EFFECTS OF CATHEPSIN PROTEOLYTIC NETWORK DYNAMICS ON EXTRACELLULAR MATRIX DEGRADATION IN BIOLOGICAL MACHINES AND INVASIVE DISEASE

A Dissertation
Presented to
The Academic Faculty

by

Meghan Cecile Ferrall-Fairbanks

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Biomedical Engineering in the
Wallace H. Coulter Department of Biomedical Engineering

Georgia Institute of Technology and Emory University
December 2017

EFFECTS OF CATHEPSIN PROTEOLYTIC NETWORK DYNAMICS ON EXTRACELLULAR MATRIX DEGRADATION IN BIOLOGICAL MACHINES AND INVASIVE DISEASE

Approved by:

Dr. Manu O. Platt, Advisor
Department of Biomedical Engineering
*Georgia Institute of Technology
and Emory University*

Dr. Melissa L. Kemp
Department of Biomedical Engineering
*Georgia Institute of Technology
and Emory University*

Dr. Eberhard O. Voit
Department of Biomedical Engineering
*Georgia Institute of Technology
and Emory University*

Dr. Rashid Bashir
Department of Bioengineering
*University of Illinois at
Urbana-Champaign*

Dr. Ellen Moomaw
Department of Chemistry and
Biochemistry
Kennesaw State University

Date Approved: July 27, 2017

To my family and friends

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ACA   aminocaproic acid

CatK   cathepsin K

CatL   cathepsin L

CatS   cathepsin S

CatV   cathepsin V

CysB   cystatin B

CysC   cystatin C

DNA   deoxyribonucleic acid

DTT   1,4-dithiothreitol

EDTA   ethylenediaminetetraacetic acid

E-64   N-[N-(L-3-trans-carboxyirane-2-carbonyl)-L-leucyl]-agmatine; [1-[N-[(L-3-trans-carboxyoxirane-2-carbonyl)-L-leucyl]amino]-4-guanidinobutane]

Fbgn   fibrinogen

FpA   fibrinopeptide A

FpB   fibrinopeptide B

HEK293T   human embryonic kidney 293T cell line

mCatL   mouse cathepsin L

MCK   muscle creatine kinase

MMP-2   matrix metalloproteinase-2

MMP-9   matrix metalloproteinase-9

PCR   polymerase chain reaction

RMSD   root-mean-square deviation

SUMMARY

Cysteine cathepsins are a family of lysosomal proteases that include the most potent mammalian collagen and elastase, meaning they degrade collagen and elastin better than any other mammalian enzyme. Normally, these proteases are involved in protein turnover and degrade proteins to maintain cellular homeostasis. In addition to their lysosomal proteolytic activity, cathepsins can be secreted from cells and degrade extracellular matrix (ECM) with differing affinities in a variety of tissue destructive diseases such as cancer metastasis, atherosclerosis, and osteoporosis.

Previous work showed cathepsin S could preferentially degrade cathepsin K in the presence of traditional substrate, a phenomenon termed cathepsin cannibalism, reducing expected substrate degradation. Moreover, once cells secret these proteins, they are subject to reaction diffusion kinetics, cell surface kinetics, and some other unidentified interactions that all must be considered beyond traditional ECM degradation kinetics. These potent enzymes are known to be upregulated in tissue destructive diseases, but researchers have been limited in their ability to successfully target cathepsin dysfunction. It is important not only to understand how these enzymes remodel the ECM, but also how they interact with each other, to effectively dose inhibitors to regulate cathepsin dysfunction.

This research develops a mechanistic model of how cathepsins interact with ECM and each other for tissue remodeling as produced and regulated by living cells, to target these cathepsin mechanisms in the treatment of tissue destructive diseases and remodeling of synthetic environments. The central hypothesis of this dissertation is that when cells secrete multiple species of cathepsins, these cathepsins exhibit complex cathepsin

cannibalism interactions, which reduces expected ECM degradation and thus concentrations of cathepsins. This dissertation is innovative because it analyzes cathepsin activity through molecular biology and computational simulation approaches to develop mechanistic models to drive targeting strategies for cathepsin-mediated diseases and sustainable fabrication of biological machines.

In summary, using experimental and computational techniques, we confirmed that these cathepsin cannibalism interactions are occurring between pairs of cathepsins, ultimately reducing cathepsin concentrations catalytically active toward substrate degradation. Furthermore, this research introduces bioinformatics tools to design mutations to interrupt these cannibalism interactions to increase substrate degradation. This brings researchers closer to reliably predicting responses to perturbations of the cathepsin proteolytic network, which is a key requirement for targeting cathepsin proteolysis in treatments for tissue destructive disease.

Understanding the cathepsin proteolytic network is also necessary for design and control of engineered living systems, such as locomoting biological machines. By characterizing the cathepsin and MMP proteolysis in the muscle strips of these engineered livings systems found that they are making increasing amounts of protease over time, which is degrading the fibrin matrix and destabilizing the biological machines. Specifically, inhibiting the cathepsin activity can reduce this cathepsin-mediated destabilization and prolong biological machine survival time. The engineered living system's functional time can be controlled by modifying the machine's proteolytic network, offering design strategies to control time until failure of these machines.

CHAPTER 1.   INTRODUCTION

When proteins are aged, defective, or taken up by the cell, they are often degraded by enzymes in the lysosomes, such as cysteine cathepsins [1, 2]. These proteases play a major role in cellular homeostasis. In addition to their proteolytic activity inside lysosomes, cathepsins can be secreted from cells and degrade extracellular matrix (ECM). Over secretion of cathepsins plays a role in a variety of tissue destructive diseases such as cancer, atherosclerosis, and osteoporosis. Cathepsins are a family of understudied proteases that include the most potent human collagenase and elastase [3-6]. Previous work shows cathepsin S preferentially degrades cathepsin K, reducing the degradation of collagen, in a process termed cathepsin cannibalism [7]. Moreover, once cells secret these proteins, they are subject to reaction diffusion kinetics, cell surface kinetics, and some other unidentified interactions, that all must be considered beyond traditional extracellular matrix degradation kinetics. Cathepsins are also secreted with a propeptide covering the enzyme's active site, which is cleaved to activate the enzyme [8-11]. Cells produce cystatins, endogenous cathepsin inhibitors, to control proteolytic activity by these enzymes [3, 12].

These potent enzymes are known to be upregulated in tissue destructive diseases, but researchers have been limited in their ability to successfully target cathepsin dysfunction in these diseases. It is important that we not only understand how these enzymes remodel the extracellular matrix, but also how these proteases interact with each other, to effectively dose pharmaceutical inhibitors to regulate cathepsin dysfunction therapeutically.

1.1 Research Objectives and Specific Aims

The main objective of this research was to develop a mechanistic model describing of how cathepsins interact with ECM and each other for tissue remodeling as produced and regulated by living cells, to target these cathepsin mechanisms in the treatment of tissue destructive diseases and remodeling of synthetic environments. Therefore, **the central hypothesis** of this dissertation was that when cells secrete multiple species of cathepsins, these cathepsins exhibit complex interactions between cathepsins K, L, S, and V, which reduce expected concentrations of cathepsins and thus the overall ECM protein degradation. The central hypothesis was tested through the following specific aims:

Aim 1: Distinguish proteolytic interactions between cathepsins K, L, S, and V among key natural substrates to develop a proteolytic network computational model. The hypothesis of this aim was that cathepsin cannibalism occurs between cathepsins K, L, S, and V, reducing target substrate degradation. To expand this idea, cannibalistic relationships between cathepsins K, L, S, and V on collagen and elastin, two major components of ECM, were explored. Time course kinetic assays of individual, pairs, triplets, and quadruplet were used to determine ECM substrate degradation over time and modeled using mass-action kinetics. This work was highlighted in Chapter 7. Furthermore, mutant cathepsins were designed and tested to verify the cathepsin cannibalism directionality as determined *in silico* in the cathepsin proteolytic network model, which was presented in Chapters 3 and 8. The goal of this model was to predict ECM degradation

due to multiple cathepsins being active extracellularly in the presence of one or more substrates.

Aim 2: Determine the role of proteolytic activity in destabilizing locomoting biological machines. The hypothesis of this aim was that proteases, specifically cysteine cathepsins, secreted by C2C12 myoblast cells in biological machines degrade collagen and other matrix proteins used in machine assembly/construction, thereby reducing stability of the machine, and ultimately shortening its lifespan. Using a locomoting biological machine constructed from a stereolithography device seeded with C2C12 myoblast cells, the machine's proteolytic network was quantified by Western blots and zymography both before and after the biological machine was electrically stimulated to trigger contractile motion. The protease activity of these biological machines and capacity to degrade the fibrin(ogen) substrates of these machines was explored in Chapters 4-6. Using computational analysis and proteolytic profiling of the biological machine, strategies using cathepsin protease inhibitors or mutant cathepsins were developed to prolong the machine's functional lifespan, which was discussed in Chapter 6.

1.2   Significance of Results

This dissertation is *innovative* because it analyzes cathepsin activity through molecular and computational approaches to develop mechanistic models to drive effective targeting strategies for cathepsin-mediated diseases and sustainable fabrication of biological machines. Cathepsin species have been found to interact with each other in a

non-intuitive way, which provides proof-of-concept that proteases bind and degrade other proteases in addition to substrate proteins, which is important for both treating tissue destructive diseases as well as developing the next generation of biological machines. The *long-term goal* of this project is to reliably predict perturbations to the proteolytic network and use this information to exploit protease-on-protease interactions in therapies for tissue destructive diseases and synthetic environments.
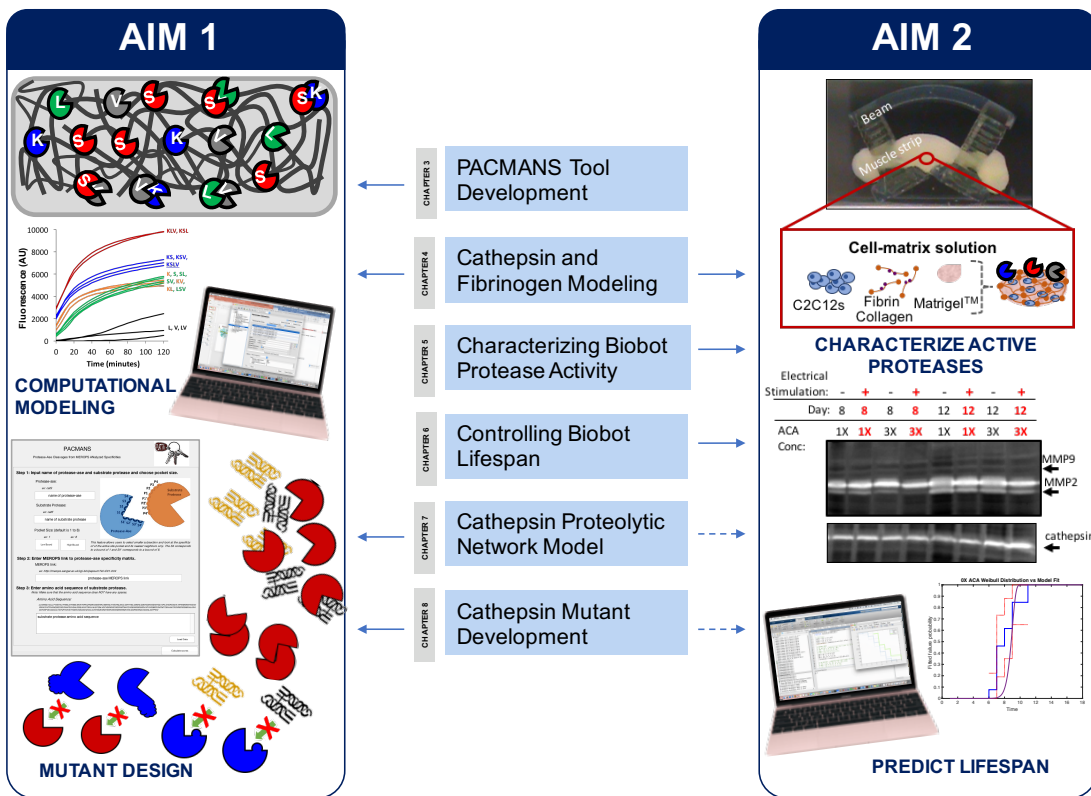
Figure 1-1. Overview of Dissertation.

CHAPTER 2.   BACKGROUND

Proteases are enzymes that degrade proteins and play a major role in cellular homeostasis through control of physiological processes including immune response, cell cycle and death, wound healing, food digestion, and protein and organelle recycling [3, 13]. Proteases composed about 2% of the human genome, accounting for almost 600 proteins, which are divided into five major classes of mammalian proteases distinguished by their catalytic mechanism: serine, cysteine, metallo, aspartic, and threonine [13, 14]. These proteases' specificities can range from highly selective, such as thrombin cleavage of fibrinogen in blood clotting, to highly non-specific, such as those involved in protein turnover [13]. Proteases "signal" through these proteolytic cleavage events which ultimately change the function of the substrate and in many pathologies proteases are dysregulated, which can result in non-specific proteases interacting with new substrates that they do not normal interact [13]. Furthermore, with the advent of the engineered living systems field, with new biomaterials and tissue engineering constructs that combine multiple cell types, it is important to consider the role of proteolysis for applications in these bioengineering constructs.

2.1   Cysteine Cathepsins

The dynamic state of protein synthesis and degradation of the structural proteins has been discovered in the last century [1, 15]. In this time, researchers have concluded that there are two major systems of intracellular protein degradation: the ubiquitin-

proteasome system and the lysosomal system [1]. In the ubiquitin-proteasome system works intracellularly where substrates to be degraded are tagged by covalent attachment of ubiquitin molecules and the subsequence tagged protein is degraded by the proteasome [15-17]. In the lysosomal degradation system, extracellular proteins are endocytosed and intracellular materials are taken up by autophagy so that hydrolases in the lysosomes can degrade these proteins [1, 18]. These so called "suicide bags" carry out this protein degradation at an acidic pH and generally the hydrolases have a very broad specificity to facilitate their role as recycling enzymes [13, 18, 19].

Cysteine cathepsins are a group of these lysosomal proteases that belong to the clan CA of cysteine peptidases, more specifically they are members of the C1 family of papain-like peptidases [1, 20]. There are eleven cysteine cathepsins in this family of potent enzymes including cathepsins B, C, F, H, K, L, O, S, V, W, and X [1] and contain the powerful human collagenase, cathepsin K, as well as the most powerful human elastase, cathepsin V [3, 5, 6, 21-24]. Since cysteine cathepsins are a part of the papain superfamily, their structure is highly conserved and like that of papain, as indicated by amino acid sequences and crystal structures [1, 25-32]. Cathepsins have two structural domains roughly equal size separated by an active site cleft [1, 20, 21]. One domain contains three α-helices with the longest element containing 40 residues, including the catalytic Cys25 at the N-terminus [20, 21]. The other domain has a β-barrel of five to six strands makes up the C-terminus [20, 21]. The two-domain interface forms the active site, which is open at the top and has reactive residues, Cys25 and His159, at the center of this cleft [1].

Substrates bind in an extended confirmation with its side chains alternating contact between the two structural domains [1, 33]. The active site cysteine and histidine form a thiolate-imidazolium ion pair, where the SH group of the Cys25 acts at the nucleophile and the adjacent imidazole from the His159 removes the proton from the SH group [21, 34]. The third member of the catalytic triad is Asn175, which acts as a polarizable acceptor for positive charge, stabilizing the thiolate-imidazolium pair for transfer of the imidazolium proton to the cleaved substrate [34]. Of interest are cathepsins K, L, S, and V which share 60% sequence identity and redundancy in target substrate proteins with different catalytic activities toward similar matrix substrates [3, 5, 35].

2.1.1   Cathepsin K

Cathepsin K is the most potent mammalian collagenase and is the only mammalian protease known to cleave in the triple helix of type I collagen at multiple sites [6, 20, 23]. It was first detected in rabbit osteoclasts and, since then, have also been found to be expressed in ovary, small intestine, colon tissue, endothelial cells, fibroblasts, macrophages, and tumor cells [20, 21, 29, 36-41]. Cathepsin K is known to be a critical enzyme in bone resorption as an enzyme secreted into an acidic milieu by osteoclasts. When individuals have an autosomal recessive genetic mutation resulting in production of an inactive form of cathepsin K, a disorder called pycnodysostosis, these individuals have unnatural bone turnover and are characterized by having brittle, dense long bones, premature closure of long bone growth, and facial hypoplasia [3, 42, 43]. Pathologically, in osteoporosis, cathepsin K overexpression, results in excessive bone breakdown [21] and

many pharmaceutical companies have tried to develop cathepsin K specific inhibitors to treat osteoporosis [44-47]. Cathepsin K is also a potent elastase and has been reported to contribute to rapid elastin degradation leading to plaque formation in cardiovascular disease [3, 48-51]. Cathepsin K's proteolytic activity has also been implicated in other tissue destructive diseases including cancer metastasis, rheumatoid arthritis, and endometriosis [36, 38, 52-56].

2.1.2 Cathepsin L

Cathepsin L is one of the ubiquitously expressed cysteine cathepsins (along with cathepsins B, C, and H) [18]. Unlike cathepsins K, S, and V, there are two proteolytically active forms of cathepsin L, a single-chain form of 28kDa and a heavy chain form of 24kDa, where a 4kDa light chain has been cleaved [20]. Cathepsin L was first identified in rat liver lysosomes and such facilitates a variety of activities both intracellularly and extracellularly [20]. In endosomes and lysosomes cathepsin L is involved with invariant chain processing and specific antigenic peptides for MHC class II molecules presentation to T-cells [20, 57, 58], recycling growth factor and growth factor receptors to the plasma membrane [20, 59, 60], and degrading the capsid and underlying proteins of viruses to initiate infection [20, 61-67]. In the extracellular space, cathepsin L has been implicated in pathologies including atherosclerosis and tumor progression, as well as neovascularization by matrix degradation and integration of circulating progenitor cells into ischemic tissue [20, 36, 38, 50, 68-71]. In the nucleus, cathepsin L has been reported to play a role in regulation of the cell cycle progression and degrade the CDP/Cux transcription factor,

accelerating cell cycle progression [20, 72-74]. Some of cathepsin L's nuclear activities have resulted in enhanced proliferation and invasiveness in tumor cells [20, 75]. Finally, cathepsin L in the cytoplasm has been reported to be both a pro-apoptotic and anti-apoptotic enzyme through cleavage of enzymes in the caspase pathway [18, 20, 76, 77].

2.1.3 Cathepsin S

Cathepsin S is unique in that it is the only cysteine cathepsin that remains stable and active at a neutral pH [3, 20, 21, 74]. Cathepsin S shares 55% sequence identity with cathepsin K and has both collagenase and elastase activities [20]. It was first identified in bovine lymph nodes and later in the spleen, but cathepsin S is also highly expressed by macrophages and in thyroid tissue and is involved with immune response [3, 20, 21]. Because of its high expression in lymph nodes and spleen, cathepsin S has been found to be involved in immune function including major histocompatibility complex (MHC) class II antigen presentation [1, 3, 20, 78]. Cathepsin S's proteolytic activity has been implicated in pathologies including emphysema, Alzheimer's disease, atherosclerosis, tumor progression and other inflammatory diseases as well as obesity [3, 20, 21, 50, 71, 79-83]. Interestingly, cathepsin S also plays a role in itch and immune response [84-88].

2.1.4 Cathepsin V

The most recently discovered member of the cysteine cathepsin family is cathepsin V and it is the most potent mammalian elastase and is about twice as efficient in cleaving insoluble elastin as pancreatic elastase [5, 20, 89]. Cathepsin V, sometimes called

9

cathepsin L2, and cathepsin L share 80% sequence identity, however mouse cathepsin L is orthologue of human cathepsin V and there is no mouse equivalent orthologue to human cathepsin L and hair follicle defect and impaired CD4+ helper cell selection in cathepsin L-deficient mice can be restored with cathepsin V-transgenic mice [20, 57, 90, 91]. Cathepsin V is selectively expressed in the thymus, testis, and cornea [5, 20, 57, 89, 92, 93]. Like cathepsins L and S, cathepsin V is involved with processing of invariant chains [20, 24, 93]. Furthermore, the elastinolytic activities of cathepsin V are increased in stenotic valves and atherosclerotic plaque development [20, 41, 94]. Finally, like the other three cathepsins, cathepsin V has also been implicated in cancer metastasis and is considered potential diagnostic markers for colon tumors [20, 38, 54, 95].

2.1.5   Cysteine cathepsin mechanisms of regulation

Cysteine cathepsins are synthesized as inactive zymogens (procathepsin) with a pro-peptide occluding the active site, which must be cleaved to have mature, active enzyme [3, 21]. This is one mechanism for how cells prevent erroneous cathepsins catalytically activity. Activation of cysteine cathepsins can occur autocatalytically, where the cathepsin is in a favorable (acidic) pH where it can cleave its own propeptide [9, 10, 96]. Alternatively, other proteases can also cleave the propeptide through intermolecular activation [10, 43, 97]. Since most cathepsins also require an acidic and reducing environment to be catalytically active towards substrates, it is hypothesized that this acts as a safety mechanisms for cells in the case of lysosomal breakdown or leakage [9, 10, 18,

96]. Moreover, the release of the propeptide has also been reported to act as a regulator of cathepsin activity and can be potent inhibitors of mature protease [10, 98].

An alternative mechanism for cysteine cathepsin regulation is the concept of cathepsin cannibalism, one species of cathepsin can preferentially degrade another, even in the presence of substrate matrix protein [7]. In the seminal study by Barry and Platt, 2012, cathepsin S was found to preferentially degrade cathepsin K, when in the presence of collagen and elastin. Furthermore, this cannibalistic relationship can protect against substrate degradation, which is important because it could be utilized as a method to preserve matrix proteins in disease states or synthetic environments. Figure 2-1A shows elastin degradation of cathepsin S (white squares), cathepsin K (white triangles), and both together (red circles) on fluorogenic elastin, which fluoresces when cleaved. Initially the predicted amount of degraded elastin by the combination of cathepsins was predicted as the sum of each cathepsin's degradation alone on the substrate, however, experimentally, the amount of degraded elastin by cathepsin K and S together was only slightly more than cathepsin K alone. Figure 2-1B shows that including cathepsin cannibalism interactions and allowing cathepsin S to bind and degrade cathepsin K, in addition to the elastin, one could accurately predict the total amount of elastin degradation. Traditionally these enzymes are investigated by biochemists in isolation; however, cells secrete many proteases simultaneously and there is evidence of protease cooperativity, it is important to understand how these proteases interact cooperatively or antagonistically to correct in pathologies where the protease activity is dysregulated.

11
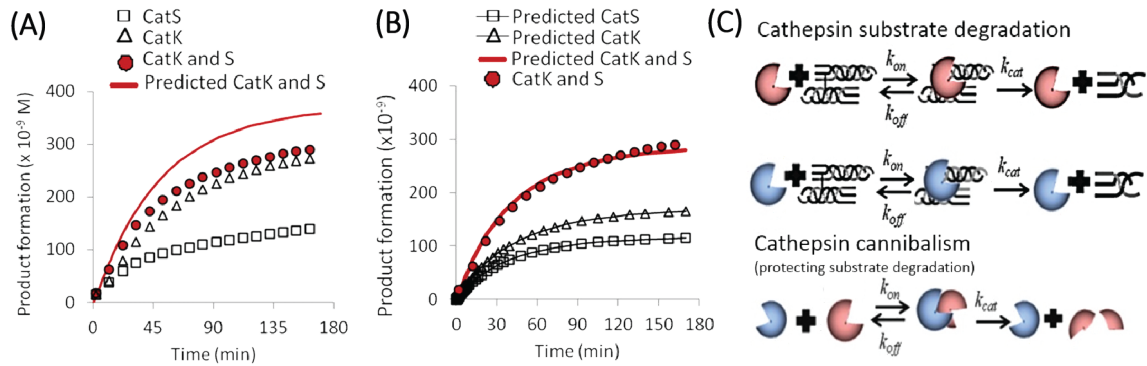
Figure 2-1. Cathepsin cannibalism terms were necessary to predict the target substrate degradation by cathepsin K and S. *(Figure adapted from Barry and Platt, JBC 2012)*

2.2 Cathepsin Inhibitors

Cells also produce the endogenous inhibitors for cathepsins, cystatin B (stefin B) and C, to regulate cathepsin activity intracellularly and extracellularly, respectively [3, 12]. These proteins serve as competitive, reversible, but tightly binding inhibitors that prevents cathepsins from binding to substrate [1]. E-64 is a small molecule inhibitor that forms a covalent bond in the active site of the cathepsin and can be used to inhibit cathepsin activity in enzymatic studies, based on yeast analog to cystatin C [99-101].

Since cysteine cathepsins are upregulated in many tissue destructive diseases including cancer metastasis [1, 36, 38, 52-54, 69-71, 75, 95, 102-115], atherosclerosis [3, 40, 41, 48, 50, 51, 68, 116-119], osteoporosis [3, 6, 42, 45, 46, 114, 120], rheumatoid arthritis [55], endometriosis [56], and tendinopathy [121, 122], pharmaceutical companies have identified the need to inhibit these potent proteases in a variety of pathologies. However, most cathepsin-specific inhibitors have failed clinical trials due to side effects,

not efficacy [47, 110, 120, 123]. Furthermore, attempts to inhibit cathepsins in cellular systems, result in non-intuitive responses. For example, during phase I trial of a cathepsin S specific inhibitor, resulted in a dose dependent increase in amount of cathepsin S and cathepsin S activity [124]. Additionally, in breast cancer cell lines, treatment with E-64 resulted in a differential cathepsin inhibition, where the amount of active cathepsin L was reduced, but the amount of active cathepsin S was increased by 24 hours [71]. Determining how these cathepsins interact with each other as a mechanism of proteolysis regulation is necessary to effectively dose pharmaceutical inhibitors to mediate cathepsin dysfunction.

2.3 Biological Machines

In the construction of engineered living systems, integrating heterotypic cells into non-natural conformations with each other and with polymers or native matrix substrates may induce emergent responses of protease secretion. Cell-cell contact, contact with matrix substrates, inflammatory cytokines, and fluid shear stresses have all been shown to induce protease expression and activity [40, 68, 94]. Their activity can subsequently degrade extracellular matrix, alter the structural configuration, and leads to failure of the machines, particularly a locomoting machine reliant on the mechanical properties and coupling that drive the locomotion.

The early walking biological machines "bio-bots" consisted of seeding rat cardiac cells on a stereo-lithographic 3D printed hydrogel "bimorph" cantilever utilizing spontaneous contraction of the cell sheet to actuate the cantilever and produce a power stroke [125]. This design was improved to control bio-bot actuation through seeding cell-

13

matrix solution of C2C12 murine myoblasts that are then differentiated into a muscle strip in the presence of insulin-like growth factor, fibrin, collagen, and Matrigel™ on a stereolithographic 3D printed polymer solution (Figure 2-2) [126, 127]. The biological machines are cultured in growth medium for three days, then switch to differentiation medium. Starting on Day 4, the biological machines can be electrically stimulated with 1 Hz frequency, 50 ms pulse width for 10 min per day to walk [126, 127].

However, without intervention, these biological machines will fail (by losing passive tension in the muscle strip and failing to contract or tissue damage in the muscle strip preventing contraction) on average after 8 days [126, 128]. Furthermore, histological staining for Masson's Trichrome reveals tissue degradation and loss of structural integrity in these biological machines [128]. It is hypothesized that this degradation is due to proteases secreted by the muscle strip and must be characterized to design effective strategies for controlling bio-bot time until failure.
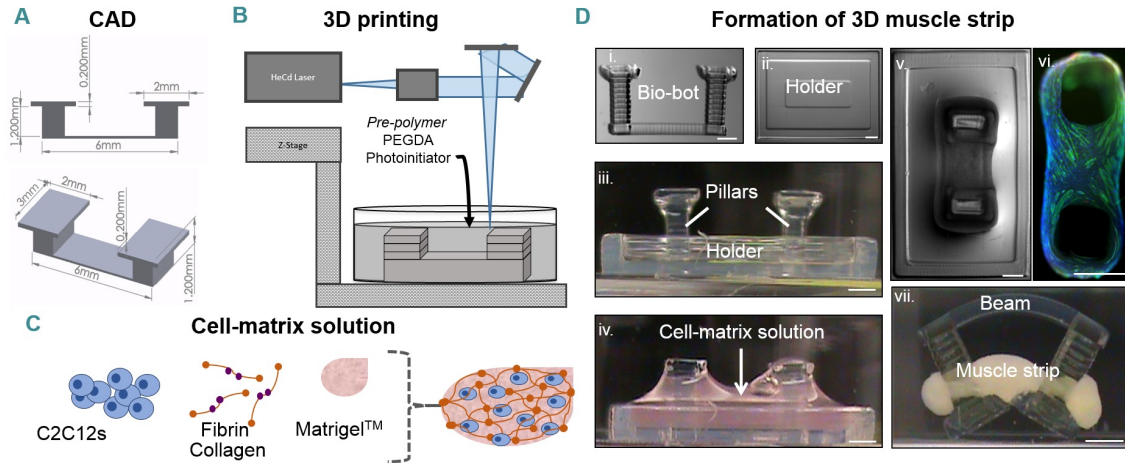
Figure 2-2. Fabrication of hydrogel structures and formation of 3D muscle strips. *(A) CAD software was used to design bio-bots with desired dimensions. (B) A Stereolithography Apparatus was used to polymerize hydrogel structures. (C) Cell-matrix solution consisted of C2C12 skeletal myoblasts, matrix proteins, and Matrigel^{TM}. (D) Fabricated bio-bot (i, side) and holder (ii, top). The solution was pipetted into a polymerized holder containing the bio-bot structure (iii-iv, side). Cells and matrix compacted around the pillars to form a solid muscle strip (v-vi, top) and the device was released (vii, side). All scale bars, 1 mm. (Figure and caption from Cvetkovic et al., PNAS 2014.)*

2.4   Motivation and Research

The ultimate goal of this project was to develop a mechanistic model of the cathepsin proteolytic network and exploiting the complex cathepsin-cathepsin interactions for applications in treatment of tissue destructive diseases and control of survival time of biological machines. Cysteine cathepsins are upregulated in a variety of tissue destructive diseases and currently researchers have had trouble specifically targeting cathepsin activity [38]. From a clinical aspect, all the well-designed cathepsin inhibitors that have been developed are effective at reducing target substrate degradation in disease states, but have failed out of phase II and III clinical trials because of off-target side-effects [47, 110]. The

results from this work offers a way to re-examine pharmaceutical inhibitors in the larger cathepsin proteolytic network context to adjust the dosing to find a safe dose for patients that still achieve the desired therapeutic effect. Using the insights gained in this research, scientists and clinicians can develop more-effective strategies at targeting the cathepsin activity in these tissue destructive illnesses.

In order to develop multicellular functional units with the ability to sense, process signals, and produce force and motion in response to external stimuli, how cells in biological machines interact with each other must first be understood. Specifically, how the cells regulate the extracellular matrix environment through proteolysis is crucial for designing biological machines that can be used for prolonged activities. Locomoting biological machines will only show electrically induced contractility for a limited amount of time, approximately 8 days following formation, until the machine fails without any additional intervention. As more advance biological machines are constructed, more cell types will be incorporated into the machine design to add additional functionality (vasculature, neuro-muscular junctions) and this heterotypic contact can change the protease signature in non-intuitive ways, which is an important element to consider in stable biological machine design and development. These proteolytic activities can be controlled as a designed benefit of the biological machines rather than a mechanism to wreak havoc on machine stability.

The research dissertation is *innovative* because it analyzes cathepsin activity through computational and molecular approaches to develop a mechanistic model of the

16

cathepsins K, L, S, and V proteolytic network and developing molecular tools to interrupt the complex cathepsin-cathepsin interactions. This work can ultimately be applied to driving effective targeting strategies for cathepsin-mediated diseases and sustainable fabrication of biological machines. Traditionally, cathepsins have been studied individually, however, cells secrete multiple cathepsin species at once, in the presence of many extracellular matrix substrates. Since these cathepsins degrade each other, the dynamics of this complex network must be understood to effectively target cathepsins whether they be upregulated in tissues destructive diseases or destabilizing engineered living systems.

CHAPTER 3.   PACMANS: A BIOINFORMATICALLY INFOMRED ALGORITHM TO PREDICT, DESIGN, AND DISRUPT PROTEASE-ON-PROTEASE HYDROLYSIS

This chapter was adapted from Ferrall-Fairbanks, M.C. et al. (2017). PACMANS: A bioinformatically informed algorithm to predict, design, and disrupt protease-on-protease hydrolysis. Protein Science, 26(4), 880-890. doi: 10.1002/pro.3113 [129]

3.1   Introduction

Cysteine cathepsins are a group of lysosomal proteases in the papain superfamily that are involved in intracellular protein degradation as well as degradation of extracellular matrix proteins when secreted by cells [1, 130]. Of particular interest, cathepsins K, L, S, and V which share 60% sequence identity and among them, include the most potent mammalian collagenase, cathepsin K, and most potent mammalian elastase, cathepsin V [3-6]. These four cathepsins have been implicated in many tissue destructive pathologies such as cancer metastasis [36, 38, 39, 53, 107, 109], osteoporosis [32, 50, 51], atherosclerosis [3, 40, 41, 50, 118, 119], rheumatoid arthritis [55], endometriosis [56], and tendinopathy [122]. Consequently, these proteases have redundancy in substrate cleavage preferences.

We recently showed that one cathepsin family member could preferentially degrade another, even in the presence of the substrate, which we termed cathepsin cannibalism [7], and explored these interactions between cathepsins K, L, S, and V in Chapter 7. It is also

important to note that this mechanism occurred between two mature enzymes, with one mature enzyme hydrolyzing the other into inactive fragments, not the well reported propeptide activation that occurs between cathepsins [9, 10, 96]. Determining that cathepsin S proteolysis of cathepsin K involved substrate degradation assays, immunoblotting, zymography, and computational modeling to test hypotheses about the cause of the unexpectedly low amount of substrate being degraded when two cathepsins were co-incubated.

Mass spectrometry can be used to sequence fragments of substrate protein after these cannibalistic hydrolysis, however this can be prohibitive due to cost and requirements of large amounts of purified protein, to test a hypothesis that may or may not be supported. We present here a computer algorithm and program to identify sites where proteases may cleave other proteases or substrate proteins, in an unbiased way, using data automatically gleaned from the online database of peptidases and their inhibitors, MEROPS (https://www.ebi.ac.uk/merops/) [131, 132]. We call this new tool PACMANS: Protease-Ase Cleavages from MEROPS Analyzed Specificities. PACMANS was validated by studying cannibalism between mature cathepsin S, as the "protease-ase," and cathepsin K, as the substrate protease, using this bioinformatically informed site-directed mutagenesis of cathepsin K to prevent its cannibalistic cleavage by cathepsin S. This bioinformatics analysis was also applied to inter-familial proteolysis of cathepsins by matrix metalloproteinases (MMPs).

3.2   Materials and Methods

3.2.1   *Algorithm and program workflow*

Protease-**A**se **C**leavages from **M**EROPS **AN**alyzed **S**pecificities (PACMANS) is a program that determines putative cleavage site locations of a dominant protease (protease-ase) hydrolyzing a substrate protease. Putative locations were determined by point-by-point comparison of a substrate protease's amino acid sequences with a protease-ase's active site specificity matrix (available from the MEROPS online peptidase database) to score the frequency each amino acid has been found in subsite P4 through P4', and is susceptible to cleavage in the protease-ase's active site [131].

PACMANS was developed using a sliding-window approach, where individual sub-sequences that fill the P4 to P4' active site pocket were scored. This pocket slides along the length of the substrate amino acid sequence and all possible sub-sequences were scored and then sorted by score. This program was implemented using MATLAB. Pseudocode for the program flow of PACMANS is given below:

Main Inputs: sequence (text string of protein under analysis) and

specmat (protease-ase matrix of specificity values)

Determine pocket size using starting and end points (testlen)

Determine the number of 1-8 amino acid sequences to be analyzed (numsequences)

for i=1:numsequnces

Determine current sequence (curseq)

for j = 1:length(curseq)

20

Determine the individual amino acid preference score for the j-th amino acid in the current sequences and store these scores in a matrix

Add individual scores together to determine cumulative sequence score

end

calculate the range and median of individual residue scores as well as subsequent fragment size for each sequence

end

Scores matrix was then sorted by maximum cumulative score and printed to a file

Specialized MATLAB functions utilized in these calculations included molweight() [from the Bioinformatics Toolbox], which calculates the molecular weight of a given amino acid sequence and quantile() [from the Statistics and Machine Learning Toolbox], which returns the quantiles of a dataset, in this case the scores, to determine the top five-percent of scores for further analysis. The computational complexity of this algorithm is $O(n)$, where n is the number of amino acids in the input sequences.

The functionality of PACMANS was further extended when used to compare substrate preferences between different protease-ases. The code was modified to calculate a normalized score, where the individual residues scores were normalized to the number of substrates used to fill the specificity matrix and then summed, like the original cumulative scoring method.

The code was packaged into a graphical user interface (GUI) using MATLAB's GUIDE Toolbox to allow use by a broader audience, which will be able to be downloaded and run as a MATLAB Application and the code was also re-implemented in PHP to host online (platt.gatech.edu/PACMANS). The functionality of the code was also adapted to allow users to vary the size of the active site picket, to a subset of the P4 to P4' used in this work.

3.2.2 Cathepsin K mutagenesis

pCMVA6 plasmid containing the gene for human cathepsin K was originally purchased from OriGene (CTSK (NM_000396) Human cDNA clone). The mutant cathepsins were made by site-directed mutagenesis using overlap extension polymerase chain reaction (PCR) with mutagenic primers to introduce point-mutations into cathepsin K gene (Supplemental Table 1) [133], which was then ligated into pcDNA 4/TO/A plasmid (Invitrogen) that contains a C-terminal polyhistidine tag for nickel-column purification. Recombined plasmid was amplified in competent bacteria cells (Strain Zymo 5α, Zymo Research) and successful mutagenesis was confirmed with DNA sequencing by Genewiz.

3.2.3 Stable expression and purification of mutant cathepsin K

HEK293T cells were stably transfected using Lipofectamine 2000 (Thermo-Fisher Scientific) and zeocin (Invivogen) as the selection marker for steady production of mutant cathepsins. C-terminal 6X polyhistidine tagged cathepsin K (wildtype and mutants) was

purified from cell lysates with the ProBond Purification System (Thermo-Fisher Scientific).

3.2.4   Multiplex cathepsin zymography

The multiplex cathepsin zymography assay was used to determine the amount of active cathepsin by the protocol previously established by our lab [134, 135].

3.2.5   Western blotting

Western blots were performed on proteins either from cell lysate (extracted using modified RIPA buffer) or purified protein in co-incubation experiments to determine levels of detectable protein. To test the cannibalism-resistance of cleavage site mutants, purified cathepsin mutants were co-incubated with or without 1 pmol of recombinant cathepsin S (Enzo) in a phosphate buffer (pH 6.0), 2mM dithiothreitol (DTT), and 1mM ethylenediamine tetraaceticacid (EDTA, Fischer Scientific) for two hours at 37°C. The reaction was quenched by the additional of a reducing loading dye and then samples were boiled for 5 minutes before loading into Western blot gels. SDS-PAGE was performed using a 12.5% SDS polyacrylamide (National Diagnostics) gel and resolved at 150V at room temperature to separate proteins by molecular weight and then transferred onto a nitrocellulose membrane (Bio-Rad) using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad). Membranes were blocked for one hour with blocking buffer (Odyssey, diluted 1:1 with PBS) and incubated over night with either a monoclonal mouse primary cathepsin K antibody (Millipore) or a polyclonal goat primary cathepsin S antibody (R&D Biosystems).

Proteins were imaged on a LI-COR Odyssey scanner using donkey anti-mouse or anti-goat secondary antibodies (Invitrogen) tagged with an infrared fluorophore. Densitometry to quantify band intensity (indicating protein amount) was performed using ImageJ (NIH). Results are presented with a mean ± standard error of the mean (SEM). Statistical analysis performed was a two-tailed, Student's t-test (Microsoft Excel).

3.2.6   Fluorescent gelatinase activity assay

Five pmol of recombinant cathepsin S (Enzo), MMP2 (Enzo), or MMP9 (Enzo) was incubated at 37°C individually or in combinations with 50 μg/ml of DQ gelatin (Thermo-Fisher) in phosphate buffer, pH 6.0, 2mM DTT, and 1mM EDTA in a 96-well plate. Changes in fluorescence were measured at an excitation in 485nm and emission of 525nm with a SpectraMax Plus microplate reader (Molecular Devices).

3.3   Results

3.3.1   Development and instructions for PACMANS

The design goals for PACMANS included (1) ease of use, (2) algorithmically efficient, (3) versatile for protease and substrate sequence input, and (4) automation. With these design goals in mind, a program using a sliding-window approach was developed, where each 8-residue segment of the amino acid sequence of the substrate protease was assessed for likelihood of cleavage in the active site pocket of the protease-ase. Individual residue scores were retrieved from MEROPS online database that populates a specificity
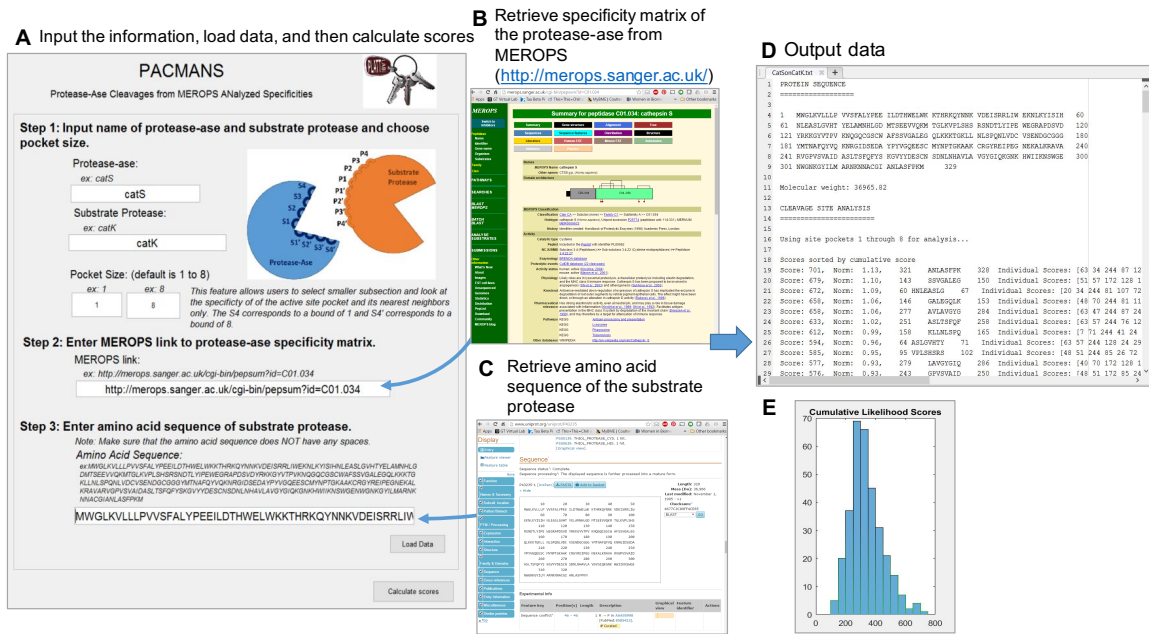
Figure 3-1. Using PACMANS to score locations of hydrolysis of the substrate protease by the protease-ase. *Input the name of hypothesized protease-ase and desired substrate protease, specificity matrix for the protease-ase, and amino acid sequence of desired substrate protease. The specificity matrix quantifies the likelihood that a given amino acid in positions P4 through P4' will result in a peptide bond cleavage between P1 and P1' in the enzyme's active site. Using the protease-ase's specificity matrix and another protease's amino acid sequence as the substrate protease, the potential locations of cleavage were scored by the sum of the individual scores that a given residue has been reported to occur in each subsite of the active site pocket. These putative sites of cleavage by protease-ase were then sorted by score and given to users in text file along with a histogram of all the scores.*

matrix for each protease that quantifies the likelihood of cleavage of each residue in each subsite of a protease's active site [131, 132]. The positions on the protease at the active site are designated with S, and for the substrate as P, and are numbered outward (typically to P4/P4') from the scissile bond between P1 and P1' positions [136], where the non-primed numbers refer to residues on the N-terminal side and the primed numbers refer to the C-terminal side of the cleaved bond (Figure 3-1). A cumulative likelihood score was

calculated by summing individual residue scores for each set of 8 amino acids being analyzed in that sliding window. A higher cumulative likelihood scored sequence would indicate higher specificity of that 8-amino acid sequence on the substrate protease for the active site in the protease-ase. Lower cumulative likelihood scores represent sequences with lower specificity for hydrolysis by the protease-ase for the substrate protease at that 8-amino acid window. This algorithm was coded and implemented in MATLAB as described in the Methods section, and is available as an online application at platt.gatech.edu/PACMANS.

Users are required to input five main pieces of information into PACMANS to generate the ranked list of putative cleavage site locations.

1) User inputs names of the protease-ase and substrate protease, which will be used to generate the output file name: "PROTEASEASEonSUBSTRATE.txt" (Figure 3-1A).

2) User defines the sliding window size of up to 8 subsites for analysis, where a lower bound of 1 would correspond to the residue in the S4 subsite and an upper bound of 8 would correspond to the residue in the S4' subsite. If these fields are left blank, PACMANS will default to the full 8 (S4-S4'), but the option is available for a smaller window.

3) User goes to MEROPS (https://www.ebi.ac.uk/merops/), click on the "Name" link under "Peptidase" on the left column on the page. Find the putative protease of interest from the Index of Peptidase Names page and click on the

"MEROPS ID" given in the right column of the table. Copy the web-address of the protease-ase of interest and paste into the link field of PACMANS interface (Figure 3-1B).

4) User inputs the FASTA-formatted amino acid sequence of the substrate protease. Amino acid sequences can be obtained from UniProt (www.uniprot.org) by searching for substrate protease of interest in the search box at the top of the page (Figure 3-1C). Select the "Entry" number of substrate protease of interest and scroll down to find the "Sequence" section. Click link to "FASTA" sequence and copy only the amino acid sequence. Be sure to remove spaces before inputting into PACMANS.

5) User clicks "Load Data" and then "Calculate Scores" to generate a text file of scores.

After execution, PACMANS will return a text file with the substrate protease amino acid sequence, molecular weight, and cleavage site analysis (Figure 3-1D). The cleavage site analysis will rank segments by cumulative sum score, and for each segment, list specific amino acid sequence being hydrolyzed, individual scores of residues (i.e. S1-P1 score), the range and median of those individual scores, as well as expected size of the two fragments of the hydrolyzed substrate protein, assuming one cleavage site. The cumulative likelihood scores will also be plotted in a histogram for quick viewing of high scoring outliers that should be considered first as most probable (Figure 3-1E).

27

Table 3-1. Top ninety-fifth percentile of cumulative cleavage scores of cathepsin S cannibalizing cathepsin K. *The table includes the highest-scored regions of where cleavage of cathepsin K by cathepsin S is most likely to occur as well as descriptive information about the range and median residue scores and a schematic of where cleavage was expected to occur within the cathepsin K protein. The pro-region of cathepsin K includes residues 1 to 114, which were excluded when determining cathepsin K mutants to interrupt cathepsin cannibalism. The '/' in the amino acid sequences indicates the scissile bond that would be hydrolyzed by cathepsin S. Schematics in the right column have a blue arrow indicating the hydrolysis would occur along with subsequence fragment sizes. Yellow stars indicate the active site residues for cathepsin K.*

	Cumulative Score	Amino Acid Sequence (P4 P3 P2 P1/P1' P2' P3' P4')	Individual Residue Specificity Scores		
			Range	Median	
1	701	(321) ANLA/SFPK (328)	[16, 244]	75	36.38 kDa 0.68 kDa
2	679	(143) SSVG/ALEG (150)	[31, 172]	74	16.91 kDa 20.13 kDa
3	672	(60) HNLE/ASLG (67)	[20, 244]	76.5	7.78 kDa 29.34 kDa
4	658	(146) GALE/BOLK (153)	[23, 244]	60.5	17.22 kDa 19.89 kDa
5	658	(277) AVLA/VGYG (284)	[16, 244]	74.5	31.49 kDa 5.57 kDa
6	633	(251) ASLT/SFQF (258)	[12, 244]	60	28.54 kDa 8.55 kDa
7	612	(158) KLLN/LSPQ (165)	[7, 244]	59	18.53 kDa 18.57 kDa
8	594	(64) ASLG/VHTY (71)	[10, 244]	48	8.11 kDa 28.94 kDa
9	585	(95) VPLS/HSRS (102)	[8, 244]	51	11.55 kDa 25.52 kDa
10	577	(279) LAVG/YGIQ (286)	[13, 172]	58.5	31.64 kDa 5.40 kDa
11	576	(243) GPVS/VAID (250)	[21, 172]	64	27.76 kDa 9.31 kDa
12	574	(169) DCVS/ENDG (176)	[20, 172]	69.5	19.69 kDa 17.39 kDa
13	554	(202) PYVG/QEES (209)	[14, 172]	46.5	23.29 kDa 13.75 kDa
14	550	(76) NHLG/DMTS (83)	[8, 244]	36	9.47 kDa 27.57 kDa
15	546	(20) EILD/THWE (27)	[8, 244]	49	2.65 kDa 34.45 kDa
16	544	(276) HAVL/AVGY (283)	[10, 172]	57	31.42 kDa 5.68 kDa

3.3.2   Application of PACMANS utility: Cathepsin S cannibalism of Cathepsin K

PACMANS was used to determine putative locations of protease-on-protease cleavage for the cathepsin S-cathepsin K relationship. Cathepsin S was input as the protease-ase and cathepsin K as the substrate protease. Top putative hydrolysis sites were ranked and output into a text file. PACMANS ranks all possible sequence segments, however only the top 5% of highest scoring sequences (16 sites) were further reviewed for structural cues that would likely affect protease-ase hydrolysis. In Table 3-1, the amino acid sequence segments are presented such that '/' indicates the scissile bond in cathepsin K. The range of the lowest and highest Individual scores as well as the median score of each 8-residue sequence is also presented. Schematics of cathepsin K with the location of hydrolysis and the fragment sizes generated were also added to Table 3-1. Five of the 16 sites were eliminated because they were in the propeptide region (the first 114 residues of
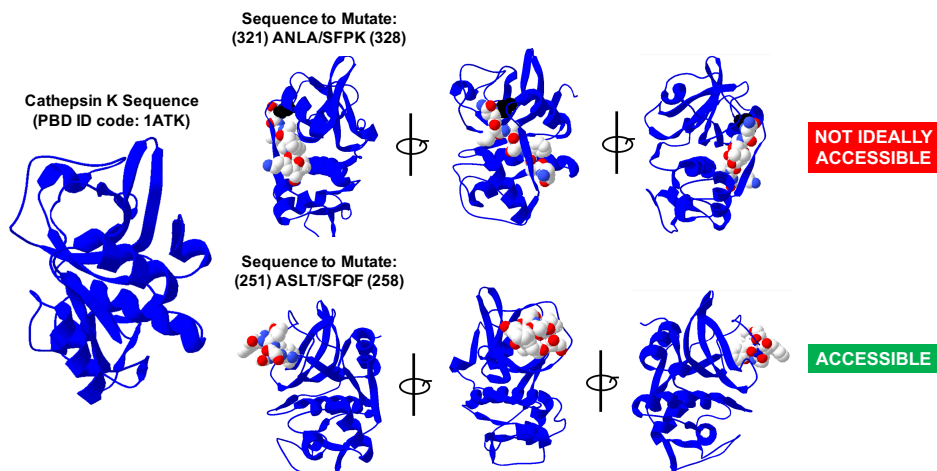
Figure 3-2. Cathepsin K putative sites of cleavage locations reviewed for accessibility to cleavage. *A cathepsin K protein data bank (PDB) model was downloaded (PDB ID: 1ATK) and sequences identified in Table 4-1 were reviewed for external accessibility by a protease-ase using Swiss-Pdb Viewer 4.1.0.*

29

cathepsin K), which would have already been cleaved off during activation to the mature enzyme (numbers 3, 8, 9, 14, and 15 in Table 3-1 with grayed text). The other eleven sites were inspected for their accessibility using Swiss-Pdb Viewer 4.1.0 of Cathepsin K (PDB ID code: 1ATK) (Figure 3-2): residues on cathepsin K's surface exposed to the aqueous environment and residues in unstructured regions were given higher priority over more internal residues because of their accessibility to cleavage by cathepsin S.

Sequence #6 ASLT/SFQF and #12 DCVS/ENDG met these criteria. L253 and V171were identified as key locations that fit all criteria and the most probable sites of cathepsin K cannibalism by cathepsin S.

3.3.3 *Validation of PACMANS predictions with site-directed mutagenesis*

To validate that these amino acid residues were important for hydrolysis of cathepsin K by cathepsin S, simulated mutations in PACMANS were analyzed. Mutating those two positions greatly reduced the overall cumulative score (L253A: 633 to 411; L253V: 633 to 561; V171A: 574 to 424) of those sequences. Site-directed mutagenesis at L253 and V171 was completed. They were mutated to alanine and valine [137] for small residues, whereas mutating to a bulky, charged residue, such as glutamic acid, would reduce the PACMANS overall score even more (L253E: 633 to 391), but could be more disruptive to protein folding which we wanted to retain.

Next, mutations were introduced to test the hypothesis that mutating PACMANS identified sites (L253, V171) would yield cannibalism-resistant cathepsin K mutants. Mutations were introduced into the cathepsin K sequence using overlap extension PCR on
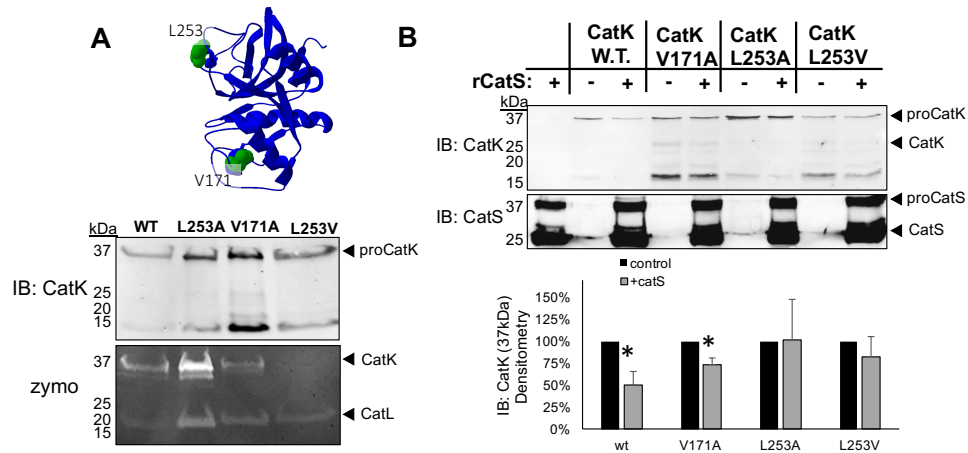
Figure 3-3. Expression of cleavage-site cathepsin K mutants yield full length protein and are protected from degradation by cathepsin S. *(A) HEK293T cells were stably transfected with a plasmid for each of the His-tagged cathepsin K mutants. Cell pellets were collected and lysed and prepared for immunoblotting to determine if protein was produced, and cathepsin zymography to test for active cathepsin K with the mutations. (B) One picomole of recombinant cathepsin S was co-incubated with purified cathepsin K wildtype or cleavage site cathepsin K mutants for 2 hours at 37°C. Cathepsin S degraded 50% of wildtype (w.t.) cathepsin K (n=3, p<0.05). Mutations at V171 showed partial protection of 25% (n=3, p<0.05), but mutations at L253 protected cathepsin K with no statistically significant difference, in the presence or absence of cathepsin S (n=3). Densitometry is quantified in the graph below.*

a plasmid with six tandem histidine residues (6xHis-tagged) at the C-terminal end of cathepsin K, confirmed by sequencing, then transfected into HEK293T cells. Translation of full length mutant cathepsin K proteins were verified by immunoblotting (reducing SDS-PAGE), and cathepsin zymography (a non-reducing SDS-PAGE method) was used to test if they were proteolytically active (Figure 3-3A). L253A, V171A, and L253V were expressed as full length proteins but, L253V did not yield an enzymatically active protein in these studies, but L253A and V171A did, as confirmed by the zymogram (Figure 3-3A).

Recombinant cathepsin K mutants were also purified from cell lysates and incubated in the presence or absence of recombinant cathepsin S to determine resistance to

degradation. Quantification of immunoblots after two hours of co-incubation with cathepsin S indicated a 50% reduction in wildtype (WT) cathepsin K (n=3, p<.02), compared to when incubated alone, only a 27% reduction in V171A cathepsin K protein (n=3, p<0.02), and no significant reduction in protein for the L253A or the L253V cathepsin K mutants (n=3) (Figure 3-3B). This indicates that PACMANS was correctly able to identify a putative site of cleavage of cathepsin K by cathepsin S, which could be mitigated by bioinformatically informed mutagenesis.

3.3.4 Inter-familial protease-on-protease interactions: cathepsin S, MMP-2, and MMP-9

In our design of PACMANS, we wanted to provide versatility for protease selection of any protease with information in MEROPS, not just cysteine cathepsins. PACMANS was applied to study interfamilial proteolysis. Matrix metalloproteinases (MMPs) are another family of proteases implicated in similar tissue destructive diseases as cysteine cathepsins, MMP-2 and MMP-9 in particular [4, 138-141], and interfamilial proteolysis can impact substrate degradation similarly to intrafamilial proteolysis. We used PACMANS to test the hypothesis that cathepsin S could degrade MMP-2 and MMP-9. To do so, cathepsin S was input into PACMANS as the protease-ase and either MMP-2 or MMP-9 as the substrate protease, the maximum cumulative scores of potential protease-on-protease interactions between cathepsin S and MMP-2 and 9 were plotted (Figure 3-4A). This procedure was repeated for MMP-2 (Figure 3-4B) and MMP-9 (Figure 3-4C) as the protease-ase as well, with cathepsin S being the substrate. Cumulative likelihood

scores for each protease-ase's hydrolysis of type 1 collagen and elastin were also analyzed

by PACMANS to compare with likelihood scores of substrate proteases as a comparative

metric for likelihood of cleavage.

