

# A Review on Molecular Functional and Structural Characterization of Human Cysteine Cathepsins in Bacterial Expression Systems

## **Authors-**

Ruwini Cooray<sup>1,\*</sup>, Hasanka Madubashetha<sup>1</sup>, Lakshan Warnakula<sup>1,2</sup>, P.D.S.U. Wickramasinghe<sup>3</sup>, Nimali De Silva<sup>4</sup>,

## **Institutional affiliations-**

- 1- Section of Genetics, Institute for Research and Development in Health and Social Care, Battaramulla, Sri Lanka.
- 2- National Science Foundation, Sri Lanka.
- 3- Department of Chemistry, Faculty of Science, University of Colombo, Sri Lanka.
- 4- Department of Nanotechnology, Faculty of Technology, Wayamba University of Sri Lanka.

**Corresponding:** krncooray@live.co.uk

## **Address for communication-**

*To the corresponding author*

Ruwini Cooray,

Research Scientist; Project Lead,

Section of Genetics,

Institute for Research and Development in Health and Social Care,

No 393/3, Lily Avenue, off Robert Gunawardena Mawatha, Battaramulla (Postal code- 10120), Sri Lanka.

Telephone- Office- +94112863084, Mobile- +94777328324

Email- krncooray@live.co.uk

## **Abstract**

Cysteine cathepsins, a class of proteinaceous enzymes, regulate a wide variety of metabolic processes in human including protein breakdown and turnover and immune functions. Eleven cysteine cathepsins have been identified so far and a wide array of studies related to identifying their specific functions, regulation and distribution patterns in tissues have been conducted. However, in recent past, the association of cysteine cathepsins in occurrence and progression of cancers have been identified and this has caused unrest in scientists triggering them to investigate the physiology, biochemical pathways and interactions of these cathepsins in cancer metastasis and therefore has become a noteworthy topic of intensive research. This review focusses and collects together the published work on molecular functional and structural characterization studies that have been done so far on *in vitro* expression of genes encoding for cysteine cathepsins in the *Escherichia coli* bacterial expression system. Accordingly, it was found out that all cathepsins except for cathepsins K, C, H, X and W have been expressed this way and the majority of them were found to be expressed in *E. coli* BL21(DE3) *pLysS* expression host via *pET3* expression vector. In addition, it was also noted that in most of the expression studies, the substrate that was used to validate the enzymatic activity of the recombinant enzyme that was produced was a cysteine residue along with a *benzyloxy-carbonyl* salt. Through this review, the authors suggest that there is a very high need that all cysteine cathepsins need to be characterized both structurally and functionally on a molecular platform to better understand their interactions including the biochemical pathways. It is also momentous that the mass production of the recombinant forms of these enzymes are facilitated via expression in such bacterial expression systems and in turn, would also provide a strong platform for the development and progression of studies related to human physiology including oncological studies such as cancer metastasis. Moreover, as per biochemical features of the enzymes that could be identified, the production of efficient inhibitors or inducers as per the necessity to improve health and promote wellbeing among the mankind could be facilitated.

**Keywords:** cathepsins; cancer; characterization; cysteine; *Escherichia coli*; expression; molecular

## **Background**

Proteases or proteolytic enzymes are a group of proteinaceous enzymes that are involved in the process of catalysing the hydrolysis process of amide bonds in biomolecules. However, unlike the reversible changes such as phosphorylation and allosteric site changes that proteins undergo, proteolysis in general is considered to be an irreversible change <sup>1</sup>. Cysteine proteases are one such group of proteolytic enzymes that involves in metabolism including protein degradation and turnover, bone remodelling and pre-hormone processing <sup>2</sup>. Apart from the structural analysis, a variety of functional analysis of these proteases have revealed, that in addition to their intracellular role in the process of recycling of proteins, they are also involved in other most usually occurring processes such as antigen presentation, bone remodelling, and prohormone activation <sup>3</sup>. Most remarkably, cysteine proteases are involved and are suggested to involve in a variety of disease processes such as pulmonary emphysema, osteoporosis, alzheimer's disease, rheumatoid arthritis, and also in cancer invasion and metastasis <sup>4</sup>.

These proteolytic enzymes can be sub divided into more than 20 classes of different families, of which the members; papain, Calpains, Streptopains, Clostripains, Viral Cysteine Proteases and Caspases are some prominent ones <sup>5</sup>. However, the largest noted family is the Papain family. This family as a whole consists of a variety of cysteine proteases that are derived from both eukaryotes and prokaryotes. As far as human cysteine proteases belonging to the papain family are concerned, it has been found to include a variety of Cathepsins <sup>6</sup>. Studies done on the structural analysis of cysteine cathepsins have proven that all of them contain a series of conserved features or rather domains which include an essential cysteine residue within each one of its active sites that determine its specific function. It has also been studied that cysteine proteases are synthesized as preproenzymes, which are processed to the corresponding proenzymes and oriented on to the lysosomes by the *mannose 6-phosphate* signal that is attached to them <sup>7</sup>. However, certain differences within these enzymes such as differences in distribution in their tissues and in certain other enzymatic properties such as exhibiting optimal activity in a slightly acidic pH and been mostly unstable at neutral pH that regulate systemic metabolism that collectively account for substrate specificities and pH stability are also noted <sup>4</sup>. In addition, it should also be noted that when cathepsins are found outside the lysosomes or found extracellularly they could rapidly irreversibly achieve an inactive status at neutral pH.

Cathepsin is derived from the Greek word “*kathapsein*” which means to digest, was proposed for the proteases that showed activity in a slightly acidic environment. Eleven human Cysteine Cathepsins have been found and studied which are Cathepsins B, C, F, H, K, L, O, S, V, X and W which was confirmed by a bioinformatic analysis of the sequence of the human genome (8). Lysosomal Cathepsins needs a reducing and a slightly acidic environment in order for them to show optimum function, such as found in the lysosomes, and hence become active <sup>9</sup>. Therefore, Cysteine Cathepsins were initially considered to be solely intracellular enzymes, responsible for non-specific and mass proteolysis in acidic environment of the endosomal or rather lysosomal compartment, where they degrade intracellular and extracellular proteins. However, this concern over these enzymes are rapidly changing and under intensive research (8).

