



HAL
open science

A review of the most important classes of serine protease inhibitors in insects and leeches

E. Clynen, L. Schoofs, M. Salzet

► To cite this version:

E. Clynen, L. Schoofs, M. Salzet. A review of the most important classes of serine protease inhibitors in insects and leeches. *Medicinal Chemistry Reviews - Online*, 2005, 2 (3), pp.197-206. hal-00167315

HAL Id: hal-00167315

<https://hal.science/hal-00167315v1>

Submitted on 18 Aug 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Trypsin and Chymotrypsin Inhibitors in Insects and Gut Leeches: an update

*E. Clynen¹, L. Schoofs¹ and M. Salzet²

1. *Laboratory for Developmental Physiology, Genomics and Proteomics, K.U.Leuven, Naamsestraat 59, B-3000 Leuven, Belgium.*
2. *Laboratoire de Neuroimmunologie des Annélides, ESA CNRS 8017, SN3,IFR 17INSERM, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cédex, France*

*Address correspondence to this author at Laboratory for Developmental Physiology, Genomics and Proteomics, K.U. Leuven, Naamsestraat 59, B-3000 Leuven, Belgium Tel 32.16.324260, Fax: 32.16.323902, e-mail: elke.clynen@bio.kuleuven.ac.be

Abstract: The constant increase of life expectancy is associated with major aging of developed populations. This indicates that the new century will have one of most epidemic progressions of cardiovascular, cancers and inflammatory diseases. The high challenge for medical research is to compress such morbidity. In these conditions, invertebrates have demonstrated to be truly useful models in drug discovery for such aging diseases. The last decade, drug discovery in leeches has opened the gate of new molecules to treat emphysema, coagulation, inflammation, dermatitis and cancer. Also other invertebrates such as insects, which evolved from the annelids, harvest potential interesting molecules, such as serine protease inhibitors that can be exploited by the medical industry.

In all metazoan species, proteases play a prominent role in a wide array of physiological processes such as food digestion, blood clotting, embryogenesis, tissue reorganization (e.g. wound healing, regeneration, molting, metamorphosis etc.), defense mechanisms and immune responses. Many of these processes are proteolytic cascades, which, once set in action, lead very rapidly and irreversibly to a specific cellular response. Activation and inactivation of protease cascades have to be closely controlled at different regulatory levels being protease gene transcription, mRNA translation, zymogen activation (all proteases are biosynthesized as large inactive precursors called proproteins or zymogens), substrate specificity, enzyme kinetics and by means of enzyme-inhibitors. Most animal species synthesize a variety of protease inhibitors with different specificities whose function is to prevent unwanted proteolysis. They act by unabling access of substrates to the proteases' active site through steric hindrance. Proteases are involved in various disease states. For instance, the destruction of the extracellular matrix of articular cartilage and bone in arthritic joints is thought to be mediated by excessive proteolytic activity [1]. In emphysema, gingivitis, tumor invasion and inflammatory infections, it is suggested that tissue destruction is caused by proteases [1, X]. Among the enzymes involved in extracellular matrix degradation, a few serine proteases (elastase, collagenase, cathepsin G) are able to solubilize fibrous proteins such as elastin and collagen [2,3]. Given the specific recognition by proteases of defined amino acid sequences, it may be possible to inhibit these enzymes when they are involved in pathological processes. Potent inhibitors have the potential to be developed as new therapeutic agents. In vertebrates, serine protease inhibitors have been studied for many years and they are known to be involved in phagocytosis, coagulation, complement activation, fibrinolysis, blood pressure regulation, *etc.* In the last decade, it became obvious that in invertebrates, serine proteases and their inhibitors are also involved in parallel physiological processes (e.g. the blood clotting cascade in *Limulus* [4] and the innate immune response [5]). Moreover, some of the protease inhibitors isolated from invertebrate sources are quite specific towards individual mammalian serine proteases. This also offers huge opportunities for medicine. Thus, the development of non-toxic protease inhibitors extracted from invertebrates for *in vivo* application may be quite important [1]. In the future, it is likely that numerous specific protease inhibitors will be

tested clinically for the treatment of human disease such like emphysema, inflammation, dermatitis and cancer.

SERINE PROTEASE INHIBITORS

The amino acid sequences of the currently identified serine protease inhibitors differ significantly, with the number of constituent amino acids ranging from 29 to about 400. Despite this, only a few fundamentally different inhibiting mechanisms seem to exist. Most inhibitors bind with their cognate enzyme(s) according to a common, substrate-like standard mechanism. They are all relatively small (from 29 to 190 amino acids) and share an exposed, rigid binding loop with a very characteristic ‘canonical’ conformation, which inserts into the active site cleft of the enzyme in a substrate-like manner [6]. At least 18 different families of standard mechanism, canonical inhibitors have been defined based on structural criteria. They all share the conformation of the combining loop but each has its own global three-dimensional structure.

The serpins (*serine protease inhibitors*) on the other hand constitute a family of large (glyco)proteins (typically about 400 amino acid residues in length) which function as suicide substrate inhibitors, causing inhibition by deformation [7]. Upon binding, they undergo a profound change in topology in order to entrap their target protease in an irreversible complex. The conformational change is initiated by reaction of the active serine of the protease with the reactive center loop of the serpin. This cleaves the reactive center, which then moves 71 Å to the opposite pole of the serpin, taking the tethered protease with it. This grossly disrupts the active site of the protease and causes a surprising 37% loss of ordered structure, which prevents the release of the protease from the complex and results in its deactivation [7,8].

Serpins are involved in diverse biological processes, including functions unrelated to protease regulation [9]. Many serpins found in human plasma regulate proteolytic reactions important in blood coagulation, fibrinolysis, the immune response, and inflammation. A serpin called maspin identified in mammalian breast tissue has been implicated in tumor suppression [10].

SERINE PROTEASE INHIBITORS IN INSECTS

Insect hemolymph, like vertebrate serum, contains several serine protease inhibitors [11]. Most of the insect serine protease inhibitors examined hitherto were identified or partially characterized from hemolymph extracts and can be grouped into two families based on their amino acid sequence and their protease inhibition characteristics: low molecular mass proteins (below 10 kDa) belonging to the standard mechanism, canonical inhibitors [6] and proteins of about 45 kDa which belong to the serpin superfamily [12]. Although details on the function of the insect serine protease inhibitors are as yet not available, evidence is accumulating that they play a role in insect anti-microbial defence mechanisms, digestion, metamorphosis and development [13].

