Mini Review

Cathepsin D: A cellular roadmap

Nousheen Zaidi, Andreas Maurer, Sebastian Nieke, Hubert Kalbacher *

Medical and Natural Sciences Research Centre, University of Tubingen, Ob dem Himmerlreich 7, 72074 Tubingen, Germany
Interfacultary Institute of Biochemistry, University of Tubingen, Germany

ARTICLE INFO

ABSTRACT

Cathepsin D is a normal and major component of lysosomes, it is found in almost all cells and tissues of mammals. Present review describes different events in cellular life of cathepsin D mainly its biosynthesis, co-translational and posttranslational modifications, targeting to lysosomes and proteolytic processing and maturation within lysosomes.

Cathepsin D (CatD, EC 3.4.23.5) is an intracellular aspartic protease of the pepsin superfamily. The involvement of CatD in several physiological functions such as protein degradation, apoptosis and autophagy [1] have been reported. Additionally, it is also found to be associated with certain pathological conditions such as cancer [2], Alzheimer's disease [3] and neuronal ceroid lipofuscinosis [4]. CatD is found in almost all mammalian cells and has a typical lysosomal localization [5,6]. Human CatD is synthesized as a single polypeptide that is proteolytically processed into mature CatD that exists as a two-chain form: an amino terminal light chain (14 kDa) and a carboxyl-terminal heavy chain (34 kDa). These chains remain associated with each other by hydrophobic interactions. Before the generation of this mature two-chain form, CatD undergoes several proteolytic stages. Therefore, the early isolations of cathepsin D showed various different forms that differed in size, number of associated polypeptides and iso-electric point but had similar activities [7]. Later on, site-directed covalent modifications demonstrated that CatD has a similar active site as pepsin, confirming its identity as an aspartic protease [8,9]. It was hypothesized that the two-chain form is derived from the single polypeptide form by proteolysis [10]. This was determined by N-terminal sequencing of different forms of porcine CatD [11]. Isolation of the cDNA clone [12] and the gene [13], detailed studies of biosynthesis [14], activation [15], substrate specificity [16] and determination of crystal structure [17,18] makes cathepsin D one of the best studied lysosomal proteases (reviewed in [19]). In this review we have dis-

cussed the journey of cathepsin D through cell; encompassing its biosynthesis, co-translational and posttranslational modifications, targeting to lysosomal compartments and proteolytic processing within lysosomes.

Transcription of cathepsin D gene

Eukaryotic gene expression is controlled by both, proximal and distal elements, generally located in the 5′ upstream region of the gene [20]. Several gene promoters contain a TATA box, that binds to the transcription factor IID (TFIID) and thus, define the transcription initiation site for that gene. Genes having such promoters are termed as facultative or regulated genes. On the contrary, the promoter region of the house keeping genes lack a recognizable TATA box and contain multiple GC boxes that are putative binding sites for transcription factors, like Sp1 [21]. The promoter region of cathepsin D has a mixed structure showing features of both housekeeping genes (high G + C content and several potential Sp1 binding sites) and regulated genes (TATAA sequence) (Fig. 1) [22].

Due to this mixed structure CatD promoter can direct both types of transcription initiations; TATA-independent and TATA-dependent [22]. Moreover, it has also been shown that transcription of CatD is initiated at five major transcription starting sites (TSS-I to TSS-V) spanning 52 bp (Fig. 1) [22]. The constitutive expression of CatD in human is TATA-independent initiating at one of the several transcription starting sites (TSS-II to -V), upstream of the TATA box, and is possibly directed by GC boxes and Sp1 factor as in many housekeeping genes. On the other hand, in estrogen stimulated breast cancer cells where CatD is reported to be overexpressed [23], transcription is TATA-dependent, starting about 28 bp downstream from the TATA box (TSS-I) [22] (Fig. 1).
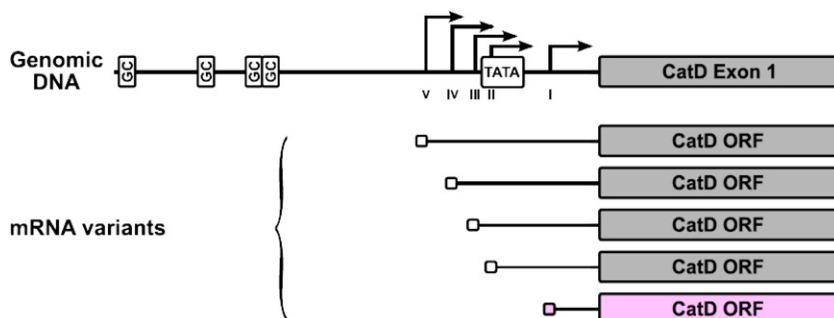
Fig. 1. Schematic diagram of the cathepsin D promoter region. Five major transcription start sites (TSS-I to -V) for CatD gene are indicated by arrowheads. These five transcription initiation sites respectively map at −20, −44, −51, −60 and −72 from the first base of the initiation codon. In basal conditions expression of human CatD is TATA-independent and initiates at one of the several transcription starting sites (TSS-II to -V), upstream of the TATA box, and is possibly directed by GC boxes and Sp1 factor. In estrogen stimulated breast cancer cells where CatD is reported to be overexpressed, transcription is TATA-dependent, starting from TSS-I. The lower part of the figure illustrates schematic comparison between the lengths of different CatD mRNAs. The difference in length is only between the 5′ untranslated regions of mRNA (*the open reading frame of CatD gene is mentioned*). The first four mRNAs with longer 5′ untranslated regions are present in higher proportion in basal conditions where TATA-independent transcription is dominant. While in breast cancer cells the proportion of mRNAs with shorter 5′ untranslated regions (*pink*) is increased. (For interpretation of colour mentioned in this figure, the reader is referred to the web version of this article.)

