Insect Biochemistry and Molecular Biology 64 (2015) 51–59

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

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

Review

Cathepsins of lepidopteran insects: Aspects and prospects

Nidhi Saikhedkar [a], Aarohi Summanwar [b], Rakesh Joshi [c, **], Ashok Giri [a, *]

[a] Biochemical Sciences Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pashan, Pune, 411008, MS, India
[b] Department of Agriculture, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, T6G 2P5, Canada
[c] Institute of Bioinformatics and Biotechnology, Savitribai Phule Pune University, Ganeshkhind, Pune, 411007, MS, India

ARTICLE INFO

Article history:
Received 21 May 2015
Received in revised form
20 June 2015
Accepted 6 July 2015
Available online 22 July 2015

Keywords:
Cathepsin
Cysteine/aspartate protease
Lepidoptera
Development
Apoptosis
Stress response
Pest control

ABSTRACT

Molecular understanding of lepidopteran physiology has revealed that proteases consist of one of the central regulatory/reacting system for insect growth and survival. Among the various proteases, cathepsins are the most crucial cellular proteases, which play vital roles during insect development. In the present review, we have discussed various aspects of the lepidopteran insect cathepsins, emphasizing their roles in processes like development, growth, metamorphosis, apoptosis and immunity. Cathepsins are categorized into different types on the basis of their sequence diversification, leading to variation in structure and catalytic function. Cathepsins exhibit tissue and stage specific expression pattern which is fine-tuned by a delicate balance of expression, compartmentalization, zymogen activation, inhibition by protein inhibitors and degradation. The indispensability of cathepsins as cellular proteases in the above mentioned processes proposes them as novel targets for designing effective and specific insect controlling strategies.

1. Introduction

Lepidoptera is the third largest order of the class Insecta consisting over 175,000 species spread all over the world (Capinera, 2008). Their life cycle is driven by a plethora of metabolic enzymes, including proteases which function in a tissue specific manner during essential processes such as digestion, development and immune response. The digestive process is dominated by serine proteases like trypsin, chymotrypsin and elastase in the alkaline pH of the midgut (Srinivasan et al., 2006). On the contrary, developmental processes like metamorphosis and eclosion are carried out by acidic pH favouring enzymes, like cathepsins, which are typically stored in the lysosomes (Turk et al., 2000). On the basis of sequence diversity and variable activity in mammals, cathepsins are categorized into different types. Cathepsin A and G are similar to serine proteases; D and E are grouped as aspartic acid proteases, whereas others, namely B, C, F, H, K, L, O, S, V, W and X are cysteine proteases (Karrer et al., 1993; Turk et al., 2012; Wex et al., 1999). Out of the 2000 arthropod cathepsin protein sequences in the NCBI, 50 belong to the lepidoptera including partial and full-length sequences. Till date, only cathepsin B, L, F, O and D have been reported in lepidoptera (Gui et al., 2006; Homma et al., 2013). In this review, we have catalogued the numerous physiological roles of cathepsins in lepidopteran insects, with sequence analysis to understand the evolutionary relationship and three dimensional structural similarities. Furthermore, we have suggested their use as a potential target(s) for development of an effective and/or specific insect controlling strategy for sustainable agriculture either alone or in combination.

2. Sequence and structural diversity

Sequence based phylogenetic analysis suggests the distribution of lepidopteran cathepsins in two major clads: Cysteine cathepsins (Cath B, L, F and O) and aspartate cathepsins (CathD) with human extracellular matrix protein as an outgroup (Fig. 1). Majority of them belong to the cysteine protease type, that is further diverged into CathB and L subgroups. The subgroup L consists of CathL, F and O cathepsins. It is observed that within the clads, different lepidopteran species possess related cathepsins which could be indicating an evolutionary divergence among them. However, a few CathL (DpCathL and CsCathL-4) are distantly related to the other CathL types. This may be because of their low sequence identity

* Corresponding author.
** Corresponding author.
E-mail addresses: rakesh.joshi@unipune.ac.in, rakeshjoshi687@gmail.com (R. Joshi), ap.giri@ncl.res.in (A. Giri).