The initial intention was to use the PACMANS scores from traditional extracellular

matrix substrates collagen I and elastin as a normalization strategy to compare between
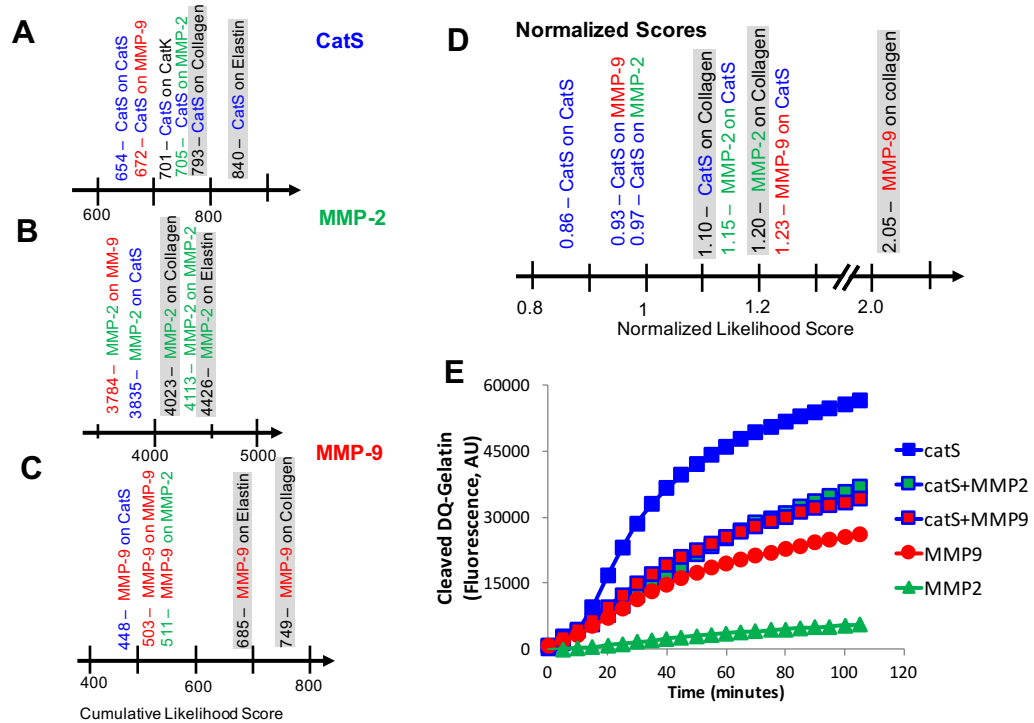
Figure 3-4. PACMANS predictions and validation of protease-on-protease interactions causing a reduction in gelatinase activity of cathepsin S when co-incubated with MMP-2 or MMP-9. *Using cathepsin S, MMP-2, and MMP-9's specificity matrices and the amino acid sequence of other enzymes as the substrate protease, the maximum cumulative cleavage likelihood was plotted, specifically focusing on (A) cathepsin S, (B) MMP-2, and (C) MMP-9 as the protease-ases of interest. PACMANS was further refined to calculate normalized scores to allow users to make comparisons among cathepsin S, MMP-2, and MMP-9 with their collagen scores as markers (D). Fluorogenic gelatin was incubated individually and in pairs of cathepsin S (catS), MMP-2, and MMP-9 (E). Cathepsin S alone degraded the most gelatin, but when co-incubated with equal amounts of either MMP-2 or MMP-9, the total gelatin degradation was approximately half that of cathepsin S alone.*

protease-ases, but the magnitude of the scores would have different ranges based on the number of papers published on the protease-ase. Maximum score was 840 for cathepsin S, 4426 for MMP-2, and 749 for MMP-9. The reason for this is that the MEROPS specificity matrix contains the frequency that a given residue has been reported in publications for each subsite. This was not constant when comparing across different proteases in this study; MMP-2, with reports as far back as 1988 [142] have more published substrate specificities analyzed and therefore generally have higher frequency values in their specificity matrix than other proteases such as cathepsin V, first reported in 1998 and have less known substrate preferences published [89]. As a result, the magnitude of the cumulative score was greater for proteases with a greater number of known substrates and preferences. To enable comparisons across protease-ases, an effective normalization strategy was developed. By dividing the individual likelihood score for a position by the total possible score of that subsite for all amino acids; $SM_{i,j}^{norm}$, the normalized score for a residue i in subsite j, can be described mathematically as:

$$SM_{i,j}^{norm} = \frac{SM_{i,j}}{\sum_{i=1}^{20} SM_{i,j}} \tag{1}$$

where i is amino acid (row), j is subsite (column), and SM is specificity matrix. Norm means normalized. The sum of the normalized scores, $SM_{i,j}^{norm}$, can be used to more accurately compare cleavage potential across multiple protease-ases, as shown below.

Using normalized scores, cathepsin S hydrolysis of MMP-2 (score of 0.97) scored higher than cathepsin K (score of 0.95), which was higher than MMP-9 (score of 0.93). MMP-2, as the protease-ase, was more likely to hydrolyze cathepsin S (score of 1.25), than

MMP-9 (score of 0.88) (Figure 3-4D). This suggested that MMP-2 would preferentially hydrolyze cathepsin S more than vice versa and that co-incubation would reduce cathepsin S's degradation ability. PACMANS-based hypotheses were then tested experimentally by co-incubating cathepsin S with MMP-2 or MMP-9 and monitoring cleavage of a fluorescently quenched gelatin substrate; increased fluorescence indicates substrate degradation. By itself, cathepsin S degraded more gelatin than the other proteases with MMP-2 degrading 10.3% as much and MMP-9 degrading 46.0% as much gelatin as cathepsin S alone (Figure 3-4E). When co-incubating with cathepsin S, only 65.2% of total gelatin was degraded with MMP-2, and only 60.5% with MMP-9 compared to cathepsin S alone (100%), even though there was initially twice the molar amount of proteases in the system (Figure 3-4E). If no interactive proteolysis was occurring, then the results from co-incubation would be expected to be the sum of each individual protease's activity. This suggests that MMP-2 and MMP-9 were reducing the amount of active cathepsin S in the system by cleaving it, or were distracting cathepsin S from degrading the gelatin serving as an alternative substrate for cathepsin S to bind, or both, which supports the hypothesis from PACMANS that MMP-2 and MMP-9 could cleave cathepsin S when co-incubated, and to some degree, vice versa.

3.4 Discussion

PACMANS uses a bioinformatics approach to analyze the primary structure of a protease (substrate protease) and determines possible locations that could be vulnerable to hydrolysis by another protease (here termed, a protease-ase), PACMANS is unique

35

compared to other putative site cleavage programs, by allowing users to define the protease of interest, and utilizing the specificity matrix from MEROPS to rank the most likely locations of cleavage. Although this program was initially conceived to analyze proteases cleaving other proteases, which could be bidirectional, it can also be used for traditional, non-protease substrate proteins, as well as short peptide sequences. Putative sites on cathepsin K susceptible to cleavage by cathepsin S were identified by PACMANS, confirming its ability to score sites based on protease-ase preferences. Cathepsin S prefers moderately-sized, branched hydrophobic residue (such as leucine in L253) in the P2 [143-145], and these were detected in the top 5% of putative sites on cathepsin K. Introducing mutations into cathepsin K at sites L253 and V171, protected cathepsin K from cleavage by cathepsin S (Figure 3-3). This validated PACMANS as a useful tool for predicting locations of protease-on-protease interactions and cleavage. Taken together, this confirms cathepsin cannibalism between cathepsin S and cathepsin K and that its outcomes can be modified by mutational analysis and application.

Other protease families may undergo similar proteolytic interactions between family members, but we also used PACMANS to demonstrate the possibility of inter-familial regulation of proteases using cathepsin S/MMP-2,-9 studies. PACMANS predicted interactions and preferential hydrolysis directions between cathepsin S vs. MMP-2 and MMP-9 (which are both gelatinases), were then tested experimentally by monitoring gelatinase activity of these proteases in isolation, in concert, or against each other. PACMANS analysis indicates that both MMP-2 and -9 have greater preference for cleaving cathepsin S, then cathepsin S has for cleaving them, and we observed

experimentally that cathepsin S gelatinase activity was reduced when co-incubated with MMP-2 or MMP-9 compared to its incubation by itself indicating interactions between MMP-2 and cathepsin S and of MMP-9 with cathepsin S were reducing cathepsin S activity. Cathepsin S could be acting as a decoy or distraction for MMPs, causing them to bind to and hydrolyze cathepsin S preferentially over the gelatin substrate, reducing the amount of gelatin being degraded over the same unit time. If MMP-2/-9 are cleaving cathepsin S, there is the additional consequence of inactivating a fraction of cathepsin S, preventing that amount of active enzyme from degrading gelatin resulting in a net reduction in total gelatinase activity. This reduced gelatinase activity observed occurred even when there was twice the molar amount of active protease present since equimolar amounts were added for each of the experimental conditions (Figure 3-4E).

Expanding PACMANS to test inter-familial protease relationships between cathepsin S and MMP-2/MMP-9, it became clear that the initial scores calculated from MEROPS were not comparable across protease-ases. The protease-ase scores clustered depending on the magnitude of scores in the MEROPS specificity matrices, rather than just biological preference. For example, MMP-9 is known to be a strong collagenase [4], however, based solely on the scores, MMP-2 preference for collagen, elastin, as well as cathepsin S, MMP-2, and MMP-9 was greater than MMP-9's score for degrading collagen. To compare across cathepsin S, MMP-2, and MMP-9, a normalized score was developed, which normalized the specificity matrix by the total number of substrates used to generate it. After normalization, protease-ase preferences could be more directly compared and conclusions made based on experimental data (Figure 3-4D-E). MMP-9 was then deemed

the strongest collagenase (score 2.05) followed by MMP-2 (score 1.20) and cathepsin S (score 1.10), which better approximates published comparisons [4].

PACMANS customizability allows users to input amino acid sequences and proteases of interest. Other programs developed to determine proteolytic cleavage sites, more commonly have a predefined, set list of proteases available for analysis. PROSPER [146] and PeptideCutter [147] are two programs that differ from PACMANS in that they analyze a set list of enzymes to cleave the amino acid sequence, while PACMANS's only limit is that the protease must have a specificity matrix characterized in MEROPS (which as of the time of this submission is 263). PeptideCutter did not have any of the cysteine cathepsins, and PROSPER only included cathepsin K.

Prediction of Protease Specificity (PoPS) is a robust program similar to PACMANS, but users must manually type each subsite's specificities for each amino acid by hand (which can be error-prone) [148] instead of directly linking to MEROPS like PACMANS. Furthermore, PoPS has the user choose between scaling methods. PACMANS normalizes the specificity matrix and computes a normalized score for comparison. PoPS gave similar results to PACMANS with the sequence *(143) SSVG/ALEG (150)* having a maximum score of 17.83, which was the second entry in PACMANS results (Table 3-1). However, the sequences containing L253 (rank 9, score: 9.28) and V171 (rank 19, score: 5.85) in the P2 locations were lower ranked in the top scores from PoPS and were shown from the PACMANS identified sites as key locations for interrupting hydrolysis.

In summary, PACMANS is now added to other tools that can be used to test basic scientific questions for proteolysis, as well as design interventions in tissue-destructive diseases. PACMANS also can be used to design mutant proteins that can be therapeutically resistant or enhanced for protease-on-protease interactions. For example, if design goals were for cathepsin K to be degraded quickly for some biotechnology application or for some biologic manufacturing tasks, then instead of expressing L253A mutants resistant to cannibalism, one could express T254G or Q257P; both of these mutants would increase likelihood scores to enhance its degradation by cathepsin S. Understanding protease-on-protease interactions can also allow for accurate dosing of protease inhibitors, considering the activities of protease-on-protease interactions that can generate off-target side effects caused by miscalculations of protease concentrations in pharmacological analyses. PACMANS offers an inexpensive *in silico* method to perform first pass analyses prior to the mutagenesis or more costly methods to confirm, which will save time and money for proteolysis researchers.

CHAPTER 4. COMPUTATIONAL PREDICTIONS OF CYSTEINE CATHEPSIN-MEDIATED FIBRINOGEN DEGRADATION

This chapter was adapted from Ferrall-Fairbanks, M.C.*, West, D. M.*, et al. Computational predictions of cysteine cathepsin-mediated fibrinogen degradation. Manuscript submitted.

4.1 Introduction

Fibrin networks form the basis of a blood clot in response to injury to the endothelial lining of blood vessels to prevent blood loss and maintain homeostasis. Maladies in this wound healing response can results in an increased risk unnecessary clot formation and thrombotic complications, such as those found in individuals with sickle cell disease, diabetes, and rheumatoid arthritis [149, 150]. Fibrinogen, the fibrin precursor, is a 340kDa protein that circulates in the blood at high concentrations [151-154] in preparation for an injury event. It is a hexameric protein, composed of two subunits, each with three polypeptide chains (α, β, and γ), held together by disulfide bonds, with a flexible region (E-domain) and with large globular regions (D-domain) on the ends [155]. It is within the E-domain that thrombin, a serine protease, binds and cleaves N-terminal fibrinopeptides A and B, producing disulfide linkages. In the presence of Factor XIIIa, fibrinogen monomers are converted into an insoluble fibrin network, which at the site of an injury, traps red blood cells and platelets to prevent blood loss [152, 156].

Fibrin-based scaffolds have emerged for a wide variety of tissue engineering constructs from hydrogels to synthetic applications in muscle strips of biological machines [128, 151, 153, 156-158], microphysiological systems [159, 160], and regenerative medicine applications as sealants during procedures and hemostatic agents in surgery [126], and perfusable microvascular networks (MVNs) [161]. Despite enhanced angiogenesis, networks destabilize within six days [159]. Plasmin is the endogenous enzyme that degrades fibrin during resolution of blood clots [159], yet even with the inhibitors of plasmin-mediated fibrin degradation, these fibrin-based constructs are being destabilized in an uncontrolled manner, indicating other proteases produced by cells may be fibrinolytic agents.

Non-plasmin fibrinolytic pathways must be considered to explain fibrin matrix destabilization. While thrombin and plasmin are known proteases involved in the coagulation cascade to cleave fibrinogen and degrade fibrin, we hypothesize that there are other proteases with similar capabilities that have been overlooked due to physiological constraints that are no longer applicable in engineered biological systems. One such candidate are cysteine cathepsins, a family of proteases that are known to degrade extracellular matrix proteins, including collagen and elastin, as well as themselves [3, 102, 130]. These enzymes are implicated in various types of tissue destructive diseases [1, 7, 130] and their substrate redundancy has been well documented [36, 38, 39, 53, 107, 109]. This family includes the most powerful mammalian collagenase and elastase [1, 19, 114,

130]; however, very little work has been done exploring cysteine cathepsins' potential to hydrolyze fibrinogen and fibrin.

Here, we use computational methods of possible molecular docking interactions combined with bioinformatics substrate specificity analysis to analyze the proteolytic interactions between fibrinogen α, β, and γ chains and cathepsins K, L, and S. Using these two independent computational techniques to implicate regions of fibrinogen susceptible to hydrolysis by cysteine cathepsins could provide a mechanistic understanding of how these cysteine cathepsins play a role in destabilizing biomaterials, tissue engineered constructs, and microphysiological systems.

4.2   Materials and Methods

4.2.1   Molecular docking simulation

Molecular docking was accomplished using the web server ClusPro 2.0 [3-6]. The atomic co-ordinates of human fibrinogen (PDB ID: 3GHG), cathepsins K (PDB ID: 4N79), cathepsin L (PDB ID: 5F02), and cathepsin S (PDB ID: 4P6E) were downloaded from the Protein Data Bank [162-166]. Briefly, the crystal structure of cathepsins were stripped of all water molecules and ligands, then uploaded to the server as a ligand. Human fibrinogen, the much larger molecule, was uploaded as a receptor. For the study of individual fibrin chains, the role of ligand and receptor remained constant, with specific chains selected during the upload of the fibrinogen model. Using ClusPro allows for selection of binding simulations that favor binding through electrostatic interactions, hydrophobic interactions,

and Van der Waals interactions. The simulations for this work were performed with a balanced approached, meaning an unbiased binding method acknowledging all bonding interactions when developing docking models. The structures were rotated and translated in three dimensions, simulating 10^9 possible binding interactions. The docking simulations followed a balanced approach, producing 10^3 energetically indistinguishable conformations, all representative of a possible binding motif. With these conformations, models in which one atom of the ligand was within 1 nm of one receptor atom were selected for clustering. Following selection, the distance between the α-carbons of the corresponding residues of the selected ligands were calculated. Ligands with a carbon-carbon bond distance within a 9-Å interface RMSD radius [166] between the ligand and macromolecule were clustered into groups, and the ligand with the most neighboring molecules were chosen as the representative model for the cluster. The models were visualized using the PyMol molecular viewer (PyMOL Molecular Graphics System; DeLano Scientific: San Carlos, CA, 2002). Although crystal structures do not capture all the different confirmations that may be observed in solution, we focused on models in which the active site of the ligand was facing the receptor, and discarded possible allosteric alternate binding modes as well as overlapping structures [167]. From the crystal structure of fibrinogen, only a portion of the overall structure was captured and further investigated, for the α chain region included was A46 through Q219, for the β chain region included was K88 through F489, and for the γ chain region included was R40 through G421.

4.2.2 Putative cleavage site determination with PACMANS

PACMANS (Protease-Ase Cleavages from MEROPS ANalyzed Specificities) [168, 169] developed by us (and available online: platt.gatech.edu/PACMANS) is an algorithm for users to input a URL link to the MEROPS specificity matrix of a protease (here cathepsin K, L, and S) and the amino acid sequence of a substrate (here fibrinogen α, β, and γ chains), and returns a ranked list scoring hydrolysis of each eight amino acid sequence in the substrate sequences. Using a sliding window approach, the program scores the likelihood of hydrolysis occurring between the P1/P1' scissile bond when a given 8 amino acid segment of the substrate was in the S4 through S4' subsites of the protease's active site using the protease's specificity matrix characterized by MEROPS. The top 5% of scores are then further analyzed for accessibility by the protease and compared with molecular docking results.

4.2.3 Fibrinogenolysis experiment

Human fibrinogen (FIB 3 Plasminogen, von Willebrand Factor and Fibronectin Depleted; Enzyme Research Laboratories) was co-incubated with recombinant human cathepsins K, L, or S (Enzo) in assay buffer (0.1M sodium phosphate buffer, pH 6.0, 1mM EDTA) with 2mM dithiothreitol (DTT) for 24 hours at 37°C. The samples were centrifuged at 7012g and the supernatant was saved. The pellet was prepped in 50mM Tris-HCl/100mM NaCl with 25% betamercaptoethanol (OmniPur) and sonicated. Samples were prepared with reduced loading dye (5X - 0.05% bromophenol blue, 10% SDS, 1.5 M Tris,

50% glycerol, 25% betamercaptoethanol) and run on 10% SDS-PAGE gels and stained with Coomassie stain (10% acetic acid, 25% isopropanol, 4.5% Coomassie blue), followed by destaining (10% isopropanol and 10% acetic acid) and imaged with an ImageQuant LAS 4000 (GE Healthcare), and densitometry performed with ImageJ (NIH) to quantify fibrin(ogen) polypeptide chains.

4.3   Results

4.3.1   *Molecular docking predicts primary cathepsin binding sites and PACMANS ranks likelihood of cleavage by cathepsins*

ClusPro molecular docking was used to calculate and visualize putative cathepsin-fibrinogen interactions based on energetically favourable binding confirmations. Our goal was to determine specific residues on fibrinogen with high probability of binding of cathepsin K, L, and S active sites. The active sites of papain family cysteine cathepsins are highly conserved [129] with a cysteine (C), histidine (H), and asparagine (N) to form the catalytic triad. For cathepsin K, its catalytic triad is formed of residues C25, H162 and N182 [170], for cathepsin L, residues C25, H163 and N187 [1], and for cathepsin S, residues C25, H164, and N184 [1]. ClusPro generated models with similar conformations closely clustered together to produce a single model representative of the group. The structures provided by ClusPro were all energetically favorable and unique (see Materials and methods), allowing for a thorough and plausible discovery of protein-substrate interactions.

45

To investigate the potential for cathepsin-mediated hydrolysis of fibrinogen (α, β, and γ) chains, the top 5% of scored PACMANS sequences were further investigated and compared against the molecular docking results. The full top 5% of scored sequences of a cathepsin on individual fibrinogen chains for cathepsins K, L, and S are in Appendix A.3. Locations that were independently identified by these two methods, molecular docking and specificity preference from PACMANS, were further analyzed to determine mechanistic interactions.

With correlation between putative cleavage sites and molecular docking binding sites, the next step focused on what possible molecular interactions that would make these sites favorable for catalysis. The top scoring sequences from each fibrinogen chain was further analyzed to determine possible interactions between amino acid residues of cathepsins K, L, and S and the cleavage site of each fibrinogen chain. For this study, correlating binding regions on structured, helical portions of the subunits was preferable to binding in the globular regions between cathepsins would have easier access for fibrinogenolysis at these locations in solution.

4.3.2   Cathepsins binding to fibrinogen

The binding of cathepsins appeared to be distributed along the length of the fibrinogen monomer, binding to all three chains, and included models with binding at both the E- and D-domains. This was a similar binding conformation in the crystal structure of thrombin bound to fibrinogen E-domain (PDB ID: 2A45), with the active site facing the

fibrinogen subunit, an advantageous conformation for cleavage. Additional models also revealed docking of cathepsin K around the D-domain of fibrinogen, regions where the fibrinogenolytic and fibrinolytic enzymes thrombin and plasmin, respectively, are known to bind for cleavage events. The helical domain of the fibrinogen subunit was where cathepsins K, L, and S hydrolysis of fibrinogen was most likely to occur. This contrasts with thrombin, where primary fibrinolysis is proposed to occur specifically in the E domain to generate fibrin monomers. The roles of thrombin, plasmin, and cathepsin as proteases associated with tissue formation implies that they can share similar binding and cleavage tendencies, indicated by their comparative interactions with fibrinogen. However, with differing amino acid compositions and substrate redundancy, cathepsins K, L, and S can display accessory roles in fibrinogen degradation or even fibrin generation.

4.3.3   *Cathepsin K cleavage likely in α and γ chains of fibrinogen*

Simulated docking showed the active site of cathepsin K directly interacting with the helical subunits in the models, binding in regions that directly correlate to the predicted cleavage sites from PACMANS (Figure 4-1). For cathepsin K (Figure 4-1B), there were five top ranked PACMANS sequences in the α chain with corresponding molecular docking models and four top ranked PACMANS sequences in the β and γ chain. From PACMANS, the greater the score, the more likely cleavage occurs and the normalized scoring allowed for comparison between cathepsins on each fibrinogen chain. Based on the culled PACMANS scoring information, the α chain in the helical region at residues #111-118 EILR/GDFS was the most likely location for cleavage by cathepsin K to occur (with a

score of 0.91). The next top scoring sequences was at residues #90-97 IQLT/YNPD in the γ chain (score of 0.90), #202-209 KQLE/QVIA in the α chain (score of 0.87), and #179-186 HQLY/IDET in the β chain (score of 0.85). Also, interestingly, the top scoring cleavage sequence in the γ chain by cathepsin K, #90-97 IQLT/YNPD (score of 0.90), was five residues shifted from a known plasmin cleavage site on fibrin (QLIK/AIQL) [171].

We focused on highest scoring PACMANS region that correlated with a docking site. For the α chain of fibrinogen (Figure 4-1A), there was only cathepsin K binding in the helical region. Within the sequence EILR/GDFS, hydrolysis was predicted between R95 and G96, in an amphipathic cleavage site (Figure 4-2A). There appeared to be a possible electrostatic interaction between E111 sidechains of catK and R114 in the cleavage site, (Figure 4-2A) and this interaction could lock these residues in an extended conformation, exposing the main chain of arginine and glycine for cleavage by the catalytic triad.

Similar trends were also seen in the β and γ chains (Figure 4-2B-C). In the β chain, there was a possible electrostatic interaction between T186 and the backbone of Y182 of the fibrinogen cleavage site HQLY/IDET. Cleavage for this site was predicted between Y182 and I183 and the cleaved isoleucine would be able to form a hydrogen bond with Q180. While the current distance was 4.4 Å, the binding of fibrinogen β chain could promote a conformational change placing these residues within a distance suitable for molecular bonding.

In the highest scoring binding site for the γ chain IQLT/YNPD, Y94 was in position

to bind with D97 within a hydrogen bonding distance of 2.9 Å, anchoring this fibrinogen

cleavage site (Figure 4-2C). The ability for both intra- and inter-molecular bonding events

to stabilize predicted cleavage sites was an interesting finding that can serve as a molecular

basis for cathepsin K fibrinogenolysis.

(B)

Fibrinogen Sequence	Chain	Cumulative Score	Normalized Score	Cluster Location
(111) EILR/GDFS (118)	α	155	0.91	C
(202) KQLE/QVIA (209)	α	152	0.87	B/E
(132) EDLR/SRIE (139)	α	139	0.81	B
(73) LIDE/VNQD (80)	α	135	0.83	D
(90) NSLF/EYQK (97)	α	128	0.75	C
(179) HQLY/IDET (186)	β	141	0.85	B
(105) LCPT/GCQL (112)	β	137	0.84	D
(200) SILE/NLRS (207)	β	144	0.83	B
(203) ENLR/SKIQ (210)	β	133	0.78	B
(90) IQLT/YNPD (97)	γ	154	0.90	C
(155) AQLE/AQCQ (162)	γ	147	0.85	B
(114) KMLE/EIMK (121)	γ	148	0.84	E
(134) RYLQ/EIYN (141)	γ	141	0.82	B/E

Figure 4-1. Top scored locations for cathepsin K binding and cleavage of fibrinogen.
*(A) Molecular docking of cathepsin K on fibrinogen monomer with the α chain is in red,
the β chain is in green, and the γ chain is in blue. Cathepsin K bind in various formations
and each color indicates a different docking model (note there is some overlap with the
model found in cluster location B and E). Structures used were PDB ID 3GHG (fibrinogen)
and 4N79 (cathepsin K). (B) Predicted fibrinogen cleavage sites by cathepsin K from
PACMANS and molecular docking models.*

(A) <u>FIBRINOGEN ALPHA CHAIN</u>

(B) <u>FIBRINOGEN BETA CHAIN</u>

(C) <u>FIBRINOGEN GAMMA CHAIN</u>
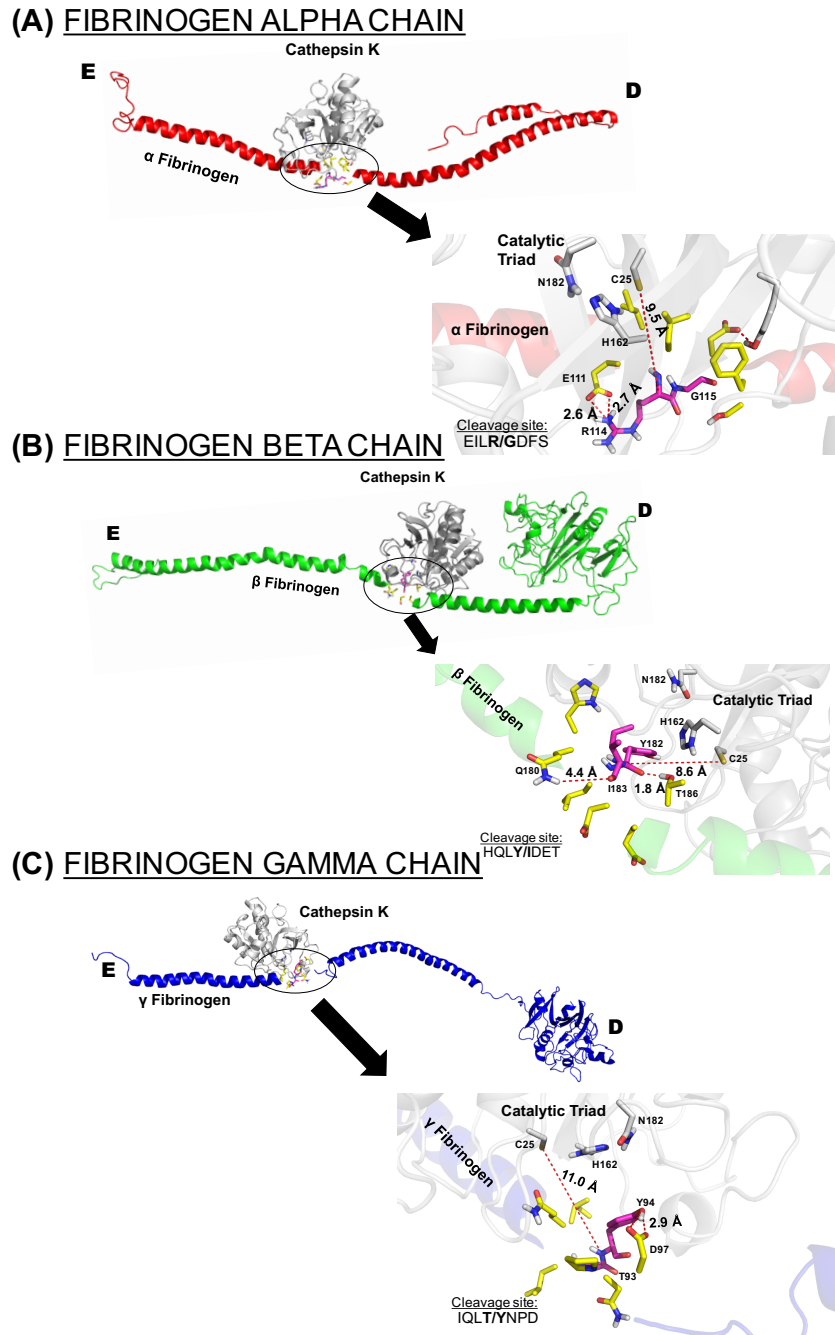
Figure 4-2. Proposed binding interfaces of cathepsin K on each fibrinogen chain. *Amino acid sequences of predicted cleavage sites for cathepsin K on fibrinogen (A) α chain, (B) β chain, and (C) γ chain were visualized in the docking models. The close up is rotate 180° and the bolded site indicates where cleavage was predicted to occur.*

50

4.3.4 *Cathepsin L has highest scored cleavage location in the α chain and greatest number of scored sequences in the β chain*

Cathepsin L docking sites multiple at cluster D where α, β, and γ cleavage as for E. A, B, and C docking sites did not have matching high PACMANS scores (Figure 4-3A). For cathepsin L (Figure 4-3B), the highest scoring sequence was found in the α chain at #111-118 EILR/GDFS (score of 0.82), and the top scoring site for cathepsin K cleavage as well (Figure 4-2A), followed by #185-192 ETVN/SNIP in the β chain (score of 0.80) and SILT/HDSS (score of 0.80) in the γ chain. IQLT/YNPD (score of 0.75) in the fibrinogen γ chain was another likely cleavage site for cathepsin L and was a predicted cathepsin K cleavage site also.

When cathepsin L binds to the α chain to sequence EILR/GDFS (Figure 4-4A), the flexible R114 can be stabilized by two cathepsin aspartates, D137 and D160, both within molecular bonding distance to the arginine amino groups, 2.8 and 2.9 Å respectively. An electrostatic interaction could ablate the flexibility of arginine, exposing the scissile bond for successful fibrinogen cleavage. This was a similar stabilizing mechanism when cathepsin L was bound to the β chain (Figure 4-4B), with E185 in proximity to N188 and S189 in the cleavage site of sequence ETVN/SNIP. For the γ chain (Figure 4-4C), the cleavage sites do not involve amino acids with flexible side chains, and therefore does not appear to be an amino acid in an advantageous position available for bond stabilization. However, T128 in the cleavage sequence SILT/HDSS is 5.2 Å from S125 and could provide a possible hydrogen bond in a natural environment.
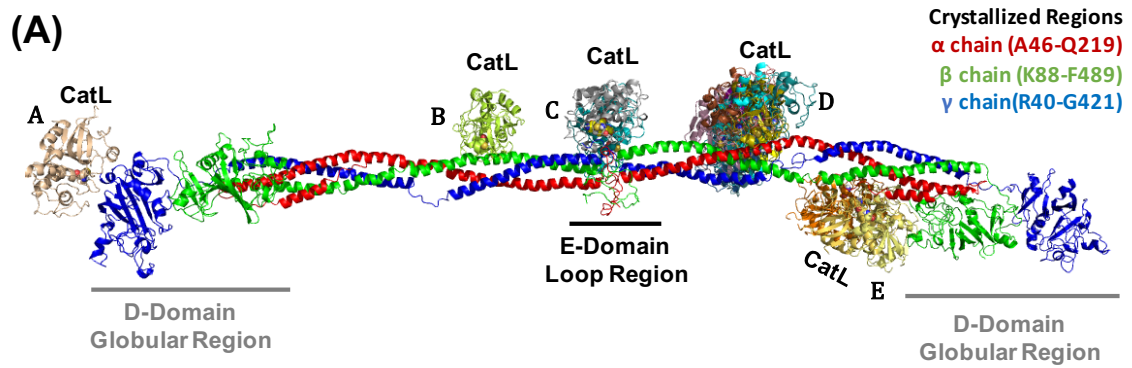
(A)

Crystallized Regions
α chain (A46-Q219)
β chain (K88-F489)
γ chain(R40-G421)

(B)

Fibrinogen Sequence	Chain	Cumulative Score	Normalized Score	Cluster Location
(111) EILR/GDFS (118)	α	863	0.82	B/D
(132) EDLR/SRIE (139)	α	789	0.75	E
(185) ETVN/SNIP (192)	β	839	0.80	E
(196) RVLR/SILE (203)	β	813	0.76	E
(135) EAVS/QTSS (142)	β	776	0.74	B/D
(200) SILE/NLRS (207)	β	750	0.71	E
(125) SILT/HDSS (132)	γ	840	0.80	E
(90) IQLT/YNPD (97)	γ	801	0.75	B/D
(108) ATLK/SRKM (115)	γ	748	0.71	E

Figure 4-3. Top scored locations for cathepsin L binding and cleavage of fibrinogen. *(A) Molecular docking of cathepsin L on fibrinogen monomer with the α chain is in red, the β chain is in green, and the γ chain is in blue. Cathepsin L binds in various formations and each color indicates a different docking model (note there is some overlap with the model found in cluster location B and D). Structures used were PDB ID 3GHG (fibrinogen) and 5F02 (cathepsin L). (B) Predicted fibrinogen cleavage sites by cathepsin L from PACMANS and molecular docking models.*

(A) FIBRINOGEN ALPHA CHAIN
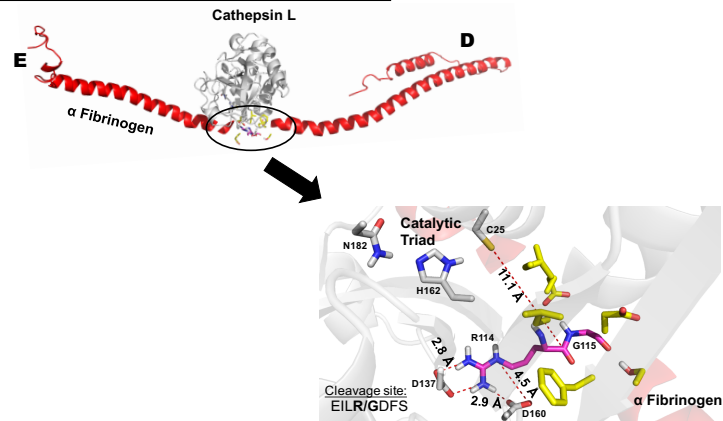
(B) FIBRINOGEN BETA CHAIN
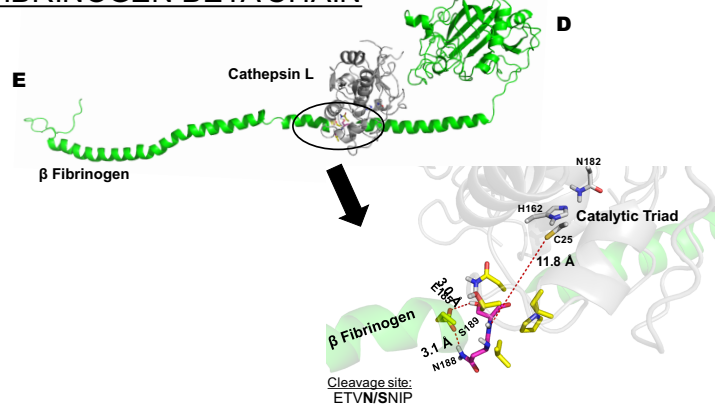
(C) FIBRINOGEN GAMMA CHAIN
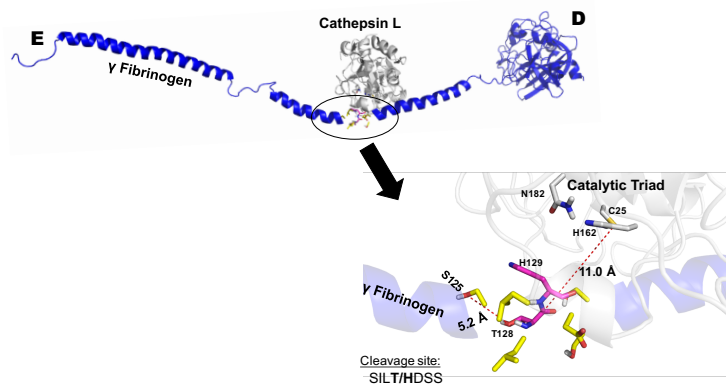
Figure 4-4. Proposed binding interfaces of cathepsin L on each fibrinogen chain. *Amino acid sequences of predicted cleavage sites for cathepsin L on fibrinogen (A) α chain, (B) β chain, and (C) γ chain were visualized in the docking models. The close up is rotate 180° and the bolded site indicates where cleavage was predicted to occur.*

4.3.5   Cathepsin S has highly scored locations of cleavage in the γ and β chains in the

helical region

For cathepsin S, clusters B, D, and E had the most probable sites of binding (Figure 4-5A). There were the most top ranked sequences from PACMANS and molecular docking models in the β chain of fibrinogen and the highest scoring sequence occurring in the γ chain at AQLE/AQCQ (with a score of 0.86), same as for cathepsins K and L. The next highly scored sequence occurred in the β chain at EAVS/QTSS (score of 0.84) and in the α chain at EILR/GDFS (both with scores of 0.81). In the β chain, the sequence ENLR/SKIQ was also predicted as a cleavage site for both cathepsin S (0.78) and cathepsin K (0.78).

The intramolecular interactions stabilizing cathepsin S when bound to fibrinogen were similar to those of cathepsins K and L. When cathepsin S binds to the α chain at sequence EILR/GDFS (Figure 4-6A), the same binding sequence of cathepsin K and cathepsin L with the highest PACMANS score, R114 is stabilized by a hydrogen bond with E111, identical to cathepsin K. For the β sequence EAVS/QTSS (Figure 4-6B) and γ sequence AQLE/AQCQ (Figure 4-6C), the scissile bond does not involve flexible side chains, however with the current rigid conformation Q139 can form a hydrogen bond with S143 in the β sequence. In the γ chain, E158 can form a possible hydrogen bond with Q162 in the cleavage sequence (Figure 4-6C). The availability of charged polar amino acids for intramolecular stabilization of flexible scissile bonds, and not necessarily for cleavage sites with smaller side chain residues, could provide validity to the idea that cleavage by

cathepsins can occur with intra-molecular bonding as a mechanism to facilitate successful

cleavage of flexible scissile bonds.
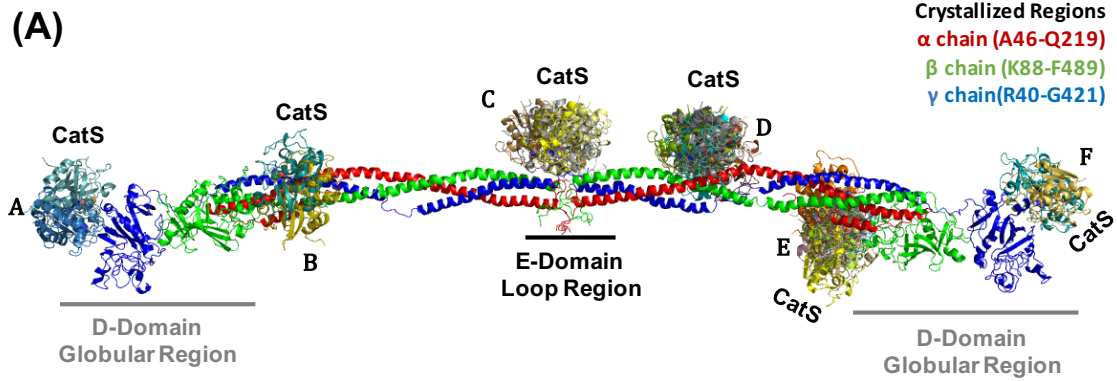
(A)

Crystallized Regions
α chain (A46-Q219)
β chain (K88-F489)
γ chain(R40-G421)

(B)

Fibrinogen Sequence	Chain	Cumulative Score	Normalized Score	Cluster Location
(111) EILR/GDFS (118)	α	592	0.81	D
(132) EDLR/SRIE (139)	α	576	0.78	B/E
(135) EAVS/QTSS (142)	β	614	0.84	D
(203) ENLR/SKIQ (210)	β	572	0.78	B/E
(185) ETVN/SNIP (192)	β	571	0.78	B/E
(196) RVLR/SILE (203)	β	575	0.77	B/E
(132) NNVE/AVSQ (139)	β	563	0.75	D
(200) SILE/NLRS (207)	β	552	0.75	B/E
(128) DELN/NNVE (135)	β	540	0.73	D
(155) AQLE/AQCQ (162)	γ	632	0.86	B/E
(125) SILT/HDSS (132)	γ	584	0.79	B/E
(53) DFLS/TYQT (60)	γ	537	0.73	C

Figure 4-5. Top scored locations for cathepsin S binding and cleavage of fibrinogen.
(A) Molecular docking of cathepsin S on fibrinogen monomer with the α chain is in red, the β chain is in green, and the γchain is in blue. Cathepsin S binds in various formations and each color indicates a different docking model (note there is some overlap with the model found in cluster location B and E). Structures used were PDB ID 3GHG (fibrinogen) and 4P6E (cathepsin S). (B) Predicted fibrinogen cleavage sites by cathepsin S from PACMANS and molecular docking models.
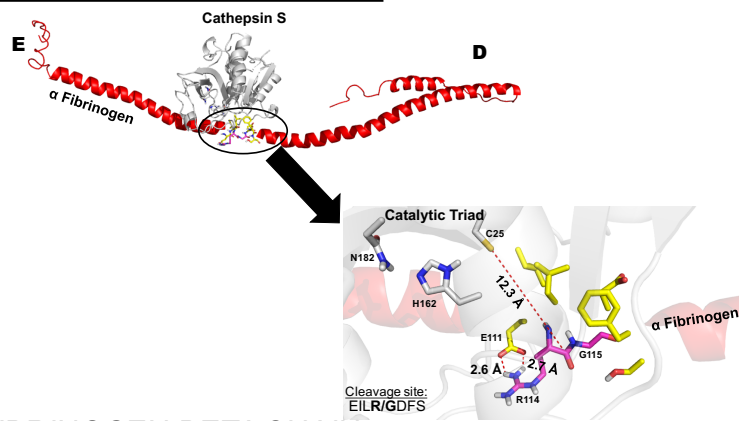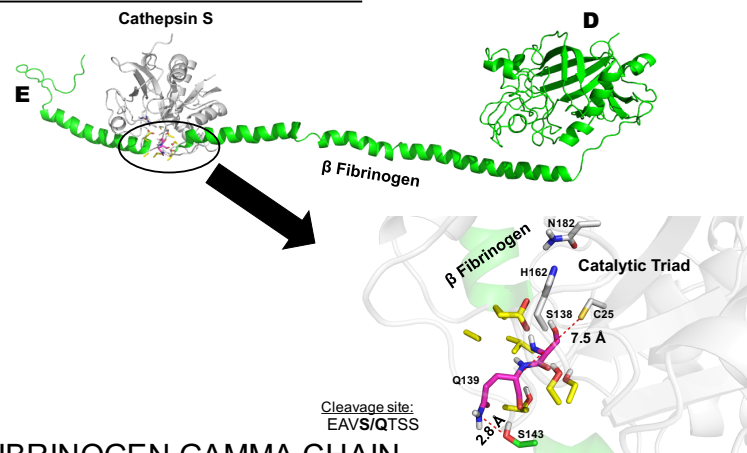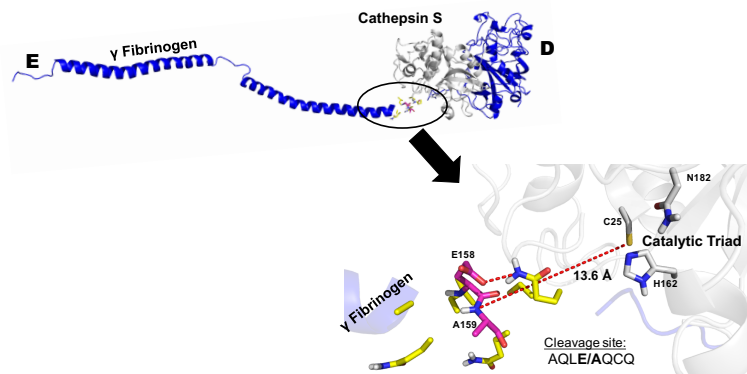
Figure 4-6. Proposed binding interfaces of cathepsin S on each fibrinogen chain. *Amino acid sequences of predicted cleavage sites for cathepsin S on fibrinogen (A) α chain, (B) β chain, and (C) γ chain were visualized in the docking models. The close up is rotate 180° and the bolded site indicates where cleavage was predicted to occur.*

56
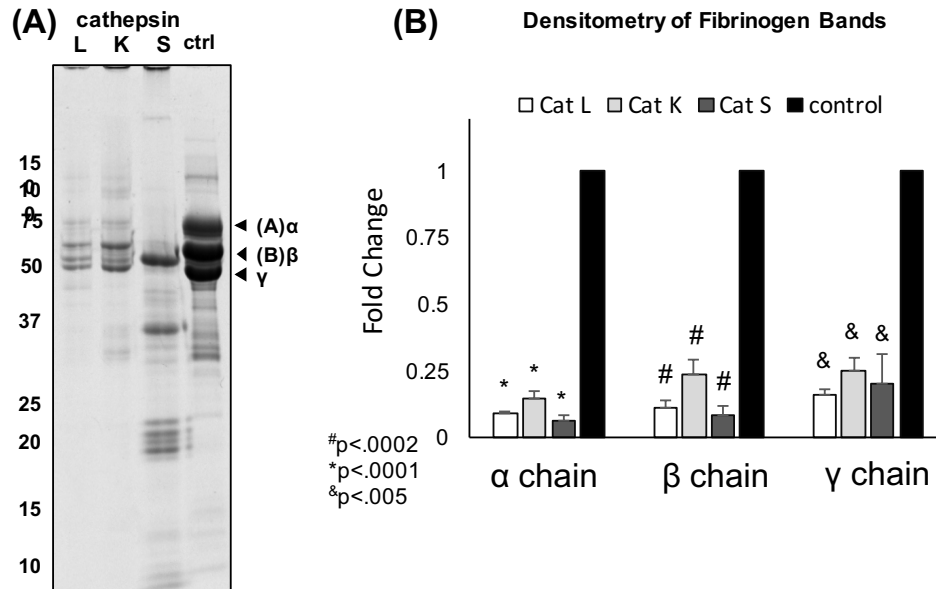
Figure 4-7. Experimental validation of cleavage of fibrinogen by cathepsins K, L, and S. *Fibrinogen was co-incubated with cathepsins K, L, and S for 24 hours at 37°C. Samples were resolved in a reduced 10% SDS-PAGE (A). Cathepsins K and L cleave fibrinogen in a similar manner to thrombin. However, cathepsin S degrades most of the Aα and Bβ polypeptide chains of fibrinogen releasing lower molecular weight fragments. (B) Densitometry was used to quantify the 64kDa A(α) chain band, 56kDa B(β) chain band, and 47kDa γ chain band and revealed that all groups (fibrin, Cat L, Cat K, and Cat S) were statistically significantly less compared to fibrinogen (pH 6). #p<.0002, *p<.0001, &p<.005, n=3.*

4.3.6 Cathepsins K, L, and S cleave all three fibrinogen chains

To experimentally confirm if cathepsins K, L, or S could cleave fibrinogen, equal amounts of fibrinogen was incubated separately with cathepsins K, L, or S for 24 hours at 37°C (Figure 4-7A) followed by reducing SDS-PAGE. α, β, and γ bands are visible in the control lane and quantification of the amount of fibrinogen remaining after incubation was done with densitometry and normalized to that of the no enzyme control. There was a significant decrease (p<0.005) after fibrinogen was co-incubated with each cathepsin, but

the banding pattern of degradation was different for each enzyme. The most reduction occurring in the α chain by all three cathepsins (CatK α: 0.15 ± 0.03-fold, CatL α: 0.09 ± 0.01-fold, CatS α: 0.06 ± 0.02-fold) (Figure 4-7B). Additionally, cathepsin S had the greatest reduction in all three fibrinogen chains (α: 0.06 ± 0.02-fold, β: 0.09 ± 0.03-fold,γ: 0.20 ± 0.11-fold) compared to cathepsins K and L. Cathepsin L showed the second most fibrinogenolysis activity (α: 0.09 ± 0.01-fold, β: 0.11 ± 0.03-fold, γ: 0.16 ± 0.02-fold). The γ chain had the most amount intact after co-incubation with cathepsins and cathepsin L hydrolyzed the γ chain the most (CatK γ: 0.25 ± 0.05-fold, CatL γ: 0.16 ± 0.02-fold, CatS γ: 0.20 ± 0.11-fold). Cathepsin K showed the least amount of hydrolyzed fibrinogen overall (α: 0.15 ± 0.03-fold, β: 0.23 ± 0.06-fold, γ: 0.25 ± 0.05-fold).

4.4   Discussion

Human cysteine cathepsins have experimentally been shown to degrade fibrinogen (Figure 4-7) and cathepsins K, L, and S appear to do so differently, generating unique cleavage profiles. We computationally assessed where fibrinogen cleavage by cathepsin K, L, and S could be occurring by combining molecular docking and putative bioinformatics cleavage site predictions as two unbiased approaches to objectively determine putative sites based on electrostatics and defined amino acids that were preferred sites fitting into the protease active site cleft. We used molecular modeling to study possible electrostatic interactions between fibrinogen and cathepsins. We visualized the binding interactions of each cathepsin on each fibrinogen chain as well as predicted putative cleavage sites based on the known cathepsin specificities. Comparing the proposed binding interactions with

predicted cleavage sites allowed for the discovery of alternative mechanisms of cathepsin-mediated fibrinogen degradation.

Much like fibrinogenolytic enzymes thrombin and plasmin, cysteine cathepsins studied here showed a preference for binding to the E-domain and the helical region (Figure 4-1A, Figure 4-3A, Figure 4-5A), the same region where hydrolytic events causing fibrin formation (thrombin) and fibrin degradation (plasmin) occurs. The binding conformation was such that the active site of the cathepsin was facing the fibrinogen subunit, indicating that the active site had a high affinity for this region. We were limited by the crystal structure of fibrinogen, which does not include portions of the fibrinogen subunit due to either resolution or flexibility, however the molecular docking proposed binding sites in the range of the amino acid sequence visualized by the crystal structure were in high agreement with top ranked predicted cleavage sites from PACMANS. The preference for the E-domain, indicated by high PACMANS scores and docking models, provides specific amino acid locations for cathepsin binding and fibrinogen cleavage. The loops that compose the E-domain indicate a highly flexible region prone to cleavage and cathepsins may bind here and initiate cleavage on its composed chains. The β chain, which forms much of the globular region of the D-domain, has a binding site directly adjacent to this portion of the protein, which was also proximal to the region where crosslinking occurs for fibrin polymerization [172, 173], as well as a known cleavage site of plasmin cleavage.

The binding regions on the α, β, and γ chains of the fibrinogen subunit showed a high correlation with the predicted cleavage sites on fibrinogen by PACMANS. Taking all of this together, the top PACMANS predicted and molecular docking fibrinogen cleavage

59

sites by cathepsin highlight the redundancy in substrate specificities of cathepsins with two α chain sequences, EILR/GDFS (catK score 0.91, catL score 0.82, catS score 0.81) and EDLR/SRIE (catK score 0.81, catL score 0.75, catS score 0.78), and one β chain sequence, SILE/NLRS (catK score 0.83, catL score 0.71, catS score 0.75), that were highly ranked cleavage sites by all three cathepsins and molecular docking models showed these as possible interactions. The redundancy in hydrolysis sites by cathepsins K, L, and S mirrors the redundancy also seen with tradition substrates [1], and the conserved nature of their active sites with the rest of protein having 60% sequence identity with this substrate and cathepsins. Cathepsins K and L were predicted to hydrolyze the fibrinogen γ chain more similarly than cathepsin S with both having a cleavage sequence, IQLT/YNPD, a few amino acids away from a known plasmin cleavage of fibrinogen. Cathepsin K had thirteen predicted cleavage sites that also had supporting molecular docking models near those sequences, while cathepsin S had twelve sequences and cathepsin L had nine sequences. Cathepsin K also had a much greater potential for cleaving multiple locations in the α chain, with five cleavages sequences with corroborating molecular docking models compared to the two cleavage sequences for cathepsins L and S. Likewise, cathepsin S had the greatest potential for cleaving in the β chain, with seven cleavage sequences with supporting molecular docking models compared to the four cleavage sequences for cathepsins K and L. Cathepsins S and L have a number of overlapping docking models that indicate cleaving in the β chain EAVS/QTSS (cathepsin S score 0.84, cathepsin L score 0.74), ETVN/SNIP (cathepsin L score 0.80, cathepsin S score 0.78), and RVLR/SILE

(cathepsin S score 0.77, cathepsin L score 0.76) and in the γ chain at SILT/HDSS (cathepsin L score 0.80, cathepsin S score 0.79). Also, it is interesting to note that although there was far more congruence in the number of β chain cleavage sequences (4), there was only one sequence found to be a top ranked PACMANS sequence for hydrolysis and had molecular docking models supporting binding at that location for cathepsins K, L, and S, whereas in the α chain, the overlap was for two cleavage sequences, which were both highly ranked cleavage locations for all three cathepsins.

The docking models and predicted cleavage sites provide computational evidence of where cathepsins K, L, and S can bind and cleave fibrinogen. Further study of cathepsins bound to individual fibrinogen chains illustrated mechanisms of intra-molecular stabilization based on the cathepsins binding to cleavage sites predicted by PACMANS (Figure 4-2, Figure 4-4, Figure 4-6). Whether the scissile bond involved a flexible or rigid side chain, the presence of nearby polar residues capable of molecular bonding portrayed fibrinogen as an easily accessible cleavage target for these cysteine cathepsins. While these models deal with rigid structures, the polarity of these residues supports the hypothesis that these residues and cleavage sites would be readily accessible for cleavage in a cellular environment. Indeed, cathepsin expression by endothelial cells, fibroblasts, osteoclasts, and immune cells degrade intracellular proteins and complicated insoluble macromolecules like type I collagen and elastin.

Defining cysteine cathepsin-mediated fibrinolysis has many benefits. As novel fibrin-based tissue engineering constructs are being used exposing fibrin to novel cellular environments, it is important that researchers understand how these synthetic environments

can be remodeled. We have recently demonstrated upregulation of cysteine cathepsins during the lifetime of a self-locomoting biological machine [128]. Further experimentation is needed to confirm the binding and cleavage sites identified by our computational techniques, however, this report, experimentally confirming cysteine cathepsin K, L, and S degradation of fibrinogen provides molecular docking evidence of its feasibility and stabilization of the intermediates and final products. Understanding this mechanism is important for regenerative medicine applications, especially when using fibrinogen polymerized gels as biocompatible, natural substrates and constructs implanted with cells, to prevent or design degradation of fibrin networks.

CHAPTER 5.   INVESTIGATING THE LIFE EXPECTANCY AND PROTEOLYTIC DEGRADATION OF ENGINEERED SKELETAL MUSCLE BIOLOGICAL MACHINES

5.1   Introduction

The modular and scalable architecture of skeletal muscle tissue makes it an ideal power supply for producing force and locomotion – on a range of forces and time scales – in engineered living systems and biological actuators. Skeletal muscle is the primary actuating source for mammals, which are evolutionarily considered to be the highest life forms and capable of many complex behaviors and high-level functionalities [174-176]. A combination of techniques from the fields of 3D printing, tissue engineering, and biomaterials has recently yielded a new class of millimeter-to-centimeter scale skeletal muscle-powered biological robots ('bio-bots') capable of dynamic and adaptive responses, reliable force production, and untethered locomotion upon applied electrical [126] or optical [127, 177] signals. This forward-engineered biomaterial-muscle platform is ideal for introducing different cell types and biomaterials, and permits control over physical, mechanical, biological, and biochemical cues [175].

In vivo, skeletal muscle is supported by a basal lamina containing layers of connective tissue that provide a barrier from physical stress and damage [178, 179]. A combination of extracellular matrix (ECM) proteins, including the natural hydrogel of fibrin, acts as a bioinspired basement membrane mimic that provides cues and physical support to the cells in the engineered solid muscle strip of the bio-bots [126]. Though synthetic polymers have been used in the fabrication of some skeletal muscle constructs, the viscoelastic mechanical properties of natural proteins, such as fibrin, can be tuned to those of native connective tissue or skeletal muscle, rendering it a useful scaffold material for engineering a myriad of tissue types [180, 181]. However, maintaining consistent results becomes challenging when sustaining a living cellular system *in vitro* for weeks or months. Skeletal muscle must be preserved in a differentiated state at environmental conditions [175]. Moreover, the system is subject to degradation by cell-secreted proteases that can break down the mechanical integrity of the ECM in tissue and thus eventually lead to device failure, despite the presence of some protease inhibitors [126].

The polymerization of fibrin into a 3D branching network is actively and equally opposed in a physiological equilibrium by fibrinolysis; despite cross-linking, the protein is extremely susceptible to degradation by cell-secreted enzymatic proteases [182, 183] (Figure 5-1). Plasmin is a serine protease secreted by many different cells to rapidly degrade fibrin [184, 185]. In skeletal muscle, inhibition of plasmin can result in fibrin build up and reduced myoblast fusion, differentiation, and regeneration [186]. Furthermore, other proteases secreted during myoblast differentiation, including cysteine cathepsins and

64
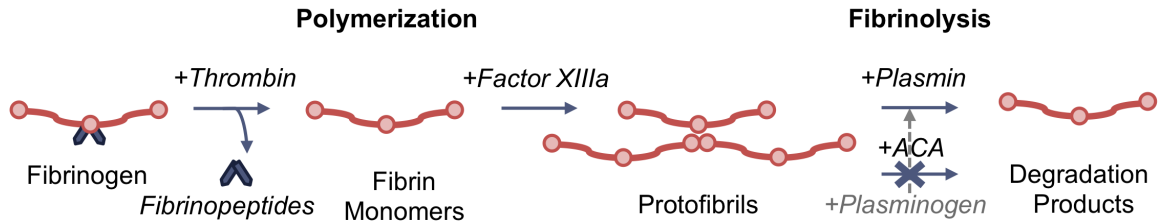
Figure 5-1. Polymerization and Degradation of Fibrin. *Fibrinogen is cleaved by thrombin to yield fibrin monomers that can be polymerized into protofibrils. The addition of the protease plasmin can lead to hydrolysis of the fibrin branching network, yielding degradation products. However, the anti-fibrinolytic inhibitor aminocaproic acid (ACA) can be added to prevent plasmin from becoming its active form. ACA binds to the lysine binding sites present on both plasmin and its activating agent plasminogen, which can be produced by C2C12s. ACA prevents plasmin from being activated and binding to (and subsequently degrading) fibrin by competitive and stoichiometric inhibition.*

matrix metalloproteinases (MMPs) must also be considered as contributors to fibrin network destabilization [187-189]. Cysteine cathepsins are powerful lysosomal proteases that degrade intracellular and extracellular matrix proteins and include the most powerful mammalian collagenase and elastase [1, 3, 5, 21, 130]; their upregulation during myogenic differentiation has been documented [187]. MMPs are another group of ECM-degrading enzymes involved in remodeling and maintenance of the ECM in both normal and pathological states [188, 189]. Specifically, MMP-2 and MMP-9 have been seen to be involved with skeletal muscle ECM remodeling in muscle satellite cell migration and differentiation, and are differentially regulated by cytokines and growth factors during these processes. While MMP-9 has been implicated in pre-fusion migration and in the early stages of regeneration, MMP-2 activation occurs at later stages of myofiber regeneration [188-191].

Here we examine the life expectancy and breakdown of engineered skeletal muscle bio-bots by loss of form and function, as a result of degradation by three major classes of proteases: plasmin, cathepsin L (CatL), and MMPs (including MMP-2 and MMP-9). We first examined the effect of adding a serine protease inhibitor, aminocaproic acid (ACA), on the maintenance of machine life expectancy and structural integrity. We also demonstrate the use of gelatin zymography, a substrate gel electrophoresis technique, to determine the effects of differentiation time, inhibitor concentration, and electrical stimulation on temporal cysteine cathepsin and MMP expression within muscle strips. An experimental window under two weeks allowed for an increased understanding of the system on a shorter time scale. Finally, we demonstrate the ability to measure CatL and MMP activity while modifying certain design parameters involved in bio-bot fabrication. With this understanding of proteases secreted by these cells and concomitant matrix degradation, we are poised to design the next generation of machines with controllable gain and loss of function. Our studies can also have important applications in skeletal muscle based tissue engineering, muscles-on-a-chip, and related fields of studies.

