STRUCTURAL REQUIREMENTS FOR THE ELASTOLYTIC AND COLLAGENOLYTIC ACTIVITIES OF CATHEPSINS AND THE IDENTIFICATION OF EXOSITE INHIBITORS

by

Xin Du

B.Sc., The University of British Columbia, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREEE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2013

Abstract

It has become increasingly transparent that substrate recognition and degradation of extracellular matrix proteins such as elastin and collagen by proteases require more than the binding to small areas around the scissile bond of the target protein. Important surface structures on proteases, known as exosites or the formation of protease complexes are likely required for the correct positioning and modification of the substrate. This thesis embarks on the identification of such exosites in lysosomal cysteine cathepsins and their involvement in protein unfolding that are required for their elastolytic and collagenolytic activities.

In chapter 2, two exosites were identified in cathepsin V that are crucial for the degradation of insoluble elastin. Both exosites are distant from the active site of the protease. Replacement of both exosites completely abolished the elastolytic activity without affecting the general proteolytic efficacy of cathepsin V. Although the exact mechanism of contribution of these exosites to elastolysis is yet to be elucidated, the finding that the double exosite variant failed to bind to insoluble elastin implies that these exosites are involved in substrate recognition.

In chapter 3, the involvement of exosites and the effect of protease oligomerization on the collagenase activity of cathepsin K was studied. Two mechanistic models including a cathepsin K/GAG tetramer and a dimer were proposed based on available crystal structures. Both models, although displaying different modes of GAG binding, share a number of important amino acid residues in their protein-protein interactions. Mutational, biochemical, and structural analysis revealed various mechanistic aspects of substrate specificity towards soluble and insoluble collagens, respectively.

In chapter 4, a library of 1280 known drug derivatives was screened using a fluorometric polarization assay to identify potential exosite inhibitors that prevent the formation of active cathepsin K/GAG complexes. Two groups of compounds were identified: 1) polyanionic and 2) polyaromatic compounds, whose IC_{50} values for the inhibition of soluble tropocollagen degradation were between 10 – 186 μM. Exosite inhibitors might have the advantage of overcoming the off-site effects of active site-directed inhibitors presently in development.

Preface

Dr. Dieter Brömme was the principle investigator of the research project. The financial support for this thesis was provided by CIHR grants held by Dr. Brömme. Xin Du has been responsible for performing the research, analysis of the research data, and manuscript preparation for all the work described in this thesis with the exception of:

Chapter 2: Nelson Chen generated and characterized mutants M1 to M5 as well as M9 and M10. Mr. Andre Wang helped to obtain scanning electron microscopy images of cathepsin-treated elastin samples. Dr. Charles Craik from UCSF kindly characterized the M11 mutant with his peptide library to determine the subsite specificity of this mutant.

Chapter 3: Dr. Adleke Aguda was responsible for the development of the cathepsin K/GAG dimer model. Dr. Vidhu Sharma generated mutant IFM1 of cathepsin K. Dr. Hongbin Li and Dr. Guillaume Lamour advised and assisted on the morphological studies of cathepsin K/GAG complexes using atomic force microscopy.

Chapter 4: Summer student Annie Hsu contributed under my supervision to this project and assisted in the preparation of the manuscript. Dr. Preety Panwar provided her SEM expertise to the characterization of collagen fibers. The screening of the drug candidate library was overseen by Dr. Tom Pfeifer and Ms. Jadwiga Kaleta at the Center for Drug Research and Development. Dr. Hongbin Li and Dr. Guillaume Lamour advised on the analysis of the atomic force microscopy data.

Table of Contents

List of Tables

List of Figures

Abbreviations

AFM	atomic force microscopy
ATC	Aurintricarboxylic acid
C4-S	chondroitin sulfate A
CAT	catalytic domain
CD	circular dichroism
CS	chondroitin sulfate
DHT	dihydrotanshinone I
DS	dermatan sulfate
DTT	Dithiothreitol
ECM	extracellular matrix
EDS	Ehlers-Danlos syndrome
EGCG	(-)-epigallocatechin gallate
FPA	High-throughput Fluorescence Polarization Assay
GAGs	Glycosaminoglycan
HLE	human leukocyte elastase
HPX	hemopexin domain
IF1	protein interface 1
IF2	protein interface 2
IF3	protein interface 3
MMPs	matrix metalloproteases
MPS	Mucopolysaccharidose
OA	Osteoarthritis
OI	osteogenesis imperfecta
PGs	Proteoglycans
SCA	substrate cleavage assay
SEM	scanning electron microscopy
SERMs	selective estrogen-receptor modulators
THP	triple helical peptide

Acknowledgements

A doctoral journey is usually accomplished with bitter and sweet moments, but my own experience has turned out rather enjoyable. This was not possible without the many people who I am deeply indebted to. They are the ones who made this thesis possible.

My first debt of gratitude must go to my supervisor, Dr. Dieter Brömme, for his excellent mentorship and brilliant guidance. It has been a great honor to work in his lab. His vision, enthusiasm, patience, inspiration and continued support are of great motivation to me throughout my graduate school career.

I am greatly thankful to my committee members, Dr. Christopher Overall, Dr. Natalie Strynadka and Dr. Gary Brayer, for their invaluable suggestions and insights. Their guidance has always been helpful, and I owe them my most sincere appreciation.

It has been a privilege to work with such an amazing group of people over the last few years; people who constantly offered me their friendship, wisdom and assistance. Their contribution to my work and life means more to me than I could ever express. I am forever grateful to Dr. Yana Selent, Dr. Luciano Puzer, Dr. Adeleke Aguda, Dr. Preety Panwar Dr. Vidhu Sharma, Dr. Nelson Chen, Jadwiga Kaleta, Paul Lythgo, Raymond Pan, Chun Mei Li, Daniel Lim and Annie Hsu. Additional thanks to our collaborators Dr. Federico I. Rosell from Dr. Grant Mauk lab, Dr. Johan Janzen from Dr. Jayachandran N Kizhakkedathu lab, Nhan T. Nguyen from Dr. Gary Brayer lab, Dr. Guillaume Lamour from Dr. Hongbin Li lab, and Andre Wong from UBC Dentistry.

I would like to thank the funding agencies Canadian Institute of Health Research.

Finally and most importantly, I wish to thank my dear grandma, my parents and my relatives. Their love is the essential driving force for my accomplishment. I owe them everything and I wish that words alone could express how much I love them.

Dedication

This work is dedicated to:

The memories of my grandfather Caizhong Du.

My grandmother, Yuzhen Zhou, despite never having a chance to receive proper education due to the unfortunate living conditions during her earlier years, has never stopped believing in the importance of education for her children and grandchildren.

My parents, Mingcong Du and Yuyan Lin, who are still making countless sacrifices to ensure the very best for me.

Chapter 1. Introduction

Overview of Extracellular Matrix

The extracellular matrix (ECM) is a cellular network of proteins, polysaccharides and lipids in various combinations that provides physical support for cells. Depending on the types of resident cells, the ECM also carries out a number of specific functions in tissue morphogenesis, differentiation, and homeostasis (1). This is achieved by initiating ECM-originated signaling to neighboring cells either biochemically through chemical mediators or biomechanically through structural constrains (2).

The ECM is a highly dynamic structure that is constantly being remodeled. ECM components are regularly replaced and subject to numerous post-translational modifications such as calcification and cross-linking (3). Specific receptors and factors such as integrins, syndecans, and various growth factors are utilized by the ECM to mediate cellular activities including growth, adhesion, migration, proliferation, survival, and invasion (3-5). Through interactions with receptors and other factors, the ECM also exerts regulatory influence on various signal transduction pathways, which, in return, give rise to tissue-specific ECM morphologies and their accompanied mechanical properties such as rigidity and elasticity. Moreover, the ECM acts as a compression buffer to maintain a homeostatic environment for the proper functioning of cellular components and enzymatic activities (1).

The ECM is composed of two classes of molecules, fibrous and non-fibrous molecules (6,7). The fibrous molecules include mainly collagen, elastin, fibronectins, tenascin, and laminin (1). These proteins form the basic architecture of the ECM with other non-fibrous molecules such as proteoglycans (PGs) and glycoproteins filling the fibrous network. The long chains of PG-

associated glycosaminoglycans have the ability to attract water molecules to form gels, which maintain cell hydration and cushion cells against hydrostatic pressure. In summary, the ECM can be viewed as a meshwork of different fibrous proteins embedded in hydrogels of PGs. Such an organization allows the ECM to resist a wide range of tensile and compressive stresses. While all constituents play irreplaceable roles in the proper functioning of the ECM, three types of molecules: collagen, elastin, and glycosaminoglycans shall receive most attention in this thesis due to their substrate functions for lysosomal cysteine proteases (cathepsins).

Collagen

Collagen, which makes up 25-35% of the total protein weight in the human body, is the most abundant fibrous protein in the ECM (8). It confers the mechanical properties of stiffness and tenacity to the connective tissues. The basic unit of collagen, also known as tropocollagen, consists of three polypeptide chains (α-chains), which form a triple helical structure of approximately 300 nm in length and 1.5 Å in diameter. The amino acid sequence of collagen usually follows the pattern of G-P-Y or G-X-Hyp (hydroxyproline), where X and Y may represent any other amino acid residues. Although each α-chain is a left-handed helix with 3.3 amino acid residues per turn, they combine into a right-handed triple helix with 30 residues per turn. Such a ropelike arrangement guarantees the integrity of collagen and prevents the structure from unraveling (9,10). There are 28 different types of collagens present in the mammalian body, and five of them have a helical structure. The most common helical collagens are type I and II collagen, which are differentiated by their amino acid sequences and the heterogeneity of the α-chains. Type I collagen, found in nearly all connective tissues except hyaline cartilage, is comprised of two α1 chains and

one α2 chain. In contrast, type II collagen consists of three identical α1 chains and is present in cartilage, (11,12).

A number of collagens including types I, II, III, V and XI can further self-assemble with staggered ends into large arrays of fibrils (13). However, fibrillogensis requires more than just the self-assembly of collagen molecules, but also the involvement of multiple coordinated enzymatic processing and modification steps at different cellular compartments. In short, pro-α-chains are synthesized by cells such as fibroblasts, osteoblasts, and chondrocytes in the endoplasmic reticulum. Upon several posttranslational modifications including the conversion of proline residues into their trans-forms by peptidylproline cis-trans isomerase and into hydroxyproline by prolyl 4-hydroxylase, pro-α-chains trimerize in the presence of collagen heat shock protein 47 to form the triple helical procollagens and are secreted into the extracellular matrix (14-19). During the secretion process, the propeptides of procollagen are cleaved by procollagen N- and C-proteinases, and thereby allowing the spontaneous self-assembly of collagen into fibrils (20). The two propeptides, though removed during fibrillogensis, aid in the proper formation of collagen fibrils. Studies have elucidated that the C-propeptide is used to maintain the soluble state of procollagen molecules during intracellular trafficking whereas the N-propeptides affect the fibrillar shape and diameter (13). Lastly, to further stabilize the fibrillar structure, lysyl oxidase creates both intermolecular and intramolecular cross-linkages at lysine and hydroxylysine residues (13).

Type I collagen is the most abundant type of collagen. It accounts for 25% of the dry protein weight and comprises up to 90% of the organic matrix in mammalian bone. It is also widely distributed in many connective tissues such as skin, tendons, blood vessels, and the cornea (6). In terms of the molecular structure, type I collagen is a heterotrimeric helix that is morphologically

the same as other collagens. It contains four cross-linkages with two in the helical domain and one at each end of the telopeptides (21). It forms the characteristic axial periodic D-banding of cross-striated fibrils that provide support, tensile strength and rigidity to the tissues. Furthermore, it forms an anchor to many macromolecules including integrins, fibronectin, fibromodulin, and decorin, and also associates with various cells such as fibroblasts and macrophages (22-27). In bone and dentin, type I collagen is calcified with hydroxyapatite to enforce their hardness.

The importance of type I collagen is illustrated by several pathological syndromes and diseases. Mutations in type I collagen genes have been demonstrated to cause osteogenesis imperfecta (OI), Ehlers-Danlos syndrome (EDS), and infantile cortical hyperostosis (28-30). Patients with OI have brittle bones caused by an abnormal collagen structure and failures in post-translational modification, folding, intracellular trafficking, and matrix incorporation (31). EDS and infantile cortical hyperostosis are characterized by hypermobility of joints and a self-limiting inflammatory disorder of infants (29,32). In addition to these genetically inheritable syndromes, type I collagen also associates with osteoporosis, cardiovascular disorders and several types of fibrotic diseases (33-35).

Elastin

Elastin is another key structural fibrous protein found in the ECM of all mammalian species. As its name suggests, elastin grants elasticity and resilience to tissues like arteries, skin, ligaments, cartilage and tendons (36). It allows tissues to recoil after repetitive stretches. The extent of stretchiness is greatly defined by the arrangement of the elastin itself (37). Elastin is the insoluble product formed by cross-linking its soluble precursor, tropoelastin. Unlike different types of tropocollagens, which are encoded by multiple genes, only one gene is responsible for the 11

isoforms of tropoelastin (38). These variations are the results of alternative splicing. Tropoelastin are composed of two distinct domain types including repetitive hydrophobic motifs primarily consisting of glycine, alanine, proline, and valine residues, and the alternating lysine-rich hydrophilic regions (39). Both of these domain types are crucial to the formation of elastin. During elastogenesis, tropoelastin is coupled by the elastin binding protein (EBP) and exocytosed into the ECM. Upon release of the EBP, tropoelastin aggregates via coacervation, which is a process of aligning tropoelastin molecules into aggregates using interactions between their hydrophobic domains (40,41). The coacervation also prepares tropoelastin for cross-linkages (e.g. desmosine and isodesmosine) between two or multiple lysine residues in their hydrophilic domains, which eventually leads to the formation of the elastin meshwork and subsequently to elastic fibers. Elastic fibers consist of elastin as the major constituent and of microfibrils, which contain mainly fibrillins (42,43). The microfibrils typically have a diameter of 10 to 15 nm and are found to spread around the periphery of elastin (42). The arrangement of the elastic fibers can vary greatly between different tissues to deliver the desired mechanistic characteristics. For example, a sheet-like lamellae assembly in blood vessels is mostly responsible for the strength and elasticity necessary for vessel expansion and contraction, whereas a filament assembly in skin and lung is more suitable for flexibility and extensibility (44-46).

Being an important structural protein of the ECM, contacts between elastin and cells are inevitable. Research has demonstrated the involvement of elastin in multiple cellular activities including phenotype definition, adhesion, proliferation, and migration (47-49). Elastin achieves these biological roles by interacting with a number of cell receptors, with EBP being the best example. Besides facilitating the secretion, EBP binds to multiple sites on the tropoelastin, and initiates downstream intracellular signaling transduction pathways involved in cell morphogenesis,

proliferation, migration, and chemotaxis in a wide range of cells such as smooth muscle cells and fibroblasts (50-52).

A few genetic diseases have been linked to the elastin gene or genes that encode proteins associated with elastogenesis and elastic fiber formation. An example represents Cutis laxa, which is characterized by inelastic and loose skin and supravalvular aortic stenosis. The later leads to a narrowing of the aorta and insufficient blood flow (53,54). Furthermore, aging and injury can also cause damage and loss of elastin, which further lead to skin wrinkling, hypertension, pulmonary emphysema, and arterial aneurysms (55-57).

Glycosaminoglycans (GAGs)

GAGs are the sugar components of proteoglycans (PGs). PGs have been shown to have a wide variety of functions in tissue buffering, hydration, storage, adhesion, regulation, and force-resistance (58,59). They are composed of a core protein with covalently attached GAG-chains via a serine residue and tetrasaccharide bridge (60). Depending on the types of monosaccharides, GAGs can be further categorized as: hyaluronan (HA), chondroitin sulfate, dermatan sulfate (DS), heparan sulfate, and keratin sulfate (61). CS is the most common GAG and made of repeating disaccharide units that consist of two negatively charged monosaccharides termed D-glucuronic acid (GlcA) and N-acetyl-D-galactosamine (GalNAc) (61). The number and positions of sulfate groups further classify CS as CSA (C4-S), CSC (C6-S), CSD (C2,6-S) and CSE (C4,6-S). Note that DS, which was formally designated as CSB, is no longer classified as a form of CS due to the differences in its sugar composition.

The presence of sulfate groups enables CS to selectively interact with various molecules including growth factors, cytokines, chemokines, adhesion molecules, and lipoproteins (62). These

specific binding interactions reflect the role of CS in many cellular activities such as cell proliferation, adhesion, migration and apoptosis (62). CS functions as a storage reservoir for a diverse array of growth factors and regulators (63). In the central nervous system, CS is directly involved in stabilizing normal brain synapses as part of perineuronal nets (64). In the skeletal system, CS participates in bone resorption through the modulation of the protease activity of cathepsin K (65). Along this line, serglycin, which contains 94% C4-S polysaccharides, has been shown to inhibit bone formation (66). The importance of CS can also be derived from its affiliation with tumor progression. Being a growth factor storage reservoir entitles CS as an effective manipulator of tumor cells. Many studies correlate cancer development to the up or down regulation of CS expression. For example, high concentrations of C4.6-S are identified in the ECM of ovarian adenocarcinomas whereas no expression is observed in normal ovaries (67). Multiple myeloma cells secrete serglycin to inhibit the complement system to protect themselves from being attacked by the innate immune system (68).

The relationship between CS and tumor progression has been utilized by therapeutic approaches to potentially prevent cancer. In situations where CS production is down-regulated, direct uptake of modified CS-chains has been demonstrated to reduce or abolish breast tumor growth in murine models (69). In situations where CS production is up-regulated, CS has been incorporated into drug delivery strategies. This method allows drugs to be specifically delivered to tumor cells while CS is recruited to the cancer site (63,70). Moreover, as CS is the prominent GAG found in articular cartilage, it has been tested as a therapeutic agent for osteoarthritis (OA). OA is a chronic wear and tear disease characterized by progressive loss of proteoglycan and collagen contents and CS supplementation is thought to be beneficial. Although this has been

demonstrated in a number of small clinical studies, large scale clinical trials have reported little or no effect of supplemental CS on OA symptoms (71).

Overview of Extracellular Proteases

There are 816 known and putative proteases in the human genome, which play important roles in a multitude of biological processes (72,73). Based on their catalytic mechanisms, seven major protease classes have been established: serine, cysteine, aspartic, threonine, glutamic, asparagine, and metallo proteases (74). A small number of proteases remain unclassified due to their unknown structures and catalytic types (72). Among these seven classes, serine and metallo proteases are traditionally considered to encompass the majority of extracellular proteases. Extracellular proteases are proteolytic enzymes whose substrate targets are extracellular matrix components. They are ingredients of a fundamental cellular toolbox for affecting environmental changes. These enzymes enable cells to alter the microenvironment of tissues by directly cleaving structural macromolecules of the ECM, as well as catalyzing the processing of matrix protein precursors. Examples include plasminogen activator and type II transmembrane serine proteases (TTSPs). Matrix metalloproteases (MMPs) and ADAMs belong to metalloprotease class. In this thesis, I focus on a newly emerging class of extracellular proteases known as cathepsins, which mostly belong to the cysteine protease class.

Cathepsins: Human Lysosomal Cysteine Proteases

Cysteine proteases constitute approximately 26% of all human proteases (74). They are grouped into seven clans (superfamilies) namely Clans CA, CD, CE, CF, CH, and two mixed Clans

PA and PB, each of them further sub-divided into families based on amino acid sequence homologies and three-dimensional structural similarities (72).

Clan CA is the largest clan of cysteine proteases with 65 human members currently listed in the MEROPS database (72). It contains the families of papain (C1), calpain (C2), autophagin (C54), and ubiquitin-related peptidases (C12, C19 and C64) (72). The papain family (C1) of Clan CA is the most extensively studied family. They are widely expressed in animals, plants, bacteria, and viruses (75). In mammals, the papain-like lysosomal cysteine proteases are named cathepsins (75). It is important to discriminate these thiol-dependent lysosomal cathepsins from proteases of other classes. For example, cathepsins D and E are aspartyl proteases and cathepsin A and G are serine proteases (76,77).

According to the human genome database, there are eleven papain-like lysosomal cathepsins (cathepsins B, C, F, H, L, K, O, S, V, X and W) (77). Sequence analysis reveals that all these cathepsins contain three functional regions including a signal peptide, a propeptide, and a catalytic peptide. The signal peptide directs the translocation of the pre-proenzyme into the endoplasmic reticulum during ribosomal translation. It is on average 10 to 20 amino acids in length, and is cleaved during ER translocation, leaving the proenzyme for further transportation (75). The propeptide region is varied in length among different cathepsins reaching from 38 amino acid residues in cathepsin X to 251 in cathepsin F (75). Several functions for the propeptide region have been reported. For example, it assists in both the proper folding of the active enzyme and the transportation of the proenzyme to the lysosome (78-81). It also acts as endogenous inhibitor to prevent maturation of the proenzyme prior to its arrival at the lysosome (82-85). The catalytic domain constitutes the active form of the mature enzyme. Like many other cysteine proteases in family C1, human papain-like cathepsins are structurally very similar to papain regarding the 3D

fold and the positioning of the catalytic residues. The three-dimensional structures of most cathepsins including cathepsins B, C, F, H, L, K, S, V, and X have been determined (72). With the exception of cathepsin C which is a tetramer, all cathepsins are monomers with molecular weights between 23 to 25 kDa (86). Overall, mature cathepsins are bilobed, containing an α-helical L- and a β-barrel R-domain with the highly conserved active site located in the cleft (Figure 1.1). The two characteristic catalytic residues do not reside on the same domain, but with the cysteine on the L-domain and the histidine on the R-domain (75).

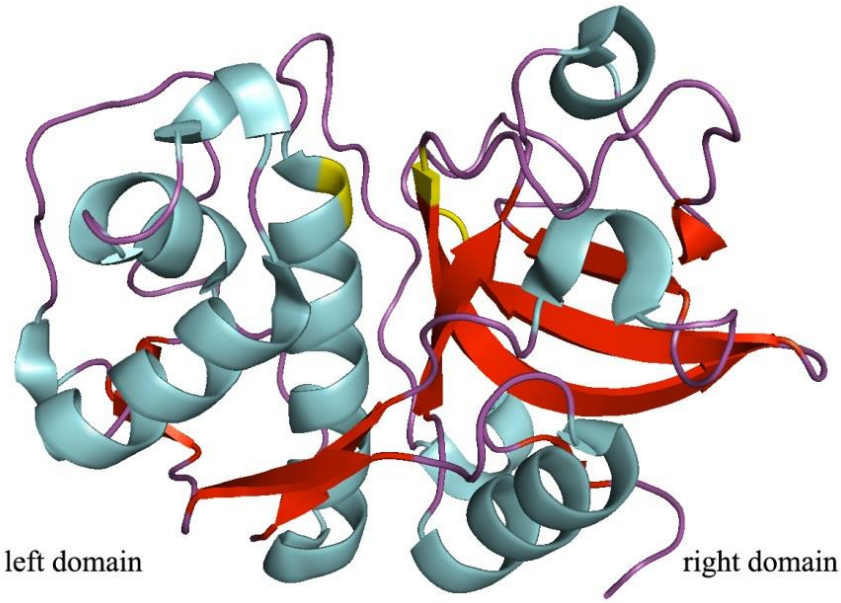


Figure 1.1. Crystal Structure of Cathepsin K. Helixes are colored in cyan, β-sheets are in red and active site residues (C_{25}, H_{162} and N_{182}) are in yellow. Crystal structure is obtained from protein data bank (1YKK).

The catalytic mechanism of papain-like cathepsins is identical to other cysteine proteases (Figure 1.2). In short, the cysteine and histidine dyad pre-form an ionic pair by transferring the

thiol proton of the cysteine residue to the histidine residue prior to the substrate binding. During catalysis, the deprotonated cysteine acts as a nucleophile attacking the carbonyl carbon of the scissile bond of the substrate and forms a tetrahedral oxyanion intermediate. This structure is usually stabilized by a glutamine residue via H-bond to form an oxyanion hole. Subsequently, the histidine donates its additional proton to the amine at the C-terminal portion of the substrate, assisting the transformation of the tetrahedral intermediate into an acyl enzyme by releasing the amine (75). In this process, an asparagine residue at the catalytic site is thought to orient the histidine during its acid-base catalysis, and thus is usually referred to as the third catalytic residue of papain-like proteases (72). To complete the substrate cleavage, a water molecule is required to release the enzyme from the N-terminus of the substrate. Again, this step is carried out via deprotonation of the water by histidine to create a stronger nucleophile.

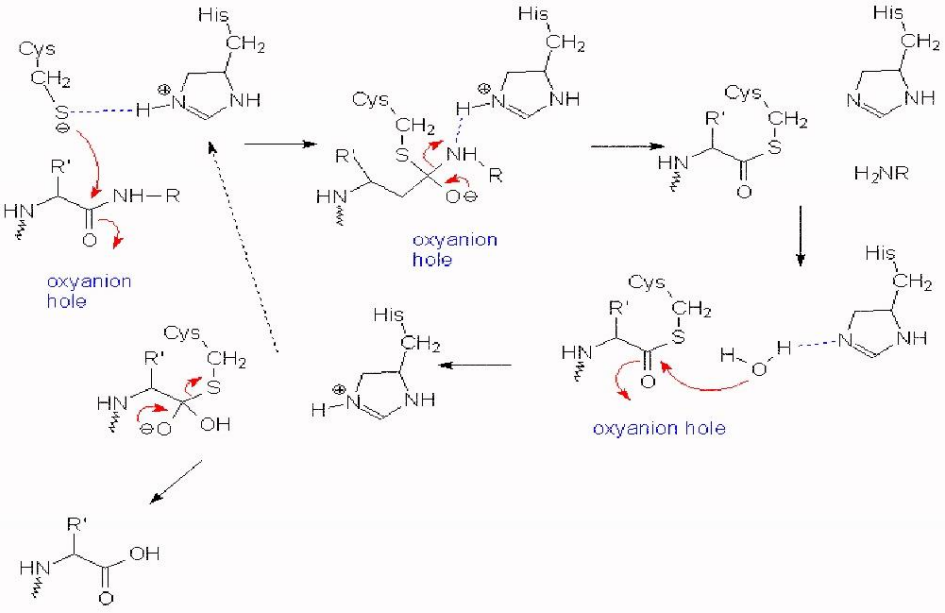


Figure 1.2. Mechanistic Reaction Scheme of Cathepsins. Picture is modified from online source http://chemistry.umeche.maine.edu/CHY431/Peptidase10.html (Department of Chemistry, University of Maine, Orono).

Although the majority of the lysosomal proteases are endopeptidases, papain-like cathepsins do not seem to strictly follow this rule (76). Among the eleven cathepsins, except cathepsin W with a yet unknown activity, there are 4 exopeptidases known as cathepsins B, C, H, and X, whereas the remaining proteases (cathepsins F, L, K, O, S and V) are endopeptidases (73). Moreover, the four exopeptidases can be further differentiated based on their substrate specificity. Cathepsin C is an aminodipeptidase; cathepsin H is an aminomonopeptidase; cathepsin B is a carboxydipetidase and cathepsin X is a carboxymonopeptidase (87). To achieve the different substrate specificities, cathepsins developed distinctive selectivity for substrates. For instance, in endopeptidases, the active site cleft extends through the interface of two domains, which provides a better accommodation for longer substrates, whereas in exopeptidases, additional structural features and specific electrostatic interactions are in place to reduce the number of substrate binding sites and to enable the proper docking of the substrate through their N- or C-termini (73,76).

The active site of cathepsins can be subdivided into seven subsites according to the Schechter and Berger model (Figure 1.3) (88). With four subsites (S1-S4) on the N-terminal side of the to-be-cleaved peptide sequence and three on the C-terminal side (S1´-S3´). Each subsite is sufficient to accommodate one amino acid residue termed P4 to P3´ residues to allow proper positioning of the substrates (88). The S2 subsite is in general the most selective site and is formed by a deep and well-defined pocket. Altogether, crystal structures of cathepsins with various substrate analogues and inhibitors have emphasized the substrate specificity sites, S2 and S1´ for substrate binding and the S1 binding site for facilitating catalysis (89-92). However, substrate specificity data obtained through the analysis of peptide libraries do not always reflect the cleavage specificity observed in protein substrates. Therefore it is difficult to define the substrate preference

for cathepsins only based on classical subsite interactions (73). It is highly likely that cathepsins bind to their substrates via additional multi-sited interactions, including exosites (92).

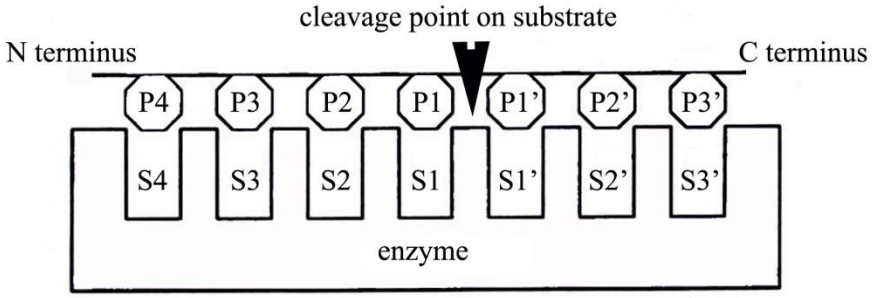


Figure 1.3. Schechter and Berger Subsite Schematic Diagram of Cathepsins. S stands for subsite and P stands for amino acid residue of peptide substrate.

Unlike MMPs which exhibit optimal activity at neutral pH, the participation of cathepsins in the ECM has to overcome two limiting factors: cathepsins are active at acidic pH (pH ~5) and require a non-oxidative microenvironment for optimal catalytic efficacy. While these two criteria contrast the conditions of the ECM, several mechanisms are discussed to make the involvement of cathepsins in ECM degradation possible. First, not all cathepsins require a strict acidic pH. Cathepsin S, which has been implicated in antigen presentation, inflammation, and tissue destruction, is efficiently active at neutral to slightly alkaline pH values (93,94). Second, the secretion of cathepsins may be coupled by the acidification of the peri- and extracellular space. Using bone resorption as an example, osteoclasts create a sealed and acidified space called resorption lacuna via H^+-ATPase pumps in order to demineralize the inorganic hydroxyapatite (95,96). This step sets the stage for the action of cathepsin K to break down collagen. Third, ECM components can be taken up by phagocytosis and intracellularly digested. Various cell types such

as macrophages, fibroblasts, endothelial, and epithelial cells have the ability to carry out endo/phagocytosis (97). With the help of phagosomes, ECM components are endocytosed and delivered to lysosomes where cathepsins are awaiting. In terms of maintaining a reduced cysteine residue for catalysis, cathepsins have been demonstrated to have a significant resistance towards oxidative stress while fulfilling their roles in the ECM (98,99).

According to Lecaille *et al*., some cathepsins exhibit tissue and cell-type selective expression (75). Whereas cathepsins B, L, and H are ubiquitously expressed, cathepsins S, F, and V are specifically expressed in macrophages, B-lymphocytes, dendritic cells, microglia and/or thymic cortical epithelial cells; cathepsin K is highly expressed in osteoclasts, macrophages, and fibroblasts, whereas cathepsin W is predominantly expressed in CD8 and natural killer cells (97,100,101). These expression pattern give rise to asymmetric contributions of cathepsins to ECM degradation in different tissues. In bone and cartilage, cathepsin K is the dominant collagenase. Cathepsins K, B, and L are also capable of cleaving proteoglycans such as aggrecan and link protein (102). In blood vessels, the involvement of cathepsins is reflected in the up-regulation of cathepsins K, L, and S in smooth muscle cells, cathepsins B, K, S, and V in macrophages, and cathepsins K and S in epithelial cells in cardiovascular diseases (103). In lung, inhibition of cathepsins B, H, and L decreases emphysema and inflammation (104). It should also be noted that cathepsins and MMPs have overlapping roles in ECM degradation in many tissues, which signifies potential cooperative degradation modes. For example, while MMPs are important in cancer progression, cathepsins B, L, H, X, S, and K also play significant roles in tumor metastasis (105-107).

In the following sections, cathepsins V and K will be reviewed in more details as they are the main subjects of this thesis.

Cathepsin V

Human cathepsin V, also known as cathepsin L2, is located on chromosome 9 at position 9q22, which is adjacent to the gene encoding cathepsin L (108). As their adjacent gene locations suggest, cathepsins V and L likely evolved from a common ancestral gene. This is also supported by the fact that cathepsins V and L share a 78% identity in their amino acid sequences (109,110). The full-length cDNA of human cathepsin V encodes a 334 amino acid long peptide with an apparent molecular weight of 37,329 Da. It contains a 17 amino acid-long signal peptide, a 96 amino acid-long propeptide, and a 221 amino acid-long mature enzyme. The active site is formed by the conserved C_{25}, H_{163} and N_{188} residues and two N-glycosylation sites at N_{108}-E-T and N_{179}-N-S are present in the mature peptide (109). Both glycosylation sites have been demonstrated to be important for the intracellular and intercellular trafficking of the enzyme (111).

In contrast to ubiquitously expressed cathepsin L, cathepsin V is specifically expressed in thymus, testis, cornea, epidermis, brain cortex and hippocampus (108,109,112,113). Studies aiming to explore its roles in these tissues have indicated several functions: 1) cathepsin V is the major protease expressed in human corneal epithelium and may be involved in controlling angiogenesis by releasing endostatin from collagen XVIII in human libocorneal epithelial cells (112,114). 2) expression of human cathepsin V in mouse thymus can mediate the positive selection of T helper cells in cathepsin L knock-out mice (115); 3) its involvement in converting the MHC class II-associated invariant chain, li, into CLIP suggests a role in regulating antigen presentation in the human thymus (116). Moreover, Fukelstein *et al*. have recently illustrated the participation of cathepsin V in the production of enkephalin which is a neuropeptide regulating analgesia, stress and brain behavior, and NPY, which is a neurotransmitter responsible for feeding behavior and thus obesity in the brain, and blood pressure regulation in the peripheral sympathetic nervous

system (113,117-121). Under pathological conditions, the expression of cathepsin V is up-regulated in both colon cancer and atherosclerosis, which can serve as a potential biological marker (103,108).

Mature cathepsin V consists of two domains, L- and R-domain, encompassing a V-shaped active cleft between these domains. The R-domain contains a twisted β-sheet whereas the L-domain is comprised of three α-helices (122). A comparison to other cathepsins reveals a highly conserved overall fold with one noticeable variation in the surface loop formed by residues G_{171} to $S_{180\,(122)}$. Interestingly, this region has also been reported to show a high degree of dissimilarities in both the number and kind of the amino acid residues between many papain-like cysteine proteases (123). Despite of the structural similarity, cathepsin V processes its own unique substrate specificity based on the configuration of its substrate-binding subsites. According to Somoza *et al*., the S1 subsite, formed by N_{66}, G_{23}, G_{67} and the carbonyl oxygen of C_{65}, is shallow and thus reveals little substrate specificity. In contrast, the S2 subsite, residing in the R-domain, is the principle specificity determinant. It is well defined by A_{135}, G_{164}, and the side chain of M_{70} at the bottom of the pocket, and K_{159}, N_{160}, L_{161}, F_{69}, the main chain of D_{162}, the carbonyl oxygen of G_{68} and the methyl group of A_{214} on the side walls of the pocket (122). Most attention falls on residues F_{69} and A_{214}, which play determining roles in defining the width and length of the S2 subsite pocket respectively. Particularly, the identity and position of the side chain of residue 69 limits the size of a potential substrate residue at P2. For example, cathepsin V has a phenylalanine, whereas cathepsin K has a leucine at this position. At the P2 position, cathepsin V exhibits similar to cathepsin L a preference for aromatic residues, whereas cathepsin K favors aliphatic amino acids such as leucine and methionine in its S2 subsite (124). The S3 subsite is also varied among cathepsins. Its shape greatly depends on the side chains of residues at position 63, 69 and 72. In

addition, the number of extra residues adjacent to G_{61} (compared to papain) matters in defining the depths of S3 pocket. Finally, the S1' binding pocket is conserved among most of the proteases in the C1 family, probably due to the fact that residues forming this pocket are also involved in the catalytic machinery (122). The analysis of the subsite structures coincides with the results of substrate specificity profiling. A library of 160,000 tetrapeptide substrate completely randomized at each of the P1, P2, P3, and P4 positions with 20 amino acids was used to probe the specificities of six cathepsins L, V, K, S, F, and B. The results of the substrate cleavage analysis revealed a similar P1 preference pattern suggesting a low selectivity at the S1 subsites. For the P2 residue, cathepsin V showed similar preferences as cathepsin L, favoring aromatic residues such as tryptophan and tyrosine over aliphatic ones such as leucine and valine (124). In contrast, cathepsin K prefers proline and aliphatic amino acids in this position. The poor adaptation of proline at the S2 subsite of cathepsin V may also explain its inability to breakdown proline-rich collagen, although it is able to form weak complexes with GAGs similar to cathepsin K (125). At the P3 position, cathepsin V showed broad specificity much like cathepsin L.

Cathepsin V is a highly potent elastase in comparison to other known elastases including cathepsin K, cathepsin S, pancreatic elastase, leukocyte elastase, and MMP-12 (125). It contributes to ECM elastin degradation via intracellular endocytosis. In atheroma tissue and monocyte-derived macrophages, cathepsin V is responsible for approximately 20% total elastolytic activity, and is regulated by the abundance of GAGs, which specifically inhibit its elastase activity (125).

