub>2</sub> of the cleaved peptides at different times of incubation with FheCL1 (top, blue) or FheCL3 (bottom, red) at 5, 15, and 60 min. Results are expressed as percentage per site. The amino acid frequency at each position within the tetradecapeptide library ranges from 4.2% to 6.8%. Met is substituted by norleucine in the library. (TIF)

## References

- Keiser J, Utzinger J (2009) Food-borne trematodiasis. *Clin Microbiol Rev* 22: 466–483.
- McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M (2006) Proteases in parasitic diseases. *Annu Rev Pathol* 1: 497–536.
- Dalton J, Caffrey C, Sajid M, Stack C, Donnelly S, et al. (2006) Proteases in trematode biology. In: Maule AG, Marks NJ, editors. *Parasitic Flatworms: molecular biology, biochemistry, immunology and physiology: CAB international*.
- Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP (1999) Proteinases and associated genes of parasitic helminths. *Adv Parasitol* 43: 161–266.
- Irving JA, Spithill TW, Pike RN, Whisstock JC, Smooker PM (2003) The evolution of enzyme specificity in *Fasciola* spp. *J Mol Evol* 57: 1–15.
- Robinson MW, Tort JF, Lowther J, Donnelly SM, Wong E, et al. (2008) Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Mol Cell Proteomics* 7: 1111–1123.
- Cancela M, Acosta D, Rinaldi G, Silva E, Duran R, et al. (2008) A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie* 90: 1461–1475.
- Cancela M, Ruetalo N, Dell'Oca N, da Silva E, Smircich P, et al. (2010) Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*. *BMC Genomics* 11: 227.
- Morphew RM, Wright HA, Lacourse EJ, Porter J, Barrett J, et al. (2011) Towards delineating functions within the fasciola secreted cathepsin L protease family by integrating in vivo based sub-proteomics and phylogenetics. *PLoS Negl Trop Dis* 5: e937.
- Robinson MW, Menon R, Donnelly SM, Dalton JP, Ranganathan S (2009) An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. *Mol Cell Proteomics* 8: 1891–1907.
- Robinson MW, Dalton JP, Donnelly S (2008) Helminth pathogen cathepsin proteases: it's a family affair. *Trends Biochem Sci* 33: 601–608.
- Turk D, Guncar G, Podobnik M, Turk B (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol Chem* 379: 137–147.
- Robinson MW, Corvo I, Jones PM, George AM, Padula MP, et al. (2011) Collagenolytic activities of the major secreted cathepsin L peptidases involved in the virulence of the helminth pathogen, *Fasciola hepatica*. *PLoS Negl Trop Dis* 5: e1012.
- Corvo I, Cancela M, Cappetta M, Pi-Denis N, Tort JF, et al. (2009) The major cathepsin L secreted by the invasive juvenile *Fasciola hepatica* prefers proline in the S2 subsite and can cleave collagen. *Mol Biochem Parasitol* 167: 41–47.
- Lecaille F, Bromme D, Lalmanach G (2008) Biochemical properties and regulation of cathepsin K activity. *Biochimie* 90: 208–226.
- Choi KH, Laursen RA, Allen KN (1999) The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*. *Biochemistry* 38: 11624–11633.
- Kim M, Hamilton SE, Guddat LW, Overall CM (2007) Plant collagenase: unique collagenolytic activity of cysteine proteases from ginger. *Biochim Biophys Acta* 1770: 1627–1635.
- Fujishima A, Imai Y, Nomura T, Fujisawa Y, Yamamoto Y, et al. (1997) The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett* 407: 47–50.
- Pauly TA, Sulea T, Ammirati M, Sivaraman J, Danley DE, et al. (2003) Specificity determinants of human cathepsin s revealed by crystal structures of complexes. *Biochemistry* 42: 3203–3213.
- Stack CM, Caffrey CR, Donnelly SM, Seshadri A, Lowther J, et al. (2008) Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. *J Biol Chem* 283: 9896–9908.
- Zhao B, Janson CA, Amegadzie BY, D'Alessio K, Griffin C, et al. (1997) Crystal structure of human osteoclast cathepsin K complex with E-64. *Nat Struct Biol* 4: 109–111.
- Kim MJ, Yamamoto D, Matsumoto K, Inoue M, Ishida T, et al. (1992) Crystal structure of papain-E64-c complex. Binding diversity of E64-c to papain S2 and S3 subsites. *Biochem J* 287 (Pt 3): 797–803.
- O'Hara BP, Hemmings AM, Buttle DJ, Pearl LH (1995) Crystal structure of glycyI endopeptidase from *Carica papaya*: a cysteine endopeptidase of unusual substrate specificity. *Biochemistry* 34: 13190–13195.
- Schroder E, Phillips C, Garman E, Harlos K, Crawford C (1993) X-ray crystallographic structure of a papain-leupeptin complex. *FEBS Lett* 315: 38–42.
- Lecaille F, Chowdhury S, Purisima E, Bromme D, Lalmanach G (2007) The S2 subsites of cathepsins K and L and their contribution to collagen degradation. *Protein Sci* 16: 662–670.
- O'Donoghue AJ, Eroy-Reveles AA, Knudsen GM, Ingram J, Zhou M, et al. (2012) Global identification of peptidase specificity by multiplex substrate profiling. *Nat Methods* 2012: 1095–1100.
- Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogoy M, et al. (2006) Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J Biol Chem* 281: 12824–12832.
- Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA, et al. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc Natl Acad Sci U S A* 97: 7754–7759.
- Norbury IJ, Hung A, Beckham S, Pike RN, Spithill TW, et al. (2012) Analysis of *Fasciola* cathepsin L5 by S2 subsite substitutions and determination of the P1–P4 specificity reveals an unusual preference. *Biochimie* 94: 1119–1127.
- Faber KN, Haima P, Harder W, Veenhuis M, Ab G (1994) Highly-efficient electrotransformation of the yeast *Hansenula polymorpha*. *Curr Genet* 25: 305–310.
- Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, et al. (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* 150: 76–85.
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195–201.
- Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714–2723.
- Portaro FC, Santos AB, Cezari MH, Juliano MA, Juliano L, et al. (2000) Probing the specificity of cysteine proteinases at subsites remote from the active site: analysis of P4, P3, P2' and P3' variations in extended substrates. *Biochem J* 347 Pt 1: 123–129.
- Menard R, Carmona E, Plouffe C, Bromme D, Konishi Y, et al. (1993) The specificity of the S1' subsite of cysteine proteases. *FEBS Lett* 328: 107–110.
- Alves MF, Puzer L, Cotrin SS, Juliano MA, Juliano L, et al. (2003) S3 to S3' subsite specificity of recombinant human cathepsin K and development of selective internally quenched fluorescent substrates. *Biochem J* 373: 981–986.
- Lecaille F, Choe Y, Brandt W, Li Z, Craik CS, et al. (2002) Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity. *Biochemistry* 41: 8447–8454.
- Smooker PM, Whisstock JC, Irving JA, Siyaguna S, Spithill TW, et al. (2000) A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*. *Protein Sci* 9: 2567–2572.

**Table S1 Generation of active site mutants by site directed mutagenesis.** Oligonucleotide primers and templates used are indicated. (DOCX)

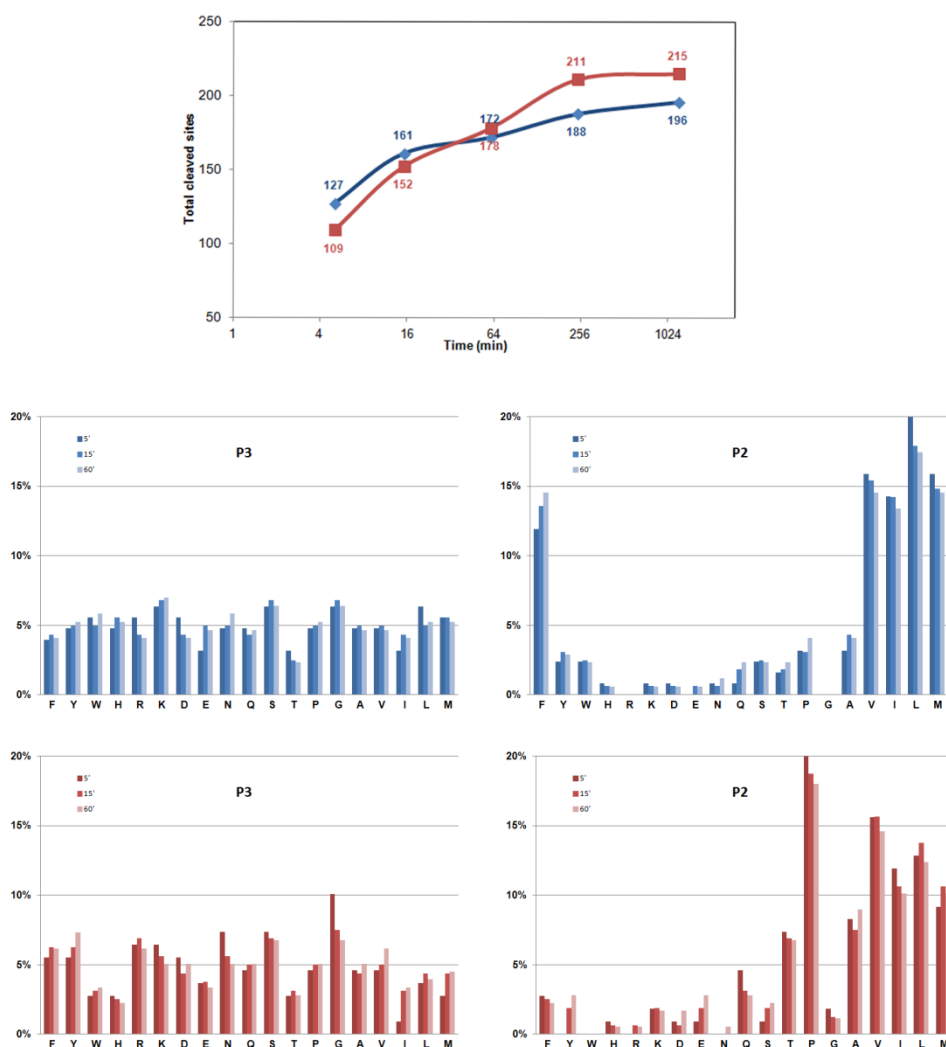
## Acknowledgments

We would like to thank Dr. R.J.S. Baerends and Dr. J.A.K.W. Kiel, Molecular Cell Biology Lab, Groningen Biomolecular Sciences and Biotechnology Institute for kindly providing the yeast expression system. We are grateful to Dr. Alberto Rascon, MSc Tatiana Basika and Brian Suzuki for their valuable help with experiments and Santiago Fontela for his kind assistance with the molecular modeling simulations. We also want to thank Dr. Gabriel Rinaldi for helpful discussion and comments to the manuscript.

## Author Contributions

Conceived and designed the experiments: IC LR CRC JFT. Performed the experiments: IC AJOD LP NPD. Analyzed the data: IC AJOD LP CRC JPD JFT. Contributed reagents/materials/analysis tools: AER JHM CSC. Wrote the paper: IC AJOD CRC JPD JFT.





**Figure S1.**

**Time course of peptide degradation by FheCL1 and FheCL3 analyzed by MSP-MS.** Top panel: total amount of cuts obtained by the different enzymes at 5 min, 15 min, 1 h, 4 h and 20 h incubation. Bottom panels: amino acids found at positions P<sub>3</sub> and P<sub>2</sub> of the cleaved peptides at different times of incubation with FheCL1 (top, blue) or FheCL3 (bottom, red) at 5, 15, and 60 min. Results are expressed as percentage per site. The amino acid frequency at each position within the tetradecapetide library ranges from 4.2% to 6.8%. Met is substituted by norleucine in the library.

TABLE S1

## Generation of active site mutants by site directed mutagenesis.

Enzyme variant	Plasmid Template	Oligonucleotide	Sequence*
FheCL3 H61N	FheCL3 wild type	FheCL3 H61N Fw	CAGAAGATTTGGCAACA <b>AA</b> CGGATGTGGAGGTGG
		FheCL3 H61N Rv	CCACCTCCACATCC <b>GTT</b> GTGCCAAATCTTCTG
FheCL3 W67L	FheCL3 wild type	FheCL3 W67L Fw	CGGATGTGGAGGTGG <b>TTT</b> GATGGAGAACGC
		FheCL3 W67L Rv	GCGTTCTCCAT <b>CAA</b> ACCACCTCCACATCCG
FheCL3 H61N W67L	FheCL3 H61N	FheCL3 W67L Fw & Rv	
FheCL1 N61H	FheCL1 wild type	FheCL1 N61H Fw	CGACCTTGGGGAAAT <b>CAT</b> GGTTGCGGTGG
		FheCL1 N61H Rv	CCACCGCAACCA <b>TG</b> ATTCCCAAGGTCG
FheCL1 L67W	FheCL1 wild type	FheCL1 L67W Fw	GTTGCGGTGGTGG <b>A</b> TGGAAAATGCTTACC
		FheCL1 L67W Rv	GGTAAGCATTTTCCAT <b>CC</b> ATCCACCACCGCAAC
FheCL1 N61H L67W	FheCL1 N61H	FheCL1 L67W Fw & Rv	

\* Residues introducing the desired mutation are indicated in italics bold

### Table S1.

**Generation of active site mutants by site directed mutagenesis.** Oligonucleotide primers and templates used are indicated.