5.2 Materials and Methods

5.2.1 Design and 3D Printing of Bio-Bot Skeleton

Bio-bot skeletons and holders were fabricated using a commercial Stereolithography apparatus (SLA 250/50, 3D Systems) [126]. Briefly, parts were designed in CAD software (AutoCAD), exported in .STL format for slicing into layers (3D Lightyear software, v1.4, 3D Systems), and built in a layer-by-layer fashion on a SLA stage that was

modified as previously described [192, 193]. To print hydrogel beams with varying stiffness profiles, the polymerization energy of the SLA was changed by controlling two parameters according to a characterized Working Curve Equation [194, 195]. Bio-bot beam skeletons were polymerized with energy doses that ranged from 108.7 to 233.3 mJ cm^{-2} in order to obtain elastic moduli of 214 to 489 kPa, as detailed earlier [126]. The liquid hydrogel formulation was not altered. A skeleton beam of 319 kPa was used unless otherwise noted.

Liquid pre-polymer hydrogel solutions were composed of bio-inert and biocompatible 20% poly(ethylene glycol) diacrylate of M_W 700 g mol^{-1} (PEGDA 700, Sigma-Aldrich) or poly(ethylene glycol) dimethacrylate of M_W 1000 g mol^{-1} (PEGDMA 1000, Polysciences, Inc.) [126, 177]. After fabrication, skeletons and holders were rinsed in PBS, sterilized in 70% EtOH, and stored in sterile PBS until use.

5.2.2   Cell Culture and Seeding

C2C12 murine myoblasts were proliferated in growth medium (GM) and passaged before confluence. For optogenetic bio-bots, ChR2-C2C12s were used, as previously described [177, 196]. GM consisted of Dulbecco's Modified Eagle Medium (DMEM, Corning Cellgro) with 10% (v/v) fetal bovine serum (FBS, VWR) and 1% (v/v) each of L-glutamine and penicillin-streptomycin (both Corning Cellgro). For the seeding of muscle strips, bio-bot skeletons were placed in polymerized holders and aspirated of excess liquid. 5 x 10^6 cells ml^{-1} (final concentration of myoblasts in cell-gel solution, unless otherwise noted) were resuspended in GM and mixed with ice-cold Matrigel^{TM} basement membrane

67

(30% of total volume, Corning), fibrinogen (4 mg ml^{-1}, Sigma-Aldrich), and thrombin from bovine plasma (0.5 U mg^{-1} fibrinogen, Sigma-Aldrich). The cell-gel solution was added to each holder (day 0) and incubated for 1.5 hr before adding warm GM. All cells and cultures were maintained at 37 ºC and 5% CO_2.
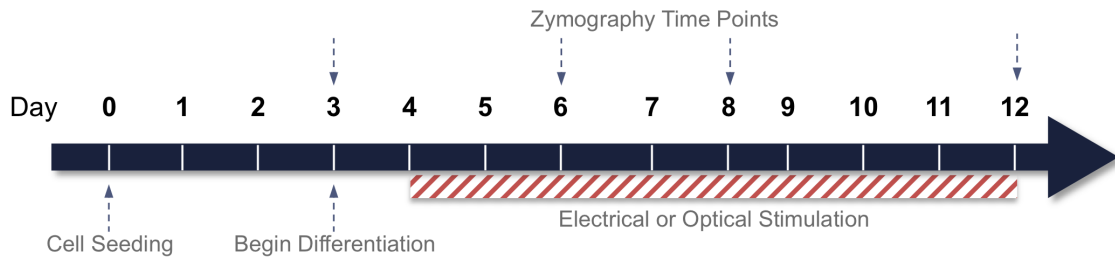
Figure 5-2. Timeline of Bio-Bot Development and Analysis. *Skeletal muscle bio-bots were fabricated and seeded on day 0. On day 3, DM was added to induce differentiation of skeletal muscle. Electrical or optical stimulation was applied on days 4-12, as noted. Samples were snap frozen for gelatin zymography on days 3, 6, 8, and 12.*

5.2.3   Skeletal Muscle Differentiation and Stimulation

After 3 days in GM, bio-bots were released from the holders and switched to differentiation media (DM), consisting of DMEM with 10% (v/v) heat-inactivated horse serum (HS, Gibco), 1% (v/v) each of L-glutamine and penicillin-streptomycin, 50 ng ml^{-1} of insulin-like growth factor (IGF-1, Sigma-Aldrich), and ACA as noted: 0 mg ml^{-1} (0x ACA), 1 mg ml^{-1} (1x ACA), or 3 mg ml^{-1} (3x ACA). Media was changed daily. Electrical stimulation was applied using a custom-built setup [126, 197] starting on day 4 (Figure 5-2). Bio-bots were placed in warm DMEM between two sterilized Pt electrodes and a current was applied perpendicular to the longitudinal axis of the muscle strip with the following parameters: 20 V amplitude, 50 ms pulse width, 1 Hz frequency, and 10 min

68

total stimulation per day. Optical stimulation of ChR2-C2C12 optogenetic bio-bots was applied as previously described [177], starting on day 4, at 1 Hz frequency for 10 min total stimulation per day.

5.2.4 Force and Diameter Measurements

Images of the bio-bots were taken using a stereomicroscope (MZ FL III, Leica Microsystems) with a digital camera (Flex, SPOT Imaging Solutions). The Measure Tool in ImageJ software (National Institutes of Health) was used to measure muscle strip dimensions from top-view (diameter) and side-view (passive tension) images. The passive tension force of the muscle strip was calculated using an equation derived from Euler-Bernoulli linear beam theory: $F_p = (8EI/lL^2)\delta_{max}$, where E is the elastic modulus (319 kPa [126], unless otherwise noted), I is the moment of inertia of the bio-bot beam, L is the beam length, l is the moment arm between the beam and muscle strip, and δ_{max} is the beam deflection.

5.2.5 Viability Assays

Cell viability was quantitatively determined using a metabolic colorimetric assay by incubating muscle strips in CellTiter 96 AQueous One Solution (MTS, Promega) and DMEM without phenol red (LifeTechnologies) in a 5:1 (v/v) ratio in the dark at 37 ºC. After 4 h, absorbance was measured at 490 nm using a microplate reader (Synergy HT, BioTek). The absorbance of a blank sample without cells was subtracted from each reading, and results were normalized to day 3.

69

5.2.6   Immunofluorescence and Histology

Muscle strips were removed from bio-bot skeletons and rinsed in PBS. For fluorescent staining, tissues were fixed in 4% paraformaldehyde (Electron Microscopy Services) for 30 min and permeabilized with 0.25% Triton X-100 (Sigma-Aldrich) for 10 min. After blocking in Image-iT® FX (Molecular Probes) overnight at 4ºC, muscle strips were incubated with MF-20 (1:400, Developmental Studies Hybridoma Bank, The University of Iowa) and anti-sarcomeric α-actinin primary antibodies (1:400, Abcam) overnight at 4ºC, rinsed 3x with PBS, and incubated with Alexa Fluor® 488 goat anti-mouse and Alexa Fluor® 568 goat anti-rabbit (both 1:400, ThermoFisher) overnight at 4ºC in the dark. Muscle strips were rinsed 3x with PBS, incubated with 4',6-diamindino-2-phenylindole (DAPI, 1:5,000 in sterile DI water, Sigma-Aldrich) for 10 min, rinsed, and imaged with a confocal microscope (LSM 710, Zeiss). For histology, frozen muscle strips were embedded in OCT compound (TissueTek), cut into 14 μm sections using a cryostat (CM3050, Leica), mounted on glass slides, stained with a Masson's Trichrome kit (Polysciences, Inc.), and imaged using a digital pathology slide scanner (C9600, NanoZoomer).

5.2.7   Muscle Strip Nucleic Acid Extraction and Assay

Muscle strips were removed from bio-bot skeletons, snap frozen in liquid nitrogen, and stored at -80 ºC until use. Before RNA isolation, tissues were lysed using a rotor-stator homogenizer (TissueRuptor, Qiagen). Total RNA was then extracted using an RNeasy Mini Kit (Qiagen). Alternatively, DNeasy Blood and Tissue Kit (Qiagen) was used to

isolate total genomic DNA from muscle strips. The concentration of DNA or RNA was measured at 260 nm with a spectrophotometer (NanoDrop 1000).

5.2.8   Muscle Strip Protein Extraction and Assay

Muscle strips were snap frozen in liquid nitrogen and stored at -80 ºC until use. Upon thawing, muscle strip cells were digested in lysis buffer with freshly added 0.1mM leupeptin (Calbiochem) for zymography and Western blotting, or RIPA buffer (Thermo Scientific) for total protein measurements and muscle creatine kinase (MCK) assays. Lysis buffer consisted of 20 nM Tris–HCl at pH 7.5, 5 mM ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid (Sigma-Aldrich), 20 mM β-glycerolphosphate (Alfa Aesar), 150 mM NaCl (BDH), 1 mM sodium orthovanadate (Sigma-Aldrich), 10 mM NaF (Sigma-Aldrich), 1% Triton X-100 (EMD Chemicals), and 0.1% Tween-20 (Fisher Scientific) [135]. After sonication and centrifugation, the protein extract supernatant was collected. Using a Pierce BCA Protein Assay Kit (Thermo Scientific), total muscle strip protein content was measured at 562 nm with a spectrophotometer. Extracted protein supernatant was used to determine MCK production using a Liquid Creatine Kinase Reagent Set (Pointe Scientific), and absorbance was measured at 340 nm with a microplate reader.

5.2.9   Cathepsin and MMP Zymography

Protocols for multiplex zymography to detect mature cathepsins K, L, S, and V have been previously optimized [134, 135]. MMP zymography has also been optimized for

detection of pro- and mature forms of MMP-2 and MMP-9 [198]. For gelatin zymography, SDS-PAGE gels (12.5% for cathepsins, 10% for MMPs) were impregnated with 5 mg ml^{-1} soluble gelatin substrate. Samples were prepared with a non-reducing loading dye (5× − 0.05% bromophenol blue, 10% SDS, 1.5 M Tris, 50% glycerol) and separated by electrophoresis resolved at 200V for 1 h at 4°C. Gels were washed in renaturing buffer and then incubated in assay buffer overnight, specific to proteases being probed. Renaturing buffer contained 65 mM Tris buffer at pH 7.4 with 20% glycerol for cathepsins, and 2.5% Triton-X 100 for MMPs. Assay buffer contained 0.1 M sodium acetate buffer at pH 4.0, 1 mM EDTA, and freshly added 2 mM dithiothreitol for cathepsins, and 50 mM Tris-HCl at pH 7.4, 10 mM CaCl$_2$, 50 mM NaCl, and 0.05% Triton X-100 for MMPs. After overnight incubation, gels were stained for 2 h in 4.5% Coomassie blue stain (Sigma-Aldrich) with 10% acetic acid and 25% isopropanol. Gels were then destained with 10% isopropanol and 10% acetic acid, and imaged with an ImageQuant LAS 4000 (GE Healthcare). Densitometry quantification of white band intensity (indicating active protease) was preformed using ImageJ, and all results were normalized to day 3 (the start of muscle differentiation). In all results, one representative gel image is shown for each experiment.

5.2.10 Western Blots

Muscle strip protein extracts were prepared with loading dye with the additional of β-mercaptoethanol, boiled for 5 min, and loaded into a 12.5% SDS-PAGE gel. Proteins were separated by molecular weight under electrophoresis resolved at 150V and then transferred to a nitrocellulouse membrane using a Trans-Blot SD Semi-Dry Transfer Cell

(Bio-Rad) at 10V for 30 min. Membranes were blocked for at least 1 h in Odyssey Blocking Buffer (LI-COR Bioscience) diluted 1:2 in PBS, then incubated with primary antibodies (1:1000) for goat polyclonal mouse cathepsin L (R&D Systems), or rabbit polyclonal cystatin C (Millipore) overnight at 4°C with constant agitation. Proteins were detected with a LI-COR Odyssey scanner using anti-goat or anti-rabbit secondary antibodies (1:5000 with 0.1% Tween, Invitrogen) tagged with an infrared fluorophore. Densitometry was preformed using ImageJ, and results were normalized to day 3.

5.2.11 *Statistical Analysis*

Results are presented as mean ± standard error of the mean (SEM) as noted in figure captions. Statistical analyses included tests for significance (one-way ANOVA followed by Tukey's Multiple Comparison Test, performed using OriginPro software, or two-tailed t-tests, performed using Microsoft Excel) and survival (Kaplan-Meier analysis, performed using OriginPro software). For muscle strip life expectancy and survival analysis, outliers were calculated using the Interquartile Range Rule. Any data points outside of $1.5 \times$ IQR (interquartile range) from the 1st or 3rd quartiles were omitted.

5.3 **Results**

5.3.1 *Development and Maturation of Skeletal Muscle Bio-Bots*

The freestanding skeletal bio-bot consisted of an engineered muscle strip coupled to a fabricated hydrogel skeleton, or a scaffold consisting of a flexible beam connected by two rigid pillars – a design inspired by the musculoskeletal physiology of tendon and bone

in which the force from a contractile muscle is transmitted to a bone through an adjoining tendon[4]. We used stereolithographic 3D printing, a rapid manufacturing technique that prints a structure with customizable geometric and mechanical parameters in a layer-by-layer fashion [194, 195, 199-201], to polymerize the bio-bot skeleton from a synthetic biocompatible hydrogel.

C2C12 myoblasts at a density of 5 x 10^6 cells ml^{-1} were mixed with ECM proteins including fibrinogen, thrombin, and Matrigel^{TM}; the liquid suspension was added to a 3D printed holder and began to cross-link into a gel. The cells applied traction forces to compact the tissue into a solid muscle strip around the pillars of the hydrogel skeleton. After three days in proliferation media, the bio-bot (consisting of the hydrogel skeleton and attached muscle strip) was released as a freestanding biological machine. Myoblasts were then induced to differentiate with the addition of horse serum and insulin-like growth factor (IGF-1) [202]. Striated, multinucleated myotubes were visible throughout the muscle strip (Figure 5-3A).

Over the course of the experiment, the total DNA present in the muscle strips increased over ten-fold, from 2.1 ± 0.8 on day 3 to 23.8 ± 3.4 µg on day 12. The total protein content per muscle strip increased as well, and was significantly amplified on day 12 (356.8 ± 33.5 µg) compared to day 3 (235.7 ± 44.7 µg, Figure 5-3B). The activity of muscle creatine kinase (MCK), which is expressed in mature skeletal muscle only after differentiation [203] and thus served as a quantitative gauge of myogenesis, was significantly increased as early as day 6 and reached a maximum output of 560.4 ± 44 U L^{-1} on day 12 compared to 4.2 ± 0.2 U L^{-1} on day 3 (C).

74
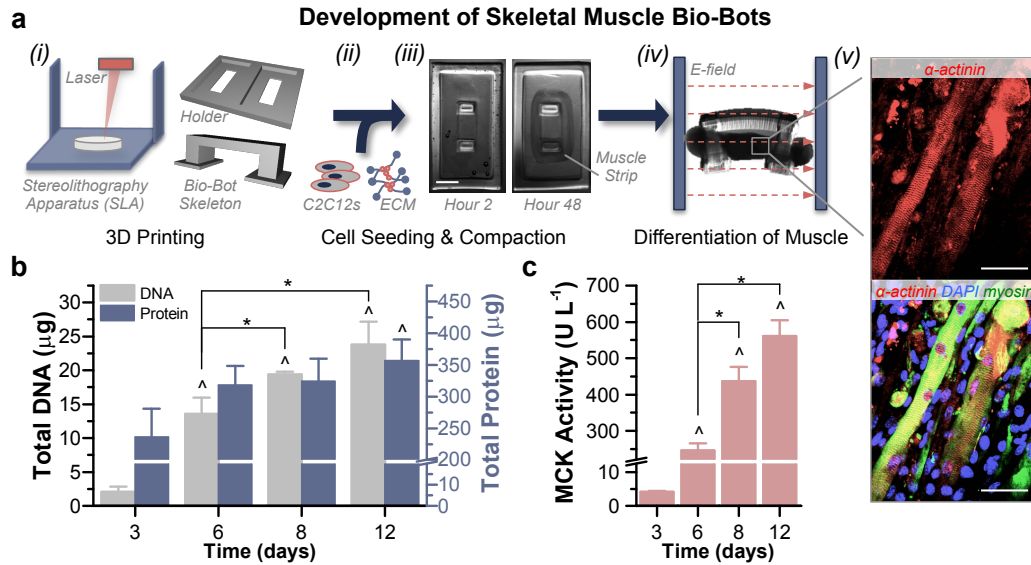
Figure 5-3. Development and Differentiation of Skeletal Muscle Bio-Bots. *(A) Modular bio-bots were assembled and differentiated in a stepwise manner. (i) A Stereolithography apparatus (SLA) was used to 3D print a millimeter-scale hydrogel skeleton and holder. (ii) C2C12s were mixed with a liquid solution of ECM proteins that included fibrinogen, thrombin, and Matrigel. (iii) When added to the holder, the cell-gel solution compacted to form a solid muscle strip. Scale bar, 2 mm. (iv) The freestanding bio-bot (consisting of the muscle strip coupled to the hydrogel skeleton) could be released from the holder and subjected to electrical or optical stimulation. (v) Immunostaining revealed the presence of striated and multinucleated myotubes (α-actinin, red; MF-20 myosin, green; DAPI nuclear stain, blue). Scale bar, 50 μm. (B) Total DNA and protein levels increased over the time course of the experiment (n = 3-4 muscle strips per time point). There was no statistically significant difference in protein/DNA ratios between days 8 (17 ± 2.8 μg protein μg DNA^{-1}) and 12 (15.7 ± 2.5 μg protein μg DNA^{-1}). (C) Muscle creatine kinase (MCK) activity was significantly increased as early as day 6, and reached a maximum output on day 12 (n = 3 muscle strips per time point). The rate of increased MCK activity slowed between days 8 and 12, confirming that this was a relevant stopping point for the experiments. All plots represent mean ± SEM. * indicates significance (p < 0.05) between conditions at the same time point; ^ indicates significance compared to initial time point.*

5.3.2   Cathepsin and MMP Zymography without ACA

Multiplex zymography was used to examine the activity of cell-secreted or cell-associated cysteine cathepsins and matrix metalloproteinases (MMPs) in muscle strips at

various time points (Figure 5-4A). Polyacrylamide gels were impregnated with gelatin to create a substrate that could be degraded by active proteases. Muscle strips were lysed and loaded for zymography. After protein separation by electrophoresis, gels were washed in a renaturing buffer that allowed proteins to re-fold into their native conformation, and then incubated in assay buffer overnight for optimal enzyme activity to degrade the substrate. Both renaturing and assay buffers were specific to the proteases examined. Gels were stained with Coomassie Blue stain, which stains all proteins, and then destained; areas of white bands indicated proteolytic activity. On the cathepsin zymogram, bands were identified for Cathepsin L, both unbound (CatL, 25-37 kDa) and bound to matrix proteins in the muscle strip tissue (CatL+tissue, 75 kDa; Figure 5-4B). On the MMP zymogram, bands were identified at 62-72 and 82-92 kDa, pertaining to MMP-2/pro-MMP-2 and MMP-9/pro-MMP-9, respectively (Figure 5-4C).

In the absence of any anti-fibrinolytic treatment (0x ACA), muscle strips produced active cathepsins and MMPs, and the amount of active enzyme increased over the course of 12 days. The activity of each protease was quantified by densitometry and normalized to the beginning of skeletal muscle differentiation on day 3. On day 12, the amount of active CatL displayed a significant (3.9 ± 1.1-fold) change, and amount of active CatL+tissue increased by 2.8 ± 0.9-fold. Electrical stimulation of muscle strips significantly increased the expression of both active CatL and CatL+tissue (5.5 ± 1.1 and 3.5 ± 0.7-fold compared to day 3, respectively; Figure 5-4B). Additionally, the amount of active MMP-2 and MMP-9 in muscle strips both increased on day 12, both with (1.2 ± 0.2

76
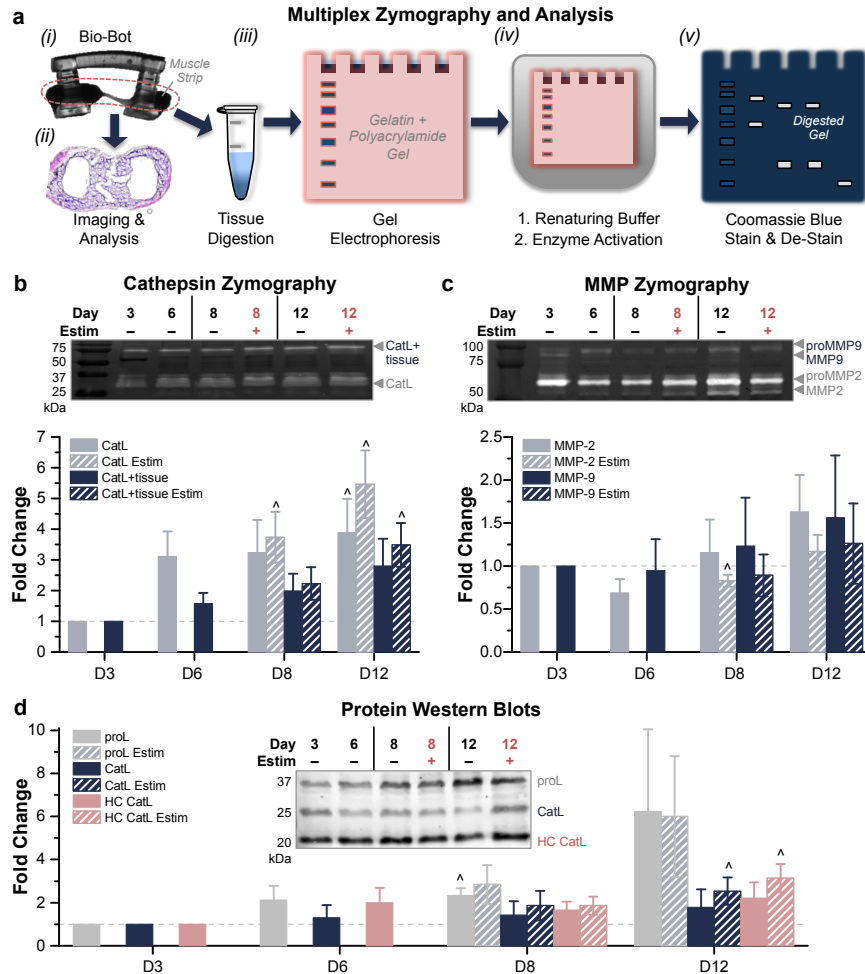
Figure 5-4. Gelatin Zymography of Cathepsins and MMPs without ACA. *(A) Muscle strip zymography process flow. (i) Degrading muscle strips were removed from hydrogel skeletons. (ii) Muscle strips could be fixed and stained for further analysis. (iii) Tissues were digested in lysis buffer. Isolated proteins were run through gel electrophoresis. (iv) Gels were incubated in renaturing and assay buffers. (v) Gels were stained with Coomassie Blue; bright bands indicated the activity of specific proteases, which were separated by molecular weight. (B) Cathepsin zymography identified the activity of CatL and CatL+tissue at days 3, 6, 8, and 12, in the presence or absence of electrical stimulation (n = 3-5 muscle strips per condition). (C) MMP zymography identified the activity of MMP-2 and MMP-9 at days 3, 6, 8, and 12, in the presence or absence of electrical stimulation (n = 3-5 muscle strips per condition). For MMP zymograms, upper bands reflect inactive enzyme (proMMP) and lower bands reflect active enzyme. (D) Western blotting confirmed the presence of cathepsins and related proteins, such as CysC (n = 3-5 muscle strips per condition). All plots represent mean ± SEM. ^ indicates significance (p < 0.05) compared to initial time point.*

and 1.3 ± 0.5-fold) and without electrical stimulation (1.6 ± 0.4 and 1.6 ± 0.7-fold; Figure 5-4C). Western blotting confirmed the identity of two mature isoforms of the cysteine protease CatL (including heavy chain [204] CatL), as well as the presence of cystatin C (CysC), an endogenous cysteine protease inhibitor. Pro-forms of both MMPs and CatL were also present (Figure 5-4D).

5.3.3 Muscle Strip Life Expectancy

Breakdown of the muscle strip was defined as detachment from the hydrogel skeleton (Figure 5-4A), rendering the bio-bot non-functional. In the absence of any anti-fibrinolytic treatment (0x ACA), muscle strips demonstrated an average life expectancy of
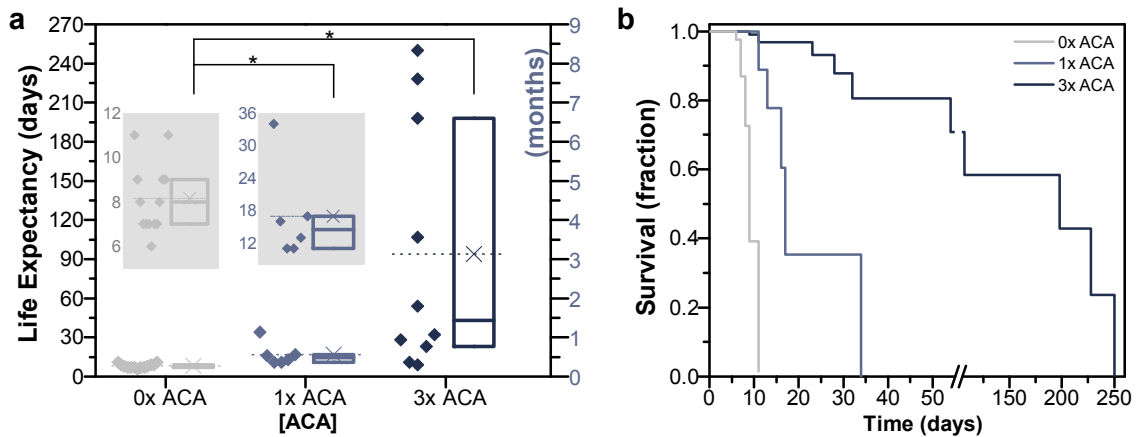
Figure 5-5. Bio-Bot Life Expectancy. *(A) In the absence of any anti-fibrinolytic treatment (0x ACA), muscle strips demonstrated an average life expectancy of 8.2 ± 0.5 days until rupture. However, incubation in medium with the serine protease inhibitor ACA lengthened the life expectancy of the muscle strips. Box plots represent 25th, 50th, and 75th percentiles, with average values marked as (x) and whiskers representing $\pm$ SEM. Data are presented to the left of the boxes (n = 6-13 muscle strips per condition, excluding outliers). * indicates significance (p < 0.05) between time points. (B) Kaplan-Meier survival analysis provided an additional method of comparison by plotting survival fraction of each treatment group as a function of time.*

8.2 ± 0.5 days until rupture (Figure 5-5A). However, incubation of the serine protease inhibitor ACA in medium lengthened the life expectancy of the muscle strips; addition of 1 mg ml^{-1} (1x) and 3 mg ml^{-1} (3x) ACA significantly increased the averages to 17 ± 3.9 and 94 ± 31.9 days until rupture, respectively. Kaplan-Meier survival analysis provided an additional method of comparison between treatment groups (Figure 5-5B). For those muscle strips which remained unruptured on the order of months, passive tension remained statistically unchanged on day 140 (4.6 months) compared to early values; furthermore, some muscle strips treated with 3x ACA continued to contract spontaneously and in response to electrical stimulation after 210 days (7 months). Finally, increasing the concentration of ACA did not significantly alter cellular viability in the muscle strips (126.6 ± 15% for 1x ACA and 139 ± 7.8% for 3x ACA on day 12 normalized to day 3).

5.3.4 Loss of Tissue Structure and Mechanical Integrity

Over the time course of the experiment, muscle strips began to degrade (most notably, in the middle region of the tissue) in the absence of anti-fibrinolytic treatment (Figure 5-6A). Five days after cell seeding, 0x ACA muscle strips displayed cross-sectional diameters already significantly reduced from those treated with the inhibitor (2.1 ± 0.08 versus 2.4 ± 0.06 mm for 3x ACA). On day 12, the diameters of all groups were significantly reduced (a range of 73.1 ± 6.8 to 78.4 ± 7% from their day 5 values). The 0x ACA muscle strips displayed a lower passive tension and faster decrease in static force over time compared to both 1x and 3x ACA treated muscle strips. As early as day 8, the passive tension was significantly lower for 0x ACA; by day 12, it had dropped to 815.5 ±
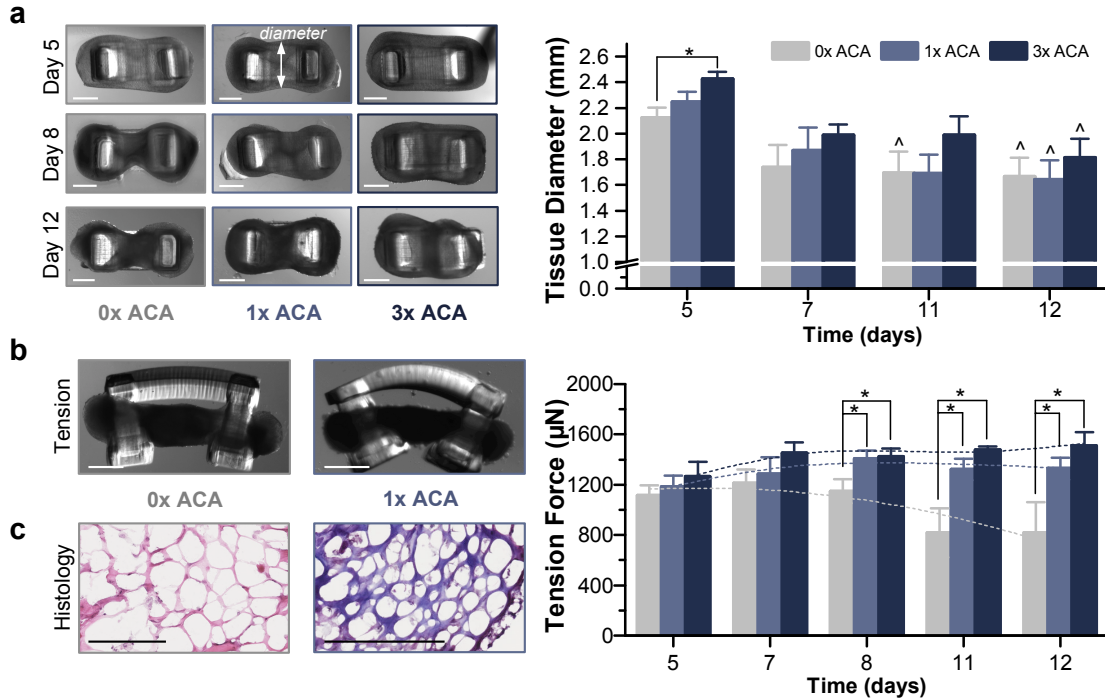
Figure 5-6. Loss of Tissue Structure and Mechanical Integrity. *(A) Muscle strips began to degrade (most notably, in the middle region of the tissue), as shown in top-view images. Though the diameters of all groups on day 12 were reduced, the addition of 3x ACA helped to maintain the integrity of the tissue longer (n = 2-9 muscle strips per condition). Scale bar, 1 mm. (B) Muscle strips cultured without ACA displayed a lower passive tension and faster decrease in static force over time. After day 8, the passive tension was significantly lower for 0x muscle strips (n = 4-7 per time point) compared to both 1x (n = 5-6 per time point) and 3x ACA (n = 5-7 per time point). Scale bar, 1 mm. All plots represent mean ± SEM. * indicates significance (p < 0.05) between conditions at the same time point; ^ indicates significance compared to initial time point. (C) Histological staining for Masson's Trichrome indicated that the addition of ACA helped to maintain the ECM in muscle strips. Scale bar, 200 μm.*

247.1 μN, compared to 1327.6 ± 86 and 1509.5 ± 106.7 μN for 1x and 3x ACA, respectively (Figure 5-6B). Neither application of daily electrical stimulation nor increasing the concentration of ACA beyond 1x caused any statistical difference in average passive tension during this time span. Additionally, histological staining for Masson's

Trichrome, a connective tissue stain, indicated that the addition of ACA helped to maintain the presence of ECM proteins supporting the differentiated myotubes (Figure 5-6C).

5.3.5   Cathepsin and MMP Zymography with 1x and 3x ACA

In the presence of a serine protease inhibitor, muscle strips cultured with 1x and 3x ACA continued to produce active cathepsins and MMPs. On day 12, the amount of active CatL displayed a 0.9 ± 0.2-fold change with 1x ACA and a 0.7 ± 0.3-fold change with 3x ACA compared to day 3. The amount of active CatL+tissue displayed a 1.2 ± 0.2-fold change with 1x ACA and a 2.3 ± 0.1-fold change with 3x ACA, significantly increased from day 3 (Figure 5-7A). Both the amount of active MMP-2 and MMP-9 increased by day 6, but decreased below day 3 levels on day 8. On day 12, MMP-2 was again upregulated (1.1 ± 0.3-fold and 1.3 ± 0.5-fold change for muscle strips cultured with 1x and 3x ACA), while levels of active MMP-9 continued to decrease significantly compared to day 3 levels (0.5 ± 0.2-fold and 0.5 ± 0.2-fold change for muscle strips cultured with 1x and 3x ACA) (Figure 5-7B).

However, when comparing muscle strips cultured with 0x ACA to those cultured in 1x and 3x ACA after 12 days, we observed a reduction in the amount of active protease that was statistically significant for CatL (0.9 ± 0.2-fold and 0.7 ± 0.3-fold, compared to 3.9 ± 1.1-fold for 0x), but not significant for MMP-2 or MMP-9 (Figure 5-8). Increasing the ACA concentration beyond 1x did not further enhance this reduction in active CatL, and electrical stimulation did not significantly alter the expression of any of the cathepsins or MMPs cultured with ACA. Finally, Western blotting also verified the presence of the
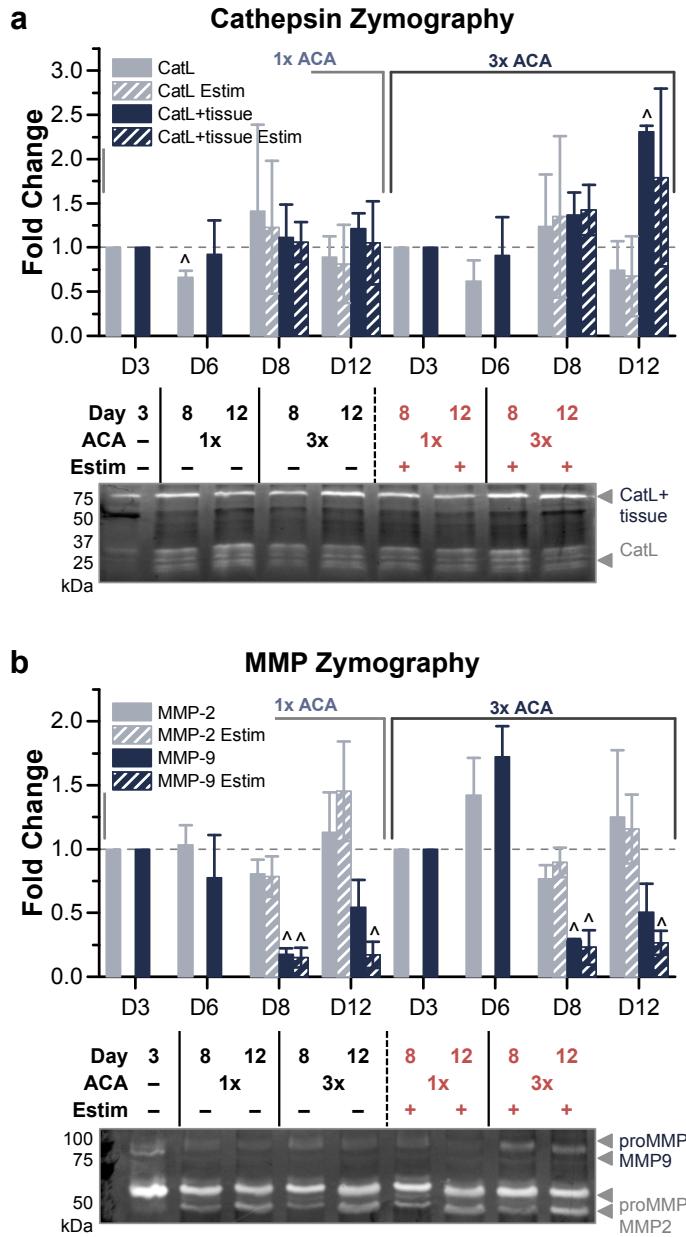
Figure 5-7. Gelatin Zymography of Cathepsins and MMPs with 1x and 3x ACA. *(A) Cathepsin zymography identified the amount of active CatL and CatL+tissue at days 3, 6, 8, and 12, in the presence or absence of electrical stimulation, for 1x and 3x ACA (n = 3-6 muscle strips per condition). (B) MMP zymography identified the amount of active MMP-2 and MMP-9 at days 3, 6, 8, and 12, in the presence or absence of electrical stimulation, for 1x and 3x ACA (n = 4-7 muscle strips per condition). All plots represent mean ± SEM. ^ indicates significance (p < 0.05) compared to initial time point.*
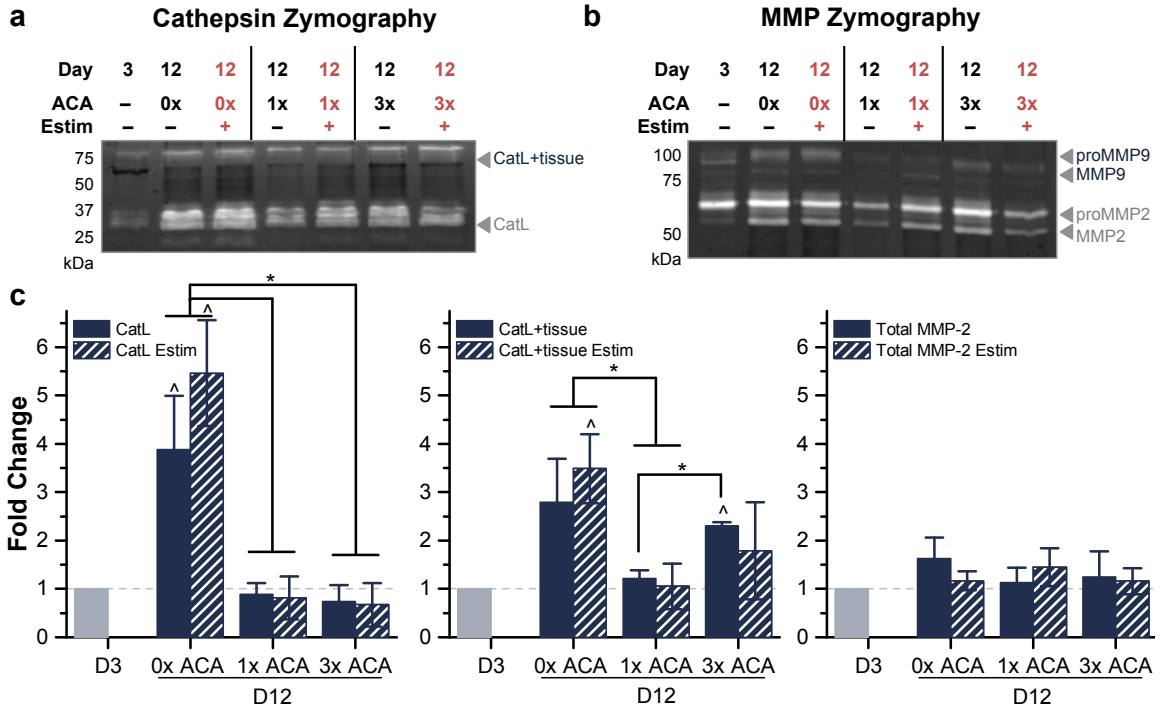
Figure 5-8. Comparison of Active Cathepsins and MMPs in 0x, 1x, and 3x ACA Cultured Muscle Strips. *(A) Cathepsin zymography identified the amount of active CatL and CatL+tissue and (B) MMP zymography identified the amount of active MMP-2, and MMP-9, with and without electrical stimulation, for day 12 conditions compared to the initial time point of day 3. (C) ACA prolonged the lifetime and reduced the amount of active CatL in locomoting biological machines. All plots represent mean ± SEM. * indicates significance (p < 0.05) between conditions at the same time point; ^ indicates significance compared to initial time point.*

pro-form of CatL, as well as the presence of CysC (Figure 5-9).

5.3.6   Varying Hydrogel Skeleton Stiffness and Cell Seeding Density

The hydrogel-muscle platform was modified to introduce tunable variations in the muscle strip as well as the 3D printed skeleton [126]. First, to test the hypothesis that concentration of cells within the muscle strip affected the amount of active proteases produced, muscle strips seeded at a density of 5×10^6 cells ml^{-1} were compared to those

Figure 5-9. Western Blotting of CatL in Muscle Strips with ACA. *(A) Western blotting confirmed the presence of cathepsins and related proteins in muscle strips cultured with 1x and 3x ACA over time. (B) Western blotting of CatL, with and without electrical stimulation, for day 12 conditions compared to the initial time point of day 3 (n = 4-7 muscle strips per condition). All plots represent mean ± SEM. * indicates significance (p < 0.05) between conditions at the same time point; ^ indicates significance compared to initial time point.*

with 2.5 and 10 x 10^6 cells ml^{-1}. On day 12, though a decreased cell density caused a significant increase in CatL, other measurements of active CatL and CatL+tissue showed no significant change compared to muscle strips with 2.5 or 10 x 10^6 cells ml^{-1} cultured in 1x ACA (Figure 5-10A, C). Likewise, MMP-2 and MMP-9 also showed no significant

84
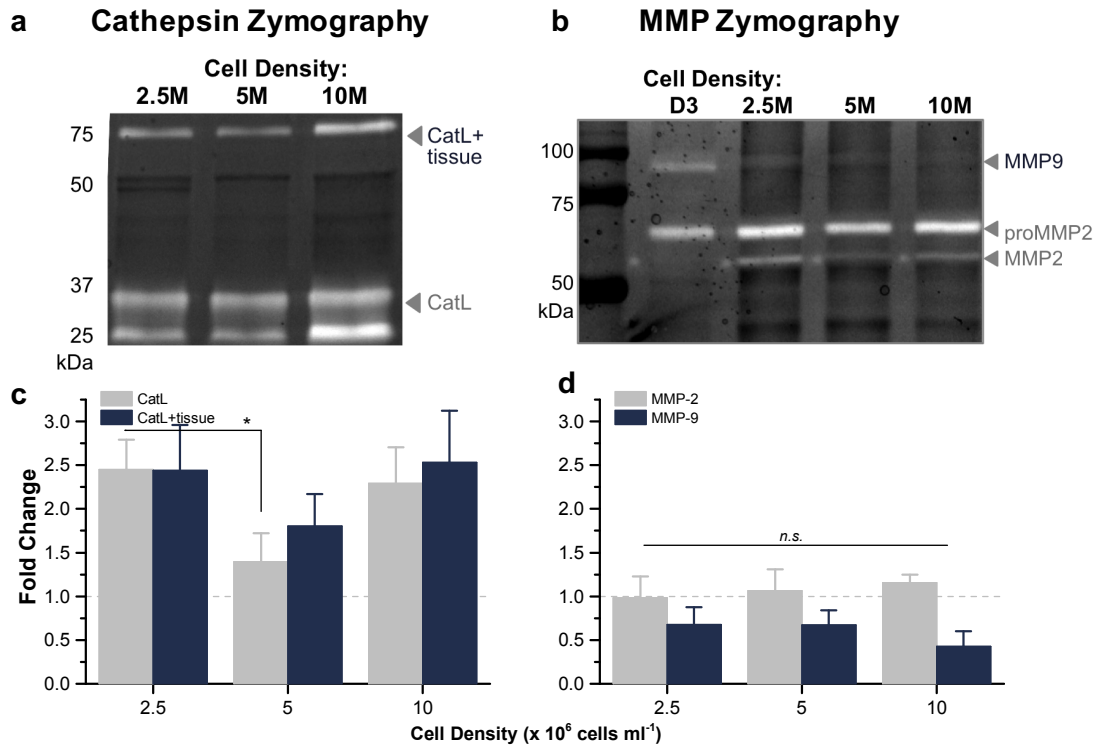
Figure 5-10. Gelatin Zymography of Cathepsins and MMPs in Muscle Strips Varying Cell Density. *(A) For day 12 muscle strips cultured with 1x ACA and fabricated with varying cell densities (2.5, 5, or 10 x 10^6 cells ml^{-1}), cathepsin zymography identified the amount of active CatL and CatL+tissue and (B) MMP zymography identified the amount of active MMP-2 and MMP-9. Zymogram signal for n = 3-4 muscle strips per condition was quantified for active cathepsins (C) and MMPs (D). All plots represent mean ± SEM. * indicates significance (p < 0.05) between conditions at the same time point; ^ indicates significance compared to initial time point.*

change compared to muscle strips with 2.5 or 10 x 10^6 cells ml^{-1} cultured in 1x ACA (Figure 5-10B, D).

Next, to assess whether cells in the muscle strip responded differently when subjected to a varying degree of static force, we altered the mechanical microenvironment felt by the myotubes during differentiation by 3D printing hydrogel skeletons with varying stiffness values (see Methods). When cultured in the serine protease inhibitor ACA, muscle

85

Cathepsin & MMP Zymography: Hydrogel Stiffness

Cathepsin Western Blot

Figure 5-11. Gelatin Zymography of Cathepsins and MMPs in Muscle Strips Varying Hydrogel Stiffness. *(A) For day 12 muscle strips cultured with 1x ACA with a hydrogel skeleton stiffness of 214, 319, or 489 kPa, cathepsin zymography identified the amount of active CatL and CatL+tissue and MMP zymography identified the amount of active MMP-2 and MMP-9. Zymogram signal for n = 3-4 muscle strips per condition was quantified for active cathepsins (B) and MMPs (C). (D) Western blotting confirmed the presence of cathepsins in muscle strips with varying hydrogel stiffness, for day 12 conditions compared to the initial time point of day 3. All plots represent mean ± SEM. * indicates significance (p < 0.05) between conditions at the same time point; ^ indicates significance compared to initial time point.*

strips coupled to hydrogel beams with a modulus of 319 kPa (an optimal stiffness used in previous bio-bot demonstrations that combined a robust mechanical structure with sufficiently high deflection necessary for locomotion [126, 177]) had retained their passive tension over time (Figure 5-6C). Muscle strips cultured in 1x ACA for 12 days in higher tension (489 kPa beam modulus) or lower tension (214 kPa beam modulus) showed no

significant change in the amount of active CatL or CatL+tissue compared to those on skeletons with a stiffness of 319 kPa (Figure 5-11A-B). Though MMP-2 showed a downward trend with increasing tension, there was no significant change for either MMP-2 or MMP-9 in an environment of higher or lower tension (Figure 5-11A, C). Western blotting corroborated the presence of the pro-form of CatL as well as CysC in these muscle strips (Figure 5-11D).

5.3.7 *Optical Stimulation of ChR2-C2C12 Muscle Strips*

To test if dynamic optical stimulation of muscle strips contributed to an increase in protease activity, bio-bots were fabricated using ChR2-C2C12s (myoblasts infected to express an optogenetic ion channel [177]). ChR2-C2C12 optogenetic bio-bots were cultured in 1x ACA and subjected to daily optical stimulation by an external blue light pulse that depolarized the cells to produce contraction. The muscle strips from ChR2-C2C12 optogenetic bio-bots were collected on day 6 (to compare to non-optogenetic bio-bots) and on day 12 (to compare to both non-optogenetic bio-bots as well as ChR2-C2C12 bio-bots without optogenetic stimulation) for zymography (Figure 5-12A). Just as daily electrical stimulation had not altered the expression of CatL or MMP-2 over 12 days, the application of daily optical stimulation for ChR2-C2C12 optogenetic bio-bots did not have a significant impact on the amount of active CatL, CatL+tissue, MMP-2 or MMP-9 (Figure 5-13). Finally, there was no fundamental difference between ChR2-C2C12 and non-optogenetic muscle strips with regards to the amount of active cathepsins or MMPs at either time point, and optical stimulation of ChR2-C2C12 optogenetic bio-bots resulted in similar
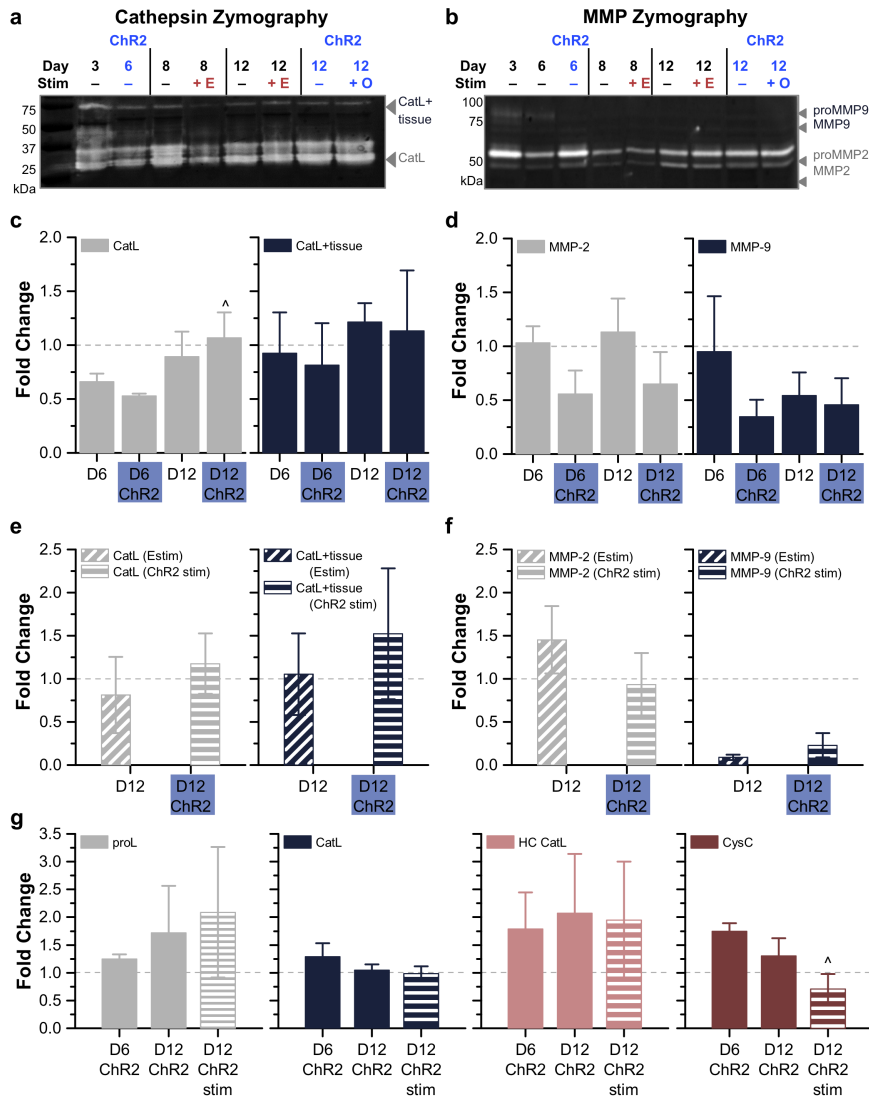
Figure 5-12. Gelatin Zymography of Cathepsins and MMPs in Muscle Strips Varying Applied Stimulation. *(A) Cathepsin zymography identified the amount of active CatL and CatL+tissue and (B) MMP zymography identified the amount of active MMP-2, and MMP-9, with and without electrical or optical stimulation. (C) Amount of active CatL, CatL+tissue, (D) MMP-2, and MMP-9 on days 6 and 12, for both non-optogenetic bio-bots as well as ChR2-C2C12 bio-bots without optogenetic stimulation. (E) Amount of active CatL, CatL+tissue, (F) MMP-2, and MMP-9 on day 12, for both non-optogenetic bio-bots subjected to electrical stimulation and ChR2-C2C12 bio-bots subjected to optical stimulation. 3 (G) Western blotting confirmed the presence of cathepsins in ChR2-C2C12 muscle strips on days 6 and 12, with and without optical stimulation (n=3 muscle strips per condition). All plots represent mean ± SEM. ^ indicates significance ($p < 0.05$) compared to initial time point (day 6 for panels C, D, and G).*

88
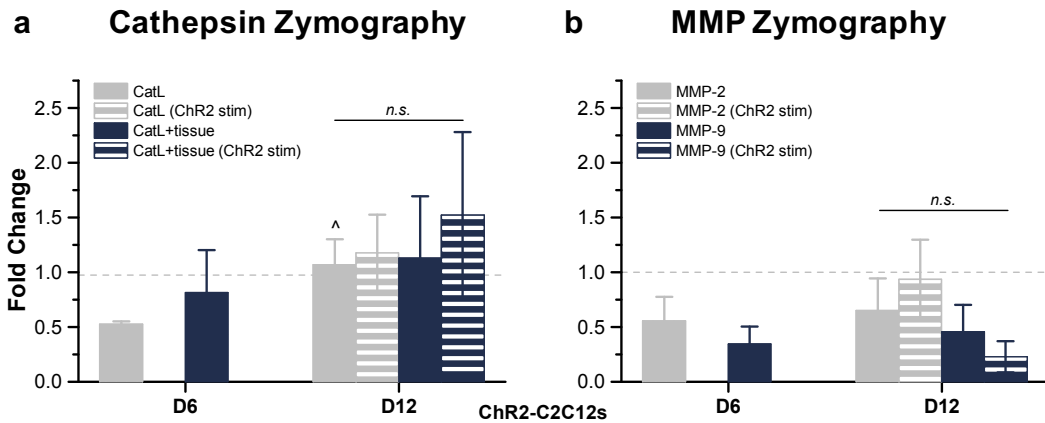
Figure 5-13. ChR2-C2C12 Optogenetic Bio-Bots. *(A) For ChR2-C2C12 optogenetic muscle strips, cathepsin zymography identified the amount of active CatL and CatL+tissue at days 6 and 12, and (B) MMP zymography identified the amount of active MMP-2 and MMP-9 at days 6 and 12, in the presence or absence of optogenetic stimulation (n=3 muscle strips per condition). All plots represent mean ± SEM. * indicates significance (p < 0.05) between conditions at the same time point; ^ indicates significance compared to initial time point (day 6).*

trends in active CatL, CatL+tissue, MMP-2 and MMP-9 expression as electrical stimulation (Figure 5-12C-F). Western blotting confirmed the presence for pro-form of CatL and CysC in the ChR2-C2C12 optogenetic bio-bots (Figure 5-12G).

5.4   Discussion

The role of the basal lamina is of critical importance in providing necessary elasticity and strength, as well as key binding sites for sequestration of myogenic growth factors, during muscle development. Laminin-2, collagen IV, and proteoglycans are present in abundance in skeletal basal lamina. In the hydrogel-muscle platform of the skeletal muscle bio-bots, natural ECM proteins comprise a bioinspired engineered system that provides mechanical and biochemical support to differentiating myotubes in a

physiological arrangement. After myoblasts withdraw from the cell growth cycle and commit to a myogenic lineage during differentiation, they migrate to (and then physically fuse with) other myoblasts to create multinucleated myotubes with one cytoplasm, and increase expression of muscle-specific genes in adult muscle fibers [187, 202, 205]. The cells in the muscle strips exhibited these morphological and genetic changes, observable by immunostaining (Figure 5-3A) and the presence of mature muscle proteins (Figure 5-3C). During injury, a similar process of migration and fusion is observed by quiescent satellite cells, or myoblast progenitors, present in the basal lamina [206].

In vivo, protease activity is regulated by a strict balance between expression, activation, and inhibition; this balance plays a major role in the degree of matrix degradation and resulting cellular processes [190]. *In vitro*, degradation and resulting instability are drawbacks to tissue engineering with fibrin, and an inhibitor such as the anti-fibrinolytic agent ACA [184] must be incorporated to prevent indiscriminate matrix digestion in systems lasting longer than a few hours. Incubation with increasing concentrations of ACA helped to preserve the structural integrity of the fibrin muscle strips and significantly increased the lifetime before rupture (Figure 5-5). Though the serine protease inhibitor aided in the maintenance of passive tension in the muscle strips over time and slowed the degradation of muscle strips at earlier time points, it did not significantly reduce tissue loss by day 12 (Figure 5-6); thus, the need for an investigation of the other protease families, including cysteine cathepsins and MMPs, was apparent.

In gel zymography is a versatile method used to determine protease activity by incubation in a gel substrate composed of degradable proteins. As it does not require

90

antibodies, is relatively inexpensive, and allows for visual confirmation of enzyme identity (and determination of active quantity) by molecular weight separation, zymography is useful in the study of both cathepsins and MMPs [134, 135] across a number of tissue types [56, 122, 207, 208]. Using a well-characterized technique wherein electrophoresis separated proteases of interest from muscle strip lysate [134, 135], we temporally quantified the amount of active cathepsins as well as MMPs in various culture conditions (Figure 5-4B, Figure 5-7, Figure 5-8, Figure 5-13). Cathepsins are implicated in a number of tissue-destructive disease states as well as mechanisms of normal tissue physiology, including myogenic differentiation [36, 39, 94, 134, 208, 209]. CatL has been recognized as a major player in muscle atrophy and wasting [187, 205]. We observed that the amount of active CatL present in the muscle strips increased steadily over time, and was most highly expressed in the absence of ACA. Its activity on day 12 was significantly higher in all 0x muscle strips compared to 1x and 3x ACA, both with and without electrical stimulation (Figure 5-8C), indicating that the presence of ACA may prevent some processes associated with physiological muscular breakdown and potentially reduce the expression of CatL. There is also prior evidence suggesting ACA may inhibit proteolytic activity by lysosomal cysteine cathepsins [210, 211]. These hypotheses could explain the extended lifetime observed in the presence of ACA and why this process is independent of electrical stimulation.

MMPs comprise a range of proteases capable of degrading all ECM proteins. Like cathepsins, they are also involved in many normal mammalian cell processes (such as migration, adhesion, and proliferation), and their upregulation has been reported in various

91

pathologies. Their role in myogenic differentiation, development, and injury is primarily extracellular, and they have been shown to interact with specific ECM proteins over varied timelines [212]. Cellular alignment, differentiation of adult myofibers, or remodeling of muscle tissue architecture is a direct result of MMPs degrading ECM proteins, which may hinder myoblast migration and fusion [189, 206]. They can also be activated (directly or indirectly) by other proteolytic enzymes including plasmin and plasminogen [191], which are present in our system. Both MMP-2 and MMP-9 can act on various collagen substrates, some of which are localized in our engineered muscle strips due to the presence of Matrigel^{TM} [189].

Expression of MMP-9 has been shown to be upregulated in C2C12 myoblast cultures during migration and before cell fusion [188, 206]. The processes concerning mature myotube formation then witness a decrease in expression over time. We observed a decline in amount of active MMP-9 at later time points in muscle strips cultured with 1x and 3x ACA (Figure 5-7B, Figure 5-12D, F, Figure 5-13B). Since upregulation of MMP-9 is also associated with early acute inflammation and initiation of muscle regeneration in proliferating and migrating satellite cells [206], it is likely that only the earlier stages of muscle strip compaction and differentiation involve processes that resemble the recruitment (and subsequent fusion) of satellite myoblasts after injury.

MMP-2, which is secreted by C2C12s, is constitutively produced in lower levels in healthy adult muscle and does not play a large role in myoblast migration or fusion. It is involved more heavily in later stages of C2C12 myofiber differentiation [188, 190, 206]. Active MMP-2 was slightly increased in all conditions on day 12 compared to day 3 (Figure

5-7B, Figure 5-8C), suggesting that while myogenic differentiation was observed, the complete extracellular remodeling of matrix proteins that accompany later stages of muscular maturation may not have been visible during this experimental timeline. Though a short-term investigation proved relevant for this application of biological machines, an investigation on a longer time scale would provide more detailed temporal data on expression of specific MMPs as well as cathepsins. Additionally, levels of MMP-2 (which can be secreted by satellite cells) increase gradually during later stages of injury, then return to baseline levels, suggesting an important role in repair and regeneration of new muscle fibers [188, 190]. The introduction of deliberate inflammation or injury to the muscle strips could test this hypothesis in future studies.

