Title	An atypical and functionally diverse family of Kunitz-type cysteine/serine proteinase inhibitors secreted by the helminth parasite Fasciola hepatica.
Author(s)	Smith, David; Cwiklinski, Krystyna; Jewhurst, Heather; Tikhonova, Irina G.; Dalton, John P.
Publication Date	2020-11-26
Publication Information	Smith, David, Cwiklinski, Krystyna, Jewhurst, Heather, Tikhonova, Irina G., & Dalton, John P. (2020). An atypical and functionally diverse family of Kunitz-type cysteine/serine proteinase inhibitors secreted by the helminth parasite Fasciola hepatica. Scientific Reports, 10(1), 20657. doi:10.1038/s41598-020-77687-7
Publisher	Nature Research
Link to publisher's version	https://doi.org/10.1038/s41598-020-77687-7
Item record	http://hdl.handle.net/10379/16350
DOI	http://dx.doi.org/10.1038/s41598-020-77687-7

1 An atypical and functionally diverse family of Kunitz-type

2 cysteine/serine proteinase inhibitors secreted by the helminth

3 parasite *Fasciola hepatica*

4

5 David Smith[1,2], Krystyna Cwiklinski[1,3]*, Heather Jewhurst[1,3], Irina G. Tikhonova[4], John P.

6 Dalton[1,3]

7 Author affiliations:

8 1. School of Biological Sciences, Queen's University Belfast BT9 7BL, Belfast, Northern

9 Ireland, United Kingdom.

10 2. Present address: Moredun Research Institute, Pentland Science Park, Penicuik, Midlothian,

11 Scotland, United Kingdom.

12 3. Centre of One Health and Ryan Institute, School of Natural Sciences, National University

13 of Ireland Galway, Galway, Ireland.

14 4. School of Pharmacy, Medical Biology Centre, Queen's University Belfast, Belfast BT9

15 7BL, Northern Ireland, United Kingdom.

16

17 *corresponding author

18 Email addresses:

19 David Smith: D.Smith@moredun.ac.uk

20   Krystyna Cwiklinski: krystyna.cwiklinski@nuigalway.ie

21   Heather Jewhurst: heather.jewhurst@nuigalway.ie

22   Irina G. Tikhonova: i.tikhonova@qub.ac.uk

23   John P. Dalton: johnpius.dalton@nuigalway.ie

24

25   **ABSTRACT**

26   *Fasciola hepatica* is a global parasite of humans and their livestock. Regulation of parasite-

27   secreted cathepsin L-like cysteine proteases associated with virulence is important to fine-

28   tune parasite-host interaction. We uncovered a family of seven Kunitz-type (FhKT) inhibitors

29   dispersed into five phylogenetic groups. The most highly expressed FhKT genes (group

30   FhKT1) are secreted by the newly excysted juveniles (NEJs), the stage responsible for host

31   infection. The FhKT1 inhibitors do not inhibit serine proteases but are potent inhibitors of

32   parasite cathepsins L and host lysosomal cathepsin L, S and K cysteine proteases (inhibition

33   constants <10 nM). Their unusual inhibitory properties are due to (a) Leu^{15} in the reactive site

34   loop P1 position that sits at the water-exposed interface of the S1 and S1' subsites of the

35   cathepsin protease, and (b) Arg^{19} which forms cation-π interactions with Trp^{291} of the S1'

36   subsite and electrostatic interactions with Asp^{125} of the S2' subsite. FhKT1.3 is exceptional,

37   however, as it also inhibits the serine protease trypsin due to replacement of the P1 Leu^{15} in

38   the reactive loop with Arg^{15}. The atypical Kunitz-type inhibitor family likely regulate parasite

39   cathepsin L proteases and/or impairs host immune cell activation by blocking lysosomal

40   cathepsin proteases involved in antigen processing and presentation.

41

44

BACKGROUND

Fasciola hepatica, the causal agent of fasciolosis, has the greatest geographical distribution and exhibits one of the broadest mammalian host ranges of all helminth (worm) parasites [1]. As a result, this disease afflicts millions of humans and their livestock on every inhabited continent [2]. Part of the parasite's success can be attributed to its ability to secrete a rich source of proteins that aid its invasion of the host, penetration and feeding of tissues, as well as counteracting and downplaying host immune responses [2-4]. The best characterised are the abundantly expressed and secreted cathepsins L and B proteases that have 23 and 11 members, respectively, with overlapping and distinct substrate and macromolecular specificities [5]. Together, these secreted cysteine proteases create a formidable digestive cocktail that allows the parasite to efficiently and rapidly tunnel through host intestinal and liver tissues during its migration to the bile ducts [5,6].

Regulation of the secretory protease activity is essential for parasites and is primarily achieved by the co-secretion of protease inhibitors [7-9]. In our search for protease inhibitors in *F. hepatica* secretions we discovered a small molecular-sized protein (6 kDa) that possessed all the structural features of Kunitz-type (FhKT1) inhibitors required to inhibit serine proteases [10]. To our surprise, however, the FhKT1 was unusual amongst KT inhibitors, as it exhibited no inhibitory activity against a wide range of standard serine proteases, including trypsin, chymotrypsin, elastase and various blood-clotting factors. Unexpectedly, FhKT1 possessed specificity and potent activity ($Ki < 0.1$ nM) against *F. hepatica* cathepsin L cysteine proteases as well as cathepsins L and cathepsins K of mammals and thus represented a specific evolutionary adaptation in their function never described previously for a KT inhibitor/I2 family inhibitor.

By interrogating the *F. hepatica* genome, transcriptome and proteome [3, 4, 11] we have uncovered a wider family of KT inhibitor genes containing five groupings or clades

70 (termed FhKT1 – FhKT5) that are differentially expressed according to the parasite's

71 development in the host. The FhKT1 group, containing three members, are the most highly

72 expressed, particularly within the early infective stages of the parasite and the only members

73 secreted by the parasite, acting at the host-parasite interface. The FhKT1 clade are associated

74 with the secretory cells of the parasite gut reproductive structures of mature adult parasites.

75 Functional expression of recombinant FhKT1 together with mutagenesis studies show that a

76 leucine present at position 15 (P1^{15}) within the RSL is critical to defining the exclusive

77 cysteine protease specificity of FhKT1.1, while arginine at this position confers FhKT1.3

78 with the ability to inhibit cysteine proteases and the serine protease trypsin. Using structural

79 modelling we also determined that position 19 (P4') is important in stabilising the RSL

80 within the cysteine protease active site. Our studies provide important insights on how

81 helminth parasites have a capacity to create novel molecules with unique and varied

82 biological activities important in host-parasite interaction by the process of gene duplication

83 and positive selection.

84

85 **RESULTS**

86 *F. hepatica* **Kunitz-type inhibitors are encoded by a multigene family**

87 By interrogating the *F. hepatica* genome we discovered a family of KT inhibitor genes

88 consisting of seven members. Phylogenetic analysis showed that three of these genes

89 clustered closely together, here termed the *fhkt1* group (*fhkt1.1, fhkt1.2 and fhkt1.3*), with the

90 remaining individual genes forming distinct branches, termed *fhkt2*, *fhkt3*, *fhkt4* and *fhkt5*

91 (Fig.1A). Analysis of the genomic organisation of these genes shows that *fhkt1.1. fhkt1.2,*

92 *fhkt1.3* and *fhkt5* are present on the same draft *F. hepatica* genome scaffold, with *fhkt2, fhkt3*

93 and *fhkt4* present on separate draft genome scaffolds. With the exception of *fhkt1.3*, all genes

94 are comprised of two exons, representing the signal peptide and KT domain, respectively,

95   separated by an intron of varying length (ranging from 500 bp to > 10 kbp; Supplementary

96   Fig. S1). The *fhkt1.3* sequence lacks the first signal peptide-encoding exon. Sequence

97   analysis of 20 cDNAs amplified from adult *F. hepatica* cDNA using a forward primer

98   encoding a consensus FhKT1 signal peptide sequence and a reverse primer corresponding to

99   the conserved KT domain did not find any *fhkt1.3* gene. This supports the idea that this gene

100   lacks a sequence associated with the first exon and, therefore, is deprived of a signal peptide.

101

102   Sequence alignments showed that six conserved cysteine residues, which form three

103   characteristic disulphide bridges, are preserved in all seven FhKT proteins (Fig.1B). Low

104   primary sequence identity was observed between the sequences (21.1%; 12 residues) but of

105   particular note was the significant variability within the P1-P4' reactive loop region that is

106   responsible for inhibitory function by KTs (Fig. 1C) that indicates important functional

107   diversity and adaptation within the family of inhibitors. Strikingly, FhKT1.1, FhKT1.2,

108   FhKT2 and FhKT5 possess a P1 leucine residue that we have shown is critical for the unique

109   inhibition of cysteine proteases [10], while FhKT1.3 and FhKT4 possess a P1 arginine

110   residue, which is more typical of classical KT inhibitors that inhibit trypsin-like and

111   chymotrypsin-like serine proteases [12-14].

112

113   Interrogation of the genomes available for related important human and animal parasites

114   revealed a surprisingly high number of diverse KTs that separate into seven groups (although

115   smaller clusters exist within these groupings) based on maximum likelihood phylogenetic

116   analysis (Supplementary Fig. S2, Supplementary Table S1). The functional diversity of the

117   trematode KT inhibitors is further highlighted by the range of P1-P4' reactive loop region

118   sequences (Supplementary Table S1). In particular, this analysis demonstrated that the

119   majority of sequences in Group A and C possess a P1 leucine residue, while Group B is

120   dominated by sequences with a P1 arginine residue.

121

122   Surprisingly, despite the presence of a large number of expanded families of KTs from a

123   diverse range of trematodes in Group D, F and G, these did not contain any sequences from

124   *F. hepatica* (Supplementary Fig. S2). The *fhkt1* (*fhkt1.1, fhkt1.2 and fhkt1.3*) and *fhkt2*

125   sequences were found within Group A with KTs from *Fasciola gigantica* and *Echinostoma*

126   *caproni* but there were no comparable sequences found within the bile dwelling flukes,

127   *Clornorchis sinensis* and *Opisthorchis viverrini*, or the blood flukes, *Schistosoma*

128   *haematobium*, *Schistosoma japonicum* and *Schistosoma mansoni*. The *fhkt3* gene clusters

129   within Group B with sequences from *E. caproni, C. sinensis* and *Schistosoma* species, while

130   *fhkt4* and *fhkt5* genes are found in Group E with sequences from *E. caproni, C. sinensis* and

131   *O. viverrini*.

132

133   **The *F. hepatica* KT inhibitor family members are under strict control of temporal**

134   **expression and secretion**

135   Of the seven *F. hepatica* KT genes, the *fhkt1* group (*fhkt1.1, fhkt1.2 and fhkt1.3*) were found

136   to be the most highly transcribed KT inhibitors in *F. hepatica* (Fig.2A). Moreover, all three

137   members exhibited a similar pattern of temporal transcription through the different

138   development stages within the host and it was noteworthy that all were significantly

139   transcribed at higher levels at 24h post-excystment and also within the mature adult parasites,

140   relative to the metacercariae (Fig. 2A; Supplementary Fig. S3). Quantitative gene expression

141   analysis (qPCR), performed on the infectious encysted metacercariae and NEJs over a time-

142   course of 48h post-excystment, revealed a marked rise in *fhkt1* transcription, particularly at

143   ~10h post-excystment, a time when the parasite is traversing the host gut wall (Fig. 2B).

144

145 The *fhkt4* gene also exhibited a high level of transcription within the metacercariae and NEJ

146 1h, 3h and 24h stages but in contrast to the *fhkt1* group, this gene is not expressed by the 21-

147 day old immature parasites that migrate within the liver tissue and mature adult parasites that

148 reside in the bile ducts (Fig.2A). By contrast, transcription of *fhkt2* and *fhkt5* are not detected

149 in the infectious NEJ stages but are required by the later stage parasites; *fhkt2* is up-regulated

150 in mature adult flukes while *fhkt5* appears earlier in 21-day old as well as mature adult

151 parasites, though at lower levels of transcription (Fig.2A). Statistically relevant levels of

152 transcription of *fhkt3* were not detected at any stage suggesting that *fhkt3* is either expressed

153 in lifecycle stages not associated with the mammalian host (e.g. during invasion and

154 migration in the intermediate snail host) or, perhaps less likely, is a pseudogene.

155

156 Analysis of the proteomic profiles of the secretomes of various stages of the parasite

157 previously reported by us [4, 11] revealed that the FhKT1 group, FhKT1.1, FhKT1.2 and

158 FhKT1.3, are the only members of the family that are secreted extra-corporeally by the

159 parasite (Supplementary Table S2). Consistent with the higher levels of transcription within

160 the 24h NEJ and the adult parasites, these inhibitors were detected in the secretomes of the

161 NEJ parasites (1h, 3h and 24h post-excystment) and mature adult parasites. Furthermore,

162 FhKT1.1 and FhKT1.2 were also detected amongst the cargo of proteins contained within the

163 extracellular vesicles (EVs) prepared from the excreted secreted (ES) products of adult flukes

164 (Supplementary Table S2, and [11]). Although FhKT1.3 lacks a signal peptide it was also

165 found in the parasite secretions, which indicates that its delivery into the host is mediated by

166 a non-classical pathway as observed for other *F. hepatica* proteins such as glutathione-S

167 transferase [3,4,6].

168

Tissue-specific expression of KT inhibitors in juvenile and adult parasites

169

170 FhKT1 was immunolocalised to the NEJ parenchymal cell bodies and within the large

171 bifurcated gut (Fig.3B-D) which also expresses the major cysteine protease *F. hepatica*

172 cathepsin L3 (FhCL3) and cathepsins B (FhCB). By altering the plane of visualisation to

173 view the surface of the NEJ parasites, we observed a complex network of thin channels never

174 described before on the underside of the NEJ surface tegument that connected the

175 parenchyma cells. The intensity of anti-FhKT1 labelling within these intracellular channels

176 was particularly abundant in the 24h and 48h NEJs (Fig.3F-H, black arrowheads). The

177 channels are not associated with the musculature nor do they have any obvious connection

178 with the oral or ventral sucker; rather, the labelling suggests that they create a transport

179 system between the parenchymal cell bodies. As a negative control, NEJ samples were also

180 probed with pre-immune sera (Fig.3A, 3E).

181

182 FhKT1 was also highly expressed in the adult parasite's reproductive organs, particularly the

183 vitelline glands, which are important for egg development, and in eggs within the ovaries

184 (Supplementary Fig. S4). Antibody binding was also observed within the parenchyma and

185 gut wall (Supplementary Fig. S5).

186

187 **Inhibition profile of the FhKT1 members and importance of P1[15] residue in**

188 **determining specificity**

189 Recombinant FhKT1.1 (rFhKT1.1) was produced as a functionally active recombinant

190 protein and isolated to homogeneity by NTA-agarose affinity chromatography (Fig.4A;

191 Supplementary Fig. S7). rFhKT1.1 was previously shown to be a potent inhibitor of two

192 major cysteine proteases secreted by adult *F. hepatica*, FhCL1 and FhCL2, although the

193 inhibition constant, K_i, is about 25-fold more for the latter enzyme (Table 1, [10]). In the

present study, we found that rFhKT1.1 is also a potent inhibitor of the NEJ-specific cathepsin

protease, FhCL3, an enzyme that has a unique activity and can digest interstitial matrix

proteins such as collagen; a reduction of activity of 93.1 % ($\pm$3.66) was observed at 2 μM

inhibitor and we determined a K_i value of 1.8 nM ($\pm$0.6) (Fig.4B, Table 1). rFhKT1.1 is also

a potent inhibitor of the human cathepsin L-like cysteine proteases human cathepsin L and K,

with similar potencies to that observed for the parasite proteases (Table 1). By contrast,

rFhKT1.1 exhibits no activity against a range of serine protease including trypsin and

chymotrypsin (Fig. 4B, Table 1) and kallikrein, thrombin, plasmin and elastase (data not

shown).