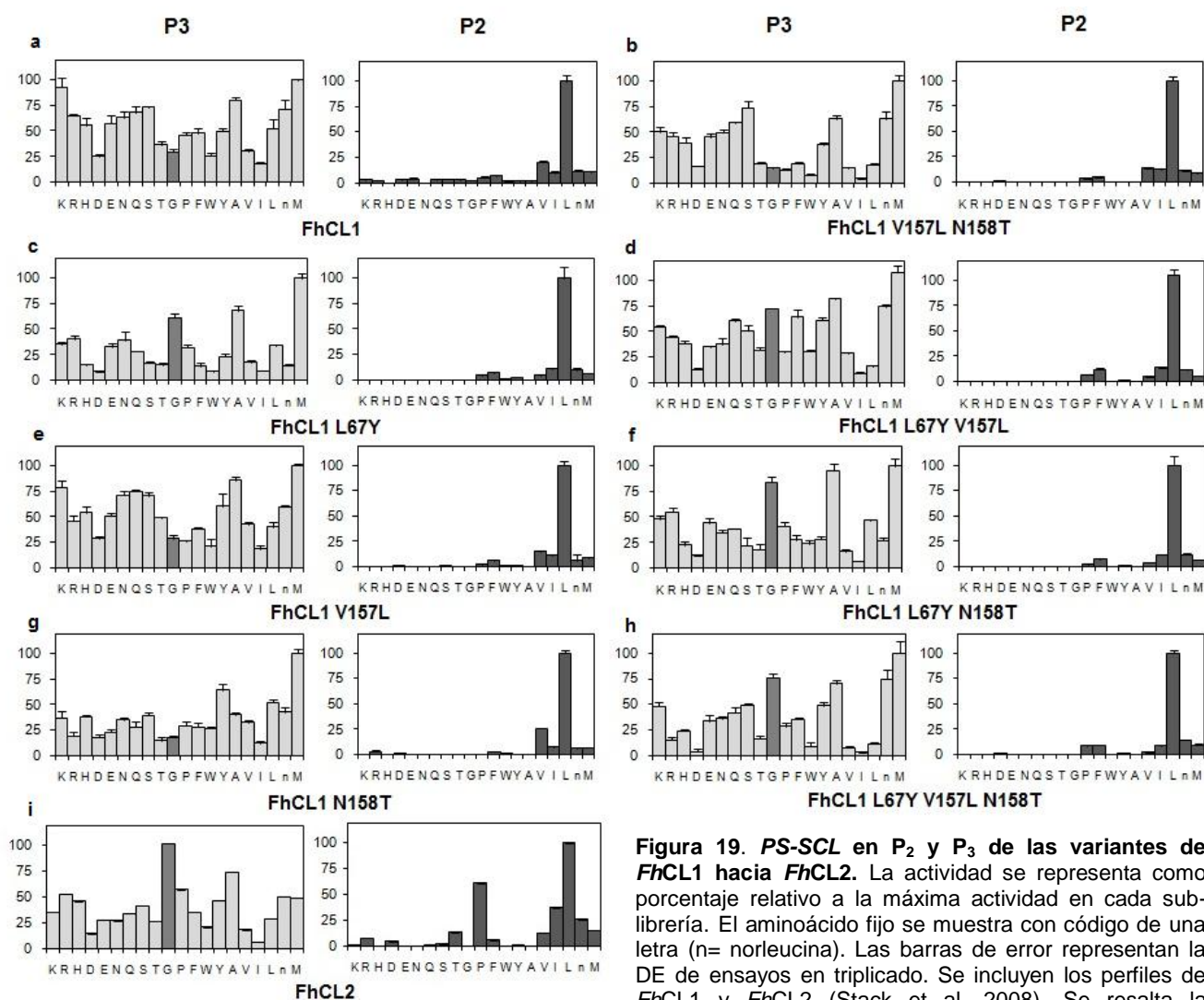
#### 4.2. Contribución de otros residuos variables del subsitio $S_2$ a la especificidad de sustrato

El análisis del sitio activo de las catepsinas de *F. hepatica* muestra otros residuos variables además de los ya estudiados. Su papel en la interacción con el sustrato no ha sido aún explorado (posiciones 157 y 158, ver tabla 1), si bien su relevancia ha sido sugerida mediante análisis estructural del sitio activo de *FhCL1* (Corvo et al. 2009; Stack et al. 2008). El residuo 157 está ubicado a la entrada del bolsillo  $S_2$  opuesto al residuo 67, y el residuo adyacente 158, se localiza junto a la histidina catalítica 159 (Stack et al. 2008). La presencia de treonina en la posición 158 es compartida por *FhCL2* y *FhCL3*, por lo que podría estar asociado a su capacidad de interactuar con prolina en  $P_2$ . Se realizaron mutantes de *FhCL1* hacia *FhCL2* con el cambio V157L y N158T, y para realizar ensayos comparativos se generó también el mutante de *FhCL1* hacia *FhCL2*, *FhCL1* L67Y, así como los tres dobles mutantes y un mutante con las tres sustituciones, cuyo bolsillo  $S_2$  tienen los mismos residuos que *FhCL2*. Todas las sustituciones en el sitio activo se generaron mediante mutagénesis sitio dirigida (la tabla con los oligonucleótidos empleados se muestra en el anexo 1). La especificidad de las enzimas mutantes se determinó empleando las *PS-SCL* (Figura 19) y mediante la hidrólisis de dos sustratos sintéticos fluorogénicos (tabla 3).

En la Figura 19 se observa que ninguno de los cambios introducidos en el bolsillo  $S_2$  modifica cualitativamente las preferencias de *FhCL1* en  $P_2$ . Incluso en la enzima mutante con las tres sustituciones, cuyo bolsillo  $S_2$  posee los mismos residuos que el de *FhCL2*, no se aprecia un incremento en la capacidad de acomodar prolina en  $P_2$  (Figura 19h). Al igual que como se observó con la sustitución de Leu67 de *FhCL1* por Trp, el cambio L67Y hace que aumente la afinidad por glicina en el bolsillo  $S_3$  (Figura 19c). Este aumento es aún mayor cuando la sustitución L67Y se combina con otra en las variantes con dos (Figura 19d y f) o tres sustituciones (Figura 19h). Así, el bolsillo  $S_3$  de *FhCL1* adquiere una conformación que facilita su interacción con el residuo glicina, asemejándose a las preferencias de *FhCL2* (Figura 19i), y más aún, a las de *FhCL3* (Corvo et al. 2013).

Al emplear sustratos peptídicos sintéticos para estimar los parámetros cinéticos de las enzimas encontramos que todas ellas causan en mayor o menor medida una reducción en la especificidad hacia el sustrato Z-VLK-AMC, ya sea disminuyendo la constante catalítica o la afinidad, pero solo provocan un leve aumento de la hidrólisis del sustrato Tos-GPR-AMC (tabla 3). La sustitución L67Y causa una disminución a la mitad en la especificidad de la enzima por el péptido Z-VLK-AMC, debido a una reducción de la constante catalítica ( $k_{cat}$ ). La  $k_{cat}$  hacia el sustrato Tos-GPR-AMC también disminuye pero es compensada con un aumento en la afinidad, manteniéndose en el mismo rango la constante de especificidad. Esto sugiere que la presencia de la tirosina en posición 67 favorece la interacción con el sustrato pero de alguna forma modifica la conformación del sitio activo hacia una configuración que dificulta la hidrólisis del enlace peptídico. En las variantes simples *FhCL1* V157L y N158T se observa una

disminución en la  $k_{cat}$  y un aumento en la  $K_M$  para el sustrato Z-VLK-AMC, respectivamente. Para estas mutaciones en combinación con la sustitución L67Y se observa el mismo efecto aunque más pronunciado, disminuyendo aún más la especificidad de las enzimas con las dobles sustituciones hacia el péptido con Leu en P<sub>2</sub> con respecto a los mutantes simples. Sin embargo, los parámetros cinéticos de la enzima con la doble sustitución V157L N158T son del mismo orden que los de las variantes únicas, mostrando que la presencia de sustituciones a ambos lados del bolsillo S<sub>2</sub> de *FhCL1* provoca una mayor distorsión en la interacción con el sustrato Z-VLK-AMC. El hecho de que el mutante triple muestre una reducción más pronunciada en la especificidad hacia este sustrato, sugiere que todas las sustituciones están contribuyendo a disminuir la preferencia por el péptido Z-VLK-AMC, y las combinaciones provocan un efecto sinérgico. En cambio, la afinidad por Tos-GPR-AMC se ve solo levemente incrementada con las sustituciones realizadas (menor  $K_M$ ). Este aumento probablemente se deba a la capacidad de las enzimas mutantes de acomodar Gly en esa posición, dado que los datos en las *PS-SCL* indican un incremento en la afinidad hacia glicina en P<sub>3</sub>.



**Figura 19.** *PS-SCL* en P<sub>2</sub> y P<sub>3</sub> de las variantes de *FhCL1* hacia *FhCL2*. La actividad se representa como porcentaje relativo a la máxima actividad en cada sub-librería. El aminoácido fijo se muestra con código de una letra (n= norleucina). Las barras de error representan la DE de ensayos en triplicado. Se incluyen los perfiles de *FhCL1* y *FhCL2* (Stack et al. 2008). Se resalta la frecuencia de glicina en P<sub>3</sub> en gris oscuro.

Tabla 3. Parámetros cinéticos de las variantes de *FhCL1* hacia *FhCL2*

Enzima	Sustrato					
	Z-VLK-AMC			Tos-GPR-AMC		
	$k_{cat}$ ( $s^{-1}$ )	$K_M$ ( $\mu M$ )	$k_{cat}/K_M$ ( $M^{-1} s^{-1}$ )	$k_{cat}$ ( $s^{-1}$ )	$K_M$ ( $\mu M$ )	$k_{cat}/K_M$ ( $M^{-1} s^{-1}$ )
<i>FhCL1</i>	$2.32 \pm 0.120$	$3.94 \pm 0.93$	$5.89 \times 10^5$	$0.0350 \pm 0.006$	$15.10 \pm 6.70$	$2.36 \times 10^3$
<i>FhCL1</i> L67Y	$0.53 \pm 0.11$	$1.72 \pm 0.06$	$3.08 \times 10^5$	$0.0240 \pm 0.003$	$8.52 \pm 2.70$	$2.82 \times 10^3$
<i>FhCL1</i> V157L	$0.72 \pm 0.07$	$5.23 \pm 0.58$	$1.38 \times 10^5$	$0.0274 \pm 0.001$	$13.16 \pm 1.51$	$2.08 \times 10^3$
<i>FhCL1</i> N158T	$2.12 \pm 0.14$	$18.38 \pm 4.1$	$1.15 \times 10^5$	$0.0227 \pm 0.001$	$6.84 \pm 1.73$	$3.32 \times 10^3$
<i>FhCL1</i> L67Y V157L	$0.17 \pm 0.010$	$3.85 \pm 0.61$	$4.42 \times 10^4$	$0.0270 \pm 0.003$	$6.57 \pm 2.40$	$4.06 \times 10^3$
<i>FhCL1</i> L67Y N158T	$1.35 \pm 0.100$	$23.30 \pm 4.00$	$5.79 \times 10^4$	$0.0310 \pm 0.004$	$8.81 \pm 3.80$	$3.51 \times 10^3$
<i>FhCL1</i> V157L N158T	$1.67 \pm 0.11$	$10.2 \pm 1.94$	$1.64 \times 10^5$	$0.0150 \pm 0.001$	$3.32 \pm 1.90$	$4.52 \times 10^3$
<i>FhCL1</i> L67Y V157L N158T	$0.65 \pm 0.07$	$25.6 \pm 6.7$	$2.54 \times 10^4$	$0.0302 \pm 0.002$	$6.86 \pm 1.85$	$4.40 \times 10^3$

## **V. DISCUSIÓN GENERAL**

## 1. La familia de catepsinas L del estadio juvenil

Los organismos parásitos han evolucionado durante millones de años para adaptarse a la vida en el entorno hostil que supone el interior de un hospedero. Este desarrolla a su vez mecanismos de defensa que apuntan a la erradicación del parásito. Estos patógenos sintetizan y secretan una batería de moléculas efectoras que utilizan para penetrar las barreras de defensa del hospedador y establecerse procurando la adquisición de nutrientes y modulando la respuesta inmune para sobrevivir dentro del mismo. En los parásitos trematodos del género *Fasciola*, es notable la predominancia de cisteína proteasas tipo catepsinas en los extractos somáticos y en los PES del gusano. Varios trabajos mostraron que las catepsinas producidas por el parásito no son las mismas a lo largo del ciclo de vida. Se observó que las proteínas secretadas durante los primeros estadios de infección en el intestino del hospedero son diferentes a las del parásito inmaduro 21 días post-infección, los cuales muestran un patrón más similar al gusano adulto residente en los canalículos biliares (Cancela et al. 2008; Tkalcevic et al. 1995). Con el advenimiento de las técnicas de secuenciación masiva y la electroforesis bidimensional para la identificación de proteínas por espectrometría de masas se realizaron varios estudios de transcriptómica y proteómica de los diferentes estadios del parásito, confirmando que las cisteína proteasas son las proteínas secretadas más representadas y que diferentes catepsinas son expresadas durante el transcurso de una infección (Cancela et al. 2010; Robinson, Tort, et al. 2008; Robinson et al. 2009). A pesar de contar con varias bibliotecas de ESTs de los diferentes estadios del parásito (Cancela et al. 2010; Robinson et al. 2009; Young et al. 2010), el genoma completo de *Fasciola hepatica* aún no ha sido secuenciado por lo que el complemento completo de proteínas que dispone el parásito es desconocido.