The fabrication of skeletal muscle bio-bots offers flexibility in varying almost any parameter in the system. We also investigated the effect of altering certain variables whose effects on bio-bot functionality we have previously reported: cell density, hydrogel stiffness [126], and dynamic optogenetic stimulation [177]. Varying these conditions revealed that the amount of active cathepsin and MMP was largely independent of these environmental changes or applied stimuli. When the cell seeding density for muscle strip formation was varied, the amount of active cathepsin and MMP was statistically unchanged (Figure 5-10). MMP-2 increased slightly as the cell concentration was increased to 10 x 10^6 cells ml^{-1}, though not significantly; it is possible that the increase in initial myoblast density accelerated the fusion and differentiation process. MMP-2 has been shown to increase with the presence of mature myotubes [190, 191], substantiating this observation.

93

When muscle strips were differentiated under higher static tension, the amount of active cathepsin L decreased, although again not significantly (Figure 5-11). Interestingly, the amount of active MMP-2 and MMP-9 both decreased in the presence of higher static tension (Figure 5-11C), despite an extracellular presence in a condition that favors higher degrees of myofiber alignment and differentiation. This may indicate reduced myotube formation in these bio-bots under higher tension.

Finally, when ChR2-C2C12 optogenetic bio-bots were subjected to optical stimulation, we observed no change in the amount of active CatL or MMPs (Figure 5-12, Figure 5-13). Though electrical stimulation is a useful tool because it allows for the global and coordinated stimulation of all excitable cells within the muscle strip, optical stimulation provides a less invasive and more specific alternative by increasing spatiotemporal control over contraction and locomotion [177, 213]. These results indicate that neither the use of optogenetic cells nor optical stimulation altered proteolytic expression profiles associated with myogenic differentiation or injury, and confirm that the electrical stimulation is not responsible for the observed proteolytic activity by cathepsins and MMPs.

In engineered living systems, biological building blocks – different cell types or tissues in an instructive environment – can be assembled to promote the emergence of cellular systems with well-defined functionality, allowing for the realization of dynamic cellular machines with the ability to interface with the environment and other living systems [174]. However, as these biological machines increase in complexity, it may be

critical to design systems with a specific life expectancy and exhibit greater consistency. Forward engineering of biological and mechanical properties, along with knowledge of how these factors affect protease activity, can dictate a system's sustainability. A machine designed to 'fail' or 'break' within a certain time frame (after performing a specific task, for example), could be programmed to upregulate expression of proteases through inflammatory cytokines such as TNF-α, which are known to accelerate overexpression or activation of some proteases [94, 188]. On the other hand, the addition of a protease inhibitor (such as the strong-binding endogenous cysC [4, 71, 134], the broad-spectrum cathepsin inhibitor E-64 [71, 94, 134], or tissue inhibitors of metalloproteinases (TIMPs) [189, 214]) could aid in prolonging the machine's lifespan by lowering the amount of active proteases. Most likely, the majority of future engineered living systems will require an intricate balance between the activity and inhibition of serine and cysteine proteases as they exist and function together in a complex, inter-connected proteolytic network [215, 216]. The ratio between proteases and their inhibitors (such as the important equilibrium required between MMPs and TIMPs in skeletal muscle [188, 189, 191, 206]) is critical to maintaining a balanced state.

Looking forward, as neuromuscular control and vascularization are introduced into engineered living systems, it will be critical that we understand what proteases are involved in these processes and how the cells may adapt to the new environment. Proteolysis of the basal lamina is critical in angiogenesis, and MMP-2 and membrane-type 1 (MT1) MMP involvement has been seen in capillary growth [188]. Furthermore, MMP-2, MMP-7, and MMP-9 are co-localized at neuromuscular junctions in healthy tissue [188]. Taken

together, these highlight the importance of recognizing that the proteases that appear to be involved in normal muscle strip development will also be involved in these other cellular processes, and must be accounted for in future development of engineered living systems.

Finally, because skeletal muscle cathepsins can also hydrolyze myogenic proteins such as myosin, troponin T, and tropomyosin [205, 217], it may also be worth examining whether the activity of these proteases in intracellular degradation (and subsequent breakdown or atrophy of muscle fibers) contributes in any way to failure of the engineered muscle system. Lastly, increased expression and regulation of cathepsins [205, 218] and MMPs [36] has been observed in dystrophic muscles; therefore this platform could also be used for the study of disease-specific models and treatments of diseases such as muscular dystrophy or other myopathies.

CHAPTER 6.   ENGINEERING BIOLOGICAL MACHINE

LIFESPAN THROUGH PROTEOLYTIC DEGRADATION

6.1   Introduction

Locomoting biological machines powered by skeletal muscle strips produce increasing amounts of cathepsin L over a 12 day timeline [128]. As described in Chapter 5, modifying different biological machine design parameters (cell seeding density, hydrogel stiffness, electrical stimulation, and optogenetic stimulation) does not alter the cysteine cathepsin protease activity of these bio-bots. Biological treatment with aminocaproic acid (ACA), which prevents plasmin, the known fibrin degrading protease, from becoming its active form (Figure 5-1) can prolong the life expectancy of the biological machines. At a low concentration of ACA (1X), the average time until bio-bot breaking went from 8.2 ± 0.5 days to 17 ± 3.8 days and at a high concentration of ACA (3X), the average time of breaking is prolonged until 105.5 ± 31.1 days (Figure 5-5). While ACA lengthens expected survival time of these biological machines, the large variability in when the bio-bots break currently cannot be controlled.

Here we examined the life expectancy of these engineered living systems to determine ways to control the time until failure. We demonstrated that specifically inhibiting the cysteine cathepsin activity of these muscle strips reduces the cathepsin bound to muscle strip matrix, without affecting overall muscle strip viability and passive tension. Furthermore, we re-examined the life expectancy data using a Weibull distribution, a

function commonly used to model lifetime distributions in reliability engineering, to describe the functional lifetime distributions of biological machines. Using the protease activity of these machines, we developed a model to predict the life expectancy of these engineered living systems, which can be used to design interventions to control the time until bio-bot failure.

6.2   Materials and Methods

6.2.1   Zymography

In-gel zymography was used to detect active cysteine cathepsins in bio-bot muscle strip samples. This modified SDS-PAGE used a gel impregnated with either 5 mg ml^{-1} soluble gelatin substrate or 5 mg ml^{-1} soluble fibrinogen substrate. Samples were prepared in a non-reducing loading dye and separated by electrophoresis at 4°C. Gels were washed in a renaturing buffer to refold enzymes and then incubated in assay buffer overnight at 37°C for optimal protease activity. The following morning gels were stained with Coomassie blue and destained with white bands indicative of protease activity. Gels were imaged with an ImageQuant LAS 4000 (GE Healthcare) and densitometry using ImageJ (NIH) was used to quantified the intensity of the white bands and is reported as the average value ± the standard error of the mean (SEM).

6.2.2   Biological machine muscle strip parameter evaluation

Biological machine cell culture development was the same as described in the Methods of Chapter 5. At day 3 in the development timelines (Figure 5-2), when switching

98

from growth media to differentiation media of DMEM with 10% (v/v) heat-inactivated horse serum (HS, Gibco), 1% (v/v) each of L-glutamine and penicillin-streptomycin, 50 ng ml^{-1} of insulin-like growth factor (IGF-1, Sigma-Aldrich). E-64, the cysteine cathepsin, broad spectrum, small molecular inhibitor was added to differentiation media at 5 μM and 10 μM concentrations. Media was changed daily. Bio-bot muscle strips samples were collected at day 3, day 12, day 30, and day 90.

The passive tension force of the muscle strip was calculated using an equation derived from Euler-Bernoulli linear beam theory as described in Section 8.2.4. The elastic modulus for these bio-bots were 319 kPa. Cell viability was compared between day 3 and day 12 samples using a metabolic colorimetric assay described in Section 8.2.5. Muscle creatine kinase (MCK), a marker of muscle cell differentiation, was assayed using Liquid Creatine Kinase Reagent Set (Pointe Scientific) as described in Section 8.2.8.

6.2.3   *Using Weibull distribution to describe biological machine survival*

Previously, the biological machine survival was determined and reported as a Kaplan Meier survival function, which converted the "age" of which individual bio-bots failed to a survival function. Kaplan Meier is popularly used because it can handle cases where participants "drop-out" and cannot be followed over the entire course of a study. However, we can track all the biological machines until they break, which opens additional methods of analysis. One such lifetime model of survival was the Weibull distribution [219]. The

Weibull frequency distribution was implemented in MATLAB and the probability density function can be described by:

$$f(x|a,b) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-(x/a)^b} \qquad (2)$$

where a was the scale parameter, b was the shape parameter (also called the Weibull slope), and x was the time until failure. The Weibull frequency distribution is often used to model the breaking strength of materials and other reliability engineering lifetime distributions. Weibull distribution parameters were determined for the 0X, 1X, and 3X biological machine time until failure data. All code associated with determining the Weibull parameters for these datasets is included in Appendix D.

6.2.4 Biological machine survival model development

A percolation theory model was developed using MATLAB to predict the bio-bot lifespan based on reported survival functions. This two-compartment model has (1) a muscle strip compartment and (2) a protease compartment. The muscle strip compartment was composed of a 2D grid representing fibrin(ogen)/extracellular matrix at each time step cycling through each node in the network, the protease compartment generates an amount of cathepsin, which if less than the $P_{cleavage}$ threshold results in a cleavage edge. At the end of the time step, the number of connected edges was calculated and if it is below a certain probability of broken, the muscle strip was broken and that time of failure was recorded. If the number of connected edges was greater than a threshold P_{broken}, the muscle strip

100

survives to the next time point. The protease compartment was composed of linear models of increasing cysteine cathepsin activity extrapolated form the protease characterization described in Chapter 8. The two main parameters of this model were the $P_{cleavage}$ and P_{broken} thresholds, which were estimated based on the Weibull distribution for the biological machines. The code describing this model is included in Appendix D.

6.3 Results

6.3.1 *E-64 treated biological machines reduced cathepsin bound to matrix*

Chapter 8 characterizes the cysteine cathepsin and MMP proteolytic activity of these biological machines [128]. Cysteine cathepsins and MMPs have known collagenase/gelatinase activity [2, 3, 21, 170, 220, 221], however the major matrix protein in these bio-bot muscle strips was fibrin [126-128]. Cysteine cathepsins are not well established in the literature as fibrinogenolytic proteases, however Chapter 6 explored and confirmed the proteolytic potential for cysteine cathepsin hydrolysis of fibrin(ogen). Since biological machines secrete cathepsin L, we adapted the multiplex cathepsin gelatin zymography protocol to use a fibrinogen substrate, instead of gelatin substrate, to verify that the cathepsins secreted by the muscle strips can degrade the fibrin(ogen) matrix (Figure 6-1).

CATHEPSIN FIBRINGOEN ZYMOGRAM
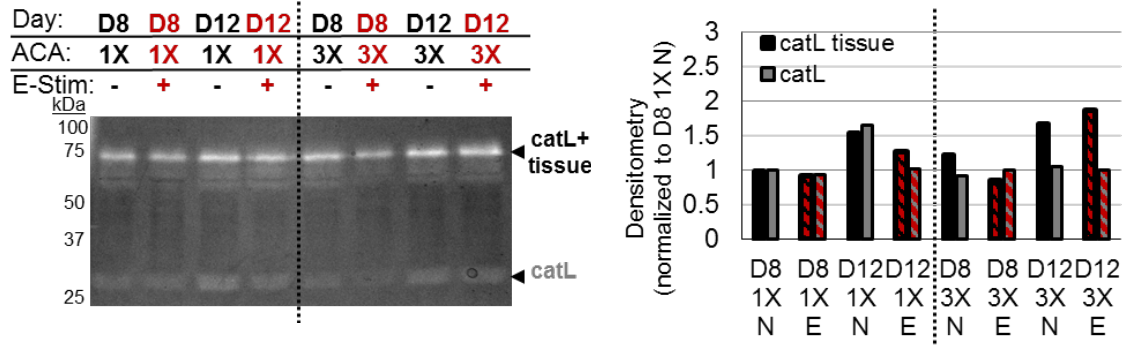
Figure 6-1. Fibrinogen zymography confirms fibrinogenolysis by bio-bot cathepsin L. *Bio-bot cathepsins have increasing fibrinogenolysis over time.*

Cathepsin Zymography
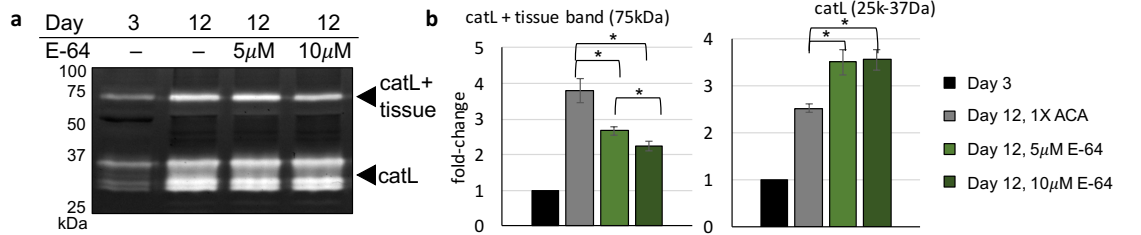
MMP Zymography
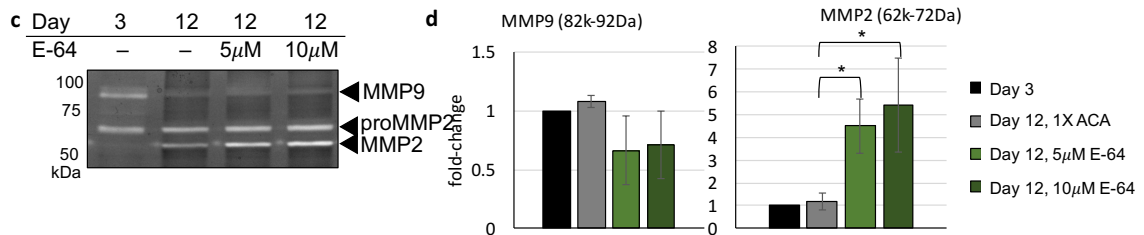
Figure 6-2. Gelatin zymography of E-64 cultured muscle strips. *(A-B) Cathepsin zymography identified a dose dependent decrease in the amount of CatL+tissue and (C-D) MMP zymography identified an increase in the amount of active MMP-2, for day 12 conditions compared to the initial time point of day 3. All plots represent mean ± SEM. * indicates significance ($p < 0.05$) between conditions at the same time point (n=5-7).*

102

Once confirming that the cathepsins secreted by the bio-bots can degrade fibrinogen, we explored the effects of specifically inhibiting the cysteine cathepsin in these biological machines. E-64, the broad spectrum, small molecule cysteine cathepsin inhibitor [99-101], was added to the bio-bot differentiation media starting on day 3 and the proteolytic activity of these machines was assayed by gelatin zymography on day 12 when treated with a low (5 µM) and high (10 µM) concentration of E-64 (Figure 6-2). The cathepsin L bound to muscle strip tissue (75kDa, catL+tissue) at Day 12 of E-64 was significantly reduced in a dose-dependent manner compared to Day 12 1X ACA bio-bots from 3.80 ± 0.34 to 2.67 ± 0.12-fold at a low concentration of E-64 and 2.24 ± 0.13-fold at a high concentration of E-64 (Figure 6-2A, B). Furthermore, there was a concurrent increase in total MMP-2 with the E-64 treatment compared to the Day 12 1X ACA bio-bots from 1.08 ± 0.05 to 4.51 ± 1.20-fold at a low concentration of E-64 and 5.43 ± 2.06-fold at a high concentration of E-64 (Figure 6-2C, D). This increase in active MMP may be a result of a shift in the muscle strip's proteolytic network through cysteine cathepsin inhibition.

6.3.2   *E-64 does not negatively impact biological machine key metrics*

The cell viability of muscle strip cells was quantified by incubating the muscle strip with a metabolic colorimetric solution (CellTiter 96 AQueous One Solution, MTS, Promega) in the dark at 37°C for 4h, then absorbance was measured at 490 nm using a microplate reader. Muscle strips showed increased in percentage of viable cells at day 12 compared to day 3 muscle strips and there was no significant difference between the day

12 with increasing concentrations of E-64 (Figure 6-3A). Furthermore, muscle creatine kinase (MCK) is a marker of mature skeletal muscle, expressed after differentiation [203]. Previously, muscle strips reached a maximum MCK activity output by day 12, which was also seen with the E-64 treated bio-bots. There was no significant difference between the MCK activity of muscle strips with increasing concentrations of E-64 at day 12 (Figure 6-3B).

E-64 treatment did not affect the functional output, muscle strip passive tension, of biological machines. Figure 6-3C compares the passive tension in the muscle strips by day 12 and there was no significant difference between the treatment conditions. Furthermore, there was no difference between adding 1X ACA or low dose (5 µM) of E-64. Long-term passive tension also did not show any significant difference due to treatment conditions (Figure 6-3D). Increasing E-64 from the low to high concentration did not greatly increase passive tension, but combination of E-64 and ACA treatments appeared to have a greater impact.

Preliminary data for the impact of E-64 on life expectancy of the biological machines was very encouraging. Table 6-1 records the known days at which bio-bots have broken with one of three treatments: (1) 5 µM and 0X ACA, (2) 5 µM and 1X ACA, or (3) 10 µM and 1X ACA. Treatment condition (1), which had comparable passive tension to 1X ACA treated bio-bots previously described, had at least a 3.5-fold increase in life expectancy (an average of 59.3 days with 5 µM E-64 only treated compared to 17 days with 1X ACA only treated bio-bots). The second row of Table 6-1 gives the age of current

bio-bots (as of 5/20/2017) for each of the conditions. It was important and encouraging to note that both the combination treatments of E-64 and ACA treated biological machines have multiple bio-bots still functioning by day 121. If we assumed that all the remaining bio-bots were to break on 5/20/2017, the average life expectancies of each E-64 treatment condition would be 2.6 to 3.8-fold improved compared to the 1X ACA treated bio-bots (treatment condition 1 would be 44.6 ± 10.6 days, treatment condition 2 would be 62.9 ± 22.6 days, and treatment condition 3 would be 64.6 ± 17.3 days).
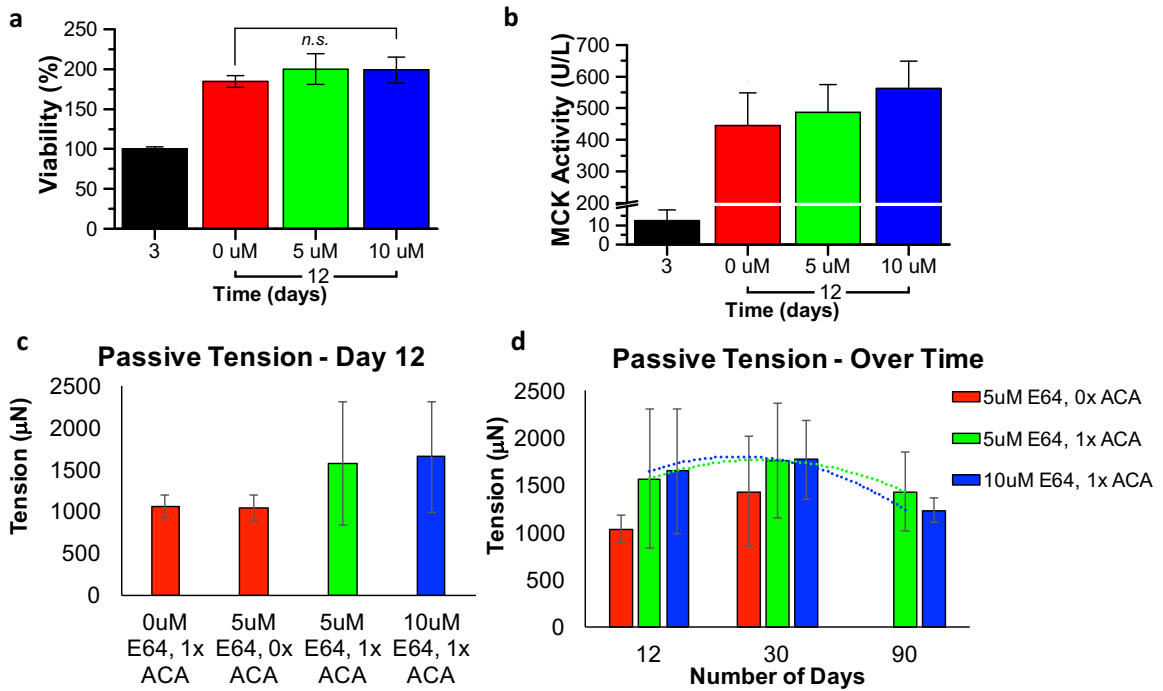
Figure 6-3. E-64 does not impact cell viability, MCK activity, and passive tension in E-64 cultured muscle strips.

Table 6-1. Life expectancy of E-64 cultured muscle strips.

	5 μM E-64 0X ACA	5 μM E-64 1X ACA	10 μM E-64 3X ACA
# days until broken	25, 64, 89 Average: 59.3± 22.8	20 Average: 20	46, 69 Average: 57.5±16.3
# days in culture (not broken as of 5/20/2017)	9, 39, 39 47	9, 9, 39, 121, 121, 121	9, 39, 47, 121, 121

6.3.3 Weibull distribution describes biological machine lifespan

The Weibull distribution is a function used in reliability engineering and used to model lifetime distributions for properties such as breaking strength of materials. Here we applied the Weibull distribution to model the life expectancy of biological machines and determine unique scale and shape parameters for each ACA concentration (

Table 6-2, Figure 6-4). While there were very different Weibull parameters for each of the

three distributions, there appeared to be a relationship between the two Weibull parameters

and the treatment concentration of ACA. Figure 6-5 shows the value of the scale parameter

and shape parameter plotted as a function of ACA concentration. Equation 3 describes the

relationship between the scale parameter (a) and ACA with an R^2 value of 1:

$$a = 8.8946 * e^{0.7838*[ACA]}$$

(3)

Equation 4 describes the relationship between the shape parameter (b) and ACA with an

R^2 value of 0.9604:

$$b = 4.8667 * e^{-0.562*[ACA]}$$

While there were currently only three data points, this trend seemed encouraging. This relationship may also extend to predicting the life expectancy of E-64 treated bio-bots, if we assumed that 1X ACA = 5 μM of E-64, then for 10 μM of E-64, the expected Weibull parameters would be: a = 42.21 and b = 1.582, which would be an average life expectancy of 37.9 days.
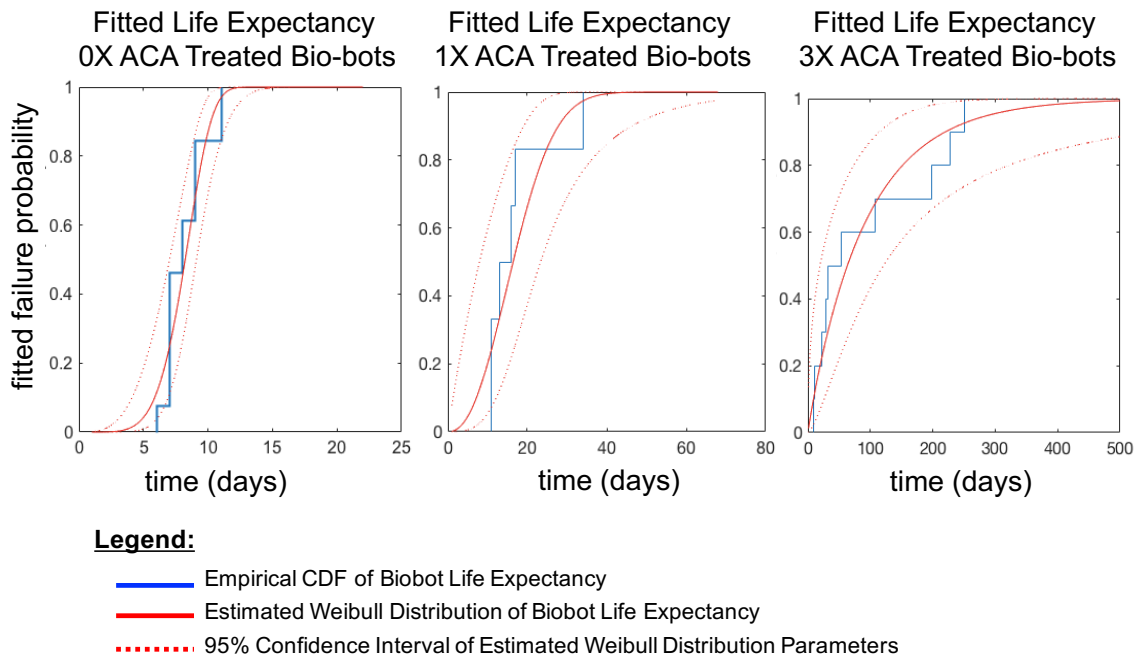
Figure 6-4. Fitted Weibull Distribution of Bio-Bot Life Expectancy. *The blue curve represents the empirical cumulative distribution function of 0X, 1X, and 3X ACA cultured bio-bots. The red curve represents the fitted Weibull distribution and the dotted red curves represent the 95% confidence bounds of the estimated Weibull distribution parameters. Weibull parameters are reported in*

Table 6-2.

107

Table 6-2. Life expectancy Weibull parameter varying ACA concentration.

	Weibull Parameter	Confidence Interval
0X Life Expectancy	a = 8.800	[7.931, 9.765]
	b = 5.552	[3.712, 8.303]
1X Life Expectancy	a = 19.29	[13.27, 28.06]
	b = 2.227	[1.289, 4.024]
3X Life Expectancy	a = 92.42	[46.78, 182.6]
	b = 0.9636	[0.590, 1.574]
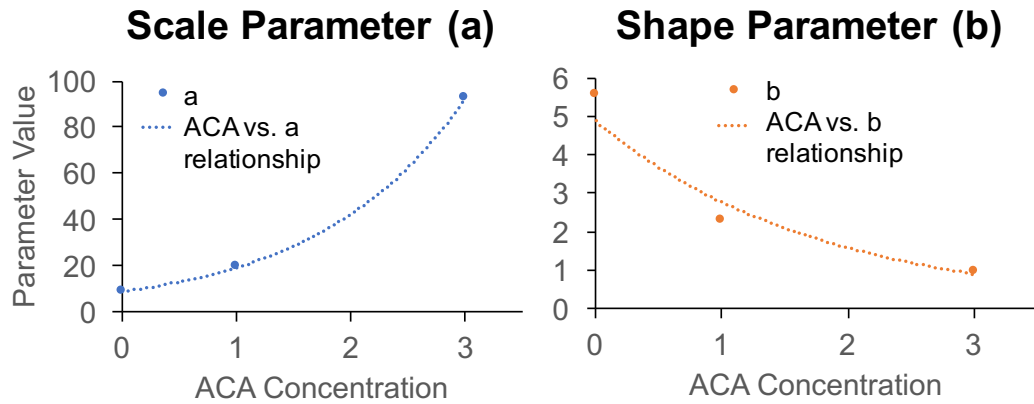
Figure 6-5. Weibull parameters represented as a function of ACA concentration.
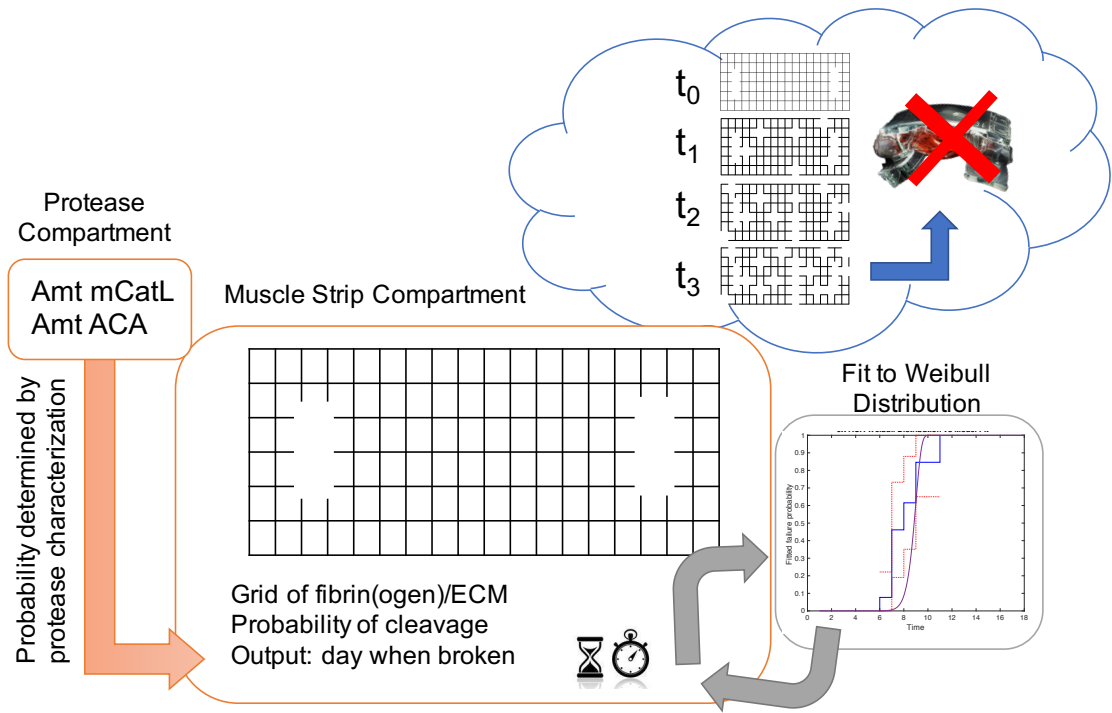
Figure 6-6. Schematic of Bio-Bot Life Expectancy Model.

6.3.4   Developing model to predict biological machine lifespan

Figure 6-6 illustrates a schematic representation of the biological machine life expectancy model and all the MATLAB code for model implementation is documented in Appendix D. There were two main compartments, the protease compartment, which output an amount of protease that was input into the muscle strip compartment, where the protease amount determines if cleavage of the matrix would occur in the fibrin(ogen)/ECM grid. The protease compartment was extrapolated from the known increases of cysteine cathepsin over time, where the fold-change increases determine in Chapter 5 were related to a physical amount of recombinant cathepsin as demonstrated in Figure 6-7. The protease amount was then assumed to increase linearly over time, with the 0X ACA model

simulations resulting in more protease than the 1X and 3X ACA models, like the relationship depicted in Figure 5-8.

a pH4 Cathepsin Zymography   **b Quantifying Protease Amount**
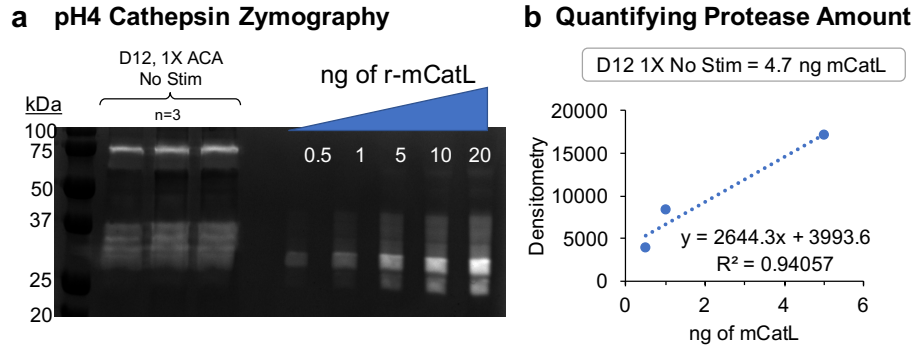
Figure 6-7. Quantified amount of active cathepsin was used to information the life expectancy model's protease compartment.

In the muscle strip compartment, a 2D lattice graph was created, where the edges represent the cleavable fibrin(ogen)/ECM matrix. This graph had post locations removed to recapitulate an approximate geometry of the bio-bot muscle strip. At each time point, cycling through each node, a protease amount was generated from the protease compartment dependent upon the current time and ACA concentration. If that amount was less than the $P_{cleavage}$ threshold, then one of the edges of that node was selected uniformly at random within the neighborhood of the selected node. The $P_{cleavage}$ was weighted by the number of edges for the node being analyzed, because it was expected that matrix degradation exhibits spatiotemporal locality. At the end of the time point, the fraction of the graph's nodes contained in the largest connected component was compared to the P_{broken} threshold. If the largest connected component contained at least P_{broken} of the nodes in the graph, the muscle strip continued to the next time step; otherwise, the muscle strip was

110

considered to have failed and that functional lifetime was collected for cumulative distribution function and Weibull distribution parameter estimation after all model simulations were completed. The $P_{cleavage}$ and P_{broken} each had ±25% variability in attempts to account of the large variability seen in the life expectancy of the biological machines.

Then a grid search was run to determine if there was a combination of $P_{cleavage}$ and P_{broken}, using this model construct, that could give the Weibull distribution parameters to describe the 0X, 1X, and 3X life expectancies. The results of the grid search are represented in Figure 6-8. The top panels describe Weibull distribution scale parameter (a, left) and shape parameter (b, right) for 0X ACA cultured bio-bots. The middle panels describe Weibull distribution parameters for 1X ACA cultured bio-bots and the bottom panels describe the parameters for 3X ACA cultured bio-bots. The relative error was calculated as described in Equation 5:

$$relative\ error = \left| \frac{true\ value - predicted\ value}{true\ value} \right| \qquad (5)$$

where the *true value* was the point-estimate of the parameters described in Table 6-2 and the *predicted value* was the estimated Weibull distribution parameter for the given $P_{cleavage}$ and P_{broken} pair determined after 200 muscle strip simulations. The shape parameters (b) are related to the inverse variance of the lifetimes, thus there were some parameters with large values of b, including cases of b=infinity, when all simulated bio-bots failed on the same day. These errors are represented on a log scale.

111

Unfortunately, using only these two thresholds as the model parameters and the protease compartment, as currently constructed, was not sufficient to recapitulate the Weibull distribution parameters estimated for the 0X, 1X, and 3X bio-bot life expectancies in

Table 6-2. However, various ranges of $P_{cleavage}$ provided the best estimates for the two different Weibull distribution parameters. The scale parameter (a) appeared to be tightly bounded to a $P_{cleavage}$ of 0.1 to 0.2 and increasing concentrations of ACA appeared to become independent of P_{broken}. Across the three ACA concentrations, if we only needed to accurately describe the scale parameter, then a $P_{cleavage}$ of 0.1 and P_{broken} on 0.25 fit the best across the life expectancies (Figure 6-9A). The shape parameter (b) had the smallest errors at larger $P_{cleavage}$ thresholds (0.25 to 0.40). The shape parameter also appeared to be independent of P_{broken} thresholds. Across the three ACA concentrations, if we only needed to (1) accurately describe the shape parameter, and (2) minimize the $P_{cleavage}$ to accommodate the scale parameter, then a $P_{cleavage}$ of 0.25 and P_{broken} on 0.25 fit the best across the life expectancies (Figure 6-9B). Comparing these solutions, the best fitting scale parameter best recovered the experiment life expectancies (Figure 6-9A).
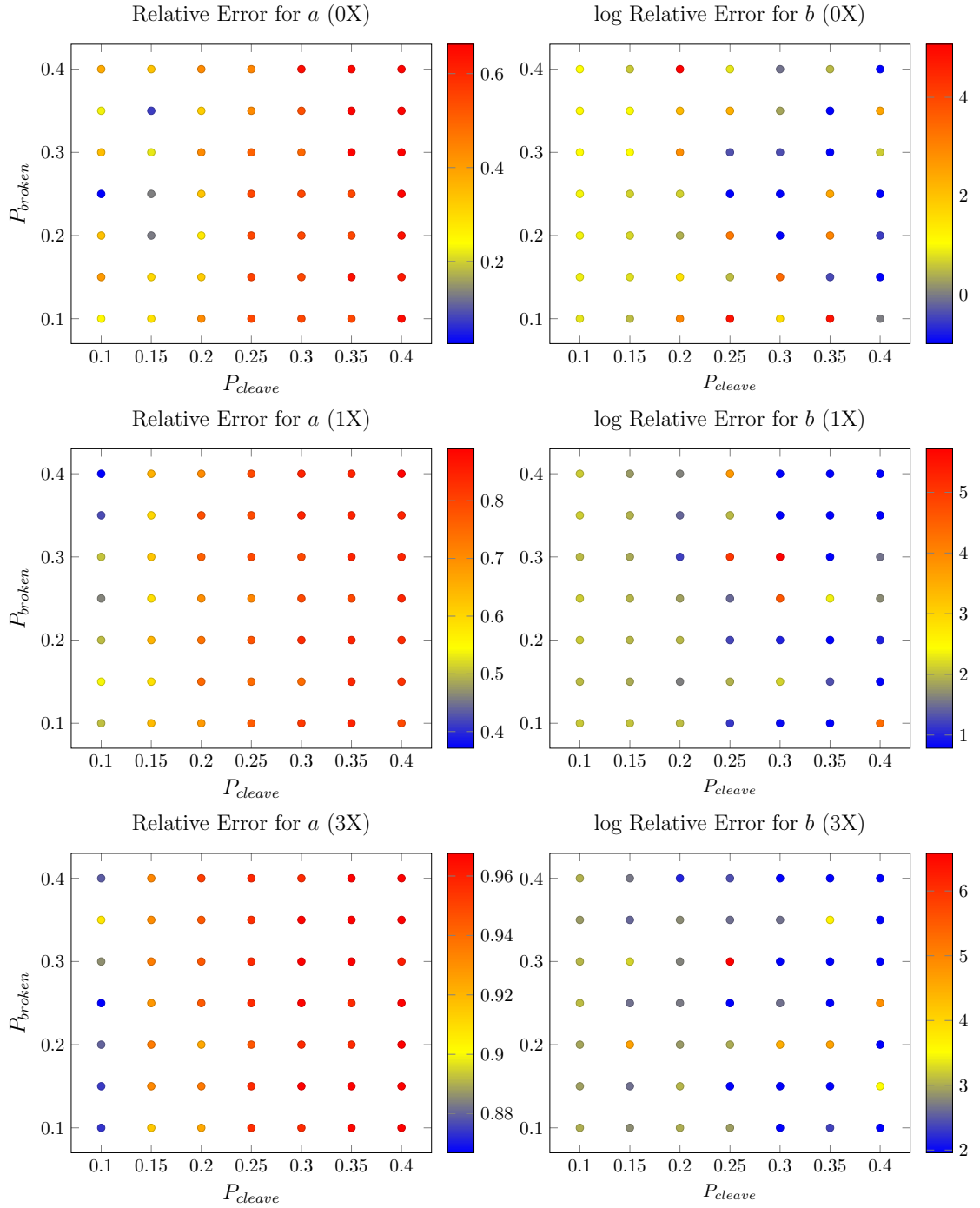
Figure 6-8. Compiled grid search results of the relative error for different combinations of P$_{cleavage}$ and P$_{broken}$ on the Weibull distribution parameters for the life expectancy of 0X, 1X, and 3X cultured biological machines.

a Best Scale Parameter (a) found at $P_{cleavage}$ = 0.1 and P_{broken} = 0.25
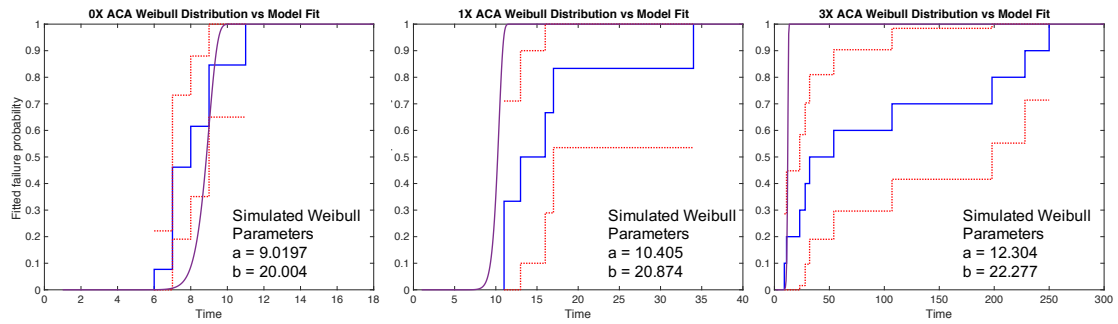
b Best Shape Parameter (b) found at $P_{cleavage}$ = 0.25 and P_{broken} = 0.25
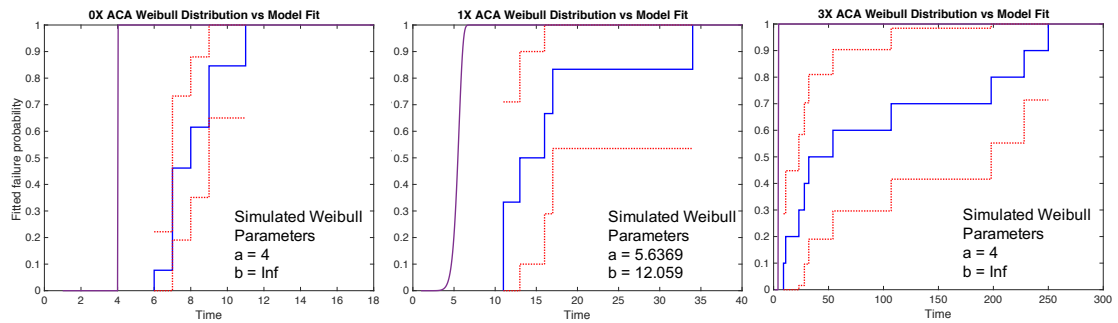
Legend:

⎯⎯ Empirical CDF of Biobot Life Expectancy

·········· 95% Confidence Interval of CDF

⎯⎯ Weibull Distribution

Figure 6-9. Best fit $P_{cleavage}$ and P_{broken} threshold pairs compared to biological machine life expectancies. *The blue curve represents the empirical cumulative distribution function of 0X, 1X, and 3X ACA cultured bio-bots. The dotted red curve represents the 95% confidence interval empirical cumulative distribution function. The purple curve represents the Weibull distribution based on the best $P_{cleavage}$ and P_{broken} threshold pair for (A) the scale parameter and (B) the shape parameter.*

6.4   Discussion

Currently, most of the locomoting biological machines will only show electrically induced contractility for 8 days following formation, without any intervention. As more advance biological machines are constructed, more cell types will be in contact with each other and may alter the machine's proteolytic profile. Therefore, protease interactions with

the machine's matrix and how they interact with each other are important in the design of stable biological machines. Then these proteolytic activities can be commandeered as a designed benefit of the biological machines rather than a destructive mechanism to wreak havoc on machine stability.

Previously we have characterized the protease activity of these biological machines (described in detail in Chapter 5), which produce an increasing amount of cathepsin L over time, independent of electrical stimulus and modulating of bio-bot design parameters (initial cell seeding density, backbone hydrogel stiffness, and optogenetic stimulus). Here we show that the cathepsin L secreted by these biological machines had fibrinogenolytic activity (Figure 6-1) as detected by a fibrinogen zymography protocol adapted from the well-established multiplex cathepsin zymography. Furthermore, culturing these biological machines with E-64, the small molecule cathepsin inhibitor, resulted in a dose-dependent, significant reduction in amount of active cathepsin L bound the muscle strip matrix proteins (Figure 6-2). Culturing biological machines with E-64 did not affect cell viability, MCK activity, and passive tension in muscle strips (Figure 6-3) and showed an increase in average lifespan of bio-bots from 8.2 days to 59.3 days, which was 3.5-fold greater than bio-bots cultured with 1X ACA, which functioned for an average of 17 days. These results further support that cathepsin L proteolysis destabilizes the bio-bot muscle strips. Here, along with some modeling described in Chapter 4, we underscore cathepsin L as an alternative fibrin(ogen)lytic protease, which is playing an important role in engineered

living systems, where fibrin-based constructs are now being exposed to non-traditional proteolytic environments.

In addition to lengthening functional time of these biological machines through inhibiting cysteine cathepsin proteolysis, we also investigated other survival analyses that would better describe life expectancies of biological machines over the Kaplan Meier analysis described in Section 5.3.3. The Weibull distribution is commonly used in reliability engineering to model lifetime distribution of different materials. Applying the Weibull distribution, we determined the Weibull parameters (a, scale parameter and b, shape parameter) for the life expectancy of biological machines cultured in 0X, 1X, and 3X ACA (Figure 6-5). For Weibull distributions, the scale parameter, a, is considered an approximation of the mean time to failure. The shape parameter, b, gives insight into the type of failure, where for b > 1 indicates an increasing failure rate over time, b = 1 indicates a constant failure rate, and b < 1 indicates a decreasing failure rate over time. From the biological machines' Weibull distributions, the shape parameter 0X and 1X ACA cultured muscle strips were 5.552 and 2.227, respectively, which suggested that these muscle strips had an increasing failure rate over time, or "wear-out" failures. However, the 3X ACA cultured muscle strips had a shape parameter of 0.96, which indicated a constant failure rate or random failures. Upon further analysis, it appeared that there was a relationship between Weibull parameters and ACA concentration. While there were only three concentrations of ACA studied and more would be need to better define this relationship, this moves us one step closing to predicting life expectancy of the bio-bots. This

116

relationship could be used to engineer the average lifetime and mechanism of failure of future biological machines. Furthermore, this relationship may also be applied to E-64 culture biological machines, since functional biological machine output appear equivalent for 1X ACA cultured and 5 μM E-64 cultured bio-bots.

Finally, a two-compartment percolation theory model, informed by the proteolytic activity described in Chapter 5 was developed to see if we could re-capitulate the Weibull distribution life expectancies for 0X, 1X, and 3X cultured bio-bots. The two main parameters of the model were $P_{cleavage}$ and P_{broken} thresholds. The model created a 2D graph, where the edges represented the matrix to be cleaved, and at each time point a protease amount was determined at each node. If that protease amount was below the $P_{cleavage}$ threshold, one of the edges of that node was randomly removed. After cycling through all the nodes, the fraction of nodes still connected was computed, if this fraction was above the P_{broken} threshold, the muscle strip went into the next time step, otherwise the failure data was recorded. Two hundred muscle strips were simulated at each $P_{cleavage}$-P_{broken} threshold pair and the Weibull distribution parameters were computed.

A grid search through various $P_{cleavage}$ and P_{broken} thresholds was performed to determine if there was a combination that would recapitulate the Weibull distributions that describe the bio-bot life expectancies (Figure 6-8). While this grid search did not return accurately predict the bio-bot life expectancies, there were some valuable insights learned that can be applied to future iterations of the model. First, the scale parameter (a) was an important driving factor in giving qualitatively good fits to the bio-bot life expectancies.

The scale parameter was also tightly constrained by the $P_{cleavage}$ between 0.1 and 0.2 giving the best fits (Figure 6-9A). Both the scale parameter and the shape parameter (b) appeared to be mostly independent of the P_{broken}, which suggest that this should be held constant in future model iterations. Higher $P_{cleavage}$ thresholds resulted in reduced error for the shape parameter, however qualitatively the fits to the bio-bot life expectancy data was far worse (Figure 6-9B). This model was only informed by the proteolytic activity of these bio-bots, which was based off a 12-day time frame, however, many of these bio-bots (especially when cultured with 1X and 3X ACA) survive much longer and we currently don't have any information on how the protease activity changes on that time scale. During bovine muscle development, there is a drop in cathepsin L activity from day 85 to day 200 [222], which approximated based on comparison of cow and mouse gestation period, would indicated a drop in bio-bot muscle strip cathepsin L activity from day 6 to day 14, just outside 12 day time frame we have characterized. Furthermore, mechanical forces and stresses may play an important role in the lifetime distribution and is another avenue that can be explored in later iterations of this model.

In summary, inhibiting cysteine cathepsin proteolytic activity can lengthen life expectancy of biological machines without negatively effecting cell viability and muscle strip passive tension. We have shown here that cathepsin L produced by these muscle strips has fibrinogenolytic activity, which further supported that cathepsin L destabilized muscle strip integrity. Our findings demonstrate that a Weibull distribution accurately models the life expectancy of bio-bots cultured in different concentrations of ACA and with a

relationship between Weibull parameters and ACA concentration, which can be used to predict biological machine life expectancy for intermediate concentrations of ACA *a priori*.

CHAPTER 7.   CATHEPSIN CANNIBALISM: A REGULATORY MECHANISM OF CATHEPSIN PROTEOLYSIS

This chapter was adapted from Ferrall-Fairbanks, M.C., Kieslich, C.A., and Platt, M.O. Cathepsin cannibalism: A regulatory mechanism of cathepsin proteolysis. Manuscript in preparation.

7.1   Introduction

Traditionally enzymes are investigated by biochemists in isolation and assumed to be inert to the reaction it is catalyzing. However, cells secrete many proteases simultaneously during normal function and disease progression. It is difficult to analyze how these proteases interact cooperatively or antagonistically and researchers have only recently started to investigate these types of interactions [7, 129, 215, 216]. Cysteine cathepsins are a potent group of lysosomal proteases that belong to the papain superfamily of peptidases that are optimally active in acidic and reducing environment with cathepsin K and V being most powerful human collagenase and elastase, respectively [3, 5, 6, 21-24]. Cysteine cathepsins are secreted as procathepsins, with a pro-peptide occluding the active site, which must be cleaved to have mature, active enzyme [3, 21]. Pro-peptide cleavage has been reported to occur through autocatalysis as well as through cleavage by another protease [5, 9, 10, 96].

Of interest are cathepsins K, L, S and V, which share 60% sequence identity and redundancy in target substrate proteins with different catalytic activities toward different

matrix substrates [3, 5, 35]. These cathepsins are upregulated in tissue destructive diseases and reliably predicting perturbations to individual cathepsins on the overall cathepsin proteolytic network is critical to treatment of many of these diseases. Pharmaceutical companies have identified this need and have well-designed cathepsin inhibitors that have proven effective at reducing target substrate degradation. However, all sixteen inhibitors that have made it to phase II and III clinical trials have failed due to side effects, not efficacy in stopping target substrate degradation [42, 120, 223, 224]. A complete network of how these cysteine cathepsins interact with each other, in addition to matrix proteins, is necessary to dose pharmaceutical inhibitors effectively to modulate cysteine cathepsin proteolysis in disease states.

Previously, we have shown that when cathepsins K and S are co-incubated together, the total amount of substrate degradation from the equimolar amounts of cathepsins was less than the sum of the individual cathepsins' proteolytic activity [7]. Computational modeling was used to test hypothetical interactions that could result in reduced substrate degradation. A "cathepsin cannibalism" interaction of cathepsin S preferentially binding and degrading cathepsin K over the substrate was introduced and could accurately capture the amount of substrate degradation [7].

To expand upon our knowledge of the cathepsin proteolytic network, here we developed a mechanistic, ordinary differential equation model characterizing the cathepsin K, L, S, and V proteolytic network with elastin and gelatin substrates. Using a systematic approach to first characterize the kinetics of individual cathepsin on elastin or gelatin, then separating the cathepsin-cathepsin (or cathepsin cannibalism) interactions in this complex

121

proteolytic network. With this cathepsin proteolytic network model, we predicted substrate degradation *a priori* of all four cathepsins at once.

7.2 Materials and Methods

7.2.1 Recombinant cathepsin and procathepsin activation

Recombinant cathepsins K, L, S, and V were used in substrate degradation studies. Mature, active recombinant cathepsin K, L, and S (Enzo) were thawed from -80°C frozen stock. Procathepsin V (Enzo) was purchased with the propeptide intact, which must be cleaved to activate before use. Procathepsin V was activated for 40 minutes at room temperature in 0.1 M sodium acetate buffer at pH 4.0, 1 mM EDTA, and freshly added 2 mM dithiothreitol before use in the time course studies.

7.2.2 Fluorogenic protein degradation

Kinetic studies of fluorogenic substrate degradation, where the substrate fluoresces when cleaved, with different combinations of recombinant cathepsins K, L, S, and V were performed to determine time series substrate degradation over 120 minutes (Figure 7-1). EnzChek® Gelatinase Assay Kit (Invitrogen) was performed in a 96-well plate with each well receiving 50μL of 150μg/mL DQ-gelatin, 5 picomoles of each enzyme, and cathepsin assay buffer (0.1 M sodium phosphate buffer, pH 6.0, 1 mM EDTA, and 2 mM dithiothreitol) used to fill each well to 150μL. Four co-incubation scenarios were examined and compared to the control of substrate only: (1) each cathepsin's gelatinase activity alone, (2) gelatinase activity by cathepsin pairs (i.e. equimolar amounts of cathepsin K and

S, S and L, etc.), (3) gelatinase activity by three cathepsins at once, and (4) gelatinase activity by all four cathepsins at once. These cases were used to determine kinetic rates in the development of the cathepsin proteolytic computational model to determine substrate degradation attributed to each cathepsin. The enzymes were added last to the reaction and then incubated at 37°C in a plate reader. The fluorescence intensity was measured every 5 minutes for 2 hours, yielding 24 measurements, with the digested products having absorption maxima at 495 nm and fluorescence emission maxima at 515nm. This experiment was also repeated using DQ-elastin, since each cathepsin has a different preference for each extracellular matrix component [3, 6, 24].
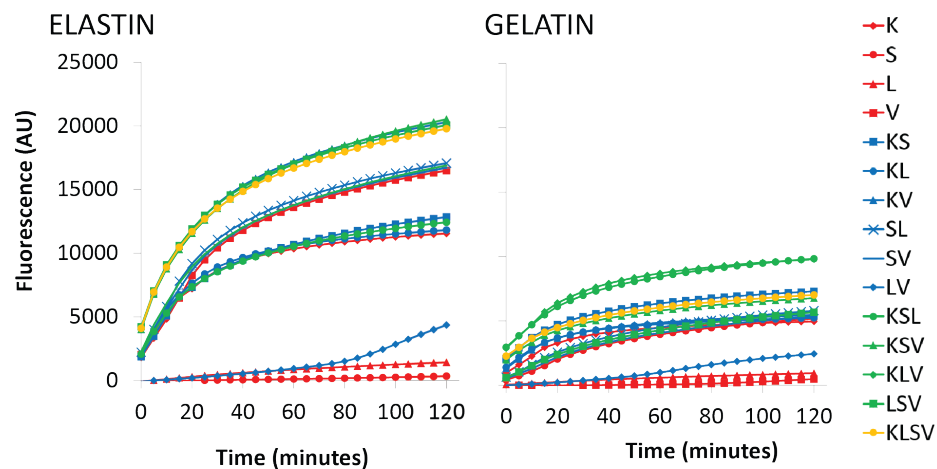
Figure 7-1. Cathepsins K, L, S, and V working in concert do not degrade the most elastin and gelatin. *Kinetic study fluorescence data of cathepsin individuals, pairs, groups of three, and all four together on DQ-elastin and DQ-gelatin. Fluorescence was indicative of degraded substrate.*

123

7.2.3 Kinetic model development

An ordinary differential equation model was developed and implemented with COPASI using general mass action kinetics to mechanistically describe the cathepsin-substrate and cathepsin-cathepsin interactions [225]. The canonical form of general mass action kinetics are described by differential equations of the form shown in Equation 6:

$$\dot{X}_i = \sum_{k=1}^{T_i} \left(\pm \gamma_{ik} \prod_{j=1}^{n+m} X_j^{f_{ikj}} \right) \tag{6}$$

where n is the number of dependent variables, m is the number of independent variables, T_i is the number of terms in the ith equation. For individual cathepsins, mass action kinetics were used to describe the reactions presented in Equations 7 and 8:

$$E + S \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} ES \overset{k_{cat}}{\to} E + P \tag{7}$$

$$E \overset{k_{deg}}{\to} E_{deg} \tag{8}$$

where E represents the enzyme (cathepsin) and S was the substrate (gelatin or elastin). ES was the complex that forms between the enzyme and substrate, which was catalyzed to free enzyme, E, and product formation (degraded substrate), P. There was also a rate at which the enzyme was inactivated over time (k_{deg}). For example, using gelatin (G) as the substrate and cathepsin S (S) as the enzyme, the mass action kinetics to describe the interactions was:

$$\frac{dS}{dt} = -k_{on}[S][G] + k_{off}[SG] + k_{cat}[SG] - k_{deg}[G] \tag{9}$$

$$\frac{dG}{dt} = -k_{on}[S][G] + k_{off}[SG] \tag{10}$$

$$\frac{dSG}{dt} = k_{on}[S][G] - k_{off}[SG] - k_{cat}[SG] \tag{11}$$

$$\frac{dP}{dt} = k_{cat}[SG] \tag{12}$$

$$\frac{dS_{deg}}{dt} = k_{deg}[S] \tag{13}$$

This set of differential equations was completed for each cathepsin (K, L, S, and V) on gelatin and elastin substrates. The kinetic parameters (k_{on}, k_{off}, k_{cat}, k_{deg}) were estimated using the dataset gathered in Section 7.2.2, with the methodology described in Section 7.2.4. The initial conditions for the systems were 25 nM of active and 8nM of inactive enzyme co-incubated with 2.5 µM of substrate.

Once kinetics for each individual cathepsin on each substrate were determined, cathepsin cannibalism interaction terms were introduced with the addition of an equimolar amount of a different cathepsin. Mass action kinetics describe this interaction as shown in Equations 14 and 15:

$$E_1 + E_2 \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} E_1E_2 \overset{k_{cat1}}{\rightarrow} E_1 + E_{2deg} \tag{14}$$

$$E_1 E_2 \xrightarrow{k_{cat2}} E_2 + E_{1deg} \tag{15}$$
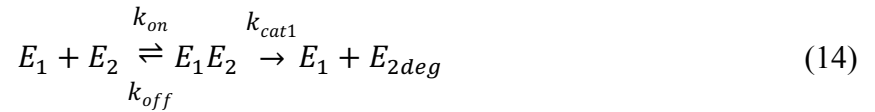
Appropriate terms were included with the differential equations and the cannibalistic rates (four parameters in total for each pair) were fit to the paired cathepsin substrate degradation data. For example, between cathepsins S and K, the cannibalism interactions added terms to the differential equations describing the mature and inactive cathepsin S pools as described as follows in Equations 16 and 17:

$$\frac{dS}{dt} = -k_{on}[S][G] + k_{off}[SG] + k_{cat}[SG] - k_{deg}[G] - k_{on_{cann}}[K][S]$$
$$+ k_{off_{cann}}[SK] + k_{cat_{cannByS}}[SK] \tag{16}$$

$$\frac{dS_{deg}}{dt} = k_{deg}[S] + k_{cat_{cannByK}}[SK] \tag{17}$$

Similar terms were included in the differential equations for the mature and inactive cathepsin K pools. Once the individual kinetics and paired cannibalism kinetic rates have been determined, elastin and gelatin substrate degradations for all four cathepsins at once could be predicted *a priori*.

7.2.4   Parameter estimation

The kinetic parameters of each cathepsin were initially fit using COPASI's genetic algorithm with a population of 50,000 for 500 generations, to the time series data of substrate degradation by individual cathepsins, which is an iterative procedure. Starting values for the optimization were either from literature or from previous iterations of the

126

model. Parameter values were constrained to be within the reported ranges of K_M and k_{cat} values [226]. Furthermore, the on-rate upper bound was set by diffusion limit [226] and the off-rate lower bound was based on biotin-streptavidin affinity, which is the strongest non-covalent biological interaction known [227, 228].

$$10^{-6}s^{-1}pM^{-1} \leq k_{on} \leq 10^{-3}s^{-1}pM^{-1} \tag{18}$$

$$10^{-5}s^{-1} \leq k_{off} \leq 10^{2}s^{-1} \tag{19}$$

$$0 \leq k_{cat} \leq 10^{3}s^{-1} \tag{20}$$

Once solutions were determined from the genetic algorithm, parameters were further refined and fit across both substrates simultaneously using a surrogate based optimization method [229] to find parameters that fit both elastin and gelatin data comparably. The objective function optimized was the sum of normalized errors as described in Equation 21:

$$error = \sum_{t=1}^{24} \frac{\left(exp_{elastin,t} - pred_{elastin,t}\right)^2}{\left(exp_{elastin,t}\right)^2} + \sum_{t=1}^{24} \frac{\left(exp_{gelatin,t} - pred_{gelatin,t}\right)^2}{\left(exp_{gelatin,t}\right)^2} \tag{21}$$

where *exp* is the experimental amount of substrate degradation (either degraded elastin or gelatin) and *pred* is the predicted amount of substrate degradation. Then, the cannibalistic terms were fit to the paired combinations of cathepsins. The cases with three and four cathepsins allowed for *a priori* confirmation of these cannibalistic relationships. Additional *in silico* scenarios were created by the computational model and the predictions were validated with kinetic studies designed based on the *in silico* simulations.