In human, Cathepsins such as B, H, L, C, X, F, O and V are highly expressed in all most all human tissues. The most basic expression profiles of these enzymes show that these enzymes are involved in normal, cellular protein degradation and turnover <sup>10</sup>. In contrast, Cathepsins K, W and S show up only in limited cells or tissues, rather show a tissue specific distribution nature, thus indicating more tissue specific functions as well. For instance, Cathepsin K is highly seen to express in osteoclasts and highly specific in function, which is to involve in bone tissue resorption <sup>10</sup>. Similarly, Cathepsin W is predominantly expressed in CD8+ lymphocytes <sup>11</sup>. As far as the physiology of the cysteine cathepsins are concerned, they show optimum activity in slightly acidic pH and mostly unstable at neutral pH <sup>12</sup>. Except for Cathepsin S, when the other cathepsins are outside the lysosomes or the lysosomal compartments, or present extracellularly, they could relatively but yet rapidly undergo irreversible inactivation, given that the pH is neutral in nature <sup>13</sup>.

## **Main text**

This review focuses on the molecular functional and structural characterization studies that have been done with regard to the *in vitro* expression of genes encoding for human cysteine proteases in bacterial expression systems. More emphasis is given to the expression of human cysteine cathepsins which is considered to be a sub class that belongs to cysteine proteases, in the *Escherichia coli* (*E. coli*) bacterial expression system as the host. The review also collects together and summarises the published expression conditions of different cathepsins in *E. coli*.

An expression study of Cathepsin O which involved the molecular cloning of genetic material from a breast cancer carcinoma and production of the active enzyme in *E. coli*, and expression analysis in human tissues done by Velasco and colleagues reflect where the Cathepsin O cDNA cloned from a human breast cancer carcinoma was expressed in *E. coli*. After optimum purification and refolding, the recombinant protein was able to degrade the synthetic peptide *benzyloxycarbonyl-Phe-Arg-7-amido-4-methylcoumarin-d-b-enzyloxycarbonyl-Arg-Arg-7-amido-4-methylcouma*, which is widely used as a substrate for the entire group of cysteine proteases enzymes<sup>14</sup>. In the meantime, it was also observed that proteolytic activity of the recombinant Cathepsin O so produced was completely terminated by *trans-epoxysuccinyl-leucylamido-(4-guanidino) buta(nE''e)*<sup>14</sup> an inhibitor of this subclass of proteolytic enzymes, further validating that the recombinant protein produced had the conserved domains that feature Cysteine Cathepsins in human. Moreover, in the work published, a 930 bp DNA fragment containing the complete coding sequence for human Cathepsin O was obtained through PCR amplification of the isolated full-length cDNA with the primers 5'-ATGGACTCCCGCGCCCC-3' and 5'-CTGTCTCTTGATCTGCCCA-3'. Accordingly, a Polymerase Chain Reaction (PCR) was carried out for a total of 25 cycles with conditions which consisted of a 95°C, 1 minute denaturation, 50°C, 1 minute annealing and 72°C, 1 minute extension phase. The PCR product was then phosphorylated with *T4 polynucleotide kinase* and ligated to the expression vector *pET3c* which was previously treated with *NdeI* and *Nuclease S1*. The resulting plasmid which was called *pETH7*, was transformed into *E. coli* strain *BL21(DE3)*, and the transformed cells were grown in Luria Bertani (LB) media containing ampicillin with an incubation at 37°C for nearly 16 hours and was screened out accordingly<sup>2</sup>. Further studies that were done in the form of enzyme activity assays confirmed that this recombinant cysteine protease produced was Cathepsin O<sup>15</sup>. This study suggested that the mass production and availability of recombinant Cathepsin O would be of highly

useful to study the potential role in the pathological protein breakdown occurring in degradative and invasive processes associated with proteins. The study also suggested that, with the availability of recombinant Cathepsin O, it would be possible to investigate the correlation between alterations in expression of this enzyme and intracellular trafficking of Cathepsin O, would have with malignancy and the formation of cancers in human <sup>16</sup>.

Furthermore, published literature showed studies expressing the Cathepsin F in *E. coli*. A cDNA encoding for Cathepsin F was obtained from a human prostate cDNA library which encoded for a polypeptide of 484 amino acids. In the study, recombinant Cathepsin F that was produced by *E. coli* as a fusion protein with *glutathione S-transferase* degraded the synthetic peptide *benzyloxycarbonyl-Phe-Arg-7-amido-4-methylcoumarin* <sup>14</sup>, which is a substrate that is commonly used for functional characterization of cysteine proteases. As far as the molecular cloning of this experimental study is concerned, a 647 bp DNA fragment containing the coding sequence for the mature human Cathepsin F was amplified by PCR amplification of the isolated full-length cDNA with the use of primers 5'-ATGGCCCCACCTGAATGGGACT-3' and 5'-TCAGTCCACCACCGCCGAG-3'. The PCR reaction was carried out for 20 cycles which consisted of the phases; denaturation at 95 °C for 30 seconds, annealing at 60°C for 30 seconds and an extension phase at 68°C for 1 minute. The PCR amplicons were then phosphorylated with *T4 polynucleotide kinase* followed by repairing with the Klenow fragment, and was ligated to the expression vector *pGEX-3X* which was previously treated with the restriction endonuclease *SmaI* and *alkaline phosphatase*. The plasmid that resulted, which was called *pGEX-3X CTSE*, was transformed into *E. coli* strain *BL21(DE3)* and the transformed cells were grown in LB media containing ampicillin at an incubation at 37°C for 16 hours <sup>4</sup>. Further enzymatic studies, chromosomal mapping and northern blot analysis confirmed that the recombinant protein so produced, absolutely was Cathepsin F. As a result of the chromosomal mapping an evolutionary study of the enzyme was also facilitated. The evolutionary study was conducted using phylogenetic analysis which validated the fact that Cathepsin F is more closely related to Cathepsin W in human <sup>17</sup>. An important outcome of this study was that recombinant Cathepsin F could be useful for searching putative genetic diseases and different forms of occurrence of cancers associated with the gene coding for Cathepsin F in human. From previous literature it was found that the levels of expression of Cathepsin F is very high compared in the human skeletal muscle and the testis in