Al inicio de este trabajo las catepsinas principales secretadas por el parásito adulto, *FhCL1* y *FhCL2*, ya habían sido purificadas (Smith et al. 1993) y sus genes clonados para permitir la caracterización fisicoquímica de las enzimas recombinantes (Dowd et al. 1997; Roche et al. 1997). En la base de datos del GenBank se encontraban depositadas varias isoformas de las catepsinas de adultos pero unas pocas secuencias correspondían a catepsinas expresadas por el estadio juvenil de *F. hepatica*. El grupo de Harmsen y cols. (Harmsen et al. 2004) identificó en juveniles una proteína cuya secuencia deducida del extremo N-terminal coincidía con el de una proteína identificada previamente en extractos somáticos de juveniles (Tkalcevic et al. 1995); a la cual denominaron *FhCL3*. En nuestro grupo se confirmó la presencia de *FhCL3* en ESTs de juveniles y se identificó una segunda catepsina a partir de ADNc de JRD que se denominó *FhCL4* (Cancela et al. 2008). Buscando clonar las catepsinas completas de juveniles generamos una mini-librería de ADNc, obteniendo 12 secuencias diferentes. El análisis filogenético de estas secuencias permitió la identificación de un tercer grupo de cisteína proteasas del estadio juvenil, al cual denominamos *FhCL6* (Figura 10). La existencia de varios genes codificantes de catepsinas en los JRD no es sorprendente ya que

en el estadio adulto del parásito también se han descrito al menos 3 grupos de catepsinas (*FhCL1: FhCL1A & B, FhCL2 y FhCL5*) (Robinson, Dalton, et al. 2008). Al igual que ocurre con las secuencias de las catepsinas de adultos, se detectaron pequeñas variaciones (polimorfismos nucleotídicos o SNP) en las secuencias de juveniles dentro de cada uno de los grupos (Figura 9). Esta heterogeneidad se observó también en estudios de proteómica de los PES de gusanos adultos, donde se identificaron múltiples catepsinas con masa similar pero puntos isoeléctricos diferentes (Morphew et al. 2007, 2011). Estos polimorfismos pueden deberse ya sea a la presencia de varios alelos de catepsinas (dado que el RT-PCR se realizó sobre ARN extraído de una población heterogénea de parásitos) o bien a la existencia de múltiples genes ortólogos codificantes de las catepsinas, lo que deberá evaluarse mediante estudios de organización del genoma del parásito. A favor de esta hipótesis, se han identificado en *F. hepatica* al menos 5 genes de catepsinas L (Heussler & Dobbelaere 1994) y en *F. gigantica* se estimó mediante *Southern blot* una familia de aproximadamente 10 genes ortólogos de catepsinas L (Grams et al. 2001). La duplicación génica y la divergencia funcional que experimentan posteriormente muchas familias génicas es un proceso fundamental en la evolución de nuevas funciones, y lleva a la aparición de genes que cumplen funciones diferentes aunque a menudo relacionadas. Es plausible que este mecanismo haya moldeado la evolución de la familia de catepsinas de *F. hepatica* (Irving et al. 2003), sin embargo dada la alta similitud de secuencias, la caracterización de esta familia génica aguarda a que se complete el genoma del parásito.

## 2. Expresión de *FhCL3* recombinante activa

Para estudiar las propiedades fisicoquímicas y la especificidad de sustrato de las catepsinas, se puso a punto un sistema eficiente para la expresión de las enzimas recombinantes en el laboratorio. Se eligió un sistema de expresión eucariota sencillo pero con la ventaja de permitir el correcto plegamiento de estas enzimas, como lo son las levaduras, el cual se optimizó para la producción de las catepsinas secretadas por el estadio adulto y juvenil de *Fasciola hepatica*, *FhCL1* y *FheCL3*. Durante el transcurso de esta tesis, estudios proteómicos demostraron que *FhCL3* es la única catepsina L que se encuentra en los productos de excreción/ secreción de JRD cultivados *in vitro* (Robinson et al. 2009), sugiriendo un papel preponderante en la interfase parásito-hospedador. Puesto que *FhCL4*, y el tercer grupo identificado, *FhCL6*, son detectadas únicamente a nivel de ARNm, es probable que participen en procesos intracelulares del metabolismo del parásito o sean secretadas en muy bajas cantidades. En base a estos resultados se decidió focalizar en el estudio de *FhCL3*. Se puso a punto la expresión de esta catepsina en la levadura *H. polymorpha*, obteniéndose la enzima madura y la proenzima secretadas al medio de cultivo. Se probaron diferentes medios de inducción, y se determinó que tanto en medio rico como en medio mínimo las enzimas son secretadas al medio de cultivo, pero solo se detectó actividad en las levaduras inducidas en



medio mínimo. El análisis del sobrenadante del cultivo en medio rico mediante SDS-PAGE mostró la presencia de la proenzima L3, sugiriendo que la misma se está expresando, pero en medio rico no ocurre el autoprocésamiento como sí sucede en el medio mínimo. Dado que los niveles de expresión resultaron similares en ambos casos y es más fácil realizar el tamizaje de clones midiendo actividad enzimática, decidimos adoptar el protocolo de expresión en medio mínimo.

Anteriormente otros grupos habían clonado y expresado *FhCL3* en forma recombinante en la levadura *S. cerevisiae* y en baculovirus, obteniéndose la proenzima en ambos casos, sin lograr procesarla a su forma activa. Este hecho, sumado a la presencia de una asparagina en la posición P<sub>1</sub> del sitio de corte del propéptido, sugería la necesidad de una asparaginil endopeptidasa para el procesamiento de *FhCL3* a su forma madura (Harmsen et al. 2004; Reszka et al. 2005). Sin embargo, en nuestro sistema parte de la enzima se activa durante o luego de su secreción al medio, y la porción que se expresa en forma de zimógeno es capaz de autoprocésarse a la forma madura al ser incubada durante 3 horas en un tampón ácido (pH 4 preferentemente) en presencia de un agente reductor como el DTT. Este resultado concuerda con lo observado para las catepsinas del estadio adulto, *FhCL1* y *FhCL2* (Stack et al. 2008), que son capaces de hidrolizar el propéptido de forma intra o intermolecular en medio reductor generando la enzima activa. Cabe destacar que la secuencia inicial del propéptido de *FhCL3* expresada en los otros 2 sistemas (H<sub>3</sub>N-SNDVSWHEWK) es idéntica a la nuestra. Por otro lado, nuestros estudios del estado de glicosilación de la enzima recombinante sí concuerdan con los resultados obtenidos en los estudios anteriores, donde encuentran que la proenzima migra con un peso molecular aparente mayor al esperado, y que se produce una forma de menor tamaño al incubar la enzima en presencia de deglicosilasas, tanto cuando es expresada en la levadura *S. cerevisiae* como en células de insecto (Harmsen et al. 2004; Reszka et al. 2005). Este hecho era predecible dado que *FhCL3* presenta sitios potenciales de *N*-glicosilación, si bien se ha descrito que la enzima nativa parece no estar glicosilada ya que su patrón de migración en SDS-PAGE coincide con el de la enzima recombinante de *S. cerevisiae* tratada con deglicosilasas (Reszka et al. 2005). Los diferentes patrones de glicosilación seguramente tengan implicancias para el uso de *FhCL3* recombinante como antígeno en la generación de vacunas o ensayos de inmunodiagnóstico, como discutiremos más adelante. Si bien puede estar influyendo en la potencia de la enzima o su estabilidad, el patrón de glicosilación no se espera que modifique la especificidad de la enzima.

Con la expresión en la levadura *H. polymorpha* también se logró un aumento significativo en la producción de la enzima *FhCL1* recombinante secretada al medio de cultivo. Roche et al., 1997 habían obtenido en el entorno de 0.5 mg/L al expresarla en *S. cerevisiae*. Inicialmente, este nivel de expresión se mejoró con rendimientos de 0.8 – 1 mg/L (Corvo et al. 2009) y las condiciones de cultivo fueron posteriormente optimizadas al protocolo descrito en la primera sección de Resultados, con el cual las cantidades de proteína obtenidas alcanzan los

10 mg/L. Además, se desarrolló un protocolo que permite obtener las enzimas con un alto grado de pureza directamente en los sobrenadantes del cultivo de las levaduras inducidas en medio con metanol. Esto tiene la ventaja de reducir enormemente los pasos de purificación de la proteína y de esta manera evitar la disminución en los rendimientos de producción inherentes a los métodos de purificación. Esto es posible pues esta levadura no produce cisteína proteasas secretadas intrínsecas, lo que dificultaría la purificación de las enzimas. Esto fue corroborado induciendo la expresión del gen ScPEX3 (Figura 8) en paralelo y en las mismas condiciones que los genes de catepsinas de *Fasciola*, y constatando que no se detecta actividad cisteína proteasa en el sobrenadante del cultivo (datos no mostrados).

Demostramos que la catepsina L3 recombinante muestra características típicas de las cisteína proteasas del tipo papaína (clan C1A), siendo activada por agentes reductores como el DTT o cisteína, e inhibida por E-64. Se determinó su estabilidad a 37°C, temperatura a la cual se encuentra dentro del hospedero mamífero. Se observó que *FhCL3* es capaz de conservar un 50% de su actividad luego de 8 horas de incubación a 37°C en un tampón de pH neutro. En comparación con la catepsina *FhCL1*, se realizó un perfil de actividad en un rango de pH 3 - 9 a concentraciones saturantes de sustrato, observando la dependencia de  $k_{cat}$  en función del pH. La curva indica la presencia de dos grupos ionizables, presumiblemente correspondientes al tiol y la histidina que forman la díada catalítica, u otros grupos cuyo estado de protonación es importante para la catálisis en el complejo enzima-sustrato. El desfasaje en la curva de una enzima con respecto a la otra sugiere diferencias en el entorno del grupo ionizable. Asimismo, *FhCL3* presenta su óptimo de actividad a pH 7, mientras *FhCL1* muestra su pico de actividad a un pH un poco menor, reflejando posibles adaptaciones a las condiciones de los diferentes microentornos en los que son secretadas. Es interesante notar que la autoactivación de la proenzima ocurre a pH ácido, mientras el máximo de actividad de las enzimas es en un entorno neutro. Algunos estudios sugieren que el pH del tubo digestivo de *F. hepatica* es ligeramente ácido (Halton, 1997) en tanto que el pH del duodeno y de los canalículos hepáticos normalmente se encuentra en valores neutros o ligeramente alcalinos (Rune, 1973). Consecuentemente, es probable que la enzima se active una vez secretada al tubo digestivo del parásito, y alcance actividad máxima al ser regurgitada, donde puede degradar proteínas del hospedero.

### 3. Especificidad de sustrato de las catepsinas de *F. hepatica*

#### 3.1. Hidrólisis de sustratos peptídicos y librerías sintéticas

En estudios previos de nuestro grupo se encontró que la actividad hidrolítica sobre péptidos sintéticos en los PES de JRD cultivados *in vitro* era en el orden tos-GPR > Z-FR > Z-RR-AMC (Cancela et al. 2008), sugiriendo la presencia de catepsinas tanto de tipo L como de tipo B. El análisis de las preferencias de sustrato de la catepsina L3 recombinante permitió

determinar que es responsable de la actividad mayor, dado que mostró una preferencia 70 veces mayor por el sustrato Tos-GPR-AMC frente a los sustratos Z-FR-AMC o Z-VLK-AMC, típicos de catepsinas L de adultos (Corvo et al. 2009). Posteriormente se evaluó un número mayor de oligopéptidos sintéticos incluyendo di, tri y tetrapéptidos con residuos de diferentes características (aromáticos, alifáticos, prolina) determinándose las constantes cinéticas mediante regresión no lineal a partir de la curva de Michaelis Menten. Al ampliar el panel de sustratos peptídicos sintéticos, se observa que *FhCL3* hidroliza con mayor eficiencia los péptidos que tienen prolina en P<sub>2</sub> (Tos-GPR-AMC > Boc-AGPR-AMC > Tos-GPK-AMC), siendo Boc-VPR-AMC la excepción, ya que la constante de especificidad para este sustrato es entre 20 y 60 veces menor que la de los otros tres péptidos conteniendo Pro en P<sub>2</sub> (Robinson et al. 2011). Cabe destacar que Boc-VPR-AMC es el único de estos péptidos que no posee una Gly en P<sub>3</sub>, lo que sugiere la relevancia del residuo acomodado en el bolsillo S<sub>3</sub> en la interacción de *FhCL3* con sus sustratos.

Para que la comparación de los parámetros cinéticos sea fiable, se requiere realizar la cinética y estimación de los parámetros en las mismas condiciones de temperatura, pH, concentración de enzima y fuerza iónica, preferentemente en el mismo instrumento. Esto se debe a que los valores absolutos de las constantes pueden variar con las condiciones experimentales. Los parámetros cinéticos se calcularon entonces en paralelo para la catepsina de juveniles y las catepsinas de adultos para todos los péptidos sintéticos evaluados (Robinson et al. 2011). Los resultados obtenidos fueron consistentes con los datos previos reportados para las enzimas *FhCL1* y *FhCL2*, mostrando que ambas son eficientes en la hidrólisis de oligopéptidos con aminoácidos hidrofóbicos en la posición P<sub>2</sub>, como Leu y Phe (Stack et al. 2008). *FhCL2* puede también acomodar péptidos con Pro en esa posición, pero en forma menos eficiente. Notoriamente, y como los resultados preliminares anticipaban, *FhCL3* resultó ser la más eficiente hidrolizando péptidos con Pro en P<sub>2</sub>, siendo los demás péptidos en general pobremente hidrolizados. Los parámetros mostraron aún más eficiencia que la catepsina K de mamíferos, considerada el prototipo de catepsina colagenolítica.