One challenge in parameter estimation and optimization is that it is difficult to confirm the solution returned is the true global minimum of the objective function and best solution. To guide the parameter estimation to find the optimal parameter set, the cathepsin proteolytic network parameters were fit systematically, compared with experimental data, and the parameters were validated with *a priori* predictions. Furthermore, bounds grounded by the best known binding affinity for biotin-streptavidin, reported ranges for K_M and k_{cat}, and the diffusion limit were used to constrain the parameter space and a two-step optimization approach was implemented with a genetic algorithm first sweeping the parameter space followed by a surrogate-based optimization to converge on an optimal solution. Parameters for cathepsin-cathepsin interactions were also constrained to agree across substrates. These cathepsin-cathepsin interactions were assumed to be inherent binding affinities of the enzymes.

The cathepsin-substrate binding and catalysis rates determine to what extent these cathepsin-cathepsin (cannibalism) interactions modulate the amount of active enzyme in the system. If the cathepsin-substrate affinities were greater than the cathepsin cannibalism interactions, then the cannibalism interactions were not play a large role. If the cathepsin-substrate affinities were less than the cannibalism interactions, then the cathepsin cannibalism interactions play a larger role in reducing the amount of cathepsin with potential to be catalytically active toward substrate.

7.3 Results

7.3.1 Stepwise determination of cathepsin K, L, and S reaction kinetics

Mass action kinetics are the basis for mathematical models of individual cathepsins K, L, and S degradation of elastin and collagen, two major mammalian extracellular matrix proteins. Gelatin is a denatured form of collagen and was used here to represent collagen in proteolytic network development. To develop a mechanistic model of the cathepsin proteolytic network, parameters were systematically estimated for a traditional Michaelis-Menten enzyme kinetics system with one enzyme and one substrate as described by Equation 7. Kinetic rates were estimated first by using a genetic algorithm with parameters bound by known physical and biology limits as described in Equations 18-20. Using cathepsin S as a model cathepsin, Michaelis-Menten kinetics alone (Model 1) could not accurately predict substrate degradation of an individual cathepsin on elastin or gelatin (Figure 7-2A, B, C; Model 1 predictions).

Michaelis-Menten kinetics alone over-predicted the proteolytic activity of cathepsins, so an inactivation rate was introduced as described by the reaction in Equation 8. Again, the kinetic rates were first estimated independently using a genetic algorithm, then further refined using a surrogate-based optimization technique, where the inactivation rate was held constant over the two substrates. The inactivation rate was fit as power law interactions as described in Equation 22 with two parameters: kinetic rate, k_{deg}, and kinetic order, g.

$$\frac{d[cathepsin_{deg}]}{dt} = k_{deg}[cathepsin]^g \tag{22}$$

The model with Michaelis-Menten kinetics and an inactivation rate (Model 2) reduced the objective normalized error for cathepsin S 28.5-fold (from 5.80×10^{-2} to 2.04×10^{-3}; Figure 7-2A, B, C; Model 2 prediction) compared to the Michaelis-Menten kinetics only case (Model 1). However, the prediction of Model 2 for the amount of inactivated cathepsin S was that 95% of the cathepsin S was inactivated within the first five minutes (Figure 7-2D; Module 2 prediction), which was not biologically reasonable. Cathepsin S activity has been reported to be dependent on pH, however varying from a pH 5.0 to 7.5, still have active cathepsin S after five minutes [96]. Furthermore, intracellularly, mature cathepsin S has been reported to have a half-life of 16-18 hours [11], which would also suggest cathepsin S would remain active for more than five minutes. This suggested that there may be another interaction to explain that cathepsin S did not immediately bind to substrate. We hypothesized a distraction interaction, defined as cathepsins reversibly binding to an inactive cathepsins; this is described by the reaction for cathepsin S reversibly binding to inactive cathepsin S (S_{deg}) in Equation 23, which adjusted the cathepsin S differential equations shown in Equation 24.

$$S + S_{deg} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} [SS_{deg}] \tag{23}$$

$$\frac{dS}{dt} = -k_{on}[S][G] + k_{off}[SG] + k_{cat}[SG] - k_{deg}[G] - k_{on_{inact}}[S][S_{deg}] \quad (24)$$

$$+ k_{off_{inact}}[SS_{deg}]$$

Incorporating these distraction terms into Model 3 further reduced the normalized error 2.9-fold (from 2.04×10^{-3} to 6.99×10^{-4}; Figure 7-2A, B, C; Model 3 prediction) for substrate degradation by cathepsin S and reduced the amount of inactive cathepsin S (Figure 7-2D; Model 2 to Model 3 prediction). This model structure also best describes cathepsins K and L on elastin and gelatin. The full list of equations with parameters for all three cathepsins on both substrates were reported in reactions 1 through 18 in Table 7-1 and corresponding parameter estimation errors were reported in Table 7-2.
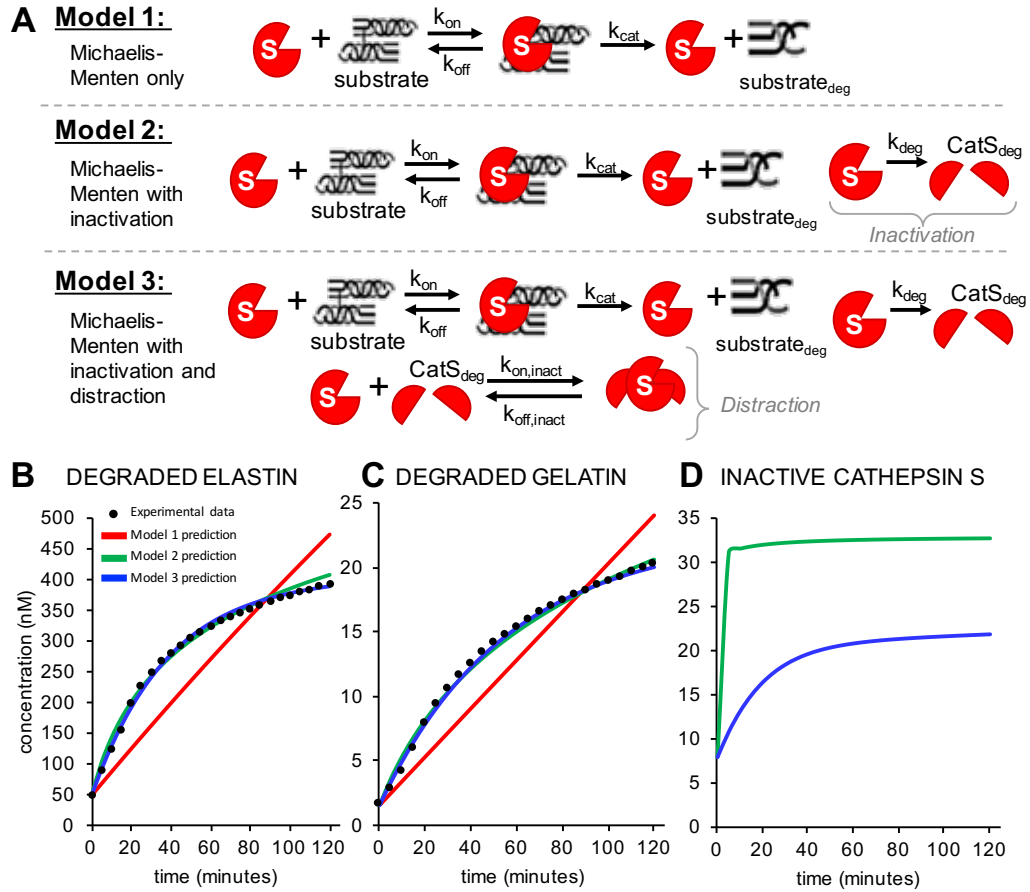
Figure 7-2. Enzyme kinetics alone was not enough to predict substrate degradation.
(A) Three model structures were evaluated to describe cathepsin S degradation of substrate. Model 1 used Michaelis-Menten kinetics only to describe substrate degradation, Model 2 included an inactivation rate of cathepsin S, and Model 3 included a distraction term of active cathepsin S reversible binding to inactive cathepsin S. Using cathepsin S as the model enzyme, the black solid points indicate the experimental degraded elastin (B) or degraded gelatin (C) over 120 minutes. Model 1 prediction (red solid line) indicates predicted substrate degradation of a Michaelis-Menten model of cathepsin S and substrate. Model 2 (green solid line) indicates the predicted substrate degradation of Michaelis-Menten kinetics model with addition of an enzyme inactivation rate. Model 3 (blue solid line) indicates a Michaelis-Menten kinetics model with an enzyme inactivation rate and a "distraction" for free enzyme to bind and un-bind to inactive enzyme. (D) In the Michaelis-Menten plus inactivation model (Model 2, green), the cathepsin S is almost all inactivated more quickly than would be expected biologically. With the addition of the distraction terms, allowing free enzyme to bind to the inactive form, (Model 3, blue), active cathepsin S remains in the system longer.

Once the individual cathepsin kinetics were established, a second cathepsin was incorporated into the proteolytic network model. Using cathepsins S and K, Model 1 described the superposition of each individual cathepsin's proteolytic activity (Figure 7-3A, B; Model 1 prediction) and predicted substrate degradation was 1.4-fold greater than the experimental data (Figure 7-3A, B; distinct points). This led to the hypothesis that there must be some interaction that reduced the amount of active enzyme available to bind to substrate. First, we introduced cathepsin-cathepsin (cathepsin cannibalism) interactions with the model, since this concept had been previously established, described by reactions in Equations 16 and 17. Cross-distraction interactions were also included in the model, defined here for the cathepsins S and K example, where cathepsin K can bind to inactive cathepsin S (S_{deg}) and cathepsin S can bind to inactive cathepsin K (K_{deg}). Parameter estimation was performed in the same two-step fashion as described above and all the rates were considered inherent binding affinity of the cathepsins and held constant across substrates. Incorporating these cathepsin cannibalism interactions reduced the normalized error compared to the case predicted by superposition (Figure 7-3A, B; Model 1 compared to Model 2 prediction) almost 410-fold for the cathepsin S and K pair.
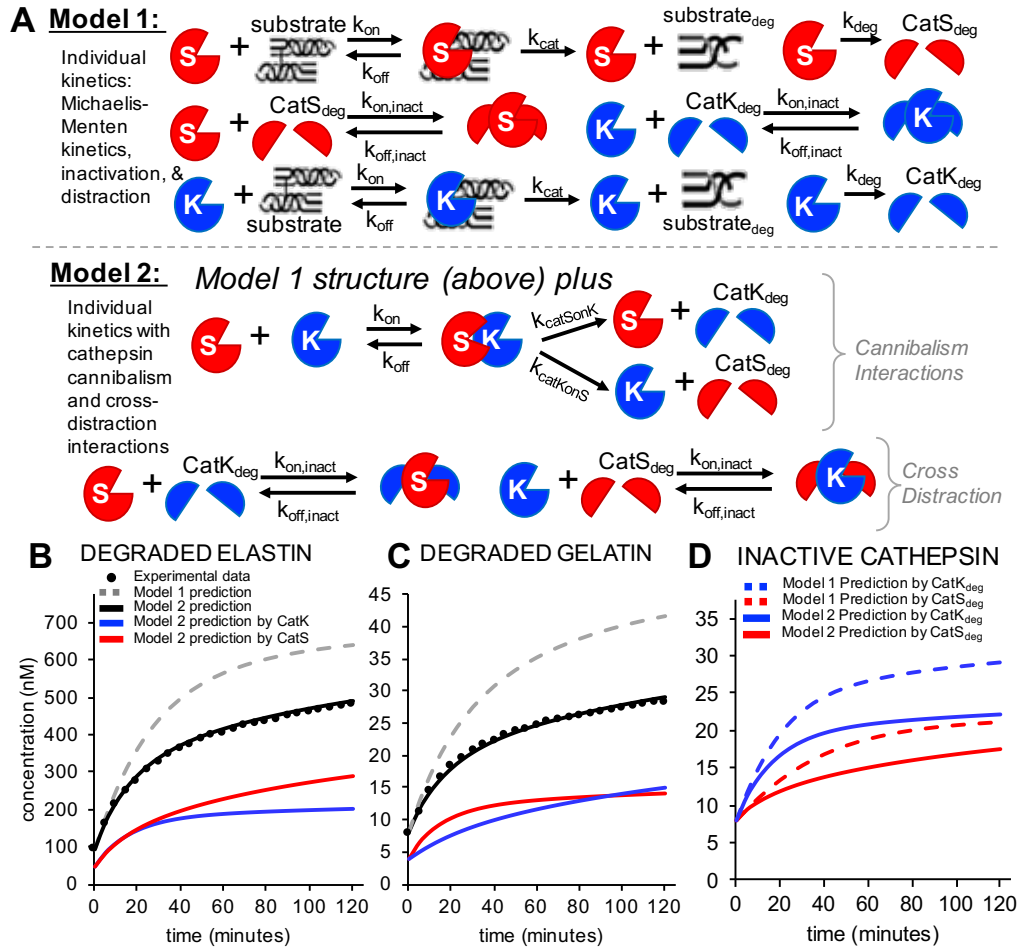
Figure 7-3. Cathepsin-cathepsin cannibalism interactions accurately described paired cathepsin substrate degradation. *(A) Two model structures were used to describe two cathepsins degrading substrate. Model 1 assumed superposition of individual cathepsin kinetics using the structure defined in Figure 7-2 and Model 2 incorporated cathepsin-cathepsin interactions (cathepsin cannibalism and cross-distractions). Using the cathepsin K (catK) and S (catS) as the model paired interaction, the black solid points indicate experimental data for degraded elastin (B) or degraded gelatin (C) over 120 minutes. Model 1 prediction (gray dashed line) indicated the expected substrate degradation assuming superposition of catK and S individual kinetics, which over predicted the experimental data for catK and S degrading substrate. Model 2 prediction (black model line) indicated the substrate degradation when the model included cathepsin cannibalism interactions between catK and S. Model 2 substrate degradation contribution by catK (blue) and catS (red) indicated the contribution of each cathepsin to the total substrate degradation model (black). (D) Total amount of inactive catK (blue) and catS (red) compared for Model 1 (dashed lines) and Model 2 (solid lines).*

134

Table 7-1. Cathepsin cannibalism reactions, rate laws, and kinetic constant values.

Name	Reaction	Rate Law	Parameters	No.
CatK binding elastin	$K + E \leftrightarrow KE$	Mass action	$k_{on} = 4.844\text{e-}4$ $k_{off} = 0.820$	1
CatK catalyzing elastin	$KE \rightarrow K + E_{deg}$	Mass action	$k_{cat} = 6.362\text{e-}3$	2
CatK inactivation	$K \rightarrow K_{inact}$	Mass action	$k_{deg} = 0.792$	3
CatK binding gelatin	$K + G \leftrightarrow KG$	Mass action	$k_{on} = 8.364\text{e-}5$ $k_{off} = 0.163$	4
CatK catalyzing gelatin	$KG \rightarrow K + G_{deg}$	Mass action	$k_{cat} = 4.592\text{e-}4$	5
Inactive CatK distraction	$K + K_{deg} \leftrightarrow KK_{deg}$	Mass action	$k_{on} = 2.037\text{e-}5$ $k_{off} = 2.025\text{e-}4$	6
CatL binding elastin	$L + E \leftrightarrow LE$	Mass action	$k_{on} = 3.915\text{e-}4$ $k_{off} = 0.993$	7
CatL catalyzing elastin	$LE \rightarrow L + E_{deg}$	Mass action	$k_{cat} = 3.604\text{e-}4$	8
CatL inactivation	$L \rightarrow L_{inact}$	Mass action with kinetic order	$k_{deg} = 7.210\text{e-}3$ $g = 2$	9
CatL binding gelatin	$L + G \leftrightarrow LG$	Mass action	$k_{on} = 9.858\text{e-}6$ $k_{off} = 0.0281$	10
CatL catalyzing gelatin	$LG \rightarrow L + G_{deg}$	Mass action	$k_{cat} = 9.961\text{e-}5$	11
Inactive CatL distraction	$L + L_{deg} \leftrightarrow LL_{deg}$	Mass action	$k_{on} = 2.679\text{e-}6$ $k_{off} = 2.553\text{e-}5$	12
CatS binding elastin	$S + E \leftrightarrow SE$	Mass action	$k_{on} = 9.930\text{e-}4$ $k_{off} = 0.603$	13
CatS catalyzing elastin	$SE \rightarrow S + E_{deg}$	Mass action	$k_{cat} = 6.013\text{e-}3$	14
CatS inactivation	$S \rightarrow S_{inact}$	Mass action	$k_{deg} = 0.861$	15
CatS binding gelatin	$S + G \leftrightarrow SG$	Mass action	$k_{on} = 1.896\text{e-}6$ $k_{off} = 8.554\text{e-}4$	16
CatS catalyzing gelatin	$SG \rightarrow S + G_{deg}$	Mass action	$k_{cat} = 3.357\text{e-}4$	17
Inactive CatS distraction	$S + S_{deg} \leftrightarrow SS_{deg}$	Mass action	$k_{on} = 8.797\text{e-}5$ $k_{off} = 1.625\text{e-}5$	18
CatK and CatL complexing	$K + L \leftrightarrow KL$	Mass action	$k_{on} = 6.642\text{e-}4$ $k_{off} = 5.500\text{e-}4$	19
CatK catalyzing CatL	$KL \rightarrow K + L_{deg}$	Mass action	$k_{cat} = 500$	20
CatL catalyzing CatK	$KL \rightarrow L + K_{deg}$	Mass action	$k_{cat} = 336.16$	21
Inactive CatK distraction of CatL	$K + L_{deg} \leftrightarrow KL_{deg}$	Mass action	$k_{on} = 1.638\text{e-}5$ $k_{off} = 1.000\text{e-}4$	22
Inactive CatL distraction of CatK	$L + K_{deg} \leftrightarrow LK_{deg}$	Mass action	$k_{on} = 5.000\text{e-}5$ $k_{off} = 500$	23
CatK and CatS complexing	$K + S \leftrightarrow KS$	Mass action	$k_{on} = 7.120\text{e-}4$ $k_{off} = 73.03$	24
CatK catalyzing CatS	$KS \rightarrow K + S_{deg}$	Mass action	$k_{cat} = 710.24$	25
CatS catalyzing CatK	$KS \rightarrow S + K_{deg}$	Mass action	$k_{cat} = 288$	26
Inactive CatS distraction of CatK	$K + S_{deg} \leftrightarrow KS_{deg}$	Mass action	$k_{on} = 1.000\text{e-}3$ $k_{off} = 9.211\text{e-}5$	27
Inactive CatK distraction of CatS	$S + K_{deg} \leftrightarrow SK_{deg}$	Mass action	$k_{on} = 3.123\text{e-}4$ $k_{off} = 1.000\text{e-}4$	28
CatL and CatS complexing	$L + S \leftrightarrow LS$	Mass action	$k_{on} = 1.000\text{e-}6$ $k_{off} = 5.500\text{e-}4$	29
CatL catalyzing CatS	$LS \rightarrow L + S_{deg}$	Mass action	$k_{cat} = 500$	30
CatS catalyzing CatL	$LS \rightarrow S + L_{deg}$	Mass action	$k_{cat} = 500$	31
Inactive CatL distraction of CatS	$L + S_{deg} \leftrightarrow LS_{deg}$	Mass action	$k_{on} = 9.373\text{e-}6$ $k_{off} = 1.000\text{e-}4$	32
Inactive CatS distraction of CatL	$S + L_{deg} \leftrightarrow SL_{deg}$	Mass action	$k_{on} = 5.000\text{e-}5$ $k_{off} = 500$	33

Table 7-2. Parameter estimation kinetic constant value errors.

Name	Reactions	Parameters Fit	Error
CatK only reactions	$K + E \leftrightarrow KE \to K + E_{deg}$ $K + G \leftrightarrow KG \to K + G_{deg}$ $K \to K_{inact}$ $K + K_{deg} \leftrightarrow KK_{deg}$	9 parameters Reactions 1-6	2.32e-4
CatL only reactions	$L + E \leftrightarrow LE \to L + E_{deg}$ $L + G \leftrightarrow LG \to L + G_{deg}$ $L \to L_{inact}$ $L + L_{deg} \leftrightarrow LL_{deg}$	9 parameters Reactions 7-12	1.60e-4
CatS only reactions	$S + E \leftrightarrow SE \to S + E_{deg}$ $S + G \leftrightarrow SG \to S + G_{deg}$ $S \to S_{inact}$ $S + S_{deg} \leftrightarrow SS_{deg}$	9 parameters Reactions 13-18	5.62e-4
CatK and CatL cannibalism reactions	$K + L \leftrightarrow KL$ $KL \to K + L_{deg}$ $KL \to L + K_{deg}$ $K + L_{deg} \leftrightarrow KL_{deg}$ $L + K_{deg} \leftrightarrow LK_{deg}$	8 parameters Reactions 19-23	8.48e-4
CatK and CatS cannibalism reactions	$K + S \leftrightarrow KS$ $KS \to K + S_{deg}$ $KS \to S + K_{deg}$ $K + S_{deg} \leftrightarrow KS_{deg}$ $S + K_{deg} \leftrightarrow SK_{deg}$	8 parameters Reactions 24-28	2.71e-4
CatL and CatS cannibalism reactions	$L + S \leftrightarrow LS$ $LS \to L + S_{deg}$ $LS \to S + L_{deg}$ $L + S_{deg} \leftrightarrow LS_{deg}$ $S + L_{deg} \leftrightarrow SL_{deg}$	8 parameters Reactions 29-33	1.19e-3

7.3.2 *Cathepsin cannibalism results in* a priori *determination of substrate degradation*

Once paired cathepsin cannibalism interactions have been established, combining the individual kinetics and paired interactions for cathepsins K, L, and S predicted the substrate degradation of all three cathepsins in combination (Figure 7-4). For elastin, the error (calculated as described by Equation 21) was reduced from 1.840×10^{-1} for Model 1 with only the individual cathepsin kinetics, to 3.334×10^{-3} with the inclusion of

cathepsin cannibalism interactions in Model 2, a 55-fold reduction in the error (Figure 7-4A). Gelatin had a similar reduction in error from 2.426×10^{-1} to 2.114×10^{-2}, an 11.5-fold reduction in error (Figure 7-4B).
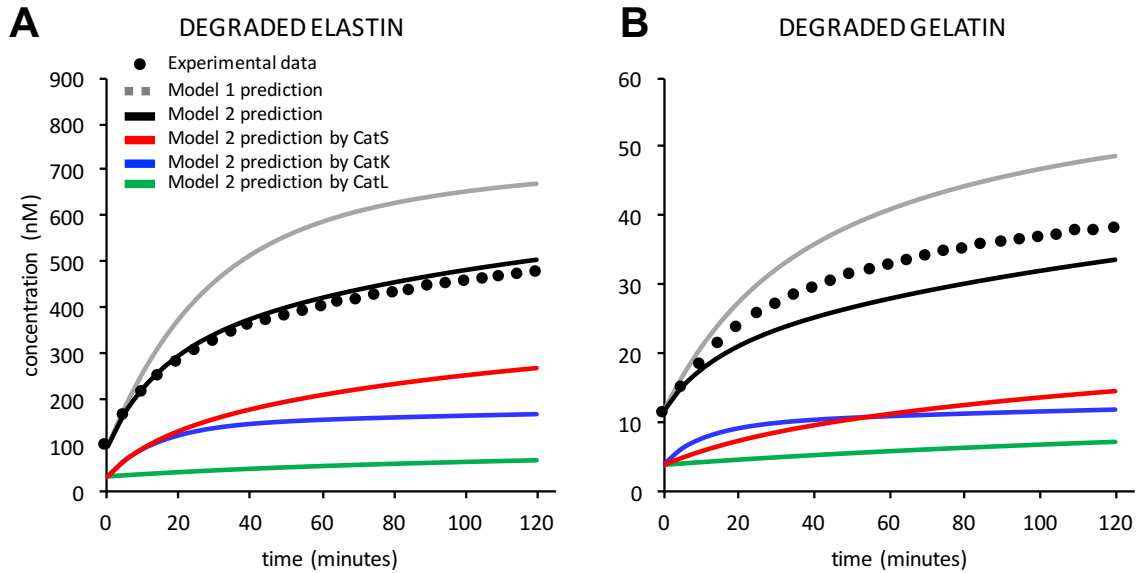
Figure 7-4. Cathepsin-cathepsin cannibalism *a priori* predicted substrate degradation by cathepsins K, L, and S. *Inclusion of cathepsin-cannibalism terms in the model predicted substrate degradation* a priori. *Two model structures were used to describe two cathepsins degrading substrate as described in Figure 7-3A. Model 1 assumed superposition of individual cathepsin kinetics and Model 2 incorporated cathepsin-cathepsin interactions (cathepsin cannibalism and cross-distractions). The black solid points indicated experimental data of degraded elastin (A) or degraded gelatin (B) over 120 minutes. Model 1 (gray dashed curve) indicated the expected substrate degradation assuming superposition of cathepsin K, L, and S individual kinetics (with no cannibalism interactions), which over-predicts the amount of substrate degradation by the three cathepsins. Model 2 (black solid curve) indicated substrate degradation when the model included cathepsin cannibalism interactions between pairs of cathepsins K, L, and S. The Model 2 degraded substrate contribution from cathepsin K (blue), cathepsin L (green), and cathepsin S (red) sums to the total substrate degradation predicted by Model 2 (black curve).*

7.3.3   Procathepsin V activation was necessary for determining amount of active enzyme

Unlike cathepsins K, L, and S, recombinant cathepsin V is only available for purchase as the inactive procathepsin form. Five picomoles of procathepsin V was activated for 40 minutes in a pH 4.0 sodium acetate buffer prior to use in DQ-substrate kinetic studies. Attempts to fit individual cathepsin V kinetics on elastin and gelatin were unsuccessful within parameter bounds when assuming all the procathepsin had been activated. However, assuming a mixed population of mature, active cathepsin V and procathepsin V that would continue to activate in the pH 6.0 kinetic study, drastically improved the individual kinetic rate fits, but the ratio of pro- to mature cathepsin V must be determined as well as activation rates. A procathepsin V activation kinetic study was run to determine (1) the distribution of mature, procathepsin, and degraded cathepsins and (2) activation rates for cathepsin V at pH 4.0 and pH 6.0 (Figure 7-5A). This kinetic study was performed with DQ-elastin as the substrate to determine the activation rates at both pH 4.0 and pH 6.0 for the following reactions in Equations 25 and 26:

$$proV + V \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} VproV \overset{k_{act,bi}}{\rightarrow} V + V \tag{25}$$

$$proV \overset{k_{act,uni}}{\rightarrow} V \tag{26}$$
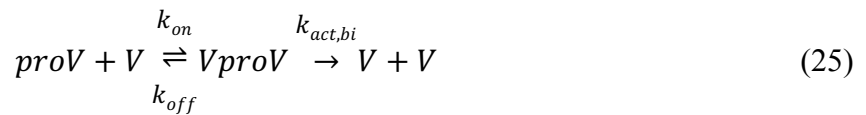
Both autocatalytic and bimolecular activation occur during cathepsin activation, so both were included in the model [10]. Models fit with only one mechanism of activation, either

only autocatalytic or only biomolecular activation, had larger errors compared to the model with both autocatalytic and bimolecular activation. Once rates were determined, a simulation of pro- to mature cathepsin V activation at pH 4.0 in the absence of substrate was run to determine the initial conditions for the cathepsin V samples from the larger KLSV experiment. Surprisingly, in the absence of substrate during activation of cathepsin V, the simulation had 10-fold less mature, active cathepsin V. Using an initial amount of 225 pM of free, mature cathepsin V, 2400 pM mature cathepsin V bound to procathepsin V, 4 pM of procathepsin V, and 30,371 pM of inactive cathepsin V as well as the activation rates for procathepsin V at pH 6 and the model structure established by cathepsins K, L, and S, the individual kinetics for cathepsin V on elastin and gelatin were determined (Figure 7-5B, C, Table 7-3 reactions 34-41, Table 7-4).
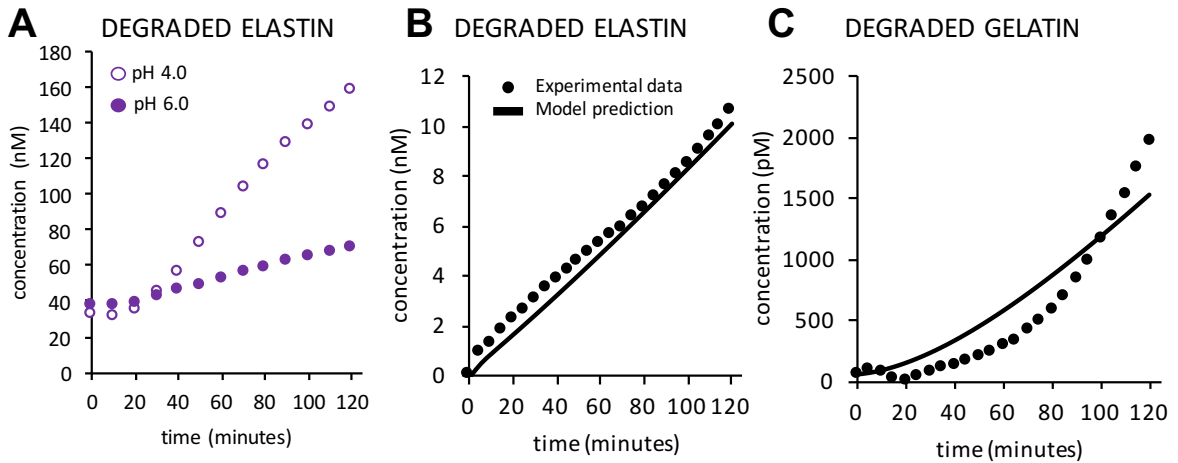
Figure 7-5. Procathepsin V activation was necessary to predict substrate degradation. *Procathepsin V activation rates were determined and compared between pH 4.0 and pH 6.0 buffers. (A) Procathepsin V activation was performed in the presence of DQ-elastin and showed that more cathepsin V was activated at pH 4.0, compared to pH 6.0, as reported by the total amount of elastin degraded after 120 minutes. After activation rates were determined, a simulation without substrate was run to determine the initial concentration of procathepsin V, mature cathepsin V, inactive cathepsin V, and complexes with cathepsin V. With these initial conditions and allowing procathepsin V to activate during the pH 6.0 experimental, individual cathepsin V kinetics were determined as demonstrated with the degradation of elastin (B) and gelatin (C) over 120 minutes, where the black solid points indicated experimental data and the black curve indicated the model predicted cathepsin V substrate degradation. Note: All cathepsin V parameter estimation was completed using a genetic algorithm with a population size of 10,000 and parameters will be further refined before manuscript is submitted for publication.*

Table 7-3. Cathepsin V reactions, rate laws, and kinetic constant values.

Name	Reaction	Rate Law	Parameters	No.
CatV autoactivation	$proV \rightarrow V$	Mass action	$k_{act} = 1.347e\text{-}5$	34
CatV binding proV	$proV + V \leftrightarrow VproV$	Mass action	$k_{on} = 6.664e\text{-}4$ $k_{off} = 0.01764$	35
CatV activating proV	$VproV \rightarrow V + V$	Mass action	$k_{act} = 0.01871$	36
CatV binding elastin	$V + E \leftrightarrow VE$	Mass action	$k_{on} = 9.033e\text{-}6$ $k_{off} = 0.007425$	37
CatV catalyzing elastin	$VE \rightarrow V + E_{deg}$	Mass action	$k_{cat} = 0.001607$	38
CatV binding gelatin	$V + G \leftrightarrow VG$	Mass action	$k_{on} = 1e\text{-}6$ $k_{off} = 1e\text{-}5$	39
CatV catalyzing gelatin	$VG \rightarrow V + G_{deg}$	Mass action	$k_{cat} = 1.119e\text{-}4$	40
Inactive CatV distraction	$V + V_{deg} \leftrightarrow VV_{deg}$	Mass action	$k_{on} = 0.001$ $k_{off} = 0.002607$	41
CatL binding proV	$L + proV \leftrightarrow LproV$	Mass action	$k_{on} = 0.001$ $k_{off} = 2.508e\text{-}4$	42
CatL activating proV	$LproV \rightarrow L + V$	Mass action	$k_{act} = 43.60$	43
CatV and CatL complexing	$V + L \leftrightarrow VL$	Mass action	$k_{on} = 0.001$ $k_{off} = 4.182$	44
CatV catalyzing CatL	$VL \rightarrow V + L_{deg}$	Mass action	$k_{cat} = 0$	45
CatL catalyzing CatV	$VL \rightarrow L + V_{deg}$	Mass action	$k_{cat} = 0$	46
Inactive CatL distraction of CatV	$V + L_{deg} \leftrightarrow VL_{deg}$	Mass action	$k_{on} = 1.208e\text{-}6$ $k_{off} = 100$	47
Inactive CatV distraction of CatL	$L + V_{deg} \leftrightarrow LV_{deg}$	Mass action	$k_{on} = 0.001$ $k_{off} = 1.092$	48
CatV and CatK complexing	$V + K \leftrightarrow VK$	Mass action	$k_{on} = 9.095e\text{-}4$ $k_{off} = 4.597e\text{-}5$	49
CatV catalyzing CatK	$VK \rightarrow V + K_{deg}$	Mass action	$k_{cat} = 1.653$	50
CatK catalyzing CatV	$VK \rightarrow K + V_{deg}$	Mass action	$k_{cat} = 0$	51
Inactive CatK distraction of CatV	$V + K_{deg} \leftrightarrow VK_{deg}$	Mass action	$k_{on} = 0.001$ $k_{off} = 0.001281$	52
Inactive CatV distraction of CatK	$K + V_{deg} \leftrightarrow KV_{deg}$	Mass action	$k_{on} = 1.048e\text{-}6$ $k_{off} = 1e\text{-}5$	53
CatV and CatS complexing	$V + S \leftrightarrow VS$	Mass action	$k_{on} = 0.001$ $k_{off} = 6.186e\text{-}5$	54
CatV catalyzing CatS	$VS \rightarrow V + S_{deg}$	Mass action	$k_{cat} = 0.1448$	55
CatS catalyzing CatV	$VS \rightarrow S + V_{deg}$	Mass action	$k_{cat} = 0$	56
Inactive CatS distraction of CatV	$V + S_{deg} \leftrightarrow VS_{deg}$	Mass action	$k_{on} = 0.001$ $k_{off} = 1e\text{-}5$	57
Inactive CatV distraction of CatS	$S + V_{deg} \leftrightarrow SV_{deg}$	Mass action	$k_{on} = 4.205e\text{-}6$ $k_{off} = 1e\text{-}5$	58

Table 7-4. Cathepsin V parameter estimation kinetic constant values errors.

Name	Reactions	Parameters Fit	Error
CatV only reactions	$V + E \leftrightarrow VE \rightarrow V + E_{deg}$ $V + G \leftrightarrow VG \rightarrow V + G_{deg}$ $V + V_{deg} \leftrightarrow VV_{deg}$ $V + proV \leftrightarrow VproV \rightarrow 2V$ $proV \rightarrow V$	12 parameters Reactions 34-41	0.0821
CatV and CatL cannibalism reactions	$V + L \leftrightarrow VL$ $VL \rightarrow V + L_{deg}$ $VL \rightarrow L + V_{deg}$ $V + L_{deg} \leftrightarrow VL_{deg}$ $L + V_{deg} \leftrightarrow LV_{deg}$ $L + proV \leftrightarrow LproV$ $\rightarrow L + V$	11 parameters Reactions 42-48	0.282
CatV and CatK cannibalism reactions	$V + K \leftrightarrow VK$ $VK \rightarrow V + K_{deg}$ $VK \rightarrow K + V_{deg}$ $V + K_{deg} \leftrightarrow VK_{deg}$ $K + V_{deg} \leftrightarrow KV_{deg}$	8 parameters Reactions 49-53	2.36e-3
CatV and CatS cannibalism reactions	$V + S \leftrightarrow VS$ $VS \rightarrow V + S_{deg}$ $VS \rightarrow S + V_{deg}$ $V + S_{deg} \leftrightarrow VS_{deg}$ $S + V_{deg} \leftrightarrow SV_{deg}$	8 parameters Reactions 54-58	6.19e-4

Similarly, to the pairs of cathepsins K, L, and S, the paired cathepsin cannibalism interactions with cathepsin V were determined, while holding the individual cathepsin kinetic rates (degradation, catalysis, inactivation, and distraction) constant. The sum of each cathepsin's individual proteolytic activity (Figure 7-6A, B; Model 1 prediction) was slightly greater than the experimental data (Figure 7-6A, B; experimental data). However, the difference in the substrate degradation between the superposition model (Model 1) and experimental data with the cathepsin V pairs was less than those for cathepsins K, L, and S pairs. Compared to the other cathepsin pairs, this reduced difference between superposition model and experimental data was most likely because cathepsin V had only

about 10% the total active amount of enzyme, compared to the other cathepsins. In all cases, the superposition model over-predicts the substrate degradation compared to those models including cathepsin-cathepsin interactions. Parameter estimation was performed in the two-step fashion as described above and all the rates, except for individual substrate kinetics, were held constant across the two substrates. The model incorporating these cathepsin cannibalism and cross-distraction interactions (allowing active cathepsin to bind to the inactive form) reduced the normalized error (Equation 21) compared to the superposition predicted case (Figure 7-6A, B; Model 1 to Model 2 prediction) 4.5-fold for the example of cathepsin V and S pair (other cathepsin V paired errors are recorded in Table 7-4).
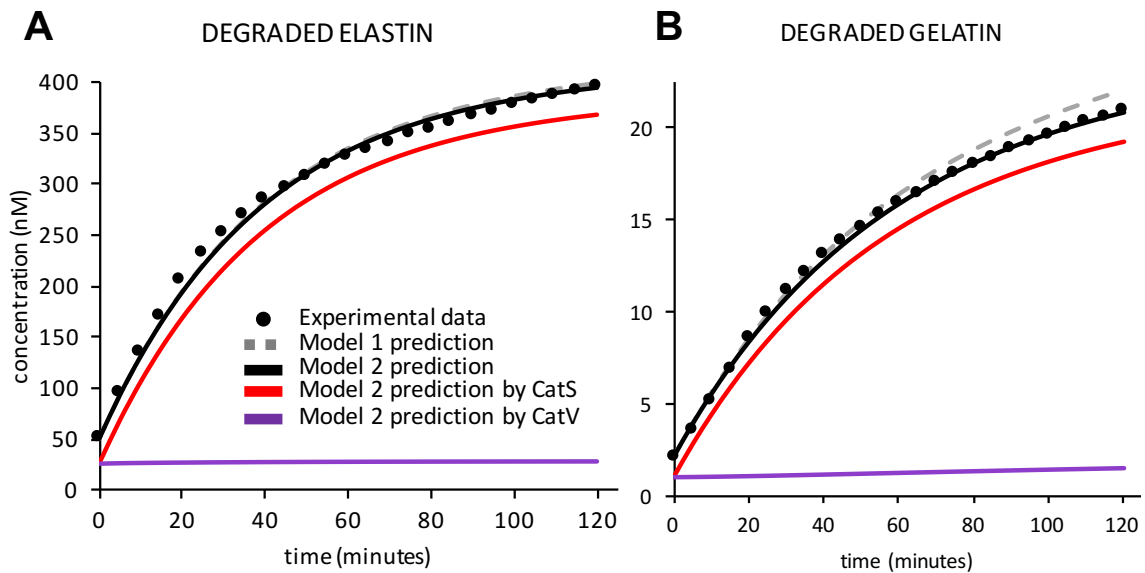
Figure 7-6. Cathepsin V cannibalism interactions predicted paired substrate degradation. *Using the cathepsin V and S as the model paired interaction, the black solid points indicated experimental data of degraded elastin (A) or degraded gelatin (B) over 120 minutes. Model 1 (gray dashed curve) indicated the predicted substrate degradation assuming superposition of cathepsin V and S individual kinetics. Model 1 over predicts*

substrate degradation, which was most evident on the gelatin substrate. Model 2 (black curve) indicated the predicted substrate degradation when the model included cathepsin cannibalism interactions between cathepsin V and S. The cathepsin V (purple) and cathepsin S (red) contribution to the total substrate degradation predicted by Model 2 (black curve). Most the substrate degradation was predicted to be due to cathepsin S catalytic activity, which may be because there was initially 10-fold more active cathepsin S compared to cathepsin V.
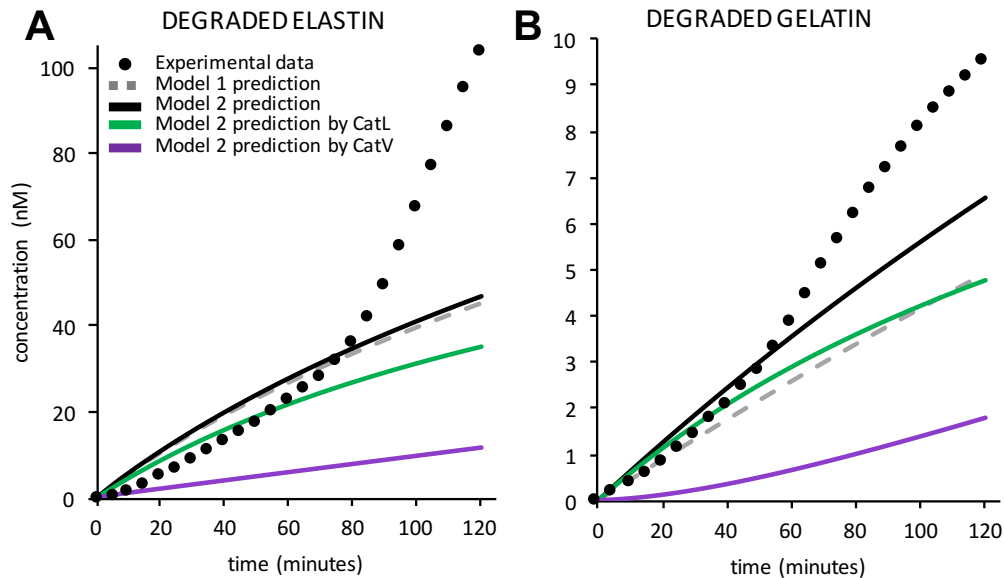
Figure 7-7. Cathepsin L also activated procathepsin V in paired interactions. *Cathepsin V and L paired interaction behaves differently compared to other cathepsin pairs, as depicted by the experimental data (black solid points) of degraded elastin (A) or degraded gelatin (B) over 120 minutes. The increasing nature of the data lead to the inclusion of cathepsin L activation of procathepsin V, ultimately reducing the error between the model system and the data. Two model structures were used to describe two cathepsins degrading substrate. Model 1 (gray dashed curve) assumed superposition of individual cathepsin kinetics using the structure defined in Figure 7-2 and Model 2 (black solid curve) incorporated cathepsin-cathepsin interactions (cathepsin cannibalism and cross-distractions) as well as cathepsin L reversible binding and activating procathepsin V. Inclusion of cathepsin L activation of procathepsin V activation (Model 2), improved the fit 1.52-fold compared to the superposition model (Model 1). The contributions of cathepsin V (purple) and cathepsin L (green) to the total substrate degradation predicted by Model 2 (black). These contributions differed compared to other paired interactions because here both cathepsins were hypothesized to continue degradation additional substrate after the 120 minutes instead of saturating as seen in the cathepsin S and K pair (Figure 7-3).*

The cathepsin V and L pair data displayed different dynamics compared to all the other cathepsin pairs. While most cathepsin pairs show a saturation of substrate degradation prior to 120 minutes, the cathepsins V and L pair continued to degrade substrate past the 120 minute observation time point (Figure 7-7; experimental data). With inclusion of cannibalism and cross-distraction interactions, the error of the superposition case, 4.278×10^{-1} was reduced to 3.116×10^{-1} (1.36-fold). This error was further improved by incorporating cathepsin L activation of procathepsin V reaction, as described by Equation 27.

$$proV + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} LproV \overset{k_{act,byL}}{\rightarrow} L + V \qquad (27)$$
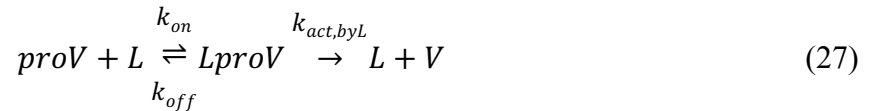
This additional model interaction reduced the error of the superposition case, 4.278×10^{-1}, to 2.820×10^{-1}, a 1.52-fold improvement. Procathepsin V activation was also included in the paired model with cathepsin K, since the model under-predicted total elastin degradation by 5.5%. However, this addition increased the overall error of the cannibalism and cross distraction interactions.

7.3.4 A priori *prediction of cathepsins K, L, S, and V substrate degradation*

Accounting for the cathepsin V individual kinetics and cathepsin-cathepsin interactions, the cathepsin proteolytic network model improves model performance for predicting the elastin and gelatin degradation of four cathepsins K, L, S, and V in combination (Figure 7-8). Compared to a model without any cathepsin-cathepsin

interactions (Fig. 3-7A, B; Model 1 prediction), the prediction accuracy in terms of degraded elastin was 30-fold (error of 7.191×10^{-3} compared to 2.131×10^{-1}) for the model including the cathepsin-cathepsin interactions (Figure 7-8A, B; Model 2 prediction) and for degraded gelatin was 16-fold better than the model with cathepsin-cathepsin interactions (error of 3.049×10^{-2} compared to 4.823×10^{-1}). For both elastin and gelatin, cathepsin S degraded the most substrate (Figure 7-8A, B; Model 2 prediction by CatS). Cathepsin K (Figure 7-8A, B; Model 2 prediction by CatK). When tracking the amount of active enzyme, there were substrate specific differences in the predicted amount of cathepsin still active at the end of the experiment (Figure 7-8C, D). Interestingly, there was on average 1.64-fold more total amount of active cathepsin present in the model including cathepsin-cathepsin interactions compared to the model relying only on the individual cathepsin kinetics to determine substrate degradation (Figure 7-8C, D). This was because through the addition of the cathepsin-cathepsin interactions, each cathepsin could bind to other cathepsins rather than substrate, effectively decreasing the amount of cathepsin catalytically active toward the substrate.
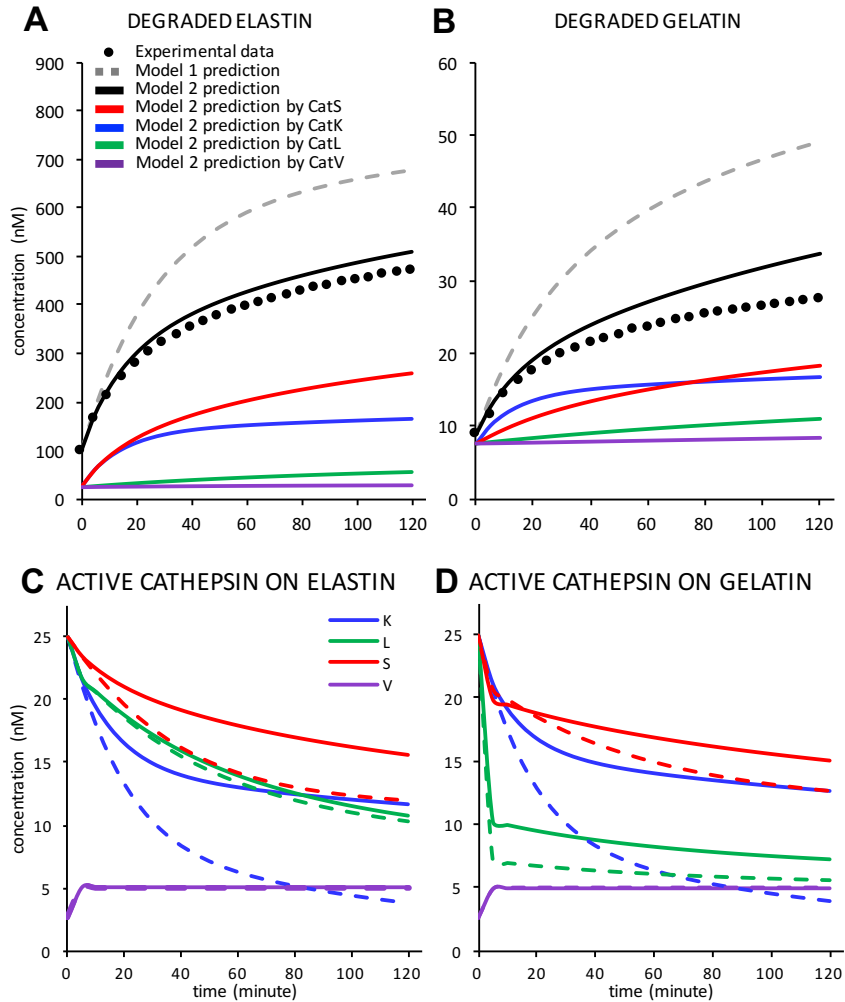
Figure 7-8. Cathepsin V cannibalism interactions *a priori* predicted substrate degradation of four cathepsins. *Inclusion of cathepsin-cannibalism terms in the model predicted substrate degradation* a priori *of cathepsins (cat) K, L, S, and V. Experimental data (black solid points) indicated degraded elastin (A) or degraded gelatin (B) over 120 minutes. Model 1 (gray dashed curve) predicted substrate degradation assuming superposition of catK, L, and S individual kinetics (with no cannibalism interactions) over-predicts total substrate degradation. Model 2 (black curve) indicated predicted substrate degradation when the model includes cathepsin cannibalism interactions between pairs of cathepsins. The contribution of catK (blue), catL (green), catS (red), and catV (purple) to the total substrate degradation predicted by Model 2 (black curve). (C, D) The total amount of active catK (blue), catL (green), catS (red), and catV (purple) can be compared between the model with cannibalism (solid line) and without cannibalism (dashed line) and how the concentration differs over time on elastin (C) and gelatin (D).*

147

7.4 Discussion

These studies demonstrate that cathepsin cannibalism occurs between cathepsins K, L, S, and V, which suppresses active enzyme in the system and protects against substrate degradation. Using a systematic approach, a cathepsin proteolytic network model of cathepsins K, L, S, and V degrading elastin and gelatin was constructed by first determining the individual substrate kinetics for each cathepsin on each substrate. In the initial layer of the model, the need for an inactivation rate was evident in addition to the traditional Michaelis-Menten kinetics to accurately capture the shape of the substrate degradation. The fits were further improved through the addition of the distraction hypothesis, where active cathepsins could reversibly bind to inactive cathepsins. Furthermore, since recombinant cathepsin V had to be activated from procathepsin V form to a catalytically active form, there was only about 10% of the expected active cathepsin available for hydrolysis at the start of the kinetic studies. Using the kinetic rates of individual and paired cathepsin interactions, the substrate of all four cathepsins could be predicted, which reduced the error 30-fold for degraded elastin and 16-fold for degraded gelatin compared to superposition of individual cathepsin kinetics.

The kinetic rates determine the directionality of cathepsin cannibalism interactions. For example, cathepsin K had the larger kinetic rate for catalysis of cathepsin L (Table 7-1, 500 s^{-1} vs. 336 s^{-1}), while the catalysis reaction between cathepsins S and L were equivalent (Table 7-1, both were 500 s^{-1}). Previously it was reported that cathepsin S hydrolyzes cathepsin K [7], however from the catalysis rates determined from these

experiments, cathepsin K had a larger rate for catalyzing cathepsin S than cathepsin S catalyzing cathepsin K (Table 7-1, 710 s^{-1} vs. 288 s^{-1}). When the Michaelis-Menten constant, K$_M$, was calculated for the two interactions, catalysis by cathepsin S had over a 2-fold smaller K_M (5.07×10^5 pM compared to 1.10×10^6 pM), indicating that cathepsin S had a higher affinity for cathepsin K. Between the pairs containing cathepsin V, cathepsin V had larger catalytic rates for cathepsins K (Table 7-3, 1.653 s^{-1}) and S (Table 7-3, 0.1448 s^{-1}) than vice versa.

Between cathepsin V pairs, the model predicted substrate degradation was improved by incorporating terms for procathepsin V activation by the other cathepsin in the pair. This improvement occurred for the cathepsins L and V pair, but not the cathepsins K and V pair. The improvement with paired cathepsin activation was hypothesized to happen with cathepsin L, and not cathepsin K, because cathepsins L and V share 80% sequence identity [20] compared to the 60% shared between cathepsins K and V . Furthermore, cathepsin V is known to activate itself, so since cathepsin L is structurally more similar to cathepsin V, it could explain why cathepsin L could also activate procathepsin V.

The cathepsin V activation experiments revealed that the duration of the procathepsin V activation protocol should be shortened, so that more than 10% of the enzyme will be catalytically active for the experiment. Furthermore, through the development of this model structure, it was clear that the initial amount of active enzyme at the start of the experiment was critical to determining reasonable rates for the cathepsin

cannibalism interactions. These experiments used 5 picomoles of recombinant enzyme in a total volume of 150 μL, which was a concentration of 33 pM. Using an initial concentration of 33 pM, the individual cathepsin kinetics were easily determined with small error compared to the experimental data, however, when the cathepsin cannibalism interaction rates were being fit, only forward binding rates that were three or more orders of magnitude above diffusion-limited binding would give a reasonable fit. Since these high forward binding rates seemed unlikely, being greater than the diffusion limit and strongest known biological interactions, reducing the initial concentration of active enzyme was the only way to achieve good fits (small error) in the paired interactions with less of these rates hit the parameter bounds. Furthermore, while this parameter set can accurately describe this combination of experiments, it was not robust enough to describe the complexity of individual cathepsin interactions. When increasing amounts of cathepsin S were co-incubated with either DQ-elastin or DQ-gelatin, the addition of more of the same cathepsin did not produce much more substrate degradation as shown by the experimental data in Figure 7-9. Specifically Figure 7-9A (discrete points) showed total degraded elastin within 5% of each other at 120 minutes when co-incubated with 1X through 4X amounts of recombinant cathepsin S. Predictions for substrate degradation with increasing amounts of cathepsin S, using the parameterized cathepsin proteolytic network model described here, still over-predicted the experimentally seen substrate degradation (Figure 7-9, prediction curves compared to discrete experimental data points). Further experiments varying the individual cathepsin concentrations should be performed, so the individual kinetics can be

150

fit simultaneously to multiple datasets and ensure robust parameters that can predict these dynamics.
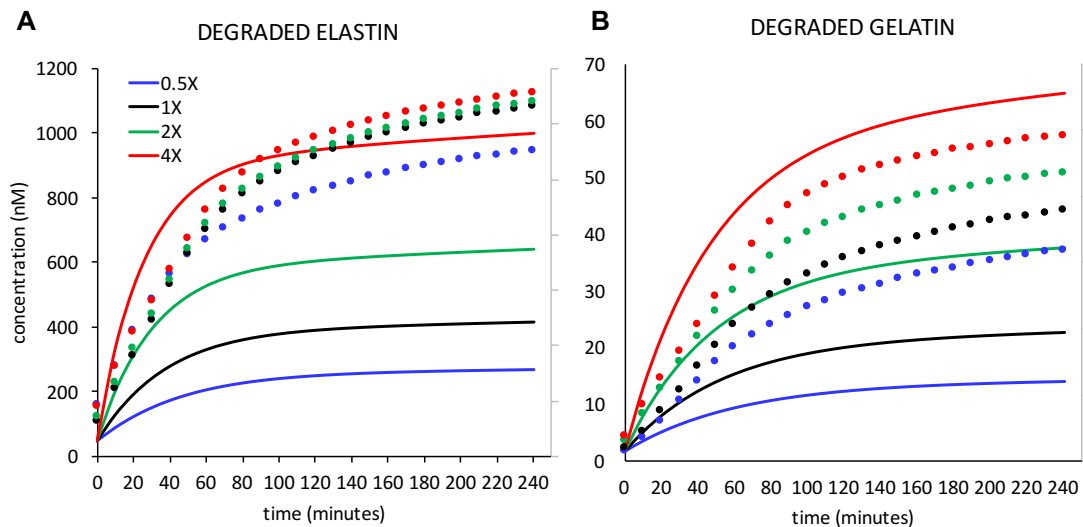
Figure 7-9. Increasing amounts of cathepsin S does not result in comparable increase in substrate degradation. *The discrete points were experimental data for substrate degradation due to increasing amounts of cathepsin S from a 0.5X concentration (2.5 pmoles of active enzyme) to a 4X concentration (20 pmoles of active enzyme) and the subsequence amount of degraded (A) elastin and (b) gelatin. Using the cathepsin S individual kinetics, model simulations with corresponding initial conditions showed a wider distribution in total substrate degradation compared to the experimental data.*

This mechanistic model of the cathepsin proteolytic activity (with interactions enumerated in Figure 7-10) brings scientists closer to successfully inhibiting cathepsins as treatment for tissue destructive diseases or controlling proteolysis in engineered living systems. This work demonstrates that complex cathepsin-cathepsin interactions are modulating the concentrations of cathepsins catalytically active towards substrates in a system. This modulation is critical for accurately targeting cathepsins with well-designed inhibitors and incorrect dosing explains why the cathepsin inhibitors that have made it to

phase II and III clinical trials have been effective at reducing target substrate degradation, but were ultimately discontinued due to off-target side effects. If the amount of cathepsins catalytically active to substrates is actually lower than anticipated, then "excess" inhibitor could precipitate these side effects.
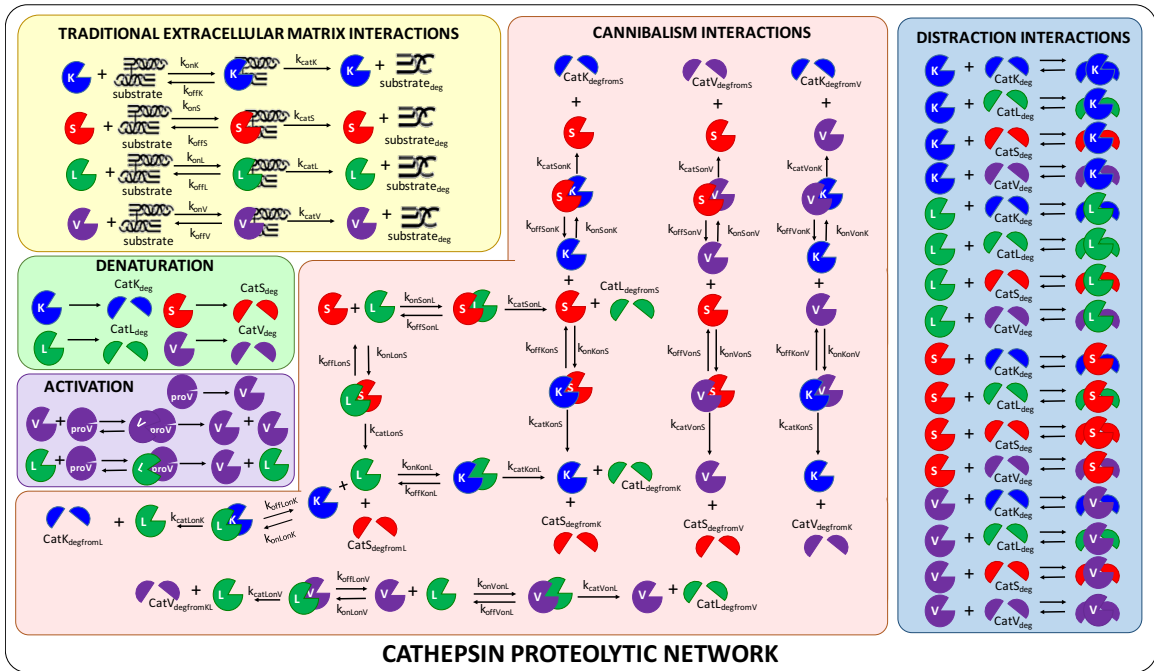
Figure 7-10. Schematic of cathepsin K, L, S, and V proteolytic network interactions.

