Insect Biochemistry and Molecular Biology 39 (2009) 782–791

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

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

Sequence and function of lysosomal and digestive cathepsin D-like proteinases of *Musca domestica* midgut

Marcelo H.P. Padilha [a], André C. Pimentel [a], Alberto F. Ribeiro [b], Walter R. Terra [a,*]

[a] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, C.P. 26077, 05513-970 São Paulo, Brazil
[b] Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, C.P. 11461, 05422-970 São Paulo, Brazil

ARTICLE INFO

Article history:
Received 18 June 2009
Received in revised form
16 September 2009
Accepted 17 September 2009

Keywords:
Aspartic proteinases
Cathepsins
Digestive cathepsin
Lysosomal cathepsin
Cathepsin secretion
Cathepsin function
Proline loop motif

ABSTRACT

Musca domestica larvae display in anterior and middle midgut contents, a proteolytic activity with pH optimum of 3.0–3.5 and kinetic properties like cathepsin D. Three cDNAs coding for preprocathepsin D-like proteinases (ppCAD 1, ppCAD 2, ppCAD 3) were cloned from a *M. domestica* midgut cDNA library. The coded protein sequences included the signal peptide, propeptide and mature enzyme that has all conserved catalytic and substrate binding residues found in bovine lysosomal cathepsin D. Nevertheless, ppCAD 2 and ppCAD 3 lack the characteristic proline loop and glycosylation sites. A comparison among the sequences of cathepsin D-like enzymes from some vertebrates and those found in *M. domestica* and in the genomes of *Aedes aegypti*, *Drosophila melanogaster*, *Tribolium castaneum*, and *Bombyx mori* showed that only flies have enzymes lacking the proline loop (as defined by the motif: DxPxPx(G/A)P), thus resembling vertebrate pepsin. ppCAD 3 should correspond to the digestive cathepsin D-like proteinase (CAD) found in enzyme assays because: (1) it seems to be the most expressed CAD, based on the frequency of ESTs found. (2) The mRNA for CAD 3 is expressed only in the anterior and proximal middle midgut. (3) Recombinant procathepsin D-like proteinase (pCAD 3), after auto-activation has a pH optimum of 2.5–3.0 that is close to the luminal pH of *M. domestica* midgut. (4) Immunoblots of proteins from different tissues revealed with anti-pCAD 3 serum were positive only in samples of anterior and middle midgut tissue and contents. (5) CAD 3 is localized with immunogold inside secretory vesicles and around microvilli in anterior and middle midgut cells. The data support the view that on adapting to deal with a bacteria-rich food in an acid midgut region, *M. domestica* digestive CAD resulted from the same archetypical gene as the intracellular cathepsin D, paralleling what happened with vertebrates. The lack of the proline loop may be somehow associated with the extracellular role of both pepsin and digestive CAD 3.

1. Introduction

The evolutionary origin of an acidic midgut region in vertebrates seems to result from an adaptation to lyse bacteria to be used as food by detritus-feeding vertebrates, a condition still found in some fishes as tilapia (Bowen, 1976). Pepsin, a proteolytic digestive enzyme, active at low pH, appeared later in evolution, probably derived from the same archetypical gene as the intracellular cathepsin D (Takahashi and Tang, 1981). This view is supported by the finding that gastric acid digestion in non-detritus-feeding vertebrates is not a necessary step in digestion, as the surgical removal of the stomach does not affect the completeness of digestion in operated animals (Vonk and Western, 1984).

The ability of Cyclorrhaphous Diptera to digest bacteria that infest their food evolved in parallel to that of vertebrates. Thus, flies are the only animals, other than vertebrates, which display such an acid region (pH 3.0–3.2) in their midgut (Vonk and Western, 1984). Furthermore, they have a cathepsin D-like proteinase and a lysozyme (absent from tissue other than the midgut) active at the acid region of the midgut (Greenberg and Paretsky, 1955; Espinoza-Fuentes et al., 1987; Lemos and Terra, 1991a,b; 1993). The adaptations to digest bacteria were probably present in the Cyclorrhapha ancestor, since these adaptations are found in both the distantly related species *Musca domestica* and *Anastrepha fraterculus*. The absence of adaptations to digest bacteria in the Nematocera detritus-feeding *Trichosia pubescens* (Sciaridae) suggests that these adaptations were not present in the Diptera ancestor.

Cathepsin D (EC 3.4.23.5) and pepsin (EC 3.4.23.1) are endopeptidases characterized by the presence of two catalytic aspartic residues, activity at acid pH values and inhibition by pepstatin. They share the same family A1 of aspartic proteinases that include only

* Corresponding author. Fax: +55 3091 2186.
E-mail address: warterra@iq.usp.br (W.R. Terra).

0965-1748/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2009.09.003

eukaryotic enzymes. Cathepsin D is a major glycosylated lysosomal aspartic proteinase, which is involved in the intracellular degradation of proteins (Tang and Wong, 1987) in species ranging from protozoa to higher animals (Barrett, 1970). Pepsin is a vertebrate digestive enzyme active in stomachs (Barrett et al., 2004).

The first report of aspartic proteinases in insects was made, as mentioned above, by Greenberg and Paretsky (1955), who found a strong proteolytic activity at pH 2.5–3.0 in homogenates of whole bodies of *M. domestica*. Lemos and Terra (1991b) showed that the enzyme is digestive and cathepsin D-like, based on pepstatin inhibition and its substrate preferences and sensibility to diazoacetylnorleucine plus $CuSO_4$.