Para reafirmar los datos obtenidos con los sustratos peptídicos, se ensayó el efecto de un par de inhibidores sintéticos selectivos de cisteína proteasas, uno irreversible, la peptidil diazometilcetona Z-Phe-Ala-CHN<sub>2</sub> y otro reversible, el inhibidor II de catepsina K (Z-LNHNHCONHNHLF-Boc, CKII). Las diazometilcetonas peptídicas inhiben específicamente las cisteína proteasas, por lo que son inhibidores selectivos (Green & Shaw 1981; Shaw & Green 1981). Su mecanismo de inhibición se basa en "imitar" al sustrato peptídico al interactuar con el sitio activo de la enzima pero una vez unido ésta no es capaz de hidrolizarlo para reanudar un nuevo ciclo de catálisis, por lo que variando los residuos de la porción peptídica es posible modular su capacidad inhibitoria frente a distintas cisteína proteasas. Los cambios en las tasas de inactivación de estos inhibidores han mostrado diferencias subyacentes en la afinidad de los subsitios de unión al sustrato de las cisteína proteasas B y L (Kirschke & Shaw 1981). Tanto

*FhCL1* como *FhCL2* fueron fuertemente inhibidas por Z-Phe-Ala-CHN<sub>2</sub>, sin embargo, la inhibición sobre *FhCL3* fue 20 veces menor, indicando que *FhCL3* no acomoda fácilmente aminoácidos hidrofóbicos en P<sub>2</sub>. Con respecto al inhibidor II de catepsina K, *FhCL2* resultó ser la enzima que mejor se inhibe ( $K_i$  0.46 nM), lo que refleja el hecho que su sitio activo comparte con la catepsina K humana aminoácidos en posiciones clave que han sido postulados como determinantes de la especificidad de sustrato. La inhibición sobre *FhCL1* fue menor pero con una  $K_i$  del mismo orden de magnitud (0.63 nM), mientras que la inhibición sobre *FhCL3* fue 20 veces menor ( $K_i$  9.05 nM) (Robinson, Corvo et al. 2011). El uso de péptidos e inhibidores sintéticos confirman la hipótesis de que la catepsina L3 secretada por el estadio juvenil de *F. hepatica* tiene diferente especificidad de sustrato que las enzimas secretadas por el estadio adulto, mostrando un comportamiento atípico, en cuanto a sus preferencias de sustrato, para las catepsinas de tipo L. Tanto las catepsinas L de mamíferos como las parasitarias prefieren en general sustratos con aminoácidos aromáticos (catepsina L y V de mamíferos) o alifáticos en P<sub>2</sub> (catepsina K, S y F de mamíferos, papaína, catepsinas *FhCL1* y *FhCL2* de *F. hepatica*) (Choe et al. 2006), mientras *FhCL3* acomoda mejor sustratos con Pro en esa posición (Corvo et al. 2009; Robinson, Corvo et al. 2011).

Al estudiar las enzimas con dos librerías peptídicas sintéticas los resultados resaltaron las mismas diferencias en la especificidad de sustrato de estas catepsinas que los datos anteriores. Muestran el rol dominante del subsitio S<sub>2</sub> así como la especificidad de *FhCL3* por residuos de Pro en P<sub>2</sub>, en contraste con *FhCL1*, la cual acomoda mejor residuos alifáticos en esa posición (Corvo et al. 2013). En general, el bolsillo S<sub>3</sub> de las catepsinas acepta bien cualquier aminoácido, si bien ha sido reportada selectividad hacia Pro en la papaína y bromelaína (Choe et al. 2006). El análisis del perfil de aminoácidos acomodados por *FhCL3* en S<sub>3</sub> nos permitió confirmar su selectividad hacia el residuo Gly como sugerían los resultados anteriores. Si comparamos los perfiles P<sub>1</sub>-P<sub>4</sub> obtenidos con las *PS-SCL* para las tres enzimas principales de *F. hepatica* es interesante notar que *FhCL2* tiene características intermedias entre *FhCL3* y *FhCL1*. Así, es capaz de acomodar Pro en P<sub>2</sub> aunque en menor medida que Leu, a la vez que Gly es el aminoácido que ocupa con mayor frecuencia el sitio S<sub>3</sub>, si bien la diferencia en frecuencia con el resto de los residuos no es tan marcada como en *FhCL3*.

Las preferencias del lado C-terminal del sitio de corte (P<sub>1</sub>'-P<sub>4</sub>') son menos conocidas pues dado el papel central que se ha atribuido al sitio P<sub>2</sub> en la especificidad de las catepsinas, en general las estrategias de estudio se han basado en sustratos fluorogénicos que ubican el fluoróforo en P<sub>1</sub>'. Los datos de *MSP-MS* muestran que a diferencia de la enzima del adulto, *FhCL3* tiene cierta restricción en el bolsillo S<sub>1</sub>' acomodando con mayor frecuencia residuos de serina y glicina (Corvo et al. 2013). En el futuro sería interesante profundizar en la especificidad del lado prima del sitio de corte de la enzima, empleando otras herramientas desarrolladas para abordar su estudio (Hemerly et al. 2003; Lanfranco et al. 2008; O'Donoghue et al. 2012).

Las diferencias de especificidad reflejan la historia evolutiva de las catepsinas de *Fasciola*. Los análisis filogenéticos muestran que los genes codificantes de *FhCL3* son el linaje más divergente, lo que llevó a plantear que probablemente sería la más antigua de las enzimas (Irving et al. 2003; Robinson, Tort, et al. 2008; Stack, Dalton & Robinson 2011), a partir de la cual evolucionaron las catepsinas de adultos posiblemente mediante eventos de duplicación y divergencia (Irving et al. 2003). En estos artículos se propone además que *FhCL1* evolucionó a través de *FhCL2*. Su sitio activo interacciona bien con residuos alifáticos en P<sub>2</sub>, siendo una enzima especializada en la digestión de hemoglobina (Lowther et al. 2009), principal fuente de nutrientes para el parásito adulto. El sitio activo de *FhCL2* muestra características de ambas enzimas, manteniendo la capacidad de acomodar residuos con Pro en P<sub>2</sub>, con menos frecuencia que los residuos alifáticos, y de digerir el colágeno, aunque en forma menos eficiente que la enzima secretada por el estadio juvenil. Sin embargo, no podemos descartar que *FhCL3* sea una enzima especializada y que la mayor divergencia se deba a selección positiva actuando a favor de la adquisición de una mayor eficiencia en la hidrólisis de los sustratos que enfrenta el parásito juvenil. De hecho, Irving y col. (Irving et al. 2003) identificaron regiones posiblemente sujetas a selección positiva que incluyen a varios de los residuos del sitio activo, entre ellos los aquí estudiados por mutagénesis.

### 3.2. Hidrólisis de sustratos macromoleculares

Durante el inicio de la infección por *Fasciola* los JRD deben atravesar la pared intestinal del hospedero, y migrar por la cavidad peritoneal y el parénquima hepático hacia los canalículos biliares del hígado donde se aloja y madura al estadio adulto. Resulta interesante entonces estudiar la capacidad de *FhCL3* de degradar componentes de la matriz extracelular con los cuales pudiera enfrentarse durante este trayecto, tales como distintos tipos de colágeno, fibronectina y laminina, entre otros.

Los colágenos están estrechamente involucrados en la formación de redes fibrilares, membranas basales así como otras estructuras de la matriz extracelular. El colágeno tipo I es el más abundante en el organismo, seguido de los colágenos tipo II, III y IV siendo el colágeno II un componente mayoritario del cartílago hialino. El colágeno se caracteriza por contener dominios con repetidos del tripéptido Gly-X-Y ricos en Pro e hidroxiprolina (Hyp), involucrados en la formación de la triple hélice del colágeno (Gelse, 2003). La habilidad de acomodar residuos de Pro en la posición P<sub>2</sub> del sustrato de *FhCL2* se asoció con su capacidad de degradar la triple hélice del colágeno tipo I, mientras que *FhCL1* solo degrada las regiones teloméricas no helicoidales de la molécula (Stack et al. 2008). Dada su alta afinidad por péptidos conteniendo Pro en P<sub>2</sub>, *FhCL3* se postulaba como una enzima que podría degradar colágeno. De acuerdo con esta hipótesis, observamos que efectivamente la enzima tiene la capacidad de degradar el colágeno tipo I en condiciones de pH y temperatura que favorecen la estabilidad de su triple hélice (pH neutro y 28°C). La actividad colagenolítica de secreciones y

extractos somáticos del parásito juvenil había sido tempranamente descrita por Howell en 1966 (Howell, 1966), utilizando un método colorimétrico de liberación de color del azocol. Nuestros resultados sugieren que la enzima principal responsable de esta actividad es *FhCL3*. Es importante notar que el pH y temperatura del medio tiene una gran influencia en la capacidad de las enzimas de digerir determinados sustratos, probablemente debido a variaciones en la conformación o plegamiento de estos. Como ejemplo, podemos mencionar la capacidad de *FhCL1* de digerir la hemoglobina y el colágeno tipo I a pH ácido pero no a pH neutro o alcalino (Lowther et al. 2009; Stack et al. 2008). Al comparar la actividad de *FhCL1*, *FhCL2* y *FhCL3* a pH ácido o neutro sobre los distintos tipos de colágeno encontramos que solo *FhCL3* y, en menor medida, *FhCL2* son capaces de hidrolizarlo a pH 7, mientras *FhCL1* incluso luego de 20 horas de incubación no es capaz de hidrolizarlo. *FhCL1* solo hidroliza el colágeno a pH 5.5 y principalmente actúa sobre los dímeros  $\beta$  siendo las cadenas  $\alpha_1$  y 2 más resistentes (Robinson, Corvo et al. 2011). Cabe mencionar que los estudios previos de degradación de moléculas de colágeno por parte de PES y de las enzimas *FhCL1* y *FhCL2* aisladas de parásitos adultos fueron realizados a pH ácido (5.5 o 6) y 37°C (Berasaín et al. 1997). Al analizar la frecuencia con que aparece cada residuo en las posiciones P<sub>1</sub>-P<sub>4</sub> de los fragmentos resultantes de la digestión del colágeno mediante HPLC-MS, se observó que en P<sub>1</sub>, P<sub>3</sub> y P<sub>4</sub> aparecen en general varios aminoácidos, siendo Gly, Pro y Ala los hallados con mayor frecuencia. Este sesgo podría deberse a que estos tres aminoácidos son los encontrados en mayor porcentaje en el colágeno (anexo 2). De acuerdo a lo esperado, tanto *FhCL2* como *FhCL3* mostraron selectividad en el subsitio S<sub>2</sub>, pero no hacia Pro como ocurre con los péptidos sintéticos, sino hacia Gly. Esto muestra que el uso de péptidos sintéticos si bien resulta útil para estudiar la selectividad de las enzimas, no necesariamente se correlaciona de forma directa con los sitios de corte observados en sustratos macromoleculares. Si bien las interacciones entre P<sub>2</sub>-S<sub>2</sub> son las más relevantes, las catepsinas no dejan de constituir un grupo promiscuo de proteasas capaces de hidrolizar una amplia gama de sustratos. Debemos considerar además que las proteínas no se presentan en una conformación extendida, sino que están plegadas en una estructura tridimensional determinada, lo que hace que ciertas regiones estén expuestas para interactuar con el sitio activo de la enzima mientras otras permanecen ocultas. La cinética de la proteólisis implica que los cortes más tempranos se realicen sobre los residuos más accesibles de la proteína, generando un cambio en la estructura del sustrato que subsecuentemente expone otros enlaces peptídicos inicialmente inaccesibles. Al realizar los ensayos de digestión del colágeno procurando conservar su estructura nativa, las regiones en contacto con el solvente no necesariamente reflejen la secuencia lineal Gly-Pro-Y que se concibe como de alta afinidad para *FhCL3*. A pesar de la concordancia entre la cinética de péptidos sintéticos y la digestión de hemoglobina por *FhCL1* (Lowther et al. 2009), también se ha reportado que los sitios de hidrólisis sobre otros sustratos proteicos, como por ejemplo las IgGs, no refleja lo observado con los péptidos más cortos

(Berasain et al. 2000). Esto puede estar relacionado con el hecho de que en estos estudios el análisis de los productos de hidrólisis se realizó a tiempo final luego de varias horas de digestión, por lo que no estamos observando cuáles son los cortes iniciales, y si estos respetan las preferencias mostradas *in vitro* con sustratos sintéticos.

Otro componente mayoritario de la matriz extracelular es la fibronectina, la cual se une a receptores de membrana del tipo integrinas, así como a otros componentes de la matriz como el colágeno, la fibrina y los proteoglicanos. La laminina es una glicoproteína que forma parte de la lámina basal asociada a otras proteínas, entre ellas la fibronectina, colágeno, entactina, y tiene una función de anclaje de las células epiteliales. Ya había sido descrita la capacidad de las enzimas nativas secretadas por el estadio adulto de degradar estas dos macromoléculas (Berasain et al. 1997). En este trabajo demostramos que las enzimas *FhCL1* y *FhCL2* recombinantes también pueden degradar fibronectina y laminina a 37°C y pH neutro, así como la catepsina L3 del estadio juvenil, la cual a la misma concentración que las enzimas del adulto parece generar un mayor número de fragmentos de menor tamaño de fibronectina y es la única capaz de digerir eficientemente la laminina. La capacidad de *FhCL3* de digerir todos los componentes de matriz extracelular analizados sugiere que juega un rol importante en la migración del parásito juvenil durante las primeras etapas de la infección por *F. hepatica*. Evidencia más directa de esto lo constituye el estudio de interferencia de JRD sobre catepsinas L y B, al notar que los juveniles tienen disminuida su capacidad de atravesar el epitelio intestinal cuando se hace un *knock down* de estos genes (McGonigle et al. 2008). Sin embargo, los autores describen una parálisis en la mayoría de los JRD 4h después de incubarlos en el medio conteniendo el ARNdc blanco de las catepsinas, aunque el movimiento normal se restablece luego de 1 hora en medio sin ARNdc. Ensayos similares en nuestro laboratorio muestran también una movilidad reducida en los gusanos tratados con ARNdc contra *FhCL2*. Estos datos indican que tanto la disminución en la motilidad, como el silenciamiento en sí de las catepsinas L de juveniles están influyendo en la reducida capacidad de los parásitos interferidos de atravesar la pared del duodeno. Para confirmarlo, sería interesante determinar si la adición de las catepsinas de juveniles (*FhCL3* y catepsinas B) a los parásitos interferidos aumenta el porcentaje de parásitos capaces de atravesar el epitelio duodenal.