Table 1. Inhibition constants (K_i) for the FhKT recombinant protein against cysteine and

serine proteases inhibited.

Enzyme	Inhibition Constant (K_i), in nM			
	rFhKT1*	rFhKT1Leu^{15}/Arg^{15}	rFhKT1.3	rFhKT1Arg^{19}/Ala^{19}
Cysteine Proteases				
F. hepatica Cathepsin L1	0.4 ($\pm$ 0.1)*	0.7 ($\pm$ 0.04)*	0.6 ($\pm$ 0.2)	4.2 ($\pm$ 0.4)
F. hepatica Cathepsin L2	10 ($\pm$ 0.3)*	27 ($\pm$ 1.5)*	10.6 ($\pm$ 0.3)	35 ($\pm$ 3.9)
F. hepatica Cathepsin L3	1.8 ($\pm$ 0.6)	3.6 ($\pm$ 0.3)	3.2 ($\pm$ 0.2)	32.8 ($\pm$ 7)
Human Cathepsin L	1.6 ($\pm$ 0.1)*	3 ($\pm$ 0.1)*	2.6 ($\pm$ 0.5)	0.3 ($\pm$ 0.004)
Human Cathepsin K	5 ($\pm$ 0.3)*	5 ($\pm$ 0.3)*	5.5 ($\pm$ 0.3)	18.5 ($\pm$ 2.9)
Serine Protease				
Bovine Trypsin	N.I.	1.5 ($\pm$ 0.7)*	1.8 ($\pm$ 0.2)	N.I.

*As reported by Smith et al. [10]

N.I: Not inhibited

209   The residue at the P1 site (position 15) is centrally poised within the reactive loop of KT

210   inhibitors and is critical for binding to and inhibiting the target protease. Sequence alignment

211   analysis shows that the P1 position in two members of the *F. hepatica* FhKT1 group,

212   FhKT1.1 and FhKT1.2, is occupied with a hydrophobic leucine residue (see Fig.1C). By

213   comparison, FhKT1.3 possesses a positively charged arginine in this position. We therefore

214   produced a recombinant form of this inhibitor, rFhKT1.3, and demonstrated that it was also a

215   potent inhibitor of the *F. hepatica* cathepsin L proteases rFhCL1, rFhCL2, rFhCL3, with K_i

216   of 0.6, 10.6 and 3.2 nM, respectively. Likewise, rFhKT1.3 was also a potent inhibitor of

217   human cathepsins K and L with K_i of 2.6 and 5.5 nM, respectively. However, unlike

218   FhKT1.1 and FhKT1.2, rFhKT1.3 showed potent inhibitory activity against the serine

219   protease trypsin with a reduction in activity of 99.23% (±0.36) at 2 μM and a K_i of 1.8 nM

220   (±0.2). rFhKT1.3 did not inhibit other serine proteases tested, such as chymotrypsin (Fig.4B,

221   Table 1).

222

223   This inhibitory profile of FhKT1.3 (i.e. potent activity against both cysteine and serine

224   proteases) is comparable to that exhibited by a variant of FhKT1.1 whereby the PI Leu^{15} was

225   purposely substituted with an P1 Arg^{15} to produce the recombinant rFhKT1.1Leu^{15}/Arg^{15} (see

226   Table 1; [10]). This data emphasises the importance of the amino acid residue at the P1 site

227   of the reactive loop in inhibition specificity.

228

229   **Arg^{19} in the C-terminus of the reactive loop of FhKT1 is important for binding to**

230   **cysteine proteases**

231   Homology models of FhKT1.1 docked to the crystal structure of FhCL1 (PDB code: 2O6X;

232   [10]) (see Fig. 5A and 5B) were used to assess the shape and electrostatic interactions that

233   take place between the two molecules. These predicted that an Arg residue situated at P4' at

234   the C-terminal end of the reactive loop (residue 19, see Fig. 1C) forms cation-π interactions

235   with Trp^{291} of the S1' subsite and electrostatic interactions with Asp^{125} of the S2' subsite of

236   the active site of the FhCL1 cysteine protease. To investigate the role of this residue in

237   cysteine and serine protease binding of FhKT1 we produced a recombinant variant inhibitor

238   whereby the positively charged arginine was replaced with a neutral alanine residue

239   (rFhKT1.1Arg^{19}/Ala^{19}; Fig.5A). The purified recombinant protein rFhKT1.1Arg^{19}/Ala^{19} was

240   shown to be a potent inhibitor of the *F. hepatica* cysteine proteases at 2 μM. However, K_i

241   values of 4.2, 35 and 32.8 nM against rFhCL1, rFhCL2 and rFhCL3, respectively,

242   demonstrated that this substitution reduced the potency of the inhibitor by 10-, 3.5- and 18-

243   fold, respectively, compared to wild-type FhKT1 (Table 1). The K_i value of

244   rFhKT1.1Arg^{19}/Ala^{19} against human cathepsin K was also increased, 3.5-fold compared to

245   wild-type FhKT1.1 but, surprisingly, the K_i for human cathepsin L was reduced 5-fold and

246   therefore binding was improved. Like wild-type FhKT1.1, rFhKT1.1Arg^{19}/Ala^{19} showed no

247   activity against bovine trypsin (Table 1) or other serine proteases examined (data not shown).

248

249   **FhKT1.1 and FhKT1.3 are potent inhibitors of native *F. hepatica* somatic and secreted**

250   **cathepsin L cysteine proteases**

251   Adult *F. hepatica* parasites express abundant cathepsin B and cathepsin L cysteine proteases

252   and the major enzymes, FhCL1, FhCL2 and FhCL5, are also excreted/secreted into the

253   culture medium in which the parasites are maintained. We examined whether recombinant

254   FhKT1.1 and FhKT1.3 could inhibit these enzymes in somatic extract and ES products.

255   Inhibition curves showed that cysteine protease activity in somatic extracts was inhibited by

256   FhKT1.1 and FhKT1.3 but this was not absolute, even when the inhibitors where added to the

257   extract at a 1 μM concentration (~20% activity remained, Fig.6A). Total cysteine protease

258   activity in the somatic extract, however, was completely inhibited by the broad-spectrum

259   cysteine protease inhibitor E-64 (at concentrations above 250 nM; Fig 6A). On the other

260   hand, FhKT1.1 and FhKT1.3 completely inhibited cysteine protease activity within the ES

261   products at concentrations of ~250 nM and above (Fig.6B). We can explain this finding by

262   our previous data showing that FhKT1.1 inhibits cathepsin L activity but not cathepsin B

263   activity [10]; while cathepsin B cysteine proteases are present within the parasite somatic

264   extracts of adult *F. hepatica,* they are not abundantly secreted by the parasite into culture

265   medium [15-18].

266

267   **Competition assays show that FhKT1.1 binding to cathepsin L is not blocked by small-**

268   **molecule inhibitors of cysteine proteases that occupy S1, S2 and S3 subsites.**

269   To gain insight into the mechanism by which FhKT1.1 inhibits cysteine proteases we

270   performed competition studies with small-molecule broad-spectrum inhibitors of cathepsin-

271   like proteases, namely E-64 and Z-Phe-Ala-CHN_2 [19, 20] (Fig. 7; Supplementary Fig. S7).

272   These assays involved first mixing the small compound at varying concentrations with adult

273   *F. hepatica* ES products containing native cysteine proteases before adding the rFhKT1.1

274   inhibitor. The complex was subsequently pulled down using NTA-beads and then analysed

275   by LDS-PAGE. Addition of E-64 or Z-Phe-Ala-CHN_2 to the ES products, even at excess

276   concentrations of 100 μM, did not prevent the binding of rFhKT1.1 to the native secreted

277   cysteine proteases (Fig.7A-D). This data demonstrates that when small molecular-sized

278   inhibitor compounds occupy the S1 and S2 subsite of the active site, FhKT1.1 can still bind

279   to the active site groove of the cysteine proteases (Fig. 5A and 5C). By comparison,

280   competition assays using recombinant human cystatin C, a cysteine protease inhibitor of 13.3

281   kDa, showed that this prevented the binding of FhKT1.1 to the active site groove (Fig. 7C).

282   Similar observations were made with rFhKT1Arg^{19}/Ala^{19} using Z-Phe-Ala-CHN_2

283   (Supplementary Fig. S6).

284

285

DISCUSSION

From infection via the intestine as a newly excysted juvenile to establishment as a mature

parasite in the bile ducts the helminth parasite *F. hepatica* embarks on a migratory path that

requires the penetration and degradation of various tissues. Digestion of the various

macromolecules encountered during this journey is accomplished by the controlled secretion

of distinct members of a complex family of cathepsin-cysteine proteases. Control of these

proteases to prevent excessive damage to parasite and host implies the involvement of

inhibitors that regulate their activity and prevent excessive hydrolysis, which is often coined

the protease/anti-protease balance [21, 22].

In this study, we found that *F. hepatica* expresses a family of Kunitz-type inhibitors

consisting of seven members. Based on phylogenetic analysis, three of these genes, *fhkt1.1*,

1.2 and *1.3*, form a single gene cluster while the four remaining genes are distinct, namely

fhkt2, *fhkt3*, *fhkt4* and *fhkt5*. Gene expression analysis using RNA-Seq data revealed that

fhkt1.1, *1.2* and *1.3* are highly expressed at all intra-mammalian stages examined, from the

NEJs that initiate infection by penetrating the gut wall to the bile duct-dwelling mature adult

worms. qPCR showed that within 6 to 10 h after the NEJ parasites emerged from their cysts,

expression of *fhkt1.1*, *1.2* and *1.3* are rapidly up-regulated. Furthermore, proteomic analysis

of NEJ and adult worm secretions (ES products) and EVs detected peptides that match only

proteins derived from the FhKT1.1, 1.2 and 1.3 gene products. Collectively, these

observations suggest that the FhKT1 group of Kunitz-protease inhibitors are the most

dominant KT inhibitors in the parasite stages examined and imply their importance in the

interaction with the mammalian host, particularly during the early invasive process.

309

310 The *fhkt4* gene is expressed primarily in the metacercariae and *in vitro*-cultured NEJs but is

311 markedly downregulated in parasites that have migrated and matured in the liver parenchyma

312 (21 days after infection). The encoded FhKT4 protein differs to the FhKT1 group of kunitz-

313 protease inhibitors as it has a P1 Arg and a P1' Ala, which predicts that this inhibitor may

314 inhibit trypsin (see below). As trypsin is a major digestive serine protease in the intestine, it

315 could be harmful to the parasite as it excysts in the duodenum and begins migration across

316 the intestinal wall. In *Fasciola gigantica*, the FhKT4 homolog, is abundantly transcribed by

317 the cercarial stage within the snail [23], indicating a possible role for the inhibition of trypsin-

318 like proteases that are released by snails with cercariae, as shown in *Schistosoma mansoni*

319 [24]. Therefore, FhKT4 may be specialised for anti-trypsin defence both within the intestine

320 and in the snail intermediate.

321

322 Although, products of the *fhkt2* and *fhkt5* genes were not found in our proteomic analysis of

323 parasite somatic and secreted proteins, transcriptomic data showed that their expression in the

324 parasite life cycle is strictly regulated. They are both up-regulated in the migrating liver stage

325 21-day old parasites and in the bile duct-dwelling adult worms. Their expression, therefore, is

326 associated with the tissue and blood-feeding life stages of *F. hepatica* and thus we propose

327 functions in anti-coagulation, similar to that suggested for KT inhibitors of the trematodes *S.*

328 *japonicum* and *S. mansoni* [25, 26]. No transcript data was identified for the *fhkt3* gene which

329 suggests it is either (a) expressed in life stages not associated with the mammalian host i.e.

330 the intermediate snail host, or (b) is redundant and not expressed in *F. hepatica*.

331

332 Confocal immunolocalisation studies using antibodies prepared against recombinant FhKT1

333 showed that the FhKT1 group members were expressed in the gut lumen of the NEJ and

334   within distinct parenchymal cell bodies. They are also observed in narrow channels that form

335   a network throughout the parasite and penetrate the underside of the surface tegument (but do

336   not appear to protrude into the tegument). Therefore, FhKT1 could be trafficked via these

337   parenchymal cell bodies to different sites within the parasite, most predominantly to the

338   digestive tract. From here FhKT1 proteins may be secreted by the parasite, explaining the

339   presence of FhKT1.1, 1.2 and 1.3 peptides within NEJ ES products, and delivered into host

340   cells and/or tissues. It is worth noting that because of the close identity in sequence/structure

341   of the three FhKT1 members, as well as FhKT4, the polyclonal antibodies employed in our

342   studies likely bind epitopes in all four inhibitors and thus the pattern of tissue localisation

343   represents a composite of these proteins. Thus, different Kunitz-type inhibitors could be

344   expressed in the cell bodies, parenchymal tissue and the digestive tract.

345

346   In adult parasites, FhKT1 group members are predominantly associated with structures of the

347   female reproductive system including vitelline cells within the vitelline glands, S1 secretory

348   cells of the Mehlis gland and within the vitelline of the eggs. While two of the major

349   cathepsin L proteases expressed by adult *F. hepatica*, FhCL1 and FhCL2, are associated with

350   the digestive tract and have not been localized in vitelline cells or parasite eggs [5, 27],

351   another adult-associated cysteine protease, FhCL5, with high activity at physiological pH was

352   suggested by Norbury et al. [28] to function in the vitelline tissues. In *S. mansoni* and *S.*

353   *japonicum*, KT inhibitors were observed between the eggshell and developing miracidia,

354   corresponding to the vitelline mass [25, 26].

355

356   Our biochemical investigations of *F. hepatica* KT inhibitors focused on the most abundant

357   and secreted members, FhKT1.1, FhKT1.2 and FhKT1.3. While the RSL of FhKT1.1 and

358   FhKT1.2 contained identical sequences, the FhKT1.3 differed in having an Arg at the P1

359 position rather than a Leu. Consistent with our earlier reports, FhKT1.1 (and by extension

360 FhKT1.2) exclusively inhibited cysteine proteases; these included the cathepsin L proteases

361 from *F. hepatica,* FhCL1, FhCL2 and FhCL3, as well as host-derived mammalian cysteine

362 proteases cathepsin L and cathepsin K (cathepsin Bs were not inhibited). FhKT1.3 also

363 inhibited these cysteine proteases but also exhibited potent inhibitory activity against the

364 serine protease trypsin. By substituting the P1 Leu of FhKT1.1 to an Arg

365 (rFhKT1Leu^{15}/Arg^{15}) we demonstrated that this variant displayed a similar inhibitory profile

366 to FhKT1.3 and thereby established the importance of P1 Arg for trypsin inhibition. FhKT1.3

367 serine protease inhibition is not broad-spectrum, however, as it did not inhibit chymotrypsin,

368 kallikrein, and thrombin. The secreted FhKT1.3 could function alongside the intracellular

369 FhKT4 to provide parasite defence against proteolytic attack within the duodenum, similar to

370 trypsin-inhibiting KT proteins from other helminth parasites [29-33].

371

372 Bozas et al. [34] first described a Kunitz-type inhibitor in extracts of adult *F. hepatica*,

373 termed Fh-KTM. Our analysis of the mass spectrometry data reported in that paper found

374 peptides matching all the FhKT1 proteins (FhKT1.1, 1.2 and 1.3), consistent with our studies

375 showing that these are the most abundant KTs in the parasite. Immunolocalisation studies

376 [34] showing that the inhibitors are dispersed throughout the parenchyma of adult worms,

377 with evidence of trafficking towards the gut, are also consistent with our studies. Bozas et al.

378 [34] reported that native Fh-KTM, isolated from soluble extracts of adult parasites reduced

379 the activity of trypsin by 93%. Since we have shown here that *F. hepatica* expresses a family

380 of KT inhibitors, this soluble extract likely contained a mix of more than one of these KT

381 inhibitors. However, given that FhKT1.1 and 1.2 are not inhibitors of trypsin, the trypsin

382 inhibition recorded by Bozas et al. [34] must be attributed to FhKT1.3.

