



Article

An Unusual Two-Domain Thyropin from Tick Saliva: NMR Solution Structure and Highly Selective Inhibition of Cysteine Cathepsins Modulated by Glycosaminoglycans

Zuzana Matoušková^{1,2}, Katarína Orsághová^{1,3} , Pavel Srb¹ , Jana Pytelková¹, Zdeněk Kukačka⁴ , Michal Buša¹ , Ondřej Hajdušek⁵, Radek Šíma^{5,6}, Milan Fábry¹, Petr Novák⁴ , Martin Horn¹ , Petr Kopáček⁵ and Michael Mareš^{1,*}

- ¹ Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo n. 2, 16610 Praha, Czech Republic; zuzana.matouskova@uochb.cas.cz (Z.M.); katarina.orsaghova@uochb.cas.cz (K.O.); michal.busa@uochb.cas.cz (M.B.); milan.fabry@uochb.cas.cz (M.F.); martin.horn@uochb.cas.cz (M.H.)
- ² Department of Biochemistry, Faculty of Science, Charles University, Hlavova 8, 12800 Praha, Czech Republic
- ³ First Faculty of Medicine, Charles University, Katerinska 32, 12108 Praha, Czech Republic
- ⁴ Institute of Microbiology, Czech Academy of Sciences, Prumyslova 595, 25250 Vestec, Czech Republic
- ⁵ Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branisovska 31, 37005 Ceske Budejovice, Czech Republic
- ⁶ Biopticka Laborator, Mikulasske Namesti 4, 32600 Plzen, Czech Republic
- * Correspondence: michael.mares@uochb.cas.cz

Abstract: The structure and biochemical properties of protease inhibitors from the thyropin family are poorly understood in parasites and pathogens. Here, we introduce a novel family member, Ir-thyropin (IrThy), which is secreted in the saliva of *Ixodes ricinus* ticks, vectors of Lyme borreliosis and tick-borne encephalitis. The IrThy molecule consists of two consecutive thyroglobulin type-1 (Tg1) domains with an unusual disulfide pattern. Recombinant IrThy was found to inhibit human host-derived cathepsin proteases with a high specificity for cathepsins V, K, and L among a wide range of screened cathepsins exhibiting diverse endo- and exopeptidase activities. Both Tg1 domains displayed inhibitory activities, but with distinct specificity profiles. We determined the spatial structure of one of the Tg1 domains by solution NMR spectroscopy and described its reactive center to elucidate the unique inhibitory specificity. Furthermore, we found that the inhibitory potency of IrThy was modulated in a complex manner by various glycosaminoglycans from host tissues. IrThy was additionally regulated by pH and proteolytic degradation. This study provides a comprehensive structure–function characterization of IrThy—the first investigated thyropin of parasite origin—and suggests its potential role in host–parasite interactions at the tick bite site.

Keywords: cathepsin; cysteine protease; tick; parasite; saliva; thyropin; protease inhibitor; protein structure



Citation: Matoušková, Z.; Orsághová, K.; Srb, P.; Pytelková, J.; Kukačka, Z.; Buša, M.; Hajdušek, O.; Šíma, R.; Fábry, M.; Novák, P.; et al. An Unusual Two-Domain Thyropin from Tick Saliva: NMR Solution Structure and Highly Selective Inhibition of Cysteine Cathepsins Modulated by Glycosaminoglycans. *Int. J. Mol. Sci.* **2024**, *25*, 2240. <https://doi.org/10.3390/ijms25042240>

Academic Editor: Antonio Lucacchini

Received: 12 December 2023

Revised: 2 February 2024

Accepted: 10 February 2024

Published: 13 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The thyroglobulin type-1 (Tg1) domain is typically a 7 kDa protein module found in a variety of unrelated multidomain proteins in the animal kingdom. It was first identified in thyroglobulin, which carries eleven Tg1 repeats [1]. The Tg1 domains, which feature a conserved pattern of cysteine residues, are divided into subgroups 1A and 1B based on the number of cysteines [1]. The variable loops in the molecular architecture of the Tg1 domain make it highly adaptive and account for the multiple functions it has acquired through evolution, including protease inhibition. Tg1 domain-containing proteins with antiproteolytic activity are known as thyropins [2], classified as protease inhibitor family I31 by the Merops database [3]. While Tg1 domains have been shown to directly inhibit cysteine and aspartic proteases, mainly those of the cathepsin type, individual thyropins differ greatly in their inhibitory specificities [4].

Most of our knowledge about thyropins and their inhibitory activity derives from the following members: the p41 invariant chain fragment, involved in major histocompatibility complex (MHC) class II maturation and antigen processing [5,6]; saxiphilin, a neurotoxin-binding protein from bullfrogs [7,8]; equistatin, derived from sea anemones, which contains domains that selectively inhibit cysteine or aspartic cathepsins [9,10]; human testican-1, a multi-domain proteoglycan from the brain [11,12]; human epithelial cell adhesion molecule (EpCAM), a carcinoma cell marker [13]; and the cysteine protease inhibitor ECI from chum salmon eggs [14]. Additionally, Tg1 domains in insulin-like growth factor binding proteins (IGFBPs) have been proposed as proteolysis regulators [15,16].

Of the thyropins whose inhibitory properties have been demonstrated, only two have been characterized at the structural level. These include crystal structures of saxiphilin, which has two Tg1 domains embedded in a large transferrin architecture [8], and the p41 fragment, a single Tg1 domain thyropin, in complex with cathepsin L [6]. This complex identified the reactive center on the Tg1 domain, which is formed by three loops that bind to the active site of cathepsin L. In addition to their inhibitory effects, Tg1 domains may establish more complex relationships with the target cathepsin. For instance, as demonstrated for cathepsin L complexed with the p41 fragment, the Tg1 domain may stabilize the protease at a neutral pH [17]. Another example is the proteolytic degradation of testican-1, an inhibitor of cathepsin L, by the same protease at high concentrations [12].

There is very little information on the structure–function relationships of thyropins in invertebrates, particularly those found in parasites. To address this gap, we performed a molecular characterization of thyropin from the hard tick *Ixodes ricinus*, a well-known vector of Lyme borreliosis and tick-borne encephalitis. Although thyropins have been recently identified by transcriptomic and proteomic analyses in several tick species [18–21], their biological roles remain unknown. Our interest in thyropins builds upon successful research into another family of protease inhibitors targeting cysteine cathepsins, the cystatins. These proteins found in tick saliva act as powerful immunomodulators of the host response, while those in the tick gut regulate blood digestion [22–25].

In this study, we provide a comprehensive functional and structural characterization of Ir-thyropin (IrThy), an unusual thyropin secreted in the saliva of *I. ricinus* ticks. Its small molecule consists of only two Tg1 domains, which are atypically disulfide-bonded. The narrow inhibitory specificity of IrThy was explained by NMR structural analysis and was further shown to be modulated by glycosaminoglycans (GAGs). Finally, we propose a role for IrThy in the initial stages of host–parasite interactions at the tick bite site.

2. Results

2.1. IrThy—A Salivary Protein Secreted by the *I. ricinus* Tick

Based on the analysis of the salivary gland transcriptome (sialome) of the hard tick *I. ricinus* [26], we identified a set of homologous sequences belonging to the I31 family of thyropin proteins. Using a consensus sequence, we designed primers for PCR amplification from the cDNA of salivary glands of half-fed *I. ricinus* females. The resulting amplicon was cloned and sequenced, yielding a sequence with an open reading frame of 501 bp (Figure S1). This sequence was deposited in GenBank under accession number PP107940. The encoded sequence of the protein, named IrThy, exhibited ~97% and ~96% identity to GenBank accessions JAA67758.1 and JAP74612.1, respectively, from *I. ricinus* [26,27].

To demonstrate the specific localization of IrThy, we applied an LC–MS/MS strategy based on enzymatic digestion of a complex proteome in conjunction with MS/MS peptide sequencing. Based on our analysis of salivary gland extract and saliva from *I. ricinus* females (Table S1), we identified IrThy in both materials, indicating that IrThy was a secreted salivary protein.

Next, we compared the mRNA transcript levels of IrThy in two tissues that play a key role in the tick feeding process (Figure 1A). Specifically, we observed substantially higher expression in the salivary glands of half-fed females than in the midgut, which is responsible for blood digestion. Figure 1B shows the dynamics of IrThy expression

during tick feeding on the host. Unfed females displayed the highest level of expression, which then gradually decreased to reach an expression level approximately one order of magnitude lower than in fed females.

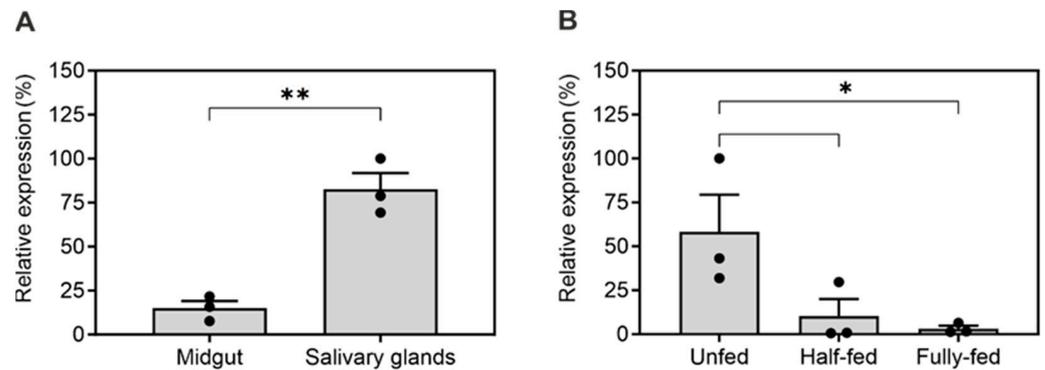


Figure 1. Transcriptional profiling of IrThy. Expression of IrThy was evaluated by qRT-PCR in adult female *Ixodes ricinus* ticks in (A) salivary glands and midgut from half-fed ticks and (B) whole-body homogenates at different stages of blood feeding. The mRNA transcript levels were normalized to the housekeeping gene *ferritin1*. Results represent the mean \pm SD of biological triplicates (pooled samples), expressed relative to the highest measured value (100%); * $p < 0.05$, ** $p < 0.005$.

In summary, we cloned IrThy from *I. ricinus* as a new member of the thyropin family. It is predominantly expressed in the salivary glands of unfed female ticks and is secreted into the tick saliva, suggesting its involvement in the early stages of tick–host interaction.

2.2. IrThy Is a Two-Domain Thyropin: Sequence Analysis, Evolutionary Distribution, and Recombinant Production

IrThy cDNA contains an open reading frame encoding a protein of 166 amino acid residues. It contains a 19-residue signal peptide followed by a mature IrThy sequence of 147 residues consisting of two consecutive Tg1 domains from the I31 thyropin family (Figure 2A). The N- and C-terminal domains have 79 and 68 residues, respectively, and share 26% sequence identity. Both domains belong to Tg1 subtype 1A [1], which features three disulfide bonds. By evaluating the connectivity of the disulfides by LC–MS/MS peptide mapping (Table S2), we found that all of the cysteines formed disulfide bridges. The connectivity is as follows: Cys17–Cys37, Cys48–Cys59, and Cys61–Cys81 in the N-domain, and Cys79–Cys103, Cys115–Cys139, and Cys122–Cys147 in the C-domain (Figure 2A). Two of them are associated with the central sequence Cys79(N-domain)–Lys80–Cys81(C-domain), where Cys79 is cross-linked to the C-domain and Cys81 to the N-domain; therefore, they can be classified as interdomain disulfide bridges (Figure 2A). Sequence comparison shows that the N-domain of IrThy shares the disulfide pattern with the prototype thyropin, the human p41 fragment, while the pattern in the C-domain is modified (Figure S2).

Given that IrThy consists of two Tg1 domains, we decided to investigate the phylogenetic distribution of IrThy homologs with analogous domain compositions. The Pfam database (hosted by InterPro) [28] lists over 20,000 proteins containing the Tg1 domain (Pfam entry: PF00086). These proteins are typically multi-domain, with up to more than a dozen Tg1 repeats, often combined with domains of other types. Figure 2C shows the distribution of Tg1 domain-containing proteins for three different search parameters. Selection I presents all sequences that have two Tg1 domains along with other Pfam domains. Most of them, 77%, belong to vertebrates, and 23% are mammalian. Selection II differs from Selection I in that it excludes proteins containing other Pfam domains. This resulted in a predominance of invertebrate sequences (70%), with over half coming from arthropods (57%). In Selection III, we restricted the sequence length to 180 residues spanning two consecutive Tg1 domains; this is consistent with the domain pattern of IrThy. These short two-domain thyropins were mostly found in invertebrates (63%), including ticks; none

were found in mammals. This indicated that IrThy-like thyropins from ticks have no analogs in mammalian hosts.