Cathepsin K

Human cathepsin K is encoded by a gene located on chromosome 1 at position 1q21 with an estimated length of 12.1kb (126). DNA sequence analysis revealed that it contains eight coding

exons that are interrupted by seven introns. Its general genomic organization is similar to that of cathepsin S (89,126). The promoter region of cathepsin K contains two consensus Sp1 binding sites, a G/C rich motif and a NF-κB binding site without the presence of a TATA box (89,126,127). Several regulatory factors such as RANKL, tumor necrosis factor-α and interleukin-1 have been demonstrated to up-regulate cathepsin K gene expression, whereas estrogen, interleukin-10 and transforming growth factor-β1 have the opposite effect during osteoclast-mediated bone resorption (104,128-130). When compared with ubiquitously expressed cathepsins B, L and H, the tissue/cell distribution of cathepsin K is relatively limited and specific. Its mRNA has been detected in tissues like bone, ovary, heart, placenta, lung, skeletal muscle, colon, and small intestine (131). In terms of cell distribution, high levels of cathepsin K expression have been detected in osteoclasts and related multinucleated cells, smooth muscle cells, synovial fibroblasts, macrophages and epithelioid cells in organs such as lung and thyroid glands (96,130-134).

The crystal structure of cathepsin K is similar to other papain-like cysteine proteases. The preproenzyme contains 329 amino acid residues with 15 in the signal sequence, 99 in the propeptide, and 215 in the mature peptide. The mature enzyme is a monomeric protein with a molecular weight of 23,495 Da containing one potential N-linked glycosylation site. A distinct characteristic of cathepsin K is a high density of positively charged amino acid residues located on the surface opposite to the active cleft (89). These positive residues have been demonstrated to interact with extracellular matrix-resident GAGs, such as C4-S via ionic interactions, which play an important role in the physiological functions of cathepsin K (135). As a potent collagenase, cathepsin K exhibits certain unique substrate specificities. It prefers aliphatic substrates containing leucine, isoleucine, and valine residues over those containing hydrophobic residues, which are preferred by cathepsins V and L (124). The S2 binding pocket of cathepsin K accommodates very

well a proline residue (89). Its S3 subsite also favorably binds glycine. These features combined empower cathepsin K for the efficient cleavage of peptide sequences present in collagens (such as the repetitive G-P-Y sequence).

Bone remodeling is a dynamic, lifelong process in which old bones are replaced by new ones. This process consists of two distinct stages called resorption and formation, and each stage involves a specific type of cells termed osteoclasts and osteoblasts, respectively. Osteoclasts are responsible for the bone resorption, while bone formation is carried out by osteoblasts. The resorption process can be further divided into two actions: 1) the solubilization of inorganic minerals by acid released from the osteoclasts, and 2) the proteolytic digestion of the organic matrix which contains predominantly fibrillar type I collagen (136). Cathepsin K, which represents 98% of the total cysteine proteases expressed in osteoclasts is critical for the resorption of the organic bone matrix (89,131). Its role in bone resorption is irreplaceable by other cathepsins and collagenases of the MMP family. For example, none of the other cathepsins exhibit any triple-helical collagenase activity and MMP-1, -8 and -13 with their hemopexin domain, are constrained to cleave collagen at one position to create the C-terminal and N-terminal fragments (137). On the contrary, cathepsin K, by forming an oligomeric complex with GAGs such as C4-S, can efficiently break down collagen at multiple sites including at both the telopeptides and within the triple helical domain (138,139).

Abnormal cathepsin K activity has been linked to various ECM related diseases. On one hand, excessive cathepsin K activity can lead to bone disorders like osteoporosis and Paget's disease, which are characterized by global and localized loss of bone mineral density, respectively (140,141). Inhibition of cathepsin K has been demonstrated to significantly reduce serum markers of osteoporosis such as the N-terminal telopeptide and C-terminal telopeptide of collagen

fragments in an estrogen-deficient monkey model (142). On the other hand, lack of cathepsin K expression can cause a severe bone resorption defect known as pycnodysostosis. Pycnodysostosis is a rare genetic bone disease resulting from non-sense, miss-sense and stop codon mutations in the cathepsin K encoding gene (133,143-145). Patients with this disease have short stature, dense but brittle bones, short fingers, wide open soft spot on the skull and several other abnormalities involving face, teeth, collar bones, skin, and nails (146). Cathepsin K has also been implicated in rheumatoid arthritis and osteoarthritis by breaking down aggrecans and fibrillar type II collagen, which are the major constituents of the ECM in articular cartilage (89). Besides being a potent collagenase, cathepsin K is also an effective elastase. Its unregulated expression has been linked to pathological elastin and collagen turnover contributing to the phenotypes of atherosclerosis and lung fibrosis (89,147).

There are a number of therapeutic treatments currently available for osteoporosis, such as bisphosphonates, and estrogen hormone replacement therapy. However, side effects including the induction of osteoclast apoptosis, jaw osteonecrosis, tissue necrosis, increasing rate of breast cancer, heart attacks, stroke and blot clot formation demand the development of more specific and safer drugs. (148-151). The discovery of cathepsin K provides the unique possibility to inhibit the single most important protease in bone resorption without inferring with more complex regulatory mechanisms. As a result, cathepsin K has received considerable attention from both academic and industrial researchers. Earlier developed irreversible inhibitors of cathepsins such as E-64 and related expoxysuccinyl derivatives, vinyl sulfones and haloketones tended to confer antigenic and immunologic complications if used chronically (75,152,153). To eliminate these side effects, subsequent drug development focused on reversible inhibitors containing amides, aliphatic ketones, nitriles, aldehydes and α-keto hetero-cycles as so-called warheads (154). At least five

compounds have advanced into clinical trials for the treatment of osteoporosis, and may potentially be used for other cathepsin K related diseases including osteoarthritis, atherosclerosis and metastatic bone diseases. Nonetheless, while all of them revealed effective outcomes in reducing biological markers of bone resorption and increasing bone mineral density in either animal models or in early stage clinical trials, at least two had caused adverse effects. For example, balicatib developed by Novartis failed in phase II trials due to various skin conditions including skin lesions, skin rashes, and rare morphea-like skin changes (155,156). Its lysosomotropic character leading to an accumulation of the drug in lysosomes and nonselective off-target effects have been implicated to be responsible for the failure of balicatib (157). In contrast, odanacatib developed by Merck Frosst is a successful example. Phase III clinical trials were completed ahead of time as a result of its high efficacy and good safety profile. It demonstrated a dosage-dependent reduction of bone resorption in postmenopausal women (158). One major advantage of odanacatib might have contributed to its success. As a non-basic compound, it maintains its potency and specificity without being trapped in lysosomes where it may inhibit other cathepsins (159).

Mechanistic Insights into the Mechanism of Elastin and Collagen Degradation

The properties of the ECM are partially maintained and controlled by the activities of elastases and collagenases. As their names suggested, elastases are responsible for breaking down fibrous elastin and tropoelastins, whereas collagenases are capable of digesting collagen fibers and tropocollagens. The following sub-sections aim to provide an overview of proteolytic mechanisms of selected elastases and collagenases.

Elastase Mechanism

Elastases mostly belong to serine proteinases (e.g., human leukocyte elastase, proteinase 3, and cathepsin G), cysteine proteinases (e.g., cathepsins V, K and S) and MMPs (e.g., MMP-12). Proteolytic degradation by elastases, also known as elastolysis, can lead to loss of tissue elasticity, which is associated with multiple pathological conditions including pulmonary emphysema, cystic fibrosis, infections, inflammation, and atherosclerosis (160). For instance, increased levels of serine neutrophil elastase and MMP-12 have been found in bronchoalveolar lavage fluid of chronic obstructive pulmonary disease patients (161-164); Up-regulation of cathepsins K and S in macrophages and smooth muscle cells, leading to excess degradation of elastin and other extracellular components, has also been identified in human atheromas (130).

The physiological relevance of elastases demands the understanding of the mechanism of elastin degradation for the development of inhibitors and thus for the development of effective therapeutic treatments. Different classes of elastases utilize distinct mechanisms to do so. For example, for the serine protease, human leukocyte elastase (HLE), a time dependent elastolytic progress consisting of an initial rapid elastase-elastin complex formation, an isomerization of the complex into a tightly bound complex, which induces local structural changes in elastin for catalysis, and a desorption process to allow elastase moving on the elastin fibrils is discussed (165). Two potential binding modes of the initial elastase-elastin complex formation are further postulated by Lonky and Wohl including a strong but non-catalytic binding via electrostatic interactions, and a weaker but catalytically-effective binding via hydrophobic interactions on the alanine-rich cross-linked elastin surface (166). The latter binding mode gives rise to the preference of HLE to cleave after small hydrophobic and aliphatic amino acids such as alanine, glycine and valine (167). Of two other serine elastases expressed by neutrophils, cathepsin G exhibits a similar

mechanism to HLE, whereas the action of proteinase 3 shows a closer resemblance to those of cysteine cathepsins (168,169).

A study by Novinec *et al*. has pointed out a similar mechanism of cathepsins to HLE, except a difference that cathepsins lack the formation of the tightly bound enzyme-substrate complex. The absence of the tightly bound complex is supported by the finding that specific macromolecular inhibitors such as stefin A and kininogen domain 3 are able to abolish the elastolytic activity of cathepsins by gaining access to the active sites (168). Based on these results, they also suggest a different mechanism: a non-catalytic adsorption onto the elastin surface followed by the formation of a catalytically competent structure to orientate the active site towards the cleavable elastin peptide bond for catalysis; upon cleavage, cathepsins can either desorb to find a distant site, or can remain adsorbed to initiate a new catalytic cycle (168).

The mechanisms of elastin degradation of both serine and cysteine elastases imply the utilization of certain surface structures in addition to the active sites in facilitating substrate binding. However, no clear elastin binding exosites have been identified for both classes of enzymes. In contrast, via NMR, bioinformatics and double mutation cycle analysis, two elastin binding exosites have been recently identified in MMP-12 (170,171). Exosite 1 comprises of mainly amino acid residues D_{124}, E_{199}, E_{201}, F_{202}, H_{206} and T_{205} from the V-P loop. Exosite 2 centers on M_{156} of the II-III loop and R_{117} of β-strand I near the back of the catalytic domain (170). Mutations created at specific amino acid residues at both exosites caused decreases of elastolysis to various extents (170). Moreover, both exosites contribute independently and partially to the binding of elastin. Although it appears that interactions conferred by each exosite are moderate, the additive effect plays a significant role in the docking and positioning of the insoluble elastin to the enzyme active site (170).

Collagenase Mechanism

Mammalian collagenases limited to two classes of proteinases including MMPs (e.g., MMP-1, MMP-8 and MMP-13) and cysteine proteases (e.g., cathepsin K). They are capable of breaking down collagen molecule within its triple helical domain, which are resistant to most proteases. Current understanding of collagenases has linked their functions with various tissue remodeling activities during morphogenesis, angiogenesis, wound healing, and bone remodeling. For example, the products of collagenolysis have been elucidated to regulate the migration of epithelial cells during wound healing and to activate/deactivate osteoclasts during bone remodeling (165,172). On the other hand, abnormal collagenase activities have been associated with a number of diseases such as osteoporosis, arthritis, atherosclerosis, lung fibrosis, and tumor development and metastasis (103,173-175).

An interesting question arises when studying the mechanism of collagen degradation by both MMPs and cysteine cathepsins: How are these proteases displaying a rather narrow active site cleft of approximately 5 Å able to cleave even just the base unit of fibrillar collagens, tropocollagen, which has a diameter of 15 Å. Recent studies on MMPs indicate that the presence of their hemopexin domain (HPX) allows MMP collagenases to stabilize and partially unwind the triple helix prior to the hydrolysis of the peptide bond (Q/L)-G#(I/L)-(A/L) (# indicates the cleaving point) at about ¾ away from the N-terminus of the collagen by the catalytic domain (CAT) (137). Using MMP-1 as the model enzyme and a number of synthesized triple helical peptide (THP) substrates, a detailed mechanistic action of MMPs has been dissected by Fields and coworkers: the initial contact occurs between the HPX domain of the MMP-1 (via residues 285-295, 302-316 and 437-457) and the THP (via analog amino acid residues 782-785 of the type I collagen). This interaction allows the presentation of the THP substrate to the CAT domain. No catalysis occurs

at this stage because the THP is still morphologically intact. In order to induce local denaturation, a back-rotation of the CAT domain takes place to release the α2- chain analog of THP, and exposes it to the CAT domain active site for subsequent hydrolysis (176,177). Interestingly, besides the HPX and the CAT domains, MMPs also utilize one of its elastin binding exosite (amino acid residues 183-191) to stabilize and position the collagen helix (178,179). This overall process is energetically favorable and provides a new approach to inhibit MMP collagenolysis by using short uncleavable peptides as competitive inhibitors for the natural substrate (176).

Whereas most cathepsins are collagenolytically inactive, cathepsin K which is predominately expressed in osteoclasts, has been recognized to be one of the most potent mammalian collagenases thus far. Its potency is evident by its proteolytic efficiency and its ability to cleave collagen at multiple sites within both telopeptides and triple helical domains (138,139). In contrast to the MMPs, cathepsin K does not possess the HPX domain or other similar structures, which implies the utilization of a distinct collagenolytic mechanism. In fact, the collagenolysis of cathepsin K is accomplished by the formation of oligomeric complexes with ECM-resident GAGs (180). Cathepsin K alone, like most cathepsins, can only cleave collagen at its telopeptides regions (65). Addition of GAGs allows this enzyme to unfold triple helical collagen as does the HPX domain in MMPs, and turns it into a powerful collagenolytically active machinery. Nonetheless, such potent activity of cathepsin K is also limited by GAGs. Excess amounts of GAGs become inhibitory, implicating a potential regulatory role of GAGs under physiological conditions. The crystal structure of cathepsin K in the presence of C4-S has been recently solved revealing a "beads-on-a-string"-like conformation, which consists of multiple cathepsin K molecules binding to one chain of GAG (135). The complex is held together by charge interactions between the positively charged cathepsin K surface and the negatively charged GAG. The binding of GAG

chain does not interference with the proper enzyme active site function as the complex still exhibits similar catalytic efficiencies as monomeric cathepsin K against gelatin and the synthetic peptide substrate, Z-FR-MCA (65,180). Although this oligomeric structure is collagenolytically inactive, it provides significant insight into the understanding of collagen degradation by cathepsin K/GAG complexes and formed the basis of my studies outlined in Chapter 3.

Hypotheses and Specific Objectives

Chapter 2

Elastin is a major ECM structural protein that is responsible for the elasticity of tissues such as skin, blood vessels, and lungs (125,181). Its resistance to proteolytic degradation greatly depends on numerous cross-links between its tropoelastin monomers to form an insoluble elastin meshwork that allows the connective tissues and organs to stretch and recoil back into their original states (182). Degradation of elastin by elastases due to an uncontrolled proteolytic degradation or its age-related depletion can lead to various symptoms and diseases from wrinkled skin to more severe conditions such rupture of blood vessels and pulmonary emphysema. Cathepsins V, K and S have been demonstrated to have potent elastolytic abilities and have been implicated in a number of elastin-associated diseases including aneurisms and neovascularization (183). However, the mechanisms employed by these elastolytic cathepsins have yet to be elucidated. Therefore, using cathepsin V as the model enzyme, one objective of this thesis was to explore the mechanism that enhances the elastolytic activity of selected cathepsins.

Cathepsin V, which is identified as the most potent mammalian elastase thus far, shares a 78% sequence identity with cathepsin L, which exhibits only minimal elastolytic capability (Figure

1.4). Active site substrate profiling using a diversified combinatorial peptide library revealed a highly similar preference between these two proteases (124). This suggests that the remaining differences in amino acid sequence may contain elastin binding exosites much like MMP-12.

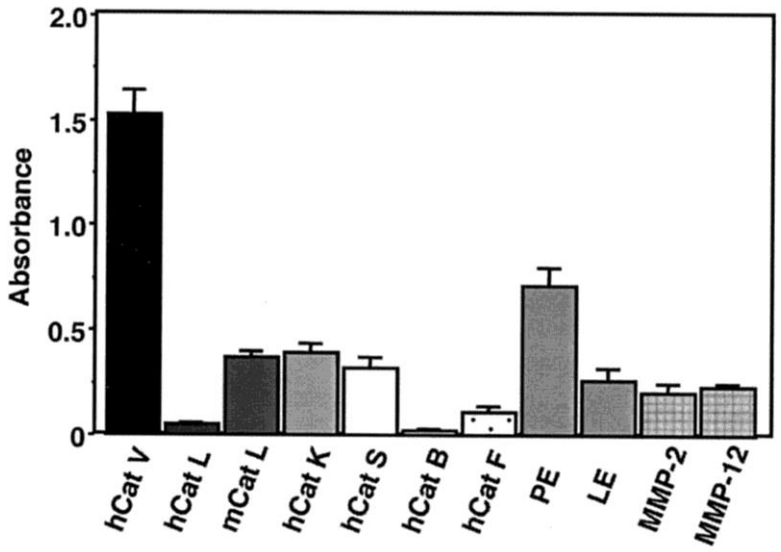


Figure 1.4. Comparison of Elastolytic Activity of Cathepsin V to Other Known Elastases. (Yasuda, Y., Li, Z., Greenbaum, D., Bogyo, M., Weber, E., and Bromme, D. (2004) The Journal of Biological Chemistry 279, 36761-36770)

Hypothesis

Cathepsin V utilizes specific exosites to bind to elastin, which facilitate its hydrolysis. The following objectives describe the experimental approach addressing this hypothesis.

Objectives

To create a series of chimeras by sequentially replacing increasing parts of cathepsin V with analogous sequences from cathepsin L. To compare the kinetic parameters (k_{cat}, K_m and

k_{cat}/K_m), the elastin-degradation-HPLC profile, and elastin-rhodamine degradation efficacy between the different chimeras and their wild-type parent enzymes to allow the localization of elastin binding exosites. To further characterize the involvement of exosites in cathepsin V-mediated elastolysis, chimeras are further characterized by substrate profiling analysis, scanning electron microscopy to study the morphological change of elastin with and without the treatment of enzymes, and protein absorption assays to verify the physical contacts between the exosites of cathepsin V and elastin.

Chapter 3

Collagen provides the main architectural and mechanical support to mammalian connective tissues. Its function is based on the formation of collagen fibrils, which are highly resistant to proteolysis. Only a few collagenases such as MMPs-1, -8, -13, and cathepsin K are capable of efficiently disassembling the collagen fibrils by digesting the tropocollagen within its triple helical domain (184). Cathepsin K is the only mammalian collagenase, which is capable to cleave tropocollagen at multiple sites. As the 3D structure of cathepsin K is very similar to those of other non-collagenolytic cathepsins and as it does not contain special structural features such as a collagen-unwinding hemopexin domain present in MMPs (137), cathepsin K must exploit a different mechanism to cleave fibrillar collagens. As described above, cathepsin K forms oligomeric complexes with glycosaminoglycans, which are required for its collagenase activity. Its monomeric form is unable to cleave collagen (180). Although a crystal structure of cathepsin K/C4-S complex was resolved (Figure 1.5) (135), the structure did not reveal the mechanism of its action towards triple helical collagens. The crystal structure showed a beads-on-a-string-like organization of multiple cathepsin K molecules attached to a single glycosaminoglycan. However,

our preliminary results indicated that an excess of cathepsin K over GAGs does not allow for a collagen cleavage. On the other hand an excess of GAGs over cathepsin K also inhibits collagen degradation (165). Only a narrow molar ratio of about 2:1 (cathepsin K:GAG) resulted in a potent collagen activity. This implies that highly specific cathepsin K/GAG complex must exist which is responsible for the collagenase activity.

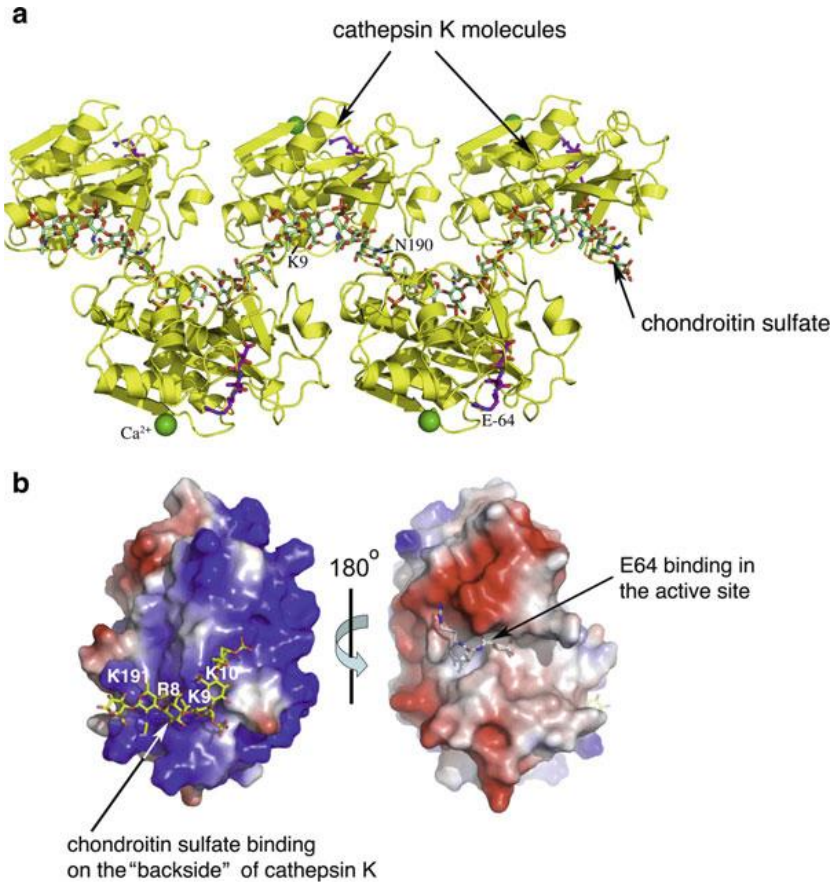


Figure 1.5. Interactions of Cathepsin K with C4-S. (a) "Beads-on-a-string"-like organization of multiple cathepsin K on a single chain of C4-S. (b) Surface electrostatic distribution of cathepsin K with C4-S binding to the positive (blue) "back" side. K_8, K_9, K_{10} and K_{191} play critical roles in binding to C4-S. (Brömme, D., and Wilson, S. (2011) Role of Cysteine Cathepsins in Extracellular Proteolysis. Springer Berlin Heidelberg. pp 23-51).

Hypotheses

There exist specific cathepsin K/GAG complexes responsible for collagenolysis. These complexes have a distinct molar ratio between cathepsin K and GAG molecules and are defined by specific GAG-protein and protein-protein-interaction sites. These complexes are not required for the non-collagenase activities of the protease.

Objective 1

Based on the two available crystal structures of cathepsin K/GAG complexes, putative cathepsin K/GAG complex models adhering to a molar 2:1 ratio were constructed. Each model, although with different cathepsin K and GAG contacting modes, shared a number of critical amino acid residues in their organizations.

Objective 2

Gradient GAGs collagenase assay with a fixed amount of cathepsin K in the presence of various concentrations of GAGs were used to determine the optimal ratio that would give rise to the highest collagenolytic efficacy. To verify the two putative models, mutations aiming to disrupt the complex formation were generated. Their collagenolytic and gelatinolytic activities were evaluated to determine their contributions in the complex. Our findings revealed that mutations on the common residues of the two models exhibited various levels of reduction in the collagenase activity. Model-specific mutations at various protein-protein interactive interfaces were also inhibitory. The results pointed to the possible existence of both complex structures *in vivo*.

Objective 3

Atomic force microscopy was utilized to identify the existence and morphologies of both complexes, and to study their interactions with tropocollagen substrate. It was found that C4-S and cathepsin K, at 2:1 molar ratio, formed ring-shaped tetrameric-like structures and preferentially sat on the helical domain of tropocollagen coinciding with the observation that cathepsin K/C4-S complex cleaved tropocollagen at multiple sites.

Objective 4

Circular dichroism spectroscopy was chosen to determine whether cathepsin K/GAG complex has a similar collagen "unwinding" ability like the hemopexin domain of the MMPs. Our results indicated that the cathepsin K/C4-S complex at 2:1 ratio exhibited detectable alternations on the CD spectrum of tropocollagen.

Chapter 4

Cathepsin K inhibitors such as odanacatib and relacatib, which were developed for the treatment of osteoporosis, are reversible active-site directed compounds. By blocking the active site of cathepsin K, they showed significant reduction of osteoporotic symptoms without affecting the quantity and functionality of the osteoclast. However, this direct blockage of the active site may show problems as it has been realized that cathepsin K also exerts non-collagenolytic functions in various cell types including synovial fibroblasts, macrophages, skin fibroblasts, dendritic cells, chondrocytes, melanocytes and epithelial cells. For example, cathepsin K expressed in mouse thyroid epithelium is involved in thyroid-globulin processing, and its inhibition may induce structural and metabolic changes in the central nervous system that are associated with

learning and memory deficits (185). The involvement of cathepsin K in cytokine secretion in bone marrow dendritic cells and in bone osteolysis also implicates its key role in autoimmune and inflammatory diseases (186-188). While many of the non-collagenolytic functions of cathepsin K are still in an early stage of characterization, they shall certainly be taken into serious consideration during drug design. Therefore, in chapter 4, based on the results obtained in chapter 3, we hypothesized that selective exosite inhibitors would only inhibit the therapeutically relevant collagenolytic activity of the protease by interfering with cathepsin K/GAG complex formation without affecting the active site functionality.

Hypothesis

There are selective exosite inhibitors which prevent the collagenolytic complex formation between cathepsin K and C4-S without affecting the active site functionality. To address this hypothesis, three specific objectives were pursued as described below.

Objective 1

A high throughput fluorescence assay was developed to screen for potential compounds that block cathepsin K/C4-S complex formation without inferring significant active site inhibition at excess concentration. 15 hits were identified among a total of 1280 known drugs. These hits were evaluated in secondary assays using *in vitro* collagenase and gelatinase assays, as well as synthetic peptide assays to test the active site and exosite activities of identified drug candidates.

Objective 2

Compounds with the strongest inhibition on the collagenolytic activity of cathepsin K were subjected to atomic force microscopy (AFM) study to determine their extents of inhibition. Our results demonstrated that suramin, aurintricarboxylic acid, dihydrotanshinone I, (-)-epigallocatechin gallate and clomiphene was able to disrupt the ring-shaped formation of cathepsin K/C4-S complex. In addition, scanning electron microscopy (SEM) revealed that dihydrotanshinone I and tamoxifen could inhibit collagen fiber degradation.

Objective 3

Molecular docking program was used to reveal potential interaction sites on cathepsin K surface with two of the most effective inhibitors. Our results revealed that dihydrotanshinone I bound remotely from the active site at one of the protein-protein interaction interface of the cathepsin K/C4-S complex, and aurintricarboxylic acid docked into a region that competitively impedes the binding of C4-S to cathepsin K. Both inhibitors, thereby, could prevent the formation of a collagenolytically functional complex.

Chapter 2. Identification of Exosites in Cathepsin V Necessary for Elastin Degradation in Comparison to Cathepsin L

Introduction

Elastin is a major structural ECM protein that provides elasticity and tensile strength to tissues such as skin, blood vessels, and lungs (181). It consists of numerous cross-linked chains of tropoelastin with an α-helical conformation (189). Cross-linking and hydrophobicity of elastin monomers result in the formation of an insoluble elastin meshwork that allows connective tissues and organs to stretch and recoil back to their original state (182). Due to the lack of *de novo* synthesis of elastin after reaching adulthood, its degradation during aging is irreversible and can lead to various pathological symptoms and diseases (190). This comprises wrinkled skin, the weakening and rupture of blood vessels, and the destruction of lung alveoli in lung emphysema (191-193).

However, elastin is highly resistant to proteolytic degradation and only very few selected serine, metallo, and cysteine proteases are capable of hydrolyzing insoluble elastin (77,194-197). The commonality of these proteases is that they prefer the binding of small hydrophobic amino acids such as alanine and valine in their substrate binding sites. When compared to known mammalian elastases, such as serine elastases, MMPs, and several lysosomal cysteine proteases (cathepsin S and K), cathepsin V revealed an exceptionally high elastolytic activity (125,195,198). Due to the insoluble nature of elastin, elastolytic activity of an enzyme is frequently measured as the release of soluble fragments from the insoluble elastin. These fragments are typically attached with a chromophore, fluorophore, or radioactive label for easy detection. Interestingly, human cathepsin L, which shares a 78% amino acid sequence identity and a very similar 3D structure with cathepsin V (109,122), exhibits only minimal cleavage activity towards insoluble elastin even

though its specific peptidolytic efficacy (k_{cat}/K_m) is higher than that of cathepsin V (109,125). The difference in their elastase activities appears unrelated to the nature of the classical subsite-defined binding areas in both proteases. Substrate profiling using a diversified combinatorial peptide library revealed similar subsite specificities between both proteases (124).

A recent study has suggested a two-step mechanism for elastin degradation by cathepsins: a non-catalytic adsorption onto the elastin surface followed by the formation of a catalytically competent structure prior to desorption (168). This implies that cathepsins may utilize certain surface structures in addition to the active cleft in facilitating substrate binding. Recent NMR and bioinformatics studies followed by mutagenesis experiments have identified two exosites in MMP-12, which are critically contributing to elastin degradation (170,171). However, no distinct elastin binding sites have been identified in cathepsins.

In this study, using cathepsin V and cathepsin L as model proteases, a series of chimeras were generated to identify non-catalytic regions that are responsible for the potent elastolytic activity of cathepsin V. Two exosites were identified, which significantly contribute to the elastolytic activity of cathepsin V. Both exosites are distant from the active site of the protease forming a triangle-like docking platform with the active cleft for elastin. The substitution of cathepsin V specific amino acid residues in either exosite 1 or 2 with analogous cathepsin L residues led to a 75% or a 43% loss in the elastolytic activity. Substitutions in both exosites resulted in an elastolytically inactive variant despite retaining its full peptidolytic ability. The identification of exosites may contribute the design of inhibitors that will only affect the elastolytic activity of cathepsin V while not interfering with the normal protease functions of the enzyme.

Materials and Methods

Materials - Benzyloxycarbonyl-Phe-Arg-7-amido-4-methylcoumarin (Z-FR-MCA) was purchased from Enzyme System Products (Dublin, USA). The cathepsin inhibitor E-64 (L-3-carboxy-trans-2-3-epoxypropionyl-leucylamido-(4guanidino)-butane) and elastin-rhodamine conjugate were purchased from Bio Basic Inc. (Markham, Canada). Bovine neck elastin was purchased from EPC (Owensville, MO). The complete diverse tetrapeptide-ACC PS-SCL library was synthesized as previously described (124).

Construction of Chimera Cathepsin V-L Expression Vectors - Previously generated human wild-type cathepsin V and cathepsin L expression pPIC9 vectors (Invitrogen, Burlington, ON) (124,131) served as templates for constructing all chimera vectors. The cDNAs were generated by PCR using *Pfu*-polymerase (Fermentas, Burlington, ON) with the primers outlined in Table 2.1. Based on the amino acid sequence alignment, the primers were designed to allow the incorporation of different parts of the cathepsin V and cathepsin L sequences. The resulting chimera cDNAs were digested with *XhoI* and *NotI* restriction enzymes and ligated into the *P. pastoris* expression vector pPIC9 (Invitrogen, Burlington, ON) with the extracellular secretion signal α-factor replacing the endogenous signal peptides of the cathepsins. Subsequently, the vectors were linearized with *SacI* and electroporated into the *P. pastoris* strain, GS115, using a standard protocol provided by Invitrogen.

Mutants	Primer Sequences
Mutant 1 (M1)	
Forward	5’ - TCA CTG AGC GAG CAG AAT CTG GTG - 3’
Reverse	5’ - CAC CAG ATT CTG CTC GCT CAG TGA - 3’
Mutant 2 (M2)	
Forward	5’ - GGC CTG GAC TCT GAG GAA TCC TAT CC - 3’
Reverse	5’ - GGA TAG GAT TCC TCA GAG TCC AGG CC - 3’
Mutant 3 (M3)	
Forward	5’ - TCT GTT GCT AAT GAC ACC GGC TTT - 3’
Reverse	5’ - AAA GCC GGT GTC ATT AGC AAC AGA - 3’
Mutant 4 (M4)	
Forward	5' - GAG AAG GCC CTG ATG AAA GCA - 3'
Reverse	5' - TGC TTT CAT CAG GGC CTT CTC - 3'
Mutant 5 (M5)	
Forward	5’ - GGT GGT CTG GTG GTT GGC TAC GGC - 3’
Reverse	5’ - GCC GTA GCC AAC CAC CAG ACC ACC- 3’
Mutant 6 (M6)	
Forward	5' - CCT ATC CAT ATG AAG CAA CGG AAG AAA GCT GTA AGT AC - 3'
Reverse	5' - GTA CTT ACA GCT TTC TTC CGT TGC TTC ATA TGG ATA GG - 3'
Mutant 7 (M7)	
Forward	5' - TGT AAG TAC AAC CCT AAG TAT TCT GTT GCT - 3'
Reverse	5' - AGC AAC AGA ATA CTT AGG GTT GTA CTT ACA - 3'
Mutant 8 (M8)	using M4 as the template
Forward	same primer as M5
Reverse	same primer as M5
Mutant 9 (M9)	
Forward	5' - ACT GGC TTC GTG GAC ATC CCT AAG GAG AAG GCC - 3'
Reverse	5' - GGC CTT CTC CTT AGG GAT GTC CAC GAA GCC AGT - 3'
Mutant 10	
Forward	5’ – ACA GTG GTC GCA CCT AAG GAG – 3’
Reverse	5’ – CTC CTT AGG TGC GAC CAC TGT – 3’
Mutant 11	using M6 as the template
Forward	same primer as M9
Reverse	same primer as M9
Flanking	
Forward	5’ - GAC TGG TTC CAA TTG ACA AGC- 3’
Reverse	5’ - GCA AAT GGC ATT CTG ACA TCC - 3’

Table 2.1. PCR Primers for Chimera Cathepsin V-L Mutants. A summary of all the primers used in constructing chimera cathepsin V-L mutants. Primers served to incorporate corresponding portions of cathepsins V and L together based on DNA and amino acid sequence alignment. Nucleotides mutated in each mutant were underlined in both forward and reverse primers.

Pichia Pastoris Protein Expression - GS115 yeast clones transformed with cathepsin V-L chimera cDNAs were screened and selected using minimal dextrose (MD, Dextrose 2% v/v) and minimal methanol (MM, methanol 0.5% v/v) agar plates. The resulting His^+ Mut^+ clones were screened for maximal cathepsin activity towards the substrate Z-FR-MCA. The clones were briefly incubated in 5 ml BMGY media ((1%/ w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate (pH 6.0), 1.34% (w/v) yeast nitrogen base, 4x 10-15% (w/v) biotin, 1% (v/v) glycerol) at 30°C until the culture reached an OD_{600} of 2.0-6.0. The cells were then transferred into 3 ml of BMMY media with 0.5% (v/v) methanol to induce protein expression. During growth, the induction was maintained by adding 0.5% (v/v) methanol every 24 hours. Cells were incubated at 30°C for 4 days to reach the approximate maximum protein production and aliquots were tested for activity after processing of the proenzyme by pepsin as previously described (131).

For large scale production, the clone expressing the highest amount of chimera protein was first inoculated in 5 mL of MD media at 30°C for 24 hours, and then followed by further inoculation in 50 mL of BMGY media for 12 hours. The cells were then pelleted and resuspended into 500 mL of BMGY and then allowed to grow until an $O.D._{600}$ of 6.0. Afterwards, the cells were transferred into 1.5 L of BMMY media with induction of 0.5% (v/v) methanol every 24 hours while aliquots of the supernatant were withdrawn every day to monitor enzyme activity and contamination. The cells were harvested on the 4^{th} day of induction. The supernatant containing the secreted protease variants was concentrated 50-fold using an Amicon Ultrafiltration membrane (EMD Millipore, Darmstadt, Germany) to around 30 mL at 4°C.

Protein Activation and Purification - The concentrated yeast culture supernatant containing the chimera proenzyme was adjusted to pH 4.0 using diluted glacial acetic acid (Fisher Scientific, Ottawa, ON) and activated with 0.6mg/mL pepsin (Sigma-Aldrich, St. Louis, MO). The

mixture was incubated at 37°C and monitored for activity using Z-FR-MCA in 100 mM sodium acetate buffer (pH 5.5, containing 2.5 mM EDTA and 2.5 mM dithiothreitol (DTT)). The activation was stopped by increasing the pH to 5.5. The activated enzyme was supplemented with ammonium sulfate to a final concentration of 2 M. After centrifugation at 4,000 g for 20 minutes at 4°C, the cleared supernatant containing the activated enzyme was loaded onto a n-butyl Sepharose column (GE Healthcare, Little Chalfont, U.K.) and washed thoroughly with the loading buffer (100 mM sodium acetate, pH 5.0, 2 M ammonium sulfate, 0.5 mM EDTA, and 0.5 mM DTT). The protein was eluted from the column using a linear gradient with reducing ammonium sulfate concentration. The presence of the chimera protein was monitored using Z-FR-MCA as a substrate. Active fractions were combined and further concentrated 10-fold using Amicon ultra concentrator (EMD Millipore, Darmstadt, Germany) to about 1.5 mL and 3x buffer exchange with 100 mM sodium acetate buffer (pH 5.5, supplemented with 0.5 mM EDTA and 0.5 mM DTT). The final solution was divided into 100 µL aliquots for storage at -80 °C.

Enzyme Titration and Kinetic Studies - Enzyme concentrations were determined by active site titration using E-64 and steady state kinetics which were performed with the fluorogenic substrate, Z-FR-MCA, as previously described (75). The excitation and emission wavelengths used for the assays were set at 380 nm and 460 nm, respectively, and the rates were recorded using the spectro-fluorometer Perkin-Elmer LB50. Michaelis-Menten constants (K_m) and specificity constants (k_{cat}/K_m) were determined using non-linear regression analysis. All enzymes were assayed at room temperature at fixed enzyme concentrations (1-5 nM) and variable substrate concentrations (1-20 µM) in 100 mM sodium acetate buffer (pH 5.5, containing 2.5 mM DTT and 2.5 mM EDTA).