male<sup>18</sup>. Therefore, now with the availability of recombinant Cathepsin F, further studies could be conducted to find out the most effective protein substrate catabolism process that is been carried out by Cathepsin F in skeletal muscles<sup>19</sup> and any suggestive role of Cathepsin F in male fertility, avenues that are very rarely thought of compared to the typical reasons for the loss of male fertility. Expression analysis of Cathepsin F has also revealed the presence of this enzyme in several human cancer cell lines, significantly high in HeLa cells<sup>20</sup>. As a result, the recombinant Cathepsin F could be used in further clinical trials that could facilitate the establishment of effective therapeutics against such cancers as per the necessity.

In another study which involved in characterizing the enzyme Cathepsin Z, a full-length cDNA encoding for this enzyme was cloned from a human brain cDNA library<sup>21</sup>. Nucleotide sequence analysis revealed that the isolated cDNA codes for a polypeptide of 303 amino acids; Cathepsin Z and this cDNA was expressed in *E. coli* as a fusion protein with *glutathione S-transferase*, and after optimum purification, the so produced recombinant protein was capable of degrading the synthetic peptide *benzyloxycarbonyl-Phe-Arg-7-amido-4 methylcoumarin*, which is effectively utilized as a substrate by cysteine proteases (14). As far as the platform of work associated with molecular cloning in this study is concerned, firstly, an appropriate expression vector that is suitable for the production of recombinant cathepsin Z in *E. coli*, was produced by generating a 727 bp DNA fragment containing the coding sequence for the mature human cathepsin Z by PCR amplification of the isolated full-length cDNA with the use of the primers 5'-ATGCTGCCCAAGAGCTGGGAC-3' and 5'-CGATGGGGTCCCCAAATG-3'. The PCR reaction was carried out for 30 cycles which consisted of the phases; denaturation at 95°C for 30 seconds, annealing at 56°C for 30 seconds and extension at 72 °C for 1 minute. The PCR product was accordingly phosphorylated with T4 polynucleotide kinase, repaired with the Klenow fragment, and ligated to the expression vector *pGEX-3X* which was previously treated with the restriction endonuclease *SmaI* and alkaline phosphatase. The plasmid that resulted was called *pGEX-3X CathZ* and was transformed into the *E. coli* strain *BL21(DE3)*, and the transformed cells were grown in LB media containing ampicillin at an incubation in 37°C for about 16 hours. Enzyme activity assays and Northern blotting analysis confirmed that the recombinant enzyme produced this way was particularly Cathepsin Z<sup>21</sup>. An important outcome of this study was that, it was found out that on analysis of the propeptide of the recombinant protein that was produced, it did

not contain any lysine residue, which is contradictory to the fact that lysine based structures that were found present in the pro region of the other different Cathepsins were proposed to act as the recognition sites for the mannose phosphorylation that is required for intracellular targeting of cysteine cathepsins<sup>22</sup>. Based on this important fact, the pro region of Cathepsin Z is markedly different from the previously characterized members of this protein family and suggests that it could be even of a different sub family of enzymes. Having considered this important hypothesis, chromosomal location of the Cathepsin Z gene has revealed that it maps to chromosome 20<sup>23</sup>. This position differs from those that have been previously studied and reported for the remaining cysteine proteinase genes of the papain family. The study therefore, strongly emphasized that further studies are required in order to characterize the enzyme and ground it in the most appropriate class of the enzyme that it would belong to. As per the evidence derived from the work it was also noted that Cathepsin Z is expressed on all normal tissues in the body, and therefore suggests that a putative general role of this enzyme is noted in intracellular protein turnover to take place in all cell types of the body. With the availability of the recombinant Cathepsin Z enzyme, the potential effect that Cathepsin Z has on malignant cancers on human and other different protein degradative processes occurring in normal and pathological health conditions could also be studied in to further depth.

In the identification of a novel type of Cathepsin from a human brain cDNA library which coded for a polypeptide of 334 amino acids a high percentage of identity was seen to be established with Cathepsin L when pairwise comparisons with the other types of cysteine proteases was done. The percentage identity was noted to be 78%<sup>24</sup>. As a result, this Cathepsin was called Cathepsin L2. When the cDNA encoding the mature Cathepsin L2 was brought to express in *E. coli*, and after sufficient and optimum purification, the so produced recombinant protein was capable of degrading the synthetic peptide *benzyloxycarbonyl-L-phenylalanyl-L-arginine-7-amido-4-methylcoumarin*, which is a common substrate used by cysteine proteases<sup>25</sup>. In addition, the proteolytic activity on the same substrate mentioned above was abolished by the compound *trans-epoxysuccinyl-L leucylamido-(4-guanidino) butane* which is a common inhibitor for cysteine proteases<sup>25</sup>. This further proved that the cDNA that was isolated in the first place encoded for as cysteine proteinase. The molecular cloning of the cDNA library obtained involved the construction of an expression vector called the *pGEX-3X CathL2* which was generated by obtaining a DNA



fragment of length 669 base pairs encoding for mature human Cathepsin L2 and PCR amplification of the isolated full-length cDNA with the use of the two primers, 5'-ATGCTTCCCAAATCTGTGGATFGG-3' and 5'-TCACACATTGGGGTAGCTGG-3'. In accordance with the design of the 5' primer, the amplified fragment contained the coding information for a methionine residue at the NH<sub>2</sub> terminal end. The PCR reaction was carried out for 30 cycles which consisted of the phases; denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds and an extension at 72°C for 1 minute. The PCR product thus obtained was phosphorylated with *T4 polynucleotide kinase* followed by repairing with Klenow fragment, and was ligated accordingly to the expression vector *pOEX-3X* which was previously treated with *SmaI* and alkaline phosphatase. The resulting recombinant plasmid, was transformed into the *E. coli* bacterial strain *BL21(DE3)*, and the transformed cells were grown in LB media<sup>26</sup>. However, it should be noted that Cathepsin L2 is called Cathepsin V in the present day. The study was capable of bringing-forth certain important considerations to the oncological research community. That is, prefatory studies and analysis from previous literature studied, was capable of showing and validating the fact that Cathepsin L2 is also produced by certain renal and ovarian cancers in human<sup>26</sup>. Therefore, Cathepsin L2 could be characterized as an emerging cancer causing proteolytic enzyme, better defined, a Cysteine Cathepsin that has the potential to induce and cause severe cancer in human<sup>26</sup>.