An aspartic proteinase similar to cathepsin D was found in families of Hemiptera Heteroptera and in several families belonging to the cucujiform series of Coleoptera (Terra and Ferreira, 1994). Some of the enzymes are lysosomal like those involved in yolk protein degradation in eggs (Fialho et al., 2005) and other are said to be midgut digestive enzymes. Among the last ones the aspartic proteinases isolated from *Callosobrucchus maculatus* (pH optimum 3.3, 62 kDa) (Silva and Xavier-Filho, 1991) and *Tribolium castaneum* (pH optimum 3.0, 22 kDa) (Blanco-Labra et al., 1996) were partially purified and shown to be similar to cathepsin D, but they have not been sequenced. As digestive cathepsin D-like enzymes are thought to be part of the mechanisms that permit insects to survive on serine proteinase inhibitor-rich seeds (Terra, 1988; Silva and Xavier-Filho, 1991), in addition to rely on bacteria exemplified by flies, a better understanding of those enzymes are required. Specifically the origin, specificity, and structure of insect cathepsin D-like enzymes (CADs) need to be clarified. The data to be gathered may be used in the development of new insect control procedures that have these enzymes as targets.

This work was undertaken with the following aims: (1) to identify the digestive *M. domestica* CAD among the CADs that were sequenced and cloned; (2) to look for structural differences between insect lysosomal and digestive CADs; (3) to confirm the occurrence of digestive CADs only in Cyclorrhapha Diptera. The results showed that digestive CADs lack glycosylation and the proline loop and that this kind of CAD is absent from Coleoptera, Nematoceran Diptera and Lepidoptera and is present in Diptera Cyclorrhapha. Furthermore, *M. domestica* digestive CADs are synthesized and secreted only by the anterior and middle midguts.

2. Materials and methods

2.1. Animals

Larvae of *M. domestica* (Diptera, Cyclorrhapha, Muscidae) were reared in a mixture of fermented commercial pig food and rice hull (1:2, v/v) (Targa and Peres, 1979). Groups of 100 larvae were rinsed with distilled water and were placed on layers (50 mL) of 10% corn starch gels in 250 mL beakers. After standing on the gels for 100 min at 25 °C, the larvae were dissected. The larvae used in this study were actively feeding individuals at third larval instar.

2.2. Preparation of midgut tissue samples

Larvae were rinsed in water, blotted with filter paper, immobilized by placing them on ice and dissected in cold 150 mM NaCl. The midgut was divided into 8 sections: anterior (sections 1–3), middle (4–6) and posterior (7–8) midgut (Fig. 1). The tissues (midgut sections, hemolymph, fat body, and Malpighian tubules) were then homogenized in double distilled water with the aid of a Potter–Elvehjem homogenizer and centrifuged at 20,000g for 30 min at 4 °C. All the above preparations (except hemolymph) were passed through glass wool. When midgut tissue was freed from contents, the homogenates were not passed through glass wool. All homogenates were stored at −20 °C.

2.3. Chemicals

The plasmid Wisard miniprep System and the p-GEM-T Easy Vector plasmid kits were purchased from Promega (Madison, WI); the DNA gel extraction kit was from Qiagen (Germany), the agar and agarose were from Life Technologies and the dNTPs and modification and restriction enzymes from New England Biolabs (Berveley, MA). All other chemicals were purchased from Merck (Darmstadt, Germany) or Aldrich–Sigma (USA) unless otherwise stated.

2.4. Molecular cloning of midgut cathepsin D-like proteinases (CADs) from M. domestica

Total RNA was extracted from midgut epithelium of *M. domestica* larvae with Trizol following the instructions of the manufacturer, Invitrogen, which are based on Chomczynski and Sacchi (1987), and sent to Stratagene (La Jolla, CA), in order to construct a cDNA library. At Stratagene the mRNAs were isolated, divided into two equal samples and used in cDNA synthesis with a poly-T and a random primer. Finally, the two cDNA pools were mixed (1:1) and non-directionally inserted in the vector λ ZAPII. The library titer is 1.1×10^6 pfu/mL.

Three cDNAs containing all the putative ORF corresponding to preprocathepsin D-like proteinases (ppCADs) were found and cloned from a cDNA library prepared from *M. domestica* larval midgut. For this, partial sequences were amplified using a degenerate primer (CADdeg, Table 1), successfully used before for cathepsin D (Harrop et al., 1996), with T3 universal primer, employing the cDNA library as a template. The PCR was performed using TAQ DNA polymerase (2.5 units) in 20 mM Tris–HCl buffer, pH 8.4 with 50 mM KCl, 0.2 mM dNTP, and 4.0 mM $MgCl_2$. Before the amplification reaction, the medium was maintained for 5 min at 94 °C. The amplification was reached using 30 cycles at the

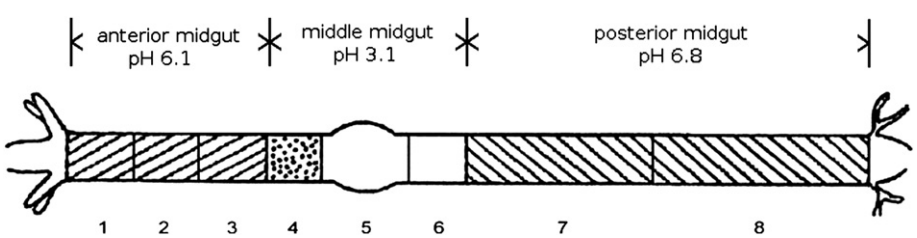

Fig. 1. Diagrammatic representation of the midgut of *M. domestica* larva showing its main regions with respective luminal pH values, based on Terra et al. (1988). The main regions were further divided into the numbered sections. All sections contain columnar cells, whereas section 4 includes also specialized cells (interstitial and oxyntic cells). Columnar cells in section 5 are flattened and almost featureless.