The substrate specificity of the S1-S4 subsites of M12 and wild-type cathepsin V were determined using four positional scanning fluorogenic 4-mer substrate combinatorial libraries, as described (124). The P1-diverse library consisted of 20 sub-libraries in which only the P1 position was systematically held constant with one of the 20 proteinogenic amino acids, omitting cysteine and including norleucine. The P2, P3, and P4 positions were randomized with an equimolar mixture of the 19 amino acids, for a total of 6,859 substrate sequences per well. The extended P4-P2 substrate specificity of proteases was profiled with the tetrapeptide P1-Arg fixed library in which the P1 position was held constant as Arg, which was determined to be the most favored amino acid at the P1 site by the P1-diverse library. An aliquot of the P1-diverse library was added to 20 wells (6,859 compounds per well) and assays were carried out in a 96-well Microfluor plate (Dynex Technologies, Chantilly, VA) for a final concentration of 7.3 nM for each compound. To determine the P2, P3, P4 specificities of both cathepsins, a P1-Arg fixed tetrapeptide library comprised of P2, P3, P4 sublibraries was used. This library had Arg at the P1 position and the P2, P3, P4 positions were spatially addressed with 19 amino acids (excluding cysteine and including norleucine). Approximately 9.5 x 10^{-9} mol aliquots of each sub-library of the P1-Arg fixed library were added to 57 wells (361 compounds per well) of a 96-well plate for a final concentration of 0.25 μM. The hydrolysis of the substrate libraries was measured in the presence of 5 nM of cathepsin V and 5 nM of M12 by monitoring the fluorescence released (spectro-microfluorimeter SpectraMax Gemini, Bioproducts, Sunnyvale , CA) at wavelengths of 380 nm for excitation and of 460 nm for emission. Assays were performed at 25°C in a buffer containing 100 mM sodium acetate (pH 5.5, 100 mM NaCl, 10 mM DTT, 1 mM EDTA, and 0.01% Brij-35).

Elastin-Rhodamine Conjugate Degradation Assay and HPLC Degradation Profile - Reaction mixture containing 1 μM enzyme and 10 mg/mL elastin-rhodamine conjugate (EMD

Millipore, Darmstadt, Germany) for a final volume of 1 mL 100 mM sodium acetate buffer (pH 5.5, containing 2.5 mM EDTA and 2.5 mM DTT) was incubated at 37°C at 200 rpm on a MaxQ 5000 floor shaker (Geneq Inc, Montreal, QC). Samples were withdrawn at time points 0, 30, 60, 90, and 120 minutes, and mixed with E-64 to stop the reaction. Samples were then centrifuged at top speed using a bench-top AccuSpin Microcentrifuge (Fisher Scientific, Ottawa, ON) to sediment all undigested elastin particles. The fluorescence of the supernatants were recorded using a 96-well plate format at excitation and absorption wavelengths of 570 nm and 590 nm, respectively, using a Gemini plate reader (Molecular Devices, Sunnyvale, CA). All measurements were carried out in triplicates in three independent experiments. The residual cathepsin activity was recorded at the same time points prior to the addition of E64 using the substrate Z-FR-MCA as described above.

The HPLC degradation profile of bovine neck elastin used 1 μM of cathepsin (final concentration) mixed with 10 mg/mL bovine neck elastin (EPC, Owensville, Missouri) in 1 mL 100 mM sodium acetate buffer (pH 5.5, containing 2.5 mM EDTA and 2.5 mM DTT) at 37°C and 200 rpm for 18 hours. The reaction was stopped with E-64, samples were centrifuged, and the supernatants were loaded onto a reversed phase C-18 analytical column (Phenomenex, Torrance, CA) on a Beckman Coulter System Gold© 126 HPLC (Brea, CA) and eluted with a linear gradient starting with 0.1% trifluoric acid and ending with 90% acetonitrile supplemented with 0.1% trifluoric acid.

Binding Assay for Cathepsins to Elastin – Aliquots of 1 mg of elastin-congo red conjugate was separately incubated with E-64 inhibited 1 μM cathepsin V, cathepsin L and M11 for 2 hours in 50 mM sodium acetate buffer (pH 5.5, containing 2.5 mM DTT and EDTA) at 600 rpm and 37ºC. Assays were stopped by centrifugation at top speed using a bench-top AccuSpin

microcentrifuge (Fisher Scientific, Ottawa, ON) to remove the insoluble contents. The soluble protein contents were then visualized and quantified on coomassie blue stained SDS-PAGE gel.

Electron Scanning Microscopy of Elastin Digests – Aliquots of 10 mg of bovine neck elastin (EPC, Owensville, Missouri) in a final volume of 1 mL 100 mM sodium acetate buffer (pH 5.5, containing 2.5 mM EDTA and 2.5 mM DTT) were separately incubated with 1 µM cathepsin V, cathepsin L and M11 at 37°C and 200 rpm on a MaxQ 5000 floor shaker (Geneq Inc, Montréal, QC) overnight. The elastin specimens were washed and pelleted at 1500 rpm for 30 seconds at room temperature. The pellets were subsequently dislodged and resuspended evenly in 20 µL of distilled water. 1 µL of the suspension was air dried on a nickel-copper EM grid and sputter-coated with gold palladium for examining with a Cambridge 260 Stereoscan SEM at 6 to 8 kv (Centre for Highthroughput Phenogenomics, Dentistry, UBC).

Statistics - Results are expressed as mean ± SD. The significance of differences of the mean values was calculated using one-way ANOVA (t-test). A p-value of less than 0.05 was considered significant.

Results

Mapping of Potential Elastin-binding Regions in cathepsin V – There is approximately 22% sequence variation between wild-type human cathepsins V and L (109). The highest incidence of sequence differences are found in the middle region of their amino acid sequences (Figure 2.1A). In order to identify potential sites that are important for the elastolytic activity of cathepsin V, a series of chimeras were generated by sequentially replacing increasing parts of cathepsin V with its analogous sequence from cathepsin L. Five chimeras, M1, M2, M3, M4 and

M5, were constructed, each of which contained various proportions of cathepsin V and cathepsin L (Figure 2.1B). The integrity of the active sites of the chimeras was determined by kinetic analysis using synthetic substrate Z-FR-MCA (Table 2.2). Chimeras M1 to M5 all possessed a functional active site which allowed the effective cleavage of the peptide substrate. The k_{cat} values for both wild-type cathepsins and their chimeras were similar. The K_m values, however, decreased and thus the k_{cat}/K_m values increased from M1 to M5. The k_{cat}/K_m values were also in line with their parent enzymes depending on the sequence contributions of cathepsin V and cathepsin L in each chimera. Chimera M2 with approximately one third of its N-terminal amino acid sequence replaced with that of cathepsin L revealed a K_m value much closer to the value of cathepsin L. This is not surprising as the main substrate specificity determining subsite S2 is greatly swayed by the amino acid residue at position 69 (phenylalanine in cathepsin V and a leucine in cathepsin L) (122,199). M2, which had a leucine at this position like cathepsin L, would therefore exhibit a preference for bulky aromatic residues in the P2 position for a substrate like Z-FR-MCA.

A

```
catV  LPKSVDWRKKGYVTPVKNQKQCGSCWAFSATGALEGQMFRKTGKLVSLSEQNLVDCSRPQ
catL  APRSVDWREKGYVTPVKNQGQCGSCWAFSATGALEGQMFRKTGRLISLSEQNLVDCSGPQ

                                        exosite 1a  exosite 1b        exosite 2
catV  GNQGCNGGFMARAFQYVKENGGLDSEESYPYVAVDEICKYRPENSVANDTGFTVVAPGKE
catL  GNEGCNGGLMDYAFQYVQDNGGLDSEESYPYEATEESCKYNPKYSVANDTGFVDI-PKQE

catV  KALMKAVATVGPISVAMDAGHSSFQFYKSGIYFEPDCSSKNLDHGVLVVGYGFEGANSNN
catL  KALMKAVATVGPISVAIDAGHESFLFYKEGIYFEPDCSSEDMDHGVLVVGYGFESTESDN
```

B

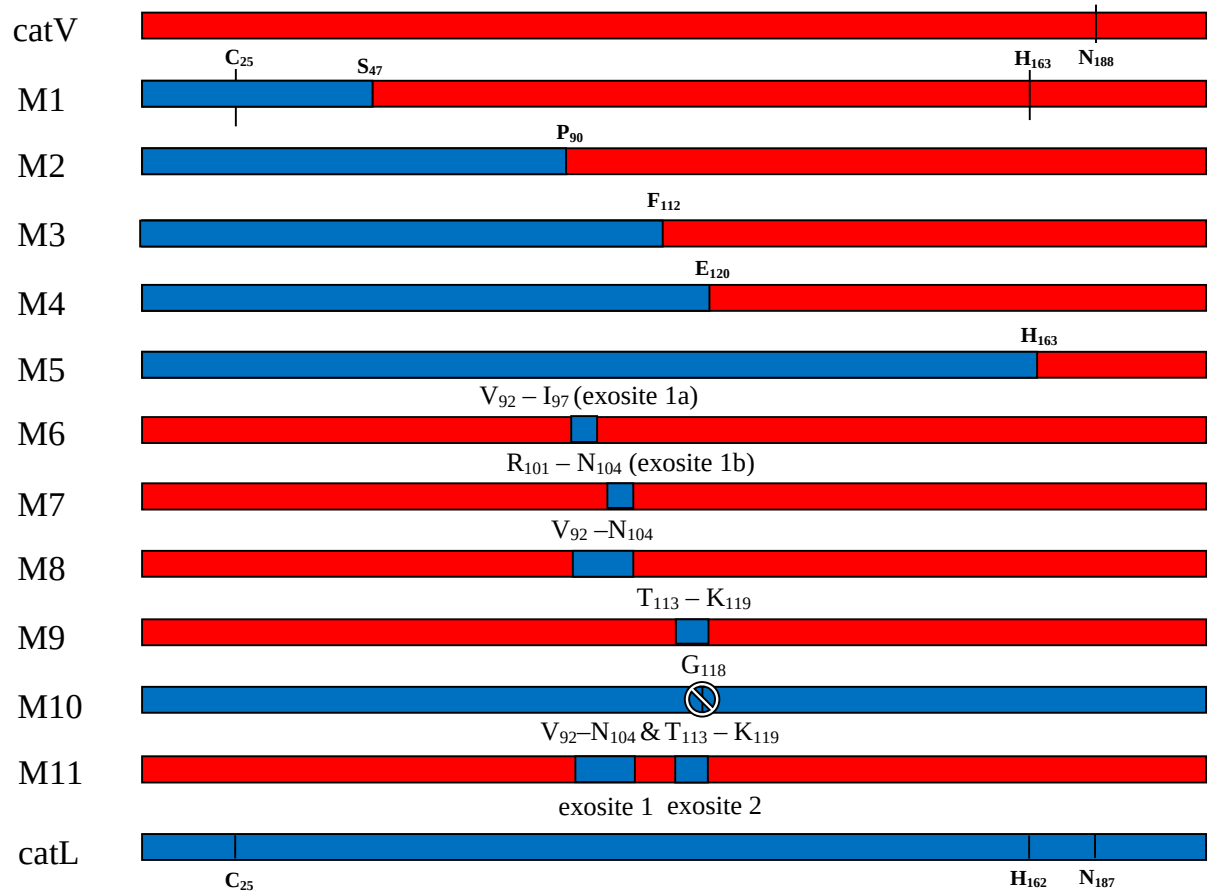


Figure 2.1. Mapping of Exosites in Cathepsin V. A. Amino acid sequence alignment of the mature wild-type cathepsin V **(**catV) and cathepsin L (catL). Exosite residues from catV are bordered whereas their analogues residues in catL are underlined. The '-' is inserted to allow proper alignment. Exosite 1a is highlighted in yellow, exosite 1b is highlighted in turquoise blue,

and G_{118} residue after the P_{117} in exosite 2 is highlighted in green. B. Schematic diagram of the wild-types and all chimeras. CatV is presented in red and catL is presented in blue, while the chimeras were generated with portions taken from catV and catL. C_{25}, H (residue 163 in catV and 162 in catL) and N (residue 188 in catV and 187 in catL) are the three signature active site amino acid residues in cathepsins. Exosite 1 spans from V_{92} – N_{104} respectively (1a: V_{92}–I_{97}, 1b: R_{101}–N_{104}), and exosite 2 contains amino acid residues T_{113} – K_{119}. The '⦸' in M10 indicates G_{118} deletion.

The elastolytic potencies of the chimeras were evaluated using elastin-rhodamine conjugate as substrate. Upon cleavage of insoluble elastin into small soluble fragments, rhodamine is released into the aqueous phase and can be monitored at excitation and absorption wavelengths of 570 nm and 590 nm, respectively. As shown in Figure 2.2A, chimeras M1 and M2 showed comparable activities to wild-type cathepsin V, whereas M3, M4, and M5, lost 41%, 64% and 85% of the elastolytic activities accordingly. The loss in activity of chimeras M3 to M5 was not due to an impairment of the active site as shown evidently by their increasing k_{cat}/K_m values. For example, M5 revealed an approximately 9-fold increase in the k_{cat}/K_m value from cathepsin V for the cleavage of Z-FR-MCA (Table 2.2).

Increasing Cathepsin L Portion

Enzyme	k_{cat} (sec^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}sec^{-1}$)
catV	25.2 ± 0.8	11.4 ± 0.1	2.2 ± 0.1 x 10^6
M10	24.4 ± 0.2	11.0 ± 0.3	2.2 ± 0.1 x 10^6
M7	24.8 ± 0.5	10.6 ± 0.7	2.3 ± 0.2 x 10^6
M6	26.2 ± 0.7	12.3 ± 0.4	2.1 ± 0.1 x 10^6
M9	25.3 ± 0.5	10.5 ± 0.2	2.4 ± 0.1 x 10^6
M8	26.1 ± 0.3	11.5 ± 0.3	2.3 ± 0.1 x 10^6
M11	22.7 ± 0.6	10.4 ± 0.5	2.2 ± 0.2 x 10^6
M1	23.2 ± 0.5	11.2 ± 0.8	2.1 ± 0.2 x 10^6
M2	24.9 ± 0.3	1.2 ± 0.1	2.1 ± 0.2 x 10^7
M3	25.2 ± 0.4	2.0 ± 0.2	1.3 ± 0.1 x 10^7
M4	21.7 ± 0.4	1.8 ± 0.6	1.2 ± 0.4 x 10^7
M5	21.5 ± 0.1	1.1 ± 0.3	1.9 ± 0.5 x 10^7
catL	20.8 ± 0.4	1.5 ± 0.2	1.4 ± 0.2 x 10^7

Table 2.2. Enzyme Kinetics of Wild-Types and Chimeras. The chimera mutants are ordered with decreasing portions of cathepsin V (catV) to show the changing trends in the k_{cat}, K_m and k_{cat}/K_m values. The kinetic values for each enzyme were determined from three independent experiments of Z-FR-MCA hydrolysis assays using a non-linear regression algorithm of the GraphPad Prism program. Cathepsin L denoted as catL.

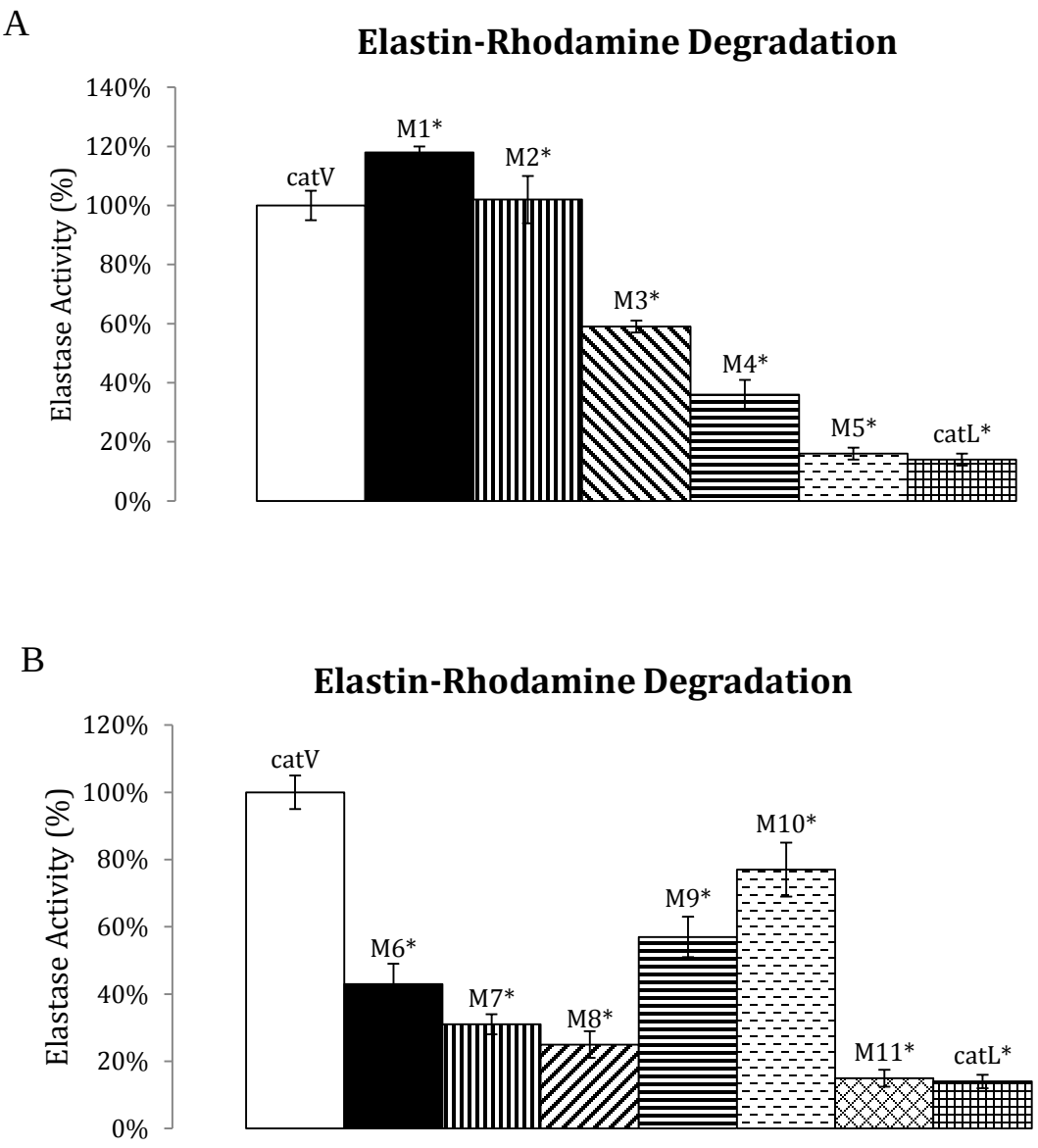


Figure 2.2. The Elastolytic Activity of Wild-Type Cathepsins V and L and Cathepsin V/L-Chimeras Using Elastin-Rhodamine Conjugate. A. Elastase activity of M1 to M5. B. Elastase activities of M6 – M11. Assays were performed at 37°C with 1 µM of enzyme at 200 rpm in the activity buffer containing 100 mM sodium acetate, 2.5 mM EDTA and 2.5 mM dithiothreitol at pH 5.5 for 120 minutes. Elastin-rhodamine cleavage was monitored at the excitation and absorption wavelengths at 570 nm and 590 nm, respectively. All measurements were done in triplicates in three independent experiments. *P*-value represents unpaired *t*-test statistical analysis relative to catV, * indicates $P < 0.05$.

Identification of Exosites 1 and 2 between Region Pro_{90} and Glu_{120} on Cathepsin V - Chimeras M1 to M5 provided important information for narrowing down the location of elastin-binding exosites in cathepsin V. Since M2, which contained the beginning 90 amino acid residues of cathepsin L from the N-terminus, preserved the same level of elastolytic activity as cathepsin V, it was unlikely that exosites resided within amino acid residues Leu_1 to P_{90} (Figure 2.1B). On the other hand, M5 which had nearly 80% of its C-terminus sequence replaced with that of cathepsin L, lost 85% of the elastolytic activity. This excluded residue H_{163} to V_{221} from being crucial in binding elastin. Together, M3 and M4, which diminished 41% and 64% of the activity conveyed the finding that amino acid residues P_{90} to E_{120} was most likely to be accountable for the elastolytic activity of cathepsin V (Figures 2.1B and 2.2A). From these 31 amino acid residues, 12 residues were different between cathepsin V and cathepsin L, and formed two clusters separated by a cathepsin V/L-identical octapeptide, S_{105}VANDTGF$_{112}$. Four more chimeras were constructed to further dissect this area. M6 was generated by the substitution of V_{92}AVDEI$_{97}$ of cathepsin V with E_{92}ATEES$_{97}$ of cathepsin L; M7 replaced R_{101}PEN$_{104}$ with N_{101}PKY$_{104}$; M8 contained both substitutions of M6 and M7; and finally M9 swapped T_{113}VVAPGK$_{119}$ with V_{113}DIPKQ$_{118}$ from cathepsin L (Figure 2.1B).

Chimeras M7 – M 9 showed k_{cat}, K_m, and k_{cat}/K_m values comparable to those of the wild-type cathepsins V and L, indicating the presence of fully functional active sites (Table 2.2). However, similar to chimeras M3 and M4, they demonstrated up to 75% loss of elastolytic activity. M6, M7, and M8 lost a 57%, 69%, and 75%, respectively, of the elastolytic activity compared to cathepsin V (Figure 2.2B). It was noted that the reduction of activities was not simply additive as M8 exhibited an activity that did not correspond to the total reduction sum of M6 and M7. This suggested a synergistic relationship between motifs V_{92}AVDEI$_{97}$ and R_{101}PEN$_{104}$. Based on this

observation and the fact that the three amino acid residues, $C_{98}KY_{100}$, were also present in cathepsin L at the same position, motifs $V_{92}AVDEI_{97}$ and $R_{101}PEN_{104}$ were combined together and termed exosite 1. In addition, M9 showed a 43% reduction in digesting elastin (Figure 2.2B) implying that motif $T_{113}VVAPGK_{119}$ was important, and was thus denoted as exosite 2 (Figure 2.1).

In order to obtain more information about the cleavage specificities of the chimeras, an HPLC profile for each chimera was completed by analyzing the solubilized peptide fragments from digesting unlabeled bovine neck elastin. Comparison of the HPLC profiles suggested that M1 to M5 generated highly identical peaks like cathepsin V with only a few noticeable variations (Figure 2.3A and 2.3B). As chimeras gained more cathepsin L characteristics, significant reductions in peak intensities were observed. The area underneath each HPLC profile was determined to quantify the strength of degradation by each chimera in reference to cathepsin V. M1 to M5 captured 56%, 43%, 41%, 38%, and 26% area of cathepsin V in comparison to 16% of cathepsin L, signifying a graduate loss of elastolytic activity (Figure 2.3F). More specifically, M1 showed nearly the same peak area as cathepsin V for peaks 1 and 2, but displayed apparent losses for the remaining peaks (Figure 2.3A). Continuing this pattern, the HPLC profiles of M2, M3, M4 and M5 became increasingly similar to that of cathepsin L. The baseline of M5 nearly overlapped the elution profile of cathepsin L with some minor spikes that were unique (Figure 2.3B). Chimeras M8 and M9 also demonstrated reduced degradation efficacies, which were in line with the results of their elastin-rhodamine assays. M8 produced less degradation products (31% area of cathepsin V) compared to M9 (53% area of cathepsin V), but both contained the characteristic elution peaks of cathepsin V with significantly less intensities (Figures 2.3C, 2.3D and 2.3F). Attempts to further analyze the HPLC derived peaks by Maldi-MS are presently in progress in collaboration with the

Schmelzer laboratory in Germany. However, we expect to identify only a limited number of cleavage fragments assigned to the known primary sequences of soluble bovine or human kappa elastin (200-205), as most fragments are likely desmosin cross-linked as cited in numerous literatures (206-211).

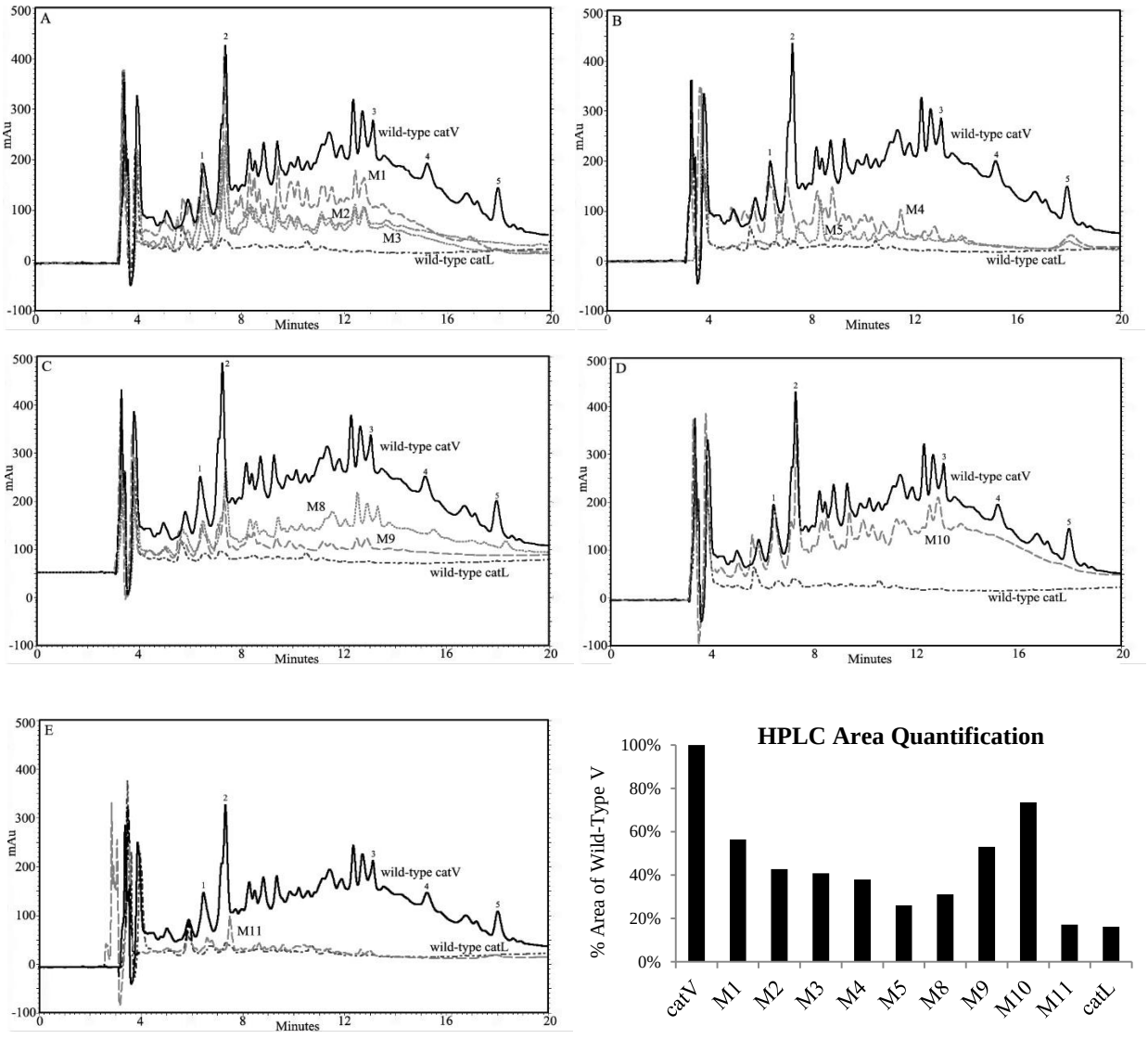


Figure 2.3. HPLC Degradation Profile of Bovine Neck Elastin with Wild-Type Cathepsins V and L and Their Chimeras. A. Overlay of wild-type catV and L, M1, M2 and M3 profiles. B. Overlay of wild-type catV and L, M4, and M5 profiles. C. Overlay of wild-type catV and L, M8 and M9 profiles. D. Overlay of wild-type catV and L and M10 profiles. E. Overlay of wild-type catV and L and M11 profiles. F. HPLC Area Quantification. Image-J (version 1.46r, developed by

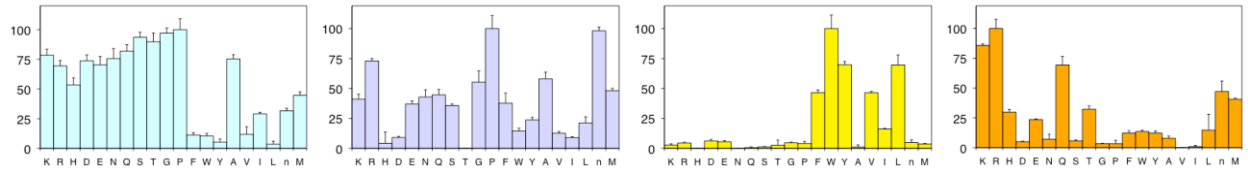
Wayne Rasband, National Institutes of Health, U.S.A) was used to quantify the areas underneath the HLPC curves from Figures 3A-3E. In general, 1 µM of each enzyme was mixed with 10 mg/mL bovine neck elastin in assay buffer at 37 °C at 200 rpm for 18 hrs. The reaction was stopped with E-64, and the supernatant was load onto an analytical reserved phase C-18 column on HPLC to generate the above spectra. A linear gradient starting with 0.1% trifluoric acid and ending with 90% acetonitrile supplemented with 0.1% trifluoric acid was used to elude the digestion products. The arrow indicates that the legend was formatted to correspond to each spetrum from top to bottom. Major peaks are labeled with numbers for easier comparison.

The Combined Effects of Exosites 1 and 2 on the Elastolytic Activity of Cathepsin V

- To further evaluate if exosites 1 and 2 were contributing to the binding to elastin, M11 was generated by integrating analogous sequences from cathepsin L to replace both exosites in cathepsin V (Figures 2.1 and 2.2). As indicated in Figures 2.2B, 2.3E and 2.3F, M11 exerted 15% residual activity compared to that of cathepsin L of 14% (in comparison to cathepsin V), and produced closely overlapping HPLC degradation profile (17% area of cathepsin V) to cathepsin L (16% area of cathepsin V) against insoluble bovine neck elastin. In other words, by swapping both exosites from cathepsin V, M11 acted like cathepsin L. Two methods were used to verify that the substitutions of both exosites only altered the elastin-binding ability of cathepsin V but had no effect on the active site. First, Michaelis-Menten kinetics showed that M11 had a very similar k_{cat}/K_m value to that cathepsin V for the cleavage of Z-FR-MCA (Table 2.2). Second, substrate profiling analysis using a complete combinatorial library containing 160,000 ACC tagged tetrapeptides revealed no major differences between M11 and cathepsin V except an increased acceptability for a few amino acid residues (A, T, M, F, W, and Y) in the S3 subsite (Figure 2.4).

P4 P3 P2 P1

Cathepsin V



M11

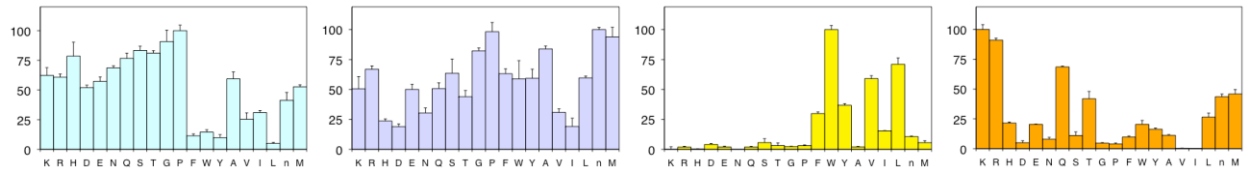


Figure 2.4. Determination of the Substrate Specificity of the Wild-Type Cathepsin V and M11 Using a Complete Diverse Peptide Substrate Library. All assays were carried out in three independent experiments. P4, P3, P2 and P1 denote the four active site binding pockets of proteases. The x-axis indicates a total of 20 amino acid residues which are ordered based on their chemical characteristics of side chain residues (acid, basic, polar, aromatic and aliphatic amino acids). The y-axis represents the picomolar fluorophore produced per second.

Involvement of G_{118} in Cathepsin V - Sequence alignments of the known cathepsin elastases including K, S, and V indicated that Pro_{117} from exosite 2 (based on the numbering of cathepsin V) was followed by a glycine residue which is missing in cathepsin L (Figure 2.1A). The comparison of the crystal structures of cathepsins V and L further revealed that K_{119} adjoining G_{118} seem to position its side chain outwards, away from the active cleft, whereas such orientation is opposite in cathepsin L (Figure 2.5) (122,212). This structural difference within exosite 2 led to the speculation that G_{118} in cathepsin V might be involved in orienting the adjacent lysine residue to create an elastin interacting point, as well as opening up the entrance for elastin into the active

cleft. To test this hypothesis, M10 was created by deleting G_{118} from cathepsin V (Figure 2.1B). The results depicted that M10 retained nearly all active site characteristics of cathepsin V including the corresponding k_{cat}, K_m, and k_{cat}/K_m values for the hydrolysis of the Z-FR-MCA substrate. However, the degradation of elastin-rhodamine showed a 23% reduction in cleavage efficacy (Figure 2.2B). The effect of the glycine$_{118}$ residue was also supported by the HPLC degradation product profile demonstrating a reduction in area underneath the curve (73% area of cathepsin V) and in peak intensities (Figures 2.3D and 2.3F). These results all support that G_{118} plays an important role in exosite 2.

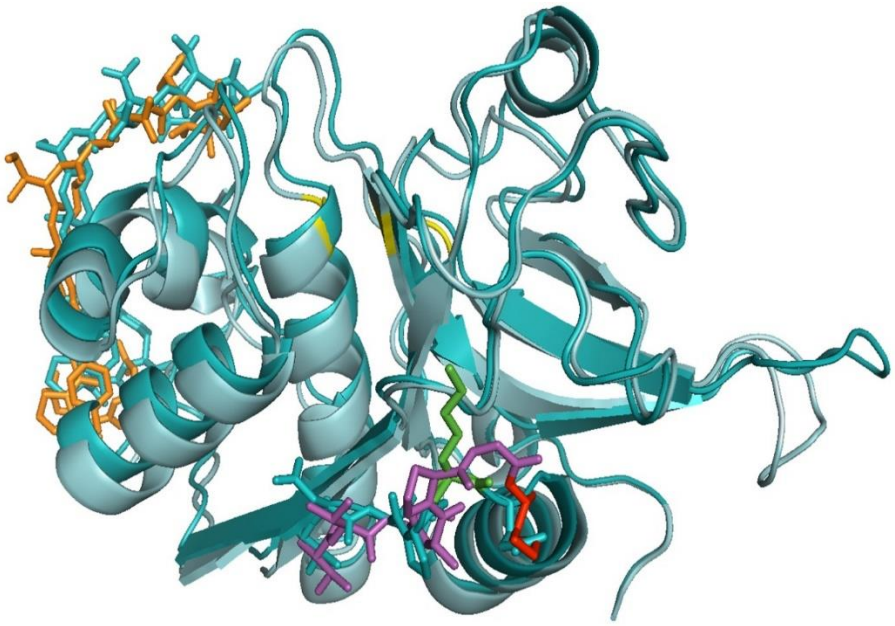
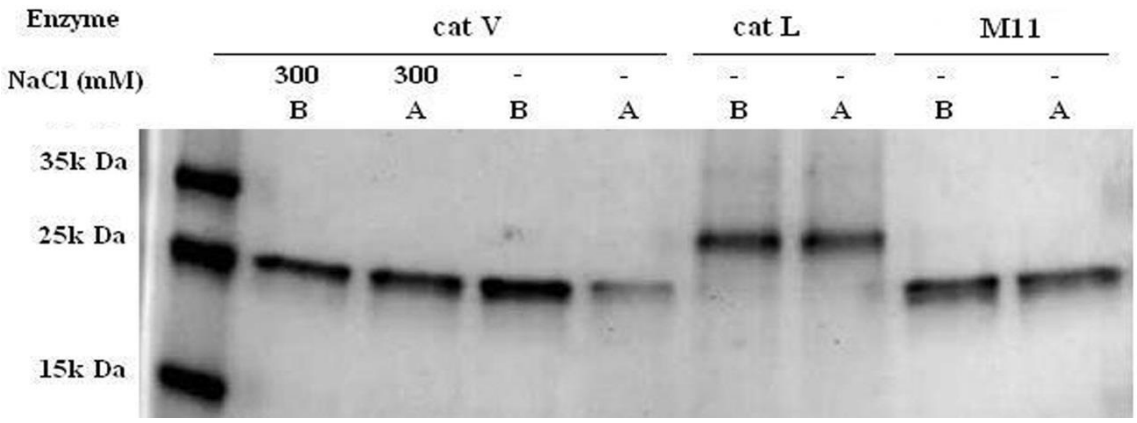


Figure 2.5. Superimposition of Human Cathepsins V and L. Cathepsin V is colored in cyan (lighter blue) with its exosite 1, shown as sticks, colored in orange, and exosite 2, colored in purple. Cathepsin L is colored in teal (darker blue) with its exosite analogous shown as sticks. Active site residues C_{25}, H_{163}, N_{188}, are colored in yellow. K_{119} of cathepsin V is highlighted in red whereas K_{118} of cathepsin L is highlighted in green. Crystal images are obtained from PDB. Cathepsin V: 1FH0, and cathepsin L: 2XU3.