Thus, estrogens not only increase transcription but also affect the pattern of initiation, with TATA-dependent transcription predominating. Therefore, in stimulated conditions, CatD mRNAs preferentially have short 5′ untranslated sequences, whereas in basal conditions where TATA-independent transcription is dominant, the proportion of longer CatD mRNAs is increased (Fig. 1).

Since the CatD gene is controlled by a mixed promoter, it has the advantage of being both constitutively expressed from TATA-independent start sites, and overexpressed in some physiological conditions such as development or tissue modeling by TATA-dependent transcription [22] and this TATA-dependent transcription is also exploited by cancer cells.

Translation, posttranslational modifications and lysosomal sorting of cathepsin D

Cathepsin D is synthesized on the rough endoplasmic reticulum as a pre-pro-enzyme. This pre-pro-cathepsin D in human comprises 412 amino acids (aa) and contains an N-terminal secretion signal peptide (S) that is composed of 20 aa. This signal peptide is removed during co-translational translocation of CatD across the ER membrane generating an inactive procathepsin D (392 aa, 52 kDa) (Fig. 2). Following this co-translational removal of signal peptide sugars are attached at two N-linked glycosylation sites at asparagine residues 134 and 263 (pre-pro-CatD numbering, UniProtKB/*Swiss-Prot*) that are widely separated on the surface of the folded protein. After glycosylation cathepsin D is transported to the Golgi stacks. Studies with mono-glycosylated and non-glycosylated CatD mutants have shown that glycosylation is not necessary for folding or enzyme activity but it is required for targeting of the enzyme to lysosomes [24].

Normal cellular life of CatD is characterized by a special feature i.e. its strict localization within acidic lysosomal compartments. Transport of pro-CatD from Golgi complex to the downstream acidic compartments is mainly mediated by the mannose-6-phosphate (M-6-P) [25] pathway. Thus, in the Golgi complex the two N-linked oligosaccharides of CatD are covalently modified and their mannose residues are phosphorylated at position six (Fig. 2) [26]. These M-6-P groups are recognized by cation-independent mannose-6-phosphate receptor (MPR300) [27] in *trans*-Golgi network that segregates lysosomal hydrolases and helps to package them into budding transport vesicles, that deliver their contents to late endosomes and thereby to lysosomes (Fig. 2).

The second M-6-P independent pathway for targeting of pro-CatD to lysosomes is carried out by interaction of pro-CatD with pro-saponin [28]. This pathway is not fully understood and is also not the major pathway for intracellular trafficking of cathepsin D. However, it is suggested that the two molecules i.e pro-CatD and pro-saponin form complexes and travel together to acidic compartments [28]. According to a very recent report transport of CatD is speculated to be partially dependent on sortilin that is a multifunctional receptor that cycles between the *trans*-Golgi network and endosomal compartments [29].

Proteolytic processing and maturation of cathepsin D in lysosomal compartments

Once segregated into the endolysosomal compartments, the pro-CatD undergoes several proteolytic processing events to generate mature cathepsin D [19]. First of all at the low pH of late endosomes the pro-CatD dissociate from M-6-P receptors and subsequently the phosphate group is removed. The initial proteolytic step is the removal of pro-peptide (44 amino acids) from pro-CatD to generate an active intermediate (48 kDa, 348 a.a) single-chain molecule (Fig. 3). The pro-peptide (also termed as activation peptide) of CatD is reported to be essential for the correct folding, activation and delivery of the protein to lysosomes [30,31].

The 48 kDa single-chain intermediate is not generated by auto-catalytic activity of CatD, because proteolytically inactive mutant of CatD was normally processed to the mature form of the enzyme when transfected in CatD-deficient cell lines. Instead, the involvement of cysteine proteases other than CatL and CatB was observed [32].

The 48 kDa intermediate is further processed into mature two-chain form of cathepsin D, this processing step is carried out by cathepsin B or L [32]. The two-chain form consists of an amino terminal light chain (14 kDa) and a carboxyl-terminal heavy chain (34 kDa) (Fig. 3). These chains remain associated with each other only by hydrophobic interactions. Additionally, seven amino acid residues between the heavy and light chain are removed [33]. Several more amino acids are removed from the carboxyl terminus of the 34 kDa heavy chain [34].

In vitro studies have demonstrated that upon acidification pro-CatD auto-catalytically generates a proteolytically active pseudo-CatD [35,36], an enzyme form that retains 18 residues (27–44) of the propeptide. But recent studies have reported that the formation of pseudo-CatD is not a required intermediate for CatD processing in cells and it is a mere *in vitro* artifact [32].

The two-chain form of human CatD is generated by the proteolytic processing of CatD at a specified region [19] that forms a puta-