So far, serpins have been reported in three different insect orders: Diptera (*Drosophila melanogaster*, *Anopheles gambiae* and *Aedes aegypti*), Lepidoptera (*Manduca sexta*, *Bombyx mori*, *Mamestra configurata* and *Hyphantria cunea*) and Hymenoptera (endoparasitic wasp *Venturia canescens*).

The genome of the fruit fly, *Drosophila melanogaster*, contains thirty serpin genes [14]. The first serpin from *Drosophila* to be reported was the Acp76A protein, a component of the male accessory gland that is transferred to the female during mating [15]. Later, six new *Drosophila* serpins were identified that according to their sequence are likely to function as serine protease inhibitors involved in dorsoventral patterning [16]. These serine proteases are perhaps acting sequentially in a cascade like the mammalian blood clotting proteases. The authors provide the first biochemical evidence that at least one of them is a potent inhibitor of trypsin-like proteases *in vitro*. One *Drosophila* serpin displays sequence similarities with human neuroserpin [17], whereas others appear to be most similar in sequence to members of the ov-serpin subfamily, which includes inhibitory serpins such as human plasminogen activator inhibitor-2 [18]. The genome of the malaria mosquito, *Anopheles gambiae*, encodes 14 serpins, 10 of which are inhibitory [19]. Recently, in the sialome of the yellow fever mosquito, *Aedes aegypti*, a novel serpin was characterized, which shows similarity to the FXa-directed salivary anticlotting

protein [20].

Serpins active against different types of serine proteases have been purified from the plasma of two lepidopteran insects, *Bombyx mori* and *Manduca sexta*. In *Manduca sexta*, alternative splicing of serpin gene-1, constitutively expressed in fat body and hemocytes, generates a group of 12 serpin-1 proteins with different protease inhibitory activities, which are secreted into the plasma [21]. Serpin-2, on the other hand, is expressed in granular hemocytes in response to bacterial challenge, suggesting that it may function in hemocyte antibacterial response [22]. In contrast to serpin-1, the amino-terminus of serpin-2 does not contain a secretion signal sequence, and serpin-2 is not present in plasma. Serpin-2 shows similarity to *Drosophila* serpin-4 and two serpin-sequences from *Anopheles gambiae* (accession numbers AJ271352 and AJ271353) also lacking signal peptides. The vertebrate protein most similar to this group of peptides is horse leucocyte elastase inhibitor (HLEI), member of the ovalbumin family, and present in the cytosol of neutrophils. Ovalbumin-type serpins also lack amino-terminal secretion signal peptides and most are cytosolic proteins. Homologues of serpin-1 have also been characterized in another Lepidopteran, *Mamestra configurata* [23].

Serine proteases play critical roles in the regulation of the invertebrate innate immune responses, as also their associated regulatory serine protease inhibitors. For example, in *Manduca sexta*, the prophenoloxidase cascade (which catalyzes the formation of melanin during the defense reaction and thus plays an important role in the encapsulation of pathogens, cuticle sclerotization and wound healing) is initiated by proteolytic processing and the activating proteases are regulated by inhibitory serpins [24]. Also in the fall webworm *Hyphantria cunea* serpin probably participates in the negative regulation of the prophenoloxidase cascade [25]. In *Drosophila*, serpin-27A regulates the melanization cascade through the specific inhibition of the terminal protease prophenoloxidase-activating enzyme [26]. Serpin-27A is required to restrict the phenoloxidase activity to the site of injury, thereby preventing systemic melanization. During the melanization process, serpin-27A is depleted from the hemolymph, This depletion is infection-dependent and controlled by the Toll pathway [27]. In *Anopheles gambiae*, the genomic

locus SRPN10 codes for four alternatively spliced serpins. At least two isoforms are transcriptionally upregulated during parasite passage through the midgut, suggesting that they may be implicated in antiparasitic action or, alternatively, parasite tolerance [28]. Endoparasitic wasps are able to develop inside permissive host insects due to their ability to overcome or evade the host's immune system. Recently evidence has arisen that ovarian calyx fluid, deposited into the host together with the parasitoid egg, contains serpin activity which might transiently inactivate host defence reactions until other means of protection are established on the egg surface [29].

Recently, two new families of serine protease inhibitors have been discovered. The first, designated as the *Bombyx* family, was discovered in silkworms [30]. The second, of which several low molecular weight peptides of around 4 kDa were characterised in locusts, was first designated as the locust serine protease inhibitor peptide family (Fig. 1). Three inhibitors were purified in *Locusta migratoria* designated as LMCI 1 and 2 (or *Locusta migratoria* chymotrypsin inhibitor 1 and 2) and HI (or hemolymph inhibitor) [31-35] and five in *Schistocerca gregaria* designated as SGPIs or *Schistocerca gregaria* protease inhibitors 1-5 [16]. In addition, 11 locust inhibitors were identified by cDNA cloning [36,37]. The locust peptide inhibitors display sequence similarities with nine cysteine rich domains (PLDs) in the light chain of pacifastin, a heterodimeric serine protease inhibitor isolated from the hemolymph of the crayfish *Pacifastacus leniusculus* [38]. The PLD's and the 4 kDa locust inhibitors share a conserved pattern of six cysteine residues (Cys-Xaa9-12-Cys-Asn-Xaa-Cys-X-Cys-Xaa2-3-Gly-Xaa3-4-Cys-Thr-Xaa3-Cys), which form three disulfide bridges (Cys1-Cys4, Cys2-Cys6, Cys3-Cys5) (Fig.1). Since pacifastin was the first reported member of this new family, the name of this family was recently changed from locust serine protease inhibitor peptide family to the pacifastin family [39]. *In silico* data mining of the expressed sequence tags databases revealed the existence of additional pacifastin-like polypeptides in two lepidopteran species *Bombyx mori* and *Manduca sexta*, suggesting a broad distribution of this peptide family among arthropods [36]. Below, we review the latest data on the pacifastin family.