Adsorption of Cathepsin V, Cathepsin L and M11 to Insoluble Elastin - The ability to exit the soluble phase and adhere to the insoluble substrates might be necessary for an effective elastase to carry out proteolysis. This implies that both exosites of cathepsin V serve as initial anchor points to elastin. To verify this hypothesis, an elastin adsorption assay was developed. Equimolar amounts of cathepsin V, cathepsin L, and M11 were inoculated with insoluble elastin-congo red. All proteases were E-64 inhibited to prevent elastolysis. This allowed an assessment on the adsorption of exosites of cathepsin V to bind to insoluble elastin. The result showed that cathepsin L and M11 did not bind to elastin as indicated by the amount of soluble protease prior and after the addition of insoluble elastin as analyzed by SDS-PAGE (Figure 2.6A). In contrast, cathepsin V revealed a 52% difference, meaning more than half of the cathepsin V molecules bound to elastin. When the reaction buffer is supplemented with 300 mM NaCl, all cathepsin V protein was recovered in the soluble phase (Figure 2.6B). This indicated that 300 mM NaCl completely abolished the exosite-mediated absorption of cathepsin V to elastin.

A



B

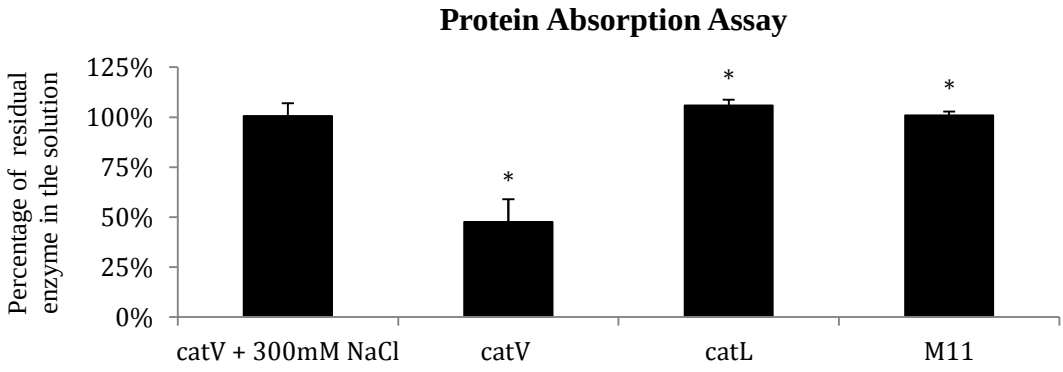


Figure 2.6. Protein Absorption Assay onto Insoluble Elastin-Rhodamine. 1 µM E-64 inhibited enzymes (wild-type catV, catL and M11) were inoculated with 10 mg insoluble elastin-rhodamine in 1 mL activity buffer for 2 hours at 600 rpm at 37˚C. A. Samples prior to contacting denoted as (**B**) and after contacting elastin-rhodamine denoted as (**A**) were compared and quantified on coomassie blue stained SDS-PAGE gel. B. Quantification of protein absorption assay of wild-types and M11. All measurements were determined in in three independent experiments. *P*-value represents unpaired *t*-test statistical analysis relative to sample "catV + 300 mM NaCl", * indicates $P < 0.05$.

Electron Microscopy Imaging of Cathepsin Digested Elastin - Morphological changes were clearly observed on bovine neck elastin digested with cathepsin V. Particles lost their surface

structure and appeared to melt with each other with undefinable surface morphology (Figure 2.7). This was noticeably different from the undigested control which had a tube–like shape with a sharp surface and larger particle size (Figure 2.7). The images of elastin particles incubated with cathepsin L and M11 were very similar to the undigested sample and thus coincided with their lack of elastolytic activities.

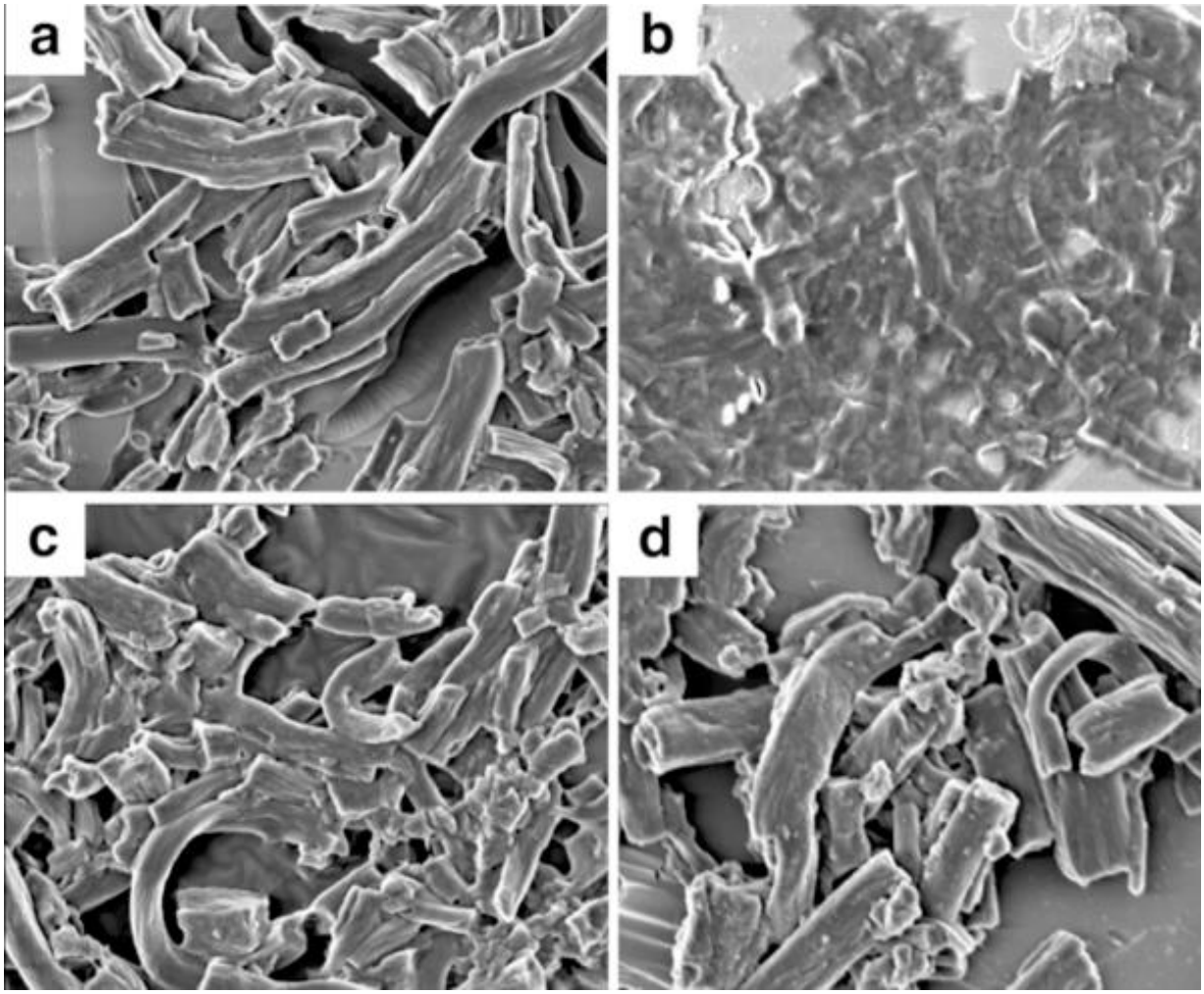


Figure 2.7. Scanning Electron Microscopy Images of Bovine Neck Elastin Powder. Incubated a) Bovine neck elastin in acetate buffer, pH 5.5, b) Bovine neck elastin with cathepsin V, c) Bovine neck elastin with M11 variant, and d) Bovine neck elastin with cathepsin L. All cathepsin concentrations were 1µM. The incubation was done overnight at 37 ºC.

Discussion

Sharing 78% amino acid sequence, cathepsins V and L are alike in many different aspects. For example, their 3-D crystal structures reveal highly superimposable secondary and tertiary arrangements with a small number of variations (Figure 2.5) (122,212). The active cleft, even though with a few differences in the constituent residues in S2 and S3 subsites (122), demonstrates very similar substrate preferences in all three subsites as shown by Choe Y. *et al.* using a library of synthetic substrates (124). Despite these similarities, many physiological functions of these two enzymes are not overlapping. Cathepsin V is described as a highly potent elastase, whereas cathepsin L demonstrates only a minimal elastolytic activity. As minor active site variations between the two enzymes do not explain the dramatic differences in their elastase activities, the presence of elastin binding exosites in cathepsin V becomes highly probable. To locate the locations of exosites, a total of 11 chimeras were made by swapping analogous sequences of cathepsin L into cathepsin V (Figure 2.1B). Evaluation of their elastolysis revealed two exosites in cathepsin V. Exosite 1 spans through the region of V_{92}AVDEICKYRPEN$_{104}$ and exosite 2 contains amino acid residues T_{113}VVAPGK$_{119}$. Both regions do not contain any amino acid residues that are involved in the formation of the active cleft or the classical substrate binding pockets (122). As shown in Figure 2.8, exosite 1 resides on the lefty lobe of the enzyme surface, which comprises 54% hydrophobic and 46% neutral amino acid residues. Cathepsin L in the same region contains 54% hydrophobic, 31% neutral and 15% hydrophilic amino acid residues respectively. GRAVY scores, which measure the average degree of hydrophobicity in peptide sequences, provide a more direct understanding on the shifts of hydropathy in this case (213). For example, replacement of exosite 1 with its analogous residues from cathepsin L reduces hydrophobicity by 187% from a GRAVY score of -0.62 to -1.78, and swapping of exosites 2

changes the GRAVY score from 0.51 to -0.63, which is a 224% decrease (Table 2.3). The significance of hydrophobicity may be of great importance for an efficient elastase. As it has been known, elastin mainly consists of neutral and hydrophobic amino acid residues (201), a higher hydrophobicity in the exosites of cathepsin V can serve as a better binding interface for elastin via van der Waals interactions (214). This rationale is supported by the dramatic deduction in the elastolytic activity of M11, implying weakening of the enzyme-substrate interaction by cathepsin L analogues (Figure 2.2B).

Cathepsin V		**GS**	**GS**	**Cathespin L**	
exosite 1a	VAVDEI	1.28	-1.70	EATEES	exosite 1a
exosite 1b	RPEN	-3.28	-2.58	NPKY	exosite 1b
exosite 1	VAVDEICKYRPEN	-0.62	-1.78	EATEESCKYNPKY	exosite 1
exosite 2	TVVAPGK	0.51	-0.63	VDIPKQ	exosite 2

Table 2.3. Comparison of Hydrophobicity between Exosites of Cathepsin V and Their Analogues in Cathepsin L. Exosite 1 of cathepsin V is divided into exosite 1a and exosite 1b. GS stands for gravy score, which is used to measure the hydrophobicity of peptide sequences. GRAVY scores were calculated using online calculator provided by Dr. Stephan Fuchs (University of Greifswald, Department of Microbiology, Greifswald, Germany). http://www.gravy-calculator.de/

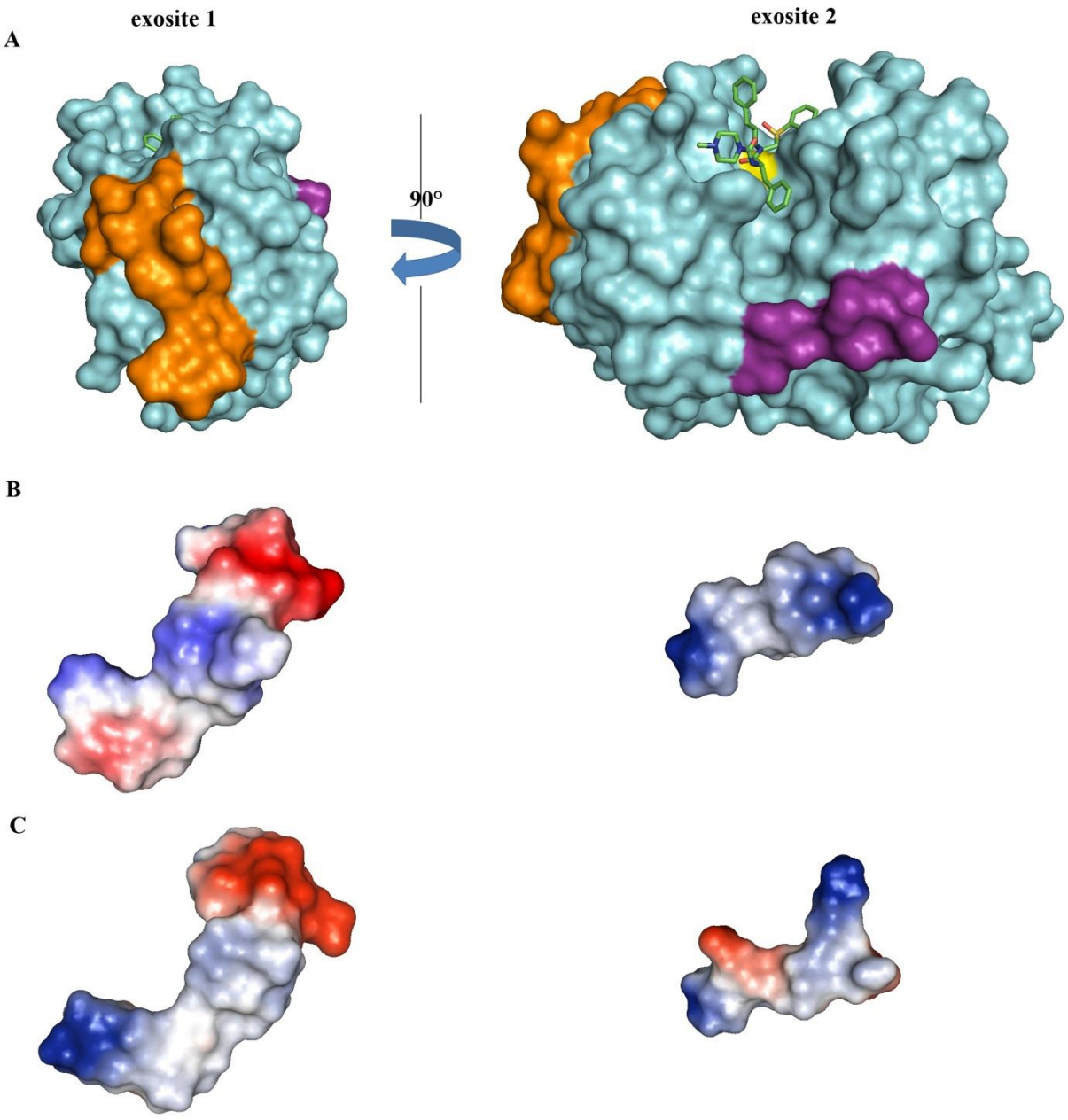


Figure 2.8. Location and Electrostatics of Exosites of Cathepsin V. A. Locations of exosites on cathepsin V. Exosite 1, colored in orange, contains residues V_{92}AVDEICKYRPEN$_{104}$. Exosite 2, colored in purple, includes residues T_{113}VVAPGK$_{119}$. Active site residue $C_{25,}$ colored in yellow, is bound to vinyl sulfone inhibitor ($C_{32}H_{40}N_4O_4S$). Crystal images are modified with Pymol from the original PDB file: 1FH0. B. Electrostatics of exosites of cathepsin V. C. Electrostatics of analogous sequences in cathepsin L: 2XU3. Positive electrostatics is shown in red; negative is shown in blue; neutral is shown in white.

Both exosites clearly contribute to the elastase activity of cathepsin V but not in an additive manner. Exosite 1 is responsible for 75% and exosite 2 for 42% as shown with the M8 and M9 chimeras. Both exosites together (M11 chimera) account for 85% of the elastolytic activity which is in line with the residual activity of cathepsin L of about 14%. This may imply a cooperative mode of enzyme-substrate interaction. As the M8 chimera deprived of exosite 1 causes a more severe reduction to the elastolytic activity of cathepsin V, it is possible that this exosite provides the initial contact point to elastin which then leads to the interaction to exosite 2. Exosite 2 sits in close proximity to the classical subsite binding region of cathepsin V. The combined binding to both exosites will thus maneuver susceptible peptide bonds of elastin into the active site for subsequent cleavage. A recent study by Novince, M. *et al.* supports this scenario (168). The authors propose that cysteine cathepsins may degrade elastin through cycles of dynamic adsorption and desorption. Such process involves a reorientation of the enzyme from a nonproductive manner to a catalytically active form upon adsorbing to elastin. Interestingly, in contrast to cathepsins K and S, cathepsin L was found to poorly absorb onto insoluble elastin (168). This finding is in agreement with the results found in the elastin-congo red adsorption assay (Figure 6). Overall, both exosites and the active cleft form a triangular arrangement on the enzyme surface of cathepsin V acting like a docking platform for elastin. More importantly, such non-linear arrangement may even force the cross-linked tropoelastin to bend and unfold itself for cleavage. Interestingly, recently identified exosites in MMP-12 form a similar triangular pattern (170).

In addition, the identification of the exosites in cathepsin V provides an explanation to the underlying mechanism where they function as the interacting points to allow enzyme from the aqueous phase to come in contact with elastin in the insoluble phase. On the contrary, cathepsin L, lacking both exosites, may not generate an adsorption strong enough to allow the subsequent

hydrolysis to occur. It is important to point out that once elastin loses its insoluble nature due to breakdown of the cross-linkages between tropoelastins, cathepsin L was reported to demonstrate a potent activity against soluble elastin such as ETNA-elastin (168). Although this observation has earned cathepsin L the name of elastase in some studies, it is clear that cathepsin L is incapable of digesting the native insoluble elastin *in vitro*. As cathepsins V and L have similar active site substrate specificities, cathepsin L may also be able to carry out proteolytic degradation in a similar fashion to cathepsin V once elastin has been solubilized. As a result, the presence of both exosites empowers cathepsin V the true functionality of an elastase.

The location and configuration of exosite 2 appears to be crucial. Exosite 2, which resembles a bridge-like structure connecting the L and R domain of cathepsin V, sits beneath the S2 subsite pocket of the active cleft (Figure 2.8). This particular subsite has in common with the other elastolytic cathepsins K and S a glycine residue in position 118 and distinguishes it from the non-elastase, cathepsin L. The additional glycine residue makes the exosite 2 bridge somewhat wider, which may be crucial to escort the elastin to the catalytic site for hydrolysis. Moreover, the glycine in cathepsin V directs the side chain of adjacent K_{119} outwardly away from the active site cleft (Figure 2.5) (122). The extending arms of K_{119} may serve two possible purposes. Firstly, this orientation leads to a clearer passage for the entrance of elastin into the active cleft. Secondly, the lysine side chain acts as a pivotal support that inserts into and holds the elastin in place. In contrast, the side chain of K_{117} of cathepsin L is orientated inwards which may impair a large substrate from entering the active cleft (Figure 2.5) (212). M10, the single point deletion at G_{118} (within exosite 2) supports this hypothesis. This mutant showed a 23% decrease in its elastolytic activity suggesting potential relevance of exosite 2 in interacting with elastin (Figure 2.2B).

The substrate specificity analysis of M11 revealed a highly similar preference to its parent protein cathepsin V except with some loss of specificity in the S3 subsite. It should be noted that none of the exosite 1 and 2 forming residues are a part of the subsite. The reason why these substitutions lead to a less specific selectivity remains elusive. However, since the S3 subsite has been less defined among cathepsins due to its shallow depth, it is possible that significant alteration in the hydrophobicity of the exosite 1 residues may have conformational consequences in the S3 subsite (20 - 30 amino acid residues away each other).

The identification of elastin-binding exosites in cathepsin V may be exploited for target specific inhibitor designing. Besides functioning as a potent elastase, cathepsin V is involved in many physiological activities such as the generation of antigen-presentable CLIP complexes in the immune system and in angiogenesis by releasing endostatin from collagen XVIII (114,116,125). Selective cathepsin V exosite inhibitors may therefore specially block the pathologically relevant matrix degradation by this protease without interfering with its other biological functions.

Chapter 3. Structural and Functional Elucidation of Collagenolytically Active Cathepsin K/GAG Complexes

Introduction

Cathepsin K is the predominant cysteine collagenase expressed in osteoclasts during bone resorption. It is capable of breaking down triple helical collagen without possessing a specific domain such as the so-called "helix unwinding" hemopexin domain known in MMPs. Unbalanced cathepsin K activity has been linked to various bone related diseases. For example, excessive cathepsin K activity can cause bone disorders like osteoporosis and Paget's disease, which are characterized by a global or localized loss of bone mineral density, respectively (140,141). In contrast, lack of cathepsin K activity causes a severe bone resorption defect known as pycnodysostosis. Pycnodysostosis is a rare genetic disease resulting from non-sense, miss-sense and stop codon mutations in the cathepsin K encoding gene (133,143-145). Patients with this disease have a short stature, brittle bones, and several other bone, skin and tooth-related abnormalities (146). Cathepsin K is also implicated in rheumatoid arthritis and osteoarthritis by breaking down aggrecans and fibrillar type II collagen, which are the major components of the ECM of articular cartilage (89).

Type I collagen, the predominant substrate of cathepsin K, represents 90% of the organic matrix in bones and is responsible for its rigidity and tensile strength. Type I tropocollagen is the basic unit of fibrillar collagens and consists of two α1 (I) and one α2 (I) chains, which are covalently cross-linked at four different loci with two in the helical domain and one at each of the telopeptides (21). It contains amino acid sequence repeats following a pattern of G-P-Y or G-X-Hyp, where X and Y can be any other amino acid residue. Tropocollagen can further self-assemble

into fibrils and fibers, which are highly resistant to proteolytic degradation (9,10). Cathepsin K, one of the most efficient collagenases, has thus developed certain substrate specificities to cleave helical collagens. It generally prefers aliphatic substrates containing leucine, isoleucine, and valine residues but also accommodates glycine (124,215). Its S2 binding pocket allows accommodating proline residues (89) and the S3 subsite is also favorable for the binding of glycine residues. These features combined empower cathepsin K with the ability to digest collagen molecules with multiple G-P-Y repeats. However, a structural examination revealed that the diameter of tropocollagen, which is approximately 15 Å, is three times larger than the width of the active cleft of cathepsin K, which is only 5 Å wide (180). This suggests that the active site of cathepsin K cannot provide sufficient space to accommodate intact tropocollagen. The tropocollagen is likely to adopt a denatured conformation mediated by cathepsin K activity to allow the cleavage. The mechanism of this remained elusive.

The crystal structure of cathepsin K does not reveal significant variations in its 3D-fold from either papain or other cysteine cathepsins. Nonetheless, while most cathepsins are negatively charged, cathepsin K has a high density of positive residues located on the surface opposite to the active cleft (89). These positive residues have been demonstrated to interact with bone and cartilage resident GAGs, such as C4-S and DS via electrostatic interactions (135). Such interactions are crucial for the degradation of type I tropocollagen by cathepsin K (65). Under different crystallization conditions, two distinct "beads-on-a-string"- like cathepsin K/GAG structures were resolved (135) (paper submitted for the cathepsin K/DS complex). These "beads-on-a-string"-like organizations were demonstrated to be collagenolytically inactive. Thus, there must exist structurally different cathepsin K/GAG complexes for efficient collagenolysis. We hypothesized a dimer and a tetramer complex model based on the two available crystallographic

structures. Both models, although with different modes of contacts between cathepsin K and GAGs, share a number of important interaction sites. Titration of the collagenase activity demonstrated that the highest activity occurred within a narrow molar cathepsin K:GAG ratio of approximately 2:1, which was supportive for both models. Atomic force microscopy of cathepsin K/GAG mixtures revealed ring-like structures similar to the tetramer model proposed. Mutational analysis further confirmed the likelihood of both models acting individually on either soluble or insoluble collagens. These results implied a possible collagenolytic process in which both dimer and tetramer active complexes function cooperatively *in vivo*.

Materials and Methods

Materials - Benzyloxycarbonyl-Phe-Arg-7-amido-4-methylcoumarin (Z-FR-MCA) was purchased from Bachem (Torrence, CA). The cathepsin inhibitor, E-64 (L-3-carboxy-trans-2-3-epoxypropionyl-leucylamido-(4guanidino)-butane) was purchased from Bio Basic Inc. (Markham, ON, Canada). Calf-skin type I collagen was purchased from Affymetrix, Inc. (Santa Clara, CA). Rat-tail type I collagen, chondroitin sulfate A (C4-S) and dermatan sulfate (DS) were from Sigma-Aldrich (St. Louis, MO).

Construction of Cathepsin K Mutant Expression Vectors - Previously generated human wild-type cathepsin K in pPIC9 vector (Invitrogen, Burlington, ON) served as a template for constructing all mutant vectors. The mutant cDNAs were generated by site-directed mutagenesis using *Pfu*-polymerase (Fermentas, Burlington, ON) with the primers outlined in Table 3.1. The primers were designed to allow either point mutations or small motif swaps in the cathepsin K sequence. The resulting mutant cDNAs were digested with *XhoI* and *NotI* restriction enzymes and

ligated into pPIC9 vector (Invitrogen, Burlington, ON) with the single α-factor replacing the endogenous signal peptides for extracellular secretion. The vectors were linearized with *SacI* prior to electroporation into the *Pichia pastoris*, GS115, using a standard protocol provided by Invitrogen.

Protein Expression - GS115 yeast clones transformed with mutant cathepsin K cDNAs were screened using minimal dextrose and minimal methanol agar plates. Subsequently, the clones were incubated in 5 ml BMGY media containing 1% yeast extract, 2% peptone, 100 mM potassium phosphate (pH 6.0), 1.34% yeast nitrogen base, 4x 10-15% biotin, and 1% glycerol at 30°C until the culture reached an OD_{600} of 2.0-6.0. The cells were then transferred into 3 ml of BMMY media supplemented with 0.5% methanol every 24 hours to induce protein expression. Maximum protein production was typically reached at day 4 of methanol induction and aliquots were tested for activity after pepsin activation as previously described (131). Positive clones were His^+ Mut^+ containing detectable proteolytic activity against synthetic substrate Z-FR-MCA.

For each mutant, the clone exhibiting the highest activity was chosen for large scale expression. The clone was inoculated in 5mL of MD media at 30°C for 24 hours, and followed by further inoculation in 50 mL of BMGY media for 12 hours. The cells were then pelleted and resuspended into 500 mL of BMGY, and allowed to grow to an $O.D._{600}$ of 6.0. The cells were transferred into 1.5 L of BMMY media with induction of 0.5% methanol every 24 hours while aliquots of the supernatant were withdrawn every day to monitor enzyme activity and contamination. The cells were harvested on the 4^{th} day of induction. The supernatant was concentrated 50-fold using an Amicon Ultrafiltration membrane (EMD Millipore, Darmstadt, Germany) to around 30 mL at 4°C.

Mutants	Primer Sequences
Q21A	
Forward	5’- C AAA AAT CAG GGT GCG TGT GGT TCC -3’
Reverse	5'- GGA ACC ACA CGC ACC CTG ATT TTT G -3'
Q92A	
Forward	5’- CA TAT GTG GGA GCG GAA GAG AGT TG -3’
Reverse	5'- CAA CTC TCT TCC GCT CCC ACA TAT G -3'
T101E	
Forward	5’-GAAGAGAGTTGTATGTACAACCCAGAAGGCAAGGCAG-3’
Reverse	5'- CTGCCTTGCCTTCTGGGTTGTACATACAAC TCT CTT C -3'
Q21A/Q92A	
Forward	primers used for Q21A and Q92A
Reverse	primers used for Q21A and Q92A
Q21D/Q92K	
Forward	5’-CAAAAATCAGGGTGAGTGTGGTTCC-3’ & 5’-CATATGTG GGAAAGGAAGAGAGTTG-3’
Reverse	5’-GGAACCACACTCACCCTGATTTTTG-3’ & 5’-CAACTCTCTTCCTTTCCCACATATG-3’
Q21A/Q92A/T101E	*use Q21AQ92A as template*
Forward	5’-GAAGAGAGTTGTATGTACAACCCAGAAGGCAAGGCAG-3’
Reverse	5'- CTGCCTTGCCTTCTGGGTTGTACATACAAC TCT CTT C -3'
M97E/N99E/T101E	
Forward	5’-GAAGAGAGTTGTGAGTACGAACCAGAAGGCAAGGCAG- 3’
Reverse	5’-CTGCCTTGCCTTCTGGTTCGTACTCACAACTCTCTTC-3’
V90E/M97K/T101K/G102Y	*use IFM1 as template*
Forward	5’- GCCTACCCATATGAGGGACAGGAAGAAAGTTGTAAGTAC-3’
Reverse	5’ – GTACTTACAACTTTCTTCCTGTCCCTCATATGGGTAGGC -3’
IFM1	
Forward	5’-GCCTACCCATATGAGGCAACAGAAGAATCCTGTAAGTACAATCCCAAGTATAAGGCAGCTAAA-3
Reverse	5'-TTTAGCTGCCTTATACTTGGGATTGTACTTACAGGATTCTTCTGTTGCCTCATATGGGTAGGC -3'
IFM2	
Forward	5’-CAAAGGATATATCCTCATGGCTAAAGACCGGAGAAACGCCTGTGGCATTGCCAACCTG -3’
Reverse	5'-CAGGTTGGCAATGCCACAGGCGTTTCTCCGGTCTTTAGCCATGAGGATATATCCTTTG-3'
Flanking Primers	
Forward	5’ - GAC TGG TTC CAA TTG ACA AGC- 3’
Reverse	5’ - GCA AAT GGC ATT CTG ACA TCC - 3’

Table 3.1. PCR Primers for Cathepsin K Mutants. A summary of all the primers used in constructing cathepsin K mutants. IFM1 replaced V_{90}GQEESCMYNPTG$_{102}$ with EATQEESCKYNPKY with its homologues sequence of KDRR from cathepsin L. IFM2 replaced

$Q_{172}KGN_{175}$ with its homologues motif ESTESDNN from cathepsin **L.** Nucleotides mutated in each mutant were underlined in both forward and reverse primers.

Protein Activation and Purification - The pH of the concentrated supernatant was lowered to 4.0 using diluted glacial acetic acid (Fisher Scientific, Ottawa, ON), and pepsin (Sigma-Aldrich, St. Louis, MO) (final concentration 0.6 mg/mL) was added to activate the cathepsin K precursor molecule. The mixture was incubated at 37°C and monitored for activity using Z-FR-MCA in 100 mM sodium acetate buffer (pH 5.5, containing 2.5 mM EDTA and 2.5 mM DTT). Upon reaching the maximum activity, the pH was raised to 5.5 to stop pepsin activation. The activated enzyme solution was supplemented with ammonium sulfate to a final concentration of 2 M and centrifuged at 4,000 g for 20 min at 4°C. The resulting supernatant was loaded onto an n-butyl Sepharose column (GE Healthcare, Little Chalfont, U.K.) and washed thoroughly with the loading buffer (100 mM sodium acetate, pH 5.0, 2 M ammonium sulfate, 0.5 mM EDTA, and 0.5 mM DTT). The protein was eluted using a linear gradient with reducing ammonium sulfate concentration. The fraction containing the desired protein was monitored using Z-FR-MCA. Active fractions were combined, concentrated 10-fold using Amicon ultra concentrator (EMD Millipore, Darmstadt, Germany) and buffer exchanged with 100mM sodium acetate buffer (pH 5.5, supplemented with 0.5 mM EDTA and 0.5 mM DTT). The final purified protein was stored at -80 °C for further usage.

Enzyme Concentration Determination and Kinetic Studies – Enzyme concentrations were determined by E-64 active site titration as previously described (216). The excitation and emission wavelengths used for the assays were 380 nm and 460 nm, respectively, using the spectro-fluorometer Perkin-Elmer LB50. Michaelis-Menten constants (k_{cat} and K_m) were

determined using non-linear regression analysis. All enzymes were assayed at diluted concentrations (~ 1-5 nM) with variable Z-FR-MCA concentrations (1-20 µM) in 100 mM sodium acetate buffer (pH 5.5, containing 2.5 mM DTT and 2.5 mM EDTA).

Protein Degradation Assay using Soluble Tropocollagen – Soluble calf-skin type I collagen (0.6 mg/mL) was incubated with 200 nM wild-type or mutant cathepsin K variants, and various concentration of C4-S depending on the assay purposes in sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol, and EDTA. For the antibody inhibitory collagenase assay, various volumes of polyclonal cathepsin K antibody (MS-2) (217) was added to the reaction mixture. The total volume of each reaction was 50 µL and incubated at 28°C. After 4 hours, 1 µL of 100 µM E-64 was added to each reaction to inhibit the residual enzyme activity. The mixtures were kept overnight at 4°C and ran on 8% SDS-PAGE gels. The bands representing the residual amount of collagen after Coomassie staining were assessed using GeneSnap program (Syngene Inc. Frederick, MD) to determine the extent of collagen degradation. All assays were repeated in three independent experiments.

Protein Degradation Assay using Insoluble Collagen – 1 mg insoluble rat-tail type I collagen fibrils were incubated with 400 nM wild-type cathepsin K in 100 mM sodium acetate buffer, pH 5.0, containing 2.5 mM DTT and EDTA for 4 hours at 28˚C. 10 µM E64 was used to terminate the reaction. Subsequently, the reaction mixture was centrifuged for 20 min. The supernatant was run on SDS-PAGE gel and Coomassie brilliant blue (R-250) stained. The band intensities were visualized and analyzed the same as described in the soluble collagenase assay. All assays were repeated in three independent experiments.

Gelatinase Assay – 1.8 mg/mL gelatin (prepared by heat denaturation of soluble calf-skin type I collagen at 95°C for 20 minutes) was incubated with 5 nM wild-type or mutant cathepsin K in sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol at 28°C. After 4 hours, 1 µL of 100 µM E-64 was added to each reaction to inhibit cathepsin K. The mixtures were kept overnight at 4°C and ran on 8% SDS-PAGE gels.

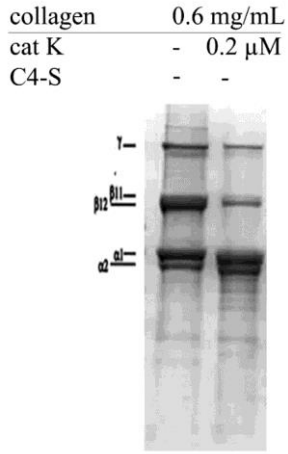
Circular Dichroism (CD) Spectroscopy – Soluble rat-tail type I collagen (0.2mg/mL) and the corresponding proteins or protein/C4-S mixture (200 – 600 nM, E-64 inhibited to prevent collagen degradation) were mixed at 28°C. Samples were subjected to ellipticity measurement (mdeg) in a 1 mm path length quartz cuvette from 190 nm to 250 nm using a J810 (Jasco) spectropolarimeter (Laboratory of molecular biophysics, University of British Columbia). The CD spectra of native collagen alone were also determined at various temperatures maintained by a computer controlled Neslab water bath to study changes in the ellipticity of collagen during thermal denaturation. Ellipticity values were average between triplicates, subtracted from the corresponding controls and plotted against wavelength.

Atomic Force Microscopy (AFM) Study – Specimens in a 10 µL volume were deposited on freshly cleaved mica for 5 minutes, rinsed with Barnsted Nanopure water (Thermo Scientific) and air dried for 30 minutes. Samples were scanned in air in tapping mode with the Cypher scanning AFM (Asylum Research, Santa Barbara, CA). Images (512 lines) were acquired at a scan rate of 3 Hz using silicon tips (Model-AC160TS, Asylum Research, radius of ~7 nm, resonance frequency of 300 kHz, and spring constant of 42 N/m). The background that corresponds to the mica surface in the images were flattened using the 1st-order flattening function provided by Igor Pro 6.22A (Wavemetrics, Inc., Portland, OR) below a threshold set at ~100-300 pm. The particle size and dimensions were measured using the same Igor software.

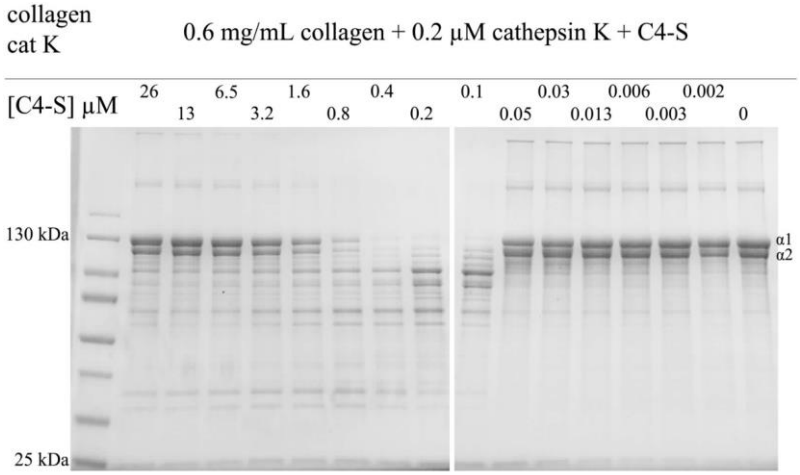
Results

Collagenolytically Active Cathepsin K/GAG Complexes – The collagenase activity of cathepsin K requires the assistance of GAGs such as C4-S and DS to form oligomeric complexes as shown in previous studies (65,135). Without GAGs, cathepsin K is incapable of cleaving tropocollagen within the triple helical domain (Figure 3.1A) due to the insufficient active cleft space (5Å) to accommodate a three times larger tropocollagen molecule (15Å). Collagenase assays performed with a fixed amount of cathepsin K (0.2 μM) revealed that only a small range of GAG concentrations allowed the efficient breakdown of soluble tropocollagen (Figures 3.1B and 3.2A). For both C4-S and DS, such windows existed between 0.1 to 0.4 μM, which could translate into cathepsin K: GAG molar ratios of 2:1, 1:1 and 1:2. Among these ratios, the 2:1 ratio was the most efficient with approximately 5% residual tropocollagen left undigested (represented as densitometric measurement of α1 tropocollagen band) whereas deviations from this ratio became gradually inhibitory (Figures 3.1C and 3.2B).