Se ha planteado que la capacidad para hidrolizar inmunoglobulinas tipo G en la región bisagra mostradas por *FhCL1* y *FhCL2*, es un posible mecanismo de evasión de la respuesta inmunitaria (Berasain et al. 2000; Smith et al. 1993). Mientras que *FhCL1* expresada en *H. polymorpha* confirma lo observado con proteasas de adultos purificadas al degradar IgG1, *FhCL3* no fue capaz de degradar este sustrato en las mismas condiciones donde sí es eficiente en la hidrólisis de colágeno tipo I (Corvo et al. 2009). Los organismos parásitos han desarrollado diferentes estrategias de defensa para evadir la respuesta inmune del hospedero, tales como la variación de los antígenos de la superficie del parásito o la interferencia directa

con las funciones efectoras del sistema inmune del hospedador. En este sentido, se ha reportado que los juveniles de *F. hepatica* pueden realizar rápidamente un recambio del glicocáliz de su superficie, lo que podrían emplear como mecanismo evasor (Hanna, 1980). La habilidad de digerir distintos tipos de IgGs descrita para las *FhCL1* y *FhCL2* podría reflejar el hecho de que el gusano adulto reside de forma permanente en una localización específica, en contacto con la sangre en los canalículos biliares del hígado, dando el tiempo suficiente al sistema inmune del hospedero de montar una respuesta efectiva. La forma juvenil por el contrario es un estado migratorio, que además va sufriendo un proceso de maduración que le lleva a modificar sus perfiles de proteínas secretadas a través del tiempo lo que podría ejercer un efecto protector en sí mismo.

#### 4. Correlación estructura-función en el sitio activo de las catepsinas

Para analizar la relación estructura-función que pudiese dar cuenta de la diferente especificidad de estas enzimas comparamos modelos moleculares de *FhCL2* y *FhCL3* con la estructura de otras catepsinas, entre ellas *FhCL1*, las catepsinas L y K humana y la zingipaína del jengibre. La catepsina K y la zingipaína se escogieron por su habilidad de acomodar Pro en  $P_2$  y de digerir el colágeno (Garnero et al. 1998; Kim et al. 2007). Las diferencias en el sitio activo de las enzimas de la familia de la papaína están dadas por unos pocos residuos que forman parte de los subsitios  $S_2$  y  $S_3$  (Drenth et al. 1968; Stack et al. 2008). Más específicamente, las posiciones variables son los residuos 61, 64, 67, 157, 158 y 205. Sorprendentemente, la mayor similitud estructural en el sitio activo de *FhCL3* se observó con la zingipaína, encontrándose los aminoácidos His61 y Trp67 conservados en ambas enzimas (Corvo et al. 2009). El residuo en la posición 67, que se localiza en el límite de los bolsillos  $S_2$  y  $S_3$ , ya ha demostrado ser relevante en la especificidad de estas enzimas (Smooker et al. 2000), mientras el residuo 61, ubicado en el piso del bolsillo  $S_3$  no había sido previamente analizado. Los modelos también revelan que la topología de los subsitios  $S_2$  y  $S_3$  de *FhCL3* es similar a la que presenta la zingipaína, siendo sus bolsillos  $S_2$  y  $S_3$  pequeños en comparación con las otras enzimas, debido a que se disponen a ambos lados de un aminoácido voluminoso como es el Trp. La presencia de una His en el bolsillo  $S_3$  contribuiría a hacerlo aún más angosto, permitiendo acomodar únicamente residuos pequeños. Sin embargo, el hecho de que el Trp ubicado a la entrada de los bolsillos  $S_2$  y  $S_3$  pueda rotar colocando su cadena lateral en una u otra dirección (Corvo et al. 2013), haría que residuos más voluminosos puedan ocupar uno de estos subsitios, siempre y cuando el residuo inmediato pueda acomodarse en el bolsillo adyacente, necesariamente restringido. La presencia de una Tyr67 en *FhCL2* y la catepsina K humana, hacen que el bolsillo  $S_2$  sea algo restringido, mientras que *FhCL1* y la catepsina L humana tienen un bolsillo más ancho delimitado por una Leu.

El análisis de HPLC-MS de los sitios de corte sobre la molécula de colágeno, es consistente con estas observaciones. Si bien el sitio  $S_3$  no muestra inicialmente selectividad, al



observar los 11 sitios donde *FhCL2* o *FhCL3* cortaron habiendo una Pro en P<sub>2</sub>, 8 contenían Gly en P<sub>3</sub>. Esto coincide con estudios previos de degradación de colágeno por la catepsina K (Kafienah et al. 1998) y la zingipaína GP11 (Kim et al. 2007), donde hallaron que de los 12 sitios de corte del colágeno I conteniendo Pro en P<sub>2</sub> (8 para la catepsina K y 4 para GP11), en 10 había una Gly en P<sub>3</sub>. Un elemento a tener en cuenta es que en estos ensayos se observan los péptidos obtenidos a tiempo final, donde se mezclan sitios de corte primarios, sobre la proteína nativa, y sitios secundarios que quedan expuestos al avanzar la proteólisis. La interrelación entre la capacidad de acomodar residuos en S<sub>2</sub>-S<sub>3</sub> también se observó para los oligopéptidos sintéticos ensayados, observando que *FhCL3* acomoda fácilmente Pro en P<sub>2</sub> sólo si está presente una glicina en P<sub>3</sub> (siendo  $k_{cat}/K_M$  60 veces mayor para Tos-GPR-AMC que para Boc-VPR-AMC). Las simulaciones moleculares dinámicas de unión a ligandos peptídicos, sustituyendo los aminoácidos variables de los sitios S<sub>1</sub>', S<sub>2</sub> y S<sub>3</sub> de *FhCL1* por los encontrados en *FhCL2* o *FhCL3* mostraron que estas últimas enzimas tienen baja afinidad de unión por el ligando AAPR\*NAA pero la afinidad se ve incrementada al sustituir la Ala por Gly en P<sub>3</sub> (ligando AGPR\*NAA). De estos estudios se desprendió también una base molecular plausible que podría en parte, dar cuenta de la mayor eficiencia de *FhCL3* para degradar el colágeno I con respecto a *FhCL2*. Se observó que *FhCL3* es capaz de unir estrechamente un ligando con hidroxiprolina en la posición P<sub>3</sub>, PPGP\*PPG (P representa hidroxiprolina) mientras en el caso de *FhCL2* el mismo ligando no logra acomodarse satisfactoriamente en su sitio activo (Robinson, Corvo et al. 2011). Nuestros resultados apuntan a que las modificaciones en los subsitios S<sub>2</sub> y S<sub>3</sub> de *FhCL2* y *FhCL3* han moldeado la especificidad de sustrato de estas dos enzimas generando selectividad diferencial sin comprometer su habilidad de degradar las moléculas de colágeno, de manera que ambas puedan contribuir a su digestión en los distintos estadios del parásito en que son secretadas. Las variaciones observadas en el sitio activo de las catepsinas de *Fasciola* son indicativas de la divergencia funcional y especialización durante la evolución de esta familia génica. Si bien la habilidad de acomodar una Pro en la posición P<sub>2</sub> del sustrato surgió de manera independiente en mamíferos, trematodos y plantas, la arquitectura del sitio activo que les confiere esa actividad resulta ser similar en todas ellas, un claro ejemplo de convergencia evolutiva.

Dado que las diferencias en la especificidad de sustrato de *FhCL1* y *FhCL3* se encuentran principalmente en los bolsillos S<sub>2</sub> y S<sub>3</sub>, decidimos sustituir mediante mutagénesis dirigida los residuos que son diferentes entre las enzimas para investigar su contribución a la especificidad de las catepsinas. Estudios previos reportaron un leve incremento de la especificidad hacia péptidos con Pro en P<sub>2</sub> en las variantes *FhCL1* L67Y (Stack et al. 2008), *FhCL5* L67Y (Smooker et al. 2000) y *FhCL5* L67W (Norbury et al. 2012). Nos sorprendió encontrar que la sustitución en la posición 67 genera efectos totalmente diferentes en *FhCL1* y *FhCL3*. Mientras el cambio de Leu a Trp en la posición 67 de *FhCL1* genera cambios sutiles en sus preferencias, como había sido reportado para *FhCL5* (leve disminución y aumento de su

afinidad hacia Leu y Pro, respectivamente, (Norbury et al. 2012), la sustitución del Trp 67 de *FhCL3* por Leu provoca una drástica alteración de la especificidad de la enzima, tanto en el subsitio  $S_2$  como en el  $S_3$ . La enzima mutante pierde su preferencia tanto por Pro en  $P_2$  como Gly en  $P_3$ , así como su capacidad de digerir el colágeno tipo I, pero sin adquirir las preferencias de sustrato de *FhCL1* (Corvo et al. 2013). Esto se debe fundamentalmente a una disminución de varios órdenes de magnitud en la  $k_{cat}$ , sugiriendo que la presencia del triptófano es necesaria para interactuar con el sustrato de forma de asegurar una orientación adecuada para la catálisis. Esta respuesta es diferente a lo observado para la sustitución del mismo aminoácido en las catepsinas L y K humanas, donde el intercambio del residuo 67 de una por otra, hace que se inviertan sus preferencias aminoacídicas en  $P_2$  (Lecaille, Choe, et al. 2002; Lecaille et al. 2007).

La sustitución L67Y (de *FhCL1* a *FhCL2*) disminuye la especificidad hacia el sustrato Z-VLK-AMC, y levemente incrementa la afinidad hacia Tos-GPR-AMC, en concordancia con lo reportado por C. M. Stack et al., 2008. La misma sustitución en *FhCL5*, sin embargo, provoca un incremento marcado en la especificidad hacia el mismo sustrato (Smooker et al. 2000). Se planteó que esta diferencia podría ser consecuencia de la presencia de una Leu en la posición 157 de *FhCL5* (Corvo et al. 2009; Stack et al. 2008). A favor de esta hipótesis se observa que en la variante de *FhCL1* con ambas sustituciones L69Y V157L disminuye aún más la especificidad hacia Z-VLK-AMC y se incrementa la afinidad por Tos-GPR-AMC como lo indica su menor valor de  $K_M$ . Por otro lado, la presencia de una Thr en la posición 158 es exclusiva de *FhCL2* y *FhCL3* entre todas las catepsinas conocidas, por lo que se hipotetizó que podría estar vinculada a la particular especificidad. Sin embargo, la contribución del residuo 158 a las interacciones en  $S_2$  es pequeña y no concuerda con las hipótesis planteadas. No obstante, analizando más en detalle el sitio activo de las enzimas, se observa que mientras el esqueleto del residuo de Thr contribuye a formar la pared del sitio  $S_2$ , la cadena lateral polar que posee el hidroxilo se orienta hacia el bolsillo  $S_1$ , por lo que podría estar influenciando las interacciones al otro lado del sitio de corte de la enzima. Como se ha mencionado, son necesarios otros enfoques con péptidos extendidos para analizar en detalle las interacciones del lado prima del sitio de corte (Hemerly et al. 2003; Juliano et al. 2004; Lanfranco et al. 2008; O'Donoghue et al. 2012; Portaro et al. 2000).

Si bien la estructura de las catepsinas y la conformación del sitio activo son similares en todas las enzimas de la familia, este resultado sugiere que existen diferencias más profundas en la arquitectura proteica que subyacen a estas observaciones. Es necesario profundizar en el estudio de estas diferencias, realizando análisis de estructura-actividad, para lo cual sería interesante obtener la estructura cristalográfica de la enzima. En este sentido, se podrían emplear diferentes estrategias para la cristalización, ya sea formando complejo con un inhibidor, como el E-64, o generando un mutante del sitio activo para impedir el autoprocésamiento a la forma madura como se hizo en el caso de pro*FhCL1* (Stack et al.

2008). Si bien nos hemos centrado en las diferencias en el sitio activo, por su obvia relevancia en la función, existen otros residuos que difieren entre estas proteínas, que podrían modular la actividad. Irving y cols. (Irving et al. 2003) identificaron varios residuos con selección positiva en la familia de las catepsinas de *Fasciola*, y si bien varios de ellos coinciden con el sitio activo, algunos mapean en otras regiones de la proteína. En este sentido, en un artículo reciente se aporta evidencia a favor de la existencia de los denominados “sectores” proteicos. Estos serían redes dispersas de aminoácidos contiguos espacialmente en la estructura de una proteína que coevolucionan, de modo que las mutaciones en residuos ubicados en un “sector” producirían cambios en la especificidad de unión de ligandos en las proteínas (McLaughlin Jr et al. 2012). Los residuos que constituyen el sector forman parte del sitio activo en muchos casos, pero también pueden localizarse en puntos distantes, siendo más difícil identificarlos como relevantes para la funcionalidad de la proteína. Sería interesante intentar identificar si existen sectores, y en caso afirmativo cuáles residuos forman parte de los mismos y si son o no compartidos por las diferentes catepsinas. Nuestros resultados sugieren la presencia de residuos clave fuera del bolsillo S<sub>2</sub> de estas enzimas, que contribuyen a delinear la arquitectura del sitio activo y por ende la especificidad de sustrato.

Nuestro trabajo apoya el papel importante que juega el residuo 67 del bolsillo S<sub>2</sub> de las catepsinas en la interacción y orientación del sustrato en una conformación adecuada para la catálisis enzimática. En las enzimas secretadas por el parásito adulto, la sustitución de este residuo provoca cambios sutiles en la especificidad de sustrato, mientras que en *FhCL3* altera drásticamente la enzima, disminuyendo su actividad hacia sustratos peptídicos y macromoleculares. Sería interesante evaluar si este cambio afecta a la estructura secundaria o la conformación de la enzima mediante técnicas espectroscópicas como dicroísmo circular, IR y fluorescencia, así como estudiar cuál o cuáles pasos de la catálisis se ven afectados al sustituir el Trp 67 en *FhCL3*. Por otra parte, la diferencia de especificidad observada ofrece una oportunidad para el desarrollo de inhibidores selectivos que exploten la particularidad de la enzima secretada por el estadio invasivo.