Several attempts were taken by early scientists to produce human Procathepsin E in *E. coli*. This was done by the use of the restriction sites generated by the restriction endonuclease enzymes *NeoI* and *SaII*. These ends were engineered into a wide variety of expression vectors, under the control of the *tac* promoters and attempts were taken to produce the recombinant protein Procathepsin E in *E. coli* cells in its authentic or rather the native form or as a fusion with proteins of the same type such as *chloramphenicol acetyl transferase* and *maltose-binding protein*. However, the resulting yields were not satisfactory, as only low quantities of the recombinant protein were produced in the authentic expression mode due to the unstable nature of the plasmid and the protein product toxicity within the *E. coli* expression system itself<sup>27</sup>. The results obtained from expressing it in the fusion form such as protein structure, specificity and purity were either not satisfying as the resulting recombinant protein exhibited a completely insoluble nature after been down stream and did not retain all properties of the authentic protein. On such grounds, there was a

very high need for a high yielding expression system that has characteristics of been under a very tight and a highly regulating transcriptional control and accordingly the *pET22b*, an evidently secretory vector exhibiting such characteristics were used to investigate the feasibility of expressing human Cathepsin E. Accordingly, in a study, the human gastric adenocarcinoma AGS cell line which previously showed to express Procathepsin E was used as the total source of RNA. Accordingly, the complementary cDNA was established. Specific amplification of the Procathepsin E cDNA was done using PCR amplification protocols by the use of a sense and an antisense primer obtained from the 5'-untranslated and 3'-coding regions of the human Procathepsin E gene. To facilitate the process of molecular cloning of these genes into an expression vector, the two oligonucleotides which were 5'-GCGAGCATGCCATGGGATCCCTTCACAGGGTG-3' and 5'-CACCGGTCGACTTAGGGGACTGCTGGGGC-3' were designed in such a way, which included the *Neol* and *Sall* restriction endonuclease sites respectively, so that primer binding to the correct sites could be specifically and tentatively achieved. Extraction and purification of the 1,100 bp fragment that resulted and authentication of the oligonucleotide sequence was done using the dideoxy sequencing method. The *Neol-Sall* fragment on both sides of the coding sequence for Procathepsin E, was reported to be subcloned into the *T7* expression vector, *pET22b* and relevant expression studies were done by introducing the so created recombinant plasmid in the *E. coli* strain *BL21(DE3) pLysS*<sup>28</sup>. As mentioned above, the study involved many optimization procedures of the molecular techniques and expression systems. This itself seems to be a very important and a ground breaking outcome of the study, so that the recombinant Procathepsin E that could now be produced in large quantities, could be used in further studies related to pathological, pathophysiological, oncological and metabolism based on various clinical scenarios and as per the requirement since this study was able to resolve two major concerns which was to produce the recombinant Procathepsin E and also to purify it to optimum levels in such a way that the authentic properties of the protein still remained. As per published literature, Cathepsin E is also reported to be expressed in activated microglia, dendritic cells, macrophages and lymphocytes and also evident to be localized in distinct cellular compartments including the plasma membrane, endoplasmic reticulum and the golgi complex<sup>29</sup>. Studies also showed that Cathepsin E is also involved in the major histocompatibility complex class 2 mediated antigen presentation<sup>30,31</sup>. Therefore this suggests that the availability of recombinant Cathepsin E paves the pathway for better understanding of various aspects of physiology in human that could eventually lead towards

the development of novel therapeutic agents that could promote human health, by combating various abnormal disease conditions that could arise as a result of deficiencies, mutations or any other abnormal presentation and activity of Cathepsin E.

Human Procathepsin S was reported to be produced in the form of inclusion bodies in *E. coli* using an inducible T7-based expression system. DNA encoding Procathepsin S was amplified from a human ileum cDNA library using PCR amplification protocols. The primers used for this amplification process were 5'-GGAATTCATATGGCACAGTTGCATAAAG-3' and 5'-CTCGGATCCGATCCTCTAGATTTCTGG-3' which were designed based on the cDNA sequence obtained<sup>32</sup>. In order to facilitate the process of cloning the PCR amplicons, with the restriction endonuclease sites *EcoRI* and *NdeI* were incorporated into the sense primer while the *BamHI* restriction site was incorporated into the antisense primer. The PCR product was then cleaved by *EcoRI* and *BamHI* and was inserted into the general expression vector *pUC19* and the so constructed recombinant plasmid was introduced into *E. coli* *DH5G* host cells. Furthermore, in another trail of the same study, Procathepsin S coding insert was completely sequenced, cleaved from the plasmid by restriction endonucleases *NdeI* and *BamHI*, and was ligated accordingly into the *pET3a* expression vector which was previously digested by the aforementioned two restriction endonucleases. The resulting plasmid, which was called *pKG2.1*, was analyzed by restriction endonuclease enzymatic digestion. The bacterial expression host strains *E. coli* *BL21(DE3) pLysS* and *BL21(DE3)*, which contained the *T7 RNA polymerase* gene, which is known to be under the control of a *lac*-type promoter on the chromosome<sup>33</sup> were transformed with *pKG2.1* vector. Accordingly, *E. coli* *BL21(DE3) (pLysS) (pKG2.1)* and *BL21(DE3J) pKG2.1* cells and were grown aerobically as one to one cultures at an incubation of 37°C in Luria-Bertani medium containing the selective antibiotic ampicillin and inoculated with 10 mL overnight culture. In addition, production of recombinant Procathepsin S was induced by addition of IPTG to the media broth to achieve a final concentration of 0.4 mM<sup>34</sup>. The study was capable of bringing forward an important outcome, that it showed the activation of the partially activated enzyme Procathepsin S has a considerable effect to some extent by Cathepsin D. In general terms, it was able to validate on a scientific platform, that there could also be synergistic or antagonistic effects on cysteine cathepsins, shown by other cysteine cathepsins. The study also observed a similar phenomenon, between recombinant human Procathepsin S and recombinant human Procathepsin L, where the