A



B



C

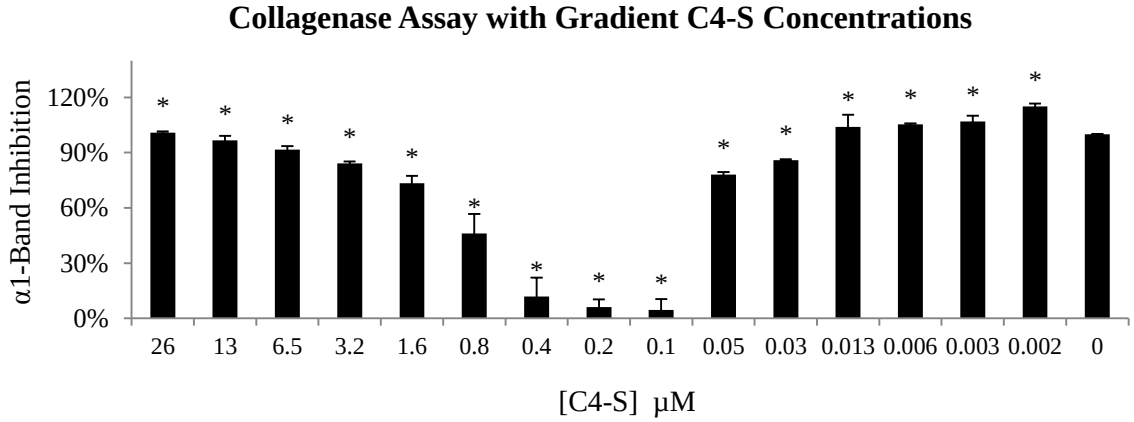
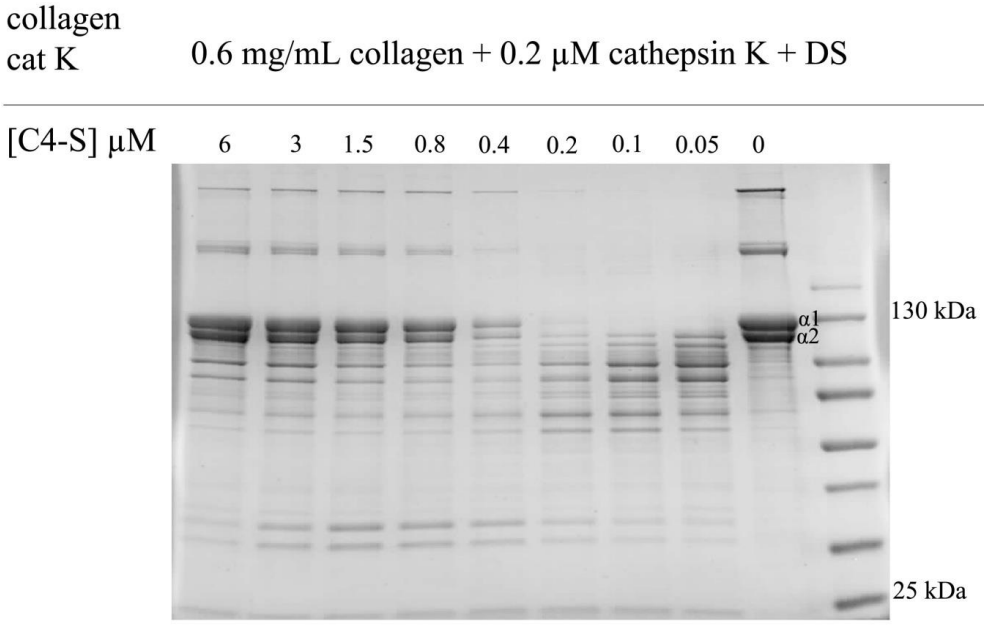


Figure 3.1. Dependency of the Collagenase Activity on the Concentration of C4-S. A. Collagenase Assay of cathepsin K without C4-S. B and C. Titration of collagenase activity of cathepsin K with C4-S. C4-S concentrations varied from 0 – 26 µM. Fixed concentration of 200 nM cathepsin K with 0.6 mg/mL calf-skin tropocollagen was used for each assay at 28 ºC for 4 hours. All assays were repeated in three independent experiments. *P*-value represents unpaired *t*-test statistical analysis relative to sample "[C4-S] = 0 µM", * indicates $P < 0.05$.

A



B

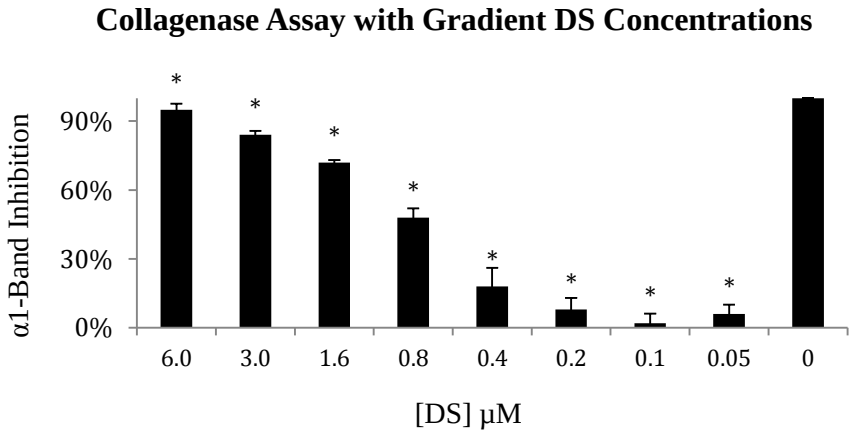


Figure 3.2. Dependency of the Collagenase Activity on the Concentration of DS. A and B. Titration of collagenase activity of cathepsin K with DS. DS concentrations varied from 0 – 6 µM. Fixed concentration of 200 nM cathepsin K with 0.6 mg/mL calf-skin tropocollagen was used for each assay at 28 ºC for 4 hours. All assays were repeated in three independent experiments. *P*-value represents unpaired *t*-test statistical analysis relative to sample "[DS] = 0 µM", * indicates $P < 0.05$.

The molar ratio-dependent collagenolytic efficacy implied the existence of structurally different cathepsin K/GAG complexes. This was confirmed by CD spectroscopy. Titration of C4-S to a fixed amount of cathepsin K (1.8 μM) gave rise to distinct CD spectra (Figure 3.3A). Cathepsin K, containing twisted β-sheets in the right domain and three α-helix in the left domain, exhibited features of both secondary structures (122). It showed two strong double minimum at 222 nm and 208-210 nm, and a stronger maximum at 191-193 nm representing α helix; a single minimum between 210 and 225 nm and a stronger positive maximum between 190 and 200 nm representing β-sheet (218,219). With the α-helix being more predominant, the combined spectrum of cathepsin K displayed a broad CD peak at 208-222 nm and a strong positive CD peak at 190-200 nm. In addition, the minimum at 208-222 nm has a higher intensity than that at 222 nm coinciding with the observation made by Manavalan and Johnson for α + β proteins (220). Upon complex formation with C4-S, CD spectra of the protein, after subtracting their corresponding C4-S controls, developed more β-sheet-like characteristics at lower C4-S concentrations (cathepsin K:C4-S molar ratio below or equal to 2:1) with a single minimum around 225 nm. However, the spectra regained more α-helix features at higher concentrations starting at 1:1 ratio indicated by two minima at 208 nm and 222 nm (Figure 3.3A). The shapes of the spectra were also more similar to that of the cathepsin K alone. These CD results supported the observation that C4-S formed structurally different complexes with cathepsin K at different molar ratios.

CD spectroscopy was also utilized to exam the effect of collagenolytically active cathepsin K/GAG complex on tropocollagen unwinding. Step-wise heat denaturation of tropocollagen served as the reference. Higher temperatures reduced the intensities of both local minimum at 198.8 nm and local maximum at 222.2 nm, resulting in flatter CD spectra (Figure 3.4A). Comparatively, a 300/150 nM cathepsin K/C4-S mixture at 2:1 molar ratio up-shifted the peak

intensity at 198.8 nm 3.2 mdeg and down-shifted 1.2 mdeg at 222.2 nm, whereas a 600/300 nM mixture caused 3.9 mdeg and 1.6 mdeg changes respectively (Figure 3.4B). These shifts in the tropocollagen CD spectra coincided with the trend of heat-induced collagen denaturation, and were supportive to the finding that the cathepsin K/GAG complex at a 2:1 molar ratio was capable of unwinding tropocollagen.

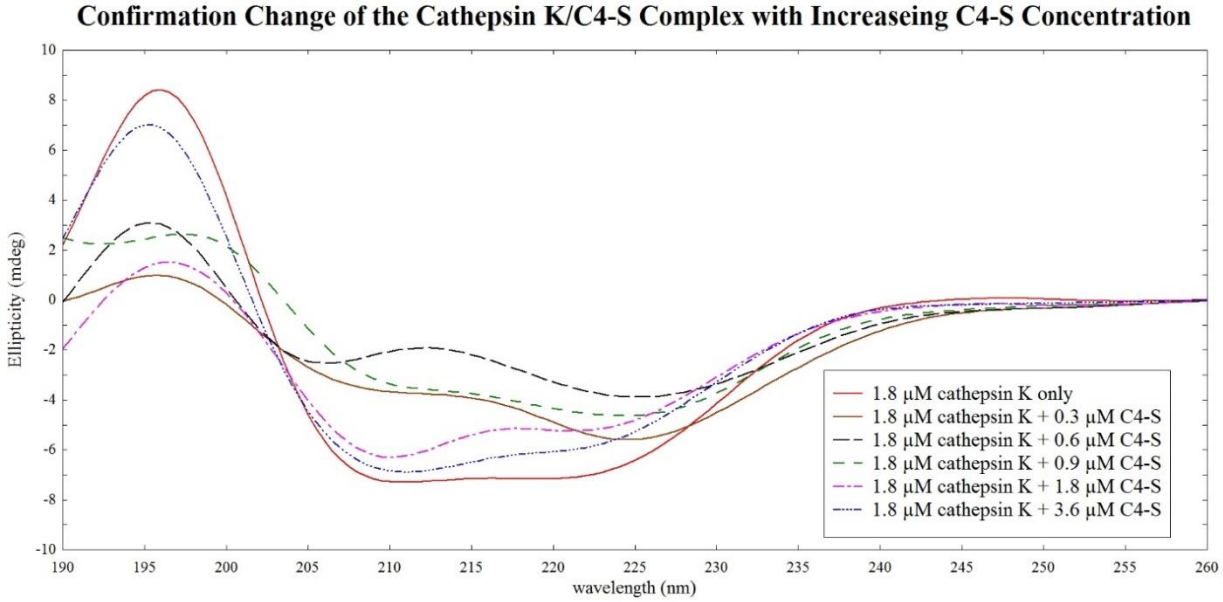
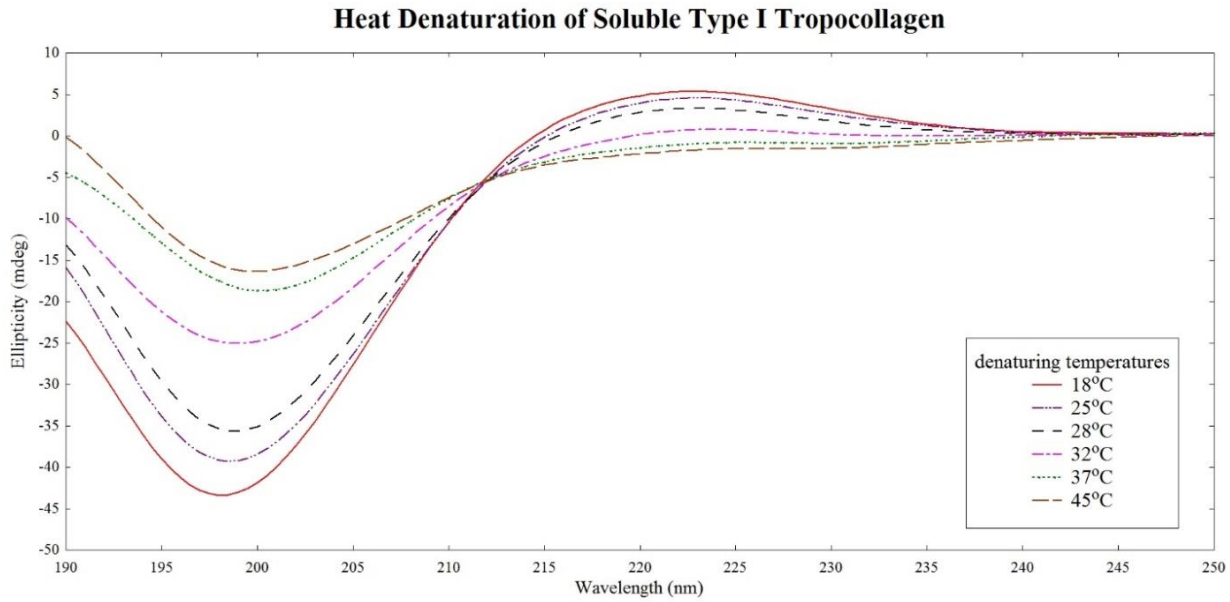


Figure 3.3. Circular Dichroism Spectroscopy Measurements. A. Conformational Changes of the Cathepsin K/C4-S Complex with Increasing Concentrations of C4-S. 1.8 µM cathepsin K (E-64 inhibited) was titrated with 0.3-3.6 µM C4-S. C4-S spectrum was subtracted to derive each spectrum. Measurements were done in three independent experiments.

A



B

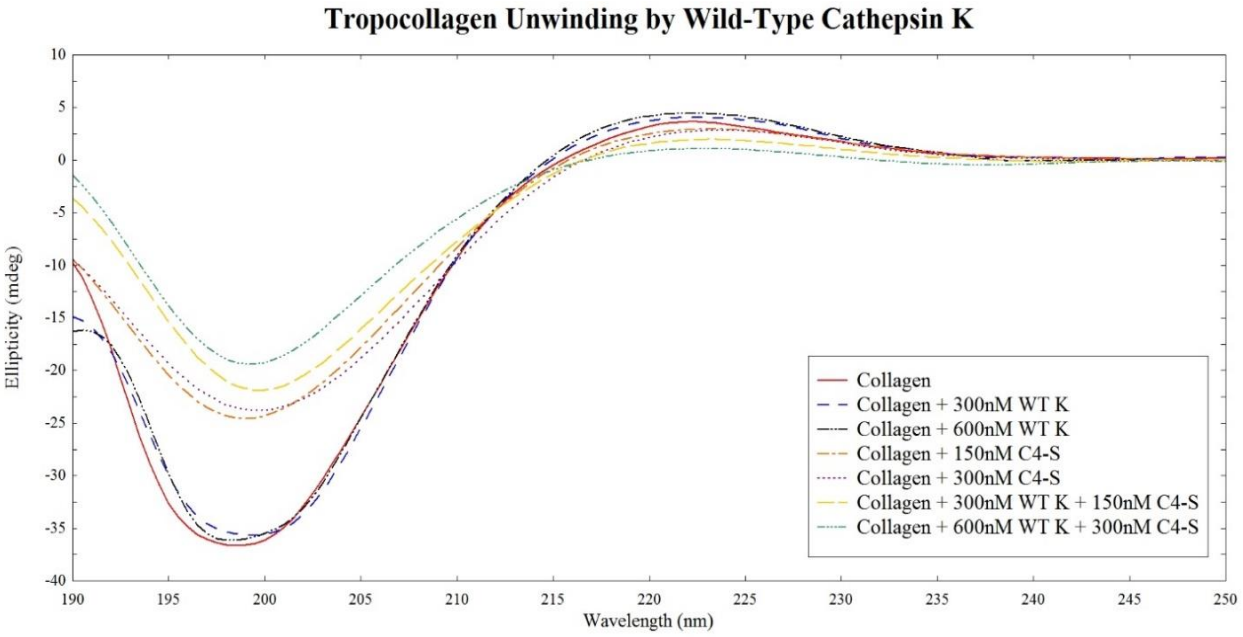


Figure 3.4. Circular Dichroism Spectroscopy Measurements. A. Heat Denaturation of Soluble Type I Tropocollagen. 0.2 mg/mL type I tropocollagen was used. B. Tropocollagen Unwinding by Wild-Type Cathepsin K/C4-S Complex. 300 and 600 nM E-64 inhibited cathepsin K mixed with C4-S in a 2:1 ratio were separately incubated with 0.2 mg/mL tropocollagen at 28ºC. C4-S spectrum was subtracted to derive each spectrum. Measurements were done in three independent experiments.

Because interactions between GAGs and collagen have previously been reported (221-224), concerns regarding a GAG-mediated modifications of tropocollagen resulting in the exposure of a single α-chain for cleavage was raised. To rule out this possibility, cathepsin V, a collagenolytically inactive protease, was used to replace cathepsin K in the collagenase assay. The result indicated that regardless of the amount of C4-S used, cathepsin V was incapable of digesting tropocollagen (Figure 3.5). This rejects the possibility that GAGs could deform tropocollagen as a main mechanism of mediating collagenolysis.

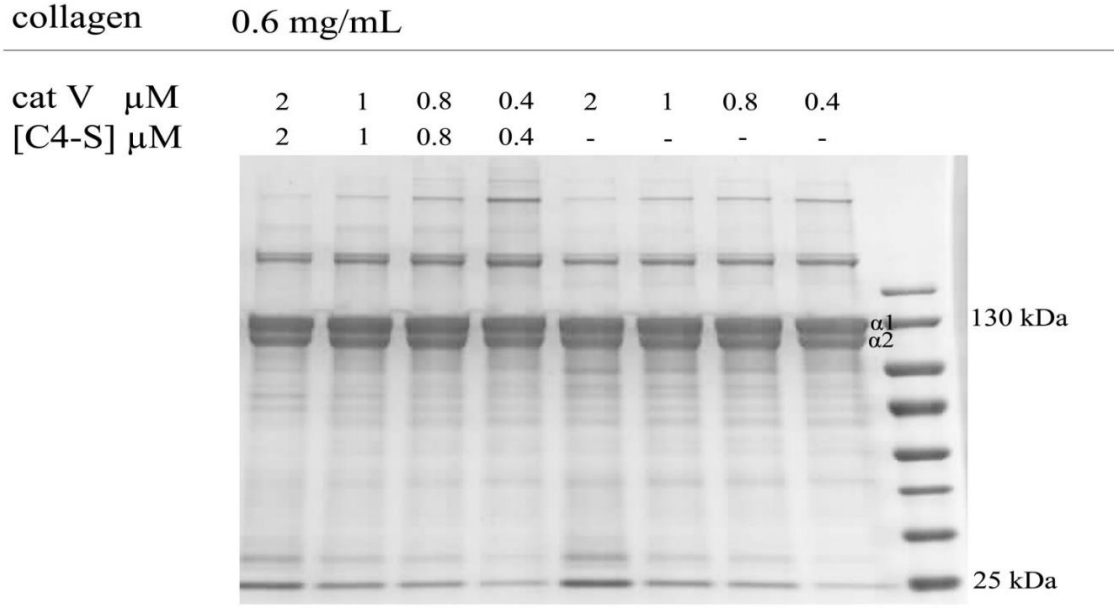


Figure 3.5. Collagenase Assay of Cathepsin V with and without C4-S. Various concentrations of cathepsin V titrated with and without C4-S. 0.6 mg/mL calf-skin tropocollagen was used for each assay at 28 ºC for 4 hours. Assays were repeated three time in three independent experiments.

Models of the Collagenolytically Active Cathepsin K/GAG Complexes – Two different crystal structures have been solved prior to this study, namely a cathepsin K/C4-S complex and a

cathepsin K/DS complex (DS structure, paper submitted) (135). Both structures exhibit a "beads-on-a-string"-like pattern with multiple cathepsin K molecules attached to a long chain of GAG but exploiting different GAG binding sites. Excess molar cathepsin K concentrations over that of GAGs as present in the beads-on-a-string structures was found to be inhibitory as demonstrated by the collagenase assay with varying GAG concentrations. Nevertheless, both crystallographic structures provided critical information on protein-protein and protein-GAG interactions, and led to the proposal of two collagenolytically active cathepsin K/GAG complexes. The dimer as well as the tetramer model conform to a molar ratio of 2:1 between cathepsin K and GAG to achieve the highest collagenase activity.

As shown in Figures 3.6A and 3.6C, the first complex model was built based on the published cathepsin K/C4-S crystal structure (135). It is a tetramer consisting of symmetrically inverted dimer cathepsin K molecules (indicated by the same color) each bridged by a single chain of GAG. This unique ring-shape organization gave rise to a central pore with three openings. The overall complex was held together by forces including protein-protein interactions between cathepsin K molecules and ionic interactions between cathepsin K and GAG molecules. In terms of the protein-protein interaction, a number of amino acid residues contributed significantly to the stabilization of the tetramer model. These included region V_{90}-G_{102} with mainly M_{97}, N_{99} and T_{101} at protein interface 1 (IF1) (Figure 3.6B, Protein Interface I Zoom), and residues K_{106}, R_{108}, S_{138}, L_{139}, and T_{140} at protein interface 3 (IF3). IF1 was previously described as the loop interface by Cherney *et al*. to be energetically favorable due to hydrophobic interactions (225). In terms of the ionic interaction, residues E_6, R_8, K_9, K_{10}, Q_{172}, N_{190}, and K_{191} provided direct contacts with the repeating disaccharides of the GAG molecule, which were previously elucidated by Li *et al*. (135). Amino acid residues Q_{21} and Q_{92} from the diagonally positioned cathepsin K molecules

(approximately 4 Å apart) extended their side chains into the center of the tetramer complex (Figure 3.6B, Central Pore Zoom), dividing the central pore into three openings of approximately 17.3 Å, 17.5 Å, and 17.3 Å in diameter. This arrangement allowed the active clefts of two cathepsin K in the ring pointing into the central pore, whereas the other two faced outwards (Figure 3.6C).

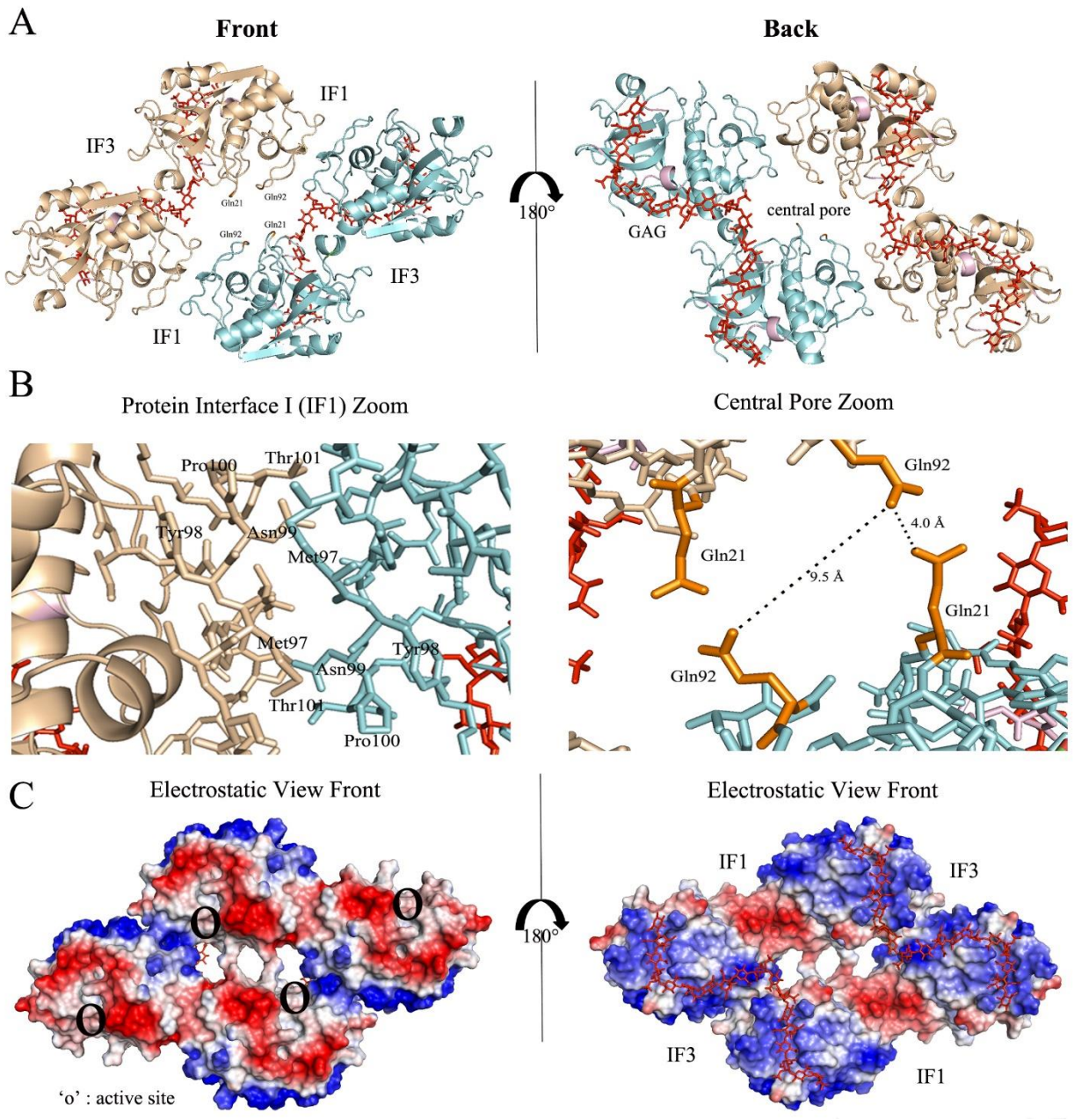


Figure 3.6. Tetramer Model of the Collagenolytically Active Cathepsin K/GAG Complex. A. Cartoon view of the cathepsin K/GAG complex tetramer model. B. Magnification showing the central pore formed by the tetramer model and protein interaction interface (IF1) between cathepsin K molecules on different GAG chains. C. Electrostatic surface presentations of the tetramer complex. IF3 represents protein interaction interface between two adjacent cathepsin K molecules residing on the same GAG chain.

The second cathepsin K/GAG complex model is a dimer, which was constructed based on the cathepsin K/DS crystal structure (paper submitted). It contained only two protein molecules with one GAG chain where the protease molecules are aligned in a head-to-tail organization (Figures 3.7A & 3.7C). This dimer is largely dependent on protein-protein interaction based on amino acid residues M_{97}, N_{99} and T_{101} of one cathepsin K molecule and residues K_{173}, G_{174}, Lys_{176}, N_{199} and K_{200} of the second cathepsin K molecule. This protein interaction was defined as protein interface 2 (IF2) (Figure 3.7B). It is important to note that IF2 of the dimer model shared the region V_{90} to G_{102} with IF1 of the tetramer model where the two proteins are positioned in a head-to-head organization. The GAG chain mainly associated with the latter molecule via the backbone of residues A_1 and P_2, and side chains of residues P_{114}, N_{117}, K_{119}, and R_{123}. Additionally, the side chain of Q_{92} from the distant cathepsin K molecule also came in contact with the GAG, which might aid in the stabilization of the overall complex (Figure 3.7B). The active clefts of the two proteins were positioned in opposite directions with the one from the first molecule residing near by the GAG chain (shown as circle in Figure 3.7C).

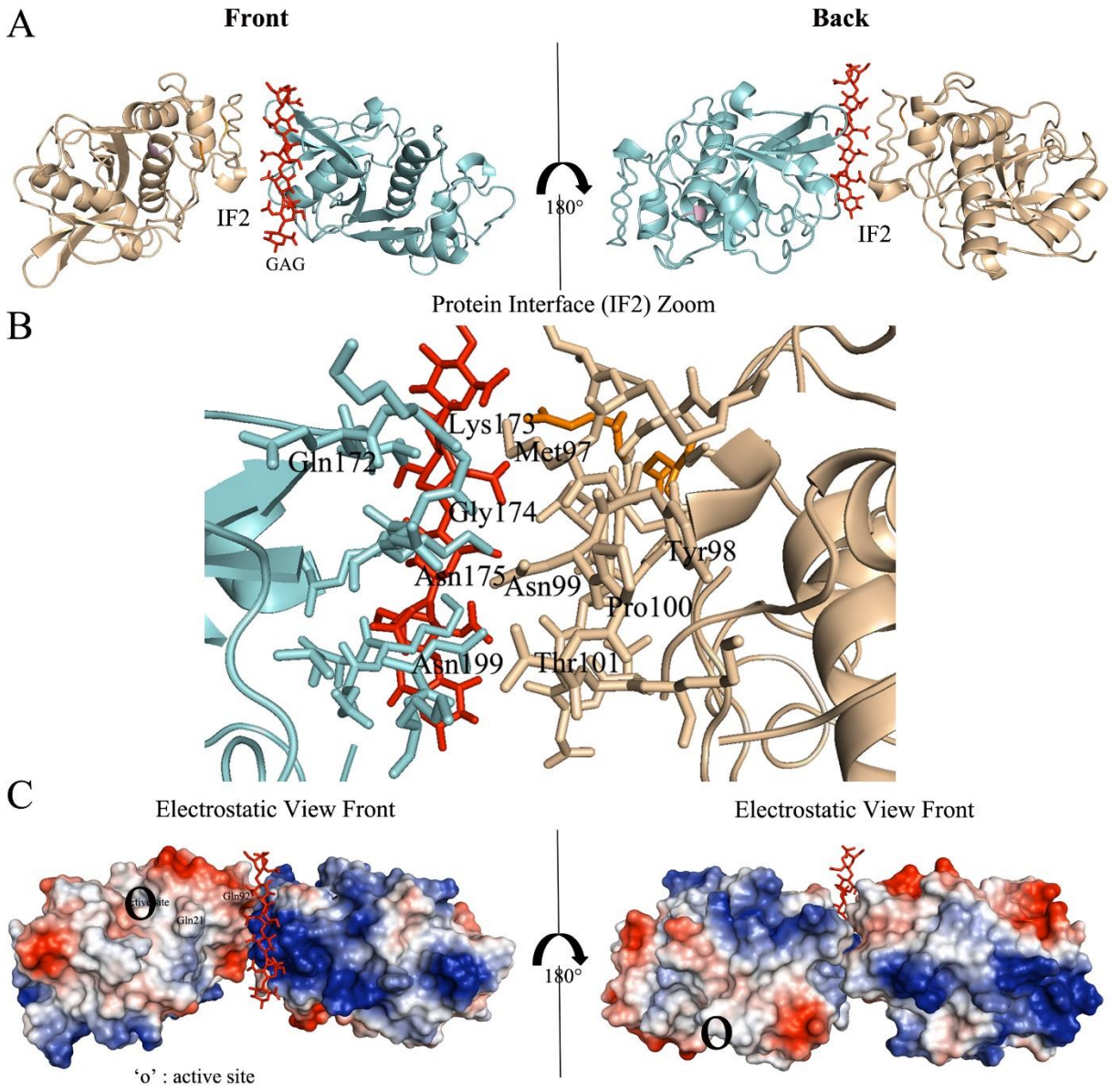


Figure 3.7. Dimer Model of the Collagenolytically Active Cathepsin K/GAG Complex. A. Cartoon view of the cathepsin K/GAG complex dimer model. B. Magnification showing the protein interaction interface (IF2) between cathepsin K molecules. C. Electrostatic surface presentation of the dimer complex. Only one cathepsin K molecule is directly bound to GAG whereas the other one interacts with the formal via its protein dimer interface. GAG chain is colored in red.

Mutational Analysis of Protein-Protein-Interface Sites – To validate both models, protease variants carrying mutations in selected amino acid residues that are involved in tetrameric and dimeric cathepsin K complexes were generated. Both models utilized a common region of V_{90}-G_{102} at their protein-protein interaction interfaces (IF1 for tetramer and IF2 for dimer). Thus, four mutants were created: single mutant T101E by mutating T_{101} to glutamic acid, M97E/N99E/T101E by mutating M_{97}, N_{99}, and T_{101} to glutamic acid to create a repulsive interface at IF1; V90E/M97K/T101K/G102Y and IFM1 by mutating V_{90}, M_{97}, T_{101} and G_{102} to EKKY and swapping $V_{90}G_{91}Q_{92}E_{93}M_{97}T_{101}G_{102}$ with EATQEKKY, respectively (work done with IFM1 attributed to Dr. Vidhu Sharma). The latter two mutants were generated by substituting the appropriate residues by amino acids present in cathepsin L. Cathepsin L does not form a complex nor does it exhibit a collagenase activity (65).

In terms of model specific residues, the potential involvement of Q_{21} and Q_{92} at the central pore of the tetramer model, and K_{173}, G_{174}, K_{176}, N_{199} and K_{200} at IF2 of the dimer model were also evaluated. More specifically, Q_{21} was mutated to either alanine or glutamic acid, and Gln_{92} was mutated to alanine or lysine. This gave rise to four mutants namely Q21A, Q21A, Q21A/Q92A, Q21D/Q92K. The Q21A/Q92A variant was designed to eliminate the dividing side chains in the central pore and the Q21D/Q92K variant to stabilize these dividers by ionic bridges. In addition, a triple mutation variant Q21A/Q92A/T101E was created to speculate on the additive effect of the three residues. At IF2, $Q_{172}KGN_{175}$ was swapped with its orthologous ESTESDNN from cathepsin L, and denoted IFM2 aiming to disrupt the protein-protein interaction at IF2.

Mutations selected were distant from the active site and were not expected to affect the overall peptidase activity and integrity of the mutants. This was supported by enzyme kinetic studies. As indicated in Table 3.2, all mutants did not significantly alter the turnover rate (k_{cat}), the

$K_{m,}$ and thus the k_{cat}/K_m values for the hydrolysis of the synthetic fluorogenic peptide substrate, Z-FR-MCA, from those of the wild-type protease. The gelatinase assay as an alternative method was also used to assess the active site functionality of the protease variants. All variants efficiently degraded gelatin at 28C° after 4 hours with an efficacy comparable to wild-type cathepsin K (Figure 3.8A)

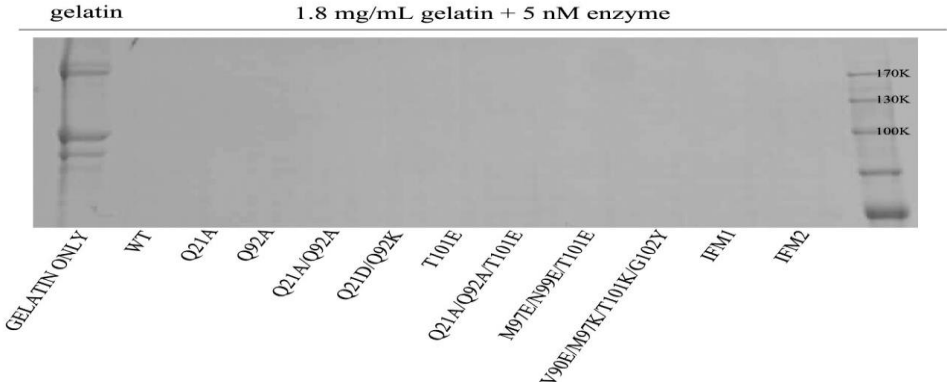
Enzyme	k_{cat} (S^{-1})	K_m (μM)	k_{cat}/K_m (10^6 $S^{-1}\cdot M^{-1}$)
wild-type	19.5 ± 1.0	7.5 ± 1.0	2.6 ± 1.4
Q21A	23.4 ± 0.6	8.2 ± 0.5	2.9 ± 0.8
Q92A	16.7 ± 0.9	7.7 ± 1.0	2.2 ± 1.3
T101E	19.4 ± 1.7	9.8 ± 1.9	2.0 ± 2.5
Q21A/Q92A	16.7 ± 0.8	8.3 ± 1.0	2.0 ± 1.3
Q21D/Q92K	15.7 ± 0.8	7.6 ± 0.5	2.1 ± 0.9
Q21A/Q92A/T101E	20.8 ± 1.9	8.9 ± 1.9	2.3 ± 1.9
M97E/N99E/T101E	18.6 ± 1.2	8.6 ± 1.4	2.2 ± 1.8
V90E/M97K/T101K/G102Y	21.2 ± 1.3	9.7 ± 1.0	2.2 ± 1.6
IFM1	16.2 ± 1.5	6.1 ± 0.3	2.7 ± 1.5
IFM2	20.4 ± 3.1	5.5 ± 2.4	3.7 ± 4.0

Table 3.2. Kinetic Values for the Peptidase Activity of Cathepsin K. The constants K_m, k_{cat}, and k_{cat}/K_m were determined according to Michaelis-Menten kinetics with procedures described in the Experimental Procedures. Varying concentrations of Z-FR-MCA as fluorogenic substrate were used. The K_m is measured in μM, k_{cat} in s^{-1}, and k_{cat}/K_m was calculated accordingly. IFM1 replaced $V_{90}G_{91}Q_{92}E_{93}M_{97}T_{101}G_{102}$ with EATQEKKY from cathepsin L. IFM2 replaced $Q_{172}KGN_{175}$ with its homologues sequence ESTESDNN from cathepsin L. Values were representatives of three independent experiments.

The stability of all variants was also addressed. Measured at room temperature, wild-type cathepsin K and the variants remained 3 - 9% residual activity after 4 hours at pH 5.5. In the presence of GAGs such as C4-S, enzyme stability was greatly enhanced to approximately 20 - 30% for all enzymes (Figure 3.8B). Mutant M97E/N99E/T101E was approximately 16% and 3% less

stable than the wild-type enzyme with and without C4-S, respectively. Introduction of three negatively charged glutamic acids might have compromised the stability of the enzyme.