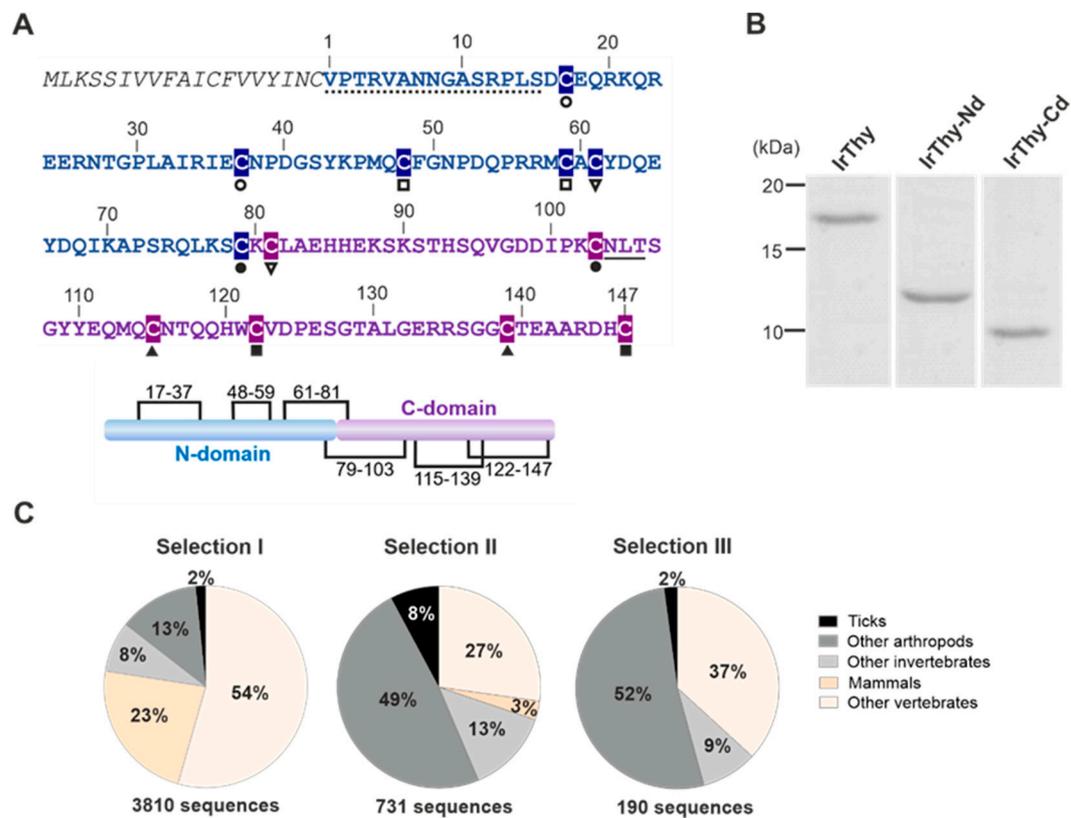


Figure 2. Sequence, evolution, and recombinant production of IrThy. **(A)** Amino acid sequence of IrThy featuring a signal peptide (italics), an unstructured N-terminal region (dotted), an N-terminal thyroglobulin-type 1 (Tg1) domain (blue), and a C-terminal Tg1 domain (purple). Residue numbering is according to the mature protein. The connectivity of cysteine residues (highlighted) is indicated by pairs of black symbols below the sequence. The predicted N-glycosylation site is underlined. The schematic diagram shows the organization of the domains (N- and C-domains) with six disulfide bridges (the black lines indicate cysteine residue connectivity). The signal peptide, the unstructured region, and the two Tg1 domains were predicted using SignalP 6.0, PrDOS, and InterPro, respectively. **(B)** Purified recombinant full-length IrThy and its individual N- and C-terminal domains (IrThy-Nd and IrThy-Cd), produced in the insect cell system, were resolved by SDS-PAGE and visualized by protein staining. **(C)** Phylogenetic distribution of proteins containing two Tg1 domains in invertebrates (including ticks and other arthropods) and vertebrates (including mammals). The sequences from three searches in the InterPro-hosted Pfam database are specified as follows: Selections I and II—the sequence length is not restricted, and other domain types may (Selection I) or may not (Selection II) be present in the molecules; Selection III—the sequence length is restricted to 180 residues, corresponding to proteins with only two consecutive Tg1 domains, such as IrThy.

Recombinant full-length IrThy was prepared as a mature protein with a removable oligohistidine tag (ultimately cleaved by a TEV protease) and expressed using the S2 insect cell system. The same procedure was used to produce individual single domains of IrThy, namely the N-domain protein IrThy-Nd and the C-domain protein IrThy-Cd. For simplicity, interdomain disulfides linked to the central sequence Cys79–Lys80–Cys81 were replaced by intradomain disulfides spanning a sequence two amino acids shorter, namely Cys61–Cys79 in IrThy-Nd and Cys81–Cys103 in IrThy-Cd. These disulfides formed correctly in the expressed single-domain proteins, as confirmed by LC-MS/MS peptide mapping (Table S2). All three recombinants were purified to homogeneity from expression media (see Section 4). They migrated on SDS-PAGE as single bands corresponding to the

predicted molecular masses of 16.9, 9.6, and 7.9 kDa for IrThy, IrThy-Nd, and IrThy-Cd, respectively (Figure 2B). To increase the homogeneity of the recombinants for structural analysis, the N-glycosylation motif on the C-domain was mutated (Figure 2A); this motif is located on the opposite side of the domain from the predicted reactive center for protease binding (see Section 2.6.).

2.3. The Unique, Narrow Inhibitory Specificity of IrThy

Purified recombinant IrThy and its individual domains, IrThy-Nd and IrThy-Cd, were screened in vitro for inhibitory activity against two panels of proteases, including archetypal representatives of major protease classes and relevant cysteine cathepsins of the mammalian host. The latter panel, selected to cover a wide range of endo- and exopeptidase cleavage specificities of cysteine cathepsins, consisted of cathepsins F, K, L, S, and V (endopeptidases), cathepsins B and X (a carboxydipeptidase/endopeptidase and carboxypeptidase, respectively), and cathepsins C and H (an aminodipeptidase and aminopeptidase, respectively).

Inhibition constants (K_i) were determined using a fluorescence activity assay; the specificity profiles are compared in Table 1. No inhibition was observed for members of aspartic and serine proteases or for cysteine proteases of the CD clan. Inhibition was observed only for certain cysteine cathepsins with endopeptidase activities from the CA clan of cysteine proteases. In general, IrThy and its two individual domains selectively inhibited cathepsins V, K, and L as follows: K_i values were within the range of 27–41 nM and 44–57 nM for cathepsin V and cathepsin K, respectively; cathepsin L was more sensitive to inhibition by IrThy and IrThy-Nd (K_i of 201 and 179 nM, respectively) than by IrThy-Cd (K_i of 795 nM). However, IrThy-Nd exhibited a broader inhibitory specificity and, in addition to the three cathepsins mentioned above, it effectively inhibited two other proteases with endopeptidase activities from the CA clan, namely cathepsin F and the archetypal plant protease papain, with K_i values of 207 and 153 nM, respectively.

Table 1. Inhibitory specificity of IrThy and its individual domains. The inhibitory potency of full-length IrThy and its N- and C-terminal domains (IrThy-Nd and IrThy-Cd) was screened against mammalian host-derived cysteine cathepsins of the CA clan and archetypal representatives of protease clans. K_i values (mean \pm SD) were determined by a kinetic activity assay using specific fluorogenic peptide substrates (see Section 4). The Merops database classification of the tested proteases (class/clan) and their mode of action are given. n.i.—no significant inhibition at 2 μ M inhibitor concentration.

Enzymes	Enzyme Specificity, Protease Class/Clan	IrThy	K_i (nM) IrThy-Nd	IrThy-Cd
Host cathepsin proteases				
Cathepsin V	Endopeptidase, Cys/CA	34.7 \pm 2.9	27.2 \pm 3.1	40.7 \pm 3.8
Cathepsin K	Endopeptidase, Cys/CA	56.9 \pm 4.3	43.7 \pm 5.4	53.1 \pm 7.7
Cathepsin L	Endopeptidase, Cys/CA	201.4 \pm 15.8	178.6 \pm 6.7	795.1 \pm 19.8
Cathepsin F	Endopeptidase, Cys/CA	n.i.	207.2 \pm 14.2	n.i.
Cathepsin S	Endopeptidase, Cys/CA	n.i.	n.i.	n.i.
Cathepsin B	Endo- and carboxydipeptidase, Cys/CA	n.i.	n.i.	n.i.
Cathepsin X	Carboxypeptidase, Cys/CA	n.i.	n.i.	n.i.
Cathepsin C	Aminodipeptidase, Cys/CA	n.i.	n.i.	n.i.
Cathepsin H	Aminopeptidase, Cys/CA	n.i.	n.i.	n.i.
Model proteases				
Papain	Endopeptidase, Cys/CA	n.i.	153.4 \pm 11.1	n.i.
Legumain	Endopeptidase, Cys/CD	n.i.	n.i.	n.i.
Cathepsin D	Endopeptidase, Asp/AA	n.i.	n.i.	n.i.
Chymotrypsin	Endopeptidase, Ser/PA	n.i.	n.i.	n.i.
Trypsin	Endopeptidase, Ser/PA	n.i.	n.i.	n.i.

We demonstrated that IrThy is capable of inhibiting the activity of model target proteases across the entire pH activity range and, moreover, that the degree of inhibition is

pH-dependent (Figure 3A). The most effective inhibition was observed at a mildly acidic pH of ~6, with reduced inhibitory potency at more acidic and neutral/alkaline pH values. The regulatory effect of pH was more pronounced for cathepsin V than for cathepsin K. We also analyzed the inhibition mode of IrThy, classifying it as competitive, which suggests that IrThy binds in the protease active site as previously described for the p41 fragment [6]. The inhibition mode was determined using a Lineweaver–Burk plot with a series of straight lines intersecting on the $1/v$ axis (Figure 3B, C).

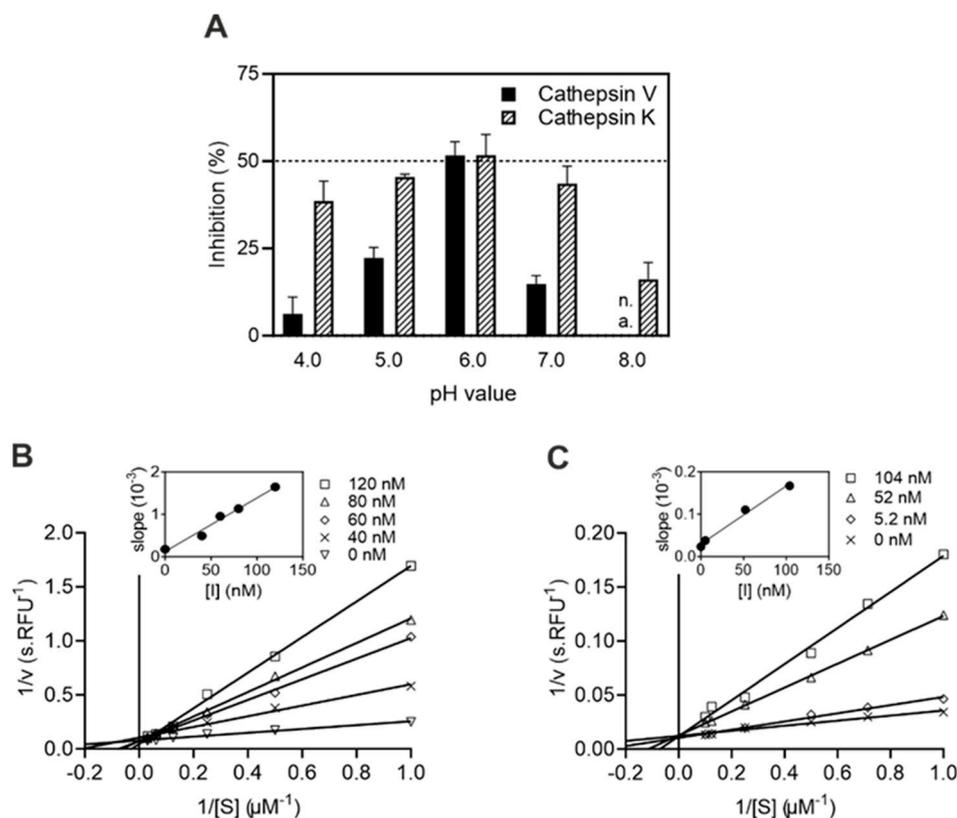


Figure 3. Characteristics of cysteine cathepsin inhibition by IrThy. (A) Effect of pH on the inhibitory potency of IrThy. The kinetic activity assay for cathepsins K and V with fluorogenic peptide substrates was performed at different pH values in the presence and absence of IrThy. The inhibitor was applied at a concentration providing ~50% inhibition at pH 6.0, and the % inhibition was calculated relative to an uninhibited control (0%) at the pH values indicated; means \pm SD are given. Note that the pH value of 8.0 is outside the functional range of cathepsin V (n.a.). (B,C) Competitive mode of inhibition by IrThy. Lineweaver–Burk plots are presented together with secondary plots of the same data (inset) for (B) cathepsin V and (C) cathepsin K. The kinetic activity assay with fluorogenic peptide substrates was performed at pH 5.5. Means \pm SD are given for triplicates.