## **VI. APLICACIONES Y PERSPECTIVAS**

## 1. Aplicaciones en la prevención, tratamiento y diagnóstico de la fascioliasis

Las cisteína proteasas han sido evaluadas como blancos para el desarrollo de vacunas y el diseño de drogas para el control de la fascioliasis, con resultados que sugieren que podrían usarse en la formulación de una vacuna efectiva contra esta parasitosis. En *F. hepatica* se ha observado que por separado las catepsinas L1 y L2 nativas provocan una disminución de alrededor del 34% en la carga parasitaria de ovejas, mientras que si se administran juntas la reducción alcanza el 60% (Bentancor et al. 2002). Un estudio más reciente empleando como antígeno de vacunación proFhCL1 recombinante expresada en *P. pastoris*, reporta una reducción de la carga parasitaria del 48% en vacunos inmunizados en relación con el grupo control (Golden et al. 2010). Ensayos de vacunación en ratas con una fracción antigénica de JRD conteniendo proteínas de peso molecular entre 12-35 KDa mostraron ser fuertemente inductores de protección contra la infección por *F. hepatica*. Mediante *western blot* con sueros de ratas vacunadas se encontró una proteína de 32 KDa como altamente inmunogénica, la cual se identificó como la proenzima de la catepsina L3 (van Milligen et al. 2000). Los ensayos de vacunación en ratas con proFhCL3 recombinante expresada en baculovirus mostraron una protección parcial (52%) frente a la infección experimental con *Fasciola*, sin embargo la proenzima expresada en *S. cerevisiae* no mostró potencial inmunoprotector (Reszka et al. 2005). A su vez reportan que se obtuvo una reducción en el nivel de protección con la enzima de baculovirus al emplear la enzima desglucosilada, indicando que el patrón de glicosilación de la catepsina influye en su capacidad antigénica. Dado que la adición de distintos azúcares depende del sistema de expresión de la proteína recombinante, no podemos suponer *a priori* cómo será el comportamiento de FhCL3 producida en nuestro sistema. Por consiguiente, sería interesante evaluar su potencial inmunoprotector, sobre todo en una vacuna en combinación con catepsinas del adulto u otras enzimas hidrolíticas, como la LAP, que ya ha demostrado un elevado nivel de protección (Acosta et al. 2008; Maggioli et al. 2011). En este sentido, otros grupos han evaluado en ratones el uso de más de una catepsina en una vacuna polivalente (una catepsina B y FhCL5 de *F. hepatica*, y una catepsina L de juveniles de *F. gigantica*), mostrando que los niveles de protección mejoran al combinar antígenos secretados por diferentes estadios (Smooker et al. 2010).

Por otro lado, en el área diagnóstica, varios trabajos reportan el uso de las catepsinas FhCL1 y FhCL2 de *F. hepatica* como antígenos para la inmuno-detección del ganado infectado (Cornelissen et al. 2001; Kuerpick et al 2013). Los anticuerpos específicos en sueros de vacas infectadas se detectan a partir de 4 semanas post-infección (pi), obteniendo una sensibilidad del 100% a partir de la semana 10 pi (Kuerpick et al. 2013). El empleo de una enzima secretada por el parásito juvenil podría contribuir al desarrollo de metodologías que permitan detectar las infecciones en forma precoz, o aumentar la sensibilidad de los ensayos. Además, la marcada especificidad de sustrato de FhCL3 hace de ella un blanco interesante para el

desarrollo de inhibidores específicos abriendo otras posibilidades al control parasitario. Las diazometilcetonas peptídicas inhiben específicamente las cisteína proteasas y no son activas frente a otras clases de proteasas como las serina o metaloproteasas, ni frente a los grupos tiol en general. Además, es posible utilizarlas *in vivo* para inactivar proteasas por ejemplo en cultivos celulares, ya que a las dosis que inhiben eficazmente a las proteasas no son tóxicas para las células (Green & Shaw 1981; Shaw & Green 1981). La acil-oximetilcetona Ac-HGPR es un inhibidor peptídico similar a las peptidil diazometilcetonas, (con prolina y glicina en P<sub>2</sub> y P<sub>3</sub>, respectivamente), selectivo e irreversible para la catepsina K humana, que no afecta a las catepsinas L y V, aunque también inhibe a la catepsina B (Choe et al. 2006). Sería interesante plantearse el desarrollo de inhibidores del tipo diazometilcetonas peptídicas específicos para *FhCL3*, que exploten su especificidad de sustrato diferencial con respecto a las otras catepsinas del parásito y a las catepsinas del hospedador. Estos inhibidores tendrían utilidad en el laboratorio para realizar estudios funcionales de la enzima, pero también presentan un uso potencial como complemento de drogas antihelmínticas que ya se utilizan para el tratamiento del ganado, o para la prevención de la infección en seres humanos que vivan en regiones endémicas o desarrollen actividades que impliquen un alto riesgo de infección. El principal desafío es generar una droga selectiva para la enzima parasitaria que no afecte al gran número de proteasas homólogas del hospedero con funciones esenciales para el organismo. En este aspecto, el conocimiento profundo de la bioquímica y estructura de las enzimas parasitarias y su comparación con las enzimas de mamíferos permite avanzar en el desarrollo de compuestos que sean inhibidores selectivos.

## 2. Aplicaciones biotecnológicas y biomédicas

La capacidad de degradar el colágeno que posee *FhCL3* abre perspectivas de uso biotecnológico de esta enzima. Son varias las aplicaciones industriales donde se usan colagenasas, como por ejemplo en la industria de limpieza de cueros, tiernizado de carnes o cosmetología, y farmacéutica donde se usan colagenasas en la formulación de dentífricos y cremas cicatrizantes entre otros productos (Kathleen et al 2005; Watanabe 2004).

En el área biomédica, se ha planteado el uso potencial de parásitos helmintos o moléculas producidas por estos para la prevención y/o tratamiento de desórdenes pro o auto-inflamatorios (como el shock séptico, rechazo de trasplantes o enfermedades autoinmunes). Esto se debe a la demostrada capacidad de estos parásitos de modular el sistema inmune de manera de volcarlo hacia una respuesta de tipo anti-inflamatoria y reguladora (Th2/Treg) (Tilp et al. 2013). En un trabajo reciente se demostró que la administración intraperitoneal de PES de *F. hepatica* disminuye de 85% a 15% la aparición de diabetes autoinmune en ratones diabéticos no-obesos, así como la capacidad de *FhCL1*, junto a otras 2 moléculas secretadas

por *F. hepatica*, de modular el sistema inmune del ratón hacia respuestas tipo Th2 (Robinson et al. 2013).

La expresión aberrante de las catepsinas se ha asociado a numerosas patologías humanas entre las que se encuentran la artritis reumatoidea, picnodisostosis (catepsina K), el síndrome de Papillon-Lefevre (catepsina C), varios desórdenes neurológicos, la osteoporosis y el cáncer (catepsinas B, F, L, S). La sobreexpresión, amplificación génica y expresión de diferentes variantes de splicing de genes de catepsinas ha sido asociada a la iniciación y progresión del cáncer. Se ha sugerido que esta expresión alterada podría participar en la degradación descontrolada de componentes de matriz extracelular y membranas basales facilitando la invasión y metástasis de células tumorales, postulándose también su participación en la proliferación de células malignas y la angiogénesis (Joyce et al. 2004; Mohamed & Sloane 2006; Turk et al 2001; Vasiljeva et al. 2007; Vasiljeva & Turk 2008). Desde esa perspectiva, el conocimiento del papel de las catepsinas en la invasión parasitaria puede contribuir a comparar mecanismos de invasión mediados por enzimas similares en procesos como el cáncer y las enfermedades degenerativas.

## VII. REFERENCIAS BIBLIOGRÁFICAS



- Acosta, D. 1991. "Epidemiología y control de *Fasciola hepatica* en Uruguay." Pp. 233–236 in *Enfermedades parasitarias de importancia económica en bovinos: Bases epidemiológicas para su prevención y control en Argentina y Uruguay*, edited by A Nari and C Fiel. Editorial Agropecuaria Hemisferio Sur.
- Acosta, D., Canela, M., Piacenza, L., Roche, L., Carmona, C., Tort, J. 2008. "*Fasciola hepatica* leucine aminopeptidase, a promising candidate for vaccination against ruminant fasciolosis." *Molecular and biochemical parasitology* 158(1):52–64.
- Alcala-Canto, Y., Ibarra-Velarde, F., Sumano-Lopez, H., Gracia-Mora, J., Alberti-Navarro, A. 2007. "Effect of a cysteine protease inhibitor on *Fasciola hepatica* (liver fluke) fecundity, egg viability, parasite burden, and size in experimentally infected sheep." *Parasitology research* 100(3):461–5.
- Alcala-Canto, Y., Ibarra-Velarde, F., Sumano-Lopez, H., Gracia-Mora, J., Alberti-Navarro, A. 2006. "Dose-response inhibition of proteolytic activity by a cysteine protease inhibitor in a murine model of fasciolosis." *Parasitology research* 98(5):438–42.
- Andrews, S. J. 1999. "The life cycle of *Fasciola hepatica*." Pp. 1–29 in *Fasciolosis*, edited by John P Dalton. Wallingford: CABI Publishing.
- Atkinson, H. J., Babbitt, P. C., Sajid, M. 2009. "The global cysteine peptidase landscape in parasites." *Trends in parasitology* 25(12):573–81.
- Bentancor, A., Piacenza, L., Carmona, C. 2002. "Immunization with cathepsin L proteinases CL1 and CL2 secreted by *Fasciola hepatica* elicit a preferential type 1 response based on IgG2a antibodies in rats." *Journal of Helminthology* 76(03):199–205.
- Berasaín, P., Carmona, C., Frangione, B., Dalton, J. P., Goñi, F. 2000. "*Fasciola hepatica*: parasite-secreted proteinases degrade all human IgG subclasses: determination of the specific cleavage sites and identification of the immunoglobulin fragments produced." *Experimental parasitology* 94(2):99–110.
- Berasaín, P., Goñi, F., McGonigle, S., Dowd, A., Dalton, J P., Frangione, B., Carmona, C. 1997. "Proteinases Secreted by *Fasciola hepatica* Degrade Extracellular Matrix and Basement Membrane Components." *Reactions* 83(1):1–5.
- Brennan, G. P., Fairweather, I., Trudgett, A., Hoey, E., McCoy, McConville, M., Meaney, M., Robinson, M., McFerran, N., Ryan, L., Lanusse, C., Mottier, L., Alvarez, L., Solana, H., Virkel, G., Brophy, P. M. 2007. "Understanding triclabendazole resistance." *Experimental and molecular pathology* 82(2):104–9.
- Campetella, O., Henriksson, J., Aslund, L., Frasc, A.C.C., Pettersson, U., Cazzulo, J. J. 1992. "The major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* is encoded by multiple polymorphic tandemly organized genes located on different chromosomes." *Molecular and biochemical parasitology* 50(2):225–34.

- Cancela, M., Acosta, D., Rinaldi, G., Silva, E., Duran, R., Roche, L., Zaha, A., Carmona, C., Tort, J. F. 2008. "A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*." *Biochimie* 90(10):1461–1475.
- Cancela, M., Ruétalo, N., Dell'Oca, N., da Silva, E., Smircich, P., Rinaldi, G., Roche, R., Carmona, C., Alvarez-Valín, F., Zaha, A., Tort, J. F. 2010. "Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*." *BMC genomics* 11:227.
- Cappetta, M., Roth, I., Díaz, A., Tort, J. F., Roche, L. 2002. "Role of the prosegment of *Fasciola hepatica* cathepsin L1 in folding of the catalytic domain." *Biological chemistry* 383(7-8):1215–21.
- Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogoyo, M., Bromme, D., Ellman, J. A., Craik, C. S. 2006. "Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities." *The Journal of biological chemistry* 281(18):12824–32.
- Choi, K. H., Laursen, R. A., Allen, K. N. 1999. "The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*." *Biochemistry* 38(36):11624–33.
- Cornelissen, J. B., Gaasenbeeka, C. P. H., Borgsteede, F. H. M., Holland, W. G., Harmsen, M. M., Boersma, W. J. A. 2001. "Early immunodiagnosis of fasciolosis in ruminants using recombinant *Fasciola hepatica* cathepsin L-like protease." *International journal for parasitology* 31(7):728–37.
- Corvo, I., O'Donoghue, A., Pastro, L., Di-Penis, N., Eroy-Reveles, A., . 2013. "Dissecting the active site of the collagenolytic cathepsin L3 protease of the invasive stage of *Fasciola hepatica*." *PLoS neglected tropical diseases* *Accepted*.
- Corvo, I., Cancela, M., Cappetta, M., Pi-Denis, N., Tort, J. F., Roche, L. 2009. "The major cathepsin L secreted by the invasive juvenile *Fasciola hepatica* prefers proline in the S2 subsite and can cleave collagen." *Molecular and biochemical parasitology* 167(1):41–7.
- Dalton, J. P., McGonigle, S., Rolph, T. P., Andrews, S. J. 1996. "Induction of protective immunity in cattle against infection with *Fasciola hepatica* by vaccination with cathepsin L proteinases and with hemoglobin." *Infection and immunity* 64(12):5066–74.
- Dalton, J. P. & Heffernan, M. 1989. "Thiol proteases released in vitro by *Fasciola hepatica*." *Molecular and biochemical parasitology* 35:161–166.
- Dalton, J. P., Skelly, P., Halton, D. W. 2004. "Role of the tegument and gut in nutrient uptake by parasitic platyhelminths." *Can. J. Zool.* 82: 211-232.
- Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K.C., Dvorak, J., Hsieh, I., Bahgat, M., Dissous, C., McKerrow, J. H. 2006. "A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite." *The Journal of biological chemistry* 281(51):39316–29.
- Donnelly, S., O'Neill, S. M., Stack, C. M., Robinson, M. W., Turnbull, L., Whitchurch, C., Dalton, J. P.. 2010. "Helminth cysteine proteases inhibit TRIF-dependent activation of