A



B

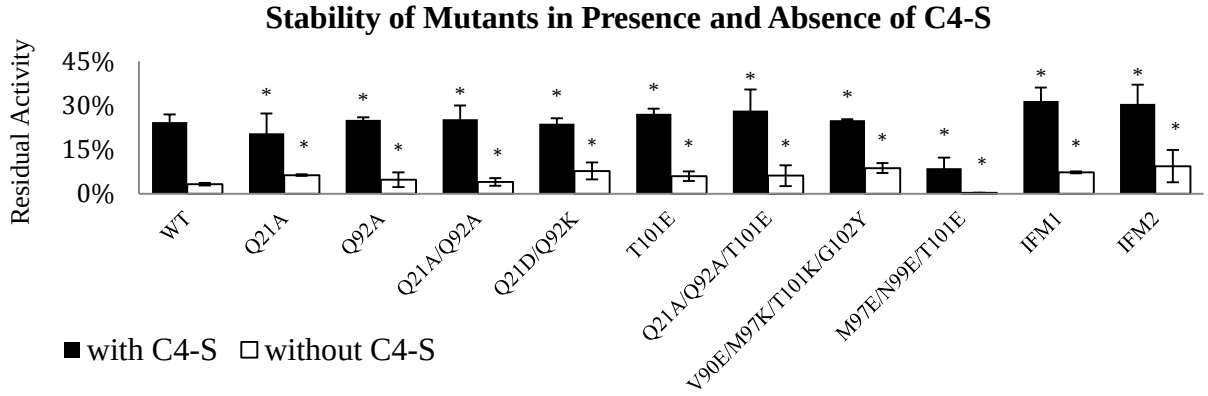


Figure 3.8. Stability of Complex Mutants and Gelatinase Assay. A. Gelatinase assay of all mutants. Gelatin was prepared by heating calf-skin collagen at 95ºC for 20 minutes. Assays were carried out with 1.8 mg/mL gelatin and 5 nm enzyme at 28ºC for 4 hours. B. Stability of mutants in the presence and absence of C4-S. The stabilities of all mutants were tested against synthetic substrate Z-FR-MCA at excitation and emission wavelengths of 380 nm and 460 nm, respectively. Samples were tested every 30 minutes at 28ºC and the residual activities after 4 hours were plotted. IFM1 replaced $V_{90}G_{91}Q_{92}E_{93}M_{97}T_{101}G_{102}$ with its homologues sequence EATQEKKY from cathepsin L. IFM2 replaced $Q_{172}KGN_{175}$ with its homologues sequence ESTESDNN from cathepsin L. All assays were done in triplets in three independent experiments. *P*-value represents unpaired *t*-test statistical analysis relative to WT, * indicates $P < 0.05$.

The collagenase activities of the variants were determined using both soluble type I tropocollagen and insoluble collagen fibers. Measured as the residual percentage of the α-1 band on coomassie blue stained SDS-PAGE gel, wild-type cathepsin K in the absence of C4-S barely degraded tropocollagen, while supplemented with C4-S allowed the nearly complete degradation (Figures 3.9). Variants Q21A, Q92A, and Q21A/Q92A at the central pore of the tetramer model did not exhibit noticeable inhibition in collagenase activity whereas double mutant Q21D/Q92K had 12% ± 1% loss in activity than that of the wild-type (Figure 3.9).

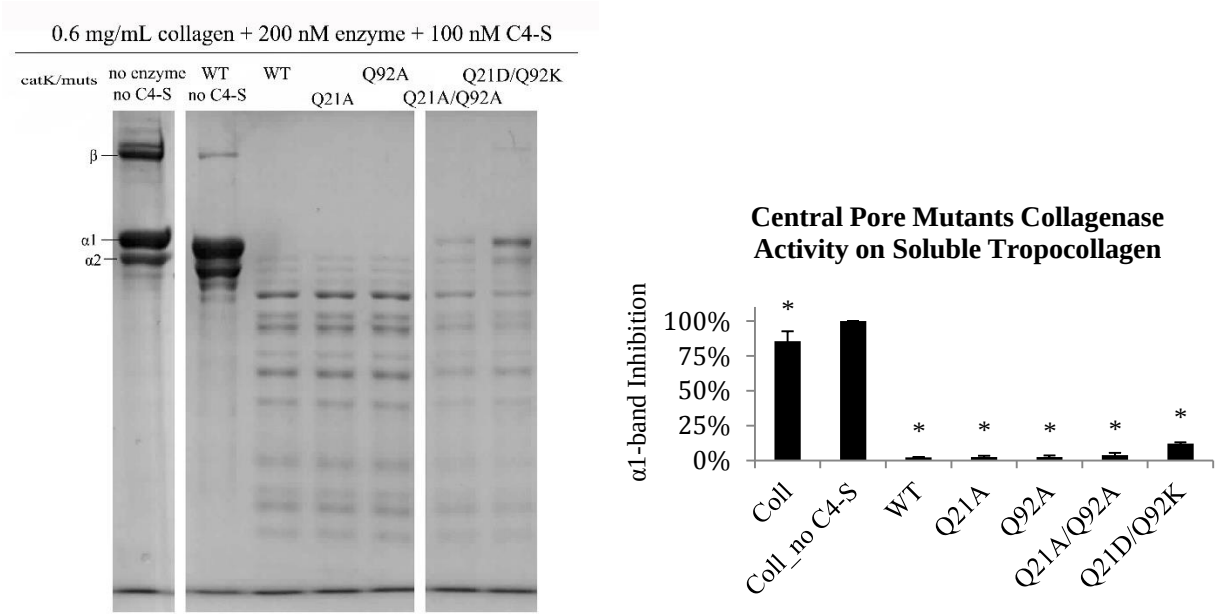
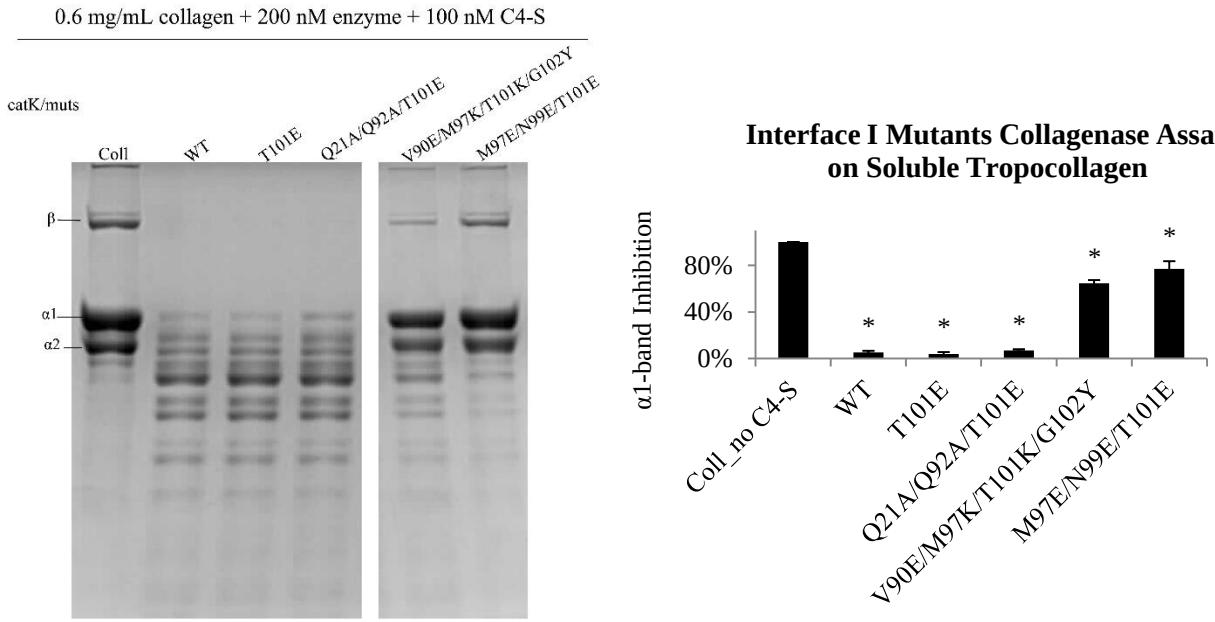


Figure 3.9. Collagenase Assay of Central Pore Mutants. Collagenase assay of central pore mutants on soluble tropocollagen created according to the tetrameric model (Q21A, Q92A, Q21A/Q92A and Q21D/Q92K) and their corresponding collagenolytic activity quantifications. Densitometry of α-bands on SDS-PAGE gels was used to determine the collagenolytic activity. Enzymes used for soluble collagenase assay were titrated to have a 200 nM and supplemented with 100 nM C4-S to ensure a 2:1 ratio for maximal activity. 1 mg mouse tail collagen fiber was used for the insoluble collagen degradation assay. Assays were repeated in three independent experiments at 28ºC for 4 hours. *P*-value represents unpaired *t*-test statistical analysis relative to sample "Coll_no_C4-S", * indicates $P < 0.05$.

In terms of the involvement of the common region V_{90}-G_{102} at IF1 of the tetramer and IF2 of the dimer, variants T101E and Q21A/Q92A/T101E revealed comparable strength to the wild-type (Figures 3.10A). Triple variant M97E/N99E/T101E resulted in a 77% ± 6% reduction in the collagenase activity, and the quadruple variant, V90E/M97K/T101K/G102Y, lessened by 65 ± 3% activity (Figures 3.10A). The weaker stability of variant M97E/N99E/T101E might have contributed to its higher loss of activity. Variant IFM1, encompassing the entire V_{90}-G_{102} region, completely diminished its collagenolytic strength, and showed a substantial accumulation of the α-1 collagen band on the SDS-PAGE gel due to telopeptide cleavage from α- and α- tropocollagen without digesting the triple helical domain (Figure 3.10B). On the other hand, variant IFM2, which as specifically created to disrupt the dimeric protein-protein interaction, displayed a nearly equal activity to wild-type cathepsin K in the presence of C4-S towards soluble tropocollagen (Figure 3.10B).

A



B

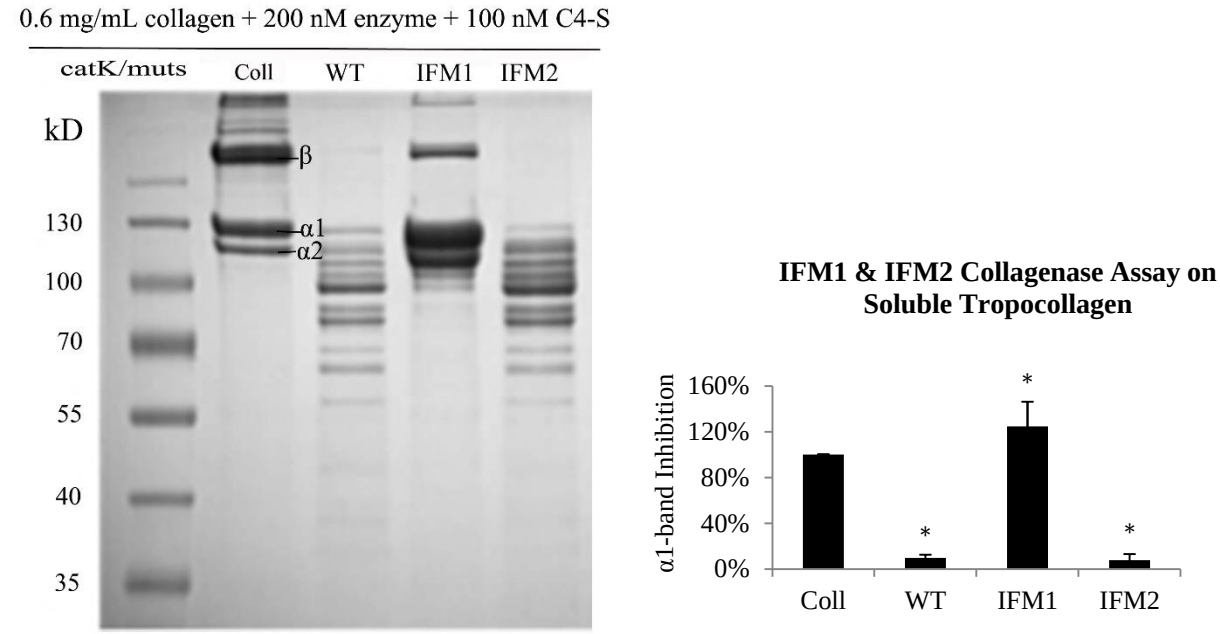


Figure 3.10. Collagenase Assay of Interface I and II Mutants. A. Collagen assay for protein interface I (IF1) mutants on soluble collagen and their corresponding collagenolytic activity quantifications. IF1 was shared by both the tetrameric and dimeric model. B. Collagenase assay using soluble collagen of IFM1 and IFM2 and their corresponding collagenolytic activity quantifications. IFM1 replaced $V_{90}G_{91}Q_{92}E_{93}M_{97}T_{101}G_{102}$ with its homologues sequence

EATQEKKY from cathepsin L. IFM2 replaced $Q_{172}KGN_{175}$ with its homologues sequence ESTESDNN from cathepsin L. Densitometry of α-bands on SDS-PAGE gels was used to determine the collagenolytic activity. Enzymes used for soluble collagenase assay were titrated to have a 200 nM and supplemented with 100 nM C4-S to ensure a 2:1 ratio for maximal activity. Assays were repeated in three independent experiments at 28ºC for 4 hours. *P*-value represents unpaired *t*-test statistical analysis relative to sample "Coll_no_C4-S"/ "Coll", * indicates $P < 0.05$.

Representative variants including IFM1, IFM2, V90E/M97K/T101K/G102Y and Q21D/Q92K were also tested against insoluble mouse tail collagen fibrils. As indicated in Figure 3.11, supernatants of insoluble collagen fibrils alone displayed no trace of soluble tropocollagen on the gel. Wild-type cathepsin K was able to liberate soluble tropocollagen from the fibrils, whereas IFM1, IFM2, and Q21D/Q92K failed to do so. In contrast, variant V90E/M97K/T101K/G102Y retained nearly 50% ± 5% of the wild-type strength.

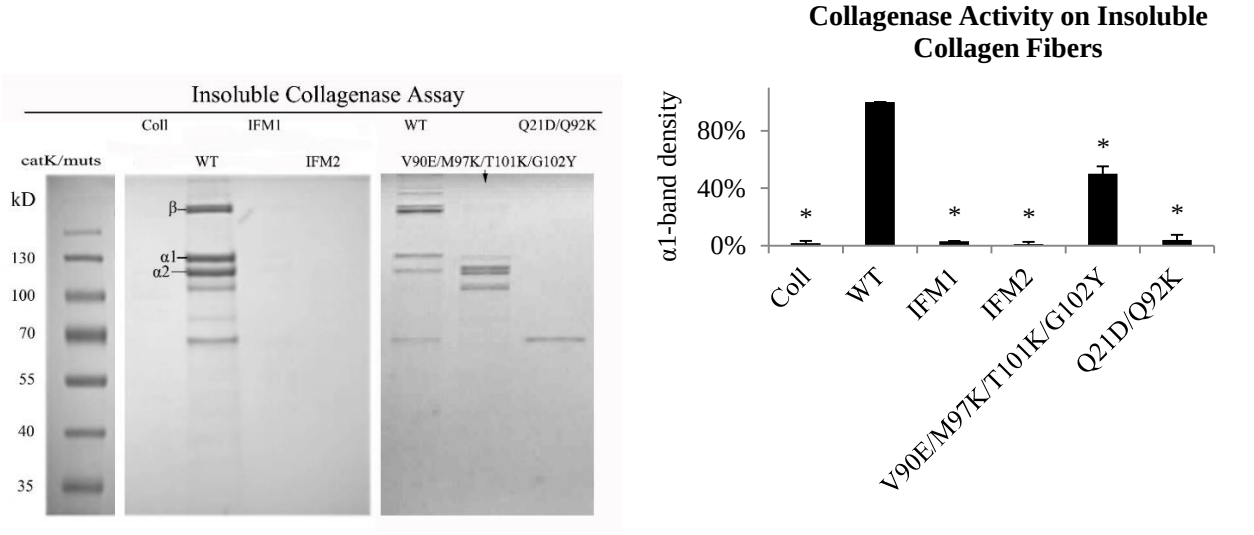


Figure 3.11. Insoluble Collagenase Assay of Complex Mutants. A. Insoluble collagen fibers degradation assay with IFM1, IFM2, V90E/M97K/T101K/G102G and Q21D/Q92K. B. Corresponding collagenolytic activity quantification of the Mutants. IFM1 replaced $V_{90}G_{91}Q_{92}E_{93}M_{97}T_{101}G_{102}$ with its homologues sequence EATQEKKY from cathepsin L. IFM2 replaced $Q_{172}KGN_{175}$ with its homologues sequence ESTESDNN from cathepsin L. Densitometry

of α-bands on SDS-PAGE gels was used to determine the collagenolytic activity. Enzymes were titrated to be 400 nM. 1 mg mouse tail collagen fiber was used for the insoluble collagen degradation assay. Assays were performed in three independent runs at 28ºC for 4 hours. *P*-value represents unpaired *t*-test statistical analysis relative to sample "Coll", * indicates $P < 0.05$.

Morphological Study of Cathepsin K/GAG Complexes with AFM - AFM was selected to provide atomic surface contour maps of the complexes. The images revealed that single cathepsin K molecules (E-64 inhibited to prevent auto-degradation) appeared as oval shaped monomers with a height of 3.3 ± 1.9 nm and width of 4.3 ± 1.2 nm (Figure 3.12 and Table 3.3). Addition of a limited amount of C4-S (8:1 molar ratio of cathepsin K to C4-S) resulted in complexes (height: 3.8 ± 1.1 nm; width: 4.0 ± 1.4) that were structurally similar to the "beads-on-a-string" organization found in the crystal structure of the cathepsin K/C4-S complex (Figure 3.12 and Table 3.3). A decrease of the molar ratio of both components to 2:1 resulted in ring-shaped assemblies with a height of 3.8 ± 1.7 nm and width of 14.6 ± 2.4 nm, which were morphologically similar to the putative tetramer (Figure 3.12 and Table 3.3). Under an excess C4-S condition of 1:8, single particles (height: 3.6 ± 1.0 nm; width: 4.0 ± 1.4 nm) resembling monomeric cathepsin K molecules were observed again. The size measured was the average of all dimensions, and the high standard deviation was the result of this averaging method. In addition, heights determined with AFM were typically smaller than of those of the crystal structures due to dehydration effect in air tapping mode. The size measurements were corrected based on a calibration curve generated with various well-defined protein molecules (Appendices Figure A1).

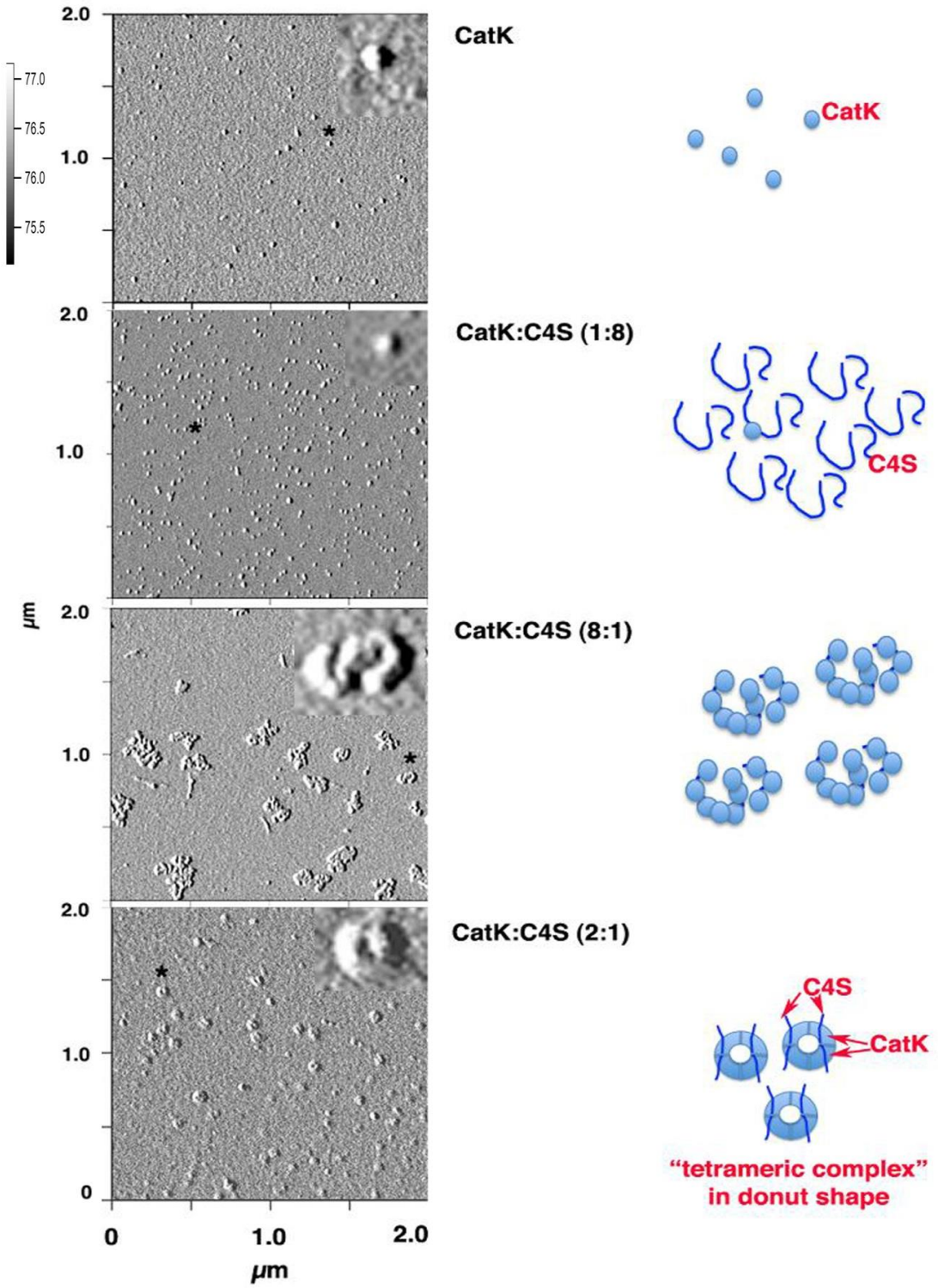


Figure 3.12. Atomic Force Microscopy Images of Cathepsin K/C4-S Mixtures at Different Molar Ratios of Cathepsin K and C4-S. The spatial organization of cathepsin K in the presence of various concentrations of C4-S. The structures changed from a single molecule distribution at

excess C4-S, to a beads-on-a-string-like organization at excess of cathepsin K, to a donut-like shape at a molar ratio of 2:1 for cathepsin K and C4-S. The right panel showed the schematic interpretation of the images. The upper right inserts are electronic magnifications of the particles indicated by *. The right panel shows the schematic interpretation of the images. Samples were scanned in air in tapping mode with the Cypher scanning AFM. Images (512 lines) were acquired at a scan rate of 3 Hz using silicon tips (Model-AC160TS, Asylum Research, radius of ~7 nm, resonance frequency of 300 kHz, and spring constant of 42 N/m).

AFM Imaging of Cathepsin K/C4-S Complex with Soluble Collagen Substrates – Collagen alone was determined to be approximately 1.5 ± 0.6 nm in diameter using AFM. When cathepsin K was added to tropocollagen in the absence of C4-S, protease molecules bound to the ends of tropocollagen. This is in agreement with the known telopeptide activity of cathepsin K (138) (Figure 3.13). The cross-sectional height and width of tropocollagen with cathepsin K were determined to be 3.7 ± 1.2 nm and 4.8 ± 1.2 nm, respectively (Table 3.3). At the molar ratio of 2:1 (cathepsin K to C4-S), protease molecules had a strong affinity for the helical domain of tropocollagen. AFM images suggest ring-shaped structures residing consecutively on the tropocollagen surface. Their morphology, height of 4.4 ± 1.2 nm and width of 12.0 ± 4.8 nm, led to the speculation that tropocollagen docked into the central pore of multiple tetramer cathepsin K/GAG complexes (Figures 3.13).

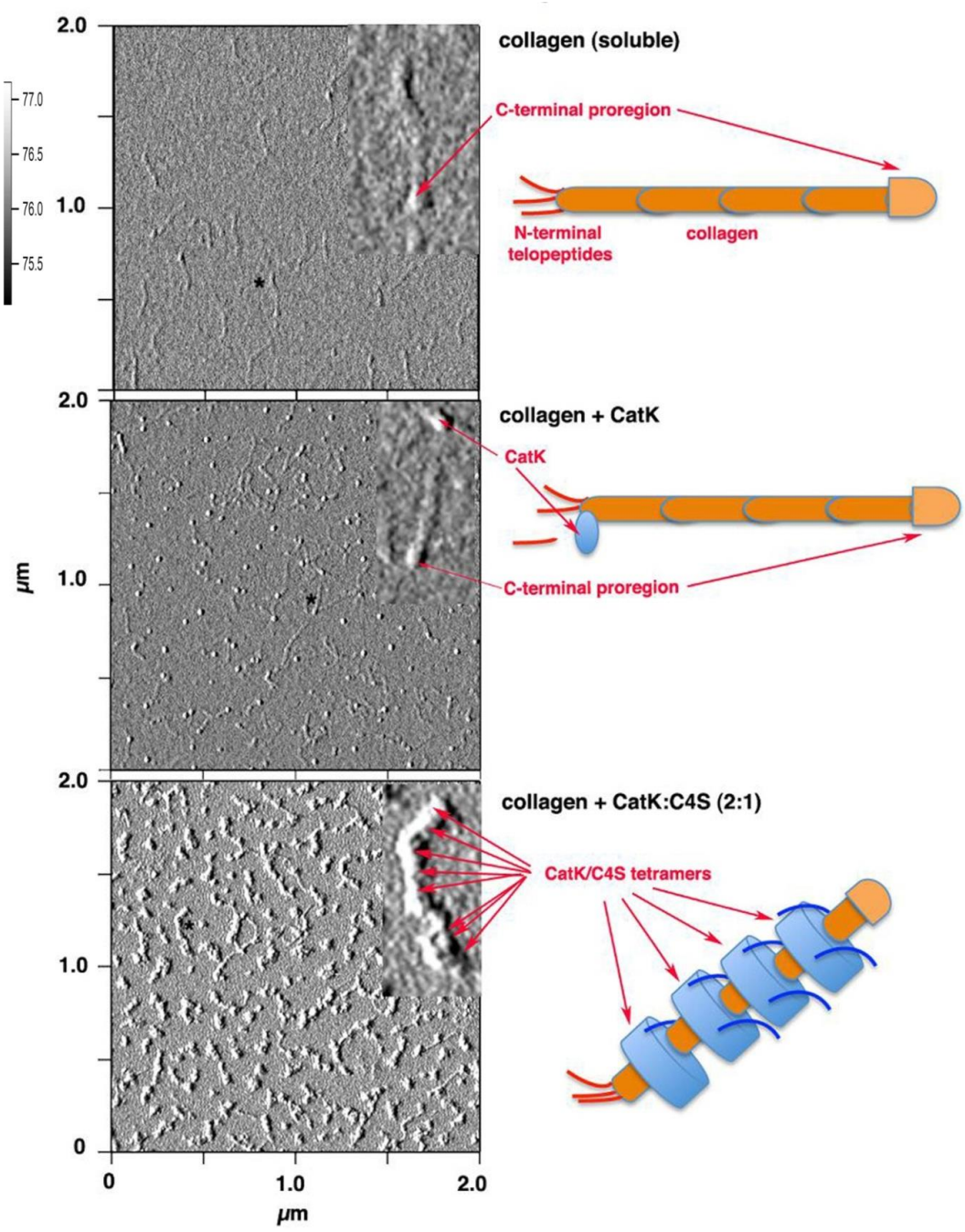


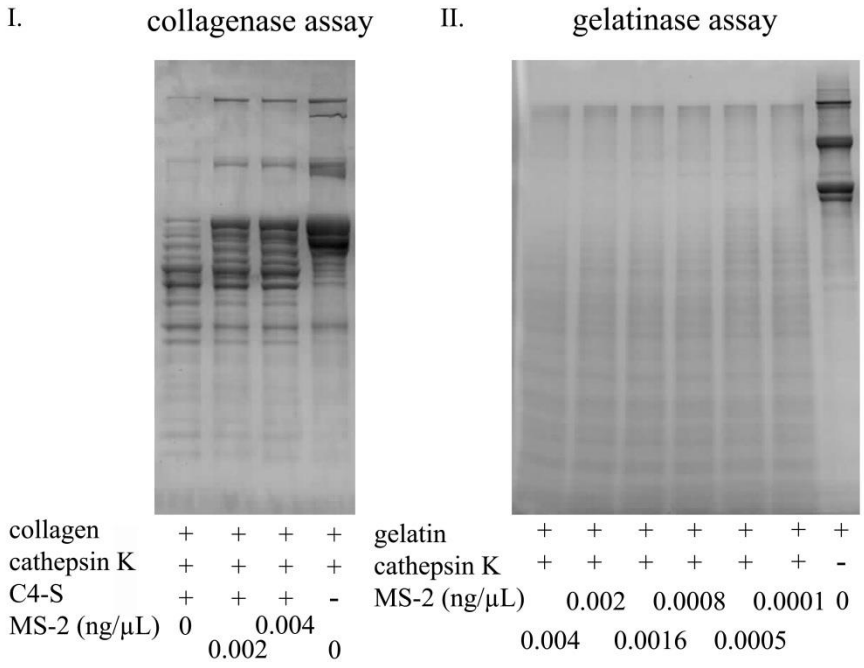
Figure 3.13. Atomic Force Microscopy Images of Active Cathepsin K/C4-S Mixture with Tropocollagen. Cathepsin K alone and cathepsin K/C4-S 2:1 molar ratio mixture interacting with tropocollagen. The upper right inserts are electronic magnifications of the particles indicated by *. The right panel shows the schematic interpretation of the images. Samples were scanned in air in tapping mode with the Cypher scanning AFM. Images (512 lines) were acquired at a scan rate of 3 Hz using silicon tips (Model-AC160TS, Asylum Research, radius of ~7 nm, resonance frequency of 300 kHz, and spring constant of 42 N/m).

	Unit: nm			
	Height	**Width**	**Height (corrected)**	**Width (corrected)**
cathepsin K alone	0.6 ± 0.3	18 ± 5	3.3 ± 1.9	4.3 ± 1.2
cathepsin K/C4-S 1:8 M/M	0.8 ± 0.1	17 ± 6	3.6 ± 1.0	4.0 ± 1.4
cathepsin K/C4-S 8:1 M/M	0.8 ± 0.1	17 ± 5	3.8 ± 1.1	4.0 ± 1.4
cathepsin K/C4-S 2:1 M/M	0.8 ± 0.2	63 ± 10	3.8 ± 1.7	14.6 ± 2.4
Collagen	0.16 ± 0.03	10 ± 4	1.5 ± 0.6	1.5 ± 0.6
collagen + cathepsin K alone	0.8 ± 0.1	20 ± 5	3.7 ± 1.1	4.8 ± 1.2
collagen + cathepsin K/C4-S 2:1 M/M	1.1 ± 0.1	50 ± 20	4.4 ± 1.2	12.0 ± 4.8

Table 3.3. Size Measurements from Atomic Force Microscopy analysis of Cathepsin K Alone, in Various Complexes with C4-S and with Collagen. Size determination of proteins have their limitations due to the hydration status of the proteins on the mica surface and the size of the AFM tip. Because the samples were imaged in air in tapping mode, standard curve was constructed to reduce dehydration effect of small protein molecules as suggested by Kang R. Cho. *et al*. (A Multistage Pathway for Huamn Prion Protein Aggregation in Vitro: From Multimeric Seeds to β-Oligomers and Nonfibrillar Structures, J.Am. Chem. Soc 2011, 133: 8586-8593). Albumin, carbonic anhydrase and cathepsin L were served as standards for the construction of the correction curve for the determination of heights. The factors for non-globular collagen were calculated based on the measurements and the known diameter of tropocollagen and are of a comparable magnitude of order as determined for the globular proteins. The width correction factors are listed as: Lateral correction factor for Cathepsin K was based on x-ray molecular dimensions of 3.5x3.5x5.5nm; Lateral correction factor for tropocollagen was based on tropocollagen diameter of 1.5nm.

Antibody Inhibition of Collagenolytic Activity - The polyclonal antibody, MS-2 (133), was used to study its effect on the collagenase activity of cathepsin K. The gelatinase assay in the presence of MS-2 was first used to rule out the possibility that the antibody blocks the active site of cathepsin K. Cathepsin K at 5 nM concentration was able to completely degrade 1.8 mg/mL of gelatin in the presence of 0.001 – 0.004 ng/µL of MS-2 suggesting that the antibody did not interfere with the proteolytic function of the enzyme (Figure 3.14A). In contrast, the degradation of collagen was inhibited by 25% and 44% when compared to the condition in the absence of the antibody. The enzyme concentration used in the collagenase assay was 200 nM and the antibody concentrations 0.002 ng/µL and 0.004 ng/µL MS-2, respectively (Figures 3.14A & 3.14B). This implied that MS-2 inhibited the collagenolytic activity of cathepsin K by hindering the formation of a collagenolytically active cathepsin K/GAG complex without impeding the functionality of its active site.

A



B

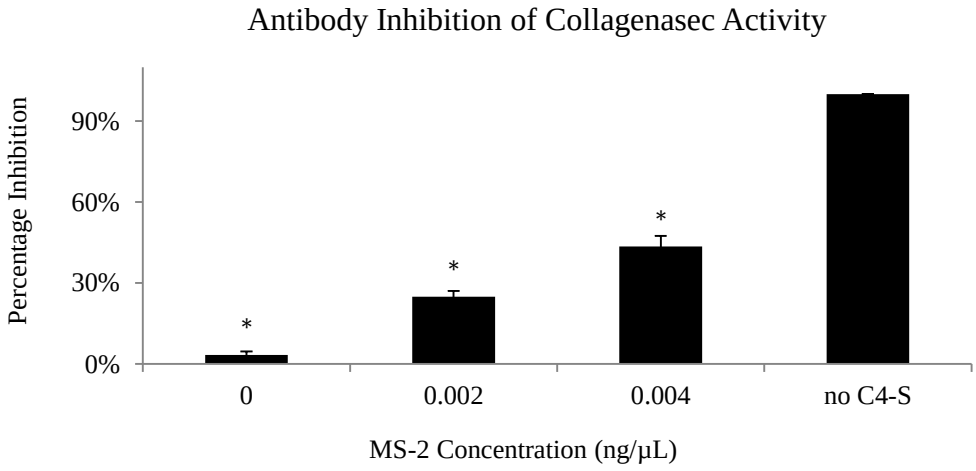


Figure 3.14. Antibody Inhibition of Collagenase Activity of Cathepsin K. A. Collagenase assay with two different concentrations of MS-2 antibody. B. Gelatinase assay with various concentrations of MS-2 antibody. B. Quantification of antibody inhibition. MS-2 is a polyclonal rabbit antibody against human cathepsin K (133). Assays were repeated in three independent experiments. *P*-value represents unpaired *t*-test statistical analysis relative to sample "no C4-S", * indicates $P < 0.05$.

Discussion

Cathepsin K, predominantly expressed in osteoclasts, exhibits an exceptional potency against type I collagen, which constitutes nearly 90% of the organic content of the bone. Earlier studies using gel filtration and fluorescence polarization have revealed that cathepsin K binds to GAGs such as C4-S (226). This interaction leads to the formation of a complex with an approximate size of 200 – 300 kDa and is crucial for its collagenase activity (65). Two recently solved crystal structures of cathepsin K with C4-S (135) and DS (paper submitted) provided detailed information on the configuration of the cathepsin K/GAG complexes. Both structures have a "beads-on-a-string"-like organization, of which multiple cathepsin K molecules bind to one single strand of GAG (135). Although different amino acid residues are involved in the GAG binding of both structures, GAG binding mostly occurs via ionic interactions at sites remote from the active cleft. This may explain why complex formation does not interfere with the efficacy and specificity of the enzyme regarding non-collagenous substrates (135). Interactions between cathepsin K and GAGs were further studied by CD spectroscopy. Cathepsin K and C4-S could form structurally different complexes at different molar ratios. The protein CD spectra revealed more β-sheet-like characteristics at lower C4-S concentrations but regained more α-helix features at cathepsin K/C4-S ratios beyond equimolar. This corresponds to a shift in complex formation: increasing concentrations of C4-S restore the CD spectrum back to that of free cathepsin K as more GAG molecules compete with each other to disrupt the protein-protein interactions that hold the complex together. These changes in protein CD spectra reflect mainly the protein/C4-S interactions upon complex formation rather than structural changes of cathepsin K as the crystal structures indicated that binding of C4-S does not cause any conformational changes within the protein (135,227).

Besides with cathepsin K, GAGs also interact with collagen molecules. According to Yang *et al*., GAGs tend to alter either the fibrillar diameter or mesh size of the collagen matrix, which leads to changes in the mechanical properties of the collagen-GAG scaffolds (161). This is supported by a study carried out by Raspanti *et al*. (223) stating that most GAGs directly influence the size, shape, and aggregation of the collagen fibrils. Scanning electron microscopy images revealed the binding of GAG molecules to the fibril surface in a highly specific manner to form regular, periodic interfibrillar links. Nonetheless, the association of GAGs with collagen is to reinforce rather than to unwind or denature collagen as is evident by the incapability of cathepsin V to cleave tropocollagen in the presence of C4-S.