In summary, we demonstrated that IrThy is a thyropin with a unique and narrow inhibitory specificity toward mammalian cysteine cathepsins. It targeted only the endopeptidase cathepsins V and K within nanomolar range, followed by cathepsin L; the other cathepsins with the spectrum of endopeptidase and exopeptidase activities proved insensitive to IrThy.

2.4. IrThy Is Prone to Proteolytic Degradation by Target Proteases at High Concentrations

The Tg1 domains proved susceptible to cleavage by cysteine cathepsins, as reported in [29]. Human testican, a cathepsin L inhibitor from the thyropin family, undergoes degradation by its target protease, cathepsin L, at high concentrations [12]. In this context, we chose to examine the sensitivity of IrThy to the proteolytic activity of cathepsins V, K, and L, which are selectively inhibited by IrThy (see Section 2.3). For comparison, we also

included an experiment with cathepsin B as a model endopeptidase/exopeptidase, which is not targeted by IrThy.

First, we applied the concentration conditions used in the kinetic assay for measuring the inhibitory potency of IrThy (Figure 4). No apparent degradation of IrThy was observed using these catalytic concentrations of cathepsins, with an enzyme:inhibitor ratio of 1:20,000 (*w/w*) (corresponding to 0.2 nM:6 μ M concentrations). On the contrary, incubation of IrThy with cathepsin concentrations increased by three orders of magnitude (1:5 ratio, *w/w*) (corresponding to 0.8 μ M:6 μ M concentrations) led to the substantial degradation of IrThy by all tested cathepsins. The time course of the degradation reaction is shown in Figure S3, which illustrates the generation of IrThy fragments. Furthermore, we used mass spectrometry for mapping the cleavage sites, which were found to be widely distributed on the IrThy molecule (Figure S4).

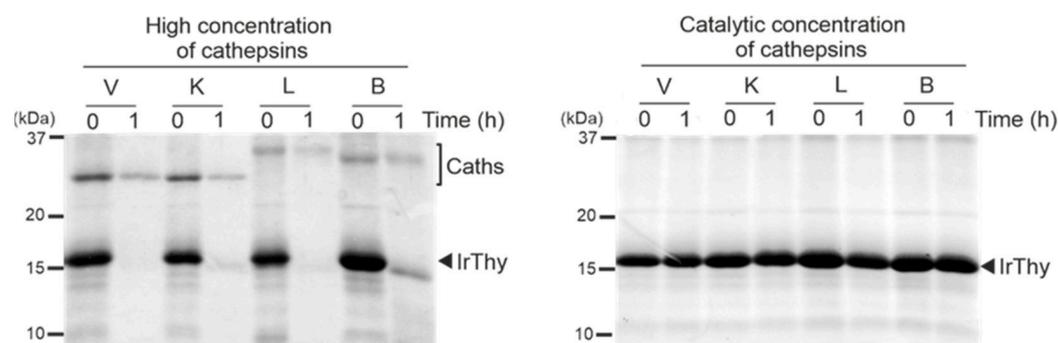


Figure 4. Proteolytic degradation of IrThy by target cathepsins V, K, and L, and a model non-target cathepsin B. IrThy was treated with high concentrations of cathepsins (an enzyme:inhibitor ratio of 1:5, *w/w*) or with catalytic concentrations of these cathepsins, corresponding to conditions in a kinetic inhibition assay (an enzyme:inhibitor ratio of 1:20,000, *w/w*). The reaction mixture was incubated at pH 5.5; the aliquots at time points 0 and 1 h were resolved by SDS–PAGE and visualized by protein staining. The positions of IrThy and cathepsins (Caths) are indicated.

To conclude, we provide evidence that IrThy can be proteolytically degraded by both its target and non-target cysteine cathepsins. We speculate that this may represent a concentration-dependent regulatory mechanism that controls the physiological inhibitory effect of IrThy.

2.5. Complex Glycosaminoglycan Modulation of IrThy Inhibitory Activity

As previously reported, GAGs, which are widely distributed in tissues, modulate interactions between several protease inhibitors of proteinaceous character and their target proteases, including cathepsins [30,31]. However, this relationship has not been investigated in thyropins. To address this shortcoming, we tested the effect of GAGs on IrThy action against cathepsins V, K, and L, which proved sensitive to IrThy inhibition (Table 1). For the screening, we used a panel of GAGs that included (i) heparin (17–19 kDa) and its disaccharide fragment (0.7 kDa), (ii) chondroitin-4-sulfate, chondroitin-6-sulfate, dermatan sulfate (all in the 20–60 kDa range), and (iii) dextran sulfate (9–20 kDa), which served as a model sulfated polysaccharide GAG analog. Individual GAGs at a concentration of 10 μ g/mL were incubated with IrThy and a target cathepsin. The resulting change in cathepsin activity was measured using a fluorescence activity assay and compared to the control experiment without GAG.

The results in Figure 5A show the dramatic effects of GAGs on the inhibition of cathepsins K and V. The inhibitory potency of IrThy against cathepsin K was substantially reduced from ~50% inhibition to ~7, 16, and 23% in the presence of heparin, chondroitin-4-sulfate, and chondroitin-6-sulfate, respectively, and was completely abolished by dermatan and dextran sulfates. The opposite effect was observed for cathepsin V, whose inhibitory potency was greatly enhanced from ~25% to within an inhibitory range of 55–70% in the

presence of all tested GAGs. The only exception was the heparin disaccharide fragment, which had no significant effect on IrThy inhibition of cathepsin V or the other cathepsins. In contrast to cathepsins K and V, the inhibition of cathepsin L by IrThy was not substantially affected by GAGs in general; changes in inhibitory potency were within 15% (Figure 5A).

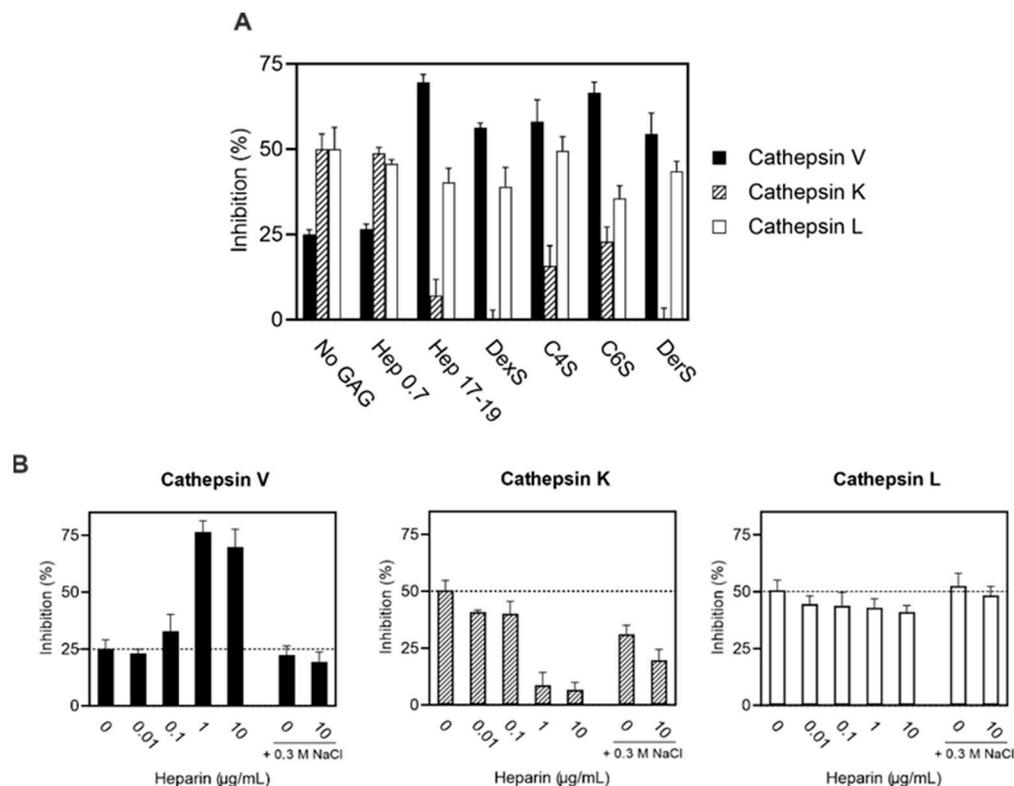


Figure 5. Effect of glycosaminoglycans (GAGs) on the inhibitory potency of IrThy. IrThy was applied at a concentration providing ~50% or ~25% inhibition in the absence of heparin. The kinetic activity assay with fluorogenic peptide substrates was performed at pH 5.5. The % inhibition was calculated relative to uninhibited controls (0%) with the same GAG concentrations (Figure S5); means \pm SD are given. Comparative experiments without GAG are indicated as No GAG and 0 heparin. (A) The inhibitory potency of IrThy against human cathepsins V, K, and L in the presence of different GAGs, including heparin (Hep 17-19) and its disaccharide fragment (Hep 0.7), chondroitin-4-sulfate (C4S), chondroitin-6-sulfate (C6S), dermatan sulfate (DerS), and the GAG analog dextran sulfate (DexS). GAGs were applied at 10 μ g/mL. (B) The inhibitory potency of IrThy against human cathepsins V, K, and L in the presence of various concentrations of heparin (0–10 μ g/mL) and 0.3 M NaCl (where indicated).

Figure 5B shows the dependence of the modulatory effect on the GAG concentration as demonstrated by heparin. For the inhibition of cathepsins K and V by IrThy, heparin was highly effective at concentrations of 1 μ g/mL and above, where the inhibitory potency of IrThy reached its minimum value against cathepsin K and its maximum value against cathepsin V. Heparin concentrations at and below 0.1 μ g/mL had a much less pronounced effect (<17% compared to the control without heparin). Inhibition of cathepsin L by IrThy, which exhibited low sensitivity to heparin, was associated with a decrease in inhibition of up to ~10%.

The electrostatic character of the heparin interactions involved in the modulatory effect was examined by increasing the ionic strength in the inhibition assay (Figure 5B). The presence of 0.3 M NaCl either substantially or completely suppressed the heparin-induced changes observed in the inhibitory potency of IrThy. In addition, the control experiment with NaCl and without heparin suggests that electrostatic interactions also contributed

to the interactions between IrThy and cathepsin K, as indicated by the ~20% reduction in inhibition.

In conclusion, GAGs modulate the inhibitory potency of IrThy against cysteine cathepsins in a complex manner. This modulatory effect is highly variable and depends on the type of target cathepsin, GAG, and GAG concentration employed. Most notably, IrThy inhibition of cathepsins K and V, both of which contain GAG-binding sites [32–35], was down- and upregulated by GAGs in opposite ways.

2.6. Spatial Structure of the IrThy C-Domain Determined by NMR Spectroscopy

To investigate the spatial structure of IrThy by NMR spectroscopy, we prepared full-length IrThy and its individual domains, IrThy-Nd and IrThy-Cd, in an *Escherichia coli* expression system for ^{13}C and ^{15}N isotope labeling. Correct protein folding was confirmed by comparing their 1D NMR spectra and inhibitory activity with recombinants produced in insect cells used in functional studies.