383

384   The specific adaptation of FhKTs to cysteine proteases suggest that these inhibitors function

385   in the regulation of the *F. hepatica* cathepsin L cysteine proteases. *F. hepatica* cysteine

386   proteases are expressed as inactive 37 kDa zymogens that activate by auto-catalytic removal

387   of an inhibitory 12 kDa propeptide to become a 25 kDa active mature form [35-37]. The need

388   for an additional regulation of these proteases suggests the importance of (a) preventing

389   uncontrolled, or leaked, auto-activation and/or (b) strict control over the proteolytic activity

390   of the cathepsin proteases following activation and secretion. FhKT1 could be viewed as a

391   "threshold inhibitor" as previously described for other regulatory inhibitors [38,39]. In this

392   scenario, the 'threshold inhibitor' co-localises with its cognate proteases, usually at lower

393   concentration, so that they can prevent undesirable premature activation. However, when

394   inhibitory potential is over-run upon bulk activation of the target protease auto-catalytic

395   activation takes place [38,39]. The *F. hepatica* cathepsin L proteases are abundant in the

396   gastrodermal cells of the parasite as well as amongst the cargo within extracellular vesicles

397   (EVs), microenvironments where FhKTs are also found [11, 15, 16, 27, 40-43]. Regulation of

398   cathepsin L cysteine proteases activity ensures the majority of cathepsin L remains in an

399   inactive form until activation is necessary, for example, upon secretion into the parasite gut

400   lumen or release of EVs into host tissue and cells. Indeed, studies by Muiño et al. [44]

401   showed that a Kunitz type inhibitor co-purifies with a mature cathepsin L secreted by adult

402   worms in culture, implying that protein-protein interactions occur between the native forms

403   of the molecules.

404

405   Secreted FhKTs taken up into host cells, from soluble secretions or within EVs, may act as

406   immunomodulatory proteins by targeting lysosomal cathepsin cysteine proteases. In this

407   study, recombinant FhKT1 proteins inhibited human cathepsin L and cathepsin K cysteine

408   proteases at sub-nanomolar concentrations indicating high potency against these enzymes.

409 Lysosomal cathepsin L plays a critical role in MHC class II antigen processing before

410 peptides are presented on the cell surface [45-47]. Inhibition of antigen processing and

411 presentation impairs T cell stimulation and differentiation, resulting in diminished adaptive

412 immunity [48, 49]. Indeed, *F. hepatica* infection in mice has been shown to have a

413 suppressive impact on the immune response [50]. Secreted and EV-contained FhKT1

414 proteins could be internalized by host immune cells and potentially interfere with cathepsin

415 L-mediated antigen processing within the lysosomal compartment of the cell.

416

417 Cathepsin K has previously been shown to function in TLR-9 mediated activation of

418 dendritic cells (DCs) [51]. Inhibition of cathepsin K in these cells results in reduced IL-6 and

419 IL-23 production, thus preventing the induction of Th-17 cells [51, 52]. Suppression of a Th-

420 17 response was also previously observed in *F. hepatica* infection in mice [50]. Interestingly,

421 Falcón et al. [53] found that a <10 kDa fraction of adult *F. hepatica* somatic extract contained

422 a FhKT1 that suppressed LPS-activated DCs *in vitro* and suppressed Th1 / Th-17 allogenic

423 response in mice. Thus, FhKT may play a role in impairing host early innate immune

424 responses by blocking cathepsin K activity, particularly given that this protease exhibits low-

425 level expression in DCs [51].

426

427 Our previous studies on the structural interaction of FhKTs with cysteine proteases predicted

428 that the P1 Leu^{15} residue of the inhibitor sits in the S2 subsite of the active site pocket [10]. In

429 this study, we revised the model of interaction in light of our new small inhibitor binding data

430 that showed the potential simultaneous binding of FhKT1.1 and Z-Phe-Ala-CHN$_2$/E-64 in the

431 active site of the protease. In our new docking model, FhKT1.1 binds to the S2, S1' and S2'

432 pockets of FhCL1 in a somewhat similar manner to the competitive binding of cystatins to

433 the cathepsin B of humans (PDB:3K9M). We found that Leu^{15} sits at the water-exposed

434   interface of the S1 and S1' subsites (near Asn[268] of S1 and Val[245] of S1', Figure 5A and 5B)

435   and that replacement of the Leu[15] with Arg[15] in FhKT1.3 (and the variant FhKT1Leu[15]/Arg[15])

436   does not change binding to the cysteine proteases.

437

438   To better understand the mechanism of cysteine protease inhibition we examined the

439   importance of Arg[19], which is conserved in FhKT1 and FhKT1.3 and predicted to make

440   interactions with Asp[125] and Trp[291] at the rim of the cysteine protease active site. While the

441   variant rFhKT1Arg[19]/Ala[19] still inhibited cathepsin L-like cysteine proteases, the inhibition

442   constant (Ki) values revealed that binding was much reduced (ranging between 5-18-fold less

443   potent) compared to rFhKT1 and rFhKT1.3. By contrast, and unexpectedly, we found that the

444   variant rFhKT1Arg[19]/Ala[19] exhibited enhanced inhibitory activity against HsCL compared to

445   the wildtype enzyme, which we could not explain using our predicted model of interaction.

446   Nevertheless, the data suggests that the residue Arg[19] plays a significant and important role in

447   the binding of the RSL.

448

449   The above data together with structural modelling of FhKT1 indicates that the RSL does not

450   interact directly with the S1 reactive site of the cysteine protease but instead forms a bridge

451   that sits across the S1 active site pocket blocking substrate access to the reactive Cys[132]. Pull-

452   down experiments demonstrated that human cystatin C, a cysteine protease inhibitor of 13.3

453   kDa, occupies the cysteine protease active site groove and prevents binding of rFhKT1 to

454   FhCL1 in a concentration-dependant manner. By contrast, low molecular weight cysteine

455   protease inhibitors E-64 (360 Da) or Z-Phe-Ala-CHN$_2$ (394 Da) that occupy the S1 active

456   pocket and penetrate the S2 space [19, 54] did not prevent the binding of rFhKT1 to the

457   enzyme active site groove. These observations prove that FhKT1 binds to the active site

458 groove but does not penetrate deeply into the S1 or S2 sub-sites of the active site region,

459 which is supported by our structural modelling.

460

461 In summary, phylogenetic analysis revealed a wide diversity of inhibitors amongst digenean

462 trematode parasites that could be categorised into seven distinct groups. *F. hepatica*

463 expresses a temporally-regulated family of Kunitz-like inhibitors with unique cysteine

464 protease-inhibiting activity. *F. hepatica* also expresses several broad-spectrum cystatins [55]

465 that inhibit the parasite cathepsin L-like cysteine proteases indicating the importance for the

466 parasite to tightly control the activity of these enzymes during migration, growth and

467 development. Moreover, inhibition of key host lysosomal cathepsin L-like cysteine proteases

468 involved in antigen processing which may be a means of controlling host responses to

469 parasite molecules. These putative pivotal roles in host-parasite interaction position the

470 FhKT1 inhibitors as viable vaccine and drug targets against the globally important zoonotic

471 parasite *F. hepatica*.

472

473 **EXPERIMENTAL PROCEDURES**

474 **Parasite material and excystment protocols**

475 *F. hepatica* metacercariae (Italian isolate; Ridgeway Research, UK) were excysted and

476 cultured in RPMI 1640 medium containing 2 mM L-glutamine, 30 mM HEPES, 0.1% (w/v)

477 glucose, and 2.5 µg/ml gentamycin for up to 48 h as described by Cwiklinski et al. [4]. Adult

478 *F. hepatica* parasites were recovered from livers of naturally infected sheep at a local

479 abattoir, washed with PBS (containing 0.1% glucose) and cultured in the same medium for 5

480 h. The parasite culture media (parasite ES proteins) were collected, centrifuged at $300 \times g$ for

481 10 min and at $700 \times g$ for 30 min and stored at -80°C. The adult parasite somatic extract was

482 isolated by homogenizing parasites in 500μl PBS and centrifugation at 4,500 x *g* for 10 min,

483 and the supernatant stored at -80°C.

484

485 **Identification of a Kunitz-type protease inhibitor gene family and phylogenetic analysis**

486 The identification of the *F. hepatica* Kunitz-type (KT) inhibitor gene (*fhkt*) family was

487 performed using BLAST analysis against the *F. hepatica* genome [3, 56] using previously

488 identified *F. hepatica* Kunitz sequences (FhKT1, [10]; Fh_Contig2704, [57]) followed by

489 manual assessment to identify the characteristic and conserved six cysteine residues that form

490 three distinctive disulphide bridges. In addition, the *F. hepatica* gene models [3] putatively

491 annotated using *in silico* tools (Uniprot, Gene Ontology (GO), and InterProScan) were

492 screened for 'kunitz-type protein' within their descriptive annotations. Homologous

493 trematode KT sequences were retrieved using BLAST, manual curation and putative

494 annotation as above from publically available transcriptome and genome databases, from the

495 following databases: (a) WormBase ParaSite [56, 58]): *Clonorchis sinensis* (PRJDA72781),

496 *Opisthorchis viverrini* (PRJNA222628), *Echinostoma caproni* (PRJEB1207), *Schistosoma*

497 *haematobium* (PRJNA78265), *Schistosoma japonicum* (PRJEA34885) and *Schistosoma*

498 *mansoni* (PRJEA36577); (b), Trematode.net [59, 60]: *Paragonimus westermani*

499 (PRJNA219632); (c) the adult *Fasciola gigantica* transcriptome [61, 62]. Maximum

500 likelihood trees were constructed with the trematode-specific KT sequences (Supplementary

501 Table 1) using MEGA v4.0 with the nucleotide sequence corresponding to central structural

502 domain of the KTs (first to last conserved Cys residue, see Fig. 1), with bootstrap values

503 calculated from 1000 iterations.

504

505 **Transcriptomic and proteomic expression analysis of the *F. hepatica* Kunitz gene family**

506　Differential gene transcription of the *F. hepatica* KT genes was investigated using the

507　available *F. hepatica* transcriptome data (European Nucleotide Archive accession number

508　PRJEB6904) as described by [3], represented as the log of the number of transcripts per

509　million (log TPM; Fig 2) and fold change relative to the metacercariae stage (Supplementary

510　Figure S3).

511

512　Quantitative real time PCR (qPCR) analysis of the *fhkt1* genes was carried out on NEJs

513　cultured for 0h, 6h, 10h, 24h and 48h in RPMI 1640 medium containing 2mM L-glutamine,

514　30mM HEPES, 0.1% glucose, 2.5 µg/ml gentamycin and 10% foetal calf serum

515　(ThermoFisher Scientific). Total RNA extraction and cDNA synthesis was carried out as per

516　Cwiklinski et al. [4]. Primers were designed to amplify all three genes based on the genomic

517　sequence data. qPCR reactions were performed in 20 µl reaction volumes in triplicate, using

518　1 µl cDNA diluted 1:2, 10 µl of Platinum SYBR Green qPCR SuperMix-UDG kit

519　(ThermoFisher Scientific) and 1 µM of each primer (*fhkt1*Forward (5'-

520　ATCCAAAAACGATGTCTTCTTCCGG-3') and *fhkt1*Reverse (5'-

521　TTGGAATCGAAAACCACAGTT-'3). A negative control (no template) was included in

522　each assay. qPCR was performed using a Rotor-Gene thermocycler (Qiagen), with the

523　following cycling conditions: 95°C: 10 min; 40 cycles: 95°C:10 s, 54°C:15 s, 72°C: 20 s;

524　72°C: 5 min. Relative expression analysis was performed manually using Pfaffl's

525　Augmented ΔΔCt method [63] whereby the comparative cycle threshold (Ct) values of

526　samples of interest were compared normalised to the housekeeping gene, Glyceraldehyde 3-

527　phosphate dehydrogenase (GAPDH). In order for this method to be valid, amplification

528　efficiencies of individual reactions were verified using the comparative quantification

529　package within the Rotor-Gene Q software v2.1.0. Annealing temperatures and melt-curve

530　analysis was also carried out to check for single DNA products produced by these primer

531   sets. Results were analysed using One Way ANOVA (P-value <0.05 was deemed

532   statistically significant) and visualised using version 6.00 for Windows, GraphPad Software

533   (https://www.graphpad.com/scientific-software/prism/).

534

535   *F. hepatica* proteomic datasets were interrogated for FhKT1 proteins within the secreted

536   products (ES proteins) from NEJs (ProteomeXchange Consortium repository: PXD007255;

537   [4]), adult parasites (ProteomeXchange Consortium repository:PXD002570; [11]) and the

538   extracellular vesicles (EVs) isolated from adult *F. hepatica* ES (ProteomeXchange

539   Consortium repository:PXD002570; [11]). The number of unique peptides identified was

540   validated using Scaffold (version 4.3.2) [11].

541

542   **Analysis of the *fhkt1.3* cDNA**

543   Total RNA was extracted from a single adult fluke using miRNeasy Mini Kit (Qiagen) and

544   cDNA synthesised using High capacity cDNA reverse transcription kit (ThermoFisher

545   Scientific). Primers were designed to bind to a consensus nucleotide sequence encoding the

546   signal peptides identified by SignalP v4.1, in *fhkt1.1* and *fhkt1.2* (fhkt1SP:

547   5'ATGCGTTGTTTCACAATCGCC 3') and to the nucleotide sequence corresponding to the

548   conserved residues at the start of the conserved KT domain (fhkt1F:

549   5'AACGATGTCTTCTTCCGGTCG 3'). The reverse primer corresponded to the conserved

550   3' end of the *fhkt1* nucleotide sequence (fhkt1R: 5'

551   TTATTGGAATCGAAAACCACAGTTG 3'). PCR products were gel purified (QIAquick

552   Gel Extraction Kit, Qiagen) and transformed into TOP10 (BL21) *Escherichia coli* competent

553   cells (TOPO cloning system, ThermoFisher Scientific). Plasmid DNA from 20 transformants

554   were sequenced by Source Bioscience (UK).

555

Whole-mount NEJ Immunolocalization of FhKT1 by confocal microscopy

NEJs were fixed with 4% paraformaldehyde in 0.1 M PBS and then incubated in 100 mM

PBS containing anti-recombinant FhKT1 antiserum at a 1:500 dilution, overnight at 4°C,

followed by three washes in AbD [4, 10]. NEJ were then incubated in a 1:200 dilution of the

secondary antibody, fluorescein isothiocyanate (FITC)-labelled goat anti-rabbit IgG (Sigma-

Aldrich) overnight at 4°C. To counter-stain muscle tissues, NEJs were incubated in AbD

containing 200 μg/ml phalloidintetramethylrhodamine isothiocyanate (TRITC) overnight at

4°C. NEJs were whole-mounted in a 9:1 glycerol solution containing 100 mM propyl gallate

and viewed using confocal scanning laser microscopy (Leica TCS SP5; Leica Microsystems,

UK) under the HCX PL APO CS ×100 oil objective lens.

Immunolocalization of FhKT1 in adult *F. hepatica* by fluorescence light microscopy.

Adult flukes were fixed in paraformaldehyde at 4°C overnight, washed with PBS, dehydrated

with ethanol and embedded in JB-4 resin (Sigma-Aldrich) [4, 10]. Sections (2 μm) were

incubated in either an anti-peptide antibody raised in mice to the following FhKT1.1 amino

acid sequence, Cys-Glu-Gly-Asn-Asp-Asn-Arg-Phe-Asp-Ser-Lys-Ser-Ser-Cys, or pre-

immune sera, each at a 1:500 dilution. Sections were then washed three times in PBS with

0.5% Triton X-100 for 30 min and incubated in a 1:500 dilution of the secondary antibody,

fluorescein isothiocyanate (FITC)-labelled goat anti-mouse immunoglobulin (Sigma-

Aldrich). After three washes in PBS with 0.5% Triton X-100 for 30 min sections were dried

and coverslips mounted using glycerol:PBS (9:1) containing 100 mM propyl gallate. Sections

were viewed using a Leica DM 2500 light microscope under the HCX PL FLUOSTAR x10

and x40 lenses.