Pacifastin Family: Structure Analysis

The studies of Mer *et al.* [40-43] contributed significantly to the elucidation of the specific structural characteristics of the locust low molecular weight inhibitors. The core region adopts a compact, globular fold, which consists of three strands (β 1, β 2 and β 3) arranged in an anti-parallel β -sheet, that demarcates a cavity and an amino-terminal segment, orientated almost perpendicular to the β -sheet. Inside the cavity, hydrophobic residues are clustered with an aromatic ring in the center of the hydrophobic core. The protease binding loop, located between two cysteine residues in the carboxy-terminal segment exhibits an extended conformation and is anchored to the core region by two cysteine bridges [35,41]. These structural characteristics are commonly encountered in small 'canonical' inhibitors. They all have a conserved ('canonical') and extended binding loop, which permits an anti-parallel β -strand interaction with the protease active site. This interaction is stabilized by intramolecular bonds (i.e. disulfide bonds) between the binding loop and the core of the inhibitor. The locust peptides are stabilized by three disulfide bridges forming a 'cysteine knot' [10,11]. They adopt a tertiary fold hitherto unobserved in the large group of 'canonical' protease inhibitors [17]. Therefore, these inhibitors constitute a new subfamily within the large group of 'canonical' inhibitors [13,41].

Four of the eight locust inhibitors (LMCI-2, SGPI 2,4,5) are post-translationally modified by a deoxyhexose moiety on the threonine residue on the fifth position after the first cysteine residue [13,32]. As the non-fucosylated peptides retain their full activity [35,44], it remains obscure why some inhibitors of this newly identified family contain this sugar while others lack it. Inspired by other conformational studies on glycosylated proteins, Mer *et al.* proposed a stabilizing effect of the fucose moiety [43].

Pacifastin Family: Specificity

The locust enzyme inhibitors display specificity differences towards locust endogenous enzymes [13] but also towards mammalian serine proteases as shown in table 1. These

differences can be mainly attributed to the amino acid sequence within the active site. Indeed, a difference of one or two amino acid residues around the reactive sites often results in considerable alteration of the inhibitory specificity [10]. The P1 residue is mainly responsible for the inhibitor's specificity for a particular protease. In general, chymotrypsin inhibitors have bulky and aromatic residues such as Phe, Tyr, Leu or Met as their P1 residue whereas trypsin require basic amino acids. Five of the 4 kDa inhibitors (LMCI-2, SGPI-2,3,4 and 5) have a Leu as P1 residue and are potent or moderately potent inhibitors of α -chymotrypsin [13,33,35]. The other inhibitors (HI, LMCI-1 and SGPI-1) with an Arg at the P1 position are potent trypsin inhibitors and display no or very weak chymotrypsin inhibiting activity [13,35]. Structure activity studies by Kellenberger *et al.* [35] provided strong evidence for the close relation between the P1 residue and the inhibitor's specificity for a particular enzyme. They demonstrated that the substitution of a P1-Arg residue for a Leu residue resulted in an enormous decrease in the K_i value (by a factor 2000) for chymotrypsin, converting both trypsin inhibitors into potent chymotrypsin inhibitors. Vice versa, the replacement of the P1-Leu by Arg in SGPI-2 converted this potent chymotrypsin inhibitor ($K_i = 6.2$ pM) into a moderate ($K_i = 51$ nM) trypsin inhibitor, whereas the activity towards chymotrypsin ($K_i = 5.5$ nM) decreased a 1000fold. The specificity of serine protease inhibitors depends not exclusively on the nature of the P1 residue. Structure-activity studies with analogues of SGPI-2 demonstrated that a Met residue instead of a Lys at the P'1 position increases the affinity of the inhibitor for trypsin almost as much as it did by changing the P1 site [44]. The rather weak trypsin inhibition ($K_i = 210$ nM) of the naturally occurring SGPI-1 can be attributed to the unfavourable interaction (repulsion) between P'1 (positively charged Lys) of SGPI-1 (CTR-**KGC**) and the S'1 residue (positively charged Lys) of trypsin.

The elucidation of the solution structure of three small serine protease inhibitors: SGPI-2, which is a good chymotrypsin inhibitor, its mutant SGPI-2 [L30R, K31M], which is a potent trypsin inhibitor and SGPI-1, which inhibits both proteases weakly, suggests that the differing conformation of the binding loops as well as the difference in flexibility of the binding loop can also, at least in part, explain the species specificity [45].

X-ray studies of the protease-inhibitor complexes are also helpful in characterizing the structural requirements for inhibition. The crystal structure of LMCI-1 [R29L, K30M] and LMCI-2 were determined in complex with bovine α -chymotrypsin. Structural divergences between the two inhibitors result in an additional interaction site between LMCI-1 [R29L, K30M] and chymotrypsin [46].

LMCI-1 and HI, which display an Arg at the P1 position have no effect on porcine trypsin and are only weak inhibitors of bovine trypsin, however, these peptide are strong inhibitors of trypsins isolated from *Locusta migratoria* [47]. Similar, SGPI-1 inhibits trypsins isolated from arthropods (crayfish and shrimp) much better than the bovine enzyme used for routine assays. This could be due to flexibility differences in the protein-protein interaction [48]. These data suggest that the more advanced trypsins of mammals could have diverged from insect/arthropod proteases, perhaps to escape from inhibitors present in their meal. This also opens possibilities for engineering the natural selectivity of HI, LMCI-1 and SGPI-1 toward targets of therapeutic or agronomical interest.

Pacifastin Family: Precursor Genes

In *L. migratoria*, LMCI 1 and 2 are derived from a single precursor polypeptide, designated as *Locusta migratoria* Pacifastin related Precursor-1 or LMPP-1 [49]. Similarly, SGPI-1 and SGPI-2 are derived from a single polypeptide precursor in *S. gregaria* [50], called *Schistocerca gregaria* Pacifastin-related Precursor-1 or SGPP-1, whereas SGPI-3 is encoded by an additional precursor, SGPP-2 [50]. The transcripts are present in several locust tissues, but not in the midgut. Important changes in transcript levels occur during development. Vanden Broeck *et al.* also provided evidence indicating that the expression of the inhibitors is hormonally regulated [50]. Recently, cDNA cloning revealed two isoforms a novel pacifastin-related precursor in *Schistocerca gregaria*, SGPP-3, which codes for three putative inhibitor peptides, two novel peptides (SGPI-4A en C) and a peptide (SGPI-4B), which most likely corresponds to the earlier purified SGPI-4. The isoforms differ at a single amino acid in the third peptide (SGPI-4Ca and SGPI-4Cp) [37]. Two different transcripts were present. Both are most abundant

in the fat body and appear to be regulated during molting and the reproductive cycle [37]. In *Locusta migratoria*, cDNA cloning revealed two novel pacifastin-related peptide precursors displaying respectively three (LMPP-2) and four (LMPP-3) inhibitor peptides [36]. Only one of the encoded peptides (HI) was identified previously.