For IrThy and IrThy-Nd, the poor quality of all standard 3D spectra indicated that the internal dynamics of both proteins contributed to a faster relaxation of the NH resonances, preventing the acquisition of reliable data. In contrast, the 3D spectra of IrThy-Cd were of sufficient quality to enable the assignment of ~85% of the backbone atoms and ~64% of all the atoms. Of the 68 IrThy-Cd residues, resonances for 17 residues were not detected. These included a stretch of Lys88–Val96, which is probably flexible and lacks contacts with more rigid structural elements. The other residues without backbone assignment were found in the proximity of cysteines Cys115 (Thr117), Cys139 (Ser136, Gly137, and Thr140), and Cys147 (Arg144, Asp145, His146, and Cys147 itself); this can be attributed to the dynamics of disulfide conformation on a millisecond timescale, resulting in a decrease in signal intensities.

Interproton distance restraints were obtained from NOESY spectra (see Section 4). A dense network of inter-residue restraints was observed within the β -sheet-containing region Tyr109–Ala130; 18, 15, and 10 restraints were generated for residues Val123, Asp124, and Pro125, respectively. This points to a significant stabilization of this region, including the L2 loop (Figure 6B). As additional restraints, we also incorporated experimental mass spectrometry data on the pairing of cysteines in IrThy-Cd, namely Cys81–Cys103, Cys115–Cys139, and Cys122–Cys147, into the structure calculation protocol (Table S2). The final refinement yielded 30 solution structures of IrThy-Cd selected on the basis of low overall energy and minimal constraint violations (Figure 6C). Atomic coordinates and experimental constraints were deposited in the Protein Data Bank (accession code: 8R6T) and the Biological Magnetic Resonance Data Bank (BMRB) (accession code: 34883).

A representative spatial structure of IrThy-Cd is presented in Figure 6A. The molecule adopts a wedge-shaped fold typical of thyropins and similar to those of the human p41 fragment (PDB code: 1ICF [6]) and bullfrog saxiphilin (PDB code: 6O0F [8]), the only thyropins whose structures are known. Figure 6B,D shows the structure-based sequence alignment and structural comparison of IrThy-Cd with p41 and the saxiphilin domains 1 and 2; the C α RMSD values are within the range of 4.9–5.6 Å, and the sequence identities are within the range of 26–32%. While the architecture of IrThy-Cd is clearly similar to that of the other thyropins (Figure 6D), the high RMSD values can be attributed to the considerable flexibility of the loops and orientation of the structural regions. The fold is characterized by a conserved N-terminal α -helix (α 1), an antiparallel β -sheet (β 1– β 2), and three loops (L1–L3), which are involved in the protease interaction (Figure 6A). In addition, IrThy-Cd incorporates another short C-terminal α -helix (α 2). In terms of spatial arrangement, these segments constitute two subdomains: the first subdomain contains α 1 and L1, and the second the rest of the thyropin molecule [6].

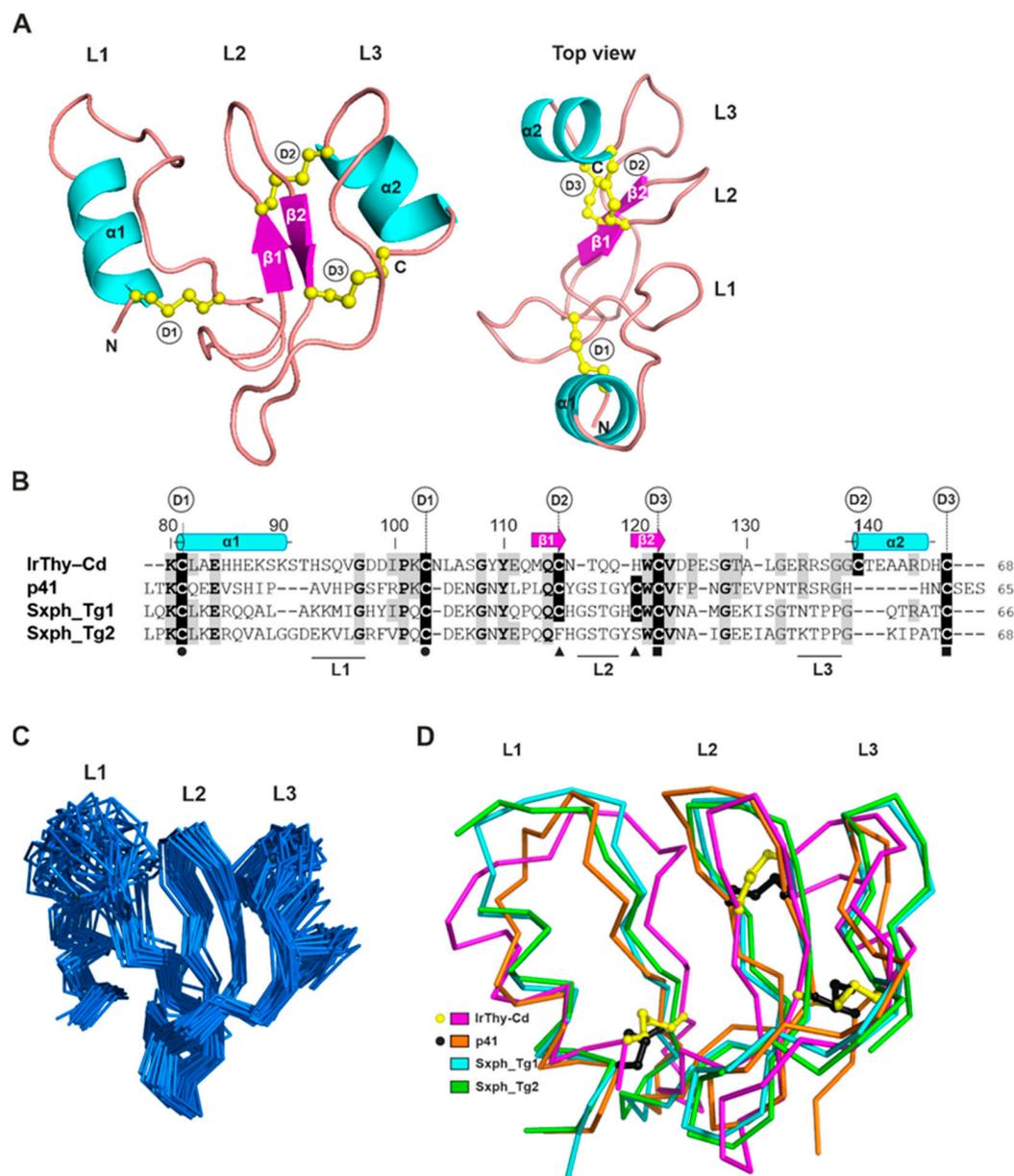


Figure 6. NMR solution structure of the IrThy C-domain (IrThy-Cd) and its comparison with other structurally characterized thyropins. **(A)** Representative three-dimensional structure of IrThy-Cd (PDB code: 8R6T) is depicted in a cartoon representation colored by secondary structural elements ($\alpha 1$ – 2 helices, cyan; $\beta 1$ – 2 strands, magenta). The N- and C-termini (N, C) and the three disulfide bridges (yellow sticks and balls) Cys81–Cys103 (D1), Cys115–Cys139 (D2), and Cys122–Cys147 (D3) are indicated. The L1, L2, and L3 loops are involved in the binding of the thyropin inhibitors to cysteine cathepsins [6]. **(B)** Structure-based sequence alignment of IrThy-Cd with structurally characterized thyropins, human p41 fragment (p41), and bullfrog saxiphilin domains (Sxph_Tg1 and Sxph_Tg2). Residues identical to those of IrThy-Cd are shaded in grey; fully conserved residues are in bold. Residue numbering is according to IrThy-Cd. Cysteine residues forming disulfide bridges are shaded in black and labeled D1–3 for IrThy-Cd; disulfide connectivity for the other thyropins is indicated by black circles, triangles, and squares below the alignment; note changes in the pairings compared to D2 of IrThy-Cd. The secondary structural elements of IrThy-Cd are depicted in cyan or magenta above the sequence and labeled the same as in **(A)**. Three regions topped with the L1, L2,

and L3 loops, which are responsible for the protease interaction, are indicated by horizontal black lines below the alignment based on the structure of the p41 complex with cathepsin L [6]. The alignment was generated using ClustalQ [36] and edited based on the structural superposition. (C) An ensemble of the 30 lowest-energy solution structures of IrThy-Cd in the ribbon representation is superimposed. The orientation of the molecule is the same as in (A) (left panel); the binding loops L1 to L3 are marked. (D) Overlay of IrThy-Cd (magenta) with structurally characterized thyropins in a ribbon representation, including the p41 fragment (orange, PDB code: 1ICF) and the saxiphilin domains 1 and 2 (cyan and green, respectively, PDB code: 6O0F). The positions of the disulfides are compared for IrThy-Cd (yellow sticks and balls) and p41 (black sticks and balls). The orientation of the molecules is the same as in (A) (left panel); the binding loops L1 to L3 are marked.

The disulfide bridges are critical to the overall fold of IrThy-Cd, considering the absence of a hydrophobic core, low content of secondary structures, and substantial conformational flexibility. Both IrThy-Cd subdomains are stabilized by the disulfide bridges Cys81–Cys103, Cys115–Cys139, and Cys122–Cys147, designated D1 through D3, respectively, in Figure 6A. Thus, IrThy-Cd conforms to the thyroglobulin subtype 1A domain, which has three disulfides [1]. In the first subdomain, the conserved D1 disulfide bridge near the N-terminus supports the formation of the L1 loop. The D2 and D3 bridges of IrThy-Cd, located in the second subdomain, connect the L2-containing segment with the L3-containing segment. Interestingly, the interloop disulfide bond D2 in IrThy-Cd is replaced by an intraloop disulfide in p41 and saxiphilin domain 1, where it forms a clamp within the L2 loop; this bridge is absent in saxiphilin domain 2. The higher number of interloop disulfides may provide greater rigidity for the second subdomain of IrThy-Cd.

Loops L1, L2, and L3 form the reactive centers of the thyropins (Figure 6A). The tripartite wedge-shaped edge fills the active site cleft of the target proteases, as previously reported for the crystal structure of the p41 fragment complex with cathepsin L [6]. As the lowest-energy NMR structures of IrThy-Cd demonstrate, the L1- and L3-containing segments are the most flexible parts of the unbound thyropin molecule in solution (Figure 6C). The large L1 loop of IrThy-Cd features the conserved residues Gly101 and Pro105 (Figure 6B), which may be important for protease recognition specificity, as previously proposed for the L1 loops of thyropins [8]. The L2 loop, rather rigid in IrThy-Cd, is stabilized by the β 1– β 2 sheet and two disulfides. The β 2 strand contains a part of the Trp–Cys–Val sequence motif conserved in the Tg1 domains (Figure 6B), which form the core of the second subdomain. The L3 loop is flanked by the D2 interloop disulfide followed by the C-terminal α 2 helix; both of these features are unique to IrThy-Cd among the known thyropin structures.

3. Discussion

In this study, we provide a comprehensive structural and functional analysis of IrThy, an unusual member of the thyropin family of protease inhibitors that is expressed and secreted by the salivary glands of the hard tick *I. ricinus*. In general, the scientific understanding of thyropins in invertebrates is limited. To our knowledge, this is the first study to characterize parasite-derived thyropin at the molecular level. The IrThy molecule consists of two consecutive Tg1 domains, the N-domain and the C-domain. We identified an atypical disulfide pattern within the C-domain and also between both domains, forming an interdomain linkage. Intriguingly, the IrThy molecule lacks the complexity of structural modules common to many other multidomain Tg1-containing proteins. Our phylogenetic distribution analysis revealed that these notably short thyropin molecules (up to 180 residues), containing only two tandem Tg1 domains, are widespread in arthropods (including ticks) and entirely absent in mammals.

IrThy is a potent inhibitor of only three human cysteine cathepsins, namely V, K, and L, which belong to the cathepsin L subfamily of endopeptidases. IrThy does not inhibit other related cathepsins that possess diverse endo- and exopeptidase activities. Therefore, IrThy possesses a unique, narrow inhibitory specificity compared to other thyropins. In

general, thyropins can be effective against endopeptidases, but also against exopeptidases. A wide range of cysteine cathepsins with endopeptidase activities, including cathepsins V, K, L, F, and S, are inhibited by the p41 fragment [5]. Inhibition of exopeptidases, such as cathepsins B and H, has been reported for the p41 fragment, equistatin, saxiphilin, and salmon egg thyropein [7,14,37,38].