activation of Procathepsin S was affected in the presence of the Procathepsin L enzyme. A recent study that was done in order to evaluate the association of Serum Cathepsin S with the Metabolic Syndrome in overweight and obese Chinese adults was capable of concluding the fact that elevated levels of serum Cathepsin S have a strong, yet an independent association with the metabolic syndrome in overweight and obese Chinese adults who were tested and the study suggested that however, further studies and investigations are required to completely establish the role and the pathophysiology related to the role of Cathepsin S in the development and the progression of Metabolic Syndrome<sup>35</sup>. Such studies emphasise the importance of the availability of methods to facilitate the recombinant production of Cathepsin S and hence with the availability of the recombinant enzyme the clinical background of such non communicable diseases could be better understood and the exact reasons for occurrence of such disease could also be understood more clearly so that the most optimum therapeutic could be brought upon for disease treatment or control that arise due to malfunctioning or hyper secretion of Cathepsin S.

The cysteine cathepsin, Cathepsin B was also expressed in *E. coli* by cloning the gene responsible for the production of the enzyme. During initial stages, according to published literature, several attempts have been taken to express human Cathepsin B in plasmid vectors like *pUC-18* and *Pkk233-2* but there is no published literature to show the output or the extent of success of such studies. However as per one study, molecular cloning and expression has been successfully achieved with the use of a *pET3* vector which contain the strong bacteriophage *T7 $\theta$ 10* promoter<sup>36</sup>. Accordingly, the 33kDa Cathepsin B protein had been expressed in sufficient quantities in this study. Human Cathepsin B cDNA containing *BamHI* linkers were cloned into the *BamHI* site of the *pET3* expression vector. The ligated plasmids were then first transformed into *E. coli HMS174* host cells, and those that contained Cathepsin B inserts were screened and transformed into other host strains for immuno-screening to verify that the protein produced was Cathepsin B. This was followed by IPTG induction and bacterial clones producing the immune reactive Cathepsin B were detected as dark spots on nitrocellulose filters as results of this study. Further characterization studies have been done on the bacterial gene products by subjecting them to protein blotting. Once they were made to undergo IPTG induction, one prominent protein band at approximately 33kDa was observed and moreover this band was absent in the lysate containing the opposite orientation plasmid and in addition, most importantly it was noted that the band obtained was also immune

reactive to the human Cathepsin B antiserum. Thus, it was concluded on a strong scientific ground that the 33kDa Cathepsin B was produced by the *pET3* expression vector system. In addition, the same study was also able to conclude that the expression of the enzyme Cathepsin B was stable in the host cells *E. coli BL21(DE3) pLysS* but not in the host cells *E. coli BL21(DE3)*. This was reported to be justified upon the fact that the host cells *E. coli BL21(DE3) pLysS* has an additional plasmid coding for the *T7* lysozyme gene that improves the expression features of the *T7* gene expression system<sup>36</sup>. However, it should be noted that under the molecular protocol and the extent of optimizations in this study, unlike the normal Cathepsin B which is glycosylated<sup>37</sup>, expressed this way, the recombinant, Cathepsin B was not glycosylated. The study also suggested that the bacterially derived Cathepsin B could be used as a substrate as well for further studies on Cathepsin B processing enzymes.

In addition, a study focused on the preparation of catalytically active human Cathepsin B from its precursor, in order to express it in an expression system of *E. coli* in the form of inclusion bodies, has also been able to show remarkable results<sup>38</sup>. Accordingly, a cDNA clone that encodes the particular human Procathepsin B was expressed in *E. coli* using a *T7* expression system which resulted in the formation of the recombinant protein in the form of insoluble cytoplasmic protein aggregates which are more or less known to be inclusion bodies<sup>38</sup>. In the study, a considerable amount of human ileum cDNA library was incubated at 70°C for duration of five minutes to break open the phages in order to facilitate the release of the DNA. This was followed by subjecting the DNA to PCR amplification using the two oligonucleotides 5'-GCGGGATCCAGAATTCATATGCGGAGCAGGCCCTCTTTC-3' and 5'-ATTCCCGGGATCCTCATAAATCGGTGCGTGGGATTCCAGC-3' as primers derived from the human Procathepsin B, that was available in literature during the time of the study. For the process of PCR amplification, a total of 30 cycles in the order 94°C for 60 seconds, 60°C for 60 seconds and 72°C for 120 seconds corresponding to denaturation, annealing and extension, respectively, were carried out. Accordingly, the DNA sequence coding for Procathepsin B was amplified. The c-terminal hexapeptide coding extension which was found in the cDNA sequence but absent in the mature enzyme was excluded from the region of amplification because it reflected to be not essential for protein folding and did not have a significant impact on the function of the mature enzyme. In order to facilitate cloning, the sense primer was designed in such a way that it