Pacifastin Family: Tissue Distribution

LMCI-1, LMCI-2 and HI were isolated from *Locusta* hemolymph and brain tissue [32,33,35]. The five *Schistocerca* inhibitors were isolated from ovarian tissue [13]. However, the spatial and temporal distribution data as analyzed by HPLC indicated the presence of each inhibitor in gonads, hemolymph and fat body of adults and in larval hemolymph and fat body [13]. The midgut was devoid of this type of 4 kDa inhibitors. Our recent data indicate that the locust midgut contains its own chymotrypsin/elastase inhibitors, which according to elution characteristics, differ from the 4 kDa locust enzyme inhibitors (Vercammen *et al.*, unpublished data).

LMCI-1 has been localized in neurosecretory cells of the brain (pars intercerebralis) as shown by immunocytochemistry [51]. Peptidomic analysis (liquid chromatography on line with mass spectrometry) unequivocally proved that LMCI-1 and 2 as well as SGPI 1 and 2 are synthesized in the brain, more specifically in the pars intercerebralis-corpora cardiacum (CC) axis (Fig. 2), which is the equivalent of the mammalian hypothalamus-pituitary system [52]. These data suggest that the inhibitors might be involved in the regulation of neuropeptide precursor processing by serine proteases (convertases). Mass spectrometric analysis also indicated the release of both LMCIs from the CC *in vitro*. Therefore, it is likely that LMCI-1 and LMCI-2 are released into the hemolymph via a neuro-endocrine pathway.

Besides the typical 4 kDa inhibitors, another 14 kDa serine protease inhibitor was isolated from *Schistocerca* ovaries by a combination of trypsin-affinity chromatography and reverse-phase high performance liquid chromatography [53]. The N-terminal sequence (Y) XAEXDELA (A) EEY (Y)Q(Q)X(I)(L)M (X being a Cys, an irregular or

modified amino acid) revealed no similarities with any other protease inhibitors isolated from invertebrate or vertebrate source. The 14 kDa inhibitor was found to be heat-stable. The purified molecule, which could be extracted in water but not in acidic methanol, potently inhibited bovine trypsin and chymotrypsin, but not pancreatic elastase. The cDNA for this 14 kDa inhibitor has not been cloned as yet.

Pacifastin Family: Biological Function

The 4 kDa inhibitors are assumed to play a role in the innate immune defense system [31,33]. They are potent inhibitors of the proteolytic cascade activating prophenoloxidase (proPO), which is present as a proenzyme in the hemolymph. Upon activation, it leads to the formation of antifungal quinones and the local production of melanine around invading parasites.

Of particular interest is that LMCI-1 and LMCI-2 have also been shown to act as inhibitors of high voltage-activated calcium channels on mammalian cells [51, 54].

As the locust inhibitors are expressed in several tissues (hemolymph, brain, fat body and gonads) we may assume that they probably are involved in many physiological processes where proteolytic cascades have to be fine-tuned.

The protease inhibitors perhaps also play a role in phase transition. SGPI-2 is apparently more abundant in the CC and hemolymph of isolated-reared locusts, than in crowded-reared locusts and this difference increases with successive generations of breeding in isolation [55,56]. This suggests that this protease inhibitor possibly inhibits a certain proteolytic cascade in solitary and not in gregarious locusts.

Recently, the recombinant production and purification of SGPI-1 and SGPI-2 via a bacterial periplasmic expression system was optimized. This could mean a major breakthrough as it is now possible to obtain sufficient biologically active material to perform physiological experiments [57].

SERINE PROTEASE INHIBITORS IN LEECHES

In hematophagous leeches, studies of serine protease inhibitors and their substrates have been extensively performed. Two groups of serine protease inhibitors can be distinguished. The first is related to specific inhibitors that interfere in the activation of the blood clotting system and the second group related to inhibitors that work on the extracellular matrix [58].

For the first group, several developments for clinical trials have been performed since the last decade by pharmaceutical companies. In fact, arterial and venous thromboses are a major cause of morbidity and mortality. Anticoagulants are a cornerstone of treatment in patients with these disorders. The two most frequently used anticoagulants, heparin and warfarin, have pharmacological and/or biophysical limitations that make them difficult to use in day-to-day clinical practice. Development of new natural anticoagulants, which were designed to overcome these limitations are in progress and new anticoagulants target various steps in the coagulation pathway are especially researched. Thrombin, the final effector in coagulation converts soluble fibrinogen into insoluble fibrin. Thus, thrombin is a key enzyme in pathogenesis of acute coronary thrombosis; therapy with heparin, an indirect thrombin inhibitor has been developed since the thirty-last decade (see, Salzet, *Curr Pharm Des.* 2002; 8(7): 493-503) but new anticoagulation drugs that target each of these clotting enzymes have been developed (**Weitz JI, Crowther MA.** *Am J Cardiovasc Drugs.* 2003; 3(3): 201-9.).

In this context, the major potent thrombin inhibitor discovered at the present time is Hirudin, isolated from the salivary glands of the medicinal leech (Haustein, K.O.; Markwardt, F. *Thromb. Diath. Haemorrh.* **1965**, 13, 60). Because, Hirudin has been plagued by bleeding complications, likely due to its high affinity for thrombin (Nawarkas and Anderson, *Heart Dis.* 2001 Mar-Apr; 3(2): 131-7.), development of recombinant

versions of hirudin (r-hirudin) has been performed from companies over the last 20 years (Deitcher, Am J Health Syst Pharm. 2003 Oct 15;60 Suppl 5:S27-31).