The recombinant C-domain of IrThy (IrThy-Cd) exhibits the same inhibitory profile as full-length IrThy. We determined the spatial structure of IrThy-Cd by NMR spectroscopy, which allowed us to analyze the structural basis of the inhibitory specificity by comparing the IrThy-Cd structure with that of the p41 fragment complexed with cathepsin L [6]. The binding mode of the p41 fragment, which carries three inhibitory loops (L1, L2, and L3), enables important interactions with the S2–S1 subsite area as well as the surrounding loops located on the protease R-domain [6]. These interactions are understood to be negatively affected by two structural changes in IrThy-Cd, including the increased size of the L1 loop in the first subdomain and the shape of the second subdomain induced by the unique disulfide D2 (Figure 6B,D). In particular, these changes prevent binding to exopeptidases such as cathepsin H (an aminopeptidase) and cathepsin B (a carboxydipeptidase), whose partially obstructed active sites restrict access [39,40].

We observed that the recombinant N-domain (IrThy-Nd) displayed a broader selectivity pattern compared to the narrow inhibitory specificity of IrThy-Cd and the parental IrThy. In addition to cathepsins V, K and L, IrThy-Nd inhibited two other homologous proteases with endopeptidase activities, cathepsin F and papain. This suggests that the binding interactions of the N-domain are somewhat restricted by the closely positioned C-domain in the full-length IrThy scaffold. Nevertheless, our AlphaFold-derived model of IrThy (Figure S6) revealed that the two reactive centers of the N- and C-domains are oriented in opposite directions and can function simultaneously. A similar arrangement has been previously documented in the tandem Tg1 domains of saxiphilin, which bind two protease molecules with different inhibitory specificities [7].

Furthermore, we performed a preliminary cross-linking mass spectrometry (XL-MS) experiment with IrThy and cathepsin V to identify proximal structural regions in the inhibitor–protease complex (Table S3). Our results indicated that IrThy preferentially binds to the active site of cathepsin V (added in an equimolar amount) through the reactive center on the N-domain, which exhibits a lower K_i value than the C-domain. However, further comparative XL-MS data collected under different concentration conditions and with other target cathepsins is required in order to better understand the IrThy binding process and the discrimination between domains.

Our results show that several GAGs in their macromolecular forms modulated the inhibitory potency of IrThy in a complex manner. Widely distributed in host tissues, these sulfated polysaccharides regulate cysteine cathepsins, particularly their autocatalytic processing, pH stability, and interactions with specific macromolecular substrates [32,35,41–45]. In the presence of GAGs, the inhibitory potency of IrThy generally decreased against cathepsin K, increased against cathepsin V, and remained relatively unaffected against cathepsin L. As previously reported, GAGs participate in a variety of functional interactions with cathepsins K and V. For example, while the collagenolytic activity of cathepsin K requires interaction with chondroitin-4-sulfate, GAGs suppress the elastinolytic activities of both cathepsins K and V [32,35,46]. The molecular surfaces of cathepsins K and V contain positively charged patches for GAG binding [32–34,45]; however, these regions are less distributed in cathepsin L, leading to reduced interactions with GAGs [33]. We hypothesize that GAGs regulate IrThy inhibition by interacting with distinct GAG-binding sites on the target cathepsins, and that GAGs may also interact with positively charged regions found on the predicted surface model of IrThy, particularly on its N-domain (Figure S6). The arrangement of these intermolecular interactions can either facilitate or hinder IrThy binding.

Protease inhibitors from tick saliva injected into the host have been shown to modulate the host immune response and suppress blood clotting at the site of tick attachment and blood feeding [47]. To date, the only studied tick salivary inhibitors of cysteine cathepsins

are members of the cystatin family, which exert immunosuppressive and anti-inflammatory effects on the host [23–25]. A recent comparative analysis of inhibitory specificity revealed that tick salivary cystatins differ from other family members in their high affinity for endopeptidases and limited effect on exopeptidases [24]. This biochemical feature makes tick salivary cystatins functionally similar to IrThy, exhibiting narrow inhibitory specificity against cysteine cathepsins with endopeptidase activities. It is therefore reasonable to assume that IrThy may play a role in immunomodulation and, based on the expression dynamics, that it acts in the initial phases of the tick–host interaction. In this context, cathepsins V, K, and L targeted by IrThy are known to be involved in the immune response and inflammation and are also present in the host skin [48–53]. Another clue to the potential physiological function of IrThy comes from a recent report on U24-ctenitoxin-Pn1a, a neurotoxin found in spider venom, that targets voltage-gated sodium channels (VGSCs) [54]. Interestingly, this protein contains two tandem Tg1 domains and is a close homolog of IrThy. The spider VGSC toxins are generally antinociceptive in an animal model [55], and it is tempting to speculate about the potential analgesic effect of IrThy at the tick bite site. Our future research will focus on identifying the exact physiological role of IrThy in tick–host interactions.

4. Materials and Methods

4.1. Materials

All protease substrates were purchased from Bachem (Bubendorf, Switzerland) with the exception of Abz-Phe-Arg-Phe(NO₂)-OH from MP Biomedicals (Irvine, CA, USA). Human cathepsins L, K, and V were produced in the *Pichia pastoris* expression system as described previously [12,56,57]. Bovine cathepsins C and H were prepared as described in [58,59]; human cathepsin D was prepared as described in [60]. Bovine trypsin, chymotrypsin, and *Carica papaya* papain were purchased from Merck (Kenilworth, NJ, USA); human cathepsins F, B, and S were obtained from Enzo Life Sciences (New York, NY, USA); and human cathepsin X and legumain were obtained from R&D Systems (Minneapolis, MN, USA). GAGs were purchased from Sigma-Aldrich (St Louis, MO, USA), including heparin (H3393), chondroitin-4-sulfate (C9819), chondroitin-6-sulfate (C4384), dermatan sulfate (C3788), and dextran sulfate (D6924); heparin disaccharide (H1002) was purchased from Dextra (Reading, UK).

Ticks and Tick-Derived Materials

Adult *I. ricinus* ticks were collected in a forest near České Budějovice, Czech Republic. Ticks were kept at 24 °C and 95% humidity under a 15/9 h day/night regime. Adult *I. ricinus* females were fed on guinea pigs and either forcibly removed on the sixth day of feeding (half-fed) or allowed to fully feed (days 7 to 8). Saliva was collected from half-fed adult *I. ricinus* females as previously described [61]. Salivation was induced by applying pilocarpine solution to the tick scutum. Saliva was collected in 10-μL capillaries for 2 h in a wet chamber kept at 30 °C, pooled, and stored at –80 °C. The salivary glands were carefully removed without disrupting the epithelium and washed in phosphate-buffered saline (PBS). Tissue extracts (150 mg protein/mL) were prepared by homogenizing pooled tissues in 0.1 M sodium acetate pH 4.5, 1% CHAPS on ice. The extract was cleared by centrifugation (16,000× g, 10 min, 4 °C), filtered through Ultrafree MC 0.22 μm (Millipore, Bedford, MA, USA), and stored at –80 °C.

4.2. IrThy Cloning and Sequencing

Total RNA was isolated from the salivary glands of five adult *I. ricinus* females fed for six days, as previously described [62]. The RNA was reverse-transcribed into cDNA (0.5 μg RNA per 20 μL reaction) using the Transcriptor High-Fidelity cDNA Synthesis Kit (Roche Applied Science, Penzberg, Germany) with random hexamers and diluted 20 times in water. A 501 bp long fragment encoding IrThy was amplified from cDNA using specific forward 5'-atgctgaagcaagtatagtagtg-3' and reverse 5'-ctttcttgattagcagtggtcgc-3'

primers, which were designed based on a consensus sequence of thyroperin homologs from *I. ricinus* sialome [26,63]. The sequence was deposited in GenBank under accession number PP107940. The resulting amplicon was isolated from the gel and cloned into the pCR4-TOPO vector (TOPO TA Cloning Kit, Thermo Fisher Scientific, Waltham, MA, USA). The isolated plasmid DNA was sequenced using the sequencing primers provided in the kit; the IrThy coding sequence resulted from the sequencing of seven individual clones.

4.3. Production of Recombinant IrThy and its Domains in Insect Cells

S2 insect cells were used to produce full-length IrThy (residues 1–147) and its individual domains, IrThy-Nd (1–80) and IrThy-Cd (80–147). They were produced as fusion proteins containing an oligohistidine tag (6xHis) and a TEV protease cleavage site with an extension of Ser-Asn-Ala-Ala-Ser residues at the N-terminus. IrThy cDNA was PCR-amplified from the pCR4-TOPO vector containing the complete IrThy insert using the forward 5'-gctagcgtgccaactcgagtggcc-3' and reverse 5'-gcgggcgctaacagtggctgcgc-3' primers, which introduced the 5' NheI and 3' NotI restriction cloning sites (underlined). The PCR product was ligated into the pUC19 vector (Thermo Fisher Scientific) using SmaI restriction endonuclease. The N-glycosylation motif Asn-Leu-Thr106 was disrupted by Thr106 to Ala106 mutagenesis as described in [64]. pUC19 vectors containing individual single-domain inserts were prepared from full-length IrThy by inverse PCR [65] using the following primers: the forward 5'-tagggcgccagggtagcagctcg-3' and reverse 5'-cttgaggactcagctgcctcgatgg-3' primers for the C-domain deletion; the forward 5'-aagtgtctggcagagcatcagagaag-3' and reverse 5'-gctagcgggggatcctctagagtcg-3' primers for the N-domain deletion; all primers were 5'-phosphorylated. The resulting vectors were digested with NheI and NotI restriction endonucleases and ligated into the pMT/BiP/V5-His A plasmid (Thermo Fisher Scientific), which was modified to contain at the N-terminus a NheI restriction site and a sequence encoding the BiP signal sequence, His6 tag, and TEV cleavage site.

Stably transfected insect cells were prepared as follows: Schneider 2 (S2) cells (Thermo Fisher Scientific) were cultured at 28 °C in Sf-900 II SFM medium (Thermo Fisher Scientific) containing 10% fetal bovine serum (Gibco, Thermo Fisher Scientific). Cells were co-transfected with the pMT/BiP/V5 vector containing IrThy-derived inserts (19 ng) and the pCoBlast selection vector containing the blasticidin resistance gene (1 ng) (Thermo Fisher Scientific) using the Calcium Phosphate Transfection Kit (Sigma Aldrich) according to the Drosophila Expression System protocol (Thermo Fisher Scientific). Stably transfected cells were selected with blasticidin (25 µg/mL) two weeks after transfection. S2 transfectants were expanded to a large volume (1.5 L) and cultured in Sf-900 II SFM medium to a density of 10⁷ cells/mL in a 6 L spinner flask, stirred at 120 rpm at 28 °C. Expression was induced with 1 mM CuSO₄, and after 10 days, cells were removed by centrifugation (1000× g, 5 min, 10 °C). The supernatant was concentrated by lyophilization to a final volume of 150 mL and desalted on a Sephadex G25 column equilibrated with 10 mM Tris-HCl pH 8.0.

Recombinant His-tagged proteins were purified by Ni²⁺ affinity chromatography on a HiTrap IMAC HP column (Cytiva, Marlborough, MA, USA) following the manufacturer's protocol for elution with imidazole at pH 8.0. The His-tags were cleaved by TEV protease (enzyme:protein ratio 1:10, w/w) in 50 mM Tris-HCl pH 8.0, 0.5 mM EDTA, and 1 mM DTT for 12 h at 26 °C and removed by Ni²⁺ affinity chromatography on a HiTrap IMAC HP column (see above). This was followed by size-exclusion chromatography on a HiLoad 16/600 Superdex 75 column (Cytiva) equilibrated with 50 mM Tris-HCl, 300 mM NaCl, pH 8.0 (for IrThy and IrThy-Cd) or pH 7.0 (for IrThy-Nd). Protein purity was monitored by Laemmli SDS-PAGE on 15% polyacrylamide gels stained with Coomassie Brilliant Blue R-250. The purified proteins were buffer-exchanged into 20 mM Tris-HCl, 20 mM NaCl, pH 8.0 (for IrThy and IrThy-Cd) or pH 7.0 (for IrThy-Nd), concentrated using an Amicon Ultra-3K centrifugal unit (Millipore), and stored at −80 °C. These recombinants were used in all experiments except for the NMR structure determination.