The association of GAGs with both collagen and cathepsin K may be of great importance for the collagenase activity of cathepsin K. During bone resorption, GAGs not only function as chaperons anchoring cathepsin K onto the collagen fiber, but also participate in complex formation with cathepsin K to breakdown triple helical collagen. Here, cathepsin K/GAG complexes are likely the structural basis for collagen unwinding and thus an equivalent to the hemopexin domain in MMPs. Furthermore, since bone resorption is a relatively slow process, the activity of cathepsin K should be tightly regulated to prevent undesired bone degradation as found in osteoporosis. Such regulation can be accomplished by controlling the GAG abundance. As demonstrated by the collagenase assay, the collagenolytic potency of cathepsin K was greatly dependent on the concentrations of GAGs. At a 2:1 cathepsin K/GAG molar ratio, the strongest activity was observed against tropocollagen whereas deviations from this ratio appeared to be inhibitory. This may have important implications in the regulation of cathepsin K activity and may explain certain bone pathological phenotypes where the accumulation of GAGs is the cause. Mucopolysaccharidoses (MPS) are autosomal recessive lysosomal storage diseases characterized

by the excessive accumulation of GAGs. For example, patients with MPS I develop short stature with severe problems in bone growth similar to those with pycnodysostosis, which is an autosomal recessive disease caused by deletion of the cathepsin K gene (165). We have previously demonstrated that excessive GAGs in the growth plate of femur of MPS I mice, osteoclastic cathepsin K activity is inhibited (165). This supports the role of GAGs in regulating the collagenase activity of cathepsin K.

The collagenolytically inactive beads-on-a-string cathepsin K/GAG complex crystal structures served as a foundation for the proposed collagenolytically active dimer and tetramer models where only direct protein-protein and protein-GAG interactions were considered. In terms of ionic interaction, the tetramer model retained amino acid residues R_8, K_9, K_{10}, Q_{172}, N_{190}, K_{191} and L_{195} as contact sites with the GAG chain. Previous results indicated that mutations created by substituting these residues with their orthologous residues from cathepsin L resulted in an altered C4-S association with cathepsin K accompanied by a significant loss of the collagenase activity towards tropocollagen without impairing the hydrolytic activities against gelatin and the synthetic substrate, Z-FR-MCA (225). These results indicated that ionic interactions incorporated in the tetramer model are not crystal artifacts but real interactions. The dimer model, on the other hand, derived from the cathepsin K/DS crystal structure, revealed alternative binding sites on cathepsin K for GAGs. These sites, mostly comprising amino acid residues A_1, P_2, Q_{92}, P_{114}, N_{117}, K_{119}, and R_{123}, have also been proven exploitable for C4-S (unpublished data of Bromme Laboratory). In terms of protein-protein interaction, variants created in V_{90}-G_{102} region, such as V90E/M97K/T101K/G102Y, M97E/N99E/T101E and IFM1 diminished 65%, 77% and 100% of the collagenase activities respectively, confirming the involvement of this region in the formation of a collagenolytically active complex. Coincidently, residue Y_{98}, which is known to cause

pycnodysostosis when mutated into a cysteine residue, is also located within this region (65,133). Mutation of this residue to cysteine caused a significant reduction in protein electrophoretic mobility in the presence of C4-S, indicating a potential loss of protein-protein interaction (65).

Model specific variant IFM2 at IF2 of the dimer model revealed interesting substrate preferences for the active cathepsin K/GAG complex. The finding indicated that IFM2, by swapping $Q_{172}KGN_{175}$ with its orthologous ESTESDNN from cathepsin L, retained 100% collagenase strength against soluble tropocollagen, but lost its activity against insoluble collagen fibers. Another variant, Q21D/Q92K also exhibited similar preference towards soluble (active) and insoluble collagen (inactive). Note that Gln_{92} is also a crucial residue interacting with the GAG chain in the dimer model (see Figure 3.7). These results lead to the speculation that the dimer model has a preferential substrate specificity towards insoluble collagen. Disruption of the dimer formation only impairs the hydrolysis of insoluble collagen without diminishing its activity towards soluble tropocollagen. The tetramer is thus selectively capable to hydrolyze soluble collagen. This explains the contrasting observation found in IFM1 and IFM2, both of which reside at IF2 of the dimer model: IFM1 was incapable of cleaving both soluble and insoluble collagen because its mutation simultaneously disrupted the protein-protein interaction at IF1 of the tetramer and IF2 of the dimer. Variant IFM2, which only disturbed the IF2 interaction of the dimer, fully maintained its collagenolytic activity towards soluble tropocollagen.

To further rationalize the validity and substrate preference of the tetramer model, a triple helical peptide was docked onto the complex. As shown in Figure 3.15, the peptide docks into the central pore of the tetramer with each of the two $\alpha 1$ chains plants into the adjacent enzyme active sites. No obvious hindrance is found as the three openings of the central pore provide sufficient space for each peptide chain. Although the docking experiment suggested that amino acid residues

Q_{21} and Q_{92} might be involved in the unraveling of the triple helical peptide, mutagenesis experiments did not reveal a significance of their roles. Nonetheless, the docking provides some supports to the substrate preferences of the complexes. The tetramer requires the substrate to be positioned in the central pore. It is sterically hindered to form around a tightly packed collagen fiber. In comparison, the size and structural configuration of the dimer allows it to better dock onto a collagen fibril and to cleave off soluble tropocollagen from the outer surface for further digestion by the tetramer.

A schematic diagram (Figure 3.16) illustrates the hypothesized collagenolysis process based on the findings of this study. The release of tropocollagen-sized fragments from insoluble collagen fibers is realized by cathepsin K dimer binding to GAGs associated with the collagen fiber. Released tropocollagen molecules are subsequently degraded by cathepsin K tetramer complexes.

Because the collagenase activity of cathepsin K depends on its complex formation with GAGs, a disruption of these interactions may represent a novel approach to selectively inhibit its collagenase activity without affecting the general hydrolytic function of the enzyme. Although high salt concentrations and negatively charged molecules such as oligonucleotides and polyglutamic acids can efficiently impede the collagenolytic activity of cathepsin K, they are not practical in treating diseases related to cathepsin K (65,228). In this project, a polyclonal antibody against cathepsin K has been demonstrated to be effective in selectively preventing collagen from being degraded. This may lead to a potential development of antibody therapy using target specific monoclonal antibodies. Furthermore, small molecule therapy targeting the GAG binding site on cathepsin K or the protein-protein interactions in the complexes might represent a novel and effective approach to specifically inhibit the collagenase activity of his protease in osteoporosis.

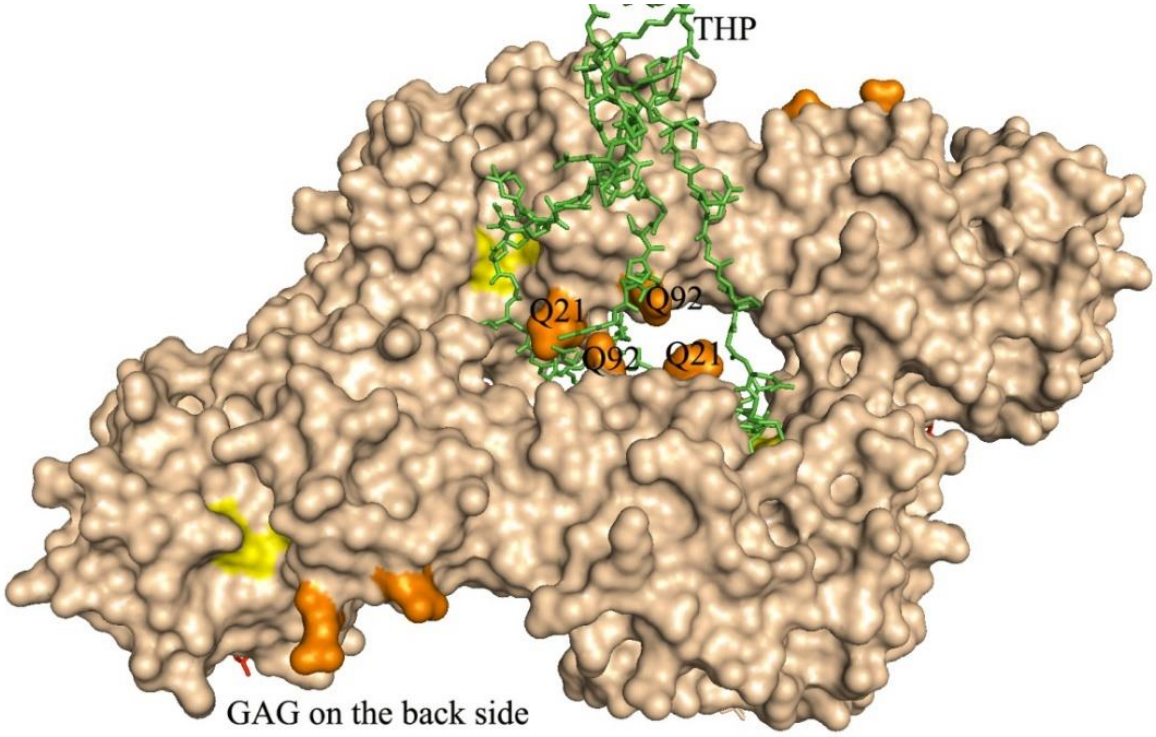


Figure 3.15. Docking Simulations of a Tropocollagen Fragment onto Cathepsin K/GAG Tetramer Complex Model. Putative docking simulation of tropocollagen with the tetramer model. Tropocollagen is colored in green. The active sites of cathepsin K is highlighted in yellow whereas Gln_{21} and Gln_{92} residues are colored in orange.

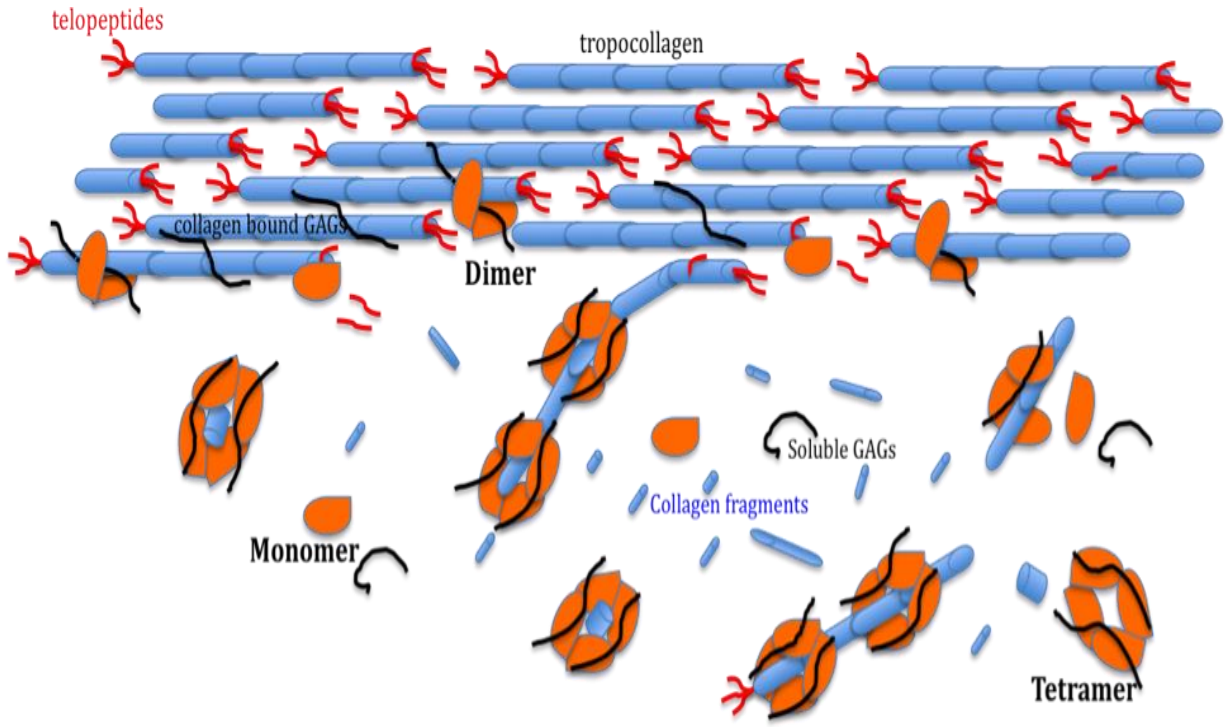


Figure 3.16. Model for Fibrillar Collagen Degradation by Cathepsin K/GAG Complexes. There are three potential modes of collagen degradation by three forms of cathepsin K: 1. Monomeric cathepsin K cleaves off telopeptides and hydrolyzes small non-helical collagen peptides. 2. Dimeric cathepsin K forms at collagen bound GAGs on the surface of the fibril and releases large tropocollagen fragments from the fibril. 3. Tetrameric cathepsin K forms around loose tropocollagen fragments with the help of soluble GAGs and hydrolyses them. It may also degrade smaller triple helical fragments.

Chapter 4. Identification of Exosite Inhibitors of Cathepsin K as Selective Collagenase Blockers

Introduction

Osteoporosis is a prevalent disease characterized by the loss of bone density leading to an increased risk of bone fractures. The bone loss associated with osteoporosis is caused by an imbalance in bone remodeling which consists of a two-step process: bone resorption and bone formation (229). Resorption involves the demineralization of inorganic bone components and stimulation of osteoclasts to create erosion cavities by removing the organic extracellular matrix. Bone formation results from the synthesis of organic matrix by osteoblasts to replace the resorbed bone. Currently, there are a number of therapeutic treatments for osteoporosis, such as bisphosphonates, estrogens, selective estrogen receptor modulators, and parathyroid hormones that target different mechanisms of the bone remodeling process (230). However, various side effects have been described in patients on these medications (231-233). Debates have arisen on the usage of alendronate, a type of bisphosphonate in causing osteonecrosis of the jaw and atypical femoral fractures (231,234). Estrogen hormone replacement therapy has been linked with increasing rates of breast cancer, heart attacks, strokes, and blot clot formations (151). Although the reasons behind these side effects are not fully understood, unspecific targeting by these medications in multiple biochemical pathways is highly probably (235).

Taking these side effects into consideration, an alternative approach that involves inhibiting cathepsin K has drawn attention. Cathepsin K is known to be the predominant collagenase produced by osteoclasts (96). After acidic demineralization, cathepsin K is secreted from the ruffled border into the resorptive pit for the degradation of the collagen matrix. The

organic extracellular matrix of bone is composed of 90% type I collagen. Excessive cathepsin K activity augments collagen breakdown resulting in osteoporosis, whereas lack of its activity can cause severe osteosclerosis known as pycnodysostosis (144). In contrast to existing drugs that focus on altering the upstream participants in the bone remodeling pathway, the inhibition of cathepsin K has three distinct advantages. First, its predominant expression in osteoclasts allows for an effective drug targeting and second, as the final executioner of the pathway that directly acts on collagen, it minimizes side effects to be expected from upstream-acting inhibitors. Third, inhibition of cathepsin K does not affect osteoclast viability and osteoclastogenesis and thus may not interrupt osteoclast-osteoblast communication which is thought to be critical for healthy bone.

A number of pharmaceutical companies have focused their efforts on the development of cathepsin K inhibitors for the therapeutic treatment of osteoporosis. However, at least two compounds failed in clinical trials. The main reasons for their downfall lied in intolerability and extraskeletal side effects (235). For instance, Balicatib developed by Novartis failed in phase II of a clinical trial due to skin morphea side effects (155). In contrast, odanacatib a nitrile peptidomimetic derivative developed by Merck & Co. successfully passed a large patient number phase III clinical trial (156,236). Nonetheless, the safety of Odanacatib still remains uncertain in selected areas as indicated by the Data Monitoring Committee who is responsible for evaluating patient safety and treatment efficacy (https://www.biopharmamarket.com/@offers/news/view/6472). In addition to this statement, recent studies have linked cathepsin K to neurobehavioral malfunctions where cathepsin K deficiency can induce changes in the central nervous system that are associated with learning and memory deficits (185). The involvement of cathepsin K in cytokine secretion in bone marrow dendritic cells and in bone osteolysis also implicates a role in autoimmune and inflammatory

diseases (186-188). Moreover, cathepsin K has been demonstrated to play a crucial role in TGF-1ß metabolism in lung and we have shown that cathepsin K deficiency leads to air way abnormalities (217). Taken together, these findings convey the message that cathepsin K is more than just a collagenase and has multiple other biological functions. Although drugs targeting the active site of cathepsin K can significantly lessen collagen degradation, they will also prevent the physiologically degradation/processing of other cathepsin K substrates. Therefore, the discovery of inhibitors specific for the collagenase activity of cathepsin K alone would be highly beneficial.

The collagenolytic activity of cathepsin K is accomplished by forming high molecular weight complexes with C4-S (135,225). A narrow range of a molar cathepsin K to C4-S ratio (~2:1) was demonstrated to have the highest collagenolytic efficacy. The cathepsin K/C4-S complex involves not only charge interactions between cathepsin K and C4-S, but also protein-protein interactions. Based on this knowledge, a library of 1280 known compounds was screened using a high throughput fluorescence polarization assay to identify inhibitors that obstruct the complex formation without impairing the active site functionality. A total of 15 hits was identified and together with a number of other anti-resorptive compounds were further studied to confirm their selective anti-collagenase activity on the complex. These exosite inhibitors of cathepsin K can therefore serve as starting templates for future drug design to treat osteoporosis without affecting other functional pathways.

Materials and Methods

Materials – The inhibitors used in this project include artemisinin, aurintricarboxylic acid, baicalein, (S)-(+)-camptothecin, dihydrotanshinone I, ellipticine, emodin, (-)-epigallocatechin gallate, exemestane, genistein, genistin, nafoxidine hydrochloride, naringin, promethazine, quercetin, quinalizarin, sanguinarine chloride, suramin sodium, tanshinone IIA, triphenylmethane, and resveratrol. All were purchased from Sigma-Aldrich (St-Louis, MO). Clomiphene citrate, raloxifene and tamoxifen were kindly donated by Linda Fung (Pharmacist, London Drugs, Vancouver, BC). Z-FR-MCA, a fluorogenic cathepsin substrate, was purchased from Enzyme System Products (Dublin, CA). The cathepsin inhibitor E-64 (L-3-carboxy-trans-2-3-epoxypropionyl-leucylamido-(4guanidino)-butane) was purchased from Bio Basic Inc. (Markham, ON). Soluble bovine type I collagen was purchased from USB (Cleveland, OH) and C4-S was purchased from Sigma-Aldrich (St-Louis, MO).

High-throughput Fluorescence Polarization Assay (FPA) – 10μM of 1280 known drugs (provided by the Center for Drug Research and Development at The University of British Columbia) were screened using the method described by Selent, J *et al*. (228). Positive hits were screened in a secondary Z-FR-MCA substrate cleavage assay (SCA) to rule out active-site interaction of the drugs.

Collagenase Assay – Soluble bovine type I collagen (0.6 mg/mL) was incubated with 200 nM human recombinant cathepsin K, 100 nM C4-S in sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol, and EDTA. Recombinant human cathepsin K was expressed in *Pichia pastoris* and purified as previously described (237). For the inhibitory collagenase assay, 1 μL of various concentrations of the inhibitors listed above in "Materials" were dissolved in their corresponding solvents and added to the reaction mixture. The total volume of each reaction was

50 µL and incubated at 28°C. After 4 hours, 1 µL of 100 µM E-64 was added to each reaction to inhibit cathepsin K. The mixtures were kept at 4°C overnight and then analyzed on 8% SDS-PAGE gels. Coomassie-stained bands representing the residual amount of α-collagen chains were scanned and assessed using the GeneSnap program (Syngene Inc. Frederick, MD) to determine the strength of each inhibitor on collagen degradation.

Determination of K_m and k_{cat} Values for the Degradation of Type I Collagen by Cathepsin K – Assays were carried out in a similar method as described in the collagenase assay but with varying concentrations of type I calf skin collagen (0.15, 0.3, 0.45, 0.6, 0.75, 0.9, 1.2 mg/mL) in the absence of inhibitors. The assay time was 1 hour and aliquots were taken out every 5 minutes from each reaction and subsequently inhibited with 1 µL of 10 µM E-64. The reaction mixtures were separated by SDS-PAGE and the α1 band density was determined using the GeneSnap program (Syngene Inc. Frederick, MD). The plotting of the time course data provided the reaction rates, which were then evaluated using non-linear regression analysis to obtain the Michaelis-Menten parameter. The α1 band densities were quantified using a calibration curve of type I collagen.

Gelatinase Assay – 1.8 mg/mL gelatin (prepared by heating soluble bovine type I collagen (0.6 mg/mL) at 95°C for 20 minutes) was incubated with 5 nM human cathepsin K in sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol, and EDTA in the presence of various inhibitors (at a 12.5 µM final concentration) at 28°C. After 4 hours, 1 µL of 10µM E-64 was added to each reaction to inhibit the remaining cathepsin K activity. The mixtures were kept overnight at 4°C and ran on 8% SDS-PAGE gels.

Cathepsin K Titration – Enzyme concentrations were determined by active site titration using E-64 with the fluorogenic substrate, Z-FR-MCA, as previously described (216).

Z-FR-MCA Activity Measurement – 5 nM cathepsin K was incubated with or without 12.5 μM inhibitor (as listed in "Materials") in 1 mL sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol, and EDTA for 1 minute prior to addition of 5 μL Z-FR-MCA. The excitation and emission wavelengths used for the assays were set at 380 nm and 460 nm, respectively, using the spectro-fluorometer Perkin-Elmer LB50.

Scanning Electron Microscopy Imaging, and Measurements – Collagen fibers were isolated from mouse tails and incubated with 3 μM cathepsin K with or without 10 μM dihydrotanshinone I in 100 mM sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol and EDTA at 28°C for 10 hours. After addition of E-64, fibers were rinsed with Barnsted Nanopure water (Thermo Scientific, Asheville, NC) and fixed with 2.5% glutaraldehyde, followed by extensive washing. Subsequently, fibers were dried by passing through increasing concentration of ethanol and transferred into a critical point dryer. Upon drying, fibers were mounted on metal stubs using double-sided carbon table and coated with Au/Pd in Hummer VI Sputtering System (Anatech, Union City, CA). Images were taken using Helios NanoLab^{TM} 650 (FEI, Hillsboro, OR) scanning electron microscope at operating voltages between 2 – 10 KV.

Atomic Force Microscopy (AFM) Study – 0.05 nM cathepsin K mixed with 0.025 nM C4-S with or without inhibitors in a 10 μL volume were deposited on freshly cleaved micas for 10 minutes, rinsed with Barnsted Nanopure water (Thermo Scientific) and dried with nitrogen for 15 minutes. Samples were scanned in air in tapping mode with the Cypher scanning AFM (Asylum Research, Santa Barbara, CA). Images (512 x 512 pixel scans) were acquired at a scan rate of 3

Hz using silicon tips (Model-AC160TS, Asylum Research, radius of ~7 nm, resonance frequency of 300 kHz, and spring constant of 42 N/m). The background that corresponds to the mica surface in the images was flattened using the 1st-order flattening function provided by Igor Pro 6.22A (Wavemetrics, Inc., Portland, OR) below a threshold set at ~100-300 pm. The particle size and dimensions were measured using the same Igor software.

Protein-Ligand Docking – The original PDB file of cathepsin K (PDB ID: 3C9E) (238) was obtained from Protein Data Bank and modified with PyMol Molecular Graphics System (version 1.3 Schrödinger, LLC) to remove the triazine ligand and water molecules. 3D ligand structures including aurintricarboxylic acid (CID 2259), tanshinone IIA (CID 164676), and para-dihydrotanshinone (CID 89406) were obtained from PubChem. Docking was carried out using ArgusLab (version 4.01, Mark A. Thompson Planaria Software LLC, Seattle, WA) under the parameters of rigid ligand structure and high precision docking.

Results

High-throughput Fluorescence Polarization Assay (FPA) – The FPA evaluates the molecular binding of a small ligand molecule to its receiver molecule by monitoring the change in fluorescent polarization light signals (228). Typically, a small molecule depolarizes polarized light due to its rapid spinning in solution whereas the bulkier complex, resulting from the attachment of the receiver, slows the spinning motion and causes the light to remain polarized (Figure 4.1) (239,240). An effective inhibitor can disturb the ligand-receiver interaction and therefore restores the light back to the depolarized state. It has been demonstrated that cathepsin K forms an oligomeric complex with C4-S to achieve efficient collagenolytic activity (135). Thus, FPA not

only allows for the evaluation on the strength of cathepsin K and C4-S interactions, but also serves as a tool for screening potential inhibitors that are able to destabilize the cathepsin K/C4-S complex formation. In this study, a 1280-member drug library was screened using a high-throughput FPA with a defined threshold of 40% inhibition. Positive hits were further screened in a secondary Z-FR-MCA substrate cleavage assay (SCA) to exclude compounds that exhibit significant interaction with the active site of the protease (SCA inhibition threshold less than 10%. 15 hits were identified which were grouped into polyanionic (e.g., suramin sodium) and polyaromatic (e.g., ellipticine) compounds (Figure 4.2 & Table 4.1). Because some of the identified inhibitors are known as anti-resorptive drugs (241,242), further effort was taken to exploit other structurally similar compounds, as well as a number of others known to affect osteoclastogenesis. All compounds used in this study were listed in Table 4.2 and Table 4.3.

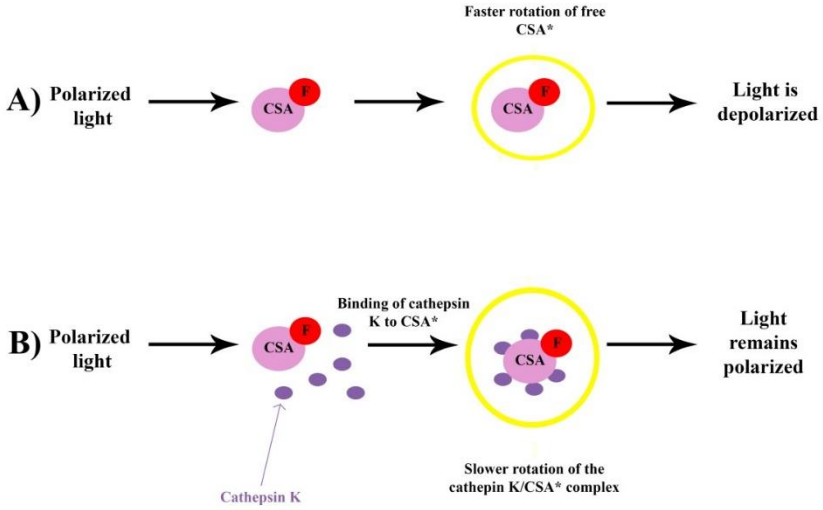


Figure 4.1. Scheme of Fluorescence Polarization Assay. C4-S used in FPA was first activated with cyanogen bromide at pH 11.0. The resulting products were fractioned and purified with a PD-10 column for the coupling of fluorescamine. The coupling was done overnight at 4°C and further fractionated with a PD-10 column in 50 mM acetate buffer, pH 5.5. Recombinant cathepsin K was expressed in *Pichia pastoris.* After pepsin activation at pH 4.0, the activated protease was purified with n-butyl and SP-Sepharose on FPLC. The polarization was performed at pH 5.5 in a 100 mM sodium acetate buffer. The fluorescence complex was excited at 485 nm and the signal was quantified at 520 nm.

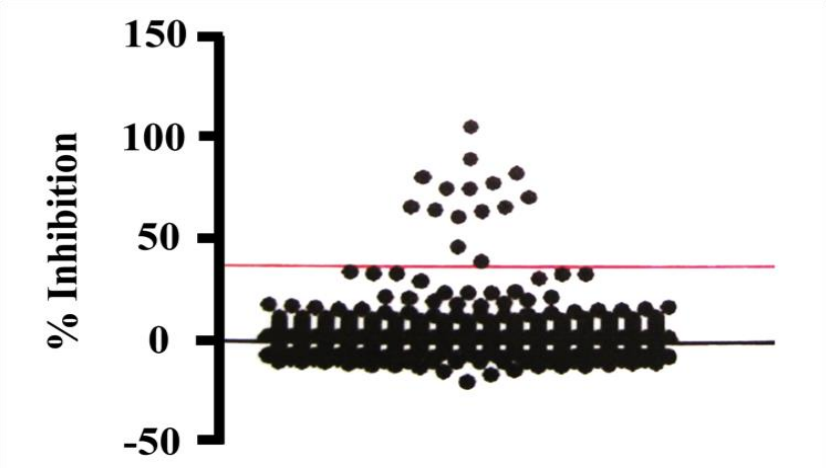
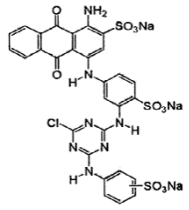
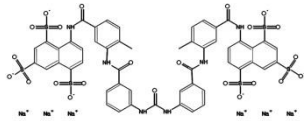
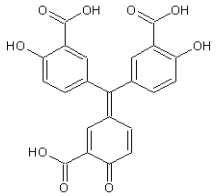
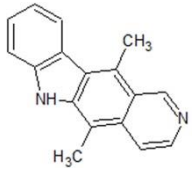
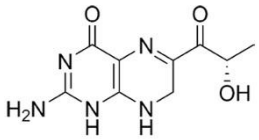


Figure 4.2. Distribution of Library Compounds as Cathepsin K/C4-S Inhibitors. Assay threshold was set at a minimum level of 40% FPA inhibition and 10% SCA inhibition. Positive compounds are shown as dots above the red line (threshold level).

Compound	Structure	FPA% Inhibition	SCA% Inhibition	MW (g/mol)	Known Activity
Reactive Blue 2		100	10	840	Potent Antagonist for P2Y Receptor and ATP-Activated Channels
Suramin Sodium		87	10	1429	P2X P2Y Receptor Antagonist and Anti-Resorption
Aurintricarboxylic Acid		90	10	422	Topoisomerase Inhibitor
Ellipticine		85	0	246	Cytochrome P450 and Topoisomerase Inhibitor
Sepiapterin		82	0	237	Co-factor of Nitric Oxide Synthase

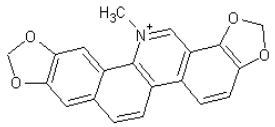
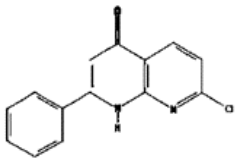
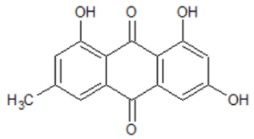
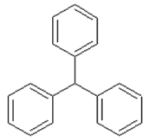
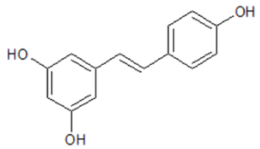
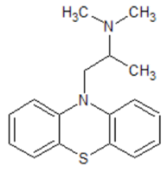
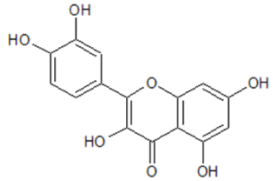
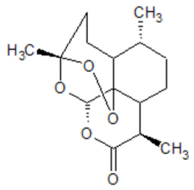
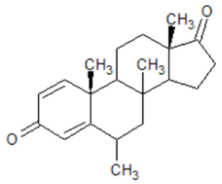
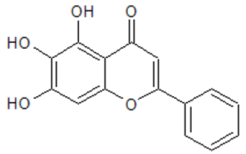
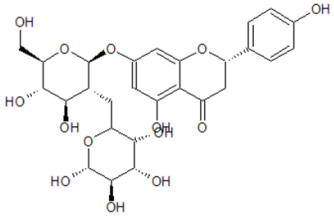
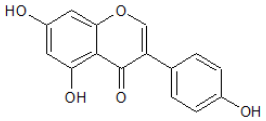
Compound	Structure	FPA% Inhibition	SCA% Inhibition	MW (g/mol)	Known Activity
Sanguinarine		51	0	368	Inhibitor of ATPases
7-Chloro-4-Hydroxy-2-Phenyl-1,8 Naphthyridine		43	0	257	A1 Adenoisine Receptor Antagonist

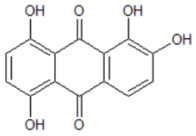
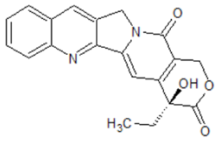
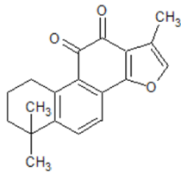
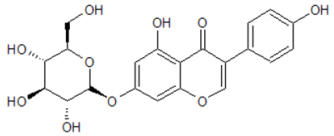
Table 4.1. Selected Positive Hits from the Highthroughput Fluorescence Polarization Assay (FPA). The maximum and minimum FPA% was defined using reactive blue 2 (100%) and 7-chloro-4hydroxy-2-phenyl-1,8 naphthyridine (43%) respectively. Compounds of interest with FPA% in between 100% to 43% were chosen and tested with a secondary Z-FR-MCA substrate cleavage assay (SCA) to verify their inhibition to the active site of cathepsin K.

Verification of Potential Inhibitors with Collagenase and Gelatinase Assays – Compounds identified via the FPA screening possessed two important characteristics. First, they had the ability to interfere with cathepsin K/C4-S complex formation. Second, they exhibited minimal or no active site inhibition. These compounds were termed exosite inhibitors of cathepsin K. To assess the effectiveness of these exosite inhibitors to block the collagenase activity of cathepsin K, the degradation of type I collagen in the presence of these inhibitors was evaluated. Serving as a control, a gelatinase assay was performed to evaluate the potential effect of inhibitors on the active site of the enzyme when a protein substrate is used. We also included structurally related compounds to those identified in the FPA assay and compounds which have reported anti-resorptive activity but no known activity towards cathepsin K. Results of the assessments are summarized in Table 4.2 and Table 4.3, representing the effective compounds and the ineffective compounds, respectively.

14 out of the 24 compounds tested did not reveal any inhibition of the collagenolytic activity of cathepsin K at concentrations of maximal solubility in their respective solvent (Table 4.2). These compounds included emodin, triphenylmethane, resveratrol, promethazine, quercetin, artemisinin, exemestane, baicalein, naringin, genistein, quinalizarin, (S)-(+)-camptothecin, tanshinone IIA, and genistin.

Compounds (ineffective)	**Structure**	**Molecular Formula**	**MW (Da)**	**Solvent**	**IC_{50} (μM)**
Emodin		$C_{15}H_{10}O_5$	270.24	DMSO	> 67
Triphenylmethane		$C_{19}H_{16}$	244.33	DMSO	> 200
Reseratrol		$C_{14}H_{12}O_3$	228.24	DMSO	> 270
Promethazine		$C_{17}H_{20}N_2S \cdot HCl$	320.85	H_2O	> 400
Quercetin		$C_{15}H_{10}O_7 \cdot H_2O$	302.24	DMSO	> 1000
Artemisinin		$C_{15}H_{22}O_5$	282.33	DMSO	> 200

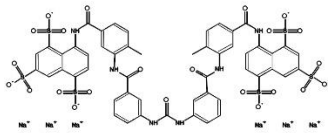
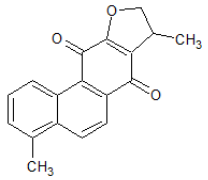
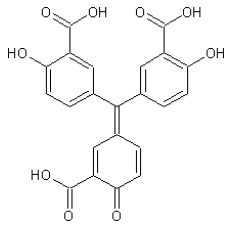
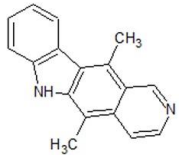
Compounds (ineffective)	Structure	Molecular Formula	MW (Da)	Solvent	IC_{50} (μM)
Exetrestane		$C_{20}H_{24}O_2$	296.4	DMSO	> 200
Baicalein		$C_{15}H_{10}O_5$	270.24	Ethanol	> 100
Naringin		$C_{27}H_{32}O_{14}$	580.54	DMSO	> 400
Genistein		$C_{15}H_{10}O_5$	270.24	DMSO	> 200

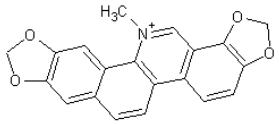
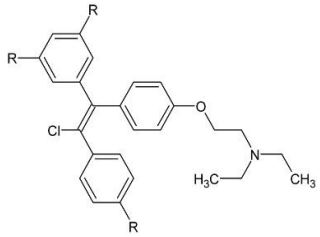
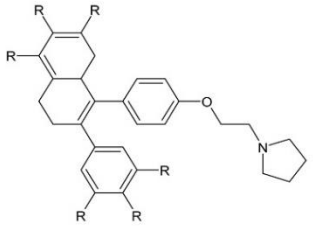
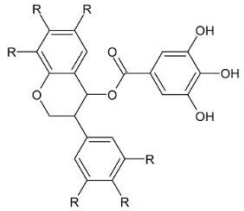
Compounds (ineffective)	Structure	Molecular Formula	MW (Da)	Solvent	IC_{50} (μM)
Quinalizarin		$C_{14}H_8O_6$	272.21	Ethanol	> 25
(S)-(+)-Camptothecin		$C_{20}H_{16}N_2O_4$	348.35	DMSO	> 200
Tanshinone IIA		$C_{19}H_{18}O_3$	294.34	Ethanol	> 260
Genistin		$C_{21}H_{20}O_{10}$	432.38	Activity buffer	> 740

***** *Inhibitors were denoted as ineffective if no inhibition of collagenase acitivity of cathepsin K was observed on SDS-PAGE As detetermined by the α1 band density at their maximum solubility in the respective solvent.*

Table 4.2. Ineffective Collagenase Exosite Inhibitors. Structural and chemical information of the compounds were obtained from PubChem. The IC_{50} value was determined by plotting the percentage of inhibition versus the concentration of inhibitor. Activity buffer contained 200 nM human recombinant cathepsin K, 100 nM C4-S in sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol, and EDTA.

