Insect Biochemistry and Molecular Biology 89 (2017) 17–30

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

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

Active subsite properties, subsite residues and targeting to lysosomes or midgut lumen of cathepsins L from the beetle *Tenebrio molitor*

Ticiane F. Damasceno [a], Renata O. Dias [a], Juliana R. de Oliveira [b], Roberto K. Salinas [a], Maria A. Juliano [b], Clelia Ferreira [a], Walter R. Terra [a, *]

[a] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Avenida Professor Lineu Prestes, 748, São Paulo 05508-000, Brazil
[b] Departamento de Biofísica, Universidade Federal de São Paulo, Rua Três de Maio, 100, São Paulo 04044-020, Brazil

ARTICLE INFO

Article history:
Received 5 July 2017
Received in revised form
14 August 2017
Accepted 20 August 2017
Available online 29 August 2017

Keywords:
Cathepsin L-like peptidase
Substrate specificity
Subsite role
Subsite hydrophobicity
Phosphotransferase
Cathepsin L activation
Tenebrio molitor

ABSTRACT

Cathepsins L are the major digestive peptidases in the beetle *Tenebrio molitor*. Two digestive cathepsins L (TmCAL2 and TmCAL3) from it had their 3D structures solved. The aim of this paper was to study in details TmCAL3 specificity and properties and relate them to its 3D structure. Recombinant TmCAL3 was assayed with 64 oligopeptides with different amino acid replacements in positions P2, P1, P1′ and P2'. Results showed that TmCAL3 S2 specificity differs from the human enzyme and that its specificities also explain why on autoactivation two propeptide residues remain in the enzyme. Data on free energy of binding and of activation showed that S1 and S2′ are mainly involved in substrate binding, S1′ acts in substrate binding and catalysis, whereas S2 is implied mainly in catalysis. Enzyme subsite residues were identified by docking with the same oligopeptide used for kinetics. The subsite hydrophobicities were calculated from the efficiency of hydrolysis of different amino acid replacements in the peptide and from docking data. The results were closer for S1 and S2′ than for S1' and S2, indicating that the residue subsites that were more involved in transition state binding are different from those binding the substrate seen in docking. Besides TmCAL1-3, there are nine other cathepsins L, most of them more expressed at midgut. They are supposed to be directed to lysosomes by a Drosophila-like Lerp receptor and/or motifs in their prodomains. The mannose 6-phosphate lysosomal sorting machinery is absent from *T. molitor* transcriptome. Cathepsin L direction to midgut contents seems to depend on overexpression.

1. Introduction

Cysteine endopeptidases (EC 3.4.22) are hydrolases in which a cysteine present in its active site is responsible for a nucleophilic attack on a substrate scissile peptide bond. All cysteine peptidases have a proton acceptor histidine and some also depend on a third residue, which orients this histidine imidazole ring (Barrett et al., 2013). Cathepsins are cysteine peptidases that were initially described in the protein degradation within the lysosomes. Because of that, at first all intracellular peptidases were named cathepsins (Barrett, 1992). Nowadays, it is known that all lysosomal cysteine peptidases are cathepsins, but not all cathepsins are lysosomal nor cysteine peptidases (Turk and Guncar, 2003).

Cathepsin L was first described as a rat liver lysosomal peptidase which preferentially cleaves substrates with a hydrophobic amino acid in P2 (namely following Schechter and Berger, 1967; see Fig. 1) and may play a role in malignant tumor growth (Barrett et al., 2013). In insects, cysteine peptidases may be involved in digestion (Terra and Ferreira, 1994), embryonic vitellin degradation (Cho et al., 1999), and in metamorphosis (Takahashi et al., 1993). Digestive cysteine peptidases occur as digestive enzymes in Hemiptera Heteroptera (bugs) and are found along with serine peptidases in the Coleoptera infraorder Cucujiformia (major beetles) (Terra and

Abbreviations: Abz, *ortho*-aminobenzoic acid; DTT, dithithreitol; E64, *trans*-epooxysuccinyl-L-leucylamido-(4-guanidino)butane; EDDnp, *N*-(2,4-dinitrophenyl) ethylenediamine; FRET, fluorescence resonance energy transfer; LERP, lysosomal enzyme receptor protein; M6P, mannose 6–phosphate; MPR, mannose-6-phosphate receptor; phosphotransferase, N-acetylglucosaminidase-1-phosphate transferase; uncovering enzyme, N-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase; TGN, trans-Golgi network; Z-FR-MCA, benzyloxycarbonyl-L-phenylalanyl-L-arginine 4-methylcoumarin-7-amide.