4.4. Production of Recombinant IrThy and its Domains in *E. coli*

The pUC19 vectors containing IrThy, IrThy-Nd, or IrThy-Cd inserts (see above) were digested with NcoI and NotI restriction endonucleases and ligated into the pETM-60 plasmid (EMBL, Heidelberg, Germany) in a frame with the NusA chaperone protein, His6-tag, and TEV protease cleavage site. Recombinant proteins were produced in SHuffle T7 Competent *E. coli* cells (New England BioLabs, Ipswich, MA, USA). Bacterial cultures were grown in ^{15}N or $^{13}\text{C}/^{15}\text{N}$ isotope-containing optimized M9 minimal media (0.5 L) with 50 $\mu\text{g}/\text{mL}$ kanamycin to an OD_{600} of 0.6. Expression was induced by the addition of isopropyl 1-thio- β -D-galactopyranoside to a final concentration of 0.4 mM. Cells were harvested by centrifugation ($4000\times g$, 10 min, 4 $^{\circ}\text{C}$) after 16 h of incubation at 16 $^{\circ}\text{C}$ with shaking at 200 rpm. Harvested *E. coli* cells were lysed using an EmulsiFlex-C3 homogenizer (Avestin, Ottawa, ON, Canada), and the lysate was clarified by centrifugation ($12,000\times g$, 10 min, 4 $^{\circ}\text{C}$). The recombinant His-tagged proteins were purified by Ni^{2+} affinity chromatography on a HiTrap IMAC HP column (Cytiva) using the manufacturer's protocol for elution with imidazole at pH 8.0. The fusion tags were cleaved by TEV protease (see above) and removed by Ni^{2+} affinity chromatography on a HiTrap IMAC HP column (see above). The purified proteins were buffer-exchanged into 50 mM Tris-HCl pH 8.0, 150 mM NaCl, concentrated using an Amicon Ultra-3K centrifugal unit (Millipore), and stored at -80°C . Protein purity was monitored by Laemmli SDS-PAGE (see above).

4.5. Expression Analysis by Quantitative Real-Time PCR

Total RNA was isolated from adult *I. ricinus* females during different stages of blood feeding and from different tissues of females fed for six days, as previously described [62]. The IrThy gene-specific qRT-PCR primers were designed using Primer3. The expression level of IrThy mRNA was measured by qRT-PCR using LightCycler 480 (Roche Applied Science) and SYBR green chemistry [62], with the forward 5'-atgagcaaatgcagtgcac-3' and reverse 5'-gcggttcactttctggat-3' primers. Relative expression was normalized to *I. ricinus ferritin1* (AF068224) [66] according to the mathematical model of Pfaffl [67]. Data were obtained from three independent biological replicates (five females per replicate). Statistical significance was analyzed using the unpaired Student's *t*-test and the one-way ANOVA test.

4.6. Mass Spectrometry Proteomic Analysis

For the identification of native IrThy, proteins from the salivary gland extract and saliva of adult *I. ricinus* female were reduced, alkylated, and digested with trypsin. LC-MS/MS analysis of the digests was performed on an Ulti-Mate 3000 RSLCnano system (Dionex, Thermo Fisher Scientific) coupled to a Triple-TOF 5600 mass spectrometer equipped with a NanoSpray III source (ABSciex, Framingham, MA, USA). The peptides were separated on an Acclaim Pep-Map100 analytical column (3 μm , 150×0.75 mm; Thermo Fisher Scientific) by gradient elution in a 0.1% formic acid-acetonitrile mobile phase. Full mass spectrometry scans were recorded from 350 to 1250 m/z ; up to 25 candidate ions per cycle were subjected to fragmentation; in MS/MS mode, fragmentation spectra were acquired from 100 to 1600 m/z . Mass data were processed using ProteinPilot 4.5 software (ABSciex).

For disulfide pairing analysis, recombinant IrThy, IrThy-Nd, or IrThy-Cd was digested with trypsin or Asp-N in the presence of cystamine; LC-MS/MS analysis of the digests was performed according to the protocol [68] described in Table S2.

4.7. Phylogenetic Distribution Analysis

Phylogenetic distribution was analyzed for proteins with Tg1 domains (Pfam entry: PF00086) in the InterPro-hosted Pfam database [28]. Proteins containing two Tg1 domains and meeting the additional criteria specified in Section 2.2. were filtered using Python scripts.

4.8. Protease Inhibition Assays

Inhibition measurements were performed in triplicates in 96-well microplates (100 μ L assay volume) at 37 °C. Recombinant inhibitors were preincubated with protease for 15 min, followed by the addition of a specific fluorogenic substrate (see below). The kinetics of product release were continuously monitored using an Infinite M1000 (Tecan, Männedorf, Switzerland) microplate reader at 365 nm excitation and 450 nm emission wavelengths (for AMC-containing substrates) or at 330 nm excitation and 410 nm emission wavelengths (for Abz-containing substrates). The progress curves of kinetic measurements were linear, indicating that the analysis was not affected by inhibitor degradation. IC₅₀ values were determined from residual velocities using dose–response plots; nonlinear regression was fitted using GraFit software Version 7 (Erithacus, East Grinstead, UK), and the inhibition constants K_i were calculated using the Cheng–Prusoff equation [41,60].

Substrates and proteases were applied in assays as follows: Z-Phe-Arg-AMC in a concentration of 2.5 μ M with 0.2 nM cathepsin L; 4.3 μ M with 0.2 nM cathepsin K; 3.6 μ M with 0.1 nM cathepsin V; 53 μ M with 0.2 nM cathepsin B; 27 μ M with 1.2 nM cathepsin F; 100 μ M with 0.5 nM papain; 17 μ M Z-Val-Val-Arg-AMC with 0.1 nM cathepsin S; 100 μ M Arg-AMC with 1.9 nM cathepsin H; 87 μ M Abz-Phe-Arg-Phe(NO₂)-OH with 1.2 nM cathepsin X; 130 μ M Gly-Arg-AMC with 0.4 nM cathepsin C; 70 μ M Z-Ala-Ala-Asn-AMC with 1.2 nM legumain; 33 μ M Abz-Lys-Pro-Ala-Glu-Phe-Phe(NO₂)-Ala-Leu-NH₂ with 3.5 nM cathepsin D; 67.8 μ M Z-Phe-Arg-AMC with 0.6 nM trypsin; and 41.5 μ M Suc-Ala-Ala-Phe-Arg-AMC with 0.02 nM chymotrypsin. The assay buffers were as follows: 50 mM sodium acetate pH 4.0 (for cathepsin D), pH 5.0 (for cathepsin X and legumain), or pH 5.5 (for cathepsins L, K, V, B, F, C, and papain); 50 mM MES pH 6.5 (for cathepsins S and H); and 50 mM Tris-HCl pH 8.0 (for trypsin and chymotrypsin). The buffers contained 0.1% polyethylene glycol 1500 and 2.5 mM dithiothreitol (for cysteine cathepsins and legumain), 50 mM NaCl (for cathepsin C), and 10 mM CaCl₂ (for trypsin and chymotrypsin).

The inhibition mode was determined using an analogous activity assay with substrate concentrations of 1–10 μ M for cathepsin K and 1–32 μ M for cathepsin V; initial velocities of the product release were interpreted using a Lineweaver–Burk plot. The effect of pH on the inhibitory potency of IrThy was measured with cathepsins K and V (as described above) in 100 mM Britton–Robinson buffers within a pH range of 4.0–8.0. The effect of GAGs on the inhibitory potency of IrThy was measured with cathepsins K, L, and V (as described above) in the presence of various GAGs added to the preincubation mixture at a final concentration of 10 μ g/mL. Heparin was also tested in a concentration range of 0.01–10 μ g/mL. Where indicated, incubation was performed in the presence of 0.3 M NaCl.

4.9. Proteolytic Degradation of IrThy

Recombinant IrThy was incubated with cathepsins B, K, L, or V in 50 mM sodium acetate pH 5.5, 2.5 mM DTT, 0.1% polyethylene glycol 1500 at 37 °C. The reaction was stopped by the addition of 10 μ M E-64 at different time points. A 100 μ L aliquot contained 10 μ g of IrThy (final concentration of 6 μ M) and 2 μ g or 0.5 ng of cathepsin (final concentrations of 0.8 μ M or 0.2 nM, respectively). The reaction mixture was acetone-precipitated, separated by Laemmli SDS-PAGE (15% polyacrylamide gel), and stained with Coomassie Brilliant Blue R-250.

4.10. NMR Spectroscopy

NMR spectra were acquired at 25 °C using an 850 MHz Bruker Avance spectrometer equipped with a triple resonance (¹⁵N/¹³C/¹H) cryoprobe. The protein sample (0.35 mL in a Shigemi tube) was measured in 50 mM deuterated Tris-HCl pH 8.0, 150 mM NaCl, and 5% D₂O/95% H₂O. A series of double- (Figure S7) and triple-resonance spectra [69,70] were recorded to obtain sequence-specific resonance assignment in NMRFAM-Sparky [71]. ¹H–¹H distance restraints were derived from spectra obtained by 3D ¹⁵N/¹H NOESY-HSQC and ¹³C/¹H NOESY-HMQC spectra using a NOE mixing time of 100 ms. The structure calculation was performed in CYANA [72] using NOESY data in combination

with backbone torsion angle restraints generated from assigned chemical shifts using the TALOS+ program [73].

The combined automated NOE assignment and structure determination protocol (CANDID) was used for automatic NOE cross-peak assignment. Five cycles of simulated annealing combined with redundant dihedral angle restraints were then used to calculate a set of converged structures with no significant restraint violations (distance and van der Waals violations $< 0.5 \text{ \AA}$ and dihedral angle constraint violations $< 5^\circ$). The 30 structures with the least restraint violations were further analyzed using the Protein Structure Validation Software suite [www.nesg.org] (accessed on 11 December 2023).

Statistics for the resulting structure are summarized in Table S4. The structures, NMR restraints, and resonance assignments were deposited in the PDB (accession code: 8R6T) and BMRB (accession code: 34883) databases. Figures with structural representations were generated using PyMOL v2.3.2 (Schrödinger, New York, NY, USA) [74]. The structural comparison of IrThy-Cd with other thyropins provided the following $C\alpha$ RMSD values (number of aligned $C\alpha$ atoms is indicated): 5.36 \AA (58 atoms) for the p41 fragment, 4.92 \AA (50 atoms) for saxiphilin domain 1, and 5.62 \AA (49 atoms) for saxiphilin domain 2.

5. Conclusions

Ticks are parasite vectors for a variety of viral and bacterial diseases in humans and domestic animals. Our study focuses on *Ixodes ricinus* ticks, vectors of Lyme borreliosis and tick-borne encephalitis. Tick saliva injected into the host modulates the physiological response at the tick bite site and facilitates the transmission of pathogens. Protease inhibitors from tick saliva are important effector proteins in this process and thus represent promising vaccination targets and pharmacological agents. Here, we characterize at the molecular level a new protease inhibitor, IrThy, found in *I. ricinus* saliva. It belongs to the thyropin family, which is poorly understood in parasites and pathogens. We provide a comprehensive analysis of IrThy, including its unique inhibitory specificity against only three human cysteine proteases, cathepsins V, K, and L, involved in immunity and inflammation, and the NMR structure that explains the functional properties. Based on the results obtained, we propose potential physiological roles for IrThy in host–parasite interaction at the tick bite site.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25042240/s1> References [6,24,68,75–79] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, M.M. and Z.M.; investigation, Z.M., K.O., P.S., J.P., Z.K., M.B., O.H., R.Š. and M.F.; writing—original draft preparation, Z.M., M.M., K.O., P.S. and M.B.; writing—review and editing, M.M., Z.M., M.H. and K.O.; visualization, Z.M., K.O. and M.M.; supervision, M.M., M.H., P.K., and P.N.; funding acquisition, M.M., P.K.; O.H., R.Š. and M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grant 21-08826S and 22-27695S from the Czech Science Foundation (CSF) (to P.K. and M.M.) and Institutional project RVO 61388963. O.H. and P.K. were supported by the Centre for Research of Pathogenicity and Virulence of Parasites (CZ.02.1.01/0.0/0.0/16_019/0000759) from the European Regional Development Fund (ERDF) and the Ministry of Education, Youth and Sports of the Czech Republic (MEYS); R.S. was supported by grant 22-30920S from the CSF; Z.K. and P.N. were supported by grant 22-27695S from the CSF, and M.H. was supported by grant LUC23037 from the MEYS.