Among the effective compounds, suramin, a polysulfonated polyaromatic symmetrical urea, is capable of suppressing collagen degradation at a relatively low concentration with an IC_{50} of 4.8 ± 0.9 µM (Figure 4.3A-1 and Table 4.3). Ranking from the most potent to the least among the polyaromatic group were: aurintricarboxylic acid (IC_{50} of 9.5 ± 3.4 µM), dihydrotanshinone I (IC_{50} of 11.0 ± 3.0 µM), nafoxidine hydrochloride (IC_{50} of 25.6 ± 3.8 µM), clomiphene citrate (IC_{50} of 30.3 ± 4.2 µM), tamoxifen (IC_{50} of 48.7 ± 10.3 µM), (-)-epigallocatechin gallate (IC_{50} of 75.3 ± 10.1 µM), ellipticine (IC_{50} of 139.3 ± 48.5 µM, raloxifene (IC_{50} of 175.5 ± 35.1 µM), and sanguinarine chloride (IC_{50} of 186.3 ± 36.8 µM) (Figures 4.3B-4.3J and Table 4.3). It should be noted that the IC_{50} values do not present true binding constants and depend on the concentration of protease. Antipain, a known active-site directed peptide aldehyde with a K_i-value of 40 nM for cathepsin K (unpublished data, Bromme lab) revealed an IC_{50} value of 36.0 ± 2.0 µM. This implies that the real binding constants for the exosite inhibitors are substantially lower than their IC_{50} values.

Compounds (effective)	Structure	Molecular Formula	MW (Da)	Solvent	IC_{50} (μM)
Polyanionic Compounds					
Suramin Sodium		$C_{51}H_{34}N_6O_{23}S_6 \cdot Na_6$	1429.2	Activity Buffer	4.8 ± 0.9
Polyaromatic Compounds					
Dihydrotanshinone I		$C_{18}H_{14}O_3$	278.3	Ethanol	11.0 ± 3.0
Aurintricarboxylic Acid		$C_{22}H_{14}O_9$	422.3	Activity Buffer	9.5 ± 3.4
Ellipticine		$C_{17}H_{14}N_2$	246.3	DMSO	88.4 ± 28.7

Compounds (effective)	Structure	Molecular Formula	MW (Da)	Solvent	IC_{50} (μM)
Sanguinarine Chloride		$C_{20}H_{14}NO_4Cl$	367.8	Methanol	186.3.0 ± 36.8
Selective Estrogen-Receptor Modulators (SERMs)					
Clomiphene Citrate		$C_{26}H_{28}ClNO \cdot C_6H_8O_7$	598.1	Methanol	30.3 ± 4.2
Nafoxidine Hydrochloride		$C_{29}H_{31}NO_2 \cdot HCl$	462.0	DMSO	25.6 ± 3.8
(-)-Epigallocatechin Gallate		$C_{22}H_{18}O_{11}$	458.4	Activity Buffer	75.3 ± 10.1

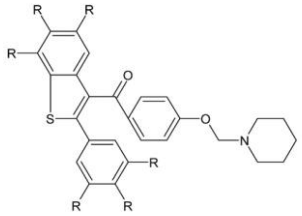
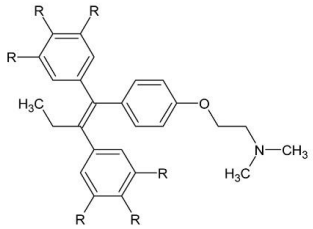
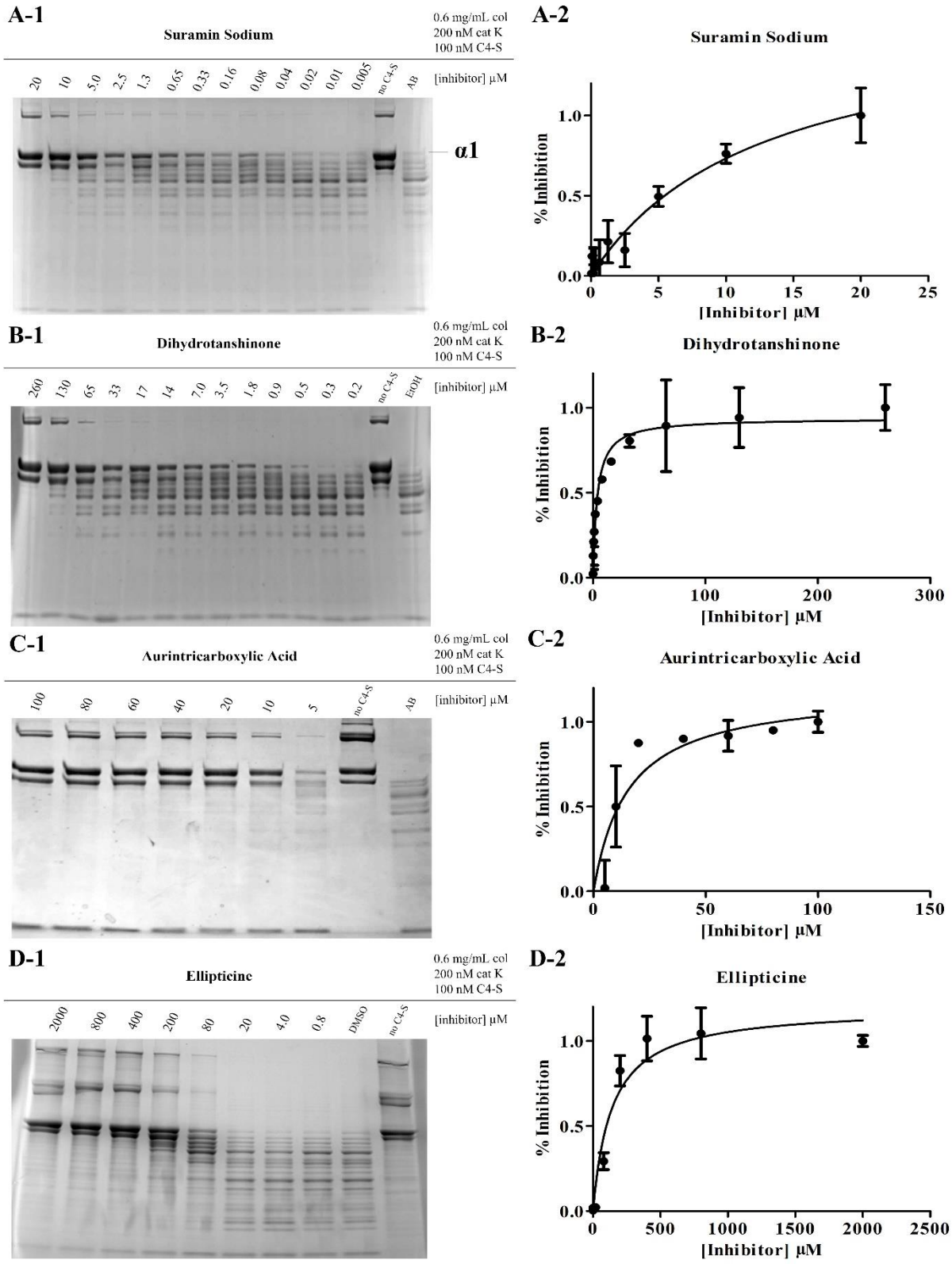
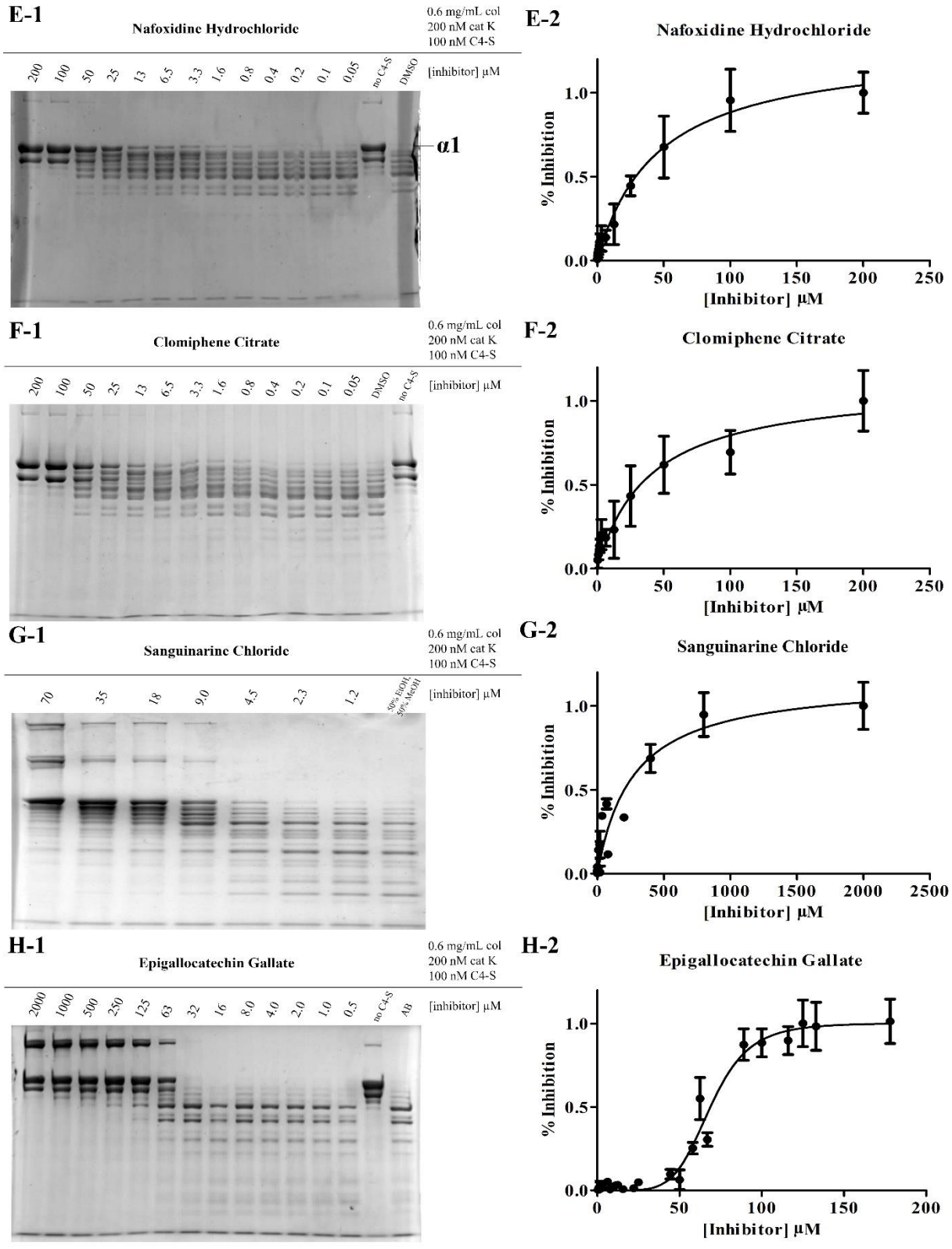
Compounds (effective)	Structure	Molecular Formula	MW (Da)	Solvent	IC_{50} (μM)
Raloxifene		$C_{28}H_{27}NO_4S$	473.6	DMSO	175.5 ± 35.1
Tamoxifen		$C_{26}H_{29}NO$	371.5	Ethanol	48.7 ± 10.3

Table 4.3. Effective Collagenase Exosite Inhibitors. Structural and chemical information of the compounds were obtained from PubChem. The IC_{50} value was determined by plotting the percentage of inhibition versus the concentration of inhibitor. To calculate the percentage of inhibition, the correlation between initial collagen concentration and the α1 band density was first established as described in the experimental procedures. By measuring the change of α1 band density relative to the control on the coomassie blue stained SDS-PAGE gel. Three replicates were carried out for each inhibitor, and the IC_{50} values were determined using GraphPad Prism software. Activity buffer contained 200 nM human recombinant cathepsin K, 100 nM C4-S in sodium acetate buffer, pH 5.5, containing 2.5 mM dithiothreitol, and EDTA.


A-1
Suramin Sodium
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
20 10 5.0 2.5 1.3 0.65 0.33 0.16 0.08 0.04 0.02 0.01 0.005 no C4-S AB
α1
A-2
Suramin Sodium
% Inhibition
[Inhibitor] µM
B-1
Dihydrotanshinone
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
260 130 65 33 17 14 7.0 3.5 1.8 0.9 0.5 0.3 0.2 no C4-S EtOH
B-2
Dihydrotanshinone
% Inhibition
[Inhibitor] µM
C-1
Aurintricarboxylic Acid
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
100 80 60 40 20 10 5 no C4-S AB
C-2
Aurintricarboxylic Acid
% Inhibition
[Inhibitor] µM
D-1
Ellipticine
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
2000 800 400 200 80 20 4.0 0.8 DMSO no C4-S
D-2
Ellipticine
% Inhibition
[Inhibitor] µM


E-1
Nafoxidine Hydrochloride
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
200 100 50 25 13 6.5 3.3 1.6 0.8 0.4 0.2 0.1 0.05 no C4-S DMSO
α1
E-2
Nafoxidine Hydrochloride
% Inhibition
[Inhibitor] µM
F-1
Clomiphene Citrate
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
200 100 50 25 13 6.5 3.3 1.6 0.8 0.4 0.2 0.1 0.05 DMSO no C4-S
F-2
Clomiphene Citrate
% Inhibition
[Inhibitor] µM
G-1
Sanguinarine Chloride
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
70 35 18 9.0 4.5 2.3 1.2 50% EtOH 50% MeOH
G-2
Sanguinarine Chloride
% Inhibition
[Inhibitor] µM
H-1
Epigallocatechin Gallate
0.6 mg/mL col
200 nM cat K
100 nM C4-S
[inhibitor] µM
2000 1000 500 250 125 63 32 16 8.0 4.0 2.0 1.0 0.5 no C4-S AB
H-2
Epigallocatechin Gallate
% Inhibition
[Inhibitor] µM

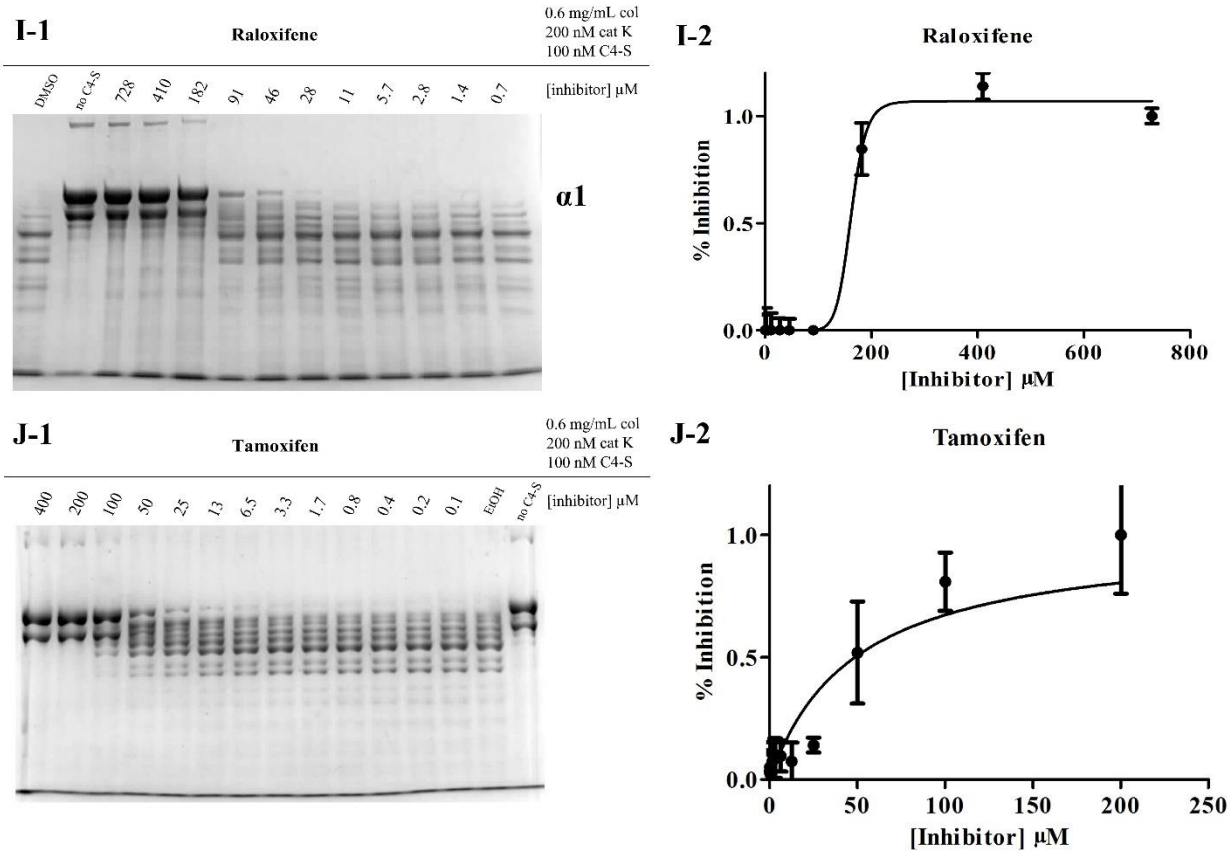


Figure 4.3. Cathepsin K Collagenase Assays in the Presence of Different Exosite Inhibitors. A-1, B-1, C-1, D-1, E-1, F-1, G-1, H-1, I-1 and J-1. SDS-PAGE images for the inhibition of type I collagen degradation for each inhibitor. A-2, B-2, C-2, D-2, E-2, F-2, G-2, H-2 and J-2. IC_{50} plots of each inhibitor. Inhibitors were diluted in the corresponding solvent as stated in Table 4.3. Gradients of inhibitors were used to study the effect of inhibition on collagenase degradation by cathepsin K. 'AB' represents the reaction buffer used in the assay, which contains 100 mM sodium acetate, 2.5 mM DTT and 2.5 mM EDTA at pH 5.5. Assays were carried out with 200 nM cathepsin K and 100 nM C4-S, and repeated in three independent runs. The lane labeled as 'AB' was used as the positive control for the inhibitors that were dissolved in 'AB'. If the solvent for the inhibitor was not reaction buffer, corresponding volume of the solvent was added to the reaction mixture and served as the positive control. The negative control 'no C4-S' removed C4-S to prevent the degradation of collagen with the corresponding volume of the inhibitor solvent. The degradation of collagen was quantified as α1 band density on the coomassie blue stained gel, and divided this density by that of the 'no C4-S' to obtain the % inhibition. The % inhibition was plotted against its corresponding inhibitor concentration to obtain the IC_{50} plot.

As per definition, exosite inhibitors bind to its protease target outside of the active site. To confirm this, the inhibitor to enzyme molar concentration ratio was increased to a $2.7x10^5$-fold excess. Six out of the ten effective compounds including aurintricarboxylic acid, clomiphene citrate, dihydrotanshinone I, ellipticine, nafoxidine hydrochloride and tamoxifen exhibited no active site inhibition. Raloxifene, sanguinarine chloride, (-)-epigallocatechin gallate and suramin sodium showed 11.1 ± 0.1%, 13.8 ± 1.8%, 24.0 ± 1.0%, and 34.0 ± 1.7% inhibition of Z-FR-MCA hydrolysis, respectively (Table 4.4). Gelatin hydrolysis, however, revealed no observable inhibition by all tested inhibitors at a 2,500-fold molar excess of the inhibitor (Figure 4.4 & Table 4.4).

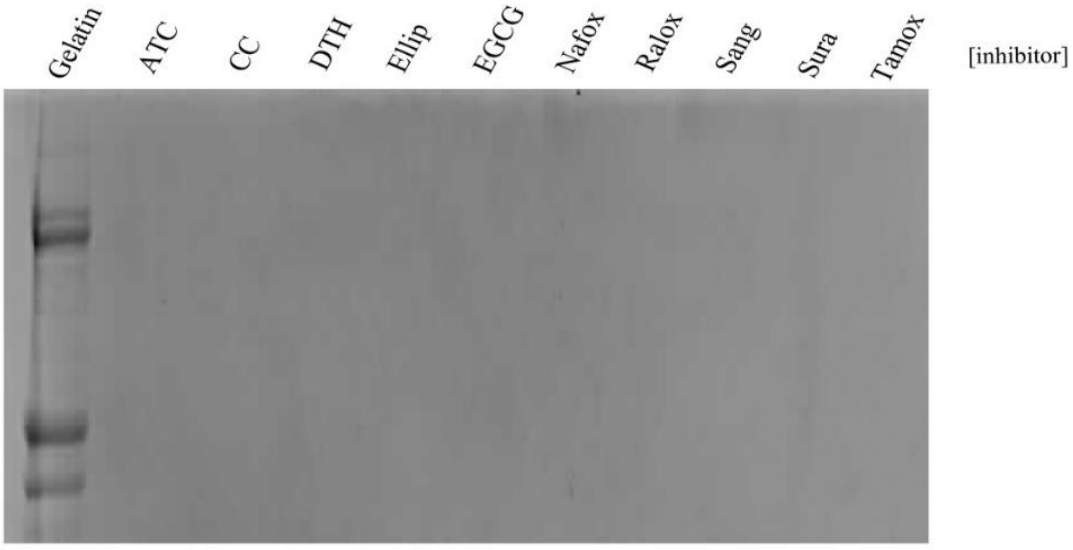


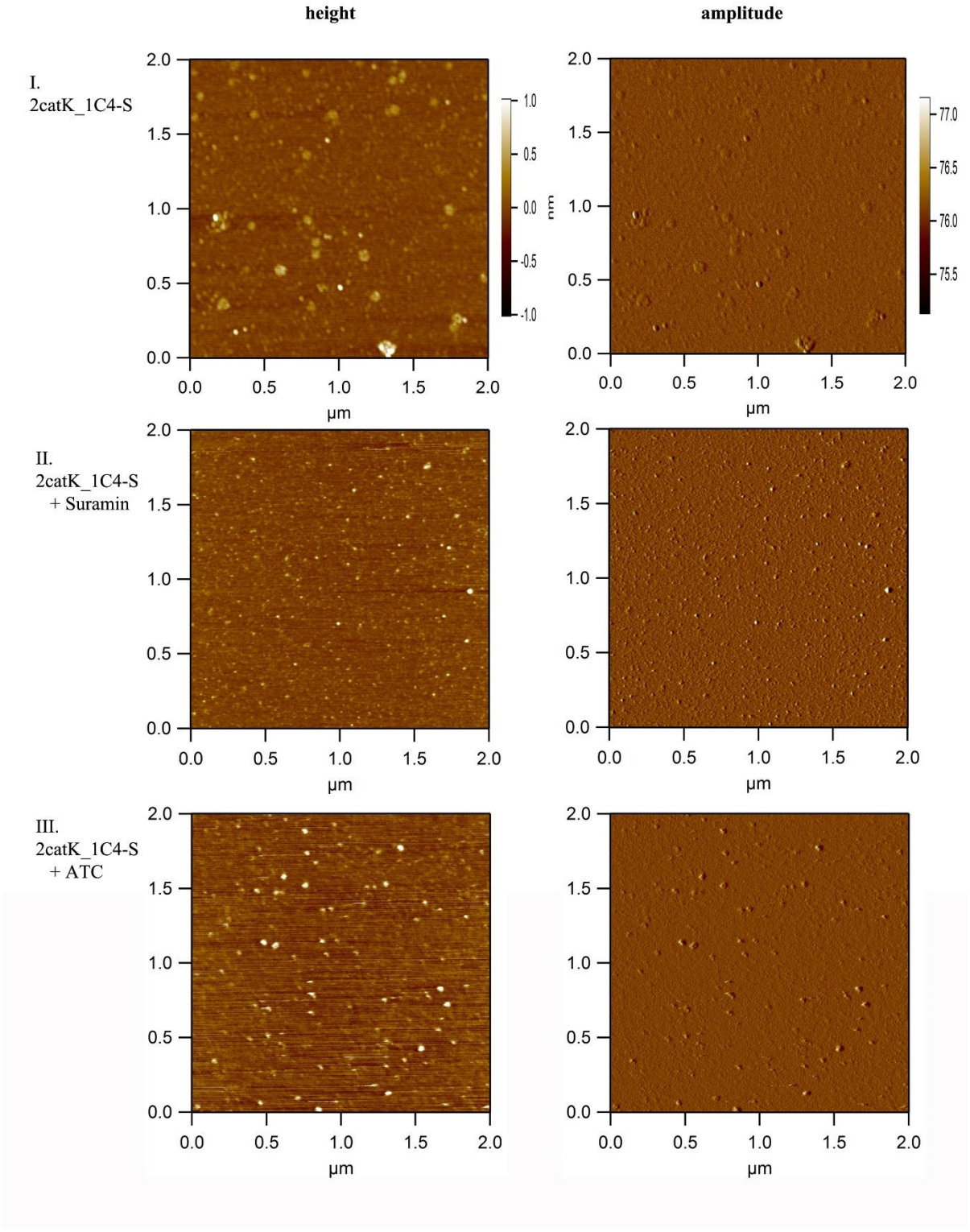
Figure 4.4. Cathepsin K Gelatinase Assay in Presence of Various Inhibitors. Assay was performed at 5 nM cathepsin K concentration without the presence of C4-S. Final concentration of the inhibitors was set at 12.5 µM. Samples were run on an 8% SDS-PAGE gel with the lane gelatin served as the positive control.

Compound	[Inhibitor]f (μM)	% Inhibition on Substrates		[Inhibitor]/[E] Ratio
		Z-FR-MCA	Gelatin	
Aurintricarboxylic Acid	12.5	0%	0%	270000
Clomiphene Citrate	12.5	0%	0%	270000
Dihydrotanshinone I	12.5	0%	0%	270000
Ellipticine	12.5	4.0 ± 1.5%	0%	270000
Epigallocatechin Gallate	12.5	24.0 ± 1.0%	0%	270000
Nafoxidine Hydrochloride	12.5	0%	0%	270000
Raloxifene	12.5	11.1 ± 0.1%	0%	270000
Sanguinarine Chloride	12.5	13.8 ± 1.8 %	0%	270000
Suramin Sodium	12.5	34.0 ± 1.7%	0%	270000
Tamoxifen	12.5	1.5 ± 0.1%	0%	270000

Table 4.4. Effects of Exosite Inhibitors on the Active Site of Cathepsin K. Inhibition was determined using the Z-FR-MCA cleavage and gelatinase assays. For both assays, each final inhibitor concentration was maintained at 12.5 μM whereas the final concentration of cathepsin K was 5 nM. The reaction time for the gelatinase assay was 30 min. The release of the flourogenic MCA group was detected at the excitation and emission wavelengths of 380 nm and 460 nm, respectively. The rates of enzymatic reaction were recorded using the spectro-fluorometer Perkin-Elmer LB50 for 30 s.

AFM Analysis of Cathepsin K/C4-S Complexes in the Presence of Exosite inhibitors – AFM was utilized to study the effects of several inhibitors including suramin, aurintricarboxylic acid, dihydrotanshinone I, (-)-epigallocatechin gallate and Clomiphene on the morphology of the collagenolytically active cathepsin K/C4-S tetrameric complexes. The images revealed that at a molar 2:1 ration of cathepsin K and C4-S, cathepsin K formed ring-shaped assemblies with an average height of 3.8 ± 1.2 nm, which was topologically similar to the putative tetramer model (Figure 4.5 I). However, in the presence of excess molar inhibitors, the ring formation was blocked to different degrees for the individual inhibitors: suramin completely prevented the ring formation and left only monomers representing single cathepsin K molecules (Figure 4.5 II); aurintricarboxylic acid gave rise to a mixture of smaller semi-circled and circled structures (Figure

4.5 III); a number of deformed rings were present dihydrotanshinone I (Figure 4.5IV) and (-)-epigallocatechin gallate (Figure 4.5V) while some monomeric cathepsin K molecules were also present; clomiphene caused proteins to aggregate, resulting in larger size clusters and scattered monomers (Figure 4.5VI). These results suggested that all selected inhibitors for AFM might interfere with the complex formation between cathepsin K and C4-S.


height
amplitude
I.
2catK_1C4-S
II.
2catK_1C4-S
+ Suramin
III.
2catK_1C4-S
+ ATC
nm
1.0
0.5
0.0
-0.5
-1.0
77.0
76.5
76.0
75.5
0.0
0.5
1.0
1.5
2.0
µm

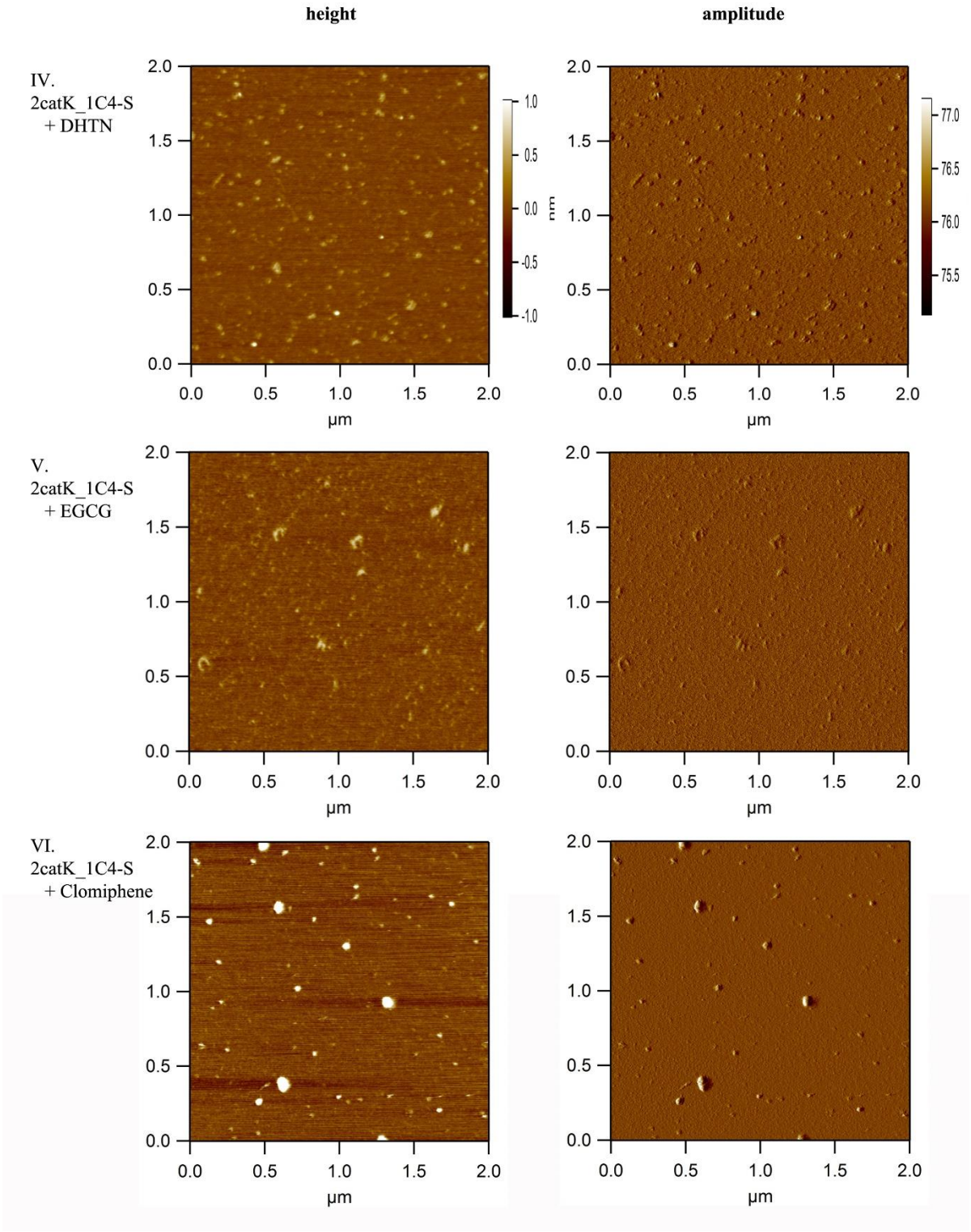


Figure 4.5. AFM Imaging of Cathepsin K/C4-S Complex with and without Inhibitors. I: cathepsin K/C4-S complex. II: cathepsin K/C4-S complex with 0.25 μM suramin. III: cathepsin

K/C4-S complex with 0.5 μM aurintricarboxylic acid (ATC). IV: cathepsin K/C4-S complex with 0.25 μM dihydrotanshinone I. V: cathepsin K/C4-S complex with 4 μM (-)-epigallocatechin gallate (EGCG). VI: cathepsin K/C4-S complex with 2 μM clomiphene. Cathepsin K/C4-S complex was set at a molar ratio of 2:1. Cathepsin K concentrations used was 0.02 nM. Specimens in a 10 μL volume were deposited on freshly cleaved mica for 5 minutes, rinsed with Barnsted Nanopure water and air dried for 30 minutes. Samples were scanned in air in tapping mode with the Cypher scanning AFM. Images (512 lines) were acquired at a scan rate of 3 Hz using silicon tips (Model-AC160TS, Asylum Research, radius of ~7 nm, resonance frequency of 300 kHz, and spring constant of 42 N/m).

Scanning Electron Microscopy Imaging and Analysis – Insoluble fibrillar collagen is the natural substrate for cathepsin K. The degradation of the fibers occurs stepwise by unraveling the macrofiber into fibrils, microfibrils and finally dissolving them (243). To evaluate whether the identified exosite inhibitors could prevent fiber collagenolysis, collagen fibers were incubated with cathepsin K in the presence of inhibitors and the fate of the fibers was subsequently analyzed by SEM. We selected two compounds, dihydrotanshinone I and tamoxifen which have an IC_{50} value of 11 ± 3 μM and 48.7 ± 10.3 μM for the inhibition of soluble tropocollagen degradation and virtually no effect on Z-FR-MCA hydrolysis. Scanning electron microscopy analyses demonstrated that collagen fibers isolated from mouse tails had an average diameter of 42.3 ± 5.8 μm (Figure 4.6A). Without cathepsin treatment, the surface of the fibers was intact, showing an organized parallel assembly of microfibrils. After digestion with 3 μM cathepsin K at 28°C for 10 hours, fibril bundles were no longer tightly packed and disintegrated into microfibrils ranging from 1.1 μm to 200 nm (Figure 4.6B). In the presence of dihydrotanshinone I and tamoxifen (50μM each), cathepsin K treated collagen fibers remained intact with average diameters of 45.2 ± 4.4 μm. (Figure 4.6C and 4.6D). Minor superficial nicks were observed, which were similar to those when collagen fibers were incubated with cathepsin K in the presence of 300 mM NaCl that prevents complex formation or with other non-collagenolytic cathepsins (243).

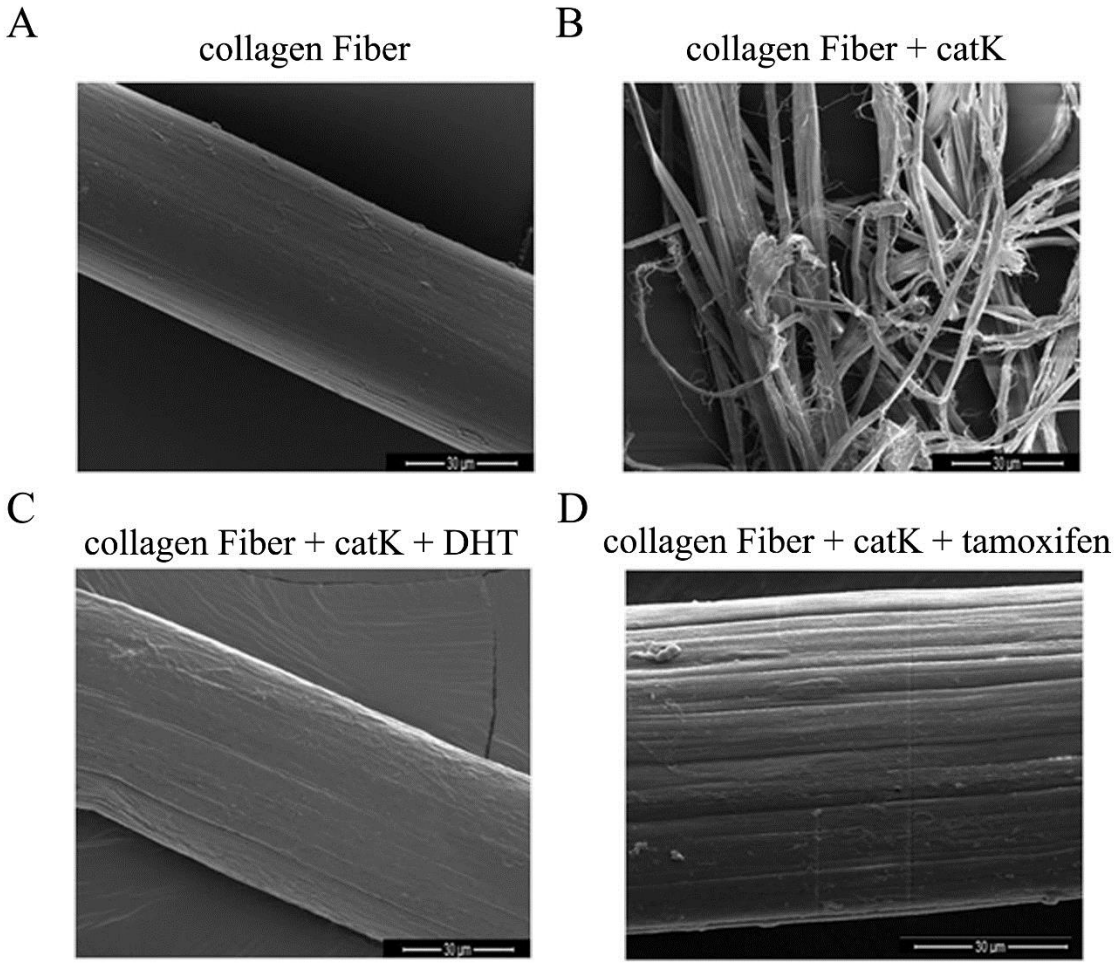