Molecular Modelling

581 The homology model of FhKT1.1 (built based on the BPTI crystal structure) was taken from

582 Smith et al. [10] to conduct docking studies using the protein-protein docking server,

583 ZDOCK [64]. Given that FhKT1 does not block the binding of the competitive antagonist, Z-

584 Phe-Ala-CHN2 to FhCL1, the initial binding hypothesis of FhKT1-FhCL1 that we proposed

585 in Smith et al. [10] was re-examined. ZDOCK produced ten docking solutions involving

586 binding to S1 and S2 pockets as well as S1' and S2' pockets. A complex of FhCL1 bound to

587 Z-Phe-Ala-CHN$_2$ was generated by superimposing the structure of FhCL1 with the crystal

588 structures of human cathepsins L and K bound to diazomethylketone or E-64 inhibitors at S1

589 and S2 pockets, respectively (Pdb codes: 3OF9 and 1ATK) using Maestro 10.2 (Schrödinger

590 [65]). The ZDOCK structure of FhKT1.1 that does not overlap with the S1, S2 and S3

591 pockets of FhCL1 was selected as a starting conformation for energy optimization. The

592 FhCL1 complex bound to FhKT1 and Z-Phe-Ala-CHN$_2$ was subjected to an optimization

593 procedure involving 2000 step minimization and 200 ps dynamic simulations using the

594 MacroModel module of Schrodinger software [65]. Graphical representations were created

595 using Schrödinger software (LLC. Maestro, Version 2018-4; https://www.schrodinger.com/).

596

597 **Production of functional recombinant Kunitz-type inhibitors in the methylotrophic**

598 **yeast *Pichia pastoris***

599 Recombinant proteins were expressed in *P. pastoris* with a C-terminal His-tag as previously

600 described in Smith et al. [10]. Protein yield was quantified by measuring the absorbance at

601 A$_{280}$ and using the Protein calculator [66]. Protein purity was visualised by NuPAGE Novex

602 4-12% BisTris protein gel (ThermoFisher Scientific).

603

604 **Determination of FhKT1.1, FhKT1.3 and FhKT1Arg[19]/Ala[19] protease inhibition profile**

605 **and kinetics**

606   Enzymes included bovine trypsin, bovine chymotrypsin, human cathepsin B, human

607   cathepsin L, and human cathepsin S (all Sigma-Aldrich) and human cathepsin K (Enzo Life

608   Sciences). Purified *F. hepatica* cathepsin L1 (FhCL1), *F. hepatica* cathepsin L2 (FhCL2) and

609   *F. hepatica* cathepsin L3 (FhCL3) [67] and recombinant FhKT1 and FhKT1Leu[15]/Arg[15] [10]

610   were expressed as active recombinant proteins in *P. pastoris*. Reaction conditions and

611   substrates employed for measuring the activity of each protease were as reported by Smith et

612   al. [10]. Additionally, rFhCL3 activity was measured using the fluorogenic substrate Z-Gly-

613   Pro-Arg-NHMec (20μM).

614   KT inhibitors (2μM) was incubated with each protease in a 100μl volume of reaction

615   buffer for 15 min at 37°C. Reaction were brought to 200μl with the addition of fluorogenic

616   substrate dissolved in reaction buffer and proteolytic activity measured as RFU (relative

617   fluorescent units) using a PolarStar Omega spectrophotometer (BMG LabTech, UK).

618   Inhibition constants were determined using the Morrison equation for tight-binding inhibition

619   as previously described [10].

620

621   **rFhKT1 and rFhKT1.3 inhibition of *F. hepatica* cysteine proteases in somatic extract**

622   **and ES proteins**

623   Decreasing concentrations of rFhKT1 and rFhKT1.3 were incubated for 10 min at 37°C with

624   2μl of *F. hepatica* somatic extract or ES diluted in 100mM sodium acetate buffer pH5.5,

625   containing 2mM DTT and 0.01% Brij L23. To measure any remaining cysteine protease

626   activity in the presence of the inhibitors, 100μl of reaction buffer (sodium acetate buffer

627   pH5.5, containing 2mM DTT and 0.01% Brij L23) containing Z-Phe-Arg-NHMec (20μM)

628   was added and monitored using a PolarStar Omega spectrophotometer. As a positive control,

629   assays used decreasing concentrations of the cysteine protease inhibitor E-64.

630

Inhibition competition and pulldown assays of FhKT1 binding to native cathepsin L-like cysteine proteases

To examine the ability of cysteine protease inhibitors to compete for active site binding with FhKT1, competition assays were carried out using three different cysteine proteases inhibitors, E-64, Z-Phe-Ala-CHN$_2$ and recombinant human cystatin C (Sigma-Aldrich). The cysteine protease inhibitors were added to adult *F. hepatica* ES proteins (~20 µg protein) at the following final concentrations; E-64 and Z-Phe-Ala-CHN$_2$: 1mM, 100µM, 5µM, 1µM, 100nM, 10nM and 1nM; recombinant human cystatin C: 500 nM, 250 nM, 100 nM, 50 nM, 10 nM and 1nM. After 15 min at 37°C, two µl of each combined *F. hepatica* ES proteins and inhibitor sample was added to 98 µl of sodium acetate buffer (containing 2 mM DTT and 0.01% Brij L23), then brought to 200 µl upon the addition of the fluorogenic peptide substrate Z-Phe-Arg-NHMec (20 µM) dissolved in the sodium acetate buffer. Fluorogenic assays were carried out in triplicate using a PolarStar Omega spectrophotometer, reported at RFU.

The remainder of each *F. hepatica* ES proteins/ inhibitor mix was added to 1 µM of rFhKT1.1 and incubated at 37°C for 30 min. Following the addition of 10 µl of Ni-NTA beads, the samples were incubated at room temperature with rotation for 30 min. The Ni-NTA beads were then pelleted by centrifugation at 100 x *g* in a bench top microcentrifuge, washed twice with wash buffer and the bound proteins eluted in 20 µl of elution buffer. Eluted proteins were analysed on NuPAGE Novex 4-12% BisTris gels (ThermoFisher Scientific), stained with Biosafe Coomassie (BioRad), and imaged using a G:BOX Chemi XRQ imager (Syngene). Additional inhibition competition assays were performed with rFhKT1Arg[19]/Ala[19] and Z-Phe-Ala-CHN$_2$-inhibited cathepsin L ES proteases as described above for rFhKT1.

656 **REFERENCES**

657 1. Robinson MW, Dalton JP. Zoonotic helminth infections with particular emphasis on
658 fasciolosis and other trematodiases. *Philos Trans R Soc Lond B Biol Sci*. **364,** 2763-2776
659 (2009).

660 2. Cwiklinski K, O'Neill SM, Donnelly S, Dalton JP. A prospective view of animal and
661 human Fasciolosis. *Parasite Immunol*. **38**, 558-568 (2016).

662 3. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, et al. The
663 *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host
664 environment and the capacity for rapid evolution. *Genome Biol*. **16**, 71-015-0632-2 (2015).

665 4. Cwiklinski K, Jewhurst H, McVeigh P, Barbour T, Maule AG, Tort J, et al. Infection by
666 the helminth parasite *Fasciola hepatica* requires rapid regulation of metabolic, virulence, and
667 invasive factors to adjust to its mammalian host. *Mol Cell Proteomics*. **17**, 792-809 (2018).

668 5. Cwiklinski K, Donnelly S, Drysdale O, Jewhurst H, Smith D, De Marco Verissimo C, et
669 al. The cathepsin-like cysteine peptidases of trematodes of the genus *Fasciola*. *Adv Parasitol*.
670 **104**, 113-164 (2019).

671 6. Robinson MW, Menon R, Donnelly SM, Dalton JP, Ranganathan S. An integrated
672 transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola*
673 *hepatica*: proteins associated with invasion and infection of the mammalian host. *Mol Cell*
674 *Proteomics*. **8**, 1891-1907 (2009).

675 7. Knox DP. Proteinase inhibitors and helminth parasite infection. *Parasite Immunol*. **29**, 57-
676 71 (2007).

677 8. Guo A. Comparative analysis of cystatin superfamily in platyhelminths. *PLoS One*. **10**,
678 e0124683 (2015).

679 9. Ranasinghe SL, McManus DP. Protease Inhibitors of Parasitic Flukes: Emerging Roles in
680 Parasite Survival and Immune Defence. *Trends Parasitol*. **33**, 400-413 (2017).

681 10. Smith D, Tikhonova IG, Jewhurst HL, Drysdale OC, Dvorak J, Robinson MW, et al.
682 Unexpected Activity of a Novel Kunitz-type Inhibitor: Inhibition Of Cysteine Proteases But
683 Not Serine Proteases. *J Biol Chem*. **291**, 19220-19234 (2016).

684 11. Cwiklinski K, de la Torre-Escudero E, Trelis M, Bernal D, Dufresne PJ, Brennan GP, et
685 al. The Extracellular Vesicles of the Helminth Pathogen, *Fasciola hepatica*: Biogenesis
686 Pathways and Cargo Molecules Involved in Parasite Pathogenesis. *Mol Cell Proteomics*. **14**,
687 3258-3273 (2015).

688 12. Grzesiak A, Helland R, Smalas AO, Krowarsch D, Dadlez M, Otlewski J. Substitutions at
689 the P(1) position in BPTI strongly affect the association energy with serine proteinases. *J Mol*
690 *Biol*. **301**, 205-217 (2000).

691   13. Laskowski M, Qasim MA. What can the structures of enzyme-inhibitor complexes tell us
692   about the structures of enzyme substrate complexes? *Biochim Biophys Acta*. **1477**, 324-337
693   (2000).

694   14. Ranasinghe S, McManus DP. Structure and function of invertebrate Kunitz serine
695   protease inhibitors. *Dev Comp Immunol*. **39**, 219-227 (2013).

696   15. Jefferies JR, Campbell AM, van Rossum AJ, Barrett J, Brophy PM. Proteomic analysis of
697   *Fasciola hepatica* excretory-secretory products. *Proteomics*. **1**, 1128-1132 (2001).

698   16. Morphew RM, Wright HA, LaCourse EJ, Woods DJ, Brophy PM. Comparative
699   proteomics of excretory-secretory proteins released by the liver fluke *Fasciola hepatica* in
700   sheep host bile and during in vitro culture ex host. *Mol Cell Proteomics*. **6**, 963-972 (2007).

701   17. Robinson MW, Tort JF, Lowther J, Donnelly SM, Wong E, Xu W, et al. Proteomics and
702   phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola
703   hepatica*: expansion of a repertoire of virulence-associated factors. *Mol Cell Proteomics*. **7**,
704   1111-1123 (2008).

705   18. Di Maggio LS, Tirloni L, Pinto AF, Diedrich JK, Yates Iii JR, Benavides U, et al. Across
706   intra-mammalian stages of the liver fluke *Fasciola hepatica*: a proteomic study. *Sci Rep*. **6**,
707   32796 (2016).

708   19. Grzonka Z, Jankowska E, Kasprzykowski F, Kasprzykowska R, Lankiewicz L, Wiczk W,
709   et al. Structural studies of cysteine proteases and their inhibitors. *Acta Biochim Pol*. **48**, 1-20
710   (2001).

711   20. Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, et al. Cysteine cathepsins: from
712   structure, function and regulation to new frontiers. *Biochim Biophys Acta*. **1824**, 68-88
713   (2012).

714   21. Janoff A. Elastases and emphysema. Current assessment of the protease-antiprotease
715   hypothesis. *Am Rev Respir Dis*. **132**, 417-433 (1985).

716   22. Gadek JE, Pacht ER. The protease-antiprotease balance within the human lung:
717   implications for the pathogenesis of emphysema. *Lung*. **168 Suppl**, 552-564 (1990).

718   23. Zhang XX, Cwiklinski K, Hu RS, Zheng WB, Sheng ZA, Zhang FK, et al. Complex and
719   dynamic transcriptional changes allow the helminth *Fasciola gigantica* to adjust to its
720   intermediate snail and definitive mammalian hosts. *BMC Genomics*. **20**, 729 (2019).

721   24. Salter JP, Lim KC, Hansell E, Hsieh I, McKerrow JH. Schistosome invasion of human
722   skin and degradation of dermal elastin are mediated by a single serine protease. *J Biol Chem*.
723   **275**, 38667-73 (2000).

724   25. Ranasinghe SL, Fischer K, Gobert GN, McManus DP. A novel coagulation inhibitor from
725   *Schistosoma japonicum*. *Parasitol*. **142**, 1663-1672 (2015).

726 26. Ranasinghe SL, Fischer K, Gobert GN, McManus DP. Functional expression of a novel
727 Kunitz type protease inhibitor from the human blood fluke *Schistosoma mansoni*. *Parasit*
728 *Vectors*. **8**, 408-015-1022-z (2015).

729 27. Dalton JP, Neill SO, Stack C, Collins P, Walshe A, Sekiya M, et al. *Fasciola hepatica*
730 cathepsin L-like proteases: biology, function, and potential in the development of first
731 generation liver fluke vaccines. *Int J Parasitol*. **33**, 1173-1181 (2003).

732 28. Norbury LJ, Beckham S, Pike RN, Grams R, Spithill TW, Fecondo JV, et al. Adult and
733 juvenile *Fasciola* cathepsin L proteases: different enzymes for different roles. *Biochimie*. **93**,
734 604-611 (2011).

735 29. Ranasinghe SL, Fischer K, Zhang W, Gobert GN, McManus DP. Cloning and
736 Characterization of Two Potent Kunitz Type Protease Inhibitors from *Echinococcus*
737 *granulosus*. *PLoS Negl Trop Dis*. **9**, e0004268 (2015).

738 30. Gonzalez S, Flo M, Margenat M, Duran R, Gonzalez-Sapienza G, Grana M, et al. A
739 family of diverse Kunitz inhibitors from *Echinococcus granulosus* potentially involved in
740 host-parasite cross-talk. *PLoS One*. **4**, e7009 (2009).

741 31. Fló M, Margenat M, Pellizza L, Grana M, Duran R, Baez A, et al. Functional diversity of
742 secreted cestode Kunitz proteins: Inhibition of serine peptidases and blockade of cation
743 channels. *PLoS Pathog*. **13**, e1006169 (2017).

744 32. Chu D, Bungiro RD, Ibanez M, Harrison LM, Campodonico E, Jones BF, et al. Molecular
745 characterization of *Ancylostoma ceylanicum* Kunitz-type serine protease inhibitor: evidence
746 for a role in hookworm-associated growth delay. *Infect Immun*. **72**, 2214-2221 (2004).

747 33. Milstone AM, Harrison LM, Bungiro RD, Kuzmic P, Cappello M. A broad spectrum
748 Kunitz type serine protease inhibitor secreted by the hookworm *Ancylostoma ceylanicum*. *J*
749 *Biol Chem*. **275**, 29391-29399 (2000).

750 34. Bozas SE, Panaccio M, Creaney J, Dosen M, Parsons JC, Vlasuk GV, et al.
751 Characterisation of a novel Kunitz-type molecule from the trematode *Fasciola hepatica*. *Mol*
752 *Biochem Parasitol*. **74**, 19-29 (1995).

753 35. Lowther J, Robinson MW, Donnelly SM, Xu W, Stack CM, Matthews JM, et al. The
754 importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine
755 protease. *PLoS Negl Trop Dis*. **3**, e369 (2009).

756 36. Collins PR, Stack CM, O'Neill SM, Doyle S, Ryan T, Brennan GP, et al. Cathepsin L1,
757 the major protease involved in liver fluke (*Fasciola hepatica*) virulence: propetide cleavage
758 sites and autoactivation of the zymogen secreted from gastrodermal cells. *J Biol Chem*. **279**,
759 17038-17046 (2004).

760 37. Nomura T, Fujisawa Y. Processing properties of recombinant human procathepsin L.
761 *Biochem Biophys Res Commun*. **230**, 143-146 (1997).