Institutional Review Board Statement: All laboratory animals were treated in accordance with the Animal Protection Law of the Czech Republic No. 246/1992 Sb., ethics approval No. 25/2018. The study was approved by the Institute of Parasitology, Biology Centre of the Czech Academy of Sciences and Central Committee for Animal Welfare, Czech Republic (protocol no. 1/2015).

Data Availability Statement: Atomic coordinates and experimental constraints were deposited in the Protein Data Bank (accession code: 8R6T) and the Biological Magnetic Resonance Data Bank (BMRB) (accession code: 34883).

Acknowledgments: The authors thank Lenka Grunclová, Daniel Sojka, and Jakub Benýšek for providing clone and enzyme products used in the study; Kateřina Faltejsková for phylogenetic analysis; Martin Hubálek for mass spectrometry analysis; Milana Štajflová and Anna Fuller for technical assistance; and Michael FitzGerald for language editing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Molina, F.; Bouanani, M.; Pau, B.; Granier, C. Characterization of the type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. *Eur. J. Biochem.* **1996**, *240*, 125–133. [[CrossRef](#)] [[PubMed](#)]
2. Lenarcic, B.; Bevec, T. Thyropins—New structurally related proteinase inhibitors. *Biol. Chem.* **1998**, *379*, 105–111.
3. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2017**, *46*, D624–D632. [[CrossRef](#)]
4. Mihelic, M.; Turk, D. Two decades of thyroglobulin type-1 domain research. *Biol. Chem.* **2007**, *388*, 1123–1130. [[CrossRef](#)] [[PubMed](#)]
5. Mihelic, M.; Dobersek, A.; Guncar, G.; Turk, D. Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation. *J. Biol. Chem.* **2008**, *283*, 14453–14460. [[CrossRef](#)]
6. Guncar, G.; Pungercic, G.; Klemencic, I.; Turk, V.; Turk, D. Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J.* **1999**, *18*, 793–803. [[CrossRef](#)] [[PubMed](#)]
7. Lenarcic, B.; Krishnan, G.; Borukhovich, R.; Ruck, B.; Turk, V.; Moczydlowski, E. Saxiphilin, a saxitoxin-binding protein with two thyroglobulin type 1 domains, is an inhibitor of papain-like cysteine proteinases. *J. Biol. Chem.* **2000**, *275*, 15572–15577. [[CrossRef](#)]
8. Yen, T.J.; Lolicato, M.; Thomas-Tran, R.; Du Bois, J.; Minor, D.L. Structure of the saxiphilin: Saxitoxin (STX) complex reveals a convergent molecular recognition strategy for paralytic toxins. *Sci. Adv.* **2019**, *5*, eaax2650. [[CrossRef](#)]
9. Lenarcic, B.; Turk, V. Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. *J. Biol. Chem.* **1999**, *274*, 563–566. [[CrossRef](#)]
10. Galesa, K.; Pain, R.; Jongsma, M.A.; Turk, V.; Lenarcic, B. Structural characterization of thyroglobulin type-1 domains of equistatin. *FEBS Lett.* **2003**, *539*, 120–124. [[CrossRef](#)]
11. Bockock, J.P.; Edgell, C.J.; Marr, H.S.; Erickson, A.H. Human proteoglycan testican-1 inhibits the lysosomal cysteine protease cathepsin L. *Eur. J. Biochem.* **2003**, *270*, 4008–4015. [[CrossRef](#)] [[PubMed](#)]
12. Meh, P.; Pavsic, M.; Turk, V.; Baici, A.; Lenarcic, B. Dual concentration-dependent activity of thyroglobulin type-1 domain of testican: Specific inhibitor and substrate of cathepsin L. *Biol. Chem.* **2005**, *386*, 75–83. [[CrossRef](#)] [[PubMed](#)]
13. Sankpal, N.V.; Brown, T.C.; Fleming, T.P.; Herndon, J.M.; Amaravati, A.A.; Loynd, A.N.; Gillanders, W.E. Cancer-associated mutations reveal a novel role for EpCAM as an inhibitor of cathepsin-L and tumor cell invasion. *BMC Cancer* **2021**, *21*, 541. [[CrossRef](#)] [[PubMed](#)]
14. Yamashita, M.; Konagaya, S. A novel cysteine protease inhibitor of the egg of chum salmon, containing a cysteine-rich thyroglobulin-like motif. *J. Biol. Chem.* **1996**, *271*, 1282–1284. [[CrossRef](#)] [[PubMed](#)]
15. Fowlkes, J.L.; Thrailkill, K.M.; Serra, D.M.; Nagase, H. Insulin-like growth factor binding protein (IGFBP) substrate zymography. A new tool to identify and characterize IGFBP-degrading proteinases. *Endocrine* **1997**, *7*, 33–36. [[CrossRef](#)] [[PubMed](#)]
16. Moreno, M.J.; Ball, M.; Rukhlova, M.; Slinn, J.; L'Abbe, D.; Iqbal, U.; Monette, R.; Hagedorn, M.; O'Connor-McCourt, M.D.; Durocher, Y.; et al. IGFBP-4 anti-angiogenic and anti-tumorigenic effects are associated with anti-cathepsin B activity. *Neoplasia* **2013**, *15*, 554–567. [[CrossRef](#)] [[PubMed](#)]
17. Fiebiger, E.; Maehr, R.; Villadangos, J.; Weber, E.; Erickson, A.; Bikoff, E.; Ploegh, H.L.; Lennon-Duménil, A.M. Invariant chain controls the activity of extracellular cathepsin L. *J. Exp. Med.* **2002**, *196*, 1263–1269. [[CrossRef](#)]
18. Porter, L.M.; Radulović, Ž.M.; Mulenga, A. A repertoire of protease inhibitor families in *Amblyomma americanum* and other tick species: Inter-species comparative analyses. *Parasit. Vectors* **2017**, *10*, 152. [[CrossRef](#)]
19. Oliveira, C.J.; Anatriello, E.; de Miranda-Santos, I.K.; Francischetti, I.M.; Sá-Nunes, A.; Ferreira, B.R.; Ribeiro, J.M. Proteome of *Rhipicephalus sanguineus* tick saliva induced by the secretagogues pilocarpine and dopamine. *Ticks Tick Borne Dis.* **2013**, *4*, 469–477. [[CrossRef](#)]
20. Bensaoud, C.; Tenzer, S.; Poplawski, A.; Medina, J.M.; Jmel, M.A.; Voet, H.; Mekki, I.; Aparicio-Puerta, E.; Cuveele, B.; Distler, U.; et al. Quantitative proteomics analysis reveals core and variable tick salivary proteins at the tick-vertebrate host interface. *Mol. Ecol.* **2022**, *31*, 4162–4175. [[CrossRef](#)]
21. Kozelková, T.; Dyčka, F.; Lu, S.; Urbanová, V.; Frantová, H.; Sojka, D.; Šíma, R.; Horn, M.; Perner, J.; Kopáček, P. Insight Into the Dynamics of the *Ixodes ricinus* Nymphal Midgut Proteome. *Mol. Cell Proteom.* **2023**, *22*, 100663. [[CrossRef](#)]
22. Kotál, J.; Buša, M.; Urbanová, V.; Řezáčová, P.; Chmelař, J.; Langhansová, H.; Sojka, D.; Mareš, M.; Kotsyfakis, M. Mialostatin, a Novel Midgut Cystatin from *Ixodes ricinus* Ticks: Crystal Structure and Regulation of Host Blood Digestion. *Int. J. Mol. Sci.* **2021**, *22*, 5371. [[CrossRef](#)]

23. Kotál, J.; Stergiou, N.; Buša, M.; Chlastáková, A.; Beránková, Z.; Řezáčová, P.; Langhansová, H.; Schwarz, A.; Calvo, E.; Kopecký, J.; et al. The structure and function of Iristatin, a novel immunosuppressive tick salivary cystatin. *Cell Mol. Life Sci.* **2019**, *76*, 2003–2013. [[CrossRef](#)]
24. Martins, L.A.; Buša, M.; Chlastáková, A.; Kotál, J.; Beránková, Z.; Stergiou, N.; Jmel, M.A.; Schmitt, E.; Chmelař, J.; Mareš, M.; et al. Protease-bound structure of Ricistatin provides insights into the mechanism of action of tick salivary cystatins in the vertebrate host. *Cell Mol. Life Sci.* **2023**, *80*, 339. [[CrossRef](#)]
25. Kotsyfakis, M.; Sá-Nunes, A.; Francischetti, I.M.; Mather, T.N.; Andersen, J.F.; Ribeiro, J.M. Antiinflammatory and immunosuppressive activity of sialostatin L, a salivary cystatin from the tick *Ixodes scapularis*. *J. Biol. Chem.* **2006**, *281*, 26298–26307. [[CrossRef](#)]
26. Schwarz, A.; von Reumont, B.M.; Erhart, J.; Chagas, A.C.; Ribeiro, J.M.; Kotsyfakis, M. De novo *Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. *FASEB J.* **2013**, *27*, 4745–4756. [[CrossRef](#)]
27. Perner, J.; Provazník, J.; Schrenková, J.; Urbanová, V.; Ribeiro, J.M.; Kopáček, P. RNA-seq analyses of the midgut from blood- and serum-fed *Ixodes ricinus* ticks. *Sci. Rep.* **2016**, *6*, 36695. [[CrossRef](#)] [[PubMed](#)]
28. Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Bork, P.; Bridge, A.; Colwell, L.; et al. InterPro in 2022. *Nucleic Acids Res.* **2022**, *51*, D418–D427. [[CrossRef](#)] [[PubMed](#)]
29. Pungercic, G.; Dolenc, I.; Dolinar, M.; Bevec, T.; Jenko, S.; Kolaric, S.; Turk, V. Individual recombinant thyroglobulin type-1 domains are substrates for lysosomal cysteine proteinases. *Biol. Chem.* **2002**, *383*, 1809–1812. [[CrossRef](#)] [[PubMed](#)]
30. Denamur, S.; Chazeirat, T.; Maszota-Zieleniak, M.; Vivès, R.R.; Saidi, A.; Zhang, F.; Linhardt, R.J.; Labarthe, F.; Samsonov, S.A.; Lalmanach, G.; et al. Binding of heparan sulfate to human cystatin C modulates inhibition of cathepsin L: Putative consequences in mucopolysaccharidosis. *Carbohydr. Polym.* **2022**, *293*, 119734. [[CrossRef](#)]
31. Rein, C.M.; Desai, U.R.; Church, F.C. Serpin-glycosaminoglycan interactions. *Methods Enzymol.* **2011**, *501*, 105–137. [[CrossRef](#)] [[PubMed](#)]
32. Li, Z.; Kienetz, M.; Cherney, M.M.; James, M.N.G.; Brömme, D. The Crystal and Molecular Structures of a Cathepsin K: Chondroitin Sulfate Complex. *J. Mol. Biol.* **2008**, *383*, 78–91. [[CrossRef](#)] [[PubMed](#)]
33. David, A.; Chazeirat, T.; Saidi, A.; Lalmanach, G.; Lecaille, F. The Interplay of Glycosaminoglycans and Cysteine Cathepsins in Mucopolysaccharidosis. *Biomedicines* **2023**, *11*, 810. [[CrossRef](#)] [[PubMed](#)]
34. Chazeirat, T.; Denamur, S.; Bojarski, K.K.; Andrault, P.M.; Sizaret, D.; Zhang, F.; Saidi, A.; Tardieu, M.; Linhardt, R.J.; Labarthe, F.; et al. The abnormal accumulation of heparan sulfate in patients with mucopolysaccharidosis prevents the elastolytic activity of cathepsin V. *Carbohydr. Polym.* **2021**, *253*, 117261. [[CrossRef](#)] [[PubMed](#)]
35. Aguda, A.H.; Panwar, P.; Du, X.; Nguyen, N.T.; Brayer, G.D.; Brömme, D. Structural basis of collagen fiber degradation by cathepsin K. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17474–17479. [[CrossRef](#)] [[PubMed](#)]
36. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [[CrossRef](#)] [[PubMed](#)]
37. Bevec, T.; Stoka, V.; Pungercic, G.; Dolenc, I.; Turk, V. Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. *J. Exp. Med.* **1996**, *183*, 1331–1338. [[CrossRef](#)]
38. Lenarcic, B.; Ritonja, A.; Strukelj, B.; Turk, B.; Turk, V. Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain. *J. Biol. Chem.* **1997**, *272*, 13899–13903. [[CrossRef](#)]
39. Musil, D.; Zucic, D.; Turk, D.; Engh, R.A.; Mayr, I.; Huber, R.; Popovic, T.; Turk, V.; Towatari, T.; Katunuma, N.; et al. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: The structural basis for its specificity. *EMBO J.* **1991**, *10*, 2321–2330. [[CrossRef](#)]
40. Guncar, G.; Podobnik, M.; Pungercar, J.; Strukelj, B.; Turk, V.; Turk, D. Crystal structure of porcine cathepsin H determined at 2.1 Å resolution: Location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure* **1998**, *6*, 51–61. [[CrossRef](#)]
41. Horn, M.; Jílková, A.; Vondrášek, J.; Marešová, L.; Caffrey, C.R.; Mareš, M. Mapping the pro-peptide of the *Schistosoma mansoni* cathepsin B1 drug target: Modulation of inhibition by heparin and design of mimetic inhibitors. *ACS Chem. Biol.* **2011**, *6*, 609–617. [[CrossRef](#)]
42. Jílková, A.; Horn, M.; Řezáčová, P.; Marešová, L.; Fajtová, P.; Brynda, J.; Vondrášek, J.; McKerrow, J.H.; Caffrey, C.R.; Mareš, M. Activation route of the *Schistosoma mansoni* cathepsin B1 drug target: Structural map with a glycosaminoglycan switch. *Structure* **2014**, *22*, 1786–1798. [[CrossRef](#)]
43. Caglic, D.; Pungercar, J.R.; Pejler, G.; Turk, V.; Turk, B. Glycosaminoglycans facilitate procathepsin B activation through disruption of propeptide-mature enzyme interactions. *J. Biol. Chem.* **2007**, *282*, 33076–33085. [[CrossRef](#)]
44. Almeida, P.C.; Nantes, I.L.; Chagas, J.R.; Rizzi, C.C.; Faljoni-Alario, A.; Carmona, E.; Juliano, L.; Nader, H.B.; Tersariol, I.L. Cathepsin B activity regulation. Heparin-like glycosaminoglycans protect human cathepsin B from alkaline pH-induced inactivation. *J. Biol. Chem.* **2001**, *276*, 944–951. [[CrossRef](#)]
45. Novinec, M.; Lenarčič, B.; Turk, B. Cysteine cathepsin activity regulation by glycosaminoglycans. *Biomed. Res. Int.* **2014**, *2014*, 309718. [[CrossRef](#)] [[PubMed](#)]
46. Yasuda, Y.; Li, Z.; Greenbaum, D.; Bogyo, M.; Weber, E.; Brömme, D. Cathepsin V, a Novel and Potent Elastolytic Activity Expressed in Activated Macrophages. *J. Biol. Chem.* **2004**, *279*, 36761–36770. [[CrossRef](#)] [[PubMed](#)]