had the restriction sites *NdeI*, *BamHI*, and *EcoRI*. However, *SmaI* and *BamHI* sites were included in the antisense primer. The PCR product was then cleaved by the restriction endonucleases *EcoRI* and *BamHI* and was accordingly inserted into the *pUC19* expression vector. The host cells which were of the *E. coli DH5alpha* strain was used to clone the recombinant plasmid that was constructed. Later, the Procathepsin B coding region was cut out of the *pUC19* vector by the use of the restriction endonuclease enzymes *NdeI* and *BamHI*, and cloned into the *pET3a* expression vector that was previously cut by the same two restriction endonuclease enzymes that were mentioned earlier. In the process of cloning, the *E. coli BL21(DE3)pLysS* expression strain was transformed with the *pET3dprocathepsin B* recombinant construct. Luria Bertani media provided with the two antibiotics ampicillin at a higher concentration (100 µg/ml) and chloramphenicol at a lower concentration (25 µg/ml) was inoculated with 2 mL of overnight culture. The culture was incubated at 37°C until an absorbance of 0.6 at A<sub>600</sub> was detected. The expression of Procathepsin B was then induced by addition of IPTG to reach a final concentration of 0.4 mM<sup>39</sup>. As per this study, the recombinant production of Cathepsin B was further optimized and the protein analysis showed a higher degree of similarity to the authentic form that occurs in human. Most importantly as per the proposed methodology by this study, no auto-processing of the enzyme Procathepsin B could occur since the refolding is carried out at alkaline pH,<sup>40</sup> which does not favour its conversion into the mature form which itself is an effective methodology that have been proposed to be carried out in the case of mass production requirements of Cathepsin B. Furthermore, Cathepsin B has been found to involve in the progression of cancer in human in aspects of caspase dependent apoptosis and in the processes of facilitating tumour neovascularization and metastasis (Mijanovic, 2019). Moreover, Cathepsin B was also found to be involved in cancer metastasis by making alterations in the extracellular matrix remodeling and facilitating invasion in different tissues of the body<sup>42</sup>. Such considerations provoke the fact that research related to investigating the most specific inhibitor for Cathepsin B needs to be encouraged and such studies could be facilitated very easily when recombinant production of Cathepsin B is made available.

In another molecular expression study, a cDNA clone encoding human cysteine protease Cathepsin L was expressed at satisfying levels in an *E. coli* expression system<sup>43</sup>. However, the recombinant protein produced was insoluble and had to be solubilized using chemical methods which involved solubilizing the protein produced in urea and facilitating to be refolded at an alkaline pH. However,

in terms of stability, the recombinant enzyme showed very less stability compared to the original human enzyme or rather the authentic form of the enzyme. Structural and functional characterization studies of Cathepsin L were performed by means of inducing mutations specifically, in either the propeptide portion or the carboxyl terminal light chain of the protein and the genes with these mutations induced were expressed in the *E. coli* expression system and further investigations were done within the same study were done to assess a variety of enzymatic features such as solubility and purity that it exhibited. In the aspect of molecular characterization, recombinant human Cathepsin L was made to produce in *E. coli* using a gene expression system based on inducible *T7 RNA* polymerase synthesis of recombinant fusion proteins, which was followed up by the IPTG induction of a *T7 RNA* polymerase gene under *lacUV5* control inserted in the bacterial chromosome. The induced *T7 polymerase* transcribed the genes cloned downstream from a polymerase-specific T7 promoter in a particular expression vector. Even though the human Cathepsin L cDNA constructions containing the entire protein-coding sequence were not expressed at appreciable levels by this system, several constructions lacking the signal peptide-coding sequence were noted to express efficiently which itself was a noteworthy outcome of the study that was conducted <sup>43</sup>. The study suggests that this ability to produce large amounts of enzymatically active Cathepsin L in *E. coli* will make it possible to produce large amounts of enzyme for structural studies and other studies with pathophysiological implications <sup>44</sup>, while making it possible to perform detailed analysis on additional mutant forms and gross studies that could relate to the metabolism and other properties of this enzyme which is known to be an acid protease <sup>45</sup>. A recent study was also able to reveal that Cathepsin L has a regulatory effect on neuroretinal homeostasis by regulating a variety of inflammatory effects that are related with age related macular degradation in human <sup>46</sup>. Such homeostatic processes that function in the body could be better understood with the availability of recombinant Cathepsin L and accordingly, efficient inducers or inhibitors that are capable of determining the rate of such enzyme catalysed biochemical processes in the body could be controlled. In addition, a study which evaluated the effect of cysteine cathepsins on the development of the brain at various stages of life showed that the human brain utilized well defined and highly balanced patterns of cathepsin gene expression throughout the different stages of the development of the brain during the different stages of life <sup>47</sup>. This study also suggested that this fact of utilization of cysteine cathepsins by the could be considered the base for other investigations to understand how cysteine cathepsins are

dysregulated in lysosomal related brain abnormalities that affect the different age categories of human; from infants to adults. Accordingly, this establishes the fact that the availability of the recombinant enzymes of cysteine cathepsins could bring about a multitude of benefits in research that involves the study of metabolic processes in the human body, including their physiology in total, the related pathophysiology and other biochemical process in the body that regulate growth, development and wellbeing.

## **Conclusion**

Cysteine Cathepsins play a vital role in the overall metabolism of human that ensure survival of them. Therefore, it is of great importance to investigate whether the genes encoding for these enzymes could be expressed in microbial expression systems, especially bacterial expression systems, to identify the production of these enzymes within such systems in order to facilitate the production of therapeutics that contain the enzyme as an active component within. As per the published literature that is available in global databases such as the NCBI and PUBMED and Google scholar it is also understood that the expression of the genes responsible for the production of the enzymes Cathepsins K, C, H, X and W, were not seen in bacterial expression systems so far. However, expression of these enzymes in *E. coli* expression systems could give rise to the manufacture of therapeutics containing these enzymes that could be used to treat or control metabolic abnormalities or any related disease, especially a wide variety of cancers ranging from breast cancers, prostate cancers to ovarian cancers. Accordingly, it could also lead towards understanding their pathophysiology, the effect of human Cathepsins on the variety of carcinomas, their regulation, control and behaviour that could shower a multitude of benefits on a global outlook since cancer has been one of the major non communicable deadly diseases throughout the world today. It could also fast track oncological research to study the correlation that exist between cancers and cathepsin expression in tissues. Therefore, it is evident that the expression of cysteine cathepsins in bacterial expression systems could also open up a pool of other potential research avenues. The authors therefore suggest the other cysteine cathepsins too need to be characterized structurally and functionally on a molecular approach or rather a platform, so that the mass production of these enzymes could create impact in many fields including in the drug and pharmaceutical industry. The authors also suggest investigating better substrates that are uniquely



utilized by each cathepsin enzyme more effectively rather than using a general substrate or rather substrates that are commonly used by many cysteine proteases when required to validate what protein has been produced by the recombinant expression. For example, in most of the studies that were discussed above, a cysteine residue along with a *benzyloxy-carbonyl* salt such as was reported to have been used <sup>9</sup>. This methodology of confirmation does not allow drawing specific conclusions of substrate consumption by the enzyme, instead they would have a more generalized response. However, if utilization of the substrate could be more specific to identify the enzyme exactly with more specificity, then the results obtained would be more reliable. Finally, the authors also suggest that research related to the distribution of Cathepsins in human tissues needs to be better understood to create a good blend of research in physiology and molecular biology related to studies on Cathepsins so that studies creating an impact to contribute towards understanding metabolic abnormalities and related novel disease treatment methods could be investigated for the betterment of the mankind.