- macrophages via degradation of TLR3." *The Journal of biological chemistry* 285(5):3383–92.
- Dowd, A. J., Dooley, M., Fagain, C., Dalton, J. P. 2000. "Stability studies on the cathepsin L proteinase of the helminth parasite, *Fasciola hepatica*." *Enzyme Microb Technol* 27:599–604.
- Dowd, A. J., Tort, J. F., Roche, L., Ryan, T., Dalton, J. P. 1997. "Isolation of a cDNA encoding *Fasciola hepatica* cathepsin L2 and functional expression in *Saccharomyces cerevisiae*." *Molecular and biochemical parasitology* 88:163–174.
- Dowd, A. J., Smith, M., McGonigle, S., Dalton, J. P. 1994. "Purification and characterisation of a second cathepsin L proteinase secreted by the parasitic trematode *Fasciola hepatica*." *European journal of biochemistry / FEBS* 223(1):91–8.
- 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., Gowen, M. 1996. "Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts." *Journal of Biological Chemistry* 271:12511.12516.
- Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M., Wolthers, B. G. 1968. "Structure of papain." *Nature* 218:929–932.
- Dzik, J. M. 2006. "Molecules released by helminth parasites involved in host colonization." *Acta biochimica Polonica* 53(1):33–64.
- Faber, K. N., Haima, P., Harder, W., Veenhuis, M., Ab, G. 1994. "Highly-efficient electrotransformation of the yeast *Hansenula polymorpha*." *Current genetics* 25(4):305–10.
- Fairweather, I., Threadgold, L. T., Hanna, R. E. 1999. "Development of *Fasciola hepatica* in the mammalian host." Pp. 47–111 in *Fasciolosis*, edited by J P Dalton. CAB International.
- Garnero, P., Bore, O., Byrjalsen, I., Ferreras, M., Drakei, F. H., McQueneyi, M. S., Foged, M. T., Delmas, P. D., Delaisse, J.. 1998. "The collagenolytic activity of cathepsin K is unique among mammalian proteinases." *The Journal of biological chemistry* 273(48):32347–52.
- Gelse, K. 2003. "Collagens—structure, function, and biosynthesis." *Advanced Drug Delivery Reviews* 55(12):1531–1546.
- Gillmor, S. A., Craik, C. S., Fletterick, R. J. 1997. "Structural determinants of specificity in the cysteine protease cruzain." *Protein Science* 6:1603–1611.
- Golden, O., Flynn, R.J., Reada, C., Sekiyaa, M., Donnelly, S. M., Stack, C., Dalton, J. P., Mulcahy, G. 2010. "Protection of cattle against a natural infection of *Fasciola hepatica* by vaccination with recombinant cathepsin L1 (rFhCL1)." *Vaccine* 28(34):5551–7.
- Grams, R., Vichasri-Grams, S., Sobhorn, P., Upatham, E. S., Viyanant, V. 2001. "Molecular cloning and characterization of cathepsin L encoding genes from *Fasciola gigantica*." *Parasitol Int* 50:105–114.

- Green, G. D. J., & Shaw, E. 1981. "Peptidyl diazomethyl ketones are specific inactivators of thiol proteinases." *Journal of Biological Chemistry* 256(4):1923–1928.
- Halton, D. W. 1997. "Nutritional Adaptations to Parasitism within the Platyhelminthes." *International Journal for Parasitology* 27(6):693–704.
- Hanna, R. E. B. 1980. "*Fasciola hepatica*: An inmunofluorescent Study of Antigenic Changes in the Tegument during Development in the Rat and the Sheep." *Experimental parasitology* 170:155–170.
- Harmsen, M. M., Cornelissen, J. B.W.J., Buijs, H. E.C.M. Boersma, W. J.A., Jeurissen, S. H.M., van Milligen, F. J. 2004. "Identification of a novel *Fasciola hepatica* cathepsin L protease containing protective epitopes within the propeptide." *International journal for parasitology* 34(6):675–82.
- Harris, J. L., Backes, B.J., Leonetti, F., Mahrus, S., Ellman, J. A., Craik, C. S.. 2000. "Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries." *Proceedings of the National Academy of Sciences of the United States of America* 97(14):7754–9.
- Hemerly, J. P., Oliveira, V., Del Nery, E., Morty, R. E., Andrews, N. W., Juliano, M. A., Juliano, L. 2003. "Subsite specificity (S3, S2, S1', S2' and S3') of oligopeptidase B from *Trypanosoma cruzi* and *Trypanosoma brucei* using fluorescent quenched peptides: comparative study and identification of specific carboxypeptidase activity." *The Biochemical journal* 373(Pt 3):933–9.
- Hernández-González, A., Valero, M. L., Sánchez del Pino, M., Oleaga, A., Siles-Lucas, M. 2010. "Proteomic analysis of *in vitro* newly excysted juveniles from *Fasciola hepatica*." *Molecular and biochemical parasitology* 172(2):121–8.
- Heussler, V. T., & Dobbelaere, D. A. E. 1994. "Cloning of a protease family of *Fasciola hepatica* by the polymerase chain reaction." *Molecular and biochemical parasitology* 64:11–23.
- Howell, R M. 1966. "Collagenase activity of immature *Fasciola hepatica*." *Nature* 209:713–714.
- Irving, J. A., Spithill, T. W., Pike, R. N., Whisstock, J. C., Smooker, P. M. 2003. "The evolution of enzyme specificity in *Fasciola spp.*" *Journal of molecular evolution* 57(1):1–15.
- Izquierdo-Martin, M., & Stein, R. L. 1992. "Transition-State Structural Features for the Association of Metalloproteases with Phosphorous-Containing Inhibitors." *J. Am. Chem. Soc.* 114:1527–1528.
- Jefferies, J. R., Campbell, A. M., van Rossum, A. J., Barrett, J., Brophy, P. M. 2001. "Proteomic analysis of *Fasciola hepatica* excretory-secretory products." *Proteomics* 9:1128–1132.
- Joyce, J. A., Baruch, A., Chehade, K., Meyer-Morse, N., Giraudo, E., Tsai, F., Greenbaum, D. C., Hager, J. H., Bogoyo, M., Hanahan, D. 2004. "Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis." *Cancer cell* 5(5):443–53.
- Juliano, M. A., Brooks, D. R., Selzer, P. M., Pandolfo, H. L., Judice, W. A. S., Juliano, L., Meldal, M., Sanderson, S. J., Mottram, J. C., Coombs, G. H. 2004. "Differences in

- substrate specificities between cysteine protease CPB isoforms of *Leishmania mexicana* are mediated by a few amino acid changes." *European journal of biochemistry / FEBS* 271(18):3704–14.
- Kafienah, W., Bromme, D., Buttle, D. J., Croucher, L. J., Hollander, A. P. 1998. "Human cathepsin K cleaves native type I and II collagens at the N-terminal end of the triple helix." *October* 732:727–732.
- Karrer, K. M., Peiffert, S. L., Ditomas, M. E. 1993. "Two distinct gene subfamilies within the family of cysteine protease genes Biochemistry." *Biochemistry* 90(April):3063–3067.
- Kathleen, N., Riley, B. S., Herman, I. M. 2005. "Collagenase promotes the cellular responses to injury and wound healing in vivo." *Journal of Burns and Wounds* 4:e1–8.
- Keiser, J. & Utzinger, J. 2009. "Food-Borne Trematodiasis." *Clinical Microbiology Reviews* 22(3):466–483.
- Kim, M., Hamilton, S. E., Guddat, L. W., Overall, C. M. 2007. "Plant collagenase: unique collagenolytic activity of cysteine proteases from ginger." *Biochimica et biophysica acta* 1770(12):1627–35.
- Kirschke, H. & Shaw, E. 1981. "Rapid interaction of cathepsin L by Z-Phe-PheCHN12 and Z-Phe-AlaCHN2." *Biochemical and biophysical research communications* 101(2):454–458.
- Kuerpick, B., Schnieder, T., Strube, C. 2013. "Evaluation of a recombinant cathepsin L1 ELISA and comparison with the Pourquier and ES ELISA for the detection of antibodies against *Fasciola hepatica*." *Veterinary parasitology* 193(1-3):206–13.
- Lanfranco, M. F., Loayza-Muroa, R., Clark, D., Núñez, R., Zavaleta, A. I., Jimenez, M., Meldal, M., Coombs, G. H., Mottram, J. C., Izidorof, M., Juliano, M. A., Juliano, L., Arévalo, J. 2008. "Expression and substrate specificity of a recombinant cysteine proteinase B of *Leishmania braziliensis*." *Molecular and biochemical parasitology* 161(2):91–100.
- Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S., Brömme, D. 2002. "Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity." *Biochemistry* 41(26):8447–54.
- Lecaille, F., Chowdhury, S., Purisima, E., Brömme, D., Lalmanach, G. 2007. "The S2 subsites of cathepsins K and L and their contribution to collagen degradation." *Protein science: a publication of the Protein Society* 16(4):662–70.
- Lecaille, F., Kaleta, J., Brömme, D. 2002. "Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design." *Chemical reviews* 102(12):4459–88.
- Li, Z., Yasuda, Y., Li, W., Bogoy, M., Katz, N., Gordon, R. E., Fields, G. B., Brömme, D. 2004. "Regulation of Collagenase Activities of Human Cathepsins by Glycosaminoglycans." *Journal of Biological Chemistry* 279(7):5470 –5479.
- Li, Z., Hou, W., Escalante-Torres, C. R., Gel, B. D., Brömme, D. 2002. "Collagenase Activity of Cathepsin K Depends on Complex Formation with Chondroitin Sulfate." *Journal of Biological Chemistry* 277(32):28669–28676.

- Li, Z., Kienetz, M., Cherney, M. M., James, M. N. G., Brömme, D. 2008. "The Crystal and Molecular Structures of a Cathepsin K: Chondroitin Sulfate Complex." *Journal of Molecular Biology* 383(1):78–91.
- Lowther, J., Robinson, M. W., Donnelly, S. M., Xu, W., Stack, C. M., Matthews, J. M., Dalton, J. P. 2009. "The importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine protease." *PLoS Neglected tropical diseases* 3(1):e369.
- Maggioli, G. Acosta, D., Silveira, F., Rossi, S., Giacaman, S., Basika, T., Gayo, V., Rosadilla, D., Roche, L., Tort, J. F., Carmona, C. 2011. "The recombinant gut-associated M17 leucine aminopeptidase in combination with different adjuvants confers a high level of protection against *Fasciola hepatica* infection in sheep." *Vaccine* 29(48):9057–63.
- Mas-Coma, S., Bargues, M. D., Valero, M. A. 2005. "Fascioliasis and other plant-borne trematode zoonoses." *International journal for parasitology* 35(11-12):1255–78.
- McGonigle, L., Mousley, A., Marks, N. J., Brennan, G. P., Dalton, J. P., Spithill, T. W., Day, T. A., Maule, A. G. 2008. "The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration." *International journal for parasitology* 38(2):149–55.
- McGrath, M. E., Klaus, J. L., Barnes, M. G., Brömme, D. 1997. "Crystal structure of human cathepsin K complex with a potent inhibitor." *Nature Structural Biology* 4(2):109–111.
- McLaughlin Jr, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S., Ranganathan, R. 2012. "The spatial architecture of protein function and adaptation." *Nature* 491:138–142.
- McManus, D. P. & Dalton, J. P. 2006. "Vaccines against the zoonotic trematodes *Schistosoma japonicum*, *Fasciola hepatica* and *Fasciola gigantica*." *Parasitology* 133:S43–S61.
- McVeigh, P., Maule, A. G., Dalton, J. P., Robinson, M. W. 2011. "*Fasciola hepatica* virulence-associated cysteine peptidases: a systems biology perspective." *Microbes and infection / Institut Pasteur*.
- van Milligen, F. J., Cornelissen, J. B., Bokhout, B. A. 2000. "*Fasciola hepatica*: an antigen fraction derived from newly excysted juveniles, containing an immunoreactive 32-kDa protein, induces strong protective immunity in rats." *Experimental parasitology* 94(3):163–71.
- Mohamed, M. M. & Sloane, B. F. 2006. "Cysteine cathepsins: multifunctional enzymes in cancer." *Nature reviews. Cancer* 6(10):764–75.
- Morphew, R. M., Wright, H. A., LaCourse, E. J., Woods, D. J., Brophy, P. M. 2007. "Comparative proteomics of excretory-secretory proteins released by the liver fluke *Fasciola hepatica* in sheep host bile and during in vitro culture ex host." *Molecular & cellular proteomics : MCP* 6(6):963–72.
- Morphew, R. M., Wright, H. A., LaCourse, E. J., Porter, J., Barrett, J., Woods, D. J., Brophy, P. M. 2011. "Towards Delineating Functions within the *Fasciola* Secreted Cathepsin L Protease Family by Integrating In Vivo Based Sub-Proteomics and Phylogenetics" edited by John Pius Dalton. *PLoS Neglected Tropical Diseases* 5(1):e937.