47. Jmel, M.A.; Aounallah, H.; Bensaoud, C.; Mekki, I.; Chmelař, J.; Faria, F.; M'Ghirbi, Y.; Kotsyfakis, M. Insights into the Role of Tick Salivary Protease Inhibitors during Ectoparasite-Host Crosstalk. *Int. J. Mol. Sci.* **2021**, *22*, 892. [[CrossRef](#)] [[PubMed](#)]
48. Panwar, P.; Hedtke, T.; Heinz, A.; Andrault, P.M.; Hoehenwarter, W.; Granville, D.J.; Schmelzer, C.E.H.; Brömme, D. Expression of elastolytic cathepsins in human skin and their involvement in age-dependent elastin degradation. *Biochim. Biophys. Acta Gen. Subj.* **2020**, *1864*, 129544. [[CrossRef](#)] [[PubMed](#)]
49. Dennemärker, J.; Lohmüller, T.; Mayerle, J.; Tacke, M.; Lerch, M.M.; Coussens, L.M.; Peters, C.; Reinheckel, T. Deficiency for the cysteine protease cathepsin L promotes tumor progression in mouse epidermis. *Oncogene* **2010**, *29*, 1611–1621. [[CrossRef](#)] [[PubMed](#)]
50. Lecaille, F.; Chazeirat, T.; Saidi, A.; Lalmanach, G. Cathepsin V: Molecular characteristics and significance in health and disease. *Mol. Aspects Med.* **2022**, *88*, 101086. [[CrossRef](#)] [[PubMed](#)]
51. Brömme, D.; Wilson, S. Role of Cysteine Cathepsins in Extracellular Proteolysis. In *Extracellular Matrix Degradation*; Parks, W.C., Mecham, R.P., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 23–51.
52. Vizovišek, M.; Fonovič, M.; Turk, B. Cysteine cathepsins in extracellular matrix remodeling: Extracellular matrix degradation and beyond. *Matrix Biol.* **2019**, *75–76*, 141–159. [[CrossRef](#)]
53. Senjor, E.; Kos, J.; Nanut, M.P. Cysteine Cathepsins as Therapeutic Targets in Immune Regulation and Immune Disorders. *Biomedicines* **2023**, *11*, 476. [[CrossRef](#)] [[PubMed](#)]
54. Khamtorn, P.; Peigneur, S.; Amorim, F.G.; Quinton, L.; Tytgat, J.; Daduang, S. De Novo Transcriptome Analysis of the Venom of *Latrodectus geometricus* with the Discovery of an Insect-Selective Na Channel Modulator. *Molecules* **2021**, *27*, 47. [[CrossRef](#)]
55. Emerich, B.L.; Ferreira, R.C.M.; Cordeiro, M.N.; Borges, M.H.; Pimenta, A.M.C.; Figueiredo, S.G.; Duarte, I.D.G.; De Lima, M.E. δ -Ctenitoxin-Pn1a, a Peptide from *Phoneutria nigriventer* Spider Venom, Shows Antinociceptive Effect Involving Opioid and Cannabinoid Systems, in Rats. *Toxins* **2016**, *8*, 106. [[CrossRef](#)] [[PubMed](#)]
56. Benýšek, J.; Buša, M.; Rubešová, P.; Fanfrlík, J.; Lepšík, M.; Brynda, J.; Matoušková, Z.; Bartz, U.; Horn, M.; Gütschow, M.; et al. Highly potent inhibitors of cathepsin K with a differently positioned cyanohydrazide warhead: Structural analysis of binding mode to mature and zymogen-like enzymes. *J. Enzyme Inhib. Med. Chem.* **2022**, *37*, 515–526. [[CrossRef](#)] [[PubMed](#)]
57. Brömme, D.; Li, Z.; Barnes, M.; Mehler, E. Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization. *Biochemistry* **1999**, *38*, 2377–2385. [[CrossRef](#)] [[PubMed](#)]
58. Horn, M.; Baudyš, M.; Voburka, Z.; Kluh, I.; Vondrášek, J.; Mareš, M. Free-thiol Cys331 exposed during activation process is critical for native tetramer structure of cathepsin C (dipeptidyl peptidase I). *Protein Sci.* **2002**, *11*, 933–943. [[CrossRef](#)] [[PubMed](#)]
59. Horn, M.; Dolečková-Marešová, L.; Rulíšek, L.; Máša, M.; Vasiljeva, O.; Turk, B.; Gan-Erdene, T.; Baudyš, M.; Mareš, M. Activation processing of cathepsin H impairs recognition by its propeptide. *Biol. Chem.* **2005**, *386*, 941–947. [[CrossRef](#)] [[PubMed](#)]
60. Máša, M.; Marešová, L.; Vondrášek, J.; Horn, M.; Ježek, J.; Mareš, M. Cathepsin D propeptide: Mechanism and regulation of its interaction with the catalytic core. *Biochemistry* **2006**, *45*, 15474–15482. [[CrossRef](#)]
61. Perner, J.; Helm, D.; Haberkant, P.; Hatalová, T.; Kropáčková, S.; Ribeiro, J.M.; Kopáček, P. The Central Role of Salivary Metalloproteases in Host Acquired Resistance to Tick Feeding. *Front. Cell Infect. Microbiol.* **2020**, *10*, 563349. [[CrossRef](#)]
62. Urbanová, V.; Šíma, R.; Šauman, I.; Hajdušek, O.; Kopáček, P. Thioester-containing proteins of the tick *Ixodes ricinus*: Gene expression, response to microbial challenge and their role in phagocytosis of the yeast *Candida albicans*. *Dev. Comp. Immunol.* **2015**, *48*, 55–64. [[CrossRef](#)]
63. Perner, J.; Kropáčková, S.; Kopáček, P.; Ribeiro, J.M.C. Sialome diversity of ticks revealed by RNAseq of single tick salivary glands. *PLoS Negl. Trop. Dis.* **2018**, *12*, e0006410. [[CrossRef](#)]
64. Jílková, A.; Řezáčová, P.; Lepšík, M.; Horn, M.; Váchová, J.; Fanfrlík, J.; Brynda, J.; McKerrow, J.H.; Caffrey, C.R.; Mareš, M. Structural basis for inhibition of cathepsin B drug target from the human blood fluke, *Schistosoma mansoni*. *J. Biol. Chem.* **2011**, *286*, 35770–35781. [[CrossRef](#)]
65. Qi, D.; Scholthof, K.B. A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis. *J. Virol. Methods* **2008**, *149*, 85–90. [[CrossRef](#)]
66. Sojka, D.; Franta, Z.; Frantová, H.; Bartošová, P.; Horn, M.; Váchová, J.; O'Donoghue, A.J.; Eroy-Reveles, A.A.; Craik, C.S.; Knudsen, G.M.; et al. Characterization of gut-associated cathepsin D hemoglobinase from tick *Ixodes ricinus* (IrCD1). *J. Biol. Chem.* **2012**, *287*, 21152–21163. [[CrossRef](#)]
67. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **2001**, *29*, e45. [[CrossRef](#)]
68. Fojtík, L.; Fiala, J.; Pompach, P.; Chmelík, J.; Matoušek, V.; Beier, P.; Kukačka, Z.; Novák, P. Fast Fluoroalkylation of Proteins Uncovers the Structure and Dynamics of Biological Macromolecules. *J. Am. Chem. Soc.* **2021**, *143*, 20670–20679. [[CrossRef](#)] [[PubMed](#)]
69. Renshaw, P.S.; Lightbody, K.L.; Veverka, V.; Muskett, F.W.; Kelly, G.; Frenkiel, T.A.; Gordon, S.V.; Hewinson, R.G.; Burke, B.; Norman, J.; et al. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J.* **2005**, *24*, 2491–2498. [[CrossRef](#)] [[PubMed](#)]
70. Veverka, V.; Lennie, G.; Crabbe, T.; Bird, I.; Taylor, R.J.; Carr, M.D. NMR assignment of the mTOR domain responsible for rapamycin binding. *J. Biomol. NMR* **2006**, *36* (Suppl. 1), 3. [[CrossRef](#)] [[PubMed](#)]
71. Lee, W.; Tonelli, M.; Markley, J.L. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015**, *31*, 1325–1327. [[CrossRef](#)] [[PubMed](#)]

72. Herrmann, T.; Güntert, P.; Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **2002**, *319*, 209–227. [[CrossRef](#)] [[PubMed](#)]
73. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **2009**, *44*, 213–223. [[CrossRef](#)] [[PubMed](#)]
74. Janson, G.; Zhang, C.; Prado, M.G.; Paiardini, A. PyMod 2.0: Improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics* **2017**, *33*, 444–446. [[CrossRef](#)] [[PubMed](#)]
75. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)] [[PubMed](#)]
76. Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L.E.; Brookes, D.H.; Wilson, L.; Chen, J.; Liles, K.; et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, *27*, 112–128. [[CrossRef](#)]
77. Meng, E.C.; Goddard, T.D.; Pettersen, E.F.; Couch, G.S.; Pearson, Z.J.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **2023**, *32*, e4792. [[CrossRef](#)]
78. Kukačka, Z.; Rosůlek, M.; Jelínek, J.; Slavata, L.; Kavan, D.; Novák, P. LinX: A Software Tool for Uncommon Cross-Linking Chemistry. *J. Proteome Res.* **2021**, *20*, 2021–2027. [[CrossRef](#)]
79. Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C.H.; Schäfer, M.; Sinz, A. Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 83–97. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.