### **List of abbreviations**

01. *E. coli- Escherichia coli*
02. LB- Luria Bertani
03. PCR- Polymerase Chain Reaction
04. IPTG- Isopropyl  $\beta$ -D-1-ThioGalactopyranoside
05. DNA- Deoxyribo Nucleic Acid
06. cDNA- Complementary Deoxyribo Nucleic Acid
07. RNA- Ribo Nucleic Acid
08. AGS- Adenocarcinoma Gastric cell line

## **References**

1. Chapman, H. A., Riese, R. J. & Shi, G.-P. Emerging Roles for Cysteine Proteases in

- Human Biology. *Annu. Rev. Physiol.* **59**, 63–88 (1997).
2. Velasco, G., Ferrandos, A., Puentes, X. S., Shchezs, L. M. & Lpez-otinn, C. Human Cathepsin 0 turnover countries. **269**, 27136–27142 (1994).
  3. Li, Z. *et al.* Regulation of Collagenase Activities of Human Cathepsins by Glycosaminoglycans. *J. Biol. Chem.* **279**, 5470–5479 (2004).
  4. Santamaría, I., Velasco, G., Pendás, A. M., Paz, A. & López-Otín, C. Molecular cloning and structural and functional characterization of human cathepsin F, a new cysteine proteinase of the papain family with a long propeptide domain. *J. Biol. Chem.* **274**, 13800–13809 (1999).
  5. Turk, V. & Turk, B. Lysosomal cysteine proteases and their protein inhibitors: Recent developments. *Acta Chim. Slov.* **55**, 727–738 (2008).
  6. Nishi, Y. *et al.* Determination of bone markers in pycnodysostosis: Effects of cathepsin K deficiency on bone matrix degradation. *J. Bone Miner. Res.* **14**, 1902–1908 (1999).
  7. Levičar, N. *et al.* Lysosomal enzymes, cathepsins in brain tumour invasion. *J. Neurooncol.* **58**, 21–32 (2002).
  8. Turk, B., Turk, D. & Turk, V. Lysosomal cysteine proteases: more than scavengers. *Biochim. Biophys. Acta* **1477**, 98–111 (2000).
  9. Grzonka, Z. *et al.* Structural studies of cysteine proteases and their inhibitors. *Acta Biochim. Pol.* **48**, 1–20 (2001).
  10. Salminen-Mankonen, H., Morko, J. & Vuorio, E. Role of Cathepsin K in Normal Joints and in the Development of Arthritis. *Curr. Drug Targets* **8**, 315–323 (2007).
  11. Stoeckle, C. *et al.* Cathepsin W expressed exclusively in CD8+ T cells and NK cells, is secreted during target cell killing but is not essential for cytotoxicity in human CTLs. *Exp. Hematol.* **37**, 266–275 (2009).
  12. Verma, S., Dixit, R. & Pandey, K. C. Cysteine proteases: Modes of activation and future prospects as pharmacological targets. *Front. Pharmacol.* **7**, 1–12 (2016).
  13. Turk, V. *et al.* Biochimica et Biophysica Acta Cysteine cathepsins : From structure ,

- function and regulation to new frontiers ☆. **1824**, 68–88 (2012).
14. Krarneff, M. Effective activation of the proenzyme form of the urokinase-type plasminogen activator ( pro-uPA ) by the cysteine protease cathepsin L L o t h a r G o r e t z k P , M a n f r e d Schmitt " ~ Karlh ø inz Matan b , J u a n Calvete b , Nicolaus C h u c h o . **297**, 112–118 (1992).
  15. Devereux, J., Haeberli, P. & Smithies, O. Prms Limited , Nucleic Acids Research. **12**, 387–395 (1984).
  16. Menzel, K. *et al.* Cathepsins B, L and D in inflammatory bowel disease macrophages and potential therapeutic effects of cathepsin inhibition in vivo. *Clin. Exp. Immunol.* **146**, 169–180 (2006).
  17. Chuaqui, R. F., Zhuang, Z., Emmert-Buck, M. R., Liotta, L. A. & Merino, M. J. Analysis of loss of heterozygosity on chromosome 11q13 in atypical ductal hyperplasia and in situ carcinoma of the breast. **150**, 297–303 (1997).
  18. Tsujinaka, T. *et al.* Interleukin 6 receptor antibody inhibits muscle atrophy and modulates proteolytic systems in interleukin 6 transgenic mice. *J. Clin. Invest.* **97**, 244–249 (1996).
  19. Takeda, A. *et al.* Demonstration of cathepsins B, H and L in xenografts of normal and Duchenne-muscular-dystrophy muscles transplanted into nude mice. *Biochem. J.* **288**, 643–648 (2015).
  20. Newsham, I. F. The long and short of chromosome 11 in breast cancer. *Am. J. Pathol.* **153**, 5–9 (1998).
  21. Santamaría, I., Velasco, G., Pendás, A. M., Fueyo, A. & López-Otín, C. Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. *J. Biol. Chem.* **273**, 16816–16823 (1998).
  22. Nishikawa, A., Nanda, A., Gregory, W., Frenz, J. & Kornfeld, S. Identification of amino acids that modulate mannose phosphorylation of mouse DNase I, a secretory glycoprotein. *J. Biol. Chem.* **274**, 19309–19315 (1999).
  23. Wex, T. *et al.* Genomic structure, chromosomal localization, and expression of human