http://dx.doi.org/10.1016/j.ibmb.2015.07.005
0965-1748/

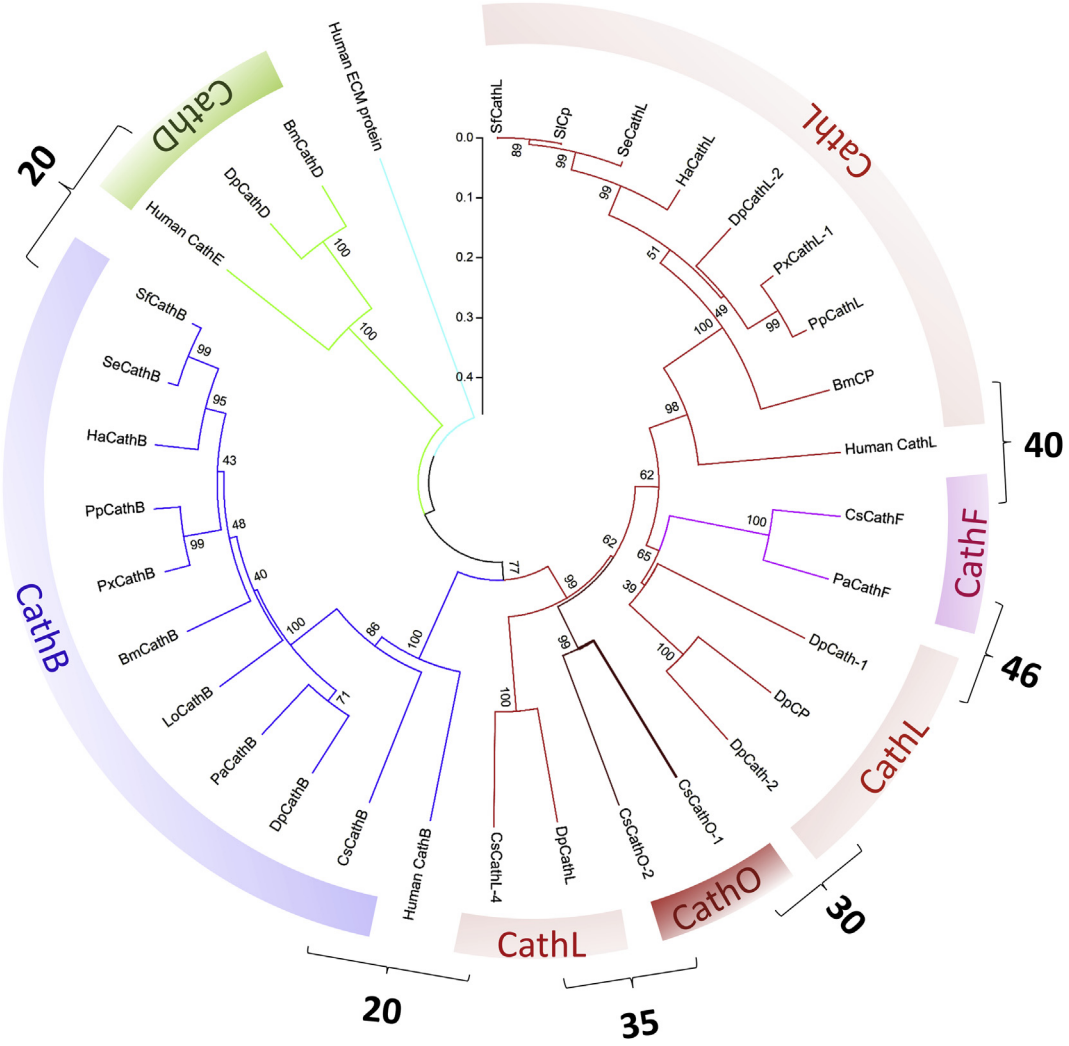

Fig. 1. Diversity of lepidopteran cathepsins: Phylogram of cathepsins from lepidopteran species (NCBI accession numbers mentioned in Supplementary table 1) was generated by MEGA program (Tamura et al., 2007) using the Gonnet protein weight matrix and clustering was done using a neighbour-joining algorithm with 1000 bootstrap iterations and p-distance amino acid substitution model. Blue, red, green, magenta, brown and cyan represent CathB, L, E, F, O, and Human ECM protein respectively. Numbers indicate percentage sequence identities among adjacent clads as calculated by BLAST tool from NCBI.

with other CathL proteases (<35%).

Sequence conservation amongst the lepidopteran cathepsins reflects in domain architecture and three dimensional structural similarities (Fig. 2). All the enzymes exhibit protease specific domain architecture with common features, like the presence of signal peptides, followed by propeptide and protease domain. Signal peptides designate their organelle targeted delivery under the control of intrinsic propeptide inhibitors. The length of the propeptide region varies among various cathepsin types, similar to human cathepsins (Novinec and Lenarčič, 2013). CathL like proteases have ~60 amino acid propeptides, while the CathB and D subgroups possess shorter domains of 40 and 30 amino acids, respectively. Also, additional cystatin like inhibitor domains are present at the N-terminus of propeptides in the CathF subgroup (Fig. 2). The proregion of insect CathL is characterized by the interspersed conserved sequence motif 'ERFNIN' (EX_3RX_2(V/I) $FX_2NX_3IX_3N$), which is absent in CathB and O (Fig. 3) (Karrer et al., 1993; Liu et al., 2006). However, lepidopteran CathF contains the ERFNIN motif instead of the 'ERFNAQ' motif reported from CathF in other species (Martinez and Diaz, 2008; Wex et al., 1999). Functional domains of insect cysteine cathepsins are homologous to "papain fold", which is made up of two subdomains, L and R, composed of three alpha helices and a beta barrel, respectively (Drenth et al., 1968; Turk et al., 1998) (Fig. 4). Thus, they are classified into the class A (CA), more specifically, to the clan 1 (C1) family of papain like enzymes (Rawlings et al., 2014). The active site of cysteine cathepsins comprises of a thiolate-imidazolium ion pair formed by Cys–His diad coming from individual subdomains (Fig. 3). Cath O, F and L are known to form a structurally similar group, correlating with their grouping in same clade of L-type cysteine cathepsins in Fig. 1 (Novinec and Lenarčič, 2013; Turk et al., 2000). However, CathB like proteases are characterized by the presence of four disulphide bonds and a 20 amino acid insertion, called the "occluding loop" in addition to the basic papain fold. The binding of occluding loop regulates specific activity of CathB at acidic and neutral pH. At low pH, histidine present at the end of the loop is protonated enabling the occluding loop to reside in the active site. Thus, it acts as a carboxydipeptidase by docking of the carboxy terminal end of the substrate in the active site. Whereas at high pH, deprotonation of His residue occurs, shifting the loop out