CHAPTER 8. GENERATING CANNIBALISM-RESISTANT AND ACTIVE SITE DEAD CATHEPSIN MUTANTS

8.1 Introduction

Chapter 7 describes computational modeling to parse the hypothesized cathepsins-cathepsin interactions between cathepsins K, L, S, and V. For pairs of cathepsins, like cathepsin K and S, there are two possible cathepsin cannibalism interactions that may be happening: (1) cathepsin S bind to cathepsin K and the complex is catalyzed to free cathepsin S and degraded cathepsin K or (2) cathepsin K can bind to cathepsins S and the complex is catalyzed to free cathepsin K and degraded cathepsin S (Figure 8-1, Reactions 1 and 2). However, to experimentally validate the cannibalism directionality revealed by the cathepsin proteolytic network model, cathepsin mutants were created using site-directed mutagenesis designed to interrupt these interactions. Two types of cathepsin mutants were used to test computational modeling informed hypotheses: (1) active site dead mutants, which were not catalytically active toward substrate, but could be degraded by other cathepsins (Figure 8-1, Reaction 3) and (2) cleavage-resistant mutants, which were immune to cathepsin cannibalism interactions without disrupting proteolytic activity (Figure 8-1, Reaction 4). Cleavage-resistant mutants can degrade substrate and other cathepsins, but are not hydrolyzed by other cathepsins. Here, we created active site dead and cannibalism-resistant mutants of cathepsins K, L, S, and V to validate the cathepsin cannibalism interactions hypothesized from the cathepsin proteolytic network model detailed in Chapter 7.

153
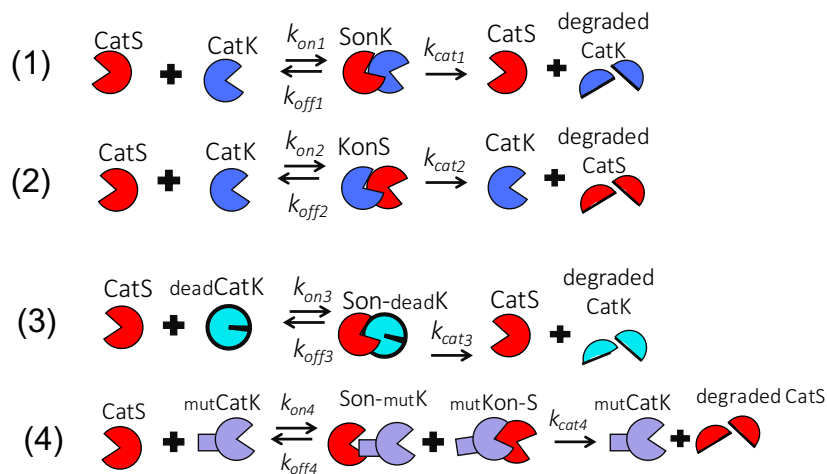
Figure 8-1. Cathepsin cannibalism interactions and mutants to disrupt cannibalism interactions. *Using cathepsins K and S as the model system, reactions (1) and (2) describe the possible cannibalism interactions between the two cathepsins. (3) Active site dead mutants are not catalytically active and can act as a substrate to a cannibal cathepsin. (4) Cannibalism-resistant mutants (or cleavage resistant mutants) can still degrade substrate but cannot be hydrolyzed by a cannibal cathepsin.*

8.2   Materials and Methods

8.2.1   Design of cannibalism-resistant mutants

Putative cleavage site locations were determined by comparing each of cathepsin K, L, S, and V's amino acid sequences with active site specificity matrix (available from the MEROPS online peptidase database) of each enzyme's active site to score the likelihood that a given amino acid in position P4 through P4' will be cleaved by the cathepsin's active site [131] using the PACMANS program [129] described in Chapter 3. Locations in the propeptide of each enzyme were excluded, since this region was cleaved during activation to the mature enzyme and would not be part of the molecule in catalytically active cathepsins. From the top scoring sequences in the PACMANS analysis,

the sequences locations were reviewed for accessibility by an external protease (as described in Figure 3-2). Once the accessible sequences were identified, each amino acid score for the 8 amino acid sequence was reviewed and the greatest scoring residue in the sequence was mutated to reduce the overall sequence score combined with the known "safe" residue substitutions [137] as shown in the example in Figure 8-2.

The active site dead mutants were designed by mutating the three active site residues in the catalytic triad that are conserved in all of the cysteine cathepsins: cysteine, histidine, and asparagine [3, 21]. The active site cysteine (C) was mutated to alanine (A) or serine (S), the active site histidine (H) was mutated to an alanine (A), asparagine (N), or glutamine (Q), and the active site asparagine (N) was mutated to alanine (A) or aspartic acid (D).

Cumulative Score for (251) ASLTSFQF (258) Sequence: 633
Individual Residue Scores:

	P4'	**P3'**	**P2'**	**P1'/**	**P1**	**P2**	**P3**	**P4**	
(251)	A	S	L	T /	S	F	Q	F	(258)
	63	57	244	76 /	128	16	37	12	

↓ *Leucine in P2' position is the highest scoring residue*

Possible mutants:

L → A: (A in P2' score = 22) Cumulative Score goes from 633 to 411

L → V: (V in P2' score = 172) Cumulative Score goes from 633 to 561

Figure 8-2. Example of how point mutations to interrupt cathepsin cannibalism interactions were determined from top scoring PACMAN sequences.

8.2.2   Site-directed mutagenesis using overlap extension PCR

The mutant cathepsins were made using site-directed mutagenesis using overlap extension polymerase chain reaction (PCR) [133]. Figure 8-3 shows a schematic of the method; first forward and reverse mutagenic primers were designed to introduce a mutation into the cathepsin gene. The primers were designed such that:

- Primers had at least 15 bases of conserved sequence (exact match) and at least 15 nucleotides overlap on their 5' ends

- Primers were between 25 and 30 nucleotides long with a high guanine/cytosine content (at least 60%) to ensure a high melting temperature (want at least 60°C)

- Primers had 3 or 4 guanine/cytosine nucleotides in the last six bases as a clamp to enhance annealing to the template
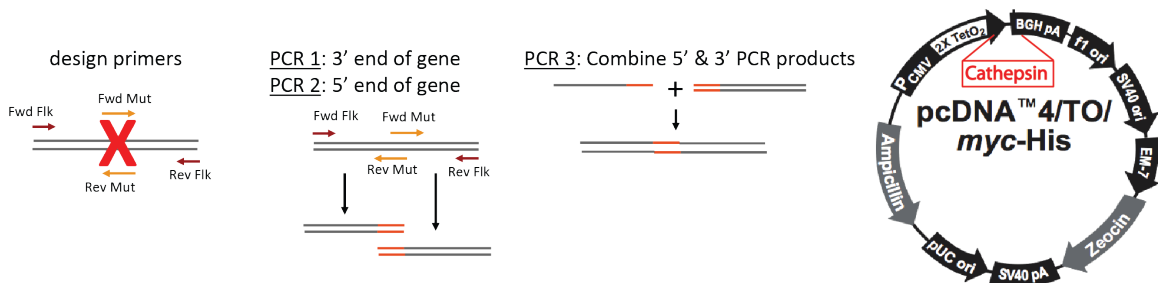
Figure 8-3. Schematic of approach for making cathepsin mutants by site-directed mutagenesis using overlap extension polymerase chain reaction. *Cathepsins mutants were made to the wildtype cathepsin gene and then inserted into the pCDNA 4/TO/myc-His vector.*

Then two rounds of PCR were carried out with the forward and reverse flanking primers to introduce the point-mutation into gene. A third round of PCR used the products from the

156

first two reactions to create the full-length cathepsin gene with the point-mutation. The gene was then ligated (T4 DNA Ligase, New England Biolabs) into the pcDNA 4/TO/A plasmid (Invitrogen), which had zeocin resistance and a C-terminal polyhistidine tag for nickel-column purification. The plasmid was then amplified in bacteria cells and sequenced (Genewiz) to verify correct point mutations were introduced.

8.2.3   *Stable expression and purification of mutant cathepsins*

Once the mutant plasmid sequences were sequenced to confirmed the designed mutation was introduced, stable cell lines to produce the mutant cathepsins were created to purify the mutant enzymes. First the plasmid was linearized with ScaI (New England Biolabs) to stably transfect HEK293T cells using Lipofectamine 2000 (Thermo Fisher Scientific). After two days, the cells were selected with zeocin (InvivoGen). After the cells were selected for four weeks for the plasmid with zeocin resistance and plasmid has been incorporated into the cell's genome through homologous recombination, the cell lysate and media were collected for purification by the polyhistidine tag with nickel columns in the ProBond Purification System (Invitrogen).

To induce cathepsin expression, cathepsin mutant cell lines were created in a T-REx inducible system (ThermoFisher Scientific) [230], where naïve HEK293T cells were transfected with 10μg of the Tet repressor plasmid (TetR) and 1μg of mutant cathepsin plasmid. Naïve HEK293T cells were simultaneously transfected with linearized TetR and mutant cathepsin plasmids and after two days, the cells were selected with both zeocin (230 μg per mL of media for the cathepsin mutant plasmid) and blasticidin (100 μg per mL of

media for the TetR plasmid). The TetR repressed the expression of target cathepsin sequence transfected and when tetracycline was added to the stable cell lines, the mutant cathepsin of interested was then transcribed and translated into protein.

8.2.4   Western blot and multiplex cathepsin zymography

Using the stable cell lines or transiently transfected cells, Western blots and multiplex cathepsin zymography were performed on cell lysate to first determine if there was mature or procathepsin detectable by Western blot and if there was gelatinase activity detectable by cathepsin zymography. Western blot samples were in a reducing loading dye with β–mercaptoethanol to break all disulfide bonds. Samples were resolved on a 12.5% SDS-PAGE gel and then transferred to a nitrocellulose membrane (BioRad), blocked for 1 hour (LI-COR Bioscience), then probed with a primary antibody overnight.   Primary antibody solutions for monoclonal mouse anti-human cathepsin K (1:500, Millipore), polyclonal goat anti-human cathepsin L (1:1000, R&D Systems), polyclonal goat anti-human cathepsin S (1:1000, R&D systems) and monoclonal mouse anti-human cathepsin V (1:1000, R&D Systems) were made in Odyssey Blocking Buffer (LI-COR Bioscience) diluted 1:2 with PBS. The following day, primary antibody solutions were saved and membranes were probed with anti-mouse or anti-goat secondary antibodies (1:5000 with 0.1% Tween, Invitrogen) tagged with an infrared fluorophore and imaged on with a LI-COR Odyssey scanner.

The multiplex cathepsin zymography assay was used to quantify cathepsin gelatinase activity [134, 135]. It is a modified SDS-PAGE method where a polyacrylamide gel was

polymerized with gelatin and cell lysate samples were loaded with a non-reducing loading dye. Then electrophoresis separated the sample proteins by molecularly weight. The gel was then put in a renaturing buffer (65 nM Tris buffer at pH 7.4 with 20% glycerol) to allow the cathepsins to re-fold into their native confirmation. Gels were then incubated in assay buffer, either 0.1 sodium acetate buffer at pH 4.0 or 0.1M sodium phosphate buffer at pH 6.0 both with 1mM EDTA and freshly added 2mM dithiothreitol (DTT), overnight at 37°C. The gel was stained the following morning with Coomassie blue (4.5% Coomassie blue, 10% acetic acid, 25% isopropanol) and destained (10% acetic acid, 10% isopropanol). On zymograms, areas with white bands indicate cathepsin activity. Zymograms were imaged with ImageQuant LAS 4000 (GE Healthcare). Densitometry quantification using ImageJ (NIH) was used to quantify white zymogram intensity bands and Western blot bands. These studies gave information about the structural stability of the mutant cathepsins.

8.3   Results

8.3.1   Verification and validation of cathepsin mutants

To verify the direction of cannibalism experimentally, cathepsin mutants were designed to test both directions:

(1)   Active-site mutant (Figure 8-1, Reaction 3): a cathepsin was mutated to be active-site-dead, meaning it does not have catalytic activity, but can still be cleaved by another active-site live cathepsins. Mutations were made in the three residues, cysteine,

histidine, and asparagine, which make up the conserved active site cleft of cysteine cathepsins, and

(2)   Cannibalism-resistant (cleavage-site) mutant (Figure 8-1, Reaction 4): mutated cathepsin is not cleavable by other cathepsins, but the mutated cathepsin can still cannibalize cathepsins.

Both active-site and cleavage-site mutations could change the folding of cathepsins. Some could be produced only the inactive procathepsin form, others may be activated to the mature form, and some of the mutations may not produce protein detectable by Western blot. The summary of cathepsin mutants generated and whether they produced positive Western blot and/or zymography signal was described in Table 8-1. Cathepsin K mutants were generated and tested first to inform other cathepsin mutant design. Seven active-site cathepsin K mutants were created and only three of those mutations produced protein detectable by Western blot and none of the mutant proteins had detectable gelatin zymography signal (Figure 8-4A). All cathepsin K cannibalism-resistant mutants were detectable by Western blot and cannibalism-resistant mutants have catalytic activity detectable by gelatin zymography (Figure 8-4A). Cathepsin L cannibalism-resistant mutants produced procathepsin L and both isoforms of mature cathepsin L were detectable by Western blot and most retained their gelatinase activity as detected by gelatin zymography (Figure 8-4B). The procathepsin form of cathepsins S and V were detected by Western blot, suggesting these mutations fold into a confirmation that could be activated and possess catalytic activity (Figure 8-4C, D). Cathepsins S and V cannibalism-resistant

160

mutants had detectable catalytic activity by gelatin zymography, while active site dead

mutants had reduced catalytic activity (Figure 8-4C, D).

Table 8-1. Summary of protease present of mutant cathepsin protein resulted in positive Western blot or gelatin zymography signal. *Active-site mutations were made of the active site cysteine (C), histidine (H), and asparagine (N). All other mutations made were cannibalism-resistant mutations. Not all mutations of the same residue yielded detectable protein. For Western blots, positive signal was distinguished between if only the procathepsin was detectable or both procathepsin and mature cathepsin detected.*

CATHEPSIN K	Western Blot	Zymography	CATHEPSIN L	Western Blot	Zymography
C139A	Pro & mature	No	L158V	Pro & mature	Yes
C139S	No	No	E176N	Pro & mature	Yes
V171A	Pro-form	Yes	V239G	Pro & mature	No
L253A	Pro-form	Yes	E289N	Pro-form	No
L253V	Pro-form	Yes	S326A	Pro & mature	Yes
H276A	No	No			
H276N	No	No	CATHEPSIN V	Western Blot	Zymography
H276Q	Pro-form	No	C138A	Pro & mature	No
N296A	No	No	H277Q	Pro & mature	No
N296D	Pro-form	No	N301D	No	No
CATHEPSIN S	Western Blot	Zymography	L158V	No	No
C139A	No	No	V219G	Pro & mature	Yes
H278Q	Pro & mature	No			
N298D	Pro & mature	No			
L159V	Pro & mature	Yes			
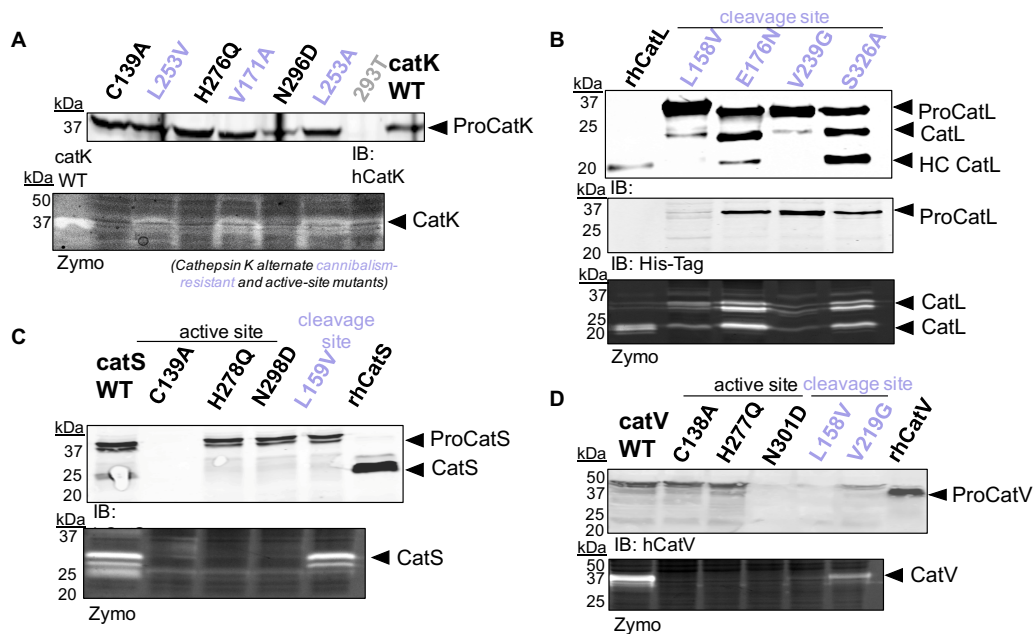
Figure 8-4. Designed cathepsin cannibalism-resistant and active site dead mutants for cathepsins K, L, S, and V. *Cathepsin (cat) mutants were verified for folding and catalytic activity with Western blots and gelatin zymography for catK (A), catL (B), catS (C), and catV (D).*

8.3.2 Cathepsin K mutants retain elastinolytic activity

Cathepsin K mutants retained elastinolytic activity (Figure 8-5). To account for the elastinolytic activity of other proteases secreted by the cell, each mutant was also inhibited with the small molecule inhibitor E-64. Naïve HEK293T cells have minimal cysteine cathepsin activity, so we assumed the elastinolytic activity inhibitable by E-64 was due to the overexpression of the mutant cathepsins. In the case of wildtype cathepsin K, 45% of the total elastin cleavage was attributed to cathepsin K. Furthermore, from the cathepsin K mutants, the cleavage-site mutants retained some of their elastinolytic activity with at least 12-28% of elastinolytic activity in these cell lines compared to 45% of elastinolytic activity

attributed to wildtype cathepsin K. Active-site mutants lost much of their elastinolytic activity, however the was site-dependent. Mutations of the active site histidine (H) resulted in the greatest loss of elastinolytic activity compared to the mutations of the cysteine (C) and asparagine (N).

elastin degradation inhibited by the E64 small molecule inhibitor
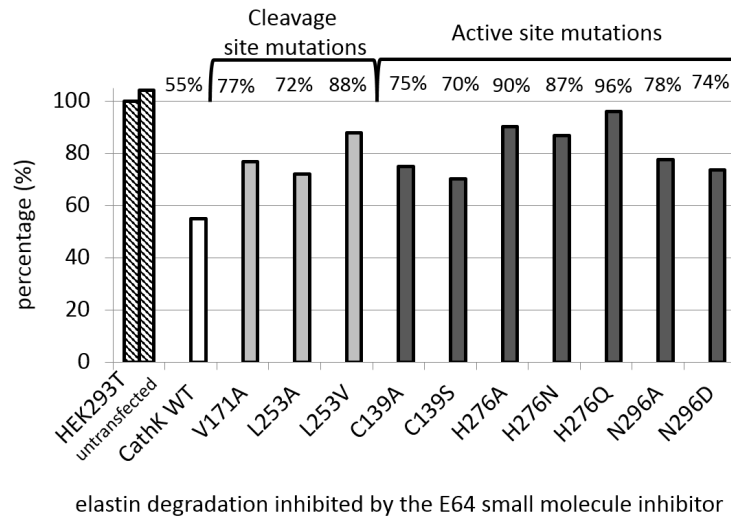
Figure 8-5. Elastinolytic activity of cathepsin K mutant stable cell lines. *Cathepsin K mutants were overexpressed in HEK293T cells and incubated with DQ-elastin in the presence and absence of the small molecule inhibitor E-64 to account for elastinolytic degradation by other proteases secreted by the cells. The y-axis shows the percentage of elastin degradation attributed to the expression of the cathepsin K mutants.*

8.3.3   Cathepsin K L253A cannibalism-resistant mutant resists hydrolysis by cathepsin S

and autodigestion by cathepsin K

Cathepsin K mutants were tested for their cleavage resistance through co-incubation studies with cannibal cathepsin (recombinant cathepsin S) as well as with autodigestion by recombinant cathepsin K. The L253 mutation protected against cannibalism hydrolysis by cathepsin S as was described in Chapter 3 and Figure 3-3, with

the L253A mutation providing the most protection from degradation by cathepsin S. The V171A mutation also provided some protection from hydrolysis by cathepsin S, when compared to wildtype cathepsin K co-incubated with cathepsin S, however there was still a significant decrease compared to V171A without cathepsin S present.
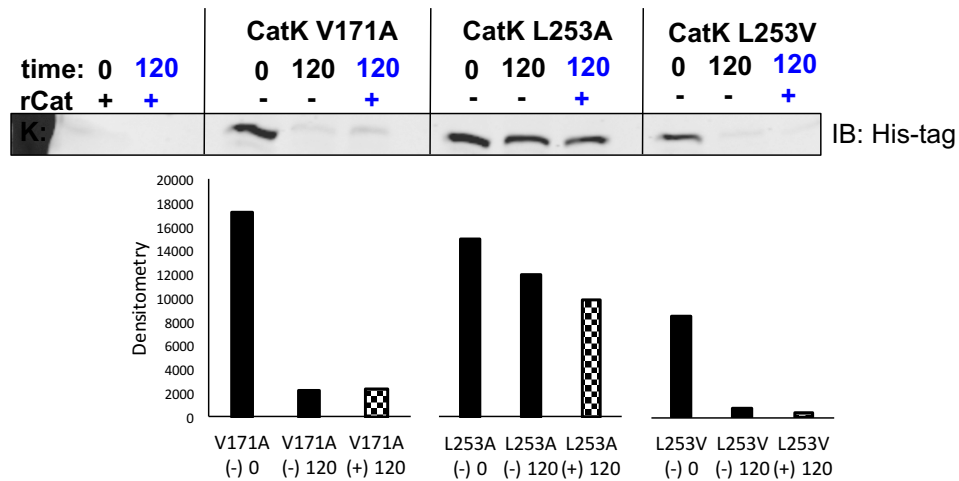
Figure 8-6. Cathepsin K L253A mutant was resistant to autodigestion. *Cathepsin K mutants were co-incubated with and without recombinant cathepsin K for 120 minutes and samples were run on a Western blot to detect amount of cathepsin remaining. Mutant cathepsins were detected with a His-tag primary antibody to distinguish from the addition of recombinant cathepsin K to the system.*

Cathepsin K is known to autodigest, or hydrolyze itself, when incubated without substrate [7]. The autodigestion potential of these cannibalism-resistant cathepsin K mutants were tested by co-incubated cathepsin K mutants with wildtype recombinant cathepsin K (Figure 8-6). In addition to resisting being cannibalized by cathepsin S, the L253A cathepsin K mutant was also resistant to autodigestion by wildtype cathepsin K over 120 minutes. Cathepsin K V171A and L253V mutants both were autodigested by their respective mutants, as shown by their Western blot signal reduction of 87.4% and 90.6%,

respectively, by 120 minutes; compared to the L253A mutant, where there was only a reduction of 19.9% by 120 minutes. At 120 minutes, recombinant wildtype cathepsin K autodigested 18.5%, leaving 65.6% of the cathepsin K L253A mutant remaining after co-incubation with recombinant wildtype cathepsin K. The cathepsin K L253A mutant was the most robust cleavage site resistant mutant by being resistant to cleavage by cathepsin S and cleavage by wildtype cathepsin K.

8.4   Discussion

This work shows that multiple cathepsin species in a single system can, in fact, alter the expected amounts of cathepsins present. Furthermore, cleavage site mutants provide experimental proof that these cathepsins are binding and degrading each other at these specific locations, which can be applied as "decoy" cathepsins to protect matrix substrates from degradation to prolong the functional lifespan of biological machines. The computational model of the cathepsin proteolytic network developed in Chapter 7 can be refined by confirming the directionality of cannibalism terms and reducing any unnecessary parameters and reactions in the network model. Kinetic studies varying the amounts and species of cathepsins present can be performed and used to determine the impact on cathepsin cannibalism binding and catalytic rates kinetic rates for computational model refinement.

165
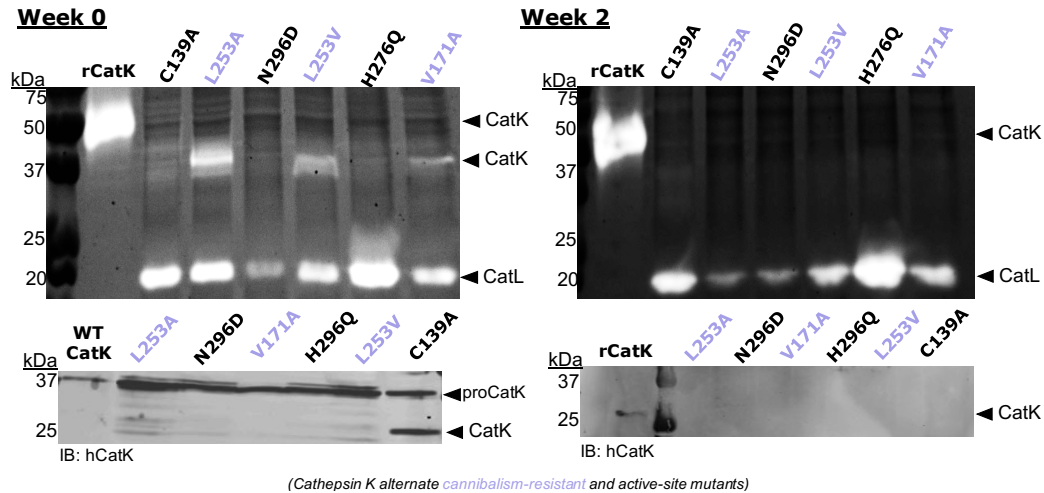
Figure 8-7. Cathepsin K mutant cell lines stop producing cathepsin K over time. *Comparing freshly thawed cells and cells in culture for two weeks, the stable cell lines stopped producing the cathepsin K mutants they were transfected to overexpress. This was detected both by gelatin zymography as well as by Western blot. Furthermore, overexpression of cathepsin K appears to induce expression of cathepsin L (20kDa) and this zymography signal also changes over time as the cathepsin K signal (37kdDa) was reduced.*

Purification of these mutant cathepsins was a significant effort in biochemistry, overexpressing cathepsin mutants was detrimental to cell health and the cell lines stopped overexpressing the cathepsins over time (Figure 8-7). However, to combat issue, we created and used tetracycline inducible (Tet-inducible) stable cell lines for each of the mutants. Using the Tet-inducible systems, we controlled when cathepsins were being produced, so that we can purify directly from the cathepsins secreted into the media for a finite amount of time, without having the continual overexpression of cathepsins becoming detrimental to cell health and minimizing the changes in cellular regulating. Based on observations of the cathepsin cell lines, we hypothesized that cathepsins can cleave cell adhesion molecules and the overexpression of cathepsins negatively impacts the HEK293T

cells ability to adhere to the cell plates. We have observed that in the overexpression cell lines, the cells visually do not have the same cobblestone morphology as the naïve HEK293T cells and the cells easily lift of the plate, suggesting they were not adhering as well as HEK293T cells without overexpression of cathepsins. This observation was also supported by cathepsins purification with the Tet-inducible cell lines, which have not yet successfully been able to be revived after collecting secreted cathepsins from these cells.

During this mutant design and protein purification process, it has become clear that much of the enzymes produced from these systems are in the procathepsin form that needs to be activated, meaning the propeptide obscuring the active site of the cathepsin must be cleaved to detect catalytic activity. We successfully activated our purified samples by incubating them in a pH 4.0 sodium acetate buffer as well as through pepsin activation. As we move forward with utilizing these cathepsin mutants to test experimental hypothesis and connect back to the cathepsin proteolytic network model, it will be important to make sure we have an accurate characterization of the amount of cathepsin used in an experiment. As demonstrated with parameter estimation and model structure of the cathepsin proteolytic network model described in Chapter 7, even the presence of inactive cathepsin can alter the expected substrate degradation. Since most readily purified cathepsins are in the procathepsin form, we can characterize the activation profile, like we did with cathepsin V (Figure 7-5A) and use that information to calculate the actual amount of active cathepsin at the start of any experiment.

These cathepsin mutants provide an exciting opportunity to tease apart the complex cathepsin-cathepsin interactions within the cathepsin proteolytic network. Just from generating these cathepsin mutants, we found that not all mutations can be introduced into the cathepsins structures, and those mutations that can successfully be introduced have varying effects on the catalytic activity of these cathepsins. Furthermore, overexpression of cathepsins in cellular systems alter the proteolytic network dynamics, as evident by the upregulation of cathepsin L in the cathepsin K mutant overexpression (Figure 8-7). This further highlights the importance of a comprehensive cathepsin proteolytic network model that reliably predicts the cathepsins' responses to perturbations to effectively target cysteine cathepsins in tissue destructive diseases and control biological machine matrix proteolysis. Altering one node in this network results in non-linear and unintuitive changes that need to be quantified to effectively inhibit cathepsins without unforeseen consequences and side effects through the delicate proteolytic environment.

CHAPTER 9.   CONCLUSIONS AND FUTURE DIRECTIONS

9.1   Major Findings

Throughout this dissertation, we have improved the current understanding of protease-protease interactions and in attempts to exploit these interactions in therapies for tissue destructive diseases and development of long-lived, stable, locomoting biological machines. This work developed a mechanistic model to quantify the interactions between cathepsins K, L, S, and V and extracellular matrix substrates collagen and elastin. Extending the previous discovery of cathepsin cannibalism between cathepsins K and S, we verified and modeled the cannibalistic interactions between cathepsins K, L, S, and V. To confirm the directionality of cathepsin cannibalism interactions, we developed a bioinformatics analysis tool, PACMANS, to compare the active site specificity matrix of a cannibal cathepsin to the sequences of a substrate cathepsin to validate the cathepsin proteolytic network model's predicted interactions. This also allowed us to determine the specific locations of where one cathepsin species binds to another, through development of 23 different cannibalism-resistant and active-site-dead mutants to study these mutant's interactions with wildtype cathepsins. Furthermore, we have shown that by introducing point-mutations we can interrupt cathepsin cannibalism interactions from occurring.

This work explored the role of proteolytic activity in destabilizing locomoting biological machines. First, we determined the cathepsin and matrix metalloproteinase (MMP) activity of locomoting biological machines over the first 12 days of a bio-bot's

functional lifespan. We reported that these engineered living systems produce an increasing amount of active cathepsin L over this time frame, independent of electrical stimulation. MMP-2 and MMP-9 were also expressed by these biological machines, but in contrast to cathepsin L, MMP-9 was reduced by day 12 and MMP-2 activity was constant during this timeframe. When these bio-bots were culture with ACA, a fibrinolysis inhibitor known to stop plasmin degradation of fibrin, although this serine protease inhibitor was not expected to affect cathepsin L, the ACA cultured bio-bots showed a reduction in active cathepsin, however the cathepsin L was still increasing over time. This temporal increase in cathepsin L proteolysis was independent of bio-bot design parameters including cell seeding density, hydrogel stiffness, and optogenetic stimulation. Furthermore, when we cultured these bio-bots with E-64, a broad spectrum cathepsin small molecule inhibitor, we found that we could reduce the amount of cathepsin bound to muscle strip matrix protein, without affecting cell viability and passive tension in the muscle strips. In fact, the E-64 cultured bio-bots had a prolonged the average life expectancy to at least 3.5-fold greater than 1X ACA cultured muscle strip. To better understand the life expectancy of these bio-bots, we analyzed different survival functions and discovered that a Weibull distribution could accurately capture variability and life expectancy of the bio-bots cultured in different amounts of ACA. Moreover, we uncovered a possible relationship that between the Weibull parameters and the ACA concentration that can be further explored to help predict the life expectancy of these biological machines.

Finally, we explored the possibility of fibrin(ogen) as a novel substrate for cysteine cathepsins. Using independent computational modeling methods, molecular docking and active site specificity analysis, we identified locations where cathepsins K, L, and S could bind to the three chains of fibrinogen. We also showed experimentally that these potent proteases can degrade fibrinogen and although these cathepsins have redundant substrate preferences, they all degraded the three chains of fibrinogen different. This novel mechanism of fibrinogenolysis has a variety of niches, including in the blood of individuals with sickle cell anemia, who have increased cathepsin activity and have been reported to have an increased concentration of circulating fibrinopeptide A. This mechanism also plays a role in biological machine and other engineering living system construct that uses fibrin networks. Many cell types produce cathepsins and in these synthetic environments where fibrinogen is combined with thrombin, other matrix proteins, and cells, cathepsins may play a larger role in that network formation than previously understood.

The cumulative findings of this doctoral thesis provide a novel understanding of the cathepsin proteolytic network and exploiting these interactions can be used in therapies for tissue destructive disease and for more control of engineered livings systems. However, there remains a large body of work yet to be completed with regards to layering on cellular complexities to this cathepsin proteolytic network. The work of this dissertation focused primarily on recombinant enzyme work, but studying these systems in cellular systems where these cathepsins can be separated by different compartments, they are exposed to a wide variety of substrate, other proteases, as well as endogenous inhibitors, and cellular

regulation can counter-act perturbations to the cathepsin proteolytic network is the next major step in controlling this proteolytic activity.

9.2 Future Research Directions

9.2.1 Cathepsin Proteolytic Network Dynamics

Cysteine cathepsin are upregulated in a variety of tissue destructive disease and currently researchers have had trouble specifically targeting cathepsin activity [38]. From the clinical aspect, all the well-designed cathepsin inhibitors have been effective at stopping disease progression, but could not advance to later-stage clinical trials because of off-target side-effects [47, 110]. The major findings of this dissertation highlight the importance of a complete picture of the proteolytic network to effectively target cathepsins. A better picture of the cathepsin proteolytic network offers a way to re-examine those pharmaceutical inhibitors to adjust the dosing to find a safe dose for patients that still achieve the desired therapeutic effect. Furthermore, with the insights gained in this doctoral thesis, researchers can develop more-effective strategies at targeting the cathepsin activity with potential combination therapies to treat tissue destructive illnesses.

9.2.1.1 Cellular effects on the cathepsin proteolytic network

In this work, we present evidence that cathepsins K, L, S, and V exhibit complex cathepsin cannibalism interactions, regulating the amount of enzyme available to degrade substrates. We have also demonstrated that even overexpressing a single cathepsin causes a shift in the cell's cathepsin proteolytic network, where the cathepsin of interested is

overexpressed and a compensatory regulator cathepsin is upregulated by the cells for a period and eventual the cathepsin of interest is no longer detected, at which point the regulatory cathepsin returns to its normal levels. This complicated cellular feedback observed within HEK293T cells, a cell line considered easy to transfect, so this underscores that complex cathepsin proteolytic dynamics that likely vary between cell types.

Previous research from our lab and from others has shown that inhibiting cathepsins activity in cellular systems does not result in a dose-dependent decrease in cathepsin activity [71, 124]. Once it is understood how a given cell type will compensatory for varying the concentration of one species of cathepsin over another, layering on cathepsin inhibitors may be easier to predict the cellular responses and that information can be used to for dosing and determining if inhibitor combination treatments are necessary to specifically control the proteolytic activity in a pathological state.

9.2.1.2   Impact of inter-familial proteases interactions on cathepsin proteolytic dynamics

Cells do not live in a cysteine cathepsin bubble, and secrete other proteases to perform other homeostatic functions. Matrix metalloproteinases (MMPs) are a class of protease with similar substrates to cathepsins and are upregulated in tissue destructive disease. We have also characterized their proteolytic activity in biological machines. We [129] as well as others [216] have started to explore the inter-familial protease interactions between cathepsins and MMPs. Future work in this area should explore these inter-familial interactions and layer on to the cathepsin proteolytic network model developed in Chapter 7. Although cathepsins and MMPs are catalytically active at different pHs (cathepsins

prefer acidic environments, while MMPs prefer neutral environments), these two classes of enzymes serve similar function and are likely involved in regulating each other. In the biological machines cultured with E-64 (discussed in Chapter 6), we observed the inhibition of cathepsin L also resulted in an increase in MMP-2 activity. As cellular regulation is incorporated in cathepsin proteolytic network model, it is important to consider what role other proteases may play a part in this expansion.

9.2.2   Biological Machine Development

In order to develop multicellular functional units with the ability to sense, process signals, and produce force and motion in response to external stimuli, the interactions between cells in the biological machine must first be understood. Specifically, how the cells regulate their extracellular matrix through proteolysis and characterizing the proteolytic network of these cells is crucial for designing biological machines that can be used for prolonged functional times. Moreover, as different cell types are incorporated into these engineered living systems, the heterotypic contact can change the protease signature in non-intuitive ways, which is an important element to consider in biological machine design and development.

9.2.2.1   Expanding life expectancy model

Developing a mechanistic model of biological machine life expectancy is important for designing interventions to control the life expectancy of future biological machines. The two-compartment model described at the end of Chapter 6 is a good starting point, but

would need to be expanded to improve its predictive power. First, the protease compartment assumes a linear increase in active cathepsin L far beyond the 12 day time frame we have characterized. To expand this model, it would be important to characterize the protease activity of these "older" bio-bots to make sure that compartment is not just linearly increasing indefinitely, especially since there is some evidence in other mammalian muscle development that cathepsin activity is turned down after a certain time point [222]. Likewise, it would be good to confirm that other proteases are not turned up at later time points to a shift in the proteolytic network if cathepsin L is turned down. A more well-defined protease compartment could allow design of different interventions to control the life expectancy.

Addition of mechanical information could also be beneficial to the predictive power of the life expectancy model. Currently this information is excluded, but especially if these bio-bots are living for months, what is the effect of the continual passive tension on the mechanical integrity of the bio-bot muscle strip? There could be a certain "age" at which extracellular matrix stiffening can prolong the functional lifespan, which would need to be further investigated. Stress information overlaid on this graph model could also help predict if muscle strips break in certain locations that would need to be appropriate reinforced for a designed life expectancy. No pattern has been seen in the simulations analyzed of the current model, edges are just cleaved until a certain proportion of all the edges are gone, but in dynamic system that is constantly held in tension, certain cleaved edges may have much more of an impact on the day the muscle strip fails. Furthermore, quantification of

histological stained Masson's Trichrome samples could be used to help information the thresholds values used.

9.2.2.2 Inducible control of cysteine cathepsin expression in bio-bots

Work discussed in this dissertation has underscored the importance and hydrolytic potential of cysteine cathepsins on fibrin(ogen). Biological machines express increasing amounts of cathepsin L over time and cathepsin L can degrade the major matrix components of the muscle strips. Through the work discussed in Chapter 8, we created tetracycline inducible wildtype and mutant cathepsins and the lab has tetracycline inducible cystatin, which is an endogenous inhibitor of cathepsin activity. Future generations of biological machines can be genetically engineered to controllably express "decoy" cathepsins or cystatins to control the proteolytic activity of these biological machines over time. For example, if a future, self-sustaining biological machine muscle strip expressed inducible cystatin C, the bio-bot can be cultured in (or pre-treated with time-release with) tetracycline to have the muscle strip express cystatin C while it performed its designed function, then when the tetracycline ran out, the cathepsin activity of the biological machine could increase and degrade the matrix. This same example would also apply to a muscle strip expressing inducible active site dead cathepsins that can act as a decoy substrate. The scenario to choose would be dictated by that would be the downstream effect of adding in an inhibitor to the system over another cathepsin (interactions that could be teased apart as described in 9.2.1.1). Moreover, if the goal is to shorten the life expectancy of a self-sustaining biological machine, then transfecting in an inducible system for a

cannibalism-resistant mutant could be employed to speed up the muscle strip destabilization.

APPENDIX A. PACMANS GUI CODE

The PACMANS GUI developed in Chapter 3 [129] is available online for download at http://platt.gatech.edu/PACMANS. This GUI was developed in MATLAB as an application that can be download and run through MATLAB. The base code for the GUI is included in this Appendix.

A.1 PACMANS User Interface

Figure A-1. User interface for PACMANS. User inputs include: (1) the MEROPS link to the specificity matrix of the protease-ase and (2) the amino acid sequence for the substrate protease.

A.2 PACMANS Main GUI Code

```matlab
function varargout = PutSiteCleavageGUI(varargin)
% PUTSITECLEAVAGEGUI MATLAB code for PutSiteCleavageGUI.fig
%      PUTSITECLEAVAGEGUI, by itself, creates a new PUTSITECLEAVAGEGUI or
raises the existing
%      singleton*.
%
%      H = PUTSITECLEAVAGEGUI returns the handle to a new
PUTSITECLEAVAGEGUI or the handle to
%      the existing singleton*.
%
%      PUTSITECLEAVAGEGUI('CALLBACK',hObject,eventData,handles,...) calls the
local
%      function named CALLBACK in PUTSITECLEAVAGEGUI.M with the given
input arguments.
%
%      PUTSITECLEAVAGEGUI('Property','Value',...) creates a new
PUTSITECLEAVAGEGUI or raises the
%      existing singleton*.  Starting from the left, property value pairs are
%      applied to the GUI before PutSiteCleavageGUI_OpeningFcn gets called.  An
%      unrecognized property name or invalid value makes property application
%      stop.  All inputs are passed to PutSiteCleavageGUI_OpeningFcn via varargin.
%
%      *See GUI Options on GUIDE's Tools menu.  Choose "GUI allows only one
%      instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help PutSiteCleavageGUI

% Last Modified by GUIDE v2.5 22-Dec-2016 15:15:01

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @PutSiteCleavageGUI_OpeningFcn, ...
                   'gui_OutputFcn',  @PutSiteCleavageGUI_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
   gui_State.gui_Callback = str2func(varargin{1});
```

```matlab
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT


% --- Executes just before PutSiteCleavageGUI is made visible.
function PutSiteCleavageGUI_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to PutSiteCleavageGUI (see VARARGIN)

% Choose default command line output for PutSiteCleavageGUI
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% Adds in graphics
imshow('PlattLogo2.jpg', 'Parent', handles.axes6);
imshow('ProteaseAseSchematic.jpg', 'Parent', handles.axes7);

% --- Outputs from this function are returned to the command line.
function varargout = PutSiteCleavageGUI_OutputFcn(hObject, eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;

% --- Executes on button press in calcScoresButton.
function calcScoresButton_Callback(hObject, eventdata, handles)
% hObject    handle to calcScoresButton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
```

```matlab
% Getting loaded components
matrixLink = get(handles.SpecMat, 'string');
if(strncmpi(matrixLink, 'https://', 5))
    matrixLink = strrep(matrixLink, 'https://merops.sanger.ac.uk/cgi-bin/pepsum?id=', ...
        'http://meghan.ferrall-fairbanks.com/merops_data/');
elseif (strncmpi(matrixLink, 'http://', 4))
    matrixLink = strrep(matrixLink, 'http://merops.sanger.ac.uk/cgi-bin/pepsum?id=', ...
        'http://meghan.ferrall-fairbanks.com/merops_data/');
else
     matrixLink = strrep(matrixLink, 'merops.sanger.ac.uk/cgi-bin/pepsum?id=', ...
    'http://meghan.ferrall-fairbanks.com/merops_data/');
end

matrixLink = strcat(matrixLink, '.csv');
specmat = getSpecifMat(matrixLink);

sequence = get(handles.AAseq, 'string');

% Removes any spaces from sequence
% temp2 = sequence;
% temp1 = '';
%
% while ~strcmp(temp1, temp2)
%     temp1 = temp2;
%     temp2 = regexprep(temp1, ' ', '');
% end
%sequence = temp2;

% saving three components into matfile needed for csa_calscore
% followed along with example:
% http://www.mathworks.com/help/matlab/import_export/load-parts-of-variables-from-
mat-files.html
save MatForScoring.mat sequence specmat -v7.3;

% create output file name
% csa_analyze will store the scoring of cleavage sites in this file
%loadedData = load('MatForScoring.mat');
outputFileName = strcat(get(handles.protease, 'string'), 'on', get(handles.substrate, ...
'string'));

% convert bounds from strings to numbers
lowB = str2num(get(handles.lowBound, 'string'));
```

```matlab
highB = str2num(get(handles.highBound, 'string'));

% account for if something weird is entered or if nothing is entered
if  isempty(lowB)
    lowB = 1;
elseif (lowB < 1) || (lowB > 8)
    lowB = 1;
end

if  isempty(highB)
    highB = 8;
elseif (highB < 1) || (highB > 8)
    highB = 8;
end

% run csa_analyze code
% looking at all 8 amino acid combinations
csa_analyze('MatForScoring.mat', outputFileName, lowB, highB)


function SpecMat_Callback(hObject, eventdata, handles)
% hObject    handle to SpecMat (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of SpecMat as text
%        str2double(get(hObject,'String')) returns contents of SpecMat as a double


% --- Executes during object creation, after setting all properties.
function SpecMat_CreateFcn(hObject, eventdata, handles)
% hObject    handle to SpecMat (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

```matlab
function AAseq_Callback(hObject, eventdata, handles)
% hObject    handle to AAseq (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of AAseq as text
%        str2double(get(hObject,'String')) returns contents of AAseq as a double


% --- Executes during object creation, after setting all properties.
function AAseq_CreateFcn(hObject, eventdata, handles)
% hObject    handle to AAseq (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


function NTseq_Callback(hObject, eventdata, handles)
% hObject    handle to NTseq (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of NTseq as text
%        str2double(get(hObject,'String')) returns contents of NTseq as a double


% --- Executes during object creation, after setting all properties.
function NTseq_CreateFcn(hObject, eventdata, handles)
% hObject    handle to NTseq (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
```

```matlab
    set(hObject,'BackgroundColor','white');
end


% --- Executes on button press in loadDataButton.
function loadDataButton_Callback(hObject, eventdata, handles)
% hObject    handle to loadDataButton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Getting loaded components
 matrixLink = get(handles.SpecMat, 'string');
 if(strncmpi(matrixLink, 'https://', 5))
    matrixLink = strrep(matrixLink, 'https://merops.sanger.ac.uk/cgi-bin/pepsum?id=', ...
        'http://meghan.ferrall-fairbanks.com/merops_data/');
 elseif (strncmpi(matrixLink, 'http://', 4))
    matrixLink = strrep(matrixLink, 'http://merops.sanger.ac.uk/cgi-bin/pepsum?id=', ...
        'http://meghan.ferrall-fairbanks.com/merops_data/');
 else
     matrixLink = strrep(matrixLink, 'merops.sanger.ac.uk/cgi-bin/pepsum?id=', ...
    'http://meghan.ferrall-fairbanks.com/merops_data/');
 end
matrixLink = strcat(matrixLink, '.csv');
SpecMatr = getSpecifMat(matrixLink);
AminoAcidSeq = get(handles.AAseq, 'string');

% saving three components into matfile needed for csa_calscore
% followed along with example:
% http://www.mathworks.com/help/matlab/import_export/load-parts-of-variables-from-
mat-files.html
save MatForScoring.mat SpecMatr AminoAcidSeq -v7.3;



function protease_Callback(hObject, eventdata, handles)
% hObject    handle to protease (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
%

% name of the cannibal proteases
get(handles.protease, 'string');
```

```matlab
% Hints: get(hObject,'String') returns contents of protease as text
%        str2double(get(hObject,'String')) returns contents of protease as a double


% --- Executes during object creation, after setting all properties.
function protease_CreateFcn(hObject, eventdata, handles)
% hObject    handle to protease (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function substrate_Callback(hObject, eventdata, handles)
% hObject    handle to substrate (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% name of the substrate proteases
get(handles.substrate, 'string');

% Hints: get(hObject,'String') returns contents of substrate as text
%        str2double(get(hObject,'String')) returns contents of substrate as a double


% --- Executes during object creation, after setting all properties.
function substrate_CreateFcn(hObject, eventdata, handles)
% hObject    handle to substrate (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function lowBound_Callback(hObject, eventdata, handles)
```

185

```matlab
% hObject    handle to lowBound (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of lowBound as text
%        str2double(get(hObject,'String')) returns contents of lowBound as a double
get(handles.lowBound, 'string');

% --- Executes during object creation, after setting all properties.
function lowBound_CreateFcn(hObject, eventdata, handles)
% hObject    handle to lowBound (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


function highBound_Callback(hObject, eventdata, handles)
% hObject    handle to highBound (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of highBound as text
%        str2double(get(hObject,'String')) returns contents of highBound as a double
get(handles.highBound, 'string');

% --- Executes during object creation, after setting all properties.
function highBound_CreateFcn(hObject, eventdata, handles)
% hObject    handle to highBound (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

```matlab
% --- Executes during object creation, after setting all properties.
function axes6_CreateFcn(hObject, eventdata, handles)
% hObject    handle to axes6 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: place code in OpeningFcn to populate axes6
axes(hObject)
imshow('PlattLogo2.jpg')

% --- Executes during object creation, after setting all properties.
function axes7_CreateFcn(hObject, eventdata, handles)
% hObject    handle to axes7 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: place code in OpeningFcn to populate axes7
axes(hObject)
imshow('ProteaseAseSchematic.jpg')
```

A.3   PACMANS Supporting Functions Code

```matlab
function [ scores ] = csa_analyze( csafile, outfile, startp, endp )
%CSA_ANALYZE csafile outfile startp endp - perform cleavage site analysis on csafile
%   csafile is a matlab .mat file with two variables - sequence and specmat
%   sequence - text string of protein under analysis
%   specmat - cell matrix of specificity values for the enzyme
%   preference for substrate
%   matrix format:  column 1: amino acid letter code
%              columns 2-x: specificity numbers for substrate sites

if ~exist(csafile,'file')
    error('Input csafile does not exist.')
end

load(csafile);

if ~exist('sequence','var') || ~exist('specmat','var')
    error('Input csafile is not valid.');
end
```

```matlab
filehdl = fopen(strcat(outfile, '.txt') ,'w');

sequence = upper(sequence);

% Export protein sequence

fprintf(filehdl,'PROTEIN SEQUENCE\r\n================\r\n\r\n');

residuesperline = 60;
residuespergroup = 10; % Must be factor of residuesperline

seqlen = length(sequence);
numlines = ceil(seqlen / residuesperline);

for i = 1:numlines
    startpos = (i-1)*residuesperline + 1;

    endpos = 0;

    if i == numlines
        remainder = mod(seqlen,residuesperline);
        if remainder == 0
            remainder = residuesperline;
        end
        endpos = startpos + remainder - 1;
    else
        endpos = i*residuesperline;
    end

    curline = sequence(startpos:endpos);

    outline = '';
    curlinelen = length(curline);
    numgroups = ceil(curlinelen / residuespergroup);
    for j = 1:numgroups;
        linestartpos = (j-1)*residuespergroup + 1;
        lineendpos = 0;

        if j == numgroups
            remainder = mod(curlinelen,residuespergroup);
            if remainder == 0
                remainder = residuespergroup;
            end
```

```matlab
            lineendpos = linestartpos + remainder - 1;
        else
            lineendpos = j*residuespergroup;
        end

        outline = [outline, curline(linestartpos:lineendpos), ' '];
    end

    fprintf(filehdl,'%i\t%s\t%i\r\n', startpos,outline,endpos);
end

fprintf(filehdl,'\r\nMolecular weight: %6.2f\r\n',molweight(sequence));

fprintf(filehdl,'\r\nCLEAVAGE SITE
ANALYSIS\r\n=====================\r\n\r\n');

fprintf(filehdl,'Using site pockets %i through %i for analysis...\r\n\r\n', startp,endp);

scores = struct('sequence',{},'score',{}, ...
    'scoreNorm', {}, 'startpos',{},'endpos',{}, ...
    'eachScore', {}, 'rangeScores', {}, 'medianScore', {}, ...
    'leftcleavagelen',{},'rightcleavagelen',{});

% pre-define matrix for individual scores


testlen = endp - startp + 1;

numsequences = seqlen - testlen + 1;
% AA = zeros(1,testlen);
% scores.eachScore = AA;


singleLetterAA = {'G';'P';'A';'V';'L';'I';'M';'F';'Y';'W';'S';'T';'C';'N';'Q';'D';'E';'K';'R';'H'};

% Specificity matrix where the columns correspond to each subsite from P4'
% to P4 and the rows correspond to each of the 20 amino acids. Scores are
% from MEROPS, which are generated based on pure counts (i.e. out of all
% the substrate cleavages known, X number are known to the Gly in the P4'
% position, etc.)
specmat = specmat.data;

% Determines the maximum scores/count for each of the subsites P4' through
```

```matlab
% P4
totalNumSubstrates = sum(specmat,1);

for i = 1:numsequences
    curseq = sequence(i:i + testlen-1);
    % Calculate scores

    curscore = 0;
    curseqscore = zeros(1, testlen);
    curseqPscore = zeros(1, testlen);

    for j = 1:length(curseq)
        correctedJ = startp - 1 + j;
        specmatrow = find(strcmp(singleLetterAA,curseq(j)));
        curseqscore(1,j) = specmat(specmatrow,correctedJ);
        curseqPscore(1,j) = ...
specmat(specmatrow,correctedJ)/totalNumSubstrates(correctedJ);
        curscore = curscore + specmat(specmatrow,correctedJ);
    end

    % Calculating range and median scores
    maxScore = max(curseqscore);
    minScore = min(curseqscore);
    range = [minScore; maxScore];

    medianSco = median(curseqscore);

%%% Normalized Cumulative Score

    % Normalize matrix scores to the total num substrates used to fill the
    % MEROPS Specificity Matrix
    % Snormi = counti / total

    Snormi = 0;
    for ii = 1:length(curseq)
        Snormi = Snormi + curseqPscore (1, ii);
    end

%%
    % convert each score matrix to a string
    EachScoreStr = mat2str(curseqscore);

    % Fill in structure array
```

```matlab
        scores(i).sequence = curseq;
        scores(i).score = curscore;
        scores(i).scoreNorm = Snormi;
        scores(i).startpos = i;
        scores(i).endpos = i + testlen - 1;
        scores(i).eachScore = EachScoreStr;
        scores(i).medianScore = medianSco;
        scores(i).rangeScores = range;
        scores(i).leftcleavagelen = i + (testlen / 2) - 1;
        scores(i).rightcleavagelen = seqlen - (i + (testlen / 2) - 1);
        scores(i).leftcleavageweight = molweight(sequence(1:i + (testlen / 2) - 1)) / 1e3;
        scores(i).rightcleavageweight = molweight(sequence(i + (testlen / 2):end)) / 1e3;

end

fprintf(filehdl, 'Scores sorted by cumulative score \n');

[unused, order] = sort([scores(:).score],2,'descend');
scores = scores(order);

cumScores = zeros(length(scores), 2);

for i = 1:length(scores)

    fprintf(filehdl, '%i, \t', i);
    fprintf(filehdl,'Score: %i,\t Norm: %5.2f,\t Start Pos: %i\t %s\t End Pos: %i\t
Individual Scores: %s\t Range: [%5.2f,\t  %5.2f]\t Median: %5.2f\t (size,molweight): L:
(%i,%6.2f)\tR: (%i,%6.2f)\r\n',...
        scores(i).score, scores(i).scoreNorm, ...
        scores(i).startpos,scores(i).sequence,scores(i).endpos,...
        scores(i).eachScore, ...
        scores(i).rangeScores(1), scores(i).rangeScores(2), scores(i).medianScore,...
        scores(i).leftcleavagelen,scores(i).leftcleavageweight,...
        scores(i).rightcleavagelen,scores(i).rightcleavageweight);
    cumScores(i, 1) = scores(i).score;
    cumScores(i, 2) = scores(i).scoreNorm;

end

if length(scores) > 8
y1 = quantile(cumScores(:,1),0.95);
fprintf(filehdl, 'Top 95th percentile of scores cutoff: %i\t \n', y1 );
```

```matlab
figure
subplot(1,2,1);
h1 = histogram(cumScores(:,1))
title('Cumulative Likelihood Scores');
xlabel('Score');
ylabel('Frequency');
subplot(1,2,2);
h2 = histogram(cumScores(:,2))
title('Normalized Likelihood Scores');
xlabel('Normalized Score');
ylabel('Frequency');
end

fclose(filehdl);

fprintf('Done. Results have been saved to %s\n',outfile);

end




function [tot,scores] = csa_calcscore(csafile, seq)

    seq = upper(seq);

    if ~exist(csafile,'file')
        error('Input csafile does not exist.')
    end

    load(csafile);

    scores = [];

    for j = 1:length(seq)
        specmatrow = find(strcmp(specmat(:,1),seq(j)));

        scores = [scores, specmat{specmatrow,1+j}];
    end

    tot = sum(scores);
end
```

```matlab
function out = getSpecifMat(url)

%webopt = weboptions('ContentType', 'table');
%out = webread(url, webopt);


% matrixLink = get(handles.SpecMat, 'string');
% matrixLink = strrep(matrixLink, 'https://merops.sanger.ac.uk/cgi-bin/pepsum?id=', ...
%     'http://meghan.ferrall-fairbanks.com/merops_data/');
% matrixLink = strcat(matrixLink, '.csv');

urlwrite(url,'myFileName.csv');
out=importdata('myFileName.csv');

%data = webread(matrixLink);
%SpecMat = readtable(data, 'Delimiter',',');
%SpecMat = cell2table(data);

%out = load('specmat.mat');

end
```

APPENDIX B. FULL PACMANS TOP SCORING TABLES FOR

CATHEPSINS ON FIBRINOGEN CHAINS

These are the supplemental tables of the full top 5% of PACMANS scores for fibrinogen α, β, and γ chains cleaved by cathepsins K (Tables B-1 through B-3), L (Tables B-4 through B-6), and S (Tables B-7 through B-9) submitted with *Ferrall-Fairbanks, M.C.*, West, D. M.*, et al. Computational predictions of cysteine cathepsin-mediated fibrinogen degradation.* Blue sequences are those with molecular docking model with the active site facing the fibrinogen chain included in the paper and Chapter 4.

B.1   Cathepsin K on Fibrinogen α, β, and γ Chains

Table B-1. Top 5% of PACMANS predicted fibrinogen α chain cleavage sites by cathepsin K from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
182	1.06	26	DFLAEGGG	33	not in crystal structure
170	1.00	821	YYPGGSYD	828	not in crystal structure
168	0.96	167	KRLEVDID	174	allosteric/hairpin
165	0.96	667	TSLGGWLL	674	not in crystal structure
156	0.90	186	RALAREVD	193	allosteric/hairpin
155	0.91	111	EILRGDFS	118	active site facing
153	0.90	239	SQLQKVPP	246	not in crystal structure
152	0.87	202	KQLEQVIA	209	active site facing
150	0.88	490	DCPEAMDL	497	not in crystal structure
150	0.88	769	GSVEEGAE	776	not in crystal structure
145	0.83	259	MELERPGG	266	not in crystal structure
145	0.84	726	VELEDWAG	733	not in crystal structure
144	0.84	262	ERPGGNEI	269	not in crystal structure
144	0.86	759	EGTAGDAL	766	not in crystal structure
143	0.82	535	PMLGEFVS	542	not in crystal structure
142	0.83	498	GTLSGIGT	505	not in crystal structure
141	0.84	396	DSPGSGNA	403	not in crystal structure
141	0.83	764	DALIEGSV	771	not in crystal structure

Table B-1 continued.