Lepirudin is one such r-hirudin that is identical to natural hirudin except for the substitution of leucine for isoleucine at the N-terminus and the elimination of a sulfate group on the tyrosine at position 63. Desirudin, is identical to hirudin except for a valine-valine in the N-terminus and the absence of the sulfate group on tyrosine at position 63 (Deitcher, Am J Health Syst Pharm. 2003 Oct 15;60 Suppl 5:S27-31). Both r-hirudins are bivalent and tightly bind to both the catalytic site and the exposit of thrombin to exert their inhibitory effects on thrombin (**Donges R, Brazel D J**, Chromatogr A. 2002 Dec 6;979(1-2):217-26.).

r-Hirudins are effective agents and do not cross-react with heparin-induced thrombocytopenia (HIT)-associated antibodies, they are excellent anticoagulants in patients with past or current HIT (Deitcher, Am J Health Syst Pharm. 2003 Oct 15;60 Suppl 5:S27-31). Clinical trials have also demonstrated the efficacy and safety of subcutaneous (s.c.) r-hirudins compared to heparins in non-HIT settings. Results of these trials support the use of r-hirudin therapy in patients with HIT or at risk of developing HIT (Deitcher, Am J Health Syst Pharm. 2003 Oct 15;60 Suppl 5:S27-31).

Bivalirudin (Angiomax, The Medicines Company) represents a new class of anticoagulant drugs that directly inhibits thrombin (Scatena, R. *Expert Opin Investig Drugs* **2000**, 9, 1119-1127). According to **White HD and Chew DP** (*Expert Opin Pharmacother.* 2002 Jun;3(6):777-88), the only direct thrombin inhibitor shown to reduce

both the ischaemic and the bleeding complications associated with percutaneous coronary intervention (PCI) is bivalirudin. This molecule is approved for this indication in the US and New Zealand. This agent is currently being studied in patients undergoing PCI with or without glycoprotein IIb/IIIa inhibitors and stenting. Bivalirudin has been shown to significantly reduce the risk of reinfarction in patients with acute myocardial infarction (MI) treated with streptokinase. It may also prove to be beneficial in patients with acute MI treated with other fibrinolytic regimens or with primary or facilitated PCI. Bivalirudin is suitable for use as an alternative to heparin in the majority of patients undergoing PCI and in patients receiving streptokinase for acute MI (**Nawarskas JJ, Anderson JR.** Heart Dis. 2001 Mar-Apr;3(2):131-7. ; **White HD and Chew DP** (Expert Opin Pharmacother. 2002 Jun;3(6):777-88).

In the case of patients affected with musculoskeletal injury (bruises) with or without hematoma, Hirudex cream extracted from *Hirudo medicinalis* (280 UI/100 g) was clinically tested (**Stamenova PK, Marchetti T, Simeonov I.** Eur Rev Med Pharmacol Sci. 2001 Mar-Apr;5(2):37-42.). Highly statistically and clinically significant improvements were noted suggesting that Hirudex is an effective local treatment in patients with mild to moderate bruises (**Stamenova PK, Marchetti T, Simeonov I.** Eur Rev Med Pharmacol Sci. 2001 Mar-Apr;5(2):37-42.).

In conclusion, unfractionated heparin and low-molecular-weight heparins (LMWHs) are widely used in medical and surgical patients to prevent and treat arterial and venous thrombotic events. Besides bleeding, the major adverse effect of heparins is heparin-induced thrombocytopenia (HIT). HIT is associated with a paradoxical hypercoagulable

state and marked risk of clinical thrombosis. Management of HIT requires the immediate cessation of all heparin exposure, and the initiation of an alternative anticoagulant. Clinical trials have demonstrated the efficacy and safety of subcutaneous r-hirudins and hirulogs compared to heparins in non-HIT settings. These molecules have shown that they are effective agents and do not cross-react with HIT-associated antibodies and they are excellent anticoagulants in patients with past or current HIT.

Concerning, the second group of inhibitors, in gut leeches (Fig. 3), Cytin [59] is the first chymotrypsin inhibitor having two chains isolated from leeches. The B chain exhibits 16% sequence identity with eglin from *H. medicinalis* [60]. No sequence homology was found with gelin from *H. manillensis* [61] or with the other chymotrypsin inhibitors isolated from jawed Hirudinae (Fig. 4) *i.e.* bdellins [62] Cytin, as does eglin [60], possesses sequence identity with the substilisin/chymotrypsin inhibitor family isolated from barley seeds (CI, [63]). Alignment of N-terminal sequences (first 22 residues) of the B chain of Cytin with CI-2a revealed a 73% residue identities (Table 3) and 52% to 67% residue identities with CI-I (a, b, c) and CI-2b, respectively. At the level of the A chain, no sequence homology was found with other chymotrypsin inhibitors. Thus, this structural feature- 2 chains linked by a disulfide bridge- is novel among chymotrypsin inhibitors.

Additionally, this structure appears to be essential for the inhibitory activity of Cytin. In fact, after reduction and S-pyridylethylation, the two chains became separated and the peptide lost its activity [59]. Furthermore, this unique structure makes Cytin different even from the potato chymotrypsin inhibitors (PI-1, [64]), of which eglin and CI-2 are members [65], as well as from other known serine proteases inhibitors. Moreover, the oxidation of Cytin provoked a loss of activity of around 85%, reflecting the fact that the Met residue in the chain A is important for the molecule's activity.

Chymotrypsin inhibitors isolated from jawed leeches (*H. medicinalis* and *H. manillensis*

), i.e., gelin and eglin [60,61] display no sequence homology with Cytin. This result is in contrast to thrombin inhibitors from these animals where there is a 70% structural homology [58]. Therefore, we can conclude that eglin and gelin did not originate from the same ancestral gene as Cytin. Besides this chymotrypsin inhibitor, a specific small peptide trypsin inhibitor (therin, [65]) as well as trypsin-chymotrypsin inhibitors ([66], tessulin) were isolated from the same species. Therin exhibits *ca.* 30% of sequence identity and spacing of the 8 cysteines residues with molecules of the antistasin-type inhibitors family [67]. The sequence similarity is much higher with the first domain of antistasin (33%) [67]. Indeed, if therin is aligned with these potentially homologous inhibitors on the basis of its cysteines, the positions of the cysteines and the putative P1 active site match best with all the antistasin-type protease inhibitors. The P1 residue of the reactive site of the inhibitor generally determines its specificity [69]. In antistasin, only the N-terminal domain is inhibitory and the P1 residue has been determined as Arg34 [70]. In the therin sequence, the P1 residue could be Lys28. Therin inhibits bovine trypsin with high affinity and specificity (K_i value of 45 ± 12 pM). This value is much higher than those obtained with other potent trypsin-chymotrypsin inhibitors found in leeches (isolated from respectively *H. medicinalis* (hirustasin; bellastasin [71,72]), *Hirudinaria nipponia* (guamerins and piguamerin [73, 74]) and the non blood sucker leech *whitmania edentula* (guamerin II, [75]) (Table 2).

