Peptides 41 (2013) 59–65

Contents lists available at SciVerse ScienceDirect

Peptides

journal homepage: www.elsevier.com/locate/peptides

# Phylogenetic distribution of protease inhibitors of the Kazal-family within the Arthropoda

Vincent van Hoef*, Bert Breugelmans, Jornt Spit, Gert Simonet, Sven Zels, Jozef Vanden Broeck

*Department of Animal Physiology and Neurobiology, Zoological Institute K.U.Leuven, Naamsestraat 59, PO Box 02465, B-3000 Leuven, Belgium*

ARTICLE INFO

*Article history:*
Received 29 May 2012
Received in revised form 30 October 2012
Accepted 30 October 2012
Available online 15 November 2012

*Keywords:*
PSTI
Kazal
Protease inhibitor
Arthropod
Phylogeny

ABSTRACT

In mammalian pancreatic cells, the *pancreatic secretory trypsin inhibitor* (PSTI) belonging to the Kazal-family prevents the premature activation of digestive enzymes and thus plays an important role in a protective mechanism against tissue destruction by autophagy. Although a similar protective mechanism exists in Arthropoda, the distribution of these inhibitors in this phylum remains obscure. A comprehensive in silico search of nucleotide databases, revealed the presence of members of the Kazal-family in the four major subphyla of the Arthropoda. Especially in the Hexapoda and the Crustacea these inhibitors are widespread, while in the Chelicerata and Myriapoda only a few Kazal-like protease inhibitors were found. A sequence alignment of inhibitors retrieved in the digestive system of insects revealed a conservation of the PSTI characteristics and strong resemblance to vertebrate PSTI. A phylogenetic analysis of these inhibitors showed that they generally cluster according to their order. The results of this data mining study provide new evidence for the existence of an ancient protective mechanism in metazoan digestive systems. Kazal-like inhibitors, which play an important protective role in the pancreas of vertebrates, also seem to be present in Arthropoda.



## 1. Introduction

A healthy mammalian pancreas delivers its stored inactive digestive enzyme precursors to the gut. Once secreted in the gut lumen, the digestive enzymes are activated in a cascade-like manner, starting with the cleavage of trypsinogen into the active trypsin by enterokinase. Subsequently, trypsin activates the other pro-proteases needed for the digestion of proteins in the food [10]. To prevent premature activation of digestive enzymes and the resulting damage to cells and tissues, mammals express a protease inhibitor in their pancreatic cells called 'pancreatic secretory trypsin inhibitor' (PSTI) [11,16]. This inhibitor is a member of the numerous and diverse Kazal-family of protease inhibitors and binds to prematurely activated trypsin, interrupting the activation cascade. Typically, a mammalian Kazal-type serine protease inhibitor domain is 40–60 amino acids long, has a molecular mass of approximately 6500 Da and has a conserved domain architecture [3,13]. This Kazal domain or motif has a general amino acid sequence of $Cys_I$-$X_a$-$Cys_{II}$-$X_b$-PVC$ys_{III}$G-$X_c$-TY-$X_d$-$Cys_{IV}$-$X_e$-$Cys_V$-$X_f$-G-$X_g$-$Cys_{VI}$ where the subscripts a, b, c, d, e, f and g are integral numbers of amino acid residues. The six cysteine residues form three intra-domain disulfide bridges in a characteristic pattern ($Cys_I$-$Cys_V$, $Cys_{II}$-$Cys_{IV}$ and $Cys_{III}$-$Cys_{VI}$) contributing to a tight three-dimensional conformation. The second amino acid C-terminal to the second cysteine is the important specificity determining P1-residue (see Fig. 2) [31].