- cathepsin W. *Biochem. Biophys. Res. Commun.* **248**, 255–261 (1998).
24. Berti, P. J. & Storer, A. C. Alignment/phylogeny of the papain superfamily of cysteine proteases. *J. Mol. Biol.* **246**, 273–283 (1995).
  25. Duffy, C., Teylert, T. J. & Shashoua, V. E. from the CAI zone should result in a 3H/. **212**, 151–154 (1981).
  26. Santamaría, I. *et al.* Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas. *Cancer Res.* **58**, 1624–1630 (1998).
  27. Athauda, S. B. P., Takahashi, T., Kageyama, T. & Takahashi, K. Autocatalytic processing of procathepsin E to cathepsin E and their structural differences. *Biochem. Biophys. Res. Commun.* **175**, 152–158 (1991).
  28. Hill, J., Montgomery, D. S. & Kay, J. Human cathepsin E produced in *E. coli*. *FEBS Lett.* **326**, 101–104 (1993).
  29. Saku, T., Sakai, H., Shibata, Y., Kato, Y. & Yamamoto, K. An immunocytochemical study on distinct intracellular localization of cathepsin E and cathepsin D in human gastric cells and various rat cells. *J. Biochem.* **110**, 956–964 (1991).
  30. Bennett, K. *et al.* Antigen processing for presentation complex E \* by class requires major histocompatibility cleavage by cathepsin. *Cell Prolif.* 1519–1524 (1992).
  31. Harada, Y. *et al.* Cathepsin E in neutrophils contributes to the generation of neuropathic pain in experimental autoimmune encephalomyelitis. *Pain* **00**, 1 (2019).
  32. Kirschke, H., Schmidt, I. & Wiederanders, B. Cathepsin S The cysteine proteinase from bovine lymphoid tissue is distinct from cathepsin L (EC 3.4.22.15). *Biochem. J.* **240**, 455–459 (2015).
  33. McNicholas, P. M., Chiang, R. C. & Gunsalus, R. P. The *Escherichia coli* modE gene: Effect of modE mutations on molybdate dependent modA expression. *FEMS Microbiol. Lett.* **145**, 117–123 (1996).
  34. Kopitar, G., Dolinar, M., Štrukelj, B., Pungerčar, J. & Turk, V. Folding and activation of human procathepsin S from inclusion bodies produced in *Escherichia coli*. *Eur. J.*

- Biochem.* **236**, 558–562 (1996).
35. Chen, L. *et al.* Elevated circulating cathepsin S levels are associated with metabolic syndrome in overweight and obese individuals. *Diabetes. Metab. Res. Rev.* **35**, (2019).
  36. Chan, M. M. Y. & Fong, D. Expression of human cathepsin B protein in *Escherichia coli*. *FEBS Lett.* **239**, 219–222 (1988).
  37. Rich, D. H., Brown, M. A. & Barrett, A. J. Purification of cathepsin B by a new form of affinity chromatography. *Biochem. J.* **235**, 731–734 (2015).
  38. Graham, J. B. *et al.* The Malmö polymorphism of coagulation factor IX, an immunologic polymorphism due to dimorphism of residue 148 that is in linkage disequilibrium with two other FIX polymorphisms. *Am. J. Hum. Genet.* **42**, 573–580 (1988).
  39. Kuhelj, R. The preparation of catalytically active human cathepsin B from its precursor expressed in. *Eur. J. Biochem.* **229**, 533–539 (1995).
  40. Mach, L., Mort, J. S. & Glossl, J. Maturation of human procathepsin B. *J. Biol. Chem.* **269**, 13030–13035 (1994).
  41. Mijanović, O. *et al.* Cathepsin B: A sellsword of cancer progression. *Cancer Lett.* **449**, 207–214 (2019).
  42. Wu, J. S. *et al.* Cathepsin B defines leader cells during the collective invasion of salivary adenoid cystic carcinoma. *Int. J. Oncol.* **54**, 1233–1244 (2019).
  43. Smith, S. M. & Gottesman, M. M. Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *J. Biol. Chem.* **264**, 20487–20495 (1989).
  44. Smith, S. M., Kane, S. E., Gal, S., Mason, R. W. & Gottesman, M. M. Glycosylation of procathepsin L does not account for species molecular-mass differences and is not required for proteolytic activity. *Biochem. J.* **262**, 931–938 (2015).
  45. Gal, S. & Gottesmant, M. M. of a cDNA for human pro- ( cathepsin L ). **253**, 303–306 (1988).
  46. Sharif, U. *et al.* Advanced glycation end products-related modulation of cathepsin L and NF-κB signalling effectors in retinal pigment epithelium lead to augmented response to

TNF $\alpha$ . *J. Cell. Mol. Med.* **23**, 405–416 (2019).

47. Hsu, A., Podvin, S. & Hook, V. Lysosomal Cathepsin Protease Gene Expression Profiles in the Human Brain During Normal Development. *J. Mol. Neurosci.* **65**, 420–431 (2018).

### **Declarations**

- **Ethics approval and consent to participate**

Not applicable.

- **Consent for publication**

Not applicable.

- **Availability of data and material**

Not applicable since this is a review article.

- **Competing interests**

The authors declare that they have no competing interests.

- **Funding**

Not applicable.

- **Authors' contributions**

Ruwini Cooray (RC), corresponding author, formulated the concept for writing, involved in proof reading the final manuscript and making all final amendments before submission. Hasanka Madubashetha (HM), involved in searching through global databases for published research work, and filtering the necessary information required for manuscript writing and was involved in writing the manuscript. Sachith Wickramasinghe (SW) was involved in searching through global databases for published research work for the preparation of the review and progressing the writing. Nimali De Silva (NS) was involved in providing the guidance and advice for the concept and scientific writing. Lakshan Warnakula (LW) was involved in providing guidance in the writing process. RC, HM, SW, NS and LW read and approved the final manuscript.

- **Acknowledgement**

All authors express their sincere thanks to the Institute for Research and Development in Health and Social Care at Battaramulla, Sri Lanka for the help in various ways throughout the preparation of the manuscript.

---