- Mulcahy, G. O'Connor, F., McGonigle, S., Dowd, A., Clery, D. G., Andrews, S. J., Dalton, J. P. 1998. "Correlation of specific antibody titre and avidity with protection in cattle immunized against *Fasciola hepatica*." *Vaccine* 16(9–10):932–939.
- Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., Bode, W. 1991. "The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity." *EMBO Journal* 10:2321–2330.
- Na, B., Kim, S., Lee, E., Kim, T., Bae, Y., Kang, I., Yu, J., Sohn, W., Cho, S., Kong, Y. 2006. "Critical roles for excretory – secretory cysteine proteases during tissue invasion of *Paragonimus westermani* newly excysted metacercariae." *Cellular Microbiology* 8:1034–1046.
- Norbury, L. J., Hung, A., Beckham, S., Pike, R. N., Spithill, T. W., Craik, C. S., Choe, Y., Fecondo, J. V., Smooker, P. M. 2012. "Analysis of *Fasciola* cathepsin L5 by S 2 subsite substitutions and determination of the P1-P4 specificity reveals an unusual preference." *Biochimie* 94(5):1119–1127.
- O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B., Greninger, A. L., Hostetter, D. R., Qu, G., Maltby, D. A., Anderson, M. O., DeRisi, J. L., McKerrow, J. H., Burlingame, A. L., Craik, C. S. 2012. "Global identification of peptidase specificity by multiplex substrate profiling." *Nature Methods* 1–9.
- Parker-Manuel, S. J., Ivens, A. C., Dillon, G. P., Wilson, R. A. 2011. "Gene Expression Patterns in Larval *Schistosoma mansoni* Associated with Infection of the Mammalian Host." *Gene Expression* 5(8):e1274, 1–19.
- Pauly, T., Sulea, T., Ammirati, M., Sivaraman, J., Danley, D. E., Griffor, M. C., Kamath, A. V., Wang, I. K., Laird, E. R., Seddon, A. P., Menard, R., Cygler, M., Rath, V. L. 2003. "Specificity determinants of human cathepsin s revealed by crystal structures of complexes." *Biochemistry* 42(11):3203–13.
- Piacenza, L., Acosta, D., Basmadjian, I., Dalton, J. P., Carmona, C. 1999. "Vaccination with cathepsin L proteinases and with leucine aminopeptidase induces high levels of protection against fascioliasis in sheep." *Infection and immunity* 67(4):1954–1961.
- Portaro, F. C. V. Santos, A. B. F., Cezari, M. H. S., Juliano, M. A., Juliano, L., Carmona, E. 2000. "Probing the specificity of cysteine proteinases at subsites remote from the active site: analysis of P4, P3, P2' and P3' variations in extended substrates." *Biochemical Journal* 347:123–129.
- Prowse, R. K., Chaplin, P., Robinson, H. C., Spithill, T. W. 2002. "*Fasciola hepatica* cathepsin L suppresses sheep lymphocyte proliferation in vitro and modulates surface CD4 expression on human and ovine T cells." *Parasite immunology* 24(2):57–66.
- Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., Barrett, A. J. 2008. "MEROPS: the peptidase database." *Nucleic acids research* 36(Database issue):D320–5.
- Reszka, N., Cornelissen, J. B.W.J., Harmsen, M. M., Bienkowska-Szewczyk, K., de Bree, J., Boersma, W. J., Rijsewijk, F. A. M. 2005. "*Fasciola hepatica* procathepsin L3 protein

- expressed by a baculovirus recombinant can partly protect rats against fasciolosis." *Vaccine* 23(23):2987–93.
- Robinson, M. W., Tort, J. F., Lowther, J., Donnelly, S. M., Wong, E., Xu, W., Stack, C. M., Padula, M., Herbert, B., Dalton, J. P. 2008. "Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors." *Molecular & cellular proteomics: MCP* 7(6):1111–23.
- Robinson, M. W., Corvo, I., Jones, P. M., George, A. M., Padula, M. P., To, J., Cancela, M., Rinaldi, G., Tort, J. F., Roche, L., Dalton, J. P. 2011. "Collagenolytic activities of the major secreted cathepsin L peptidases involved in the virulence of the helminth pathogen, *Fasciola hepatica*." *PLoS neglected tropical diseases* 5(4):e1012.
- Robinson, M. W. & Dalton, J. P. 2009. "Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiasis." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364(1530):2763–76.
- Robinson, M. W., Dalton, J. P., Donnelly, S. 2008. "Helminth pathogen cathepsin proteases: it's a family affair." *Trends in biochemical sciences* 33(12):601–8.
- Robinson, M. W., Dalton, J. P., O'Brien, B. A., Donnelly, S. 2013. "*Fasciola hepatica*: the therapeutic potential of a worm secretome." *International journal for parasitology* 43(3-4):283–91.
- Robinson, M. W., Menon, R., Donnelly, S., Dalton, J. P., Ranganathan, S. 2009. "An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host." *Molecular & cellular proteomics: MCP* 8(8):1891–907.
- Roche, L., Dowd, A. J., Tort, J. F., McGonigle, S., McSweeney, A., Curley, G. P., Ryan, T., Dalton, J. P. 1997. "Functional expression of *Fasciola hepatica* cathepsin L1 in *Saccharomyces cerevisiae*." *European journal of biochemistry / FEBS* 245(2):373–80.
- Ruettger, A., Schueler, S., Mollenhauer, J. A., Wiederanders, B. 2008. "Cathepsins B, K, and L are regulated by a defined collagen type II peptide via activation of classical protein kinase C and p38 MAP kinase in articular chondrocytes." *The Journal of biological chemistry* 283(2):1043–51.
- Rune, S. J. 1973. "pH in human duodenum: its physiological and pathophysiological significance." *Digestion* 8(3):261–268.
- Rzychon, M., Chmiel, D., Stec-Niemczyk, J. 2004. "Modes of inhibition of cysteine proteases." *Acta biochimica Polonica* 51(4):861–73.
- Sambrook, J., Fritsch, E., Maniatis, T. 1989. *Molecular Cloning: a laboratory manual*. 2nd Editio. Cold Spring Harbor Laboratory.
- Schechter & Berger, A. 1967. "On the size of the active site in proteases, I papain." *Biochemical and biophysical research communications* 27(2):157–162.



- Schwede, T., Kopp, J., Guex, N., Peitsch, M. C. 2003. "SWISS-MODEL: an automated protein homology-modelling server." *Nucleic acids research* 13:3381–3385.
- Selent, J., Kaleta, J., Li, Z., Lalmanach, G., Brömme, D. 2007. "Selective Inhibition of the Collagenase Activity of Cathepsin K." *Journal of Biological Chemistry* 282(22):16492–16501.
- Shaw, E. & Green, G. D. J. 1981. "Inactivation of thiol proteases with peptidyl diazomethyl ketones." *Methods in Enzimology* 80:82–826.
- Smith, A. M., Dowd, A. J., McGonigle, S., Keegan, P. S., Brennan, G., Trudgett, A., Dalton, J. P. 1993. "Purification of a cathepsin L proteinase secreted by adult *Fasciola hepatica*." *Molecular and biochemical parasitology* 62:1–8.
- Smith, S. M. & Gottesman, M. M. 1989. "Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*." *Journal of Biological Chemistry* 264:20847–20495.
- Smith, A. M., Dowd, A. J., Heffernan, M., Robertson, C. D., Dalton, J. P. 1993. "*Fasciola hepatica*: a secreted cathepsin L-like proteinase cleaves host immunoglobulin." *International journal for parasitology* 23(8):977–83.
- Smooker, P. M., Whisstock, J. C., Irving, J. A., Siyaguna, S., Spithill, T. W., Pike, R. N. 2000. "A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*." *Protein science : a publication of the Protein Society* 9(12):2567–72.
- Smooker, P. M., Jayaraj, R., Pike, R. N., Spithill, T. W. 2010. "Cathepsin B proteases of flukes: the key to facilitating parasite control?" *Trends in parasitology* 26(10):506–14.
- Stack, C. M., Dalton, J. P., Robinson, M. W. 2011. "The phylogeny, structure and function of trematode cysteine proteases, with particular emphasis on the *Fasciola hepatica* cathepsin L family." Pp. 116–135 in *Cysteine proteases of pathogenic organisms.*, vol. Chapter 8, edited by John P Dalton and Mark W Robinson. Landes Biosciences, Springer.
- Stack, C. M., Caffrey, C. R., Donnelly, S. M., Seshadri, A., Lowther, J., Tort, J. F., Collins, P.R., Robinson, M. W., Xu, W., McKerrow, J. H., Craik, C. S., Geiger, S. R., Marion, R., Brinen, L. S., Dalton, J. P. 2008. "Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*." *The Journal of biological chemistry* 283(15):9896–908.
- Stein, C., Kuchenmuller, T., Hendrickx, S., Pruss, U., Wolfson, L., Engels, D., Schlundt, J. 2007. "The Global Burden of Disease Assessments — WHO Is Responsible?" *PLoS Neglected Tropical Diseases* 1(3):e161 1–8.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. 2011. "MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance and Maximum Parsimony Methods." *Mol. Biol. Evol.* 28(10):2731–2739.
- Tao, K., Stearns, N. A., Dong, J., We, Q. L., Sahagian, G. G. 1994. "The proregion of cathepsin L is required for proper folding, stability, and ER exit." *Archives of Biochemistry and Biophysics* 311:19–27.

- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G. 1997. "The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." *Nucleic acids research* 25(24):4876–4882.
- Tilp, C., Kapur, V., Loging, W., Erb, K. J. 2013. "Prerequisites for the pharmaceutical industry to develop and commercialise helminths and helminth-derived product therapy." *International journal for parasitology* 43(3-4):319–25.
- Tkalcevic, J., Ashman, K., Meeusen, E. 1995. "*Fasciola hepatica*: rapid identification of newly excysted juvenile proteins." *Biochemical and biophysical research communications* 213(1):169–174.
- Turk, B., Turk, D., Turk, V. 2000. "Lysosomal cysteine proteases: more than scavengers." *Biochimica et biophysica acta* 1477(1-2):98–111.
- Turk, D., Guncar, G., Podobnik, M., Turk, B. 1998. "Revised Definition of Substrate Binding Sites of Papain-Like Cysteine Proteases." *Biological Chemistry* 379:137 –147.
- Turk, Ī., Turk, V., Turk, B. 2001. "Lysosomal cysteine proteases: facts and opportunities." *EMBO Journal* 20(17):4629–4633.
- Tyndall, J. D. A., Nall, T., Fairlie, D. P. 2005. "Proteases universally recognize beta strands in their active sites." *Chemical reviews* 105(3):973–99.
- Vasiljeva, O., Reinheckel, T., Peters, C., Turk, D., Turk, V., Turk, B. 2007. "Emerging roles of cysteine cathepsins in disease and their potential as drug targets." *Current pharmaceutical design* 13(4):387–403.
- Vasiljeva, O. & Turk, B. 2008. "Dual contrasting roles of cysteine cathepsins in cancer progression: apoptosis versus tumour invasion." *Biochimie* 90(2):380–6.
- Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, R., Storer, A. C., Thomas, D. Y. 1995. "Processing of the Papain Precursor." *The Journal of biological chemistry* 270(18):10838–10846.
- Watanabe, K. 2004. "Collagenolytic proteases from bacteria." *Appl. Microbiol. Biotechnol* 63:520–526.
- Young, N. D., Hall, R. S., Jex, A. R., Cantacessi, C., Gasser, R. B. 2010. "Elucidating the transcriptome of *Fasciola hepatica* - a key to fundamental and biotechnological discoveries for a neglected parasite." *Biotechnology advances* 28(2):222–31.

## **VIII. ANEXOS**

**ANEXO 1. Oligonucleótidos empleados en la mutagénesis del bolsillo S<sub>2</sub> de *FhCL1* hacia *FhCL2*.**

Enzima	Plásmido molde	Oligonucleótidos	Secuencia
<i>FhCL1</i> L67Y	<i>FhCL1</i>	<i>FhCL1</i> L67Y Fw	GGTTGCGGTGGTGGAT <b>AT</b> ATGGAAAATGCTTACC
		<i>FhCL1</i> L67Y Rv	GGTAAGCATTTCAT <b>AT</b> ATCCACCACCGCAACC
<i>FhCL1</i> V157L	<i>FhCL1</i>	<i>FhCL1</i> V157L Fw	G TTCACCGCTTCGT <b>TTG</b> AACCATGCAGTCTTGG
		<i>FhCL1</i> V157L Rv	CCAAGACTGCATGGTT <b>CAA</b> ACGAAGCGGTGAAC
<i>FhCL1</i> N158T	<i>FhCL1</i>	<i>FhCL1</i> N158T Fw	CACCGCTTCGTGTG <b>ACC</b> CATGCAGTCTTGG
		<i>FhCL1</i> N158T Rv	CCAAGACTGCATG <b>GGT</b> CACACGAAGCGGTG
<i>FhCL1</i> L67Y V157L	<i>FhCL1</i> L67Y	<i>FhCL1</i> V157L Fw & Rv	
<i>FhCL1</i> L67Y N158T	<i>FhCL1</i> L67Y	<i>FhCL1</i> N158T Fw & Rv	
<i>FhCL1</i> V157L N158T	<i>FhCL1</i>	<i>FhCL1</i> V157L N158T Fw	G TTCACCGCTTCGT <b>TTGACC</b> CATGCAGTCTTGGCTG
		<i>FhCL1</i> V157L N158T Rv	CAGCCAAGACTGCATG <b>GGTCAA</b> ACGAAGCGGTGAAC
<i>FhCL1</i> L67Y V157L N158T	<i>FhCL1</i> V157L N158T	<i>FhCL1</i> L67Y Fw & Rv	

## ANEXO 2. Frecuencia de aminoácidos en el colágeno I

Utilizando la herramienta ProtParam de ExPASy (<http://web.expasy.org/protparam/>) se calcula el porcentaje de cada aminoácido en las cadenas de colágeno alfa1 y alfa 2 bovino y humano. El contenido es muy similar al comparar cada cadena y entre las dos especies, por lo que en la gráfica superior el porcentaje (%) de cada aminoácido (aa) se estimó como un promedio entre ambas cadenas en las 2 especies.