762 38. Turk B, Turk D, Salvesen GS. Regulating cysteine protease activity: essential role of
763 protease inhibitors as guardians and regulators. *Curr Pharm Des*. **8**, 1623-1637 (2002).

764 39. Deveraux QL, Reed JC. IAP family proteins--suppressors of apoptosis. *Genes Dev*. **13**,
765 239-252 (1999).

766 40. Marcilla A, Trelis M, Cortes A, Sotillo J, Cantalapiedra F, Minguez MT, et al.
767 Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins
768 and are internalized in intestinal host cells. *PLoS One*. 7, e45974 (2012).

769 41. Silverman JM, Clos J, de'Oliveira CC, Shirvani O, Fang Y, Wang C, et al. An exosome-
770 based secretion pathway is responsible for protein export from Leishmania and
771 communication with macrophages. *J Cell Sci*. **123**, 842-852 (2010).

772 42. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. *J
773 Cell Biol*. **200**, 373-383 (2013).

774 43. Mulcahy LA, Pink RC, Carter DR. Routes and mechanisms of extracellular vesicle
775 uptake. *J Extracell Vesicles*. **3**, 10.3402/jev.v3.24641 (2014).

776 44. Muiño L, Perteguer MJ, Garate T, Martinez-Sernandez V, Beltran A, Romaris F, et al.
777 Molecular and immunological characterization of *Fasciola* antigens recognized by the MM3
778 monoclonal antibody. *Mol Biochem Parasitol*. **179**, 80-90 (2011).

779 45. Ishidoh K, Kominami E. Gene regulation and extracellular functions of procathepsin L.
780 *Biol Chem*. **379**, 131-135 (1998).

781 46. Honey K, Rudensky AY. Lysosomal cysteine proteases regulate antigen presentation. *Nat
782 Rev Immunol*. **3**, 472-482 (2003).

783 47. Barrett AJ, Woessner JF, Rawlings ND. *Handbook of Proteolytic Enzymes, Volume 1:*
784 3rd ed. Elsevier; 2012.

785 48. Maizels RM, Yazdanbakhsh M. Immune regulation by helminth parasites: cellular and
786 molecular mechanisms. *Nat Rev Immunol*. **3**, 733-744 (2003).

787 49. Zavasnik-Bergant T, Turk B. Cysteine cathepsins in the immune response. *Tissue
788 Antigens*. **67**, 349-355 (2006).

789 50. Walsh KP, Brady MT, Finlay CM, Boon L, Mills KH. Infection with a helminth parasite
790 attenuates autoimmunity through TGF-beta-mediated suppression of Th17 and Th1
791 responses. *J Immunol*. **183**, 1577-1586 (2009).

792 51. Asagiri M, Hirai T, Kunigami T, Kamano S, Gober HJ, Okamoto K, et al. Cathepsin K-
793 dependent toll-like receptor 9 signaling revealed in experimental arthritis. *Science*. **319**, 624-
794 627 (2008).

795 52. Takayanagi H. The unexpected link between osteoclasts and the immune system. *Adv Exp
796 Med Biol*. **658**, 61-68 (2010).

797 53. Falcón CR, Masih D, Gatti G, Sanchez MC, Motran CC, Cervi L. *Fasciola hepatica*
798 Kunitz type molecule decreases dendritic cell activation and their ability to induce
799 inflammatory responses. *PLoS One*. **9**, e114505 (2014).

800   54. Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, et al. Cysteine cathepsins: from
801   structure, function and regulation to new frontiers. *Biochim Biophys Acta*. **1824,** 68-88
802   (2012).

803   55. Cancela M, Corvo I, DA Silva E, Teichmann A, Roche L, Diaz A, et al. Functional
804   characterization of single-domain cystatin-like cysteine proteinase inhibitors expressed by the
805   trematode *Fasciola hepatica. Parasitol*. **144,** 1695-1707 (2017).

806   56. WormBase ParaSite. http://parasite.wormbase.org/index.html. Version: WBPS14
807   (WS27).

808   57. Young ND, Hall RS, Jex AR, Cantacessi C, Gasser RB. Elucidating the transcriptome of
809   *Fasciola hepatica* - a key to fundamental and biotechnological discoveries for a neglected
810   parasite. *Biotechnol Adv*. **28**, 222-231 (2010).

811   58. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite - a
812   comprehensive resource for helminth genomics. *Mol Biochem Parasitol*. **215**, 2-10 (2017).

813   59. Martin J, Rosa BA, Ozersky P, Hallsworth-Pepin K, Zhang X, Bhonagiri-Palsikar V, et
814   al. Helminth.net: expansions to Nematode.net and an introduction to Trematode.net. *Nucleic*
815   *Acids Res*. **43**, 698-706 (2015).

816   60. Trematode.net. http://trematode.net/TN_frontpage.cgi.

817   61. Young ND, Jex AR, Cantacessi C, Hall RS, Campbell BE, Spithill TW, et al. A portrait
818   of the transcriptome of the neglected trematode, *Fasciola gigantica*-biological and
819   biotechnological implications. *PLoS Negl Trop Dis*. **5**, e1004 (2011).

820   62. Gasser lab webpage. http://bioinfosecond.vet.unimelb.edu.au/index.html.

821   63. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR.
822   *Nucleic Acids Res*. **29**, e45 (2001).

823   64. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. ZDOCK Server:
824   Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers.
825   *Bioinformatics*. **30**, 1771-3 (2014).

826   65. Small-Molecule Drug Discovery Suite release 2014-4, Schrödinger, LLC, New York, NY
827   (2014).

828   66. Protein calculator. https://www.mrc-lmb.cam.ac.uk/ms/methods/proteincalculator.html.

829   67. Stack CM, Caffrey CR, Donnelly SM, Seshaadri A, Lowther J, Tort JF, et al. Structural
830   and functional relationships in the virulence-associated cathepsin L proteases of the parasitic
831   liver fluke, *Fasciola hepatica. J Biol Chem*. **283**, 9896-9908 (2008).

832

833   **FIGURE LEGENDS**

834 **Fig. 1. A family of Kunitz-type inhibitors is present in the *F. hepatica* genome.** (A)

835 Maximum-likelihood phylogenetic tree based on sequences encoding the kunitz domain in *F.*

836 *hepatica* genome. Bootstrap values >50% from 1000 replicate iterations are shown. (B)

837 Amino acid sequence alignment of the translated amino acid sequence of each *fhkt* gene. The

838 P1, residue 15, within the reactive site loop is indicated with an arrow. Secretory signal

839 peptides are shown in red and lines show the three conserved disulphide bonds that occur

840 between the six conserved cysteine residues, highlighted in black, specifically between Cys^1

841 and Cys^6, Cys^2 and Cys^4, and Cys^3 and Cys^5. The Kunitz reactive loop region is shaded in

842 grey and P1 and P4' residues predicted to interact with residues at the S2 and S2' sites of

843 cathepsin L-like cysteine proteases are highlighted in blue. (C) Amino acid sequence

844 alignment of the reactive loop region of the seven members of the Kunitz family. The P1 and

845 P4' residues are highlighted red and blue, respectively.

846

847 **Fig. 2. Temporal regulation of *F. hepatica* KT gene expression during the mammalian**

848 **host-associated parasite stages.** (A) Graphical representation using Graphpad software (v6

849 for Windows) of relative gene expression for each member of the *F. hepatica* KT family

850 represented by log transcripts per million (log TPM) at various stages of the parasite's

851 lifecycle including the infective metacercariae (met), newly excysted juveniles (NEJ) at 1 h, 3

852 h and 24 h post-excystment, liver stage juvenile parasites 21-days post infection (Juv 21 d)

853 and bile ducts stage mature adult parasites (adult). The encoded P1 residue and the absence or

854 presence of a signal peptide (Sig. Pep) is shown alongside each graph. (B) Relative fold

855 expression of the genes representing the *fhkt1* group during the first 48h post-excystment by

856 NEJ (normalised to expression at NEJ excystment and relative to a GAPDH reference, with

857 SEM). Statistical analysis was carried out using One Way ANOVA with Tukey's post hoc

858   test, with all samples showing statistical differences in fold expression compared with the

859   levels at excystment ($p<0.001$: ***), visualised using GraphPad Software (v6 for Windows).

860

861   **Fig. 3. Immunolocalization of *F. hepatica* KT in infective NEJs in the first 48 h of**

862   **infection.** Panels A-D show a plane within the interior of the NEJ, whereas panels E-H show

863   a plane at the surface of the same individual NEJs. Following excystment, NEJs were

864   maintained in culture and sample parasites taken at 3 h post-excystment (panels B and F), 24

865   h post-excystment (panels C and G) and 48 h post-excystment (panels D and H). Parasites

866   were probed with rabbit pre-immune serum (A and E) or anti-FhKT1 antibodies followed by

867   FITC-labelled secondary antibodies. The presence of FhKT within *F. hepatica*-specific

868   structures are indicated by arrows; parenchymal cell bodies, white arrows; NEJ digestive

869   tract, grey arrows; network of channels below surface tegument, black arrows. All specimens

870   were counter-stained with phalloidin-tetramethylrhodamine isothiocyanate (TRITC) to stain

871   muscle tissue (red fluorescence) and provide structure. OS, oral sucker; VS, ventral sucker.

872   Scale bars = 20 μM.

873

874   **Fig. 4. Purification of recombinant Kunitz-type inhibitors and their inhibition profiles**

875   **against cysteine and serine proteases.** (A) LDS-PAGE analysis of the recombinantly yeast-

876   expressed *F. hepatica* KT inhibitors, rFhKT1.1 (lane 1); rFhKT1.1Leu^{15}/Arg^{15} (lane 2);

877   rFhKT1.3 (lane 3) and, rFhKT1Arg^{19}/Ala^{19} (lane 4). M, molecular size markers. (B)

878   Inhibitory activity (2 μM) of rFhKT1.1 (red bars), rFhKT1Leu^{15}/Arg^{15} (blue bars); rFhKT1.3

879   (green bars) and rFhKT1Arg^{19}/Ala^{19} (black bars) against a range of cysteine proteases

880   including *F. hepatica* cathepsin L1 (FhCL1), *F. hepatica* cathepsin L2 (FhCL2), *F. hepatica*

881   cathepsin L3 (FhCL3), human cathepsin L (HsCL), human cathepsin K (HsCK), human

882   cathepsin B (HsCB), human cathepsin S (HsCS) and the serine proteases trypsin and

883   chymotrypsin. Inhibition is presented relative to the activity of each enzyme in the absence of

884   inhibitors ± SD, visualised using GraphPad Software (v6 for Windows).

885

886   **Fig. 5. 3-D model of FhCL1 interactions with FhKT1.1 and Z-Phe-Ala-CHN₂.** (A) The

887   overall view of the 3-D model of the FhCL1 tertiary complex in surface representation.

888   Regions forming the S1, S2, S1' and S2' active site subsites of FhCL1 are shown in

889   alternating pink and cyan surface. The reactive Cys residue within the S1 pocket of the

890   cysteine protease is shown in yellow. FhKT1 is shown by the dark blue cartoon with the

891   reactive site loop (Leu^{15}-Arg^{19}) in a stick-like representation, whereas Z-Phe-Ala-CHN$_2$ is

892   shown by the green stick representation. (B) FhKT1.1 reactive site loop residues Leu^{15}-Gly^{16}-

893   Gly^{17} are shown in stick format (dark blue) within the active site of the *F. hepatica* cathepsin

894   L1. Leu^{15} sits near Val^{245} of S1' and Asn^{268} of S1. Arg^{19} forms hydrogen bonds and salt

895   bridge interactions with Asp^{125} and cation-π interactions with Trp^{291}, shown in by the dotted

896   lines in black, pink and green, respectively. (C) The cysteine protease inhibitor Z-Phe-Ala-

897   CHN$_2$ is shown in stick format (green) within the active site of the *F. hepatica* cathepsin L1.

898   The inhibitor forms hydrogen bonds with the backbone of Asn^{268} and Gly^{175} from the S1 and

899   S2 subsites, respectively, depicted by the black dotted lines. Graphical representations were

900   created using Schrödinger software (LLC. Maestro, Version 2018-4;

901   https://www.schrodinger.com/).

902

903   **Fig. 6. *F. hepatica* KT inhibitors inhibit all secreted cysteine protease activity, but not**

904   **somatic extract activity.** (A) Cysteine protease activity within somatic extracts of adult *F.*

905   *hepatica* were measured using the fluorogenic peptide substrate Z-Phe-Arg-NHMec (relative

906   fluorescent units, RFU/min) in the presence of FhKT1.1 (red line), FhKT1.3 (blue line) and

907   the cysteine protease inhibitor E-64 (black line) at a range of concentrations [*I*] (1 μM, 500

908   nM, 250 nM, 125 nM, 62.5 nM, 31.25 nM and 15.625 nM). The *F. hepatica* KT inhibitors

909   do not completely inhibit all cysteine protease within the somatic extract. (B) Cysteine

910   protease activity within the ES proteins of adult *F. hepatica* measured in the presence of

911   FhKT1.1 (red line) and FhKT1.3 (blue line) at a range of concentrations [*I*] (1 μM, 500 nM,

912   250 nM, 125 nM, 62.5 nM, 31.25 nM and 15.625 nM). Graphical representations produced

913   using GraphPad Software (v6 for Windows).

914

915   **Fig. 7. rFhKT1.1 binds to the active site groove but not within the active site pockets.**

916   Cysteine protease activity in adult *F. hepatica* excretory/secretory (ES) products measured in

917   the presence of increasing concentrations of cysteine protease inhibitors E-64 (A), Z-Phe-

918   Ala-CHN$_2$ (C) and human cystatin C (E) (% Activity, relative to the cysteine protease activity

919   of ES containing no inhibitor ± SD). rFhKT1.1 (10 μM) was added to replicate reaction

920   samples and then pull-down using NTA-beads (panels B, D and F). In the presence of the low

921   molecular weight inhibitors, E-64 (B) and Z-Phe-Ala-CHN$_2$ (D), rFhKT1.1 (black arrows, B

922   and D) is not prevented from interacting with the cathepsin L cysteine proteases (white

923   arrows, B and D). By contrast, interactions between rFhKT1.1 (black arrow, F) and cathepsin

924   cysteine protease (white arrow, F) are blocked by the recombinant human cystatin C (grey

925   arrow, F) observed by LDS-PAGE. Graphical representations produced using GraphPad

926   Software (v6 for Windows).

927

928   **SUPPLEMENTARY MATERIAL**

929   **Supplementary Table S1. Trematode Kunitz type inhibitor gene sequences used for**

930   **phylogenetic analysis.**

931 **Supplementary Table S2. FhKT proteins present in the excretory/secretory products of**

932 **NEJ and adult *F. hepatica* represented as Normalized Spectral Abundance Factor**
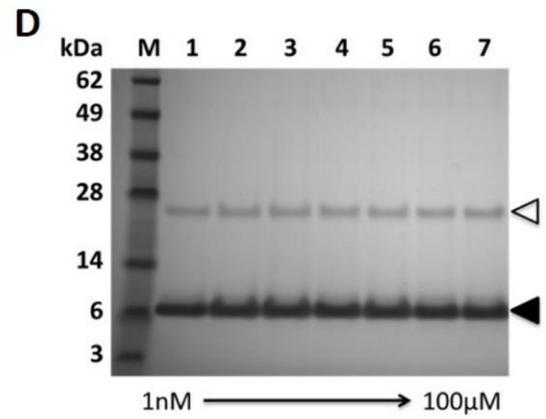
933 **(NSAF).**

934 **Supplementary Figure S1. Schematic representation of the *F. hepatica* Kunitz-type gene**

935 **structure**

936 **Supplementary Figure S2. Phylogenetic analysis of helminth Kunitz-type inhibitors.**

937 **Supplementary Figure S3. Graphical representation of differential gene expression**

938 **represented as fold change compared to the metacercariae stage.**

939 **Supplementary Figure S4. Immunolocalization of FhKT1 proteins in the reproductive**

940 **organs of adult *F. hepatica.***

941 **Supplementary Figure S5. Immunolocalization of FhKT1 proteins in the gut of adult *F.***

942 ***hepatica.***

943 **Supplementary Figure S6. rFhKT1.1Arg^{19}/Ala^{19} binding to the active site groove of**

944 **native cathepsin Ls is not prevented by Z-Phe-Ala-CHN$_2$.**

945 **Supplementary Figure S7. Full length gel photos as shown in (A) Fig. 6; (B) Fig. 9B, D**

946 **and F; (C) Supplemental Figure 5.**

947

948 **DECLARATIONS**

949 **Ethics approval and consent to participate**

950 Not applicable.

951

952 **Consent for publication**

953 Not applicable.

954

955 **Availability of data and materials**

956 The transcriptome data sets supporting the conclusions of this article are available in the

957 European Nucleotide Archive repository, PRJEB6904;

958 http://www.ebi.ac.uk/ena/data/view/PRJEB6904, previously reported by Cwiklinski et al. [3].