Cumulative Score	Normalized Score		Fibrinogen Sequence		Inclusion/Exclusion Rationale
139	0.81	132	EDLRSRIE	139	active site facing
138	0.82	244	VPPEWKAL	251	not in crystal structure
138	0.82	563	HHPGIAEF	570	not in crystal structure
136	0.79	27	FLAEGGGV	34	not in crystal structure
136	0.80	694	RGFGSLND	701	not in crystal structure
135	0.83	73	LIDEVNQD	80	active site facing
135	0.80	192	VDLKDYED	199	allosteric/hairpin
135	0.78	650	IKLPGSSK	657	not in crystal structure
135	0.79	765	ALIEGSVE	772	not in crystal structure
134	0.80	715	HLLTQRGS	722	not in crystal structure
133	0.79	21	DSGEGDFL	28	not in crystal structure
133	0.79	211	DLLPSRDR	218	no docking models near
133	0.81	699	LNDEGEGE	706	not in crystal structure
132	0.80	844	VSFRGADY	851	not in crystal structure
131	0.79	9	LVLSVVGT	16	not in crystal structure
131	0.77	565	PGIAEFPS	572	not in crystal structure
130	0.78	103	HSLTTNIM	110	no docking models near
130	0.75	495	MDLGTLSG	502	not in crystal structure
129	0.77	32	GGVRGPRV	39	not in crystal structure
128	0.75	90	NSLFEYQK	97	active site facing
128	0.76	377	SSVSGSTG	384	not in crystal structure
128	0.75	627	RPVRDCDD	634	not in crystal structure
127	0.75	300	SGPGSTGN	307	not in crystal structure
127	0.75	326	SGPGSTGS	333	not in crystal structure
127	0.77	724	LRVELEDW	731	not in crystal structure

Table B-2. Top 5% of PACMANS predicted fibrinogen β chain cleavage sites by cathepsin K from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
169	0.98	110	CQLQEALL	117	no docking models near
158	0.92	210	QKLESDVS	217	no docking models near
156	0.92	94	GCLHADPD	101	no docking models near
155	0.91	339	ELLIEMED	346	globular
148	0.85	174	SELEKHQL	181	no docking models near
146	0.86	388	SQLMGENR	395	globular
144	0.84	57	PSLRPAPP	64	not in crystal structure
144	0.83	200	SILENLRS	207	active site facing
141	0.85	179	HQLYIDET	186	active site facing
139	0.82	376	RGTAGNAL	383	globular
137	0.82	28	VKSQGVND	35	not in crystal structure
137	0.84	105	LCPTGCQL	112	active site facing
136	0.82	59	LRPAPPPI	66	not in crystal structure
133	0.78	203	ENLRSKIQ	210	active site facing
132	0.78	304	GNVATNTD	311	globular
132	0.78	335	MGPTELLI	342	globular
129	0.75	316	CGLPGEYW	323	globular
127	0.76	89	APDAGGCL	96	no docking models near
127	0.75	317	GLPGEYWL	324	globular
127	0.73	322	YWLGNDKI	329	globular
126	0.73	115	ALLQQERP	122	no docking models near
125	0.73	254	MYLIQPDS	261	globular
125	0.73	414	GWLTSDPR	421	globular
124	0.76	20	LLLLCVFL	27	not in crystal structure
124	0.77	116	LLQQERPI	123	no docking models near
123	0.71	100	PDLGVLCP	107	no docking models near

Table B-3. Top 5% of PACMANS predicted fibrinogen γ chain cleavage sites by cathepsin K from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
166	0.98	426	HHLGGAKQ	433	not in crystal structure
160	0.92	222	KRLDGSVD	229	globular
156	0.89	64	KDLQSLED	71	allosteric
156	0.91	274	VELEDWNG	281	globular
154	0.90	90	IQLTYNPD	97	active site facing
153	0.91	445	DSLYPEDD	452	not in crystal structure
150	0.89	305	AYFAGGDA	312	globular
148	0.84	114	KMLEEIMK	121	active site facing
147	0.86	67	QSLEDILH	74	allosteric
147	0.85	155	AQLEAQCQ	162	active site facing
143	0.83	416	NRLTIGEG	423	globular
141	0.82	134	RYLQEIYN	141	active site facing
139	0.80	35	CILDERFG	42	not in crystal structure
137	0.83	438	HPAETEYD	445	not in crystal structure
136	0.83	244	LSPTGTTE	251	globular
135	0.80	148	VNLKEKVA	155	no docking models near
135	0.80	293	VGPEADKY	300	globular
134	0.82	317	DGFDFGDD	324	globular
132	0.78	26	AYVATRDN	33	not in crystal structure
130	0.78	18	LFLSSTCV	25	not in crystal structure
128	0.75	3	WSLHPRNL	10	not in crystal structure
127	0.77	272	LRVELEDW	279	globular
125	0.73	368	GHLNGVYY	375	globular

B.2 Cathepsin L on Fibrinogen α, β, and γ Chains

Table B-4. Top 5% of PACMANS predicted fibrinogen α chain cleavage sites by cathepsin L from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
1006	0.95	498	GTLSGIGT	505	not in crystal structure
962	0.90	667	TSLGGWLL	674	not in crystal structure
944	0.89	377	SSVSGSTG	384	not in crystal structure
941	0.88	432	KLVTSKGD	439	not in crystal structure
921	0.88	765	ALIEGSVE	772	not in crystal structure
914	0.86	742	FRVGSEAE	749	not in crystal structure
909	0.86	239	SQLQKVPP	246	not in crystal structure
897	0.84	694	RGFGSLND	701	not in crystal structure
888	0.83	26	DFLAEGGG	33	not in crystal structure
887	0.83	495	MDLGTLSG	502	not in crystal structure
878	0.83	481	EVVTSEDG	488	not in crystal structure
878	0.82	634	DVLQTHPS	641	not in crystal structure
877	0.83	447	EKVTSGST	454	not in crystal structure
863	0.82	111	EILRGDFS	118	active site facing
852	0.80	8	CLVLSVVG	15	not in crystal structure
852	0.81	32	GGVRGPRV	39	not in crystal structure
846	0.80	326	SGPGSTGS	333	not in crystal structure
844	0.80	313	SGTGGTAT	320	not in crystal structure
837	0.80	769	GSVEEGAE	776	not in crystal structure
834	0.78	715	HLLTQRGS	722	not in crystal structure
832	0.79	756	SSYEGTAG	763	not in crystal structure
830	0.78	467	KTVIGPDG	474	not in crystal structure
829	0.78	259	MELERPGG	266	not in crystal structure
829	0.79	297	SGSSGPGS	304	not in crystal structure
828	0.79	414	EEVSGNVS	421	not in crystal structure
828	0.78	535	PMLGEFVS	542	not in crystal structure
823	0.78	697	GSLNDEGE	704	not in crystal structure
820	0.77	347	QNPGSPRP	354	not in crystal structure
816	0.77	501	SGIGTLDG	508	not in crystal structure
816	0.78	545	ESRGSESG	552	not in crystal structure
815	0.76	295	WNSGSSGP	302	not in crystal structure
806	0.76	323	PGSSGPGS	330	not in crystal structure

Table B-4 continued.

Cumulative Score	Normalized Score	Fibrinogen Sequence			Inclusion/Exclusion Rationale
805	0.76	12	SVVGTAWT	19	not in crystal structure
805	0.75	186	RALAREVD	193	no docking models near
805	0.76	300	SGPGSTGN	307	not in crystal structure
804	0.76	9	LVLSVVGT	16	not in crystal structure
803	0.76	650	IKLPGSSK	657	not in crystal structure
803	0.76	815	ANLNGIYY	822	not in crystal structure
800	0.76	753	LQVSSYEG	760	not in crystal structure
794	0.75	504	GTLDGFRH	511	not in crystal structure
791	0.75	339	SGTGSTGN	346	not in crystal structure
789	0.75	132	EDLRSRIE	139	active site facing
789	0.74	321	WKPGSSGP	328	not in crystal structure

Table B-5. Top 5% of PACMANS predicted fibrinogen β chain cleavage sites by cathepsin L from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
936	0.89	115	ALLQQERP	122	no docking models near
894	0.85	94	GCLHADPD	101	no docking models near
894	0.85	414	GWLTSDPR	421	globular
885	0.83	57	PSLRPAPP	64	not in crystal structure
862	0.81	26	FLVKSQGV	33	not in crystal structure
842	0.79	210	QKLESDVS	217	no docking models near
840	0.79	100	PDLGVLCP	107	no docking models near
839	0.80	185	ETVNSNIP	192	active site facing
823	0.78	330	SQLTRMGP	337	globular
822	0.78	339	ELLIEMED	346	globular
815	0.78	38	EGFFSARG	45	not in crystal structure
813	0.76	196	RVLRSILE	203	active site facing
802	0.77	58	SLRPAPPP	65	not in crystal structure
795	0.76	203	ENLRSKIQ	210	no docking models near
789	0.74	64	PPISGGGY	71	not in crystal structure
782	0.74	233	IPVVSGKE	240	globular
782	0.74	301	QGFGNVAT	308	globular
782	0.74	388	SQLMGENR	395	globular
776	0.74	135	EAVSQTSS	142	active site facing
776	0.73	381	NALMDGAS	388	globular
764	0.71	110	CQLQEALL	117	no docking models near
750	0.71	200	SILENLRS	207	active site facing
745	0.71	214	SDVSAQME	221	no docking models near
744	0.70	316	CGLPGEYW	323	globular
742	0.70	234	PVVSGKEC	241	globular
739	0.70	102	LGVLCPTG	109	no docking models near

Table B-6. Top 5% of PACMANS predicted fibrinogen γ chain cleavage sites by cathepsin L from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
989	0.92	426	HHLGGAKQ	433	not in crystal structure
915	0.87	239	EGFGHLSP	246	globular
894	0.83	64	KDLQSLED	71	no docking models near
877	0.83	16	ALLFLSST	23	not in crystal structure
861	0.81	18	LFLSSTCV	25	not in crystal structure
861	0.80	222	KRLDGSVD	229	globular
840	0.80	125	SILTHDSS	132	active site facing
828	0.78	259	IHLISTQS	266	globular
827	0.78	39	ERFGSYCP	46	not in crystal structure
804	0.76	305	AYFAGGDA	312	globular
803	0.76	155	AQLEAQCQ	162	allosteric facing
801	0.75	90	IQLTYNPD	97	active site facing
792	0.75	209	EIDGSGNG	216	globular
786	0.75	190	SGLYFIKP	197	globular
785	0.75	242	GHLSPTGT	249	globular
785	0.75	368	GHLNGVYY	375	globular
770	0.72	306	YFAGGDAG	313	globular
769	0.73	17	LLFLSSTC	24	not in crystal structure
766	0.72	262	ISTQSAIP	269	globular
759	0.71	416	NRLTIGEG	423	globular
758	0.71	291	FKVGPEAD	298	globular
748	0.70	53	DFLSTYQT	60	allosteric facing
748	0.71	108	ATLKSRKM	115	active site facing

B.3 Cathepsin S on Fibrinogen α, β, and γ Chains

Table B-7. Top 5% of PACMANS predicted fibrinogen α chain cleavage sites by cathepsin S from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
735	0.99	26	DFLAEGGG	33	not in crystal structure
696	0.94	498	GTLSGIGT	505	not in crystal structure
685	0.93	377	SSVSGSTG	384	not in crystal structure
682	0.93	481	EVVTSEDG	488	not in crystal structure
674	0.91	769	GSVEEGAE	776	not in crystal structure
665	0.89	742	FRVGSEAE	749	not in crystal structure
642	0.87	239	SQLQKVPP	246	not in crystal structure
638	0.86	259	MELERPGG	266	not in crystal structure
637	0.85	432	KLVTSKGD	439	not in crystal structure
636	0.86	634	DVLQTHPS	641	not in crystal structure
636	0.86	726	VELEDWAG	733	not in crystal structure
627	0.85	447	EKVTSGST	454	not in crystal structure
624	0.84	495	MDLGTLSG	502	not in crystal structure
621	0.84	697	GSLNDEGE	704	not in crystal structure
621	0.83	715	HLLTQRGS	722	not in crystal structure
618	0.83	667	TSLGGWLL	674	not in crystal structure
607	0.82	9	LVLSVVGT	16	not in crystal structure
597	0.81	753	LQVSSYEG	760	not in crystal structure
596	0.80	535	PMLGEFVS	542	not in crystal structure
595	0.81	414	EEVSGNVS	421	not in crystal structure
595	0.80	764	DALIEGSV	771	not in crystal structure
592	0.81	111	EILRGDFS	118	active site facing
591	0.79	249	KALTDMPQ	256	not in crystal structure
590	0.80	12	SVVGTAWT	19	not in crystal structure
588	0.79	296	NSGSSGPG	303	not in crystal structure
588	0.80	765	ALIEGSVE	772	not in crystal structure
583	0.78	8	CLVLSVVG	15	not in crystal structure
581	0.79	800	AEVYGGGW	807	not in crystal structure
579	0.79	365	SERGSAGH	372	not in crystal structure
578	0.79	539	EFVSETES	546	not in crystal structure
576	0.78	132	EDLRSRIE	139	active site facing
572	0.77	211	DLLPSRDR	218	no docking models near

Table B-7 continued.

Cumulative Score	Normalized Score	Fibrinogen Sequence			Inclusion/Exclusion Rationale
572	0.76	467	KTVIGPDG	474	not in crystal structure
567	0.76	751	YALQVSSY	758	not in crystal structure
567	0.77	756	SSYEGTAG	763	not in crystal structure
565	0.77	326	SGPGSTGS	333	not in crystal structure
565	0.76	605	DEAGSEAD	612	not in crystal structure
564	0.77	313	SGTGGTAT	320	not in crystal structure
560	0.76	815	ANLNGIYY	822	not in crystal structure
559	0.75	322	KPGSSGPG	329	not in crystal structure
556	0.76	339	SGTGSTGN	346	not in crystal structure
555	0.74	186	RALAREVD	193	no docking models near
553	0.75	545	ESRGSESG	552	not in crystal structure

Table B-8. Top 5% of PACMANS predicted fibrinogen β chain cleavage sites by cathepsin S from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
652	0.89	115	ALLQQERP	122	no docking models near
646	0.87	94	GCLHADPD	101	no docking models near
644	0.87	414	GWLTSDPR	421	globular
639	0.86	57	PSLRPAPP	64	not in crystal structure
614	0.84	135	EAVSQTSS	142	active site facing
611	0.82	381	NALMDGAS	388	globular
605	0.82	210	QKLESDVS	217	no docking models near
585	0.80	114	EALLQQER	121	no docking models near
579	0.77	110	CQLQEALL	117	no docking models near
578	0.79	339	ELLIEMED	346	globular
577	0.78	330	SQLTRMGP	337	globular
575	0.77	196	RVLRSILE	203	active site facing
573	0.77	100	PDLGVLCP	107	no docking models near
572	0.78	203	ENLRSKIQ	210	active site facing
571	0.78	185	ETVNSNIP	192	active site facing
570	0.77	26	FLVKSQGV	33	not in crystal structure
567	0.77	214	SDVSAQME	221	no docking models near
563	0.76	132	NNVEAVSQ	139	active site facing
563	0.76	233	IPVVSGKE	240	globular
552	0.75	200	SILENLRS	207	active site facing
548	0.74	64	PPISGGGY	71	not in crystal structure
546	0.74	25	VFLVKSQG	32	not in crystal structure
546	0.74	234	PVVSGKEC	241	globular
546	0.74	388	SQLMGENR	395	globular
540	0.73	128	DELNNNVE	135	active site facing
539	0.72	254	MYLIQPDS	261	globular

Table B-9. Top 5% of PACMANS predicted fibrinogen γ chain cleavage sites by cathepsin S from PACMANS and how sequences relate to molecular docking models.

Cumulative Score	Normalized Score	Fibriongen Sequence			Inclusion/Exclusion Rationale
687	0.92	426	HHLGGAKQ	433	not in crystal structure
634	0.86	18	LFLSSTCV	25	not in crystal structure
632	0.86	155	AQLEAQCQ	162	active site facing
630	0.84	64	KDLQSLED	71	no binding
604	0.81	90	IQLTYNPD	97	no docking models near
595	0.80	416	NRLTIGEG	423	globular
593	0.80	274	VELEDWNG	281	globular
589	0.80	209	EIDGSGNG	216	globular
584	0.79	125	SILTHDSS	132	active site facing
581	0.79	16	ALLFLSST	23	not in crystal structure
580	0.77	222	KRLDGSVD	229	globular
579	0.78	242	GHLSPTGT	249	globular
570	0.77	259	IHLISTQS	266	globular
561	0.76	305	AYFAGGDA	312	globular
550	0.74	67	QSLEDILH	74	no docking models near
537	0.73	53	DFLSTYQT	60	active site facing
536	0.72	71	DILHQVEN	78	no docking models near
536	0.72	291	FKVGPEAD	298	globular
535	0.73	368	GHLNGVYY	375	globular
532	0.71	35	CILDERFG	42	not in crystal structure
531	0.71	196	KPLKANQQ	203	globular
529	0.72	445	DSLYPEDD	452	not in crystal structure
524	0.70	306	YFAGGDAG	313	globular

APPENDIX C. BIOBOT LIFE EXPECTANCY MODEL CODE

This appendix has the MATLAB code for estimating the Weibull distribution parameters for the 0X, 1X, and 3X ACA treated biological machine life expectancy data. The code for the protease-informed life expectancy model and corresponding grid search are also included in this Appendix.

C.1   Weibull Distribution MATLAB CODE

```matlab
function pd = fittingWeibull(ACA)
% Biobot time until breaking can be estimated as a Weibull distribution
%
% Weibull: y = a*b*x^(b-1)*exp(-1*a*x^b)
% parameters to fit: a & b

% Determining Weibull parameters for each dataset:
%     data is the age at which the biobots break
%     outliers removed: 0X - 1 lived to 38days
%                       1X - 1 lived to 164 & 1 to 233 days

data0X = [6, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 11, 11]';
data1X = [11, 11, 13, 16, 17, 34]';
data3X = [9, 11, 23, 28, 32, 54, 107, 198, 228, 250]';

if ACA == 0
   pd= fitdist(data0X, 'Weibull')
elseif ACA == 1
   pd= fitdist(data1X, 'Weibull')
elseif ACA == 3
   pd= fitdist(data3X, 'Weibull')
end

% Results for Fitted Weibull Distribution
%
% 0X Weibull Distribution
% A = 8.80037   [7.93074, 9.76535]
% B = 5.55191   [3.71228, 8.30317]
```

```matlab
%
% 1X Weibull Distribution
% A = 19.2933   [13.2657, 28.0598]
% B = 2.27744   [1.28885, 4.02432]
%
% 3X Weibull Distribution
% A =  92.4197   [46.7841, 182.571]
% B = 0.963603   [0.589746, 1.57446]
%
end


function plotBiobotWeibull()

% 0X ACA
obstime0X = [6, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 11, 11]';
figure()
T = 11;
subplot(1,2,1);
[empF,x,empFlo,empFup] = ecdf(obstime0X);
stairs(x,empF);
hold on;
stairs(x,empFlo,':'); stairs(x,empFup,':');
hold off
xlabel('Time'); ylabel('Proportion failed'); title('Empirical CDF')
subplot(1,2,2);
paramEsts = wblfit(obstime0X);
[nlogl,paramCov] = wbllike(paramEsts,obstime0X);
xx = linspace(1,2*T,500);
[wblF,wblFlo,wblFup] = wblcdf(xx,paramEsts(1),paramEsts(2),paramCov);
stairs(x,empF);
hold on
handles = plot(xx,wblF,'r-',xx,wblFlo,'r:',xx,wblFup,'r:');
hold off
xlabel('Time'); ylabel('Fitted failure probability'); title('Weibull Model vs. Empirical')

% 1X ACA
obstime1X = [11, 11, 13, 16, 17, 34]';
figure()
T = 34;
subplot(1,2,1);
[empF,x,empFlo,empFup] = ecdf(obstime1X);
stairs(x,empF);
```

```matlab
hold on;
stairs(x,empFlo,':'); stairs(x,empFup,':');
hold off
xlabel('Time'); ylabel('Proportion failed'); title('Empirical CDF')
subplot(1,2,2);
paramEsts1X = wblfit(obstime1X);
[nlogl,paramCov1X] = wbllike(paramEsts1X,obstime1X);
xx = linspace(1,2*T,500);
[wblF,wblFlo,wblFup] = wblcdf(xx,paramEsts1X(1),paramEsts1X(2),paramCov1X);
stairs(x,empF);
hold on
handles = plot(xx,wblF,'r-',xx,wblFlo,'r:',xx,wblFup,'r:');
hold off
xlabel('Time'); ylabel('Fitted failure probability'); title('Weibull Model vs. Empirical')

% 3X ACA
obstime3X = [9, 11, 23, 28, 32, 54, 107, 198, 228, 250]';
figure()
T = 250;
subplot(1,2,1);
[empF,x,empFlo,empFup] = ecdf(obstime3X);
stairs(x,empF);
hold on;
stairs(x,empFlo,':'); stairs(x,empFup,':');
hold off
xlabel('Time'); ylabel('Proportion failed'); title('Empirical CDF')
subplot(1,2,2);
paramEsts3X = wblfit(obstime3X);
[nlogl,paramCov3X] = wbllike(paramEsts3X,obstime3X);
xx = linspace(1,2*T,500);
[wblF,wblFlo,wblFup] = wblcdf(xx,paramEsts3X(1),paramEsts3X(2),paramCov3X);
stairs(x,empF);
hold on
handles = plot(xx,wblF,'r-',xx,wblFlo,'r:',xx,wblFup,'r:');
hold off
xlabel('Time'); ylabel('Fitted failure probability'); title('Weibull Model vs. Empirical')
end
```

C.2 Bio-Bot Life Expectancy Model MATLAB Code

```matlab
function lifeSpan = biobotModel20170626(ProbCleav,ProbDead, ACA, writegraph, plotgraph)
% Biobot Data
```

```matlab
% 0X outlier - removed 1 bot that lived for 38 days
t0X = [6, 7, 8, 9, 11];
d0X = [1, 5, 2, 3, 2];
r0X = [13, 12, 7, 5, 2];
lifeTable0X = biobotSurvival(t0X,d0X,r0X);
average0X = mean(t0X);
stdDev0X = std(t0X);
SEM0X = stdDev0X/sqrt(size(t0X,2)-1);

% 1X outlier - removed 2 bot that lived for 233 & 164 days
t1X = [11, 13, 16, 17, 34];
d1X = [2, 1, 1, 1, 1];
r1X = [6, 4, 3, 2, 1];
lifeTable1X = biobotSurvival(t1X,d1X,r1X);
average1X = mean(t1X);
stdDev1X = std(t1X);
SEM1X = stdDev1X/sqrt(size(t1X,2)-1);

% 3X outlier - additional timepoints and no outliers
t3X = [9, 11, 23, 28, 32, 54, 107, 198, 228, 250];
d3X = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1];
r3X = [10, 9, 8, 7, 6, 5, 4, 3, 2, 1];
lifeTable3X = biobotSurvival(t3X,d3X,r3X);
average3X = mean(t3X);
stdDev3X = std(t3X);
SEM3X = stdDev3X/sqrt(size(t3X,2)-1);

lifeSpanData = [average0X, stdDev0X, SEM0X;...
    average1X, stdDev1X, SEM1X;...
    average3X, stdDev3X, SEM3X];

if ACA == 0
    lifeSpanData(1,:);
elseif ACA == 1
    lifeSpanData(2,:);
elseif ACA == 3
    lifeSpanData(3,:);
end

% Data for formatted for Weibull distribution
if ACA == 0
    obstime = [6, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 11, 11]';
elseif ACA == 1
```

```matlab
    obstime = [11, 11, 13, 16, 17, 34]';
elseif ACA == 3
    obstime = [9, 11, 23, 28, 32, 54, 107, 198, 228, 250]';
end

% Define muscle strip dimensions (scale 1/10 mm, 2D)
RanNum = 1;
MSwidth = 40;
MSlength = 80;
MSdepth = 1;

iterNum = 200;
lifeSpanVec = zeros(iterNum,1);
if ACA == 0
    fileName = 'Biobot-Failure-Sim_0X_2017-06-26.txt';
elseif ACA == 1
    fileName = 'Biobot-Failure-Sim_1X_2017-06-26.txt';
elseif ACA == 3
    fileName = 'Biobot-Failure-Sim_3X_2017-06-26.txt';
end

% M = 'Biobot Lifespan Failure Model Simulation';
% dlmwrite(fileName,M);

fnameBase = 'Graph_Edges_2017-06-26_';
g=graph(laplacian3(MSlength,MSwidth,MSdepth), 'OmitSelfLoops');
g=remove_posts(g, MSlength, MSwidth, MSdepth);
fname = strcat(fnameBase,num2str(0),'.txt');
write_graph(g,fname);
        if plotgraph
            gplot(adjacency(g), graphCoords);
            graphvizfname = strcat(fnameBase, 'viz.', ...
                        num2str(0), '.fig');
            savefig(graphvizfname);
        end
variabilityCleave = 0.75 + (1.25-0.75)*rand;
for a = 1:iterNum
    % Set loop conditions
    cond = 0;
    condStop = 1;
    p = ProbCleav * variabilityCleave;
    % create 3D grid
    g=graph(laplacian3(MSlength,MSwidth,MSdepth), 'OmitSelfLoops');
```

```matlab
g = remove_posts(g, MSlength, MSwidth, MSdepth);
if writegraph
    [X,Y] = ind2sub([MSlength, MSwidth], 1:size(g.Nodes,1));
    graphCoords = [transpose(X), transpose(Y)];
    fnameCoords = strcat(fnameBase,'Coordinates.csv');
    dlmwrite(fnameCoords, graphCoords);
end
iteration_number = 0;
while (cond < condStop)
    iteration_number = iteration_number + 1;
    if ACA == 1 && iteration_number > 164
        disp(['strip reached maximum iteration_number: ', num2str(iteration_number)])
        break
    end
    if ACA == 3 && iteration_number > 250
        disp(['strip reached maximum iteration_number: ', num2str(iteration_number)])
        break
    end
    for v = 1:numnodes(g)
        % need to search through length & width to get to all graph
        % vertices
        % identify the neighbors
        neighb = neighbors(g,v);
        incIfLessEdges = size(neighb,1);
        % Generate random number based on protease
        RanNum = proteaseComp20170521(iteration_number, ACA);
        % increase prob of cleavage proportional to number of
        % edges cleaved
        if incIfLessEdges ~= 0
            RanNum = RanNum*(1-(1/incIfLessEdges)+(1/4));
        end
        % Determine if cleavage occurs
        if RanNum <= p
            % if cleaved remove one edge from (i,j)'s
            % neighborhood
            % if there are no more neighbors skip it.
            if size(neighb,1) == 0
                continue
            end
            % if there is/are neighbors, choose one at random
            nN = randi([1,size(neighb,1)],1);
            %nN = 1;
            assert(nN <= size(neighb, 1), 'random vertex too big')
```

```matlab
                neighRm = neighb(nN);
                % remove that neighbor
                g = rmedge(g, v, neighRm);
            end
        end
        %% check if graph is still connected
        % first computer connected components
        cc = conncomp(g);
        % record the edges for the graph
        if writegraph;
            fname = strcat(fnameBase,num2str(iteration_number),'.txt');
            write_graph(g,fname);
            if plotgraph
                gplot(adjacency(g), graphCoords);
                graphvizfname = strcat(fnameBase, 'viz.', ...
                                num2str(iteration_number), '.fig');
                savefig(graphvizfname);
            end
        end
        % the biobot is broken if the largest component contains no more
        % than 1/4 of the fibrin molecules
        variabilityBreak = 0.75 + (1.25-0.75)*rand;
        if max(histc(cc, unique(cc))) < numnodes(g)*ProbDead*variabilityBreak
            disp('Graph is disconnected making plot');
            %figure()
            %gplot(adjacency(g), gcoords); % turn on here to see the final graph
            %figure()
            %plot(g)
            cond = condStop;
            X = ['strip broke on iteration: ', num2str(iteration_number)];
            disp(X)
        end
    end

    lifeSpanVec(a,1) = iteration_number;

end

dlmwrite(fileName, lifeSpanVec, '-append');

% figure()
% T = 11;
% xx = linspace(1,2*T,500);
```

212

```matlab
%
% subplot(1,2,1)
% obstime0X = [6, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 11, 11]';
% [empF0X,x0X,empFlo0X,empFup0X] = ecdf(obstime0X);
% params0X = wblfit(obstime0X);
% [nlogl,paramCov0X] = wbllike(params0X,obstime0X);
% [wblF0X,wblFlo0X,wblFup0X] =
wblcdf(xx,params0X(1),params0X(2),paramCov0X);
% stairs(x0X,empF0X);
% hold on
% handles = plot(xx,wblF0X,'r-',xx,wblFlo0X,'r:',xx,wblFup0X,'r:');
% hold off
% xlabel('Time'); ylabel('Fitted failure probability'); title('Weibull Model vs. Empirical')
% title('0X Biobot Data')
%
% subplot(1,2,2)
% [empF,x,empFlo,empFup] = ecdf(lifeSpanVec);
WeibullParams = wblfit(lifeSpanVec);
% [nlogl,paramCov] = wbllike(WeibullParams,lifeSpanVec);
% [wblF,wblFlo,wblFup] = wblcdf(xx,WeibullParams(1),WeibullParams(2),paramCov);
% stairs(x,empF);
% hold on
% handles = plot(xx,wblF,'r-',xx,wblFlo,'r:',xx,wblFup,'r:');
% hold off
% xlabel('Time'); ylabel('Fitted failure probability'); title('Weibull Model vs. Empirical')
% title('Simulated 0X Biobot Data')

lifeSpan = [WeibullParams(1), WeibullParams(2)];

end


function amt = proteaseComp20170521 (day, ACA)

% 0X ACA
mCatL0Xdays = [3 6 6 8 8 12 12];
mCatL0Xng = [5.25922 12.0436 20.6688 12.7273 ...
   21.3525 14.8836 26.0332];

MMP20Xdays = [3 6 6 8 8 12 12];
MMP20Xng = [0.0057 0.002988 0.004812 0.0044 ...
   0.0088 0.0068 0.0118];
```

```matlab
mCatL0Xhi = @(t) 2.14*t + 2.8135;
mCatL0Xlo = @(t) 1.002*t + 3.9638;

MMP20Xhi = @(t) 0.0008*t + 0.0023;
MMP20Xlo = @(t) 0.0002*t + 0.0037;

% 1X ACA
mCatL1Xdays = [3 6 6 8 8 12 12];
mCatL1Xng = [5.25922 3.103 3.8392 2.2615 ...
    12.5695 3.4711 5.8903];

MMP21Xdays = [3 6 6 8 8 12 12];
MMP21Xng = [0.0057 0.005045 0.006755 0.003916 ...
    0.005284 0.0046 0.0082];

mCatL1Xhi = @(t) 0.2399*t + 5.1503;
mCatL1Xlo = @(t) -0.1882*t + 4.8883;

MMP21Xhi = @(t) 0.0002*t + 0.0047;
MMP21Xlo = @(t) -0.0001*t + 0.0058;

% 3X ACA
mCatL3Xdays = [3 6 6 8 8 12 12];
mCatL3Xng = [5.25922 2.0511 4.4703 3.4185 ...
    9.6243 2.1563 5.6273];

MMP23Xdays = [3 6 6 8 8 12 12];
MMP23Xng = [0.0057 0.0064 0.0098 0.003773 ...
    0.005027 0.0041 0.0101];

mCatL3Xhi = @(t) 0.1405*t + 5.2263;
mCatL3Xlo = @(t) -0.2833*t + 5.2749;

MMP23Xhi = @(t) 0.0004*t + 0.0051;
MMP23Xlo = @(t) -0.0002*t + 0.0067;

% Determine probability of cleavage
amt = 0;
if day<3
   amt = rand;
else
   if ACA == 0
      maxCatL = mCatL0Xhi(day);
```

```matlab
            minCatL = mCatL0Xlo(day);
            rangeCatL = maxCatL - minCatL;
            maxMMP2 = MMP20Xhi(day);
            minMMP2 = MMP20Xlo(day);
            rangeMMP2 = maxMMP2 - minMMP2;
            catL = mCatL0Xlo(day) + rangeCatL*rand(1,1);
            MMP2 = MMP20Xlo(day) + rangeMMP2*rand(1,1);

        elseif ACA == 1
            maxCatL = mCatL1Xhi(day);
            minCatL = mCatL1Xlo(day);
            rangeCatL = maxCatL - minCatL;
            maxMMP2 = MMP21Xhi(day);
            minMMP2 = MMP21Xlo(day);
            rangeMMP2 = maxMMP2 - minMMP2;
            catL = mCatL1Xlo(day) + rangeCatL*rand(1,1);
            MMP2 = MMP21Xlo(day) + rangeMMP2*rand(1,1);

        elseif ACA == 3
            maxCatL = mCatL3Xhi(day);
            minCatL = mCatL3Xlo(day);
            rangeCatL = maxCatL - minCatL;
            maxMMP2 = MMP23Xhi(day);
            minMMP2 = MMP23Xlo(day);
            rangeMMP2 = maxMMP2 - minMMP2;
            catL = mCatL3Xlo(day) + rangeCatL*rand(1,1);
            MMP2 = MMP20Xlo(day) + rangeMMP2*rand(1,1);

        end

        protease = catL+MMP2;
        minProtease = minCatL + minMMP2;
        maxProtease = maxCatL + maxMMP2;
        rangeProtease = maxProtease - minProtease;
        proportion = (protease - minProtease)/rangeProtease;
        amt = 1-proportion;
end
end

function lifeTable = biobotSurvival(t, d, r)

% Populating the Life Table
% First column in failure time (t), number of failed (d), number at risk
```

```matlab
% (r), hazard rate (h(t)), cumulative hazard rate, survival probability
% (S(t)), cumulative distribution function (CDF)
lifeTable = zeros(length(t),7);

lifeTable(:,1) = t;     % failure time
lifeTable(:,2) = d;     % number of failed
lifeTable(:,3) = r;     % number at risk

% Hazard Rate (h(t))
for i=1:length(t)
    lifeTable(i,4) = d(i)/r(i);
end

% Cumulative Hazard Rate
for i=1:length(t)
    if i == 1
        lifeTable(i,5) = 0;
    else
        lifeTable(i,5) = lifeTable(i-1,5) + lifeTable(i,4);
    end
end

% Survival Probability (S(t))
for i=1:length(t)
    if i == 1
        lifeTable(i,6) = 1 - lifeTable(i,4);
    else
        lifeTable(i,6) = lifeTable(i-1,6)*(1 - lifeTable(i,4));
    end
end

% Cumulative Distribution Function (CDF)
for i=1:length(t)
    lifeTable(i,7) = 1 - lifeTable(i,6);
end

end

function A = laplacian3(nx,ny,nz)
ex=ones(nx,1);
Lx=spdiags([ex -3*ex ex],[-1 0 1],nx,nx);
ey=ones(ny,1);
Ly=spdiags([ey -3*ey ey],[-1 0 1],ny,ny);
```

```matlab
Ix=speye(nx);
Iy=speye(ny);
L2=kron(Iy,Lx)+kron(Ly,Ix);

N=nx*ny*nz;
e=ones(N,1);
L=spdiags([e e],[-nx*ny nx*ny],N,N);
Iz=speye(nz);

A=kron(Iz,L2)+L;
end


function [ g ] = remove_posts( g, length, width, depth)
%UNTITLED Summary of this function goes here
%   Detailed explanation goes here
nv = size(g.Nodes);
[X,Y,Z] = ind2sub([length, width, depth], 1:size(g.Nodes,1));
for i = 1:nv
    % post 1
    if (1/4 * width < Y(i) && Y(i) < 3/4 * width) && (1/8*length <  X(i) && X(i) < 1/4*
length)
        for j = neighbors(g, i)
            g = rmedge(g, i,j);
        end
    end
    % post 2
    if 1/4 * width < Y(i) && Y(i) < 3/4 * width && 3/4*length <  X(i) && X(i) < 7/8*
length
        for j = neighbors(g, i)
            g = rmedge(g, i,j);
        end
    end
end
%fprintf('number of edges remaining %d', size(g.Edges,1));
end


function [  ] = write_graph( graph, fname )
%write_graph writes a graph to a file in edgelist form
%   Detailed explanation goes here
```

```
writetable(graph.Edges, fname);

end
```

C.3   Grid Search MATLAB Code

```matlab
function biobotGridSearch20170626()

% settings

probCleavMin = 0.10;
probCleavMax = 0.40;
probCleavStep = 0.05;

probCleav = probCleavMin:probCleavStep:probCleavMax;

probDeadMin = 0.10;
probDeadMax = 0.40;
probDeadStep = 0.05;

probDead = probDeadMin:probDeadStep:probDeadMax;

parpool(32)

parfor a = 1:length(probCleav)
    for b = 1:length(probDead)

        % Simulated biobots & determine failure times
        % Weibull's distribution parameters are the output
        res0X = biobotModel20170625(probCleav(a), probDead(b), 0, 0, 0);
        res1X = biobotModel20170625(probCleav(a), probDead(b), 1, 0, 0);
        res3X = biobotModel20170625(probCleav(a), probDead(b), 3, 0, 0);
        % Store combo of biobot parameter values and Weibull parameters
        resAll = [probCleav(a), probDead(b), res0X(1), res0X(2), res1X(1), res1X(2),
res3X(1), res3X(2)];

        % to avoid issues with the parfor loop & writing to files
        % concurrently
        t = getCurrentTask();
        tID = t.ID;

        dlmwrite(strcat('BiobotGS_20170626_', num2str(tID), '.txt'), resAll, '-append')
```
218

```
        end
end

    end
```

APPENDIX D. CATHEPSIN MUTANT SEQUENCES

This appendix contains specific sequences information for plasmids, cathepsin genes, as well as primers used to design and generate cathepsin mutants (Chapter 8).

D.1 Plasmid Sequence

D.1.1 pCMV6 Plasmid

The original cathepsin wildtype plasmids were purchased from OriGene and were on the pCMV6 plasmid (Figure D-1).
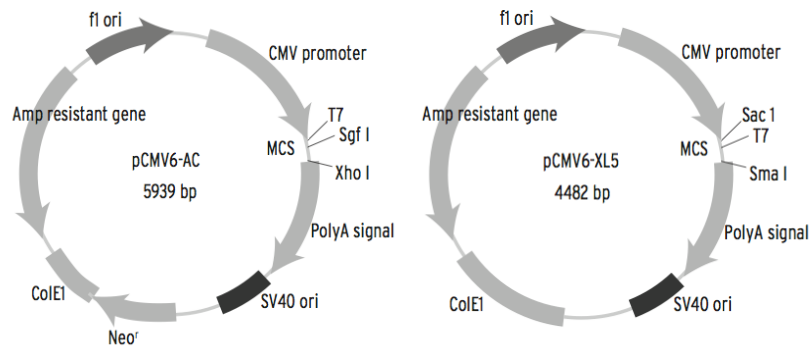
Figure D-1. pCMV6 plasmid map (from OriGene) with the wildtype cathepsin sequences. Cathepsins K, L, and S were on the pCMV6-AC backbone and cathepsin V was on the pCMV-XL5 backbone.

D.1.2 pcDNA 4/TO/Myc-His A Plasmid

The wildtype cathepsin genes were PCR'ed off the pCMV6 plasmids and were mutated or inserted as wildtype cathepsin into Invitrogen's pcDNA 4/TO/Myc-His A vector (Figure D-2).
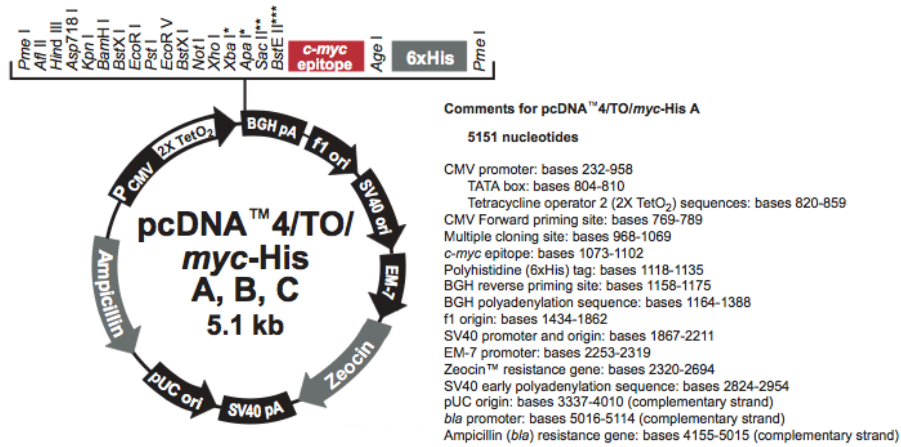
Figure D-2. Invitrogen's pcDNA 4/TO/myc-His plasmid map that the backbone for the mutant cathepsin plasmids.

The multiple cloning sites of version A is give in Figure D-3. Cathepsins K and S were cloned between EcoRI and AgeI and cathepsins L and V were cloned between BamHI and AgeI, which removed the c-myc epitope (which if included would prevent the cathepsins from folding).
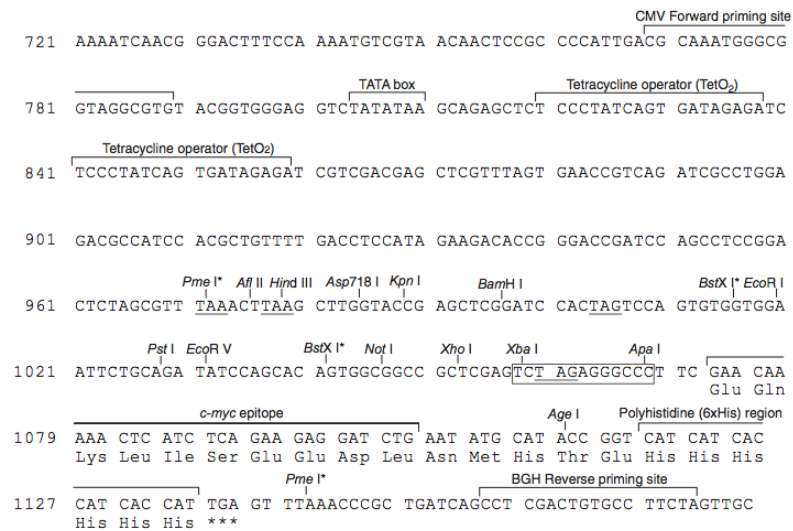
Figure D-3. Multiple cloning sites for pcDNA 4/TO/myc-His plasmid version A.

Key sites:

CMV priming site   c-myc isotope   His tagging   Stop codon

BGH reverse priming site

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTG
CTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGG
TCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGAC
CGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCG
ATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAG
TAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTAC
ATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCA
TTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATG
GCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG
TACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC
ACCCCATTGACGTCAATGGGAGTTTGTTTTGGAACCAAAATCAACGGGACTTT
CCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTG
TACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCT
CCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATC
GCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACC
GATCCAGCCTCCGGACTCTA

GCGTTTAAACTTAAGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTG
GAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCT
TCGAACAAAAACTCATCTCAGAAGAGGATCTGAATATGCATACCGGTCATCA
TCACCATCACCATTGAGTTTAAACC

CGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCC
CTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCA
GCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGC
GGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTA
GCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTT
CCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTT
ACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGG

CCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTT
AATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTA
TTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATG
AGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGT
TAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGC
ATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAG
CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCC
GCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTC
CGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCT
GCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTT
TTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACG
TGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAG
GTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCG
CGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTC
CCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACC
CTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCT
GGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGT
GTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAG
CAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGC
ACTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTCGATTCCAC
CGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCT
GGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAA
CTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATT
TCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTC
ATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGC
GTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC
CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATG
AGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG
GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCT
CGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACG
GTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGG
CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCAT
AGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCT
TTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTC
AGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG
AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTAC
GGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTA
CCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG

TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGA
AAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT
AGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGC
GATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAA
CTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCG
AGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA
AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTA
TTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGC
AACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTAT
GGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCA
TGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGT
AAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCA
AGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCA
ATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATC
CAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA
AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCA
ATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTC

D.1.3 pCDNA6-TETR

This is the sequence of the Tet Repressor plasmid used in combination with the

pCDNA 4/TO/Myc-His A cathepsin plasmids to induce cathepsin expression with

tetracycline.

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTG
CTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGG
TCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGAC
CGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCG
ATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAG
TAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTAC

ATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCA
TTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATG
GCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG
TACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC
ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTT
CCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTG
TACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCAC
TGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCT
GGCTAGCGTTTAAACTTAAGCTTGGTACCCGGGGATCCTCTAGGGCCTCTGAG
CTATTCCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAA
AGCTCCGGATCGATCCTGAGAACTTCAGGGTGAGTTTGGGGACCCTTGATTGT
TCTTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGAGGGGGCAAAGTTTTC
AGGGTGTTGTTTAGAATGGGAAGATGTCCCTTGTATCACCATGGACCCTCATG
ATAATTTTGTTTCTTTCACTTTCTACTCTGTTGACAACCATTGTCTCCTCTTATT
TTCTTTTCATTTTCTGTAACTTTTTCGTTAAACTTTAGCTTGCATTTGTAACGA
ATTTTTAAATTCACTTTTGTTTATTTGTCAGATTGTAAGTACTTTCTCTAATCA
CTTTTTTTTCAAGGCAATCAGGGTATATTATATTGTACTTCAGCACAGTTTTAG
AGAACAATTGTTATAATTAAATGATAAGGTAGAATATTTCTGCATATAAATTC
TGGCTGGCGTGGAAATATTCTTATTGGTAGAAACAACTACATCCTGGTCATCA
TCCTGCCTTTCTCTTTATGGTTACAATGATATACACTGTTTGAGATGAGGATA
AAATACTCTGAGTCCAAACCGGGCCCCTCTGCTAACCATGTTCATGCCTTCTT
CTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTT
TGGCAAAGAATTGTAATACGACTCACTATAGGGCGAATTGATATGTCTAGAT
TAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGG
AATCGAAGGTTTAACAACCCGTAAACTCGCCCAGAAGCTAGGTGTAGAGCAG
CCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCTCGACGCCTTAGC
CATTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAA
AGCTGGCAAGATTTTTTACGTAATAACGCTAAAAGTTTTAGATGTGCTTTACT
AAGTCATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAA
CAGTATGAAACTCTCGAAAATCAATTAGCCTTTTTATGCCAACAAGGTTTTTC
ACTAGAGAATGCATTATATGCACTCAGCGCTGTGGGGCATTTTACTTTAGGTT
GCGTATTGGAAGATCAAGAGCATCAAGTCGCTAAAGAAGAAAGGGAAACAC
CTACTACTGATAGTATGCCGCCATTATTACGACAAGCTATCGAATTATTTGAT
CACCAAGGTGCAGAGCCAGCCTTCTTATTCGGCCTTGAATTGATCATATGCGG
ATTAGAAAACAACTTAAATGTGAAAGTGGGTCCGCGTACAGCGGATCCCGG
GAATTCAGATCTTATTAAAGCAGAACTTGTTTATTGCAGCTTATAATGGTTAC
AAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGC
ATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGTCGA
ATCCATCACACTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCGAAGGTA
AGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCAT

225

CACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAG
GTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTA
TGGCTTCTGAGGCGGAAAGAAC

D.2   Wildtype Cathepsin Sequences

D.2.1   Cathepsin K

EcoRI   Start/Stop   Primer binding   Cathepsin K Sequence

Starting with where the forward binding primer site:

AGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGC
GGCCGGGAATTCGGCACGAGGGATCACTGGAGCTGACTTCCGCAATCCCGAT
GGAATAAATCTAGCACCCCTGATGGTGTGCCCACACTTTGCTGCCGAAACGA
AGCCAGACAACAGATTTCCATCAGCAGG

ATGTGGGGGCTCAAGGTTCTGCTGCTACCTGTGGTGAGCTTTGCTCTGTACCC
TGAGGAGATACTGGACACCCACTGGGAGCTATGGAAGAAGACCCACAGGAA
GCAATATAACAACAAGGTGGATGAAATCTCTCGGCGTTTAATTTGGGAAAAA
AACCTGAAGTATATTTCCATCCATAACCTTGAGGCTTCTCTTGGTGTCCATAC
ATATGAACTGGCTATGAACCACCTGGGGGACATGACCAGTGAAGAGGTGGTT
CAGAAGATGACTGGACTCAAAGTACCCCTGTCTCATTCCCGCAGTAATGACA
CCCTTTATATCCCAGAATGGGAAGGTAGAGCCCCAGACTCTGTCGACTATCG
AAAGAAAGGATATGTTACTCCTGTCAAAAATCAGGGTCAGTGTGGTTCCTGT
TGGGCTTTTAGCTCTGTGGGTGCCCTGGAGGGCCAACTCAAGAAGAAACTG
GCAAACTCTTAAATCTGAGTCCCCAGAACCTAGTGGATTGTGTGTCTGAGAAT
GATGGCTGTGGAGGGGGCTACATGACCAATGCCTTCCAATATGTGCAGAAGA
ACCGGGGTATTGACTCTGAAGATGCCTACCCATATGTGGGACAGGAAGAGAG
TTGTATGTACAACCCAACAGGCAAGGCAGCTAAATGCAGAGGGTACAGAGA
GATCCCCGAGGGGAATGAGAAGCCCTGAAGAGGGCAGTGGCCCGAGTGGG
ACCTGTCTCTGTGGCCATTGATGCAAGCCTGACCTCCTTCCAGTTTTACAGCA
AAGGTGTGTATTATGATGAAAGCTGCAATAGCGATAATCTGAACCATGCGGT
TTTGGCAGTGGGATATGGAATCCAGAAGGGAAACAAGCACTGGATAATTAAA
AACAGCTGGGGAGAAAACTGGGGAAACAAAGGATATATCCTCATGGCTCGA

226

AATAAGAACAACGCCTGTGGCATTGCCAACCTG<mark style="background:lime">GCCAGCTTCCCCAAGATG</mark><mark style="background:magenta">T</mark><mark style="background:magenta">GA</mark>

CTCCAGCCAGCCAAATCCATCCTGCTCTTCCATTTCTTCCACGATGGTGCAGT
GTAACGATGCACTTTGGAAGGGAGTTGGTGTGCTATTTTTGAAGCAGATGTG
GTGATACTGAGATTGTCTGTTCAGTTTCCCCATTTGTTTGTGCTTCAAATGATC
CTTCCTACTTTGCTTCTCTCCACCCATGACCTTTTTCACTGTGGCCATCAGGAC
TTTCCCTGACAGCTGTGTACTCTTAGGCTAAGAGATGTGACTACAGCCTGCCC
CTGACTGTGTTGTCCCAGGGCTGATGCTGTACAGGTACAGGCTGGAGATTTTC
ACATAGGTTAGATTCTCATTCACGGGACTAGTTAGCTTTAAGCACCCTAGAGG
ACTAGGGTAATCTGACTTCTCACTTCCTAAGTTCCCTTCTATATCCTCAAGGT
AGAAATGTCTATGTTTTCTACTCCAATTCATAAATCTATTCATAAGTCTTTGGT
ACAAGTTTACATGATAAAAGAAATGTGATTTGTCTTCCCTTCTTTGCACTTT
TGAAATAAAGTATTTATCTCCTGTCTACAGTTTAATAAATAAAAAAAAAAAA
AAAAAAAAAAAAAA

D.2.2   *Cathepsin L*

<mark style="background:magenta">Start/Stop</mark>   <mark style="background:lime">Primer binding</mark>   Cathepsin L Sequence

The PCR subcloning primer will insert the BamHI site before the cathepsin L gene. Starting

with where the forward binding primer site:

<mark style="background:lime">AGAGCTCGTTTAGTGAACCGTCAG</mark>AATTTTGTAATACGACTCACTATAGGGC
GGCCGGGAATTCGGCACGAGGGGGGGGCCGGACAGGGACTGGAAGAGAGGA
CGCGGTCGAGTAGGTGTGCACCAGCCCTGGCAACGAGAGCGTCTACCCCGAA
CTCTGCTGGCCTTGAGGTGGGGAAGCCGGGGAGGGCAGTTGAGGACCCCGCG
GAGGCGCGTGACTGGTTGAGCGGGCAGGCCAGCCTCCGAGCCGGGTGGACA
CAGGTTTTAAAAC

<mark style="background:magenta">ATG</mark>AATCCTACACTCATCCTTGCTGCCTTTTGCCTGGGAATTGCCTCAGCTAC
TCTAACATTTGATCACAGTTTAGAGGCACAGTGGACCAAGTGGAAGGCGATG
CACAACAGATTATACGGCATGAATGAAGAAGGATGGAGGAGAGCAGTGTGG
GAGAAGAACGTGAAGATGATTGAACTGCACAATCAGGAATACAGGGAAGGG
AAACACAGCTTCACAATGGCCATGAACGCCTTTGGAGACATGACCAGTGAAG
AATTCAGGCAGGTGATGAATGGCTTTCAAAACCGTAAGCCCAGGAAGGGGA
AAGTGTTCCAGGAACCTCTGTTTTATGAGGCCCCCAGATCTGTGGATTGGAGA

227

GAGAAAGGCTACGTGACTCCTGTGAAGAATCAGGGTCAGTGTGGTTCTTGTT
GGGCTTTTAGTGCTACTGGTGCTCTTGAAGGACAGATGTTCCGGAAAACTGG
GAGGCTTATCTCACTGAGTGAGCAGAATCTGGTAGACTGCTCTGGGCCTCAA
GGCAATGAAGGCTGCAATGGTGGCCTAATGGATTATGCTTTCCAGTATGTTCA
GGATAATGGAGGCCTGGACTCTGAGGAATCCTATCCATATGAGGCAACAGAA
GAATCCTGTAAGTACAATCCCAAGTATTCTGTTGCTAATGACACCGGCTTTGT
GGACATCCCTAAGCAGGAGAAGGCCCTGATGAAGGCAGTTGCAACTGTGGG
GCCCATTTCTGTTGCTATTGATGCAGGTCATGAGTCCTTCCTGTTCTATAAAG
AAGGCATTTATTTTGAGCCAGACTGTAGCAGTGAAGACATGGATCATGGTGT
GCTGGTGGTTGGCTACGGATTTGAAAGCACAGAATCAGATAACAATAAATAT
TGGCTGGTGAAGAACAGCTGGGGTGAAGAATGGGGCATGGGTGGCTACGTA
AAGATGGCCAAAGACCGGAGAAACCATTGTGGAATTGCCTC<mark style="background:#00ff00">AGCAGCCAGCT</mark>
<mark style="background:#00ff00">ACCCCACTGTG</mark><mark style="background:#ff00ff">TGA</mark>

GCTGGTGGACGGTGATGAGGAAGGACTTGACTGGGGATGGCGCATGCATGG
GAGGAATTCATCTTCAGTCTACCAGCCCCCGCTGTGTCGGATACACACTCGA
ATCATTGAAGATCCGAGTGTGATTTGAATTCTGTGATATTTTCACACTGGTAA
ATGTTACCTCTATTTTAATTACTGCTATAAATAGGTTTATATTATTGATTCACT
TACTGACTTTGCATTTTCGTTTTTAAAAGGATGTATAAATTTTTACCTGTTTAA
ATAAAATTTAATTTCAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AA

D.2.3   Cathepsin S

<mark style="background:#ff0000">EcoRI</mark>   <mark style="background:#ff00ff">Start/Stop</mark>   <mark style="background:#00ff00">Primer binding</mark>   Cathepsin S Sequence

Starting with where the forward binding primer site:

<mark style="background:#00ff00">AGAGCTCGTTTAGTGAACCGTCAG</mark>AATTTTGTAATACGACTCACTATAGGGC
GGCCGG<mark style="background:#ff0000">GAATTC</mark>GGCACGAGGTTAGAAGAGAGCCCACTAATTCAAGGACTCT
TACCGTGGGAGCAACTGCTGGTTCTATCACA

<mark style="background:#ff00ff">ATG</mark>AAACGGCTGGTTTGTGTGCTCTTGGTGTGCTCCTCTGCAGTGGCACAGTT
GCATAAAGATCCTACCCTGGATCACCACTGGCATCTCTGGAAGAAAACCTAT
GGCAAACAATACAAGGAAAAGAATGAAGAAGCAGTACGACGTCTCATCTGG
GAAAAGAATCTAAAGTTTGTGATGCTTCACAACCTGGAGCATTCAATGGGAA
TGCACTCATACGATCTGGGCATGAACCACCTGGAGACATGACCAGTGAAGA
AGTGATGTCTTTGATGAGTTCCCTGAGAGTTCCCAGCCAGTGGCAGAGAAAT

228

ATCACATATAAGTCAAACCCTAATTGGATATTGCCTGATTCTGTGGACTGGAG
AGAGAAAGGGTGTGTTACTGAAGTGAAATATCAAGGTTCTTGTGGTGCTTGC
TGGGCTTTCAGTGCTGTGGGGGCCCTGGAAGCACAGCTGAAGCTGAAAACAG
GAAAGCTGGTGTCTCTCAGTGCCCAGAACCTGGTGGATTGCTCAACTGAAAA
ATATGGAAACAAAGGCTGCAATGGTGGCTTCATGACAACGGCTTTCCAGTAC
ATCATTGATAACAAGGGCATCGACTCAGACGCTTCCTATCCCTACAAAGCCA
TGGATCAGAAATGTCAATATGACTCAAAATATCGTGCTGCCACATGTTCAAA
GTACACTGAACTTCCTTATGGCAGAGAAGATGTCCTGAAAGAAGCTGTGGCC
AATAAAGGCCCAGTGTCTGTTGGTGTAGATGCGCGTCATCCTTCTTTCTTCCT
CTACAGAAGTGGTGTCTACTATGAACCATCCTGTACTCAGAATGTGAATCATG
GTGTACTTGTGGTTGGCTATGGTGATCTTAATGGGAAAGAATACTGGCTTGTG
AAAAACAGCTGGGGCCACAACTTTGGTGAAGAAGGATATATTCGGATGGCAA
GAAATAAAGGAAATCATTGTGGGATTGCTAGCTTT<mark style="background-color:#00FF00">CCCTCTTACCCAGAAATC</mark><mark style="background-color:#FF00FF">TAG</mark>

AGGATCTCTCCTTTTTATAACAAATCAAGAAATATGAAGCACTTTCTCTTAAC
TTAATTTTTCCTGCTGTATCCAGAAGAAATAATTGTGTCATGATTAATGTGTA
TTTACTGTACTAATTAGAAAATATAGTTTGAGGCCGGGCACGGTGGCTCACG
CCTGTAATCCCAGTACTTGGGAGGCCAAGGCAGGCATATCAACTTGAGGCCA
GGAGTTAAAGAGCAGCCTGGCTAACATGGTGAAACCCCATCTCTACTAAAAA
TACAAAAAATTAGCCGAGCACGGTGGTGCATGCCTGTAATCCCAGCTACTTG
GGAGGCTGAGGCACGAGATTCCTTGAACCCAAGAGGTTGAGGCTATGTTGAG
CTGAGATCACACCACTGTACTCCAGCCTGGATGACAGAGTGGAGACTCTGTT
TCAAAAAAACAGAAAGAAAATATAGTTTGATTCTTCATTTTTTAAATTTGC
AAATCTCAGGATAAAGTTTGCTAAGTAAATTAGTAATGTACTATAGATATAA
CTGTACAAAAATTGTTCAACCTAAAACAATCTGTAATTGCTTATTGTTTTATT
GTATACTCTTTGTCTTTTAAGACCCCTAATAGCCTTTTGTAACTTGATGGCTTA
AAAATACTTAATAAATCTGCCATTTCAAATTTCAAAAAAAAAAAAAAAAAAAA
AAA

D.2.4   Cathepsin V

<mark style="background-color:#FF0000">GC/CCGGCG NOTI</mark>   <mark style="background-color:#FF00FF">Start/Stop</mark>   <mark style="background-color:#00FF00">Primer binding</mark>   Cathepsin V Sequence

Starting with where the forward binding primer site:

<mark style="background-color:#00FF00">AGCAGAGCTCGTTTAGTGAACCGT</mark>CAGAATTTTGTAATACGACTCACTATAG
G<mark style="background-color:#FF0000">GCGGCCGC</mark>GCTCTAAGCGCCCGGGGCCCCGCCCAGTGGCCGGCACAGCCAA