Tessulin [66] is a 9-kDa peptide of 81 amino acids. It possesses 16 cysteines and displays 16% sequence similarity with antistasin-type inhibitors (Fig. 5). It inhibits trypsin (K_i 1 pM) and chymotrypsin (K_i 150 pM) and exhibits no activity with thrombin, factor Xa, cathepsin G and elastase. Tessulin, like Cytin [59] and Therin [64] are the only trypsin-chymotrypsin inhibitors isolated from leeches that do not inhibit elastase or cathepsin G. Furthermore, Tessulin, in conjunction with other serine-protease inhibitors isolated from *Theromyzon* i.e. Therin, Cytin, Therostasin and Theromin [65,66,68-70], significantly diminishes the level of human granulocyte and monocyte activation induced by lipopolysaccharides (10 μ g). The combined level of inhibition is higher than that of aprotinin, another serineprotease inhibitor used biomedically. Sequence comparisons were carried out for Tessulin with the five different protease inhibitors isolated from the

leech *T. tessulatum*: Therin, Therostasin, Cytin, antitrypsin A, B and Tessulin [63]. This revealed that, for three of the five peptides, being Therostasin, Theromin and Tessulin, display a high degree of sequence similarities (>80%), except for the amino acid residues surrounding the putative active site. Therefore, they probably constitute a new protease inhibitor family (Fig. 6).

Apart from this family of small peptides (< 10 kDa), we also demonstrated the presence of two specific trypsin-chymotrypsin inhibitors with a molecular mass ranging between 14 and 15 kDa, from head parts of the rhynchobdellid leech *Theromyzon tessulatum* (Fig. 7, [69]). Two proteins, anti-trypsin-chymotrypsin A : ATCA (14636.6 ± 131 Da) and anti-trypsin-chymotrypsin B : ATCB (14368 ± 95 Da) were purified by size exclusion and anion-exchange chromatography followed by reversed-phase HPLC. Based on amino-acid composition, N-terminal sequence determination (MELCELGQSCSRDNPQPSNM), matrix assisted laser desorption-time of flight measurement (MALDI-TOF), trypsin mapping comparison, inhibition constant determination (K_i), and influence on amidolytic activity of different serine proteases, we demonstrate that ATCA and ATCB are novel and highly potent serine protease inhibitors of trypsin and chymotrypsin (ATCA : 350 fM towards trypsin and chymotrypsin; ATCB : 400 and 75 fM towards trypsin and chymotrypsin, respectively). We further surmise that ATCA and ATCB are linked, in that ATCB would lead to the formation of ATCA after loss of few amino acid residues.

Taken together, these data reflect that invertebrates contain several highly specific serine protease inhibitors that could be used as lead compounds for the development of powerful drugs.

ACKNOWLEDGEMENTS

This work was supported in part by the Centre National de la Recherche Scientifique, and the MNERT. L. Schoofs acknowledges funding by the Fund for Scientific Research-Flanders (Belgium) or FWO and the Research Council of the K.U.Leuven. E. Clynen is a postdoctoral fellow of the FWO.

REFERENCES

- [1] Roston, D. *Int. J. Cardiol.*, **1996**, *53 suppl*, S11-S37. [20] Boigegrain, R.A., Mattras, H., Brehelin, M., Paroutaud, P., Coletti-Previero, M.A. *Biochem. Biophys. Res. Commun.* **1992**, *189*,790-793.
- [2] Sloane, B.F., Rozhin, J., Johnson, K., Taylor, H., Crissman, J.D., Honn, K.V. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 2483-2487.
- [3] Berquin, I.M ; Sloane, B.F. *Adv. Exp. Med. Biol.* **1996**, *389*, 281-94.
- [4] Imler, J.-L., Hoffmann, J.A. *Curr. Opin. Microbiol.* **2000**, *3*,16-22.
- [5] Iwanaga, S. *Curr. Opin. Immunol.* **1993**, *5*,74-82. [23] Liang, Z., Sottrup-Jensen, L., Aspán, A., Hall, M., Söderhäll, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6682-6687.
- [6] Laskowski Jr., M. *Adv. Exp. Med. Biol.* **1980**, *199*, 1-17.
- [7] Huntington, J.A., Read, R.J., Carrell, R.W. *Nature* **2000**, *407*, 923-926.
- [8] Carrell, R.W., Huntington, J.A. *Biochem. Soc. Symp.* **2003**, *70*, 163-178.
- [9] Gettins, P.G.W., Patston, P.A., Olson, S.T In: *Serpins: Structure, Function and Biology* **1996**, R.G. Landes, Austin, TX. [25] Mer, G., Hietter, H., Kellenberger, C., Renatus, M., Luu, B., Lefèvre, J.F. *J. Mol. Biol.* **1996**, *58*, 158-171.
- [10] Sager, R., Sheng, S., Pemberton, P., Hendrix, M.J. *Adv. Exp. Med. Biol.* **1997**, *425*,77-88. [26] Mer, G., Dejaegere, A; Stote, R., Kieffer, B., Lefevre, J.F. *J. Phys. Chem.* **1996**, *100*, 2667-2674.
- [11] Polanowski, A., Wilusz, T. *Acta. biochim. Pol.* **1996**, *43*, 445-453.