959 The mass spectrometry proteomics data analysed as part of this study have been deposited to

960 the ProteomeXchange Consortium via the PRIDE partner repository with the following data

961 set identifiers (a) NEJ specific datasets [4]: PXD007255 and 10.6019/PXD007255; (b) adult

962 ES and EV datasets [11]: PXD002570 and 10.6019/ PXD002570.

963

980

981 **Acknowledgements**

982 Not applicable.

983

984   Fig.1

985

986

987

988   Fig.2
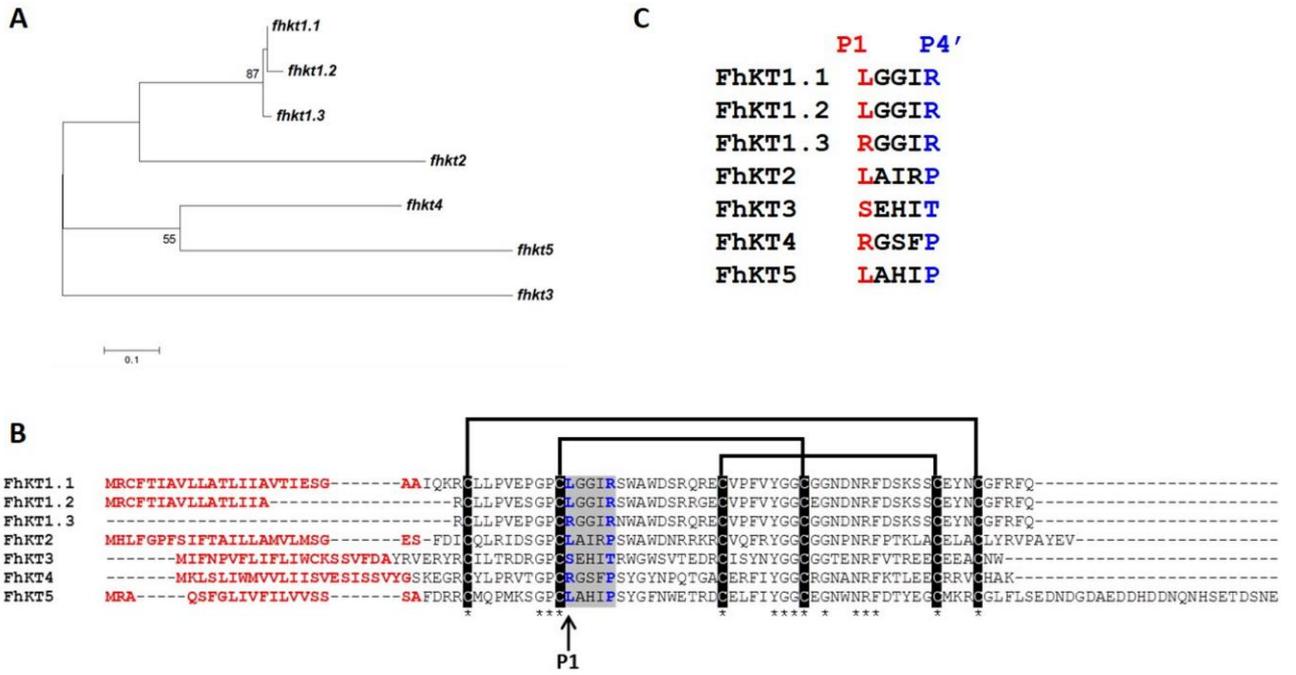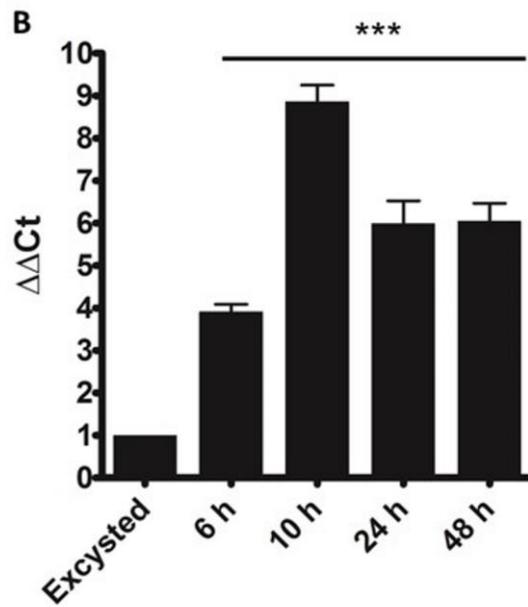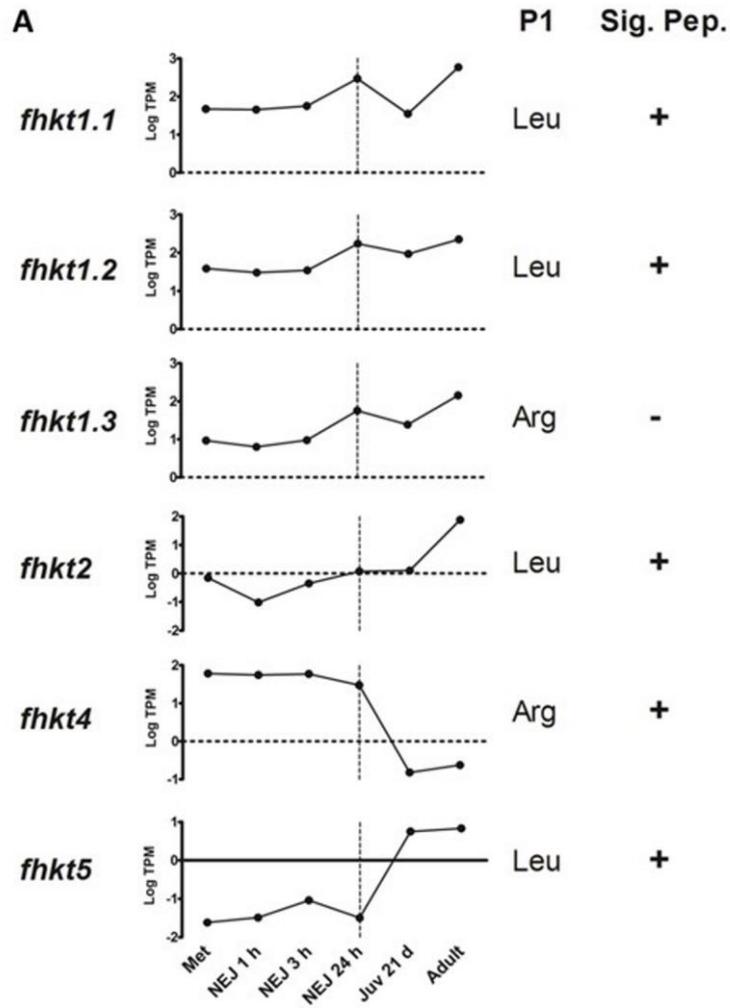
991   Fig.3