* Corresponding author.
E-mail address: warterra@iq.usp.br (W.R. Terra).
URL: http://www2.iq.usp.br/docente/warterra/

http://dx.doi.org/10.1016/j.ibmb.2017.08.004
0965-1748/

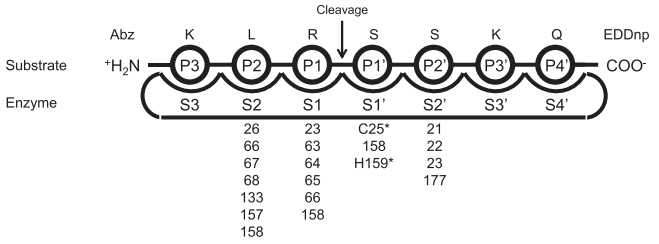

Fig. 1. Diagrammatic representation of an insect CAL bound to a fluorescence energy transfer (FRET) peptide. The peptide bond cleaved by CAL is that between R and S. Varying the amino acid residues at the positions P2, P1, P1′, and P2′ it is possible to describe the specificities of the subsites S2, S1, S1′, and S2′, respectively. Figures below each subsite correspond to the positions of the residues. Residues are numbered according to papain.

Ferreira, 1994). In this last case, the presence of digestive cysteine peptidase is thought to be an adaptation to the ingestion of seeds rich in serine peptidase inhibitors (Terra and Cristofoletti, 1996). Cathepsin L-like (CAL) enzymes were revealed as the only quantitatively important insect digestive cysteine peptidase (Terra and Ferreira, 1994, 2012).

Because of the importance of insect CALs, which seem to be involved in resistance mechanisms against natural inhibitors (Hilder and Boulter, 1999) and may have peculiar secretory routes (Cristofoletti et al., 2005), there is a need for a better characterization of these enzymes. Cathepsin L-like peptidases are mainly present in the anterior and middle midgut of the beetle *Tenebrio molitor*, an important model in Coleoptera and a cosmopolitan pest of stored products (Thie and Houseman, 1990; Terra and Cristofoletti, 1996; Vinokurov et al., 2006). There are at least three isoforms of a truly lysosomal CAL (TmCAL1a, TmCAL1b, TmCAL1c) and two digestive ones (TmCAL2 and TmCAL3), based on immunoblotting and immunocytolocalization in different *T. molitor* tissues (Cristofoletti et al., 2005; Beton et al., 2012). N-terminal sequencing of the purified *T. molitor* major digestive CAL performed in our group confirmed that TmCAL2 is the major CAL recovered activity in its midgut (Cristofoletti et al., 2005), while the major CAL recovered activity in USDA (Manhattan) strain of *T. molitor* is TmCAL3 (Prabhakar et al., 2007).

Sorting of lysosomal cathepsin L in mammalian species is mainly correlated to the mannose 6-phosphate (M6P) pathway. In this system, N-linked oligosaccharides on the enzyme surface have a mannose residue phosphorylated by the N-acetylglucosamine-1-phosphate transferase (phosphotransferase). This enzyme recognizes a lysine-based signal on the cathepsin L surface, which is composed by two lysine residues separated by a distance of ~34 Å (Warner et al., 2002). In a second step, an enzyme in the trans-Golgi network (TGN), the N-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase (uncovering enzyme) acts in the final uncovering of M6P removing a covering *N*-acetylglucosamine from the M6P recognition marker (Rohrer and Kornfeld, 2001). Finally, two M6P receptors (MPRs), MPR46 and MPR300, recognize this signal and transport their ligands from the TGN to the endosomal/lysosomal system (Ghosh et al., 2003). In contrast, primitive eukaryotes present a M6P-independent pathway to cathepsin L trafficking, which is dependent on a nine-amino acid loop motif in the enzyme prodomain region (Huete-Pérez et al., 1999). In insects, the mechanisms involved in the lysosomal sorting of these enzymes are yet unclear, as *Drosophila melanogaster* presents only a truncated homologue form of the MPR300 receptor, the lysosomal enzyme receptor protein (LERP), which does not bind the M6P, but partially rescues missorted soluble lysosomal enzymes in MPR-deficient mammalian cells (Dennes et al., 2005).