- [12] Bode W., Huber, R. *Eur. J. Biochem.* **1992**, *204*, 433-451 [24] Mer, G., Kellenberger, C., Koehl, P., Stote, R., Sorokine, O., Van Dorsselaer, A., Luu, B., Hietter, H., Lefevre, J.F. *Biochemistry* **1994**, *33*, 15397-15407.
- [13] Hamdaoui, A., Watalev, S., Devreese, B., Chiou, S.J., Vanden Broeck, J., Van Beeumen, J., De Loof, A., Schoofs, L. *FEBS Lett.* **1998**, *422*, 74-78.
- [14] Han, J.-H., Zhang, H.,-Y., Min, G.,-S., Kemler, D., Hashimoto, C. *Science* **2000**, *287*, 2185-98.
- [15] Coleman, S., Drahn, B., Petersen, G., Stolorov, J; Kraus, K. *Insect Biochem. Mol. Biol.* **1995**, *25*, 203-207.
- [16] Han, J.-H., Zhang, F.-Y., Min, G.-S., Kemler, D., Hashimoto, C. *FEBS Letters* **2000**, *468*, 194-198.
- [17] Schrimpf, S.P., Bleiker, A.J., Brecevic, L., Kozlov, S.V., Berger, P., Osterwalder, T., Krueger, S.R., Schinzel, A. Sonderegger, P. *Genomics* **1997**, *40*, 55-62.
- [18] Remold-O'Donnell, E. *FEBS Lett.* **1993**, *315*, 105-108.
- [19] Christophides, G.K. *et al. Science* **2002**, *298*, 159-165.
- [20] Valenzuela, J.G., Pham, V.M., Garfield, M.K., Francischetti, I.M.B., Ribeiro, J.M.C. *Insect Biochem. Mol. Biol.* **2002**, *32*, 1101-1122.
- [21] Jiang, H., Kanost, M.R. *J. Biol. Chem.* **1997**, *272*, 1082-1087.
- [22] Gan, H., Wang, Y., Jiang, H., Mita, K., Kanost, M.R. *Insect Biochem. Mol. Biol.* **2001**, *31*, 887-898.

- [23] Chamankhah, M., Braun, L., Visal-Shah, S., O'Grady, M., Baldwin, D., Shi, X., Hemmingsen, S.M., Alting-Mees, M., Hegedus, D.D. *Insect Biochem. Mol. Biol.* **2003**, *33*, 355-369.
- [24] Kanost, M.R., Jiang, H., Wang, Y., Yu, X.Q., Ma, C., Zhu, Y. *Adv. Exp. Med. Biol.* **2001**, *484*, 319-328.
- [25] Park, D.S., Shin, S.W., Hong, S.D., Park, H.Y. *Mol. Cells.* **2000**, *10*, 186-192.
- [26] De Gregorio, E., Han, S.J., Lee, W.J., Baek, M.J., Osaki, T., Kawabata, S.I., Lee, B.L., Iwanaga, S., Lemaitre, B., Brey, P.T. *Dev. Cell* **2002**, *3*, 581-592.
- [27] Ligoxygakis, P., Pelte, N., Ji, C., Leclerc, V., Duvic, B., Belvin, M., Jiang, H., Hoffmann, J.A., Reichhart, J.M. *The EMBO Journal* **2002**, *21*, 6330-6337.
- [28] Danielli, A., Kafatos, F.C., Thanasis, G.L. *J. Biol. Chem.* **2003**, *278*, 4184-4193.
- [29] Beck, M., Theopold, U., Schmidt, O. *J. Insect Physiol.* **2000**, *46*, 1275-1283.
- [30] Pham, T.N., Hayashi, K., Takano, R., Itoh, M., Eguchi, M., Shibata, H., Tanaka, T., Hara, S. *J. Biochem.* **1996**, *119*, 428-434.
- [31] Brehélin, M., Boigegrain, R.A., Drif, L., Coletti-Previero, M.A. *Biochem. Biophys. Res. Commun.* **1991**, *179*, 841-846.
- [32] Nakakura, N., Hietter, H., Van Dorsselaer, A., Luu, B. *Eur. J. Biochem.* **1992**, *204*, 147-153.
- [33] Boigegrain, R.A., Matras, H., Brehelin, M., Paroutaud, P., Coletti-Previero, M.A. *Biochem. Biophys. Res. Commun.* **1992**, *189*, 790-793.

- [34] Boigegrain, R.A., Pugnieri, M., Paroutaud, P; Castro, B., Brehélin, M. *Insect Biochem. Mol. Biol.* **2000**, *30*,145-152.
- [35] Kellenberger, C., Boudier, C., Bermudez, I., Biet, J.G., Luu, B., Hietter, H. *J. Biol. Chem.* **1995**, *270*, 25514-25519.
- [36] Simonet, G., Claeys, I., Vanderperren, H., November, T., De Loof, A., Vanden Broeck, J. *Insect Mol. Biol.* **2002**, *11*, 249-256.
- [37] Simonet G., Claeys, I., November, T., Wataleb, S., Janssen, T., Maes, R., De Loof, A., Vanden Broeck, J. *Insect Mol. Biol.* **2002**, *11*, 353-360.
- [38] Liang, Z., Sottrup-Jensen, L., Aspán, A., Hall, M., Söderhäll, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6682- 6687.
- [39] Simonet, G., Claeys, I., Vanden Broeck, J. *Comp. Biochem. Physiol. B* **2002**, *132*, 247-255.
- [40] Mer, G., Kellenberger, C., Koehl, P., Stote, R., Sorokine, O., Van Dorsselaer, A., Luu, B., Hietter, H., Lefevre, J.F. *Biochemistry* **1994**, *33*, 15397-15407.
- [41] Mer, G., Hietter, H., Kellenberger, C., Renatus, M., Luu, B., Lefèvre, J.F. *J. Mol. Biol.* **1996**, *58*, 158-171.
- [42] Mer, G., Dejaegere, A; Stote, R., Kieffer, B., Lefevre, J.F. *J. Phys. Chem.* **1996**, *100*, 2667-2674.
- [43] Mer, G. ; Hietter, H., Lefevre, J.F. *Nat. Struct. Biol.* **1996**, *3*, 45-53.
- [44] Malik, Z., Amir, S., Pal, G., Buzas, Z., Varallyay, E., Antal, J., Szilagyi, Z., Vekey, K., Asboth, B., Pathy, A., Graf, L. *Biochim. Biophys. Acta* **1999**, *1434*, 143-150.