992

993

994   Fig.4
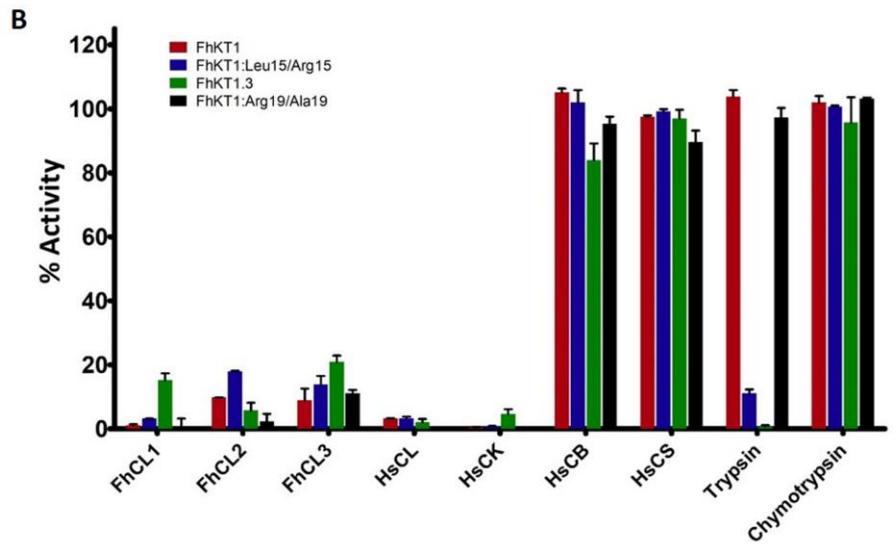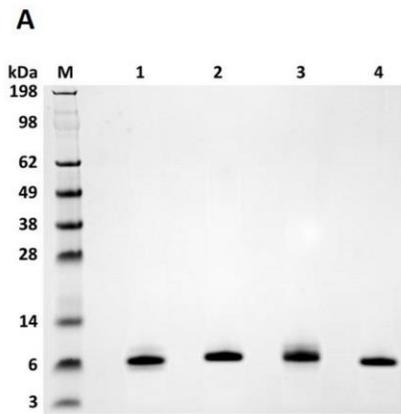
997   Fig.5

998

999

1000   Fig.6

A. Somatic Extract

B. ES products

1001

1002

1003   Fig.7

1004

1005
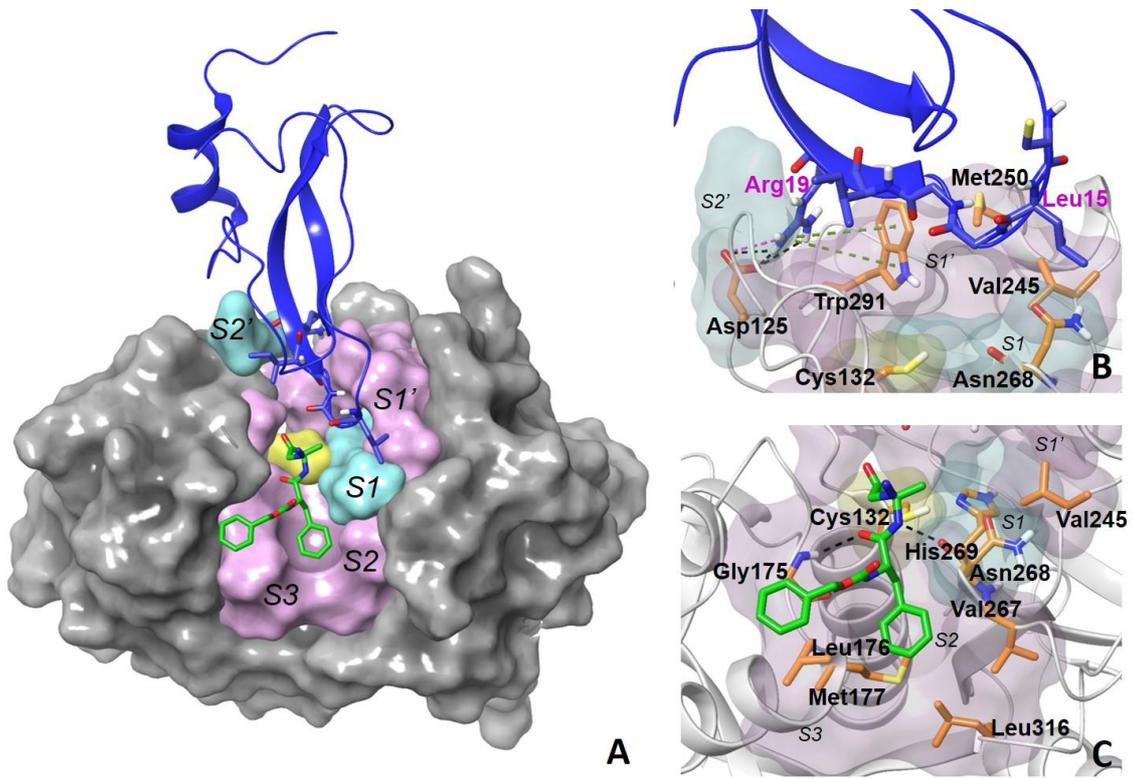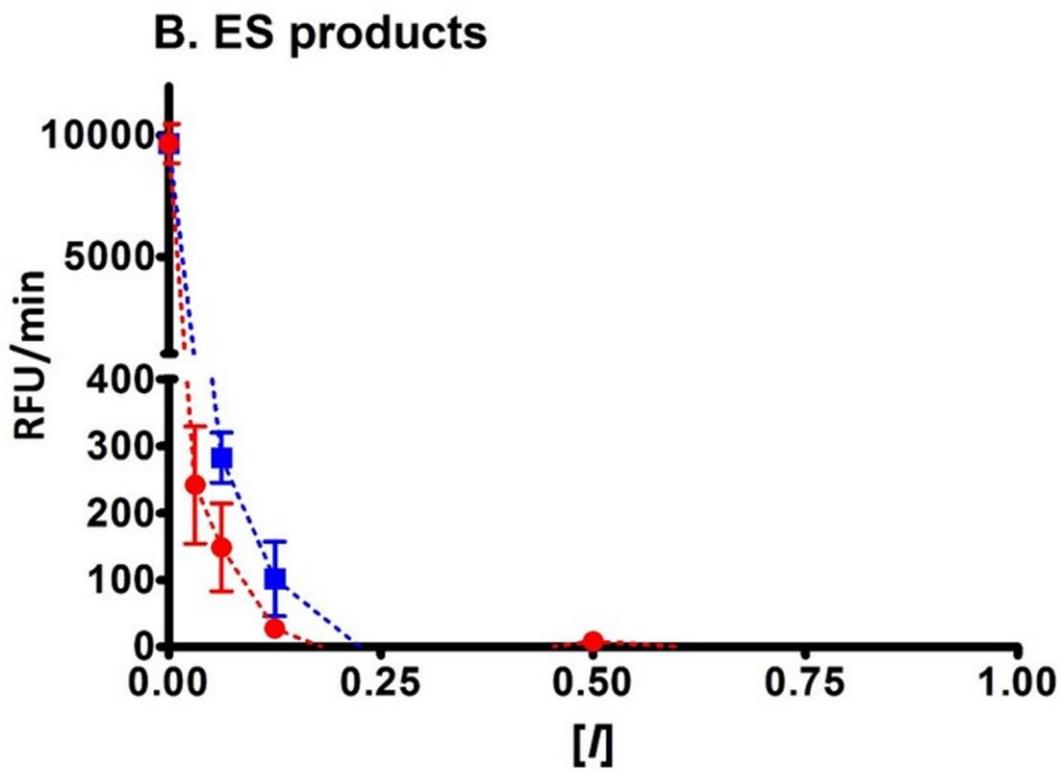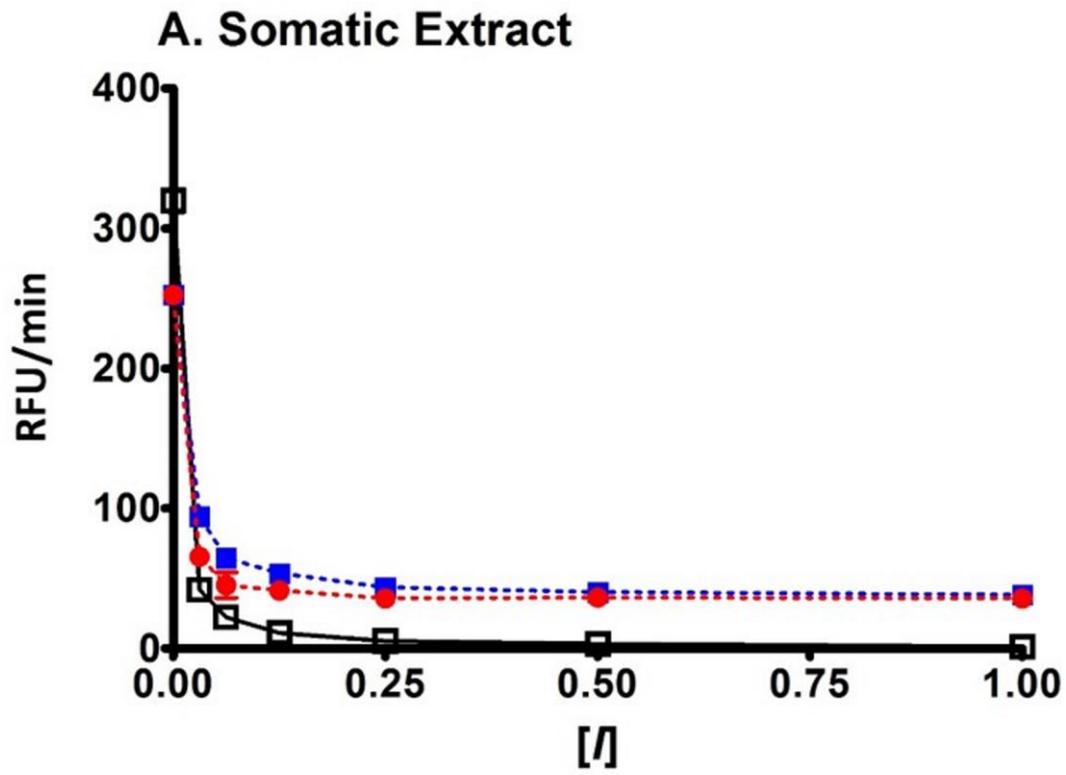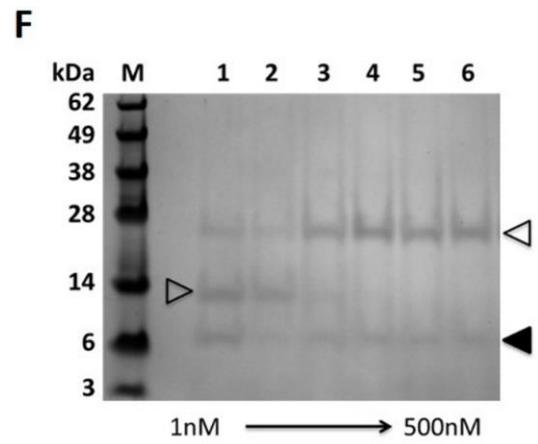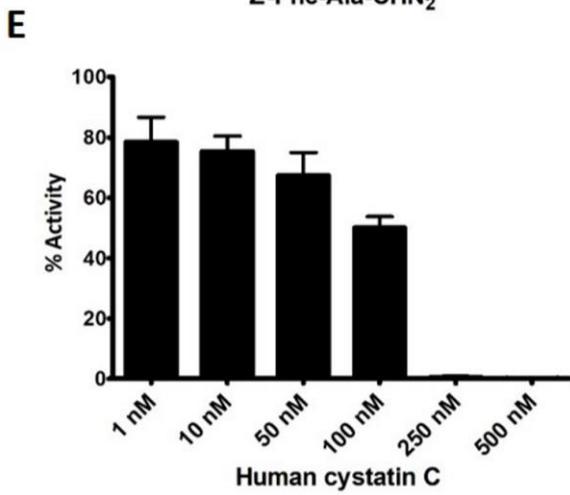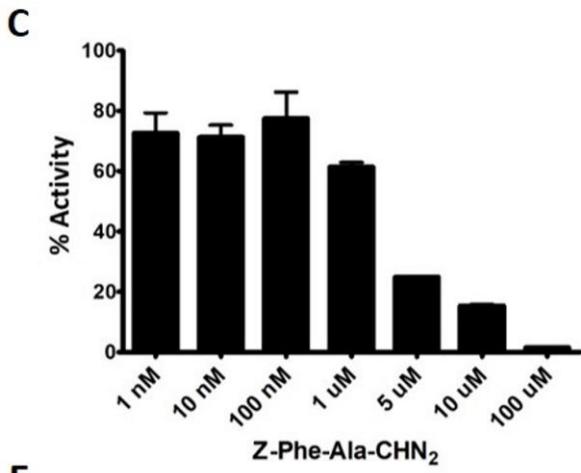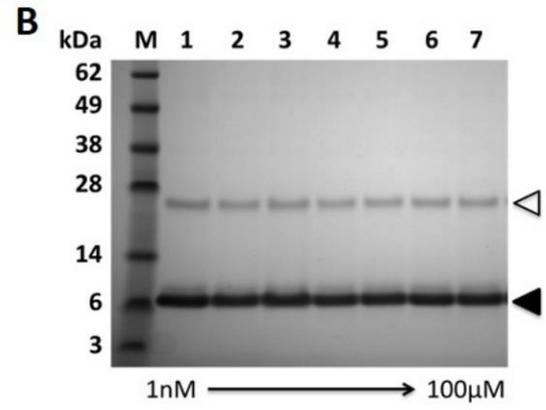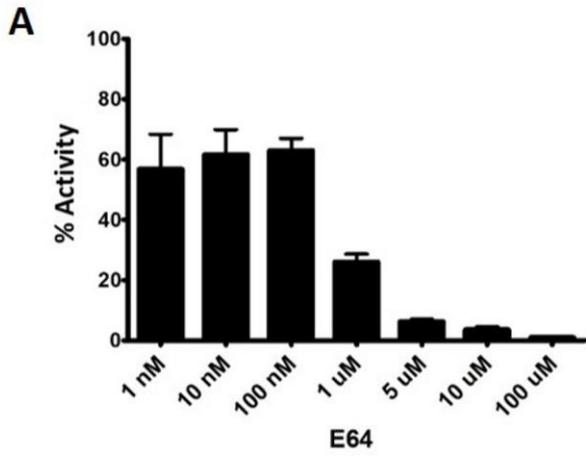
Supplementary Table S1: Trematode Kunitz type inhibitor gene sequences used for phylogenetic analysis

Nomenclature on Phylogram	Species	Gene/Scaffold ID	P1-P4'
Group A			
fgkt1	*Fasciola gigantica*	Contig12	LGGIR
fhkt1.2	*Fasciola hepatica*	BN1106_s318B000274	LGGIR
fhkt1.1	*Fasciola hepatica*	BN1106_s8826B000029	LGGIR
fhkt1.3	*Fasciola hepatica*	BN1106_s11518B000016	RGGIR
eckt1	*Echinostoma caproni*	ECPE_0001647201	FRGGI
eckt2	*Echinostoma caproni*	ECPE_0001016301	LAIHY
eckt3	*Echinostoma caproni*	ECPE_0001278301	LAIHY
fgkt2	*Fasciola gigantica*	Contig38896	LAIRP
fhkt2	*Fasciola hepatica*	BN1106_s6608B000014	LAIRP
Group B			
cskt1	*Clonorchis sinensis*	csin108828	HENYT
fhkt3	*Fasciola hepatica*	scaffold5597	SEHIT
cskt2	*Clonorchis sinensis*	csin107698	AENLR
eckt4	*Echinostoma caproni*	ECPE_0001105001	RGYHV
eckt5	*Echinostoma caproni*	ECPE_0000795001	RAAIT
sjkt1	*Schistosoma japonicum*	Sjp_0020270	RASLL
shkt1	*Schistosoma haematobium*	MS3_09801	RSKLH
smkt3	*Schistosoma mansoni*	Smp_139840	RASFN
Group C			
sjkt4	*Schistosoma japonicum*	Sjp_0097640	RYNYH
sjkt5	*Schistosoma japonicum*	Sjp_0117580	GNNST
sjkt6	*Schistosoma japonicum*	Sjp_0024620	LKRHP
shkt3	*Schistosoma haematobium*	MS3_09688	LQNIP
sjkt7	*Schistosoma japonicum*	Sjp_0024630	LHNKP
shkt4	*Schistosoma haematobium*	MS3_10748	LQKKP
smkt2	*Schistosoma mansoni*	Smp_179120	LQNKP
Group D			
sjkt2	*Schistosoma japonicum*	Sjp_0030350	RASIQ
cskt4	*Clonorchis sinensis*	csin103940	RGDVT
ovkt2	*Opisthorchis viverrini*	T265_11148	FVTAT
Group E			
cskt6	*Clonorchis sinensis*	csin112642	KAYMP
cskt7	*Clonorchis sinensis*	csin106214	LASMP
ovkt1	*Opisthorchis viverrini*	T265_11147	RAMIP
cskt3	*Clonorchis sinensis*	csin102310	RAMIP
eckt8	*Echinostoma caproni*	ECPE_0000615701	FHIFI
fhkt4	*Fasciola hepatica*	BN1106_s3911B000104	RGSFP
eckt6	*Echinostoma caproni*	ECPE_0001134901	GANIL

fhkt5	*Fasciola hepatica*	BN1106_s4272B000063	LAHIP
fgkt5	*Fasciola gigantica*	Contig26461	LALVP
eckt7	*Echinostoma caproni*	ECPE_0001663401	HTLLL
Group F			
pwkt5	*Paragonimus westermani*	comp19852	SDSIT
pwkt4	*Paragonimus westermani*	comp14876	GESLT
pwkt3	*Paragonimus westermani*	comp20534	MGHST
pwkt1	*Paragonimus westermani*	comp21326	RALIK
pwkt2	*Paragonimus westermani*	comp15715	RALMK
Group G			
smkt1	*Schistosoma mansoni*	Smp_147730	RALLK
shkt2	*Schistosoma haematobium*	MS3_09690	RALIK
cskt5	*Clonorchis sinensis*	csin102323	NFRTR
sjkt3	*Schistosoma japonicum*	Sjp_0076670	RGYFR
eckt9	*Echinostoma caproni*	ECPE_0000445201	SQFIT

1006
1007

1008

1009 **Supplementary Table S2:** FhKT proteins present in the excretory/secretory products of NEJ

1010 and adult *F. hepatica* represented as Normalized Spectral Abundance Factor (NSAF).

Inhibitor	NEJ 1h**	NEJ 3h**	NEJ 24h**	Adult	Adult EVs
FhKT1.1	-	-	0.00330	-	0.00416
FhKT1.2	0.01131	0.00482	0.00967	0.18816	0.01258
FhKT1.3	0.00447	0.00293	-	0.00615	-

1011 **Average NSAF value from triplicate samples

1012

1013 **Supplementary Figure S1. Schematic representation of the *F. hepatica* Kunitz-type gene**

1014 **structure.** Exons are represented by the coloured boxes and the introns are depicted as a

1015 black line. The nucleotide sequence encoding the signal peptide (exon 1) and Kunitz domain

1016 (exon 2) is represented by the blue exon and red exons, respectively. *fhkt1.3* consists of only

1017 one exon, with no sequence encoding a signal peptide. Scale bar indicates the length of 100

1018 nucleotide base pairs. The size of the respective introns in the *F. hepatica* genome assembly

1019 (PRJEB6687) is shown.

1020
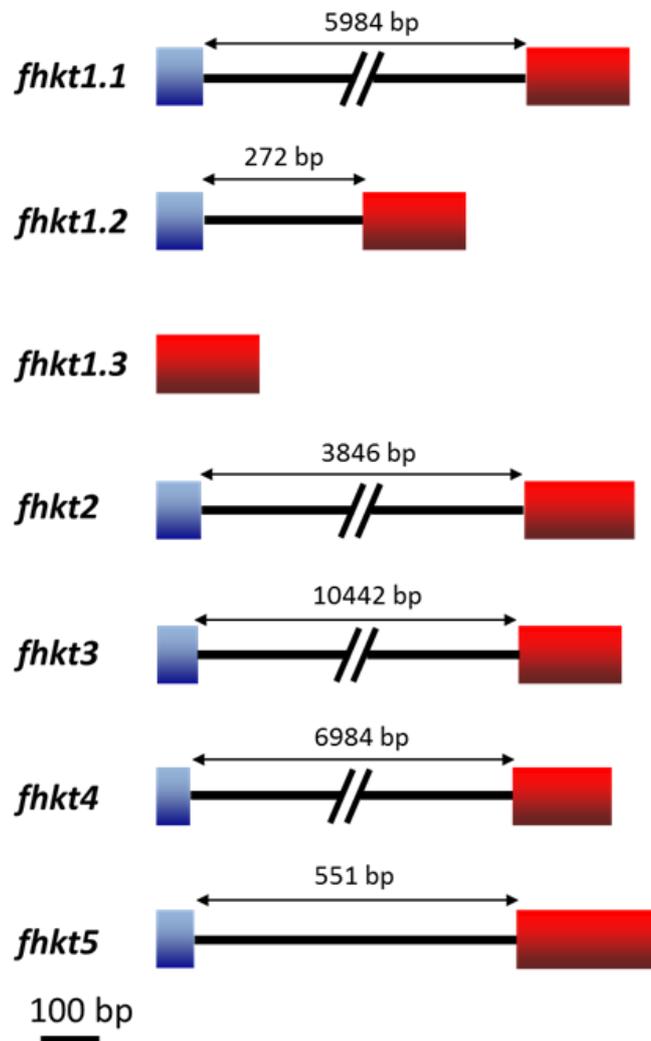
1021
1022
1023

1024   **Supplementary Figure S2. Phylogenetic analysis of helminth Kunitz-type inhibitors.**

1025   Maximum-likelihood phylogenetic tree computed using 1000 bootstrap replicates based on

1026   the sequence encoding the kunitz domain between cysteine residues 1 and 6 (Cys^1 and Cys^6),

1027   from nine helminth species: *Clonorchis sinensis* (*cskt1-7*), *Echinostoma caproni* (*eckt1-9*),

1028   *Fasciola hepatica* (*fhkt1-5*), *F. gigantica* (*fgkt1, 2* and *5*), *Opisthorchis viverrini* (*ovkt1-2*),

1029   *Paragonimus westermani* (*pwkt1-5*), *Schistosoma haematobium* (*shkt1-4*), *S. japonicum*

1030   (*sjkt1-7*) and *S. mansoni* (*smkt1-3*). Bootstrap values >50% are shown. The green coloured

1031   blocks highlight the presence of the *F. hepatica* KT sequences within 3 distinct groups

1032   formed by phylogenetic analysis (A, B, and E). The black lines represent the other clusters

1033   generated by phylogenetic analysis that do not contain *F. hepatica* sequences (clusters C, D,

1034   F and G). The accession numbers of the sequences represented by this phylogenetic tree are

1035   included in Supplementary Table S1.
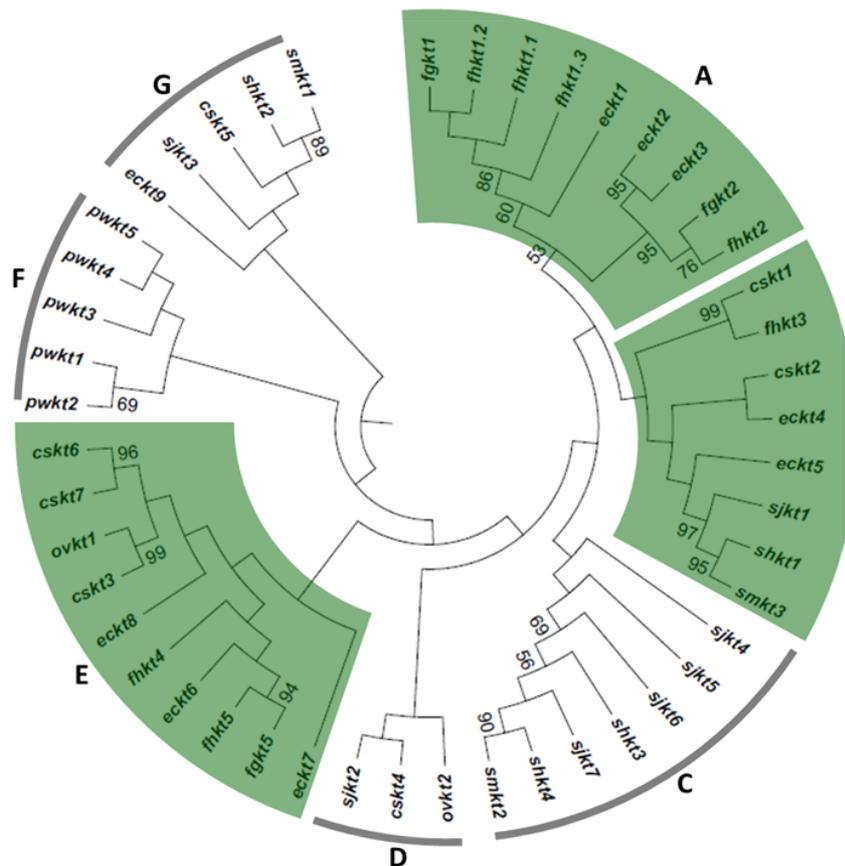
1036

1037

1038 **Supplementary Figure S3. Graphical representation of differential gene expression of**

1039 **the FhKT family represented as log fold change compared to the metacercariae life**

1040 **cycle stage.** Differential expression was calculated as reported by Cwiklinski et al. [3] using

1041 negative binomial model of successive developmental stages relative to metacercariae and

1042 tagwise dispersion estimated from all samples in edge R. ($p<0.01$: **; $p<0.001$: ***).