Recently, the 3D structures of the pro-cathepsins L TmCAL2 and TmCAL3, as the mutants pCAL2C25S and pCAL3C26S, were solved by X-ray crystallography in our group showing striking differences regarding the human lysosomal CAL (Beton et al., 2012).

Despite the results on insect CALs gathered till now, there are no data on substrate specificities for these enzymes. These kind of data, if obtained by using a large number of substrates with amino acid replacements in positions P2, P1, P1′, and P2', may be used to identify the role of active subsites in catalysis or substrate binding (Marana et al., 2002) and their hydrophobicity. The latter is an index that evaluates the hydrophobic character of the subsites and was used to compare different insect trypsins, disclosing one of the reasons why lepidopteran trypsins are resistant to a large number of plant protein inhibitors (Lopes et al., 2006). As the TmCAL3 3D structure did not include a substrate (or inhibitor), the identification of active site residues may be improved by appropriate docking modeling of TmCAL3 in complex with the same substrate used in the kinetic studies. An extensive study of TmCAL3 substrate specificity may permit to compare subsite hydrophobicities obtained by enzyme kinetics with those estimated from the 3D models obtained in the docking simulations.

More recently a large number of *T. molitor* midgut CALs were described, in addition to the cited TmCAL1, TmCAL2, and TmCAL3 (Martynov et al., 2015). The authors also proposed which were the key residues in S1 and S2 for all the CALs based on alignment with the human cathepsin L. This identification can be improved by a better identification of CAL subsites as described above. Finally, the determination by RNA-seq of the expression of *T. molitor* CALs in different midgut regions and carcass may support biological roles for those enzymes, guided by the previous findings concerning TmCAL1, TmCAL2, and TmCAL3 (Cristofoletti et al., 2005; Beton et al., 2012).

The present work was undertaken with three aims: (1) identify the role and hydrophobicity of S2, S1, S1′, and S2′ of TmCAL3 by using fluorescence energy transfer (FRET) peptides with 19 amino acid replacements in positions, P2, P1, P1′, and P2'; (2) propose the subsite residues of all *T. molitor* CALs from docking models built with the use of the same substrate for TmCAL3 and HsCAL1 (lysosomal human CAL, see Puzer et al., 2004) and approximations in the case of the other CALs; (3) compare subsite hydrophobicity obtained by enzyme kinetics with those obtained from TmCAL3 and HsCAL1 3D structure and docking procedures. The results showed important differences between the subsites of *T. molitor* digestive CALs and the lysosomal ones (including the human one), in agreement with the fact that there were several events of gene duplication with the lysosomal CALs being more ancestral. Subsite hydrophobicity calculated from the free energy of substrate binding (obtained by kinetics) agrees quite well with those calculated from the identified subsites in docking models. Finally, a mechanism of targeting TmCALs to midgut lumen was proposed.

2. Material and methods

2.1. Enzymes

Recombinant *Tenebrio molitor* pCAL3 was expressed in *Escherichia coli* and purified as previously described (Beton et al., 2012). The molar concentration was determined by active site titration with E−64 (Barrett et al., 1982). The fragment coding for the mutant pCAL3W177A was obtained by site-directed mutagenesis through overlap extension (Higuchi et al., 1988) using the plasmid pCAL3/pAE as a template and the primers: 5′-AACTGCAGCCTGCCAAAATCGCTCTTCC-3′, 5′-CCCAAGCTTTTACAAAGCAGGGTATGAAGCAG-3′, 5′-CCTGGGGCTCTGGAGCGGGAGAGAGCGGA-3′, 5′-TCCGCTCTCTCCCGCTCCAGAGCCCCAGG-3'. The mutant plasmid