229

TCGCAGCGCGGGAAGGCGGTGGGGGCGGGGAAGGCCGCCTGGAAACTTAAA
TCCCGAGGCGGGCGAACCTGCACCAGACCGCGGACGTCTGTAATCTCAGAGG
CTTGTTTGCTGAGGGTGCCTGCGCAGCTGCGACGGCTGCTGGTTTTGAAAC

ATGAATCTTTCGCTCGTCCTGGCTGCCTTTTGCTTGGGAATAGCCTCCGCTGTT
CCAAAATTTGACCAAAATTTGGATACAAAGTGGTACCAGTGGAAGGCAACAC
ACAGAAGATTATATGGCGCGAATGAAGAAGGATGGAGGAGAGCAGTGTGGG
AAAAGAATATGAAAATGATTGAACTGCACAATGGGGAATACAGCCAAGGGA
AACATGGCTTCACAATGGCCATGAATGCTTTTGGTGACATGACCAATGAAGA
ATTCAGGCAGATGATGGGTTGCTTTCGAAACCAGAAATTCAGGAAGGGGAAA
GTGTTCCGTGAGCCTCTGTTTCTTGATCTTCCCAAATCTGTGGATTGGAGAAA
GAAAGGCTACGTGACGCCAGTGAAGAATCAGAAACAGTGTGGTTCTTGTTGG
GCTTTTAGTGCGACTGGTGCTCTTGAAGGACAGATGTTCCGGAAAACTGGGA
AACTTGTCTCACTGAGCGAGCAGAATCTGGTGGACTGTTCGCGTCCTCAAGG
CAATCAGGGCTGCAATGGTGGCTTCATGGCTAGGGCCTTCCAGTATGTCAAG
GAGAACGGAGGCCTGGACTCTGAGGAATCCTATCCATATGTAGCAGTGGATG
AAATCTGTAAGTACAGACCTGAGAATTCTGTTGCTAATGACACTGGCTTCACA
GTGGTCGCACCTGGAAAGGAGAAGGCCCTGATGAAAGCAGTCGCAACTGTG
GGGCCCATCTCCGTTGCTATGGATGCAGGCCATTCGTCCTTCCAGTTCTACAA
ATCAGGCATTTATTTTGAACCAGACTGCAGCAGCAAAAACCTGGATCATGGT
GTTCTGGTGGTTGGCTACGGCTTTGAAGGAGCAAATTCGAATAACAGCAAGT
ATTGGCTCGTCAAAAACAGCTGGGGTCCAGAATGGGGCTCGAATGGCTATGT
AAAAATAGCCAAAGACAAGAACAACCACTGTGGAATCGCCACAGCAGCCAG
CTACCCCAATGTGTGA

GCTGATGGATGGTGAGGAGGAAGGACTTAAGGACAGCATGTCTGGGGAAAT
TTTATCTTGAAACTGACCAAACGCTTATTGTGTAAGATAAACCAGTTGAATCA
TTGAGGATCCAAGTTGAGATTTTAATTCTGTGACATTTTTACAAGGGTAAAAT
GTTACCACTACTTTAATTATTGTTATACACAGCTTTATGATATCAAAGACTCA
TTGCTTAATTCTAAGACTTTTGAATTTTCATTTTTTAAAAAGATGTACAAAAC
AGTTTGAAATAAATTTTAATTCGTATATAAA

GCGGCCGC

D.3 Mutant Cathepsin Primers

D.3.1 Wildtype cathepsin subcloning primers

- Forward binding primer for cathepsins K, S:
  - 5'-AGAGCTCGTTTAGTGAACCGTCAG-3'
  - This primer binds upstream of an EcoRI restriction site that will be used to clone the gene into pcDNA4-TO-A
- Forward binding primer for cathepsin L1:
  - 5'-GATAGGATCCAGAGCTCGTTTAGTGAACCGTCAG-3'
  - Primer generates a BamHI site upstream of the gene
- Forward binding primer for cathepsin V (L2):
  - 5'-ATATGGATCCAGCAGAGCTCGTTTAGTGAACCGT-3'
  - Primer generates a BamHI site upstream of the gene
- Cathepsin K AgeI reverse primer:
  - 5'-GCTGACCGGTCATCTTGGGGAAGCTGGC-3'
- Cathepsin S AgeI reverse primer:
  - 5'-TCTAACCGGTGATTTCTGGGTAAGAGGG-3'
- Cathepsin L AgeI reverse primer:
  - 5'-GTGGACCGGTCACAGTGGGGTAGCTGGCTGCT-3'
- Cathepsin V (L2) AgeI reverse primer:
  - 5'-GTAGACCGGTCACATTGGGGTAGCTGGC-3'

D.3.2 Cathepsin mutagenic primers

D.3.2.1 Cathepsin S Active Site Mutations

Mutation	Primers	Restriction Enzyme	Characteristics
C139A	Fwd: 5'-tct tgt ggg gcg gcc tgg gct ttc agt gc-3'	NarI site added	65.5%; 70.9°C
	Rev: 5'-agc cca ggc cgc ccc aca aga acc ttg a-3'		66.7%; 70.2°C
H278Q	Fwd: 5'-aat gtg aac cag ggt gta ctt gtg gtt ggc-3'	HinfI site removed	50.0%; 63.9°C
	Rev: 5'-c acc ctg gtt cac att ctg agt aca gga gt-3'		50.0%; 61.1°C
N298D	Fwd: 5'-tgg ctt gtg aag gat agc tgg ggc cac aac-3'	PvuII site removed	56.7%; 66.8°C
	Rev: 5'-gct atc ctt caca ag cca gta ttc ttt ccc-3'		46.7%; 60.5°C

D.3.2.2 Cathepsin S Cleavage Site Mutations

Cathepsins L/V on Cathepsin S

Mutation	Primers	Restriction Enzyme	Characteristics
L159V	Fwd: 5'-aaa acc ggg aag gtg gtg tct ctc agt gcc-3'	NciI site added	56.7%; 66.5°C
	Rev: 5'-ga cac cac ctt ccc ggt ttt cag ctt cag c-3'		56.7%; 65.8°C
L288V	Fwd: 5'-ggt gac gtt aac ggg aaa gaa tac tgg c-3'	HpaI site removed	50.0%; 61.0°C
	Rev: 5'-ttt ccc gtt aac gtc acc ata gcc aac cac-3'		50.0%; 63.6°C
S324A	Fwd: 5'-ggg att gct gcc ttt ccc tct tac cca g-3'	NheI site removed	57.1%; 64.2°C
	Rev: 5'-ggg aaa ggc agc aatccc aca atg aat tcc-3'		46.7%; 61.5°C

- Here I assumed that catL & catV have the same specificity matrix
- Didn't make any cathepsin K on cathepsin S mutants

D.3.2.3 Cathepsin L Active Site Mutations

Mutation	Primers	Restriction Enzyme	Characteristics
C138A	Fwd: 5'-tgt ggt agc gct tgg gct ttt agt gct ac-3'	AfeI site added	51.7%; 63.8°C
	Rev: 5'-agc cca agc gct acc aca ctg acc aca ctg-3'		60.0%; 68.1°C
H276Q	Fwd: 5'-atg gac cag ggt gtc ctg gtg gtt ggc tac -3'	DrdI site added	60.0%; 67.7°C
	Rev: 5'-c cag gac acc ctg gtc cat gtc ttc act cg-3'		60.0%; 66.6°C
N300D	Fwd: 5'-gtg aag gat agc tgg ggt gaa gaa tgg ggc-3'	PvuII site removed	56.7%; 65.2°C
	Rev: 5'-c acc cca gct atc ctt cac cag cca ata-3'		56.0%; 64.0°C

232

D.3.2.4 Cathepsin L Cleavage Site Mutations

Cathepsins K on Cathepsin L

Mutation	Primers	Restriction Enzyme	Characteristics
E176N	Fwd: 5'-ggc aat aac ggt tgg aat ggt ggc cta atg g-3'	AclI site added	51.6%; 64.2°C
	Rev: 5'-att cca acc gtt att gcc ttg agg gcc aga-3'		50.0%; 64.8°C
E289N	Fwd: 5'-~~gaa~~ agt act aac tcc gat aac aat aaa tat tgg-3'	ScaI site added	30.3%; 54.4°C
	Rev: 5'-atc gga gtt agt act ttc aaa tcc gta gcc-3'		43.3%; 59.5°C

E289N – doesn't really have good primers, very low GC content, partially because of location in the gene; also the forward primer is too many nucleotides (30nt – without gaa – sequence, 30%, 53.2deg)

Cathepsins S on Cathepsin L

Mutation	Primers	Restriction Enzyme	Characteristics
S326A	Fwd: 5'-att gcc gca gcg gcc agc tac ccc act gtg-3'	BglI site added	66.7%; 71.8°C
	Rev: 5'-gct ggc cgc tgc ggc aat tcc aca atg g-3'		64.3%; 69.3°C
L158V	Fwd: 5'-ggc agg gtg atc agc ctg agt gag cag aat-3'	BclI site added	56.7%; 66.3°C
	Rev: 5'-ct cag gct gat cac cct gcc agt ttt ccg g-3'		60.0%; 66.9°C
~~**V239A**~~	~~Fwd: 5'-atg aag gcg gcc gca act gtg ggg ccc-3'~~	~~NotI site added~~	~~70.4%; 72.3°C~~
	~~Rev: 5'-agt tgc ggc cgc ctt cat cag ggc ttc ctc-3'~~		~~63.3%; 69.9°C~~
V239G	Fwd: 5'-atg aag gcc ggg gca act gtg ggg ccc-3'	NciI site added	70.4%; 72.1°C
	Rev: 5'-c agt tgc ccc ggc ctt cat cag ggc c-3'		69.2%; 69.3°C

Cathepsin V on Cathepsin L: Sites have not yet been identified since catV does not have a defined specificity matrix.

D.3.2.5 Cathepsin V Active Site Mutants

Mutation	Primers	Restriction Enzyme	Characteristics
C138A	Fwd: 5'-gt ggt agc gct tgg gct ttt agt gcg act gg-3'	AfeI site added	58.1%; 66.9°C
	Rev: 5'-agc cca agc gct acc aca ctg ttt ctg-3'		55.6%; 64.6°C
H277Q	Fwd: 5'-aac ctt gat caa ggt gtt ctg gtt ggc tac-3'	BclI site added	46.7%; 61.9°C
	Rev: 5'-aac acc ttg atc aag gtt ttt gct gct gc-3'		44.8%; 61.5°C
N301D	Fwd: 5'-gtc aag gat agc tgg ggt cca gaa tgg ggc-3'	PvuII site removed	60.0%; 66.6°C
	Rev: 5'-c cca gct atc ctt gac gag cca ata ctt gc-3'		53.3%; 63.2°C

D.3.2.6 Cathepsin V Cleavage Site Mutations

Cathepsins K on Cathepsin V

Mutation	Primers	Restriction Enzyme	Characteristics
E287N	Fwd: 5'-ggc ttc aac ggc gca aat tcg aat aac agc-3'	BstBI site removed	50.0%; 63.7°C
	Rev: 5'-att tgc gcc gtt gaa gcc gtc gcc aac cac-3'		60.0%; 69.4°C
L114V	Fwd: 5'-t ctt gac gtc ccc aaa tct gtg gat tgg ag-3'	ZraI site added	50.0%; 63.1°C
	Rev: 5'-tt ggg gac gtc aag aaa cag agg ctc acg g-3'		56.7%; 66.3°C
E306N	Fwd: 5'-tgg gtt ccc aat tgg ggc tcg aat ggc-3'	MfeI site added	63.0%; 68.3°C
	Rev: 5'-cc cca att ggg acc cca gct gtt ttt gac g-3'		56.7%; 66.2°C

Cathepsins S on Cathepsin V

Mutation	Primers	Restriction Enzyme	Characteristics
L158V	*Also location for catL on catV, see table below for primers*		
L280V	Fwd: 5'-gat cac ggc gtg gtg gtg gtt ggc tac ggc-3'	BsiEI site added	66.7%; 70.3°C
	Rev: 5'-c cac cac gcc gtg atc cag gtt ttt gct gc-3'		60.0%; 68.0°C
G288A	*Also location for catL on catV, see table below for primers*		
~~V219A~~	~~Fwd: 5'-cct gag aat agc gct gct aat gac act ggc-3'~~	~~AfeI site added~~	~~53.3%; 63.7°C~~
	~~Rev: 5'-agc agc gct att ctc agg tct gta c-3'~~		~~52.0%; 60.1°C~~
V219G	Fwd: 5'-cct gag aat agc ggg gct aat gac act ggc-3'	EcoRI site removed	56.7%; 65.0°C
	Rev: 5'-agc ccc gct att cac agg tct gta c-3'		56.0%; 62.2°C
L236V	Fwd: 5'-g gag aag gcc gtc atg aaa gca gtc gca ac-3'	BspHI site added	56.7%; 65.8°C
	Rev: 5'-cat gac ggc ctt ctc ctt tcc agg cgt gac-3'		60.0%; 66.9°C

Cathepsins L on Cathepsin V

Mutation	Primers	Restriction Enzyme	Characteristics
L158V	Fwd: 5'-aaa acc ggg aag gtg gtc tca ctg agc gag-3'	NciI site added	56.7%; 66.1°C
	Rev: 5'-gac cac ctt ccc ggt ttt ccg gaa cat ctg-3'		56.7%; 65.4°C
G288A	Fwd: 5'-ggc ttt gag gcg gcc aat tcg aat aac agc-3'	ApoI site removed	53.3%; 64.9°C
	Rev: 5'-cga att ggc cgc ctc aaa gcc gta gcc aac-3'		60.0%; 67.8°C

D.3.2.7 Cathepsin K Cleavage Site Mutants

(Designed by and created in Ellen Moomaw's Lab Class)

Mutation	Primers	Restriction Enzyme	Characteristics
C139A	Fwd: 5' – tgt ggt agc gct tgg gct ttt agc tct gtg – 3'	AfeI site	53.3%; 65.2°C
	Rev: 5' – agc cca agc gct acc aca ctg acc ctg – 3'		63.0%; 67.7°C
C139S	Fwd: 5' – ggt tcc agc tgg gct ttt agc tct gtg – 3'	PvuII site added	55.6%; 63.2°C
	Rev: 5' – agc cca gct gga acc aca ctg acc ctg – 3'		63.0%; 67.7°C
H276A	Fwd: 5' – ctg aac gct gca gtt ttg gca gtg gga – 3'	PstI site	55.6%; 65.1°C
	Rev: 5' – c tgc caa aac tgc agc gtt cag att atc gct – 3'		48.4%; 63.9°C
H276N	Fwd: 5' – ctg aac cag gca gta ctg gca gtg gga – 3'	ScaI site added	59.3%; 65.5°C
	Rev: 5' – gc cag tac tgc gtt gtt cag att atc gc – 3'		50.0%; 61.3°C
H276Q	Fwd: 5' – ctg aac cag gca gta ctg gca gtg gga – 3'	ScaI site added	59.3%; 65.5°C
	Rev: 5' – gc cag tac tgc ctg gtt cag att atc gc – 3'		53.6%; 62.5°C
N296A	Fwd: 5' – atc aag gcg agc tgg gga gaa aac tgg gga – 3'	PvuII site removed	56.7%; 67.5°C
	Rev: 5' – cca gct cgc ctt gat tat cca gtg ctt g – 3'		53.6%; 62.9°C
N296D	Fwd: 5' – aag gat agc tgg gga gaa aac tgg gga – 3'	PvuII site removed	51.9%; 62.9°C
	Rev: 5' – cca gct atc ctt gat tat cca gtg ctt g – 3'		46.4%; 59.0°C

D.3.2.8 <u>Cathepsin K Cleavage Site Mutants</u>

Cathepsins S on Cathepsin K (Designed by and created in Ellen Moomaw's Lab Class)

Mutation	Primers	Restriction Enzyme	Characteristics
L62A	Fwd: 5' – cat aac gct gag gct agc ctt ggt gtc cat ac – 3'	NheI site	53.1%; 64.3°C
	Rev: 5' – acc aag gct agc ctc agc gtt atg gat gg – 3'		55.2%; 64.9°C
L62V	Fwd: 5' – cat aac gtg gag gct agc cct ggt gtc cat ac – 3'	NheI site	56.2%; 65.5°C
	Rev: 5' – acc aag gct agc ctc gtg gtt atg gat gg – 3'		55.2%; 64.8°C
L97A	Fwd: 5' – gta ccc gcg tct cat tcc cgc agt – 3'	BsmBI site	62.5%; 65.0°C
	Rev: 5' – gga at gaga cgc ggg tac ttt gag tcc – 3'		55.6%; 62.4°C
L97V	Fwd: 5' – ctc aag gta ccg gtg tct cat tcc cgc agt – 3'	AgeI site added	56.7%; 65.8°C
	Rev: 5' – gga at gaga cac cgg tac ctt gag tcc agt c – 3'		54.8%; 64.0°C
V171A	Fwd: 5' – cta gtg gat tgc gca tct gag aat gat – 3'	FspI site	44.4%; 58.7°C
	Rev: 5' – ctc aga tgc gca atc cac tag gtt ctg ggg – 3'		56.7%; 64.9°C
L253A	Fwd: 5' – gcc att gat gca agc gct acc tcc ttc cag – 3'	AfeI site	56.7%; 65.7°C
	Rev: 5' – gga ggt agc gct tgc atc aat ggc cac aga – 3'		56.7%; 66.8°C
L253V	Fwd: 5' – gat gct agc gtg acct cc ttc cag ttt tac agc – 3'	NheI site	51.5%; 64.3°C
	Rev: 5' – gga ggt cac gct agc atc aat ggc cac – 3'		59.3%; 64.8°C

Mutation	Primers	Restriction Enzyme	Characteristics
L159V	*To be designed*		
L274V	Fwd: 5'-gc gat aac gtt aac cat gcg gtt ttg gca g-3'	AclI site added	50.0%; 63.4°C
	Rev: 5'-c cgc atg gtt aac gtt atc gct aat gca gc-3'		50.0%; 62.9°C
G284A	Fwd: 5'-ggc taa gca atc cag aag gga aac aag cac-3'	BlpI site added	50.0%; 62.9°C
	Rev: 5'-ctg gat tgc tta gcc cac tgc caa aac cgc-3'		56.7%; 66.4°C

Here we assumed that catL & catV have the same specificity matrix

All characterization (GC content percentage and Melting temperature) were determined

using the Integrated DNA Technologies, OligoAnalyzer 3.1

(https://www.idtdna.com/calc/analyzer).

REFERENCES

1.	Turk, V., et al., *Cysteine cathepsins: from structure, function and regulation to new frontiers.* Biochim Biophys Acta, 2012. **1824**(1): p. 68-88.

2.	Turk, V., B. Turk, and D. Turk, *Lysosomal cysteine proteases: facts and opportunities.* EMBO J, 2001. **20**(17): p. 4629-33.

3.	Chapman, H.A., R.J. Riese, and G.P. Shi, *Emerging roles for cysteine proteases in human biology.* Annu Rev Physiol, 1997. **59**: p. 63-88.

4.	Garnero, P., et al., *The collagenolytic activity of cathepsin K is unique among mammalian proteinases.* J Biol Chem, 1998. **273**(48): p. 32347-52.

5.	Bromme, D., et al., *Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization.* Biochemistry, 1999. **38**(8): p. 2377-85.

6.	Aguda, A.H., et al., *Structural basis of collagen fiber degradation by cathepsin K.* Proc Natl Acad Sci U S A, 2014. **111**(49): p. 17474-9.

7.	Barry, Z.T. and M.O. Platt, *Cathepsin S cannibalism of cathepsin K as a mechanism to reduce type I collagen degradation.* J Biol Chem, 2012. **287**(33): p. 27723-30.

8.	Brömme, D., F.S. Nallaseth, and B. Turk, *Production and activation of recombinant papain-like cysteine proteases.* Methods, 2004. **32**(2): p. 199-206.

9.	McQueney, M.S., et al., *Autocatalytic activation of human cathepsin K.* J Biol Chem, 1997. **272**(21): p. 13955-60.

10.	Menard, R., et al., *Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in the processing of procathepsin L in vitro.* J Biol Chem, 1998. **273**(8): p. 4478-84.

11.	Nissler, K., et al., *The half-life of human procathepsin S.* Eur J Biochem, 1999. **263**(3): p. 717-25.

12.	Abrahamson, M., *Cystatins*, in *Methods in Enzymology*. 1994, Academic Press, Inc. p. 685-700.

13.   Turk, B., D. Turk, and V. Turk, *Protease signalling: the cutting edge.* EMBO J, 2012. **31**(7): p. 1630-43.

14.   Pislar, A., M. Perisic Nanut, and J. Kos, *Lysosomal cysteine peptidases - Molecules signaling tumor cell death and survival.* Semin Cancer Biol, 2015. **35**: p. 168-79.

15.   Ciechanover, A., *Intracellular protein degradation: from a vague idea through the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting.* Bioorg Med Chem, 2013. **21**(12): p. 3400-10.

16.   Ciechanover, A., *The ubiquitin-proteasome proteolytic pathway.* Cell, 1994. **79**(1): p. 13-21.

17.   Glickman, M.H. and A. Ciechanover, *The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction.* Physiol Rev, 2002. **82**(2): p. 373-428.

18.   Repnik, U., et al., *Lysosomes and lysosomal cathepsins in cell death.* Biochim Biophys Acta, 2012. **1824**(1): p. 22-33.

19.   Turk, B. and V. Turk, *Lysosomes as "suicide bags" in cell death: myth or reality?* J Biol Chem, 2009. **284**(33): p. 21783-7.

20.   Barrett, A.J., N.D. Rawlings, and J.F. Woessner, *Handbook of Proteolytic Enzymes.* 3 ed. Vol. 1. 2012, Amsterdam, GB: Academic Press. 4097.

21.   McGrath, M.E., *The lysosomal cysteine proteases.* Annu Rev Biophys Biomol Struct, 1999. **28**: p. 181-204.

22.   Li, Z., et al., *Regulation of collagenase activities of human cathepsins by glycosaminoglycans.* J Biol Chem, 2004. **279**(7): p. 5470-9.

23.   Panwar, P., et al., *Effects of cysteine proteases on the structural and mechanical properties of collagen fibers.* J Biol Chem, 2013. **288**(8): p. 5940-50.

24.   Yasuda, Y., et al., *Cathepsin V, a novel and potent elastolytic activity expressed in activated macrophages.* J Biol Chem, 2004. **279**(35): p. 36761-70.

25.   Drenth, J., et al., *Structure of papain.* Nature, 1968. **218**(5145): p. 929-32.

26.   Kamphuis, I.G., et al., *Structure of papain refined at 1.65 A resolution.* J Mol Biol, 1984. **179**(2): p. 233-56.

27.   Fujishima, A., et al., *The crystal structure of human cathepsin L complexed with E-64.* FEBS Lett, 1997. **407**(1): p. 47-50.

28.   McGrath, M.E., et al., *Crystal structure of human cathepsin K complexed with a potent inhibitor.* Nat Struct Biol, 1997. **4**(2): p. 105-9.

29.   Zhao, B., et al., *Crystal structure of human osteoclast cathepsin K complex with E-64.* Nat Struct Biol, 1997. **4**(2): p. 109-11.

30.   Somoza, J.R., et al., *Crystal structure of human cathepsin V.* Biochemistry, 2000. **39**(41): p. 12543-51.

31.   Turkenburg, J.P., et al., *Structure of a Cys25-->Ser mutant of human cathepsin S.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 3): p. 451-5.

32.   Lecaille, F., D. Bromme, and G. Lalmanach, *Biochemical properties and regulation of cathepsin K activity.* Biochimie, 2008. **90**(2): p. 208-26.

33.   Turk, D., et al., *Revised definition of substrate binding sites of papain-like cysteine proteases.* Biol Chem, 1998. **379**(2): p. 137-47.

34.   Metzler, D.E., *The Cysteine Proteases (Thiol Proteases)*, in *Biochemistry: The Chemical Reactions of Living Cells*, J. Hayhurst, Editor. 2003, Academic Press / Elsevier: San Diego, California p. 618-619.

35.   Reiser, J., B. Adair, and T. Reinheckel, *Specialized roles for cysteine cathepsins in health and disease.* J Clin Invest, 2010. **120**(10): p. 3421-31.

36.   Chen, B. and M.O. Platt, *Multiplex zymography captures stage-specific activity profiles of cathepsins K, L, and S in human breast, lung, and cervical cancer.* J Transl Med, 2011. **9**: p. 109.

37.   Keegan, P.M., S. Surapaneni, and M.O. Platt, *Sickle cell disease activates peripheral blood mononuclear cells to induce cathepsins k and v activity in endothelial cells.* Anemia, 2012. **2012**: p. 201781.

38.   Mohamed, M.M. and B.F. Sloane, *Cysteine cathepsins: multifunctional enzymes in cancer.* Nat Rev Cancer, 2006. **6**(10): p. 764-75.

39.   Park, K.Y., G. Li, and M.O. Platt, *Monocyte-derived macrophage assisted breast cancer cell invasion as a personalized, predictive metric to score metastatic risk.* Sci Rep, 2015. **5**: p. 13855.

40.   Platt, M.O., et al., *Expression of cathepsin K is regulated by shear stress in cultured endothelial cells and is increased in endothelium in human atherosclerosis.* Am J Physiol Heart Circ Physiol, 2007. **292**(3): p. H1479-86.

41.   Platt, M.O. and W.A. Shockey, *Endothelial cells and cathepsins: Biochemical and biomechanical regulation.* Biochimie, 2016. **122**: p. 314-23.

42.   Bromme, D., P. Panwar, and S. Turan, *Cathepsin K osteoporosis trials, pycnodysostosis and mouse deficiency models: Commonalities and differences.* Expert Opin Drug Discov, 2016. **11**(5): p. 457-72.

43.   Hou, W.S., et al., *Characterization of novel cathepsin K mutations in the pro and mature polypeptide regions causing pycnodysostosis.* J Clin Invest, 1999. **103**(5): p. 731-8.

44.   Gauthier, J.Y., et al., *The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K.* Bioorganic & Medicinal Chemistry Letters, 2008. **18**(3): p. 923-928.

45.   Lewiecki, E.M., *Odanacatib, a cathepsin K inhibitor for the treatment of osteoporosis and other skeletal disorders associated with excessive bone remodeling.* IDrugs, 2009. **12**(12): p. 799-809.

46.   Bone, H.G., et al., *Odanacatib for the treatment of postmenopausal osteoporosis: development history and design and participant characteristics of LOFT, the Long-Term Odanacatib Fracture Trial.* Osteoporos Int, 2015. **26**(2): p. 699-712.

47.   Mullard, A., *Merck &Co. drops osteoporosis drug odanacatib.* Nat Rev Drug Discov, 2016. **15**(10): p. 669.

48.   Barascuk, N., et al., *Human macrophage foam cells degrade atherosclerotic plaques through cathepsin K mediated processes.* Bmc Cardiovascular Disorders, 2010. **10**.

49.   Donners, M.M., et al., *Cathepsin K Deficiency Prevents the Aggravated Vascular Remodeling Response to Flow Cessation in ApoE-/- Mice.* PLoS One, 2016. **11**(9): p. e0162595.

50.   Lafarge, J.C., et al., *Cathepsins and cystatin C in atherosclerosis and obesity.* Biochimie, 2010. **92**(11): p. 1580-6.

51.   Lutgens, E., et al., *Disruption of the cathepsin K gene reduces atherosclerosis progression and induces plaque fibrosis but accelerates macrophage foam cell formation.* Circulation, 2006. **113**(1): p. 98-107.

52.   Andrade, S.S., et al., *Cathepsin K induces platelet dysfunction and affects cell signaling in breast cancer - molecularly distinct behavior of cathepsin K in breast cancer.* BMC Cancer, 2016. **16**: p. 173.

53. Brubaker, K.D., et al., *Cathepsin K mRNA and protein expression in prostate cancer progression.* J Bone Miner Res, 2003. **18**(2): p. 222-30.

54. Mason, S.D. and J.A. Joyce, *Proteolytic networks in cancer.* Trends in Cell Biology, 2011. **21**(4): p. 228-237.

55. Hou, W.S., et al., *Comparison of cathepsins K and S expression within the rheumatoid and osteoarthritic synovium.* Arthritis Rheum, 2002. **46**(3): p. 663-74.

56. Porter, K.M., et al., *Cathepsin Protease Inhibition Reduces Endometriosis Lesion Establishment.* Reprod Sci, 2016. **23**(5): p. 623-9.

57. Sevenich, L., et al., *Expression of human cathepsin L or human cathepsin V in mouse thymus mediates positive selection of T helper cells in cathepsin L knock-out mice.* Biochimie, 2010. **92**(11): p. 1674-1680.

58. Nakagawa, T., et al., *Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus.* Science, 1998. **280**(5362): p. 450-3.

59. Hiwasa, T., et al., *Inhibition of cathepsin L-induced degradation of epidermal growth factor receptors by c-Ha-ras gene products.* Biochem Biophys Res Commun, 1988. **151**(1): p. 78-85.

60. Reinheckel, T., et al., *The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling.* J Cell Sci, 2005. **118**(Pt 15): p. 3387-95.

61. Akache, B., et al., *A two-hybrid screen identifies cathepsins B and L as uncoating factors for adeno-associated virus 2 and 8.* Mol Ther, 2007. **15**(2): p. 330-9.

62. Hood, C.L., et al., *Biochemical and structural characterization of cathepsin L-processed Ebola virus glycoprotein: implications for viral entry and immunogenicity.* J Virol, 2010. **84**(6): p. 2972-82.

63. Huang, I.C., et al., *SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells.* J Biol Chem, 2006. **281**(6): p. 3198-203.

64. Pager, C.T., et al., *A mature and fusogenic form of the Nipah virus fusion protein requires proteolytic processing by cathepsin L.* Virology, 2006. **346**(2): p. 251-7.

65. Pager, C.T. and R.E. Dutch, *Cathepsin L is involved in proteolytic processing of the Hendra virus fusion protein.* J Virol, 2005. **79**(20): p. 12714-20.

66.   Shah, P.P., et al., *A small-molecule oxocarbazate inhibitor of human cathepsin L blocks severe acute respiratory syndrome and ebola pseudotype virus infection into human embryonic kidney 293T cells.* Mol Pharmacol, 2010. **78**(2): p. 319-24.

67.   Simmons, G., et al., *Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry.* Proc Natl Acad Sci U S A, 2005. **102**(33): p. 11876-81.

68.   Platt, M.O., R.F. Ankeny, and H. Jo, *Laminar shear stress inhibits cathepsin L activity in endothelial cells.* Arterioscler Thromb Vasc Biol, 2006. **26**(8): p. 1784-90.

69.   Lankelma, J.M., et al., *Cathepsin L, target in cancer treatment?* Life Sci, 2010. **86**(7-8): p. 225-33.

70.   Sudhan, D.R. and D.W. Siemann, *Cathepsin L targeting in cancer treatment.* Pharmacol Ther, 2015. **155**: p. 105-16.

71.   Wilder, C.L., et al., *Differential cathepsin responses to inhibitor-induced feedback: E-64 and cystatin C elevate active cathepsin S and suppress active cathepsin L in breast cancer cells.* Int J Biochem Cell Biol, 2016. **79**: p. 199-208.

72.   Goulet, B., et al., *A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor.* Mol Cell, 2004. **14**(2): p. 207-19.

73.   Goulet, B., M. Truscott, and A. Nepveu, *A novel proteolytically processed CDP/Cux isoform of 90 kDa is generated by cathepsin L.* Biol Chem, 2006. **387**(9): p. 1285-93.

74.   Ceru, S., et al., *Stefin B interacts with histones and cathepsin L in the nucleus.* J Biol Chem, 2010. **285**(13): p. 10078-86.

75.   Goulet, B., et al., *Increased expression and activity of nuclear cathepsin L in cancer cells suggests a novel mechanism of cell transformation.* Mol Cancer Res, 2007. **5**(9): p. 899-907.

76.   Cirman, T., et al., *Selective disruption of lysosomes in HeLa cells triggers apoptosis mediated by cleavage of Bid by multiple papain-like lysosomal cathepsins.* J Biol Chem, 2004. **279**(5): p. 3578-87.

77.   Ishisaka, R., et al., *Participation of a cathepsin L-type protease in the activation of caspase-3.* Cell Struct Funct, 1999. **24**(6): p. 465-70.

78.   Beck, H., et al., *Cathepsin S and an asparagine-specific endoprotease dominate the proteolytic processing of human myelin basic protein in vitro.* European journal of immunology, 2001. **31**: p. 3726-3736.

79.   Lemere, C.A., et al., *The lysosomal cysteine protease, cathepsin S, is increased in Alzheimer's disease and Down syndrome brain. An immunocytochemical study.* The American journal of pathology, 1995. **146**: p. 848-860.

80.   Munger, J.S., et al., *Lysosomal processing of amyloid precursor protein to Aβ peptides: a distinct role for cathepsin S.* Biochemical Journal, 1995. **311**: p. 299-305.

81.   Naour, N., et al., *Cathepsins in human obesity: Changes in energy balance predominantly affect cathepsin S in adipose tissue and in circulation.* Journal of Clinical Endocrinology and Metabolism, 2010. **95**: p. 1861-1868.

82.   Ou, Z., et al., *The genetic polymorphisms of cathepsin S were associated with metabolic disorders in a Chinese Han population.* Gene, 2013. **526**: p. 385-389.

83.   Taleb, S., et al., *Cathepsin S promotes human preadipocyte differentiation: Possible involvement of fibronectin degradation.* Endocrinology, 2006. **147**: p. 4950-4959.

84.   Kim, N., et al., *Overexpression of Cathepsin S Induces Chronic Atopic Dermatitis in Mice.* Journal of Investigative Dermatology, 2012. **132**: p. 1169-1176.

85.   Reddy, V.B., et al., *Cathepsin S Elicits Itch and Signals via Protease-Activated Receptors.* Journal of Investigative Dermatology, 2010. **130**: p. 1468-1470.

86.   Reddy, V.B., et al., *Redefining the concept of protease-activated receptors: cathepsin S evokes itch via activation of Mrgprs.* Nature communications, 2015. **6**: p. 7864.

87.   Cattaruzza, F., et al., *Cathepsin S is activated during colitis and causes visceral hyperalgesia by a PAR2-dependent mechanism in mice.* Gastroenterology, 2011. **141**: p. 1864-74.e1-3.

88.   Zhao, P., et al., *Cathepsin S causes inflammatory pain via biased agonism of PAR2 and TRPV4.* Journal of Biological Chemistry, 2014. **289**: p. 27215-27234.

89.   Adachi, W., et al., *Isolation and characterization of human cathepsin V: a major proteinase in corneal epithelium.* Invest Ophthalmol Vis Sci, 1998. **39**(10): p. 1789-96.

90.   Hagemann, S., et al., *The human cysteine protease cathepsin V can compensate for murine cathepsin L in mouse epidermis and hair follicles.* European journal of cell biology, 2004. **83**(11-12): p. 775-780.

91.   Roth, W., et al., *Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and pertubation of hair follicle cycling.* FASEB J, 2000. **14**(13): p. 2075-86.

92.   Kenney, M.C., et al., *Increased Levels of Catalase and Cathepsin V/L2 but Decreased TIMP-1 in Keratoconus Corneas: Evidence that Oxidative Stress Plays a Role in This Disorder.* Investigative Opthalmology & Visual Science, 2005. **46**(3): p. 823-10.

93.   Tolosa, E., et al., *Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis.* J Clin Invest, 2003. **112**(4): p. 517-26.

94.   Keegan, P.M., C.L. Wilder, and M.O. Platt, *Tumor necrosis factor alpha stimulates cathepsin K and V activity via juxtacrine monocyte-endothelial cell signaling and JNK activation.* Mol Cell Biochem, 2012. **367**(1-2): p. 65-72.

95.   Santamaría, I., et al., *Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas.* Cancer research, 1998. **58**(8): p. 1624-1630.

96.   Vasiljeva, O., et al., *Recombinant human procathepsin S is capable of autocatalytic processing at neutral pH in the presence of glycosaminoglycans.* FEBS Lett, 2005. **579**(5): p. 1285-90.

97.   Dahl, S.W., et al., *Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing.* Biochemistry, 2001. **40**(6): p. 1671-8.

98.   Coulombe, R., et al., *Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment.* EMBO J, 1996. **15**(20): p. 5492-503.

99.   Hanada, K., et al., *Structure and Synthesis of E-64, a New Thiol Protease Inhibitor.* Structure and Synthesis of E-64, a New Thiol Protease Inhibitor, 1978. **42**(3): p. 529-536.

100.   Hanada, K., et al., *Isolation and Characterization of E-64, a New Thiol Protease Inhibitor.* Agricultural and Biological Chemistry, 1978. **42**(3): p. 523-528.

101.   Matsumoto, K., et al., *Structural basis of inhibition of cysteine proteases by E-64 and its derivatives.* Biopolymers, 1999. **51**(1): p. 99-107.

102.   Berdowska, I., *Cysteine proteases as disease markers.* Clin Chim Acta, 2004. **342**(1-2): p. 41-69.

103.   Burden, R.E., et al., *Antibody-Mediated Inhibition of Cathepsin S Blocks Colorectal Tumor Invasion and Angiogenesis.* Clinical Cancer Research, 2009. **15**.

104.   Gocheva, V., et al., *Distinct roles for cysteine cathepsin genes in multistage tumorigenesis.* Genes Dev, 2006. **20**(5): p. 543-56.

105.   Grotsky, D.A., et al., *BRCA1 loss activates cathepsin L-mediated degradation of 53BP1 in breast cancer cells.* J Cell Biol, 2013. **200**(2): p. 187-202.

106.   Hira, V.V., et al., *Cathepsin K cleavage of SDF-1alpha inhibits its chemotactic activity towards glioblastoma stem-like cells.* Biochim Biophys Acta, 2017. **1864**(3): p. 594-603.

107.   Littlewood-Evans, A.J., et al., *The osteoclast-associated protease cathepsin K is expressed in human breast carcinoma.* Cancer Res, 1997. **57**(23): p. 5386-90.

108.   Ohrvik, H., et al., *Cathepsin Protease Controls Copper and Cisplatin Accumulation via Cleavage of the Ctr1 Metal-binding Ectodomain.* J Biol Chem, 2016. **291**(27): p. 13905-16.

109.   Olson, O.C. and J.A. Joyce, *Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response.* Nat Rev Cancer, 2015. **15**(12): p. 712-29.

110.   Palermo, C. and J.A. Joyce, *Cysteine cathepsin proteases as pharmacological targets in cancer.* Trends Pharmacol Sci, 2008. **29**(1): p. 22-8.

111.   Park, K.Y., W.A. Li, and M.O. Platt, *Patient specific proteolytic activity of monocyte-derived macrophages and osteoclasts predicted with temporal kinase activation states during differentiation.* Integr Biol (Camb), 2012. **4**(12): p. 1459-69.

112.   Small, D.M., et al., *Cathepsin S from both tumor and tumor-associated cells promote cancer growth and neovascularization.* International Journal of Cancer, 2013. **133**: p. 2102-2112.

113.   Sobotič, B., et al., *Proteomic Identification of Cysteine Cathepsin Substrates Shed from the Surface of Cancer Cells.* Molecular & Cellular Proteomics, 2015. **14**: p. 2213-2228.

114.   Vasiljeva, O., et al., *Emerging roles of cysteine cathepsins in disease and their potential as drug targets.* Curr Pharm Des, 2007. **13**(4): p. 387-403.

115. Wilkinson, R.D.A., et al., *Cathepsin S: therapeutic, diagnostic, and prognostic potential.* Biological Chemistry, 2015. **396**: p. 867-882.

116. Jormsjo, S., et al., *Differential expression of cysteine and aspartic proteases during progression of atherosclerosis in apolipoprotein E-deficient mice.* Am J Pathol, 2002. **161**(3): p. 939-45.

117. Rodgers, K.J., et al., *Destabilizing Role of Cathepsin S in Murine Atherosclerotic Plaques.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2006. **26**.

118. Samokhin, A.O., et al., *Pharmacological inhibition of cathepsin S decreases atherosclerotic lesions in Apoe-/- mice.* J Cardiovasc Pharmacol, 2010. **56**(1): p. 98-105.

119. Sukhova, G.K., et al., *Deficiency of cathepsin S reduces atherosclerosis in LDL receptor-deficient mice.* J Clin Invest, 2003. **111**(6): p. 897-906.

120. Bromme, D. and F. Lecaille, *Cathepsin K inhibitors for osteoporosis and potential off-target effects.* Expert Opin Investig Drugs, 2009. **18**(5): p. 585-600.

121. Parks, A.N., et al., *Supraspinatus tendon overuse results in degenerative changes to tendon insertion region and adjacent humeral cartilage in a rat model.* J Orthop Res, 2016.

122. Seto, S.P., et al., *Cathepsins in Rotator Cuff Tendinopathy: Identification in Human Chronic Tears and Temporal Induction in a Rat Model.* Ann Biomed Eng, 2015. **43**(9): p. 2036-46.

123. Garber, K., *Two pioneering osteoporosis drugs finally approach approval.* Nat Rev Drug Discov, 2016. **15**(7): p. 445-6.

124. Payne, C.D., et al., *Pharmacokinetics and pharmacodynamics of the cathepsin S inhibitor, LY3000328, in healthy subjects.* Br J Clin Pharmacol, 2014. **78**(6): p. 1334-42.

125. Chan, V., et al., *Development of miniaturized walking biological machines.* Sci Rep, 2012. **2**: p. 857.

126. Cvetkovic, C., et al., *Three-dimensionally printed biological machines powered by skeletal muscle.* Proc Natl Acad Sci U S A, 2014. **111**(28): p. 10125-30.

127. Raman, R., C. Cvetkovic, and R. Bashir, *A modular approach to the design, fabrication, and characterization of muscle-powered biological machines.* Nat Protoc, 2017. **12**(3): p. 519-533.

128. Cvetkovic, C., et al., *Investigating the Life Expectancy and Proteolytic Degradation of Engineered Skeletal Muscle Biological Machines.* Sci Rep, 2017. **7**: p. 3775.

129. Ferrall-Fairbanks, M.C., et al., *PACMANS: A Bioinformatically Informed Algorithm to Predict, Design, and Disrupt Protease-on-Protease Hydrolysis.* Protein Sci, 2017.

130. Brix, K., et al., *Cysteine cathepsins: cellular roadmap to different functions.* Biochimie, 2008. **90**(2): p. 194-207.

131. Rawlings, N.D., et al., *MEROPS: the database of proteolytic enzymes, their substrates and inhibitors.* Nucleic Acids Res, 2014. **42**(Database issue): p. D503-9.

132. Rawlings, N.D., A.J. Barrett, and R. Finn, *Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors.* Nucleic Acids Res, 2016. **44**(D1): p. D343-50.

133. Ho, S.N., et al., *Site-directed mutagenesis by overlap extension using the polymerase chain reaction.* Gene, 1989. **77**(1): p. 51-9.

134. Li, W.A., et al., *Detection of femtomole quantities of mature cathepsin K with zymography.* Anal Biochem, 2010. **401**(1): p. 91-8.

135. Wilder, C.L., et al., *Manipulating substrate and pH in zymography protocols selectively distinguishes cathepsins K, L, S, and V activity in cells and tissues.* Arch Biochem Biophys, 2011. **516**(1): p. 52-7.

136. Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain.* Biochem Biophys Res Commun, 1967. **425**(3): p. 497-502.

137. Bordo, D. and P. Argos, *Suggestions for "safe" residue substitutions in site-directed mutagenesis.* J Mol Biol, 1991. **217**(4): p. 721-9.

138. Abdul-Hussien, H., et al., *Collagen degradation in the abdominal aneurysm: a conspiracy of matrix metalloproteinase and cysteine collagenases.* Am J Pathol, 2007. **170**(3): p. 809-17.

139. Emmert-Buck, M.R., et al., *Increased gelatinase A (MMP-2) and cathepsin B activity in invasive tumor regions of human colon cancer samples.* Am J Pathol, 1994. **145**(6): p. 1285-90.

140. Mason, S.D. and J.A. Joyce, *Proteolytic networks in cancer.* Trends Cell Biol, 2011.

141. Platt, M.O., et al., *Cyclic pressure and shear stress regulate matrix metalloproteinases and cathepsin activity in porcine aortic valves.* J Heart Valve Dis, 2006. **15**(5): p. 622-9.

142. Collier, I.E., et al., *H-ras oncogene-transformed human bronchial epithelial cells (TBE-1) secrete a single metalloprotease capable of degrading basement membrane collagen.* J Biol Chem, 1988. **263**(14): p. 6579-87.

143. Choe, Y., et al., *Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities.* J Biol Chem, 2006. **281**(18): p. 12824-32.

144. Menard, R., et al., *The specificity of the S1' subsite of cysteine proteases.* FEBS Lett, 1993. **328**(1-2): p. 107-10.

145. Bromme, D., et al., *Peptidyl vinyl sulphones: a new class of potent and selective cysteine protease inhibitors: S2P2 specificity of human cathepsin O2 in comparison with cathepsins S and L.* Biochem J, 1996. **315 (Pt 1)**: p. 85-9.

146. Song, J., et al., *PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites.* PLoS One, 2012. **7**(11): p. e50300.

147. Gasteiger, E.H., C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D., Bairoch, A. , *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomic Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press: Totowa, New Jersey. p. 571-607.

148. Boyd, S.E., et al., *PoPS: a computational tool for modeling and predicting protease specificity.* J Bioinform Comput Biol, 2005. **3**(3): p. 551-85.

149. Averett, R.D., et al., *Computational imaging analysis of fibrin matrices with the inclusion of erythrocytes from homozygous SS blood reveals agglomerated and amorphous structures.* J Thromb Thrombolysis, 2017. **43**(1): p. 43-51.

150. Norton, D.G., et al., *Computational imaging analysis of glycated fibrin gels reveals aggregated and anisotropic structures.* J Biomed Mater Res A, 2017.

151. Mosesson, M.W., *Fibrinogen and fibrin structure and functions.* J Thromb Haemost, 2005. **3**(8): p. 1894-904.

152. Herrick, S., et al., *Fibrinogen.* Int J Biochem Cell Biol, 1999. **31**(7): p. 741-6.

153. Stubbs, M.T., et al., *The interaction of thrombin with fibrinogen. A structural basis for its specificity.* Eur J Biochem, 1992. **206**(1): p. 187-95.

154. Kamath, S. and G.Y. Lip, *Fibrinogen: biochemistry, epidemiology and determinants.* QJM, 2003. **96**(10): p. 711-29.

155. Doolittle, R.F., G. Spraggon, and S.J. Everse, *Three-dimensional structural studies on fragments of fibrinogen and fibrin.* Curr Opin Struct Biol, 1998. **8**(6): p. 792-8.

156. Weisel, J.W. and R.I. Litvinov, *Fibrin Formation, Structure and Properties.* Subcell Biochem, 2017. **82**: p. 405-456.

157. Blomback, B., et al., *A two-step fibrinogen--fibrin transition in blood coagulation.* Nature, 1978. **275**(5680): p. 501-5.

158. Binnie, C.G. and S.T. Lord, *The fibrinogen sequences that interact with thrombin.* Blood, 1993. **81**(12): p. 3186-92.

159. Whisler, J.A., M.B. Chen, and R.D. Kamm, *Control of perfusable microvascular network morphology using a multiculture microfluidic system.* Tissue Eng Part C Methods, 2014. **20**(7): p. 543-52.

160. Moya, M.L., et al., *In vitro perfused human capillary networks.* Tissue Eng Part C Methods, 2013. **19**(9): p. 730-7.

161. Ahmed, T.A., E.V. Dare, and M. Hincke, *Fibrin: a versatile scaffold for tissue engineering applications.* Tissue Eng Part B Rev, 2008. **14**(2): p. 199-215.

162. Comeau, S.R., et al., *ClusPro: an automated docking and discrimination method for the prediction of protein complexes.* Bioinformatics, 2004. **20**(1): p. 45-50.

163. Comeau, S.R., et al., *ClusPro: a fully automated algorithm for protein-protein docking.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W96-9.

164. Kozakov, D., et al., *PIPER: an FFT-based protein docking program with pairwise potentials.* Proteins, 2006. **65**(2): p. 392-406.

165. Kozakov, D., et al., *How good is automated protein docking?* Proteins, 2013. **81**(12): p. 2159-66.

166. Kozakov, D., et al., *The ClusPro web server for protein-protein docking.* Nat Protoc, 2017. **12**(2): p. 255-278.

167. Malave, I., et al., *Levels of tumor necrosis factor alpha/cachectin (TNF alpha) in sera from patients with sickle cell disease.* Acta Haematol, 1993. **90**(4): p. 172-6.

168. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.

169. Rose, P.W., et al., *The RCSB Protein Data Bank: views of structural biology for basic and applied research and education.* Nucleic Acids Res, 2015. **43**(Database issue): p. D345-56.

170. Turk, B., D. Turk, and V. Turk, *Lysosomal cysteine proteases: more than scavengers.* Biochim Biophys Acta, 2000. **1477**(1-2): p. 98-111.

171. Kirschbaum, N.E. and A.Z. Budzynski, *A unique proteolytic fragment of human fibrinogen containing the A alpha COOH-terminal domain of the native molecule.* J Biol Chem, 1990. **265**(23): p. 13669-76.

172. Spraggon, G., S.J. Everse, and R.F. Doolittle, *Crystal structures of fragment D from human fibrinogen and its crosslinked counterpart from fibrin.* Nature, 1997. **389**(6650): p. 455-62.

173. Pandya, B.V., et al., *Polymerization site in the beta chain of fibrin: mapping of the B beta 1-55 sequence.* Biochemistry, 1991. **30**(1): p. 162-8.

174. Kamm, R.D. and R. Bashir, *Creating living cellular machines.* Ann Biomed Eng, 2014. **42**(2): p. 445-59.

175. Duffy, R.M. and A.W. Feinberg, *Engineered skeletal muscle tissue for soft robotics: fabrication strategies, current applications, and future challenges.* Wiley Interdiscip Rev Nanomed Nanobiotechnol, 2014. **6**(2): p. 178-95.

176. Dennis, R.G. and P.E. Kosnik, 2nd, *Excitability and isometric contractile properties of mammalian skeletal muscle constructs engineered in vitro.* In Vitro Cell Dev Biol Anim, 2000. **36**(5): p. 327-35.

177. Raman, R., et al., *Optogenetic skeletal muscle-powered adaptive biological machines.* Proc Natl Acad Sci U S A, 2016. **113**(13): p. 3497-502.

178. Khodabukus, A. and K. Baar, *Regulating fibrinolysis to engineer skeletal muscle from the C2C12 cell line.* Tissue Eng Part C Methods, 2009. **15**(3): p. 501-11.

179. Watras, J.M., *Berne & Levy Physiology*. 2008: Mosby Elsevier.

180. Janmey, P.A., J.P. Winer, and J.W. Weisel, *Fibrin gels and their clinical and bioengineering applications.* J R Soc Interface, 2009. **6**(30): p. 1-10.

181. Weisel, J.W., *Structure of fibrin: impact on clot stability.* J Thromb Haemost, 2007. **5 Suppl 1**: p. 116-24.

182. Weisel, J.W., *Advances in Protein Chemistry* Vol. 70. 2005: Academic Press.

183. Lorentz, K.M., et al., *Engineered aprotinin for improved stability of fibrin biomaterials.* Biomaterials, 2011. **32**(2): p. 430-8.

184. Prentice, C.R., *Basis of antifibrinolytic therapy.* J Clin Pathol Suppl (R Coll Pathol), 1980. **14**: p. 35-40.

185. Kupcsik, L., M. Alini, and M.J. Stoddart, *Epsilon-aminocaproic acid is a useful fibrin degradation inhibitor for cartilage tissue engineering.* Tissue Eng Part A, 2009. **15**(8): p. 2309-13.

186. Suelves, M., et al., *Plasmin activity is required for myogenesis in vitro and skeletal muscle regeneration in vivo.* Blood, 2002. **99**(8): p. 2835-44.

187. Ebisui, C., et al., *Changes of proteasomes and cathepsins activities and their expression during differentiation of C2C12 myoblasts.* J Biochem, 1995. **117**(5): p. 1088-94.

188. Carmeli, E., et al., *Matrix metalloproteinases and skeletal muscle: a brief review.* Muscle Nerve, 2004. **29**(2): p. 191-7.

189. Chen, X. and Y. Li, *Role of matrix metalloproteinases in skeletal muscle: migration, differentiation, regeneration and fibrosis.* Cell Adh Migr, 2009. **3**(4): p. 337-41.

190. Kherif, S., et al., *Expression of matrix metalloproteinases 2 and 9 in regenerating skeletal muscle: a study in experimentally injured and mdx muscles.* Dev Biol, 1999. **205**(1): p. 158-70.

191. Lluri, G. and D.M. Jaworski, *Regulation of TIMP-2, MT1-MMP, and MMP-2 expression during C2C12 differentiation.* Muscle Nerve, 2005. **32**(4): p. 492-9.

192. Chan, V., et al., *Three-dimensional photopatterning of hydrogels using stereolithography for long-term cell encapsulation.* Lab Chip, 2010. **10**(16): p. 2062-70.

193. Chan, V., et al., *Multi-material bio-fabrication of hydrogel cantilevers and actuators with stereolithography.* Lab Chip, 2012. **12**(1): p. 88-98.

194. Jacobs, P.F. *Rapid Prototyping and Manufacturing: Fundamentals of StereoLithography.* in *Society of Manufacturing Engineers.* 1992.

195. Arcaute, K., B.K. Mann, and R.B. Wicker, *Stereolithography of three-dimensional bioactive poly(ethylene glycol) constructs with encapsulated cells.* Ann Biomed Eng, 2006. **34**(9): p. 1429-41.

196.   Uzel, S.G., et al., *Microfluidic device for the formation of optically excitable, three-dimensional, compartmentalized motor units.* Sci Adv, 2016. **2**(8): p. e1501429.

197.   Bajaj, P., et al., *Patterning the differentiation of C2C12 skeletal myoblasts.* Integr Biol (Camb), 2011. **3**(9): p. 897-909.

198.   Toth, M., A. Sohail, and R. Fridman, *Assessment of gelatinases (MMP-2 and MMP-9) by gelatin zymography.* Methods Mol Biol, 2012. **878**: p. 121-35.

199.   Melchels, F.P., J. Feijen, and D.W. Grijpma, *A review on stereolithography and its applications in biomedical engineering.* Biomaterials, 2010. **31**(24): p. 6121-30.

200.   Bartolo, P.J., *Stereolithography: Materials, Processes and Applications* 2011: Springer US.

201.   Raman, R. and R. Bashir, *Stereolithographic 3D Bioprinting for Biomedical Applications. Essentials of 3D Biofabrication and Translation.* . 2015: Elsevier Inc.

202.   Duan, C., H. Ren, and S. Gao, *Insulin-like growth factors (IGFs), IGF receptors, and IGF-binding proteins: roles in skeletal muscle growth and differentiation.* Gen Comp Endocrinol, 2010. **167**(3): p. 344-51.

203.   Chamberlain, J.S., J.B. Jaynes, and S.D. Hauschka, *Regulation of creatine kinase induction in differentiating mouse myoblasts.* Mol Cell Biol, 1985. **5**(3): p. 484-92.

204.   Szlachcic, A., M. Zakrzewska, and J. Otlewski, *Longer action means better drug: tuning up protein therapeutics.* Biotechnol Adv, 2011. **29**(4): p. 436-41.

205.   Bechet, D., et al., *Lysosomal proteolysis in skeletal muscle.* Int J Biochem Cell Biol, 2005. **37**(10): p. 2098-114.

206.   Lewis, M.P., et al., *Gelatinase-B (matrix metalloproteinase-9; MMP-9) secretion is involved in the migratory phase of human and murine muscle cell cultures.* J Muscle Res Cell Motil, 2000. **21**(3): p. 223-33.

207.   Nguyen, A.H., et al., *MMP-mediated mesenchymal morphogenesis of pluripotent stem cell aggregates stimulated by gelatin methacrylate microparticle incorporation.* Biomaterials, 2016. **76**: p. 66-75.

208.   Platt, M.O., et al., *Low-Cost Method to Monitor Patient Adherence to HIV Antiretroviral Therapy Using Multiplex Cathepsin Zymography.* Mol Biotechnol, 2016. **58**(1): p. 56-64.

209. Nemori, R. and T. Tachikawa, *A review for in situ zymography: method for localization of protease activities in a tissue.* The Tissue Culture Engineering, 1999. **25**(9).

210. Weissmann, G. and I. Spilberg, *Breakdown of cartilage proteinpolysaccharide by lysosomes.* Arthritis Rheum, 1968. **11**(2): p. 162-9.

211. Brown, S.I., C.W. Hook, and M.P. Tragakis, *Presence, significance, and inhibition of lysosomal proteoglycanases.* Invest Ophthalmol, 1972. **11**(3): p. 149-52.

212. Fanjul-Fernandez, M., et al., *Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models.* Biochim Biophys Acta, 2010. **1803**(1): p. 3-19.

213. Sakar, M.S., et al., *Formation and optogenetic control of engineered 3D skeletal muscle bioactuators.* Lab Chip, 2012. **12**(23): p. 4976-85.

214. Kjaer, M., *Role of extracellular matrix in adaptation of tendon and skeletal muscle to mechanical loading.* Physiol Rev, 2004. **84**(2): p. 649-98.

215. Fortelny, N., et al., *Network analyses reveal pervasive functional regulation between proteases in the human protease web.* PLoS Biol, 2014. **12**(5): p. e1001869.

216. Christensen, J. and V.P. Shastri, *Matrix-metalloproteinase-9 is cleaved and activated by cathepsin K.* BMC Res Notes, 2015. **8**: p. 322.

217. Matsukura, U., et al., *Mode of degradation of myofibrillar proteins by an endogenous protease, cathepsin L.* Biochim Biophys Acta, 1981. **662**(1): p. 41-7.

218. Tjondrokoesoemo, A., et al., *Cathepsin S Contributes to the Pathogenesis of Muscular Dystrophy in Mice.* J Biol Chem, 2016. **291**(19): p. 9920-8.

219. Pinder, J.E., J.G. Wiener, and M.H. Smith, *The Weibull Distribution: A New Method of Summarizing Survivorship Data.* Ecology, 1978. **59**(1): p. 175-179.

220. Chakraborti, S., et al., *Regulation of matrix metalloproteinases: an overview.* Mol Cell Biochem, 2003. **253**(1-2): p. 269-85.

221. Massova, I., et al., *Matrix metalloproteinases: structures, evolution, and diversification.* FASEB J, 1998. **12**(12): p. 1075-95.

222. Bechet, D.M., et al., *Developmental control of cathepsin B expression in bovine fetal muscles.* Arch Biochem Biophys, 1996. **334**(2): p. 362-8.

223. Le Gall, C., E. Bonnelye, and P. Clezardin, *Cathepsin K inhibitors as treatment of bone metastasis.* Curr Opin Support Palliat Care, 2008. **2**(3): p. 218-22.

224. Brömme, D. and F. Lecaille, *Cathepsin K inhibitors for osteoporosis and potential off-target effects.* Expert Opinion on Investigational Drugs, 2009. **18**(5): p. 585-600.

225. Hoops, S., et al., *COPASI--a COmplex PAthway SImulator.* Bioinformatics, 2006. **22**(24): p. 3067-74.

226. Milo, R. and R. Phillips, *Cell Biology By the Numbers.* 1 ed. 2015: Garland Science.

227. Green, N.M., *Avidin.* Adv Protein Chem, 1975. **29**: p. 85-133.

228. Holmberg, A., et al., *The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures.* Electrophoresis, 2005. **26**(3): p. 501-10.

229. Boukouvala, F., R. Misener, and C.A. Floudas, *Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO.* European Journal of Operational Research, 2016. **252**(3): p. 701-727.

230. Yao, F., et al., *Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells.* Hum Gene Ther, 1998. **9**(13): p. 1939-50.