The 'authentic' PSTI occurs only in vertebrate species [1,8,11,23,24,41], which is logical since they are the only animal group with an actual pancreas. However, a study in 2006 revealed the presence of the PSTI-homolog Kazal1 in the endodermal cell layer of the body column of the freshwater polyp, *Hydra magnipapillata* [6]. The cells expressing Kazal1 showed the specific morphology of gland cells, with a multi-vacuolated cytoplasm corresponding to large secretory vesicles. Considering the similar structure and cellular localization, the authors speculated that Kazal1's function was to prevent the premature activation of digestive enzymes. An RNAi-mediated knockdown of this inhibitor resulted in severe autophagy of gland cells and epithelial digestive cells. This phenotype resembles pancreatitis – a disease which is associated with dysfunctional PSTI in vertebrates. So, even though *Hydra* does not possess a pancreas, a similar protease inhibitor seems to have the same protective role as PSTI has in vertebrates. Cnidaria – the phylum *Hydra* belongs to – are an evolutionary sister group to the Bilateria (including vertebrates and arthropods),

* Corresponding author. Tel.: +32 16324260; fax: +32 16323902.
*E-mail addresses:* Vincent.vanHoef@bio.kuleuven.be, vvanhoef@hotmail.com (V. van Hoef), Bert.Breugelmans@bio.kuleuven.be (B. Breugelmans), Jornt.Spit@bio.kuleuven.be (J. Spit), Gert.Simonet@bio.kuleuven.be (G. Simonet), Sven.Zels@bio.kuleuven.be (S. Zels), Jozef.VandenBroeck@bio.kuleuven.be (J. Vanden Broeck).

0196-9781/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.peptides.2012.10.015

suggesting an ancient origin for this protective mechanism in a common ancestor. Since Arthropoda emerged only after the divergence of Cnidaria and Bilateria more than 580 Mya [5], a comparable defensive role for PSTI-like proteins in arthropods seems plausible. Further evidence for a similar function in arthropods was observed in the locust *Locusta migratoria* [36]. Again, an RNAi-mediated knockdown of a locust-specific PSTI-like protein (LmPSTI) resulted in a phenotype not unlike that observed in mammals suffering from pancreatitis (i.e. the occurrence of autophagy in the insect's caeca).

This manuscript reports on a detailed, comparative in silico data mining study of PSTI-like protein precursors among Arthropoda – with a special emphasis on the distribution in Insecta. To our knowledge, no arthropod PSTI-homologs have been characterized as such so far (except for LmPSTI in *L. migratoria*). However, there have been numerous reports of Kazal-type family protease inhibitors (PSTI being an important member of the Kazal-type family inhibitors) with a variety of suggested functions in various arthropod classes and orders including species, such as *Bombyx mori* [42], *Rhodnius prolixus* [9], *Aedes aegyptii* [39] and several crustaceans [4,7,19,38].

## 2. Materials and methods

### 2.1. Database searches

Arthropod homologs of PSTI were searched for in silico in EST (Expressed Sequence Tags) databases using a list of 15 known vertebrate PSTI precursor sequences from the Uniprot protein database (http://www.uniprot.org) as a query (see additional data). To check and further complement previous search results, novel arthropod sequences were used as additional queries. Detailed searches were performed on a vast number of nucleotide sequences available via the BLAST (Basic Local Alignment Search Tool) function accessible on the NCBI platform (http://www.ncbi.nlm.nih.gov/BLAST/). Complementary to this, various databases of individual species containing new sequences that are not yet incorporated into the NCBI platform were analyzed (BeetleBase; beetlebase.org, FlyBase; flybase.org).

### 2.2. Strategy

The data mining search was performed in a systematic way, covering as many classes and orders as possible. Per arthropod subphylum, EST sequence data sets were analyzed for the presence of PSTI homologs and, whenever this was possible, additional searches were performed in the classes, orders and families of the subphylum. Since Kazal-like inhibitors (the peptide family PSTI belongs to) are a very numerous family with many and diverse functions, a distinction was made between EST sequences originating from the organisms' digestive system and others. This distinction was made to focus on the novel inhibitors that potentially have a similar role as the mammalian PSTI – namely to protect the organism from damage due to prematurely activated digestive enzymes.