- [45] Gáspári, Z., Patthy, A., Gráf, L., Perczel, A. *Eur. J. Biochem.* **2002**, 269, 527-537.
- [46] Roussel, A., Mathieu, M., Dobbs, A., Luu, B., Cambillau, C., Kellenberger, C. *J. Biol. Chem.* **2001**, 276, 38893-38898.
- [47] Kellenberger, C., Ferrat, G., Leone, P., Darbon, H., Roussel, A. *Biochemistry* **2003**, 42, 13605-13612.
- [48] Patthy, A., Amir, S., Malik, Z., Bódi, A., Kardos, J., Asbóth, B., Gráf, L. *Arch. Biochem. Biophys.* **2002**, 398, 179-187.
- [49] Kromer, E., Nakakura, N., Lagueux, M. *Insect. Biochem. Molec. Biol.* **1994**, 24, 329-331.
- [50] Vanden Broeck, J., Chiou, S.J., Schoofs, L., Hamdaoui, A., Vandenbussche, F.; Simonet, G., Wataleb, S., De Loof A. *Eur. J. Biochem.* **1998**, 254, 90-95.
- [51] Harding, L., Scott, R.H., Kellenberger, C., Hietter, H., Luu, B., Beadle, D.J., Bermudez, I. *J. Recept. Signal. Transduct. Res.* **1995** 15, 355-364.
- [52] Clynen, E.; Baggerman, G., Veelaert, D., Cerstiaens, A., Van derHorst, D., Harthoorn, L., Derua, R., Waelkens, E., De Loof, A., Schoofs, L. *Eur. J. Biochem.* **2001**, 268, 1929-1939.
- [53] Hamdaoui, A., Schoofs, L., Wataleb, S., Devreese, B., Van Beeumen, J., Chiou, S., Vanden Broeck, J., De Loof A. *Biochem. Biophys. Res. Commun.* **1997**, 238, 357-360.
- [54] Scott, R.H., Gorton, V.J., Harding, L., Patel, D., Pacey, S., Kellenberger, C., Hietter, H., Bermudez, I. *Neuropharmacology* **1997**, 36, 195-208.
- [55] Rahman, M.M., Vanden Bosch, L., Baggerman, G., Clynen, E., Hens, K., Hoste, B.,

Meylaers, K., Vercammen, T., Schoofs, L., De Loof, A., Breuer, M. *Peptides* **2002**, *23*, 1907-1914.

[56] Clynen, E., Stubbe, D., De Loof, A., Schoofs, L. *Comp. Biochem. Physiol. B* **2002**, *132*, 107-115.

[57] Simonet, G., Claeys, I., Huybrechts, J., De Loof, A., Vanden Broeck, J. *Protein Expr. Purif.* **2003**, *31*, 188-196.

[58] Salzet, M. *FEBS Lett.* **2001**, *492*, 187-192.

[59] Chopin, V., Bilfinger, T.V., Stefano, G.B., Matias, I., Salzet, M. *Eur J. Biochem.* **1997**, *249*, 733-738

[60] Seemüller, U., Meier, M., Ohlsson, K., Muller, H.-P., Fritz, H. *Hoppe-seyler Z. physiol. chem.* **1977**, *358*, 1105-1117.

[61] Electricwala, A., Von Sicard, N.A.E., Sawyer, R.T., Atkinson, T. *J. Enzyme. Inhibition* **1993**, *6*, 293-302.

[62] McPhalen, C.A., Schnebli, H. P., James, M.N.G. *FEBS Lett* . **1985**, *188*, 55-58.
, *1434*, 143-150

[63] Laskowski, M., Qasim, MA. *Biochim. Biophys. Acta.* **2000**, *1477*, 324-37

[64] Chopin, V., Matias, I., Stefano, G.B., Salzet, M. *Eur. J. Biochem.* **1998**, *254*, 565-70.

[65] Chopin, V., Stefano, G.B., Salzet, M. *Eur. J. Biochem.* **1998**, *256*, 662-668

[66] Chopin, V., Salzet, M., Baert, J.L., Vandenbulcke, F., Sautière, P.E., Kerkaert, J.P., Malecha J. *J. Biol. Chem.* **2000**, *275*, 32701-32107.

- [67] Salzet, M., Vieau, D., Stefano, G.B. *Immunol. Today* **1999**, *20*, 541-544
- [68] Chopin V, Stefano GB, Salzet M. *J. Enzy. Inhib.* **2000**, *15*, 367-379.
- [69] Salzet, M., Chopin, V., Baert, J.L., Matias, I., Malecha J. *J. Biol. Chem.* **2000**, *275*, 0774-30780.
- [70] Söllner, C., Mentele, R., Eckerskorn, C. *Eur. J. Biochem.* **1994**, *219*, 937-943.
- [71] Moser, M., Auerswald, E., Mentele, R., Eckerskorn, C., Fritz, H., Fink, E. *Eur. J. Biochem.* **1998**, *253*, 212-20.
- [72] Jung, H.I., Kim, S.I., Ha, K.S. *J. Biol. Chem.* **1995**, *70*, 13879-13884.
- [73] **Kim, K.Y., Lim, H.K., Lee, K.J., Park, D.H., Kang, K.W., Chung, S.I., Jung, K.H.** *Protein Expr Purif.*, **2000**, *20*, 1-9.
- [74] Kim, D.R., Kang, K.W. *Eur. J. Biochem.* **1998**, *254*, 692-697.
- [75] Kim, D.R., Hong, S.J., Ha, K.S., Joe, C.O., Kang, K.W. *J. Enzyme Inhib.* **1996**, *10*, 81-91.
- [X] Salzet, M. *Curr. Pharm. Des.* **2002**, *8*, 493-503.

Fig. (1). Sequence alignment of the locust 4 kDa serine protease inhibitors. The conserved (cysteine) residues are in red. A consensus sequence displaying the disulfide bridges is below. The P1 residue is indicated by an arrow. Fucosylated Thr-residues are in italic.

Fig. (2). MALDI-TOF spectrum of *Schistocerca corpora cardiaca* indicating the presence of the SGPI-2 ion peak.

Fig. (3). Photographs of *Theromyzon tessulatum* gut.

Fig. (4). Photographs of *Hirudo medicinalis* jaws.

Fig. (5). SDS-PAGE analysis of tessulin.

Fig. (6). SDS-PAGE analysis of anti-trypsin A (1) and anti-trypsin B (2)

Fig. (7). Sequence alignment between Theromin, Therostasin and Tessulin obtained using Multalin software : <http://www.toulouse.inra.fr/centre/serWWW.htm>

Table 1. Comparison of Active Site Sequences of the Insect 4 kDa Serine Protease Inhibitors