1043
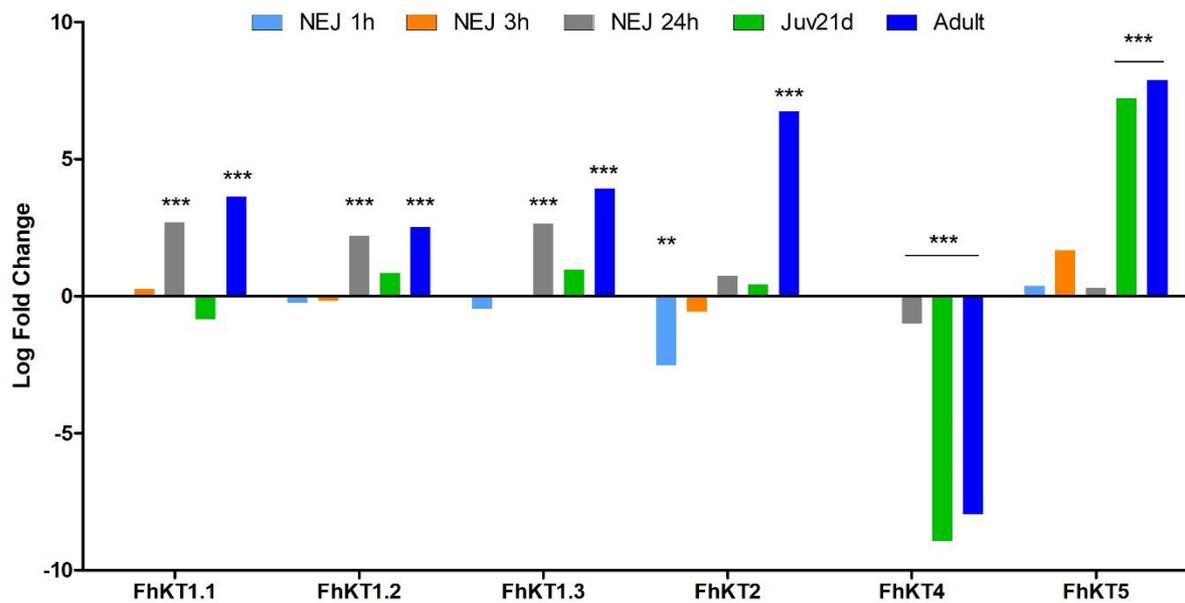
1044
1045
1046

1047 **Supplementary Figure S4. Immunolocalization of FhKT1 proteins in the reproductive**
1048 **organs of adult *F. hepatica*.** Serial sections through a JB-4 embedded adult *F. hepatica* were
1049 stained with Toluidine Blue or anti-FhKT1 antibodies. Toluidine Blue staining highlights the
1050 vitelline glands (A) and ovaries containing parasite eggs (E) (10x). Sections probed with
1051 mouse pre-immune serum (negative control) and stained with anti-mouse FITC show light
1052 background fluorescence in the testes (B) and eggshell (F) (10x). Serial sections probed with
1053 polyclonal anti-FhKT1 antibodies show strong staining in the vesicular structures within the
1054 vitelline cells contained in the vitelline glands (C) (10x), as also shown at higher
1055 magnification of 40x (D). Intense staining was also observed in the vitelline cell-derived yolk
1056 mass within eggs (G and H, 10X and 100x, respectively). E: egg; G: gut; T: testes; Te:
1057 tegument; V: vitelline glands containing vitelline cells. Scale bar: panels A-C and E-G, 200
1058 μm; panels D and H, 50 μm.
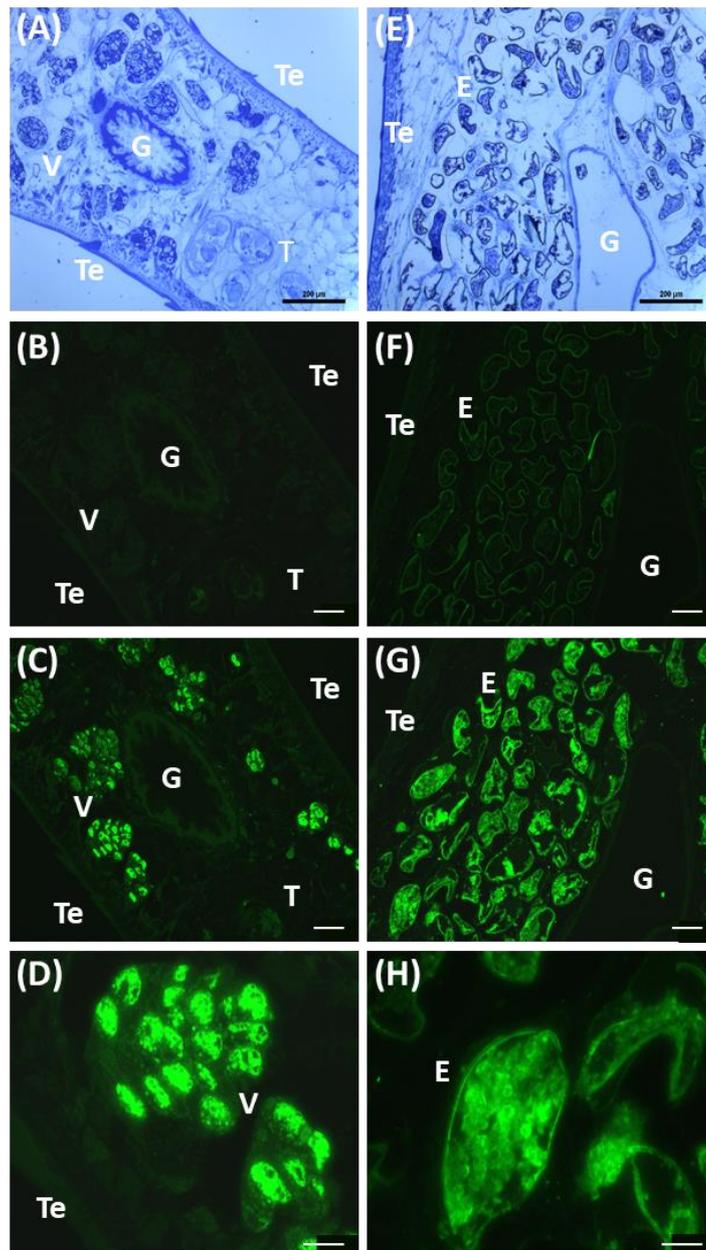
1059

1060

Supplementary Figure S5. Immunolocalization of FhKT1 proteins in the gut of adult *F. hepatica.* (A) Cross-section of a JB-4 section of adult *F. hepatica* stained with Toluidine Blue highlighting the digestive gut, parenchyma, tegument and the ventral sucker. (B) Section probed with mouse pre-immune serum (negative control) (C-D) or anti-FhKT1 antibodies followed by with anti-mouse-FITC. Visible staining in the within the ventral sucker and parenchymal cell bodies that are concentrated around near the gut, with diffuse staining throughout the parenchyma and gut (C and D, 10x and 40 x, respectively). G, gut; PC, Parenchymal cell body; Te, tegument; VS, ventral sucker. Scale bar: panels A-C, 200 μm; panel D, 50 μm.
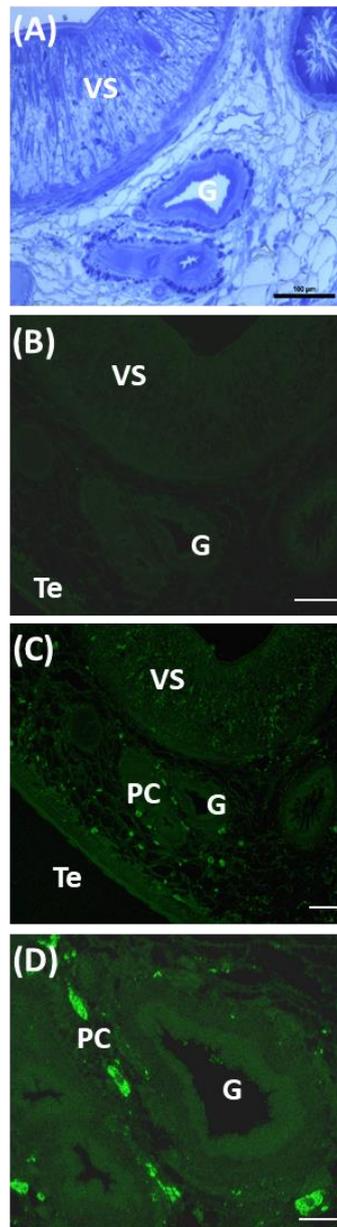
1073 **Supplementary Figure S6. rFhKT1.1Arg[19]/Ala[19] binding to the active site groove of**
1074 **native cathepsin Ls is not prevented by Z-Phe-Ala-CHN$_2$** (A) Cysteine protease activity
1075 within adult *F. hepatica* excretory/secretory (ES) products measured in the presence of Z-
1076 Phe-Ala-CHN$_2$, at a concentration ranging from 1 nM to 100μM (% Activity, relative to the
1077 cysteine protease activity of ES containing no inhibitor ± SD). (B) rFhKT1.1Arg[19]/Ala[19] (10
1078 μM) was added to replicate reaction samples and then pull-down using NTA-beads.
1079 rFhKT1.1Arg[19]/Ala[19] (black arrow) is not prevented from interacting with the cathepsin L
1080 cysteine proteases by Z-Phe-Ala-CHN$_2$ (white arrows).
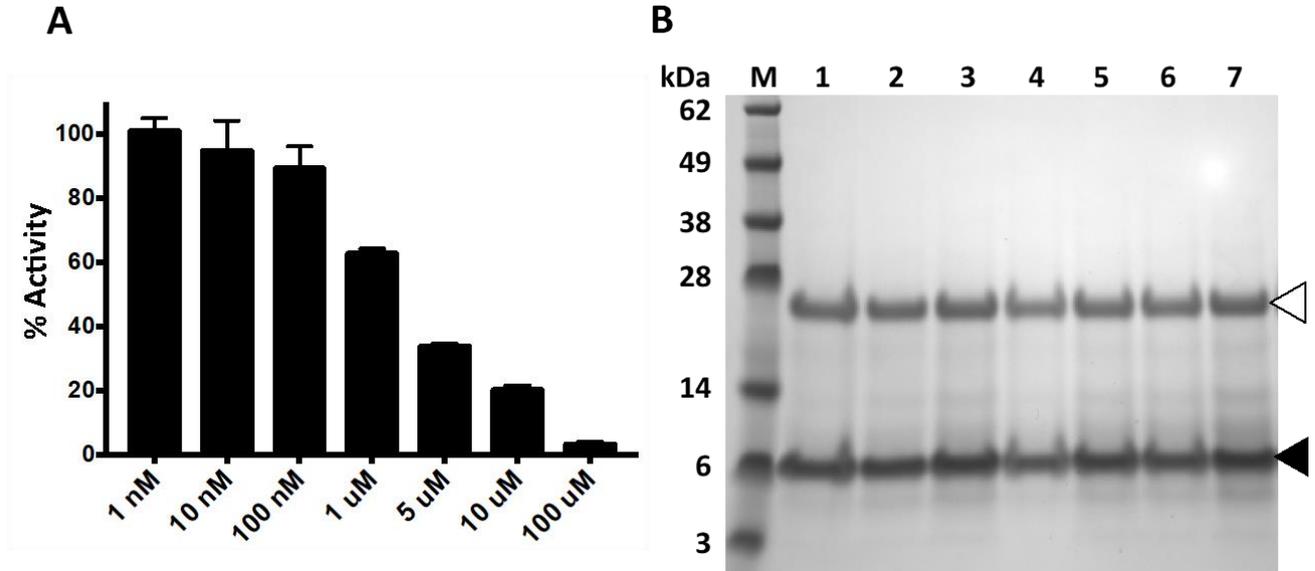
1081

1082
1083

1084 **Supplementary Figure S7. Full length gel photos as shown in (A) Fig. 4A; (B) Fig. 7B, D**

1085 **and F; (C) Supplemental Figure 6B.**

(A) Fig 4A
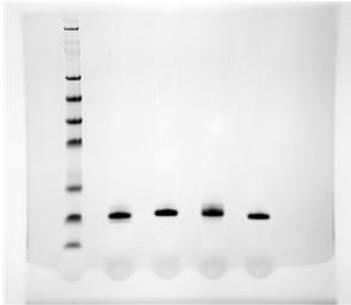
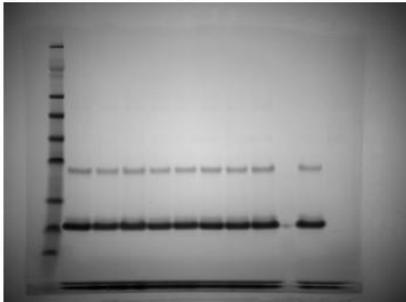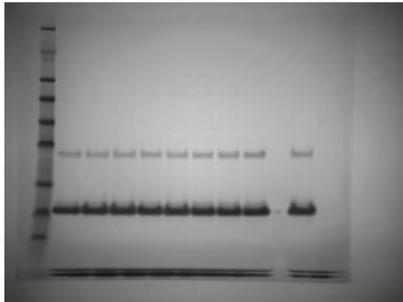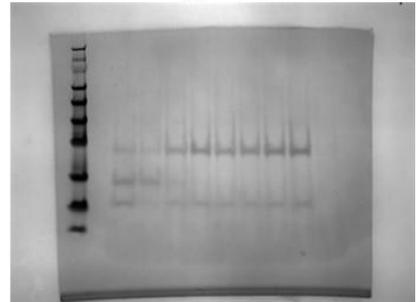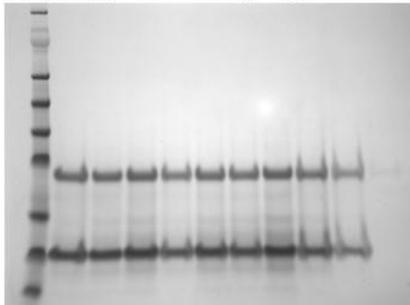(B)   Fig. 7B   Fig. 7D   Fig. 7F

(C) Supplementary Fig. 6B

1086
1087
1088
1089

1090