### 2.3. Sequence comparison and phylogenetic analysis of insect PSTI-like proteins

The PSTI-like precursors obtained from insect EST sequences originating from the digestive system were restricted in length to the two exterior Cys residues ($Cys_I$ to $Cys_{VI}$) and aligned using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and visualized in the perl/tk-based tool Aline (http://crystal.bcs.uwa.edu.au/px/charlie/software/aline/). The alignment file which resulted from the ClustalW alignment was used as an input file to visualize the phylogeny of the PSTI-like precursors in the free software program PHYLIP (PHYLogeny Inference Package) (http://evolution.genetics.washington.edu/phylip.html) using the maximum parsimony method.

## 3. Results

By almost any measure, arthropods are the most diverse and successful group of animals on the planet. They make up over three quarters of all known currently living and fossil organisms, and are represented by – according to scientists' recent estimates – 3.7 million species [14]. So far, however, 'only' around 855,000 species have been described, which leaves approximately 75% of the contemporary arthropod species awaiting scientific description. This large number of species is divided into four subphyla: Myriapoda (centipedes, millipedes, etc.), Chelicerata (horseshoe crabs, spiders and pycnogonids), Crustacea (shrimps, crabs, etc.) and Hexapoda (insects and springtails). As described in Section 2, a similarity search was performed to discover novel Kazal-like inhibitors and potential PSTI homologs. The results are visualized in Fig. 1.

### 3.1. PST-like inhibitors in Myriapoda

Myriapoda is a subphylum of the Arthropoda containing among others the centipedes and the millipedes (hence the name Myriapoda, which means 10,000 legs in Greek). They are subdivided in four subphyla: the Chilopoda, the Diplopoda, the Pauropoda and the Symphyla. Genetic data of this animal group are scarce, with only 4461 EST sequences and 3893 nucleotide sequences submitted to the NCBI platform (and almost completely confined to the diplopodans). A similarity search through EST data, revealed one significant hit in the African giant black millipede, *Archispirostreptus gigas*. Since the *A. gigas* EST sequences originated from a whole animal, a further tissue distribution was not available. Although several Kazal-like inhibitor features are present (6 cysteines, conserved amino acids) and the P1-residue is a Lys (suggesting trypsin inhibitory specificity), other characteristics differ from a typical Kazal-domain (most notably, a remarkably short amino acid stretch between $Cys_V$ and $Cys_{VI}$). In the Chilopoda, Pauropoda and Symphyla, no PSTI-homolog could be discovered.

### 3.2. PSTI-like inhibitors in Chelicerata

The subphylum of Chelicerata is currently subdivided in three classes: the Arachnida (arachnids), the Pycnogonida (sea spiders) and the Merostomata (horseshoe crabs). Genetic data about this subphylum are more abundant in comparison with data about Myriapoda – especially the Arachnida are well-represented. Searching this relative abundance of nucleotide information did, however, not result in a large number of Kazal-like inhibitors. In the Southeast Asian horseshoe crab, *Carcinoscorpius rotundicauda* (a member of the Merostomata), and the Black-legged tick, *Ixodes scapularis* (a member of the Arachnida), respectively two and one putative Kazal-like inhibitors were found. Not surprisingly, these two species are greatly overrepresented in the nucleotide information available about their class (46.5% and 93.5% for Ixodes and Carcinoscorpius, respectively). A further localization in the digestive system could not be deduced for these inhibitors. A search through the very recently sequenced genome of the spider mite, *Tetranychus urticae* revealed the presence of two putative Kazal-type inhibitors [12]. Searching through currently available pycnogonid nucleotide information did not yield potential Kazal-type inhibitors.

### 3.3. PSTI-like inhibitors in Crustacea

Crustacea is a large and diverse subphylum of the Arthropoda, including many commercially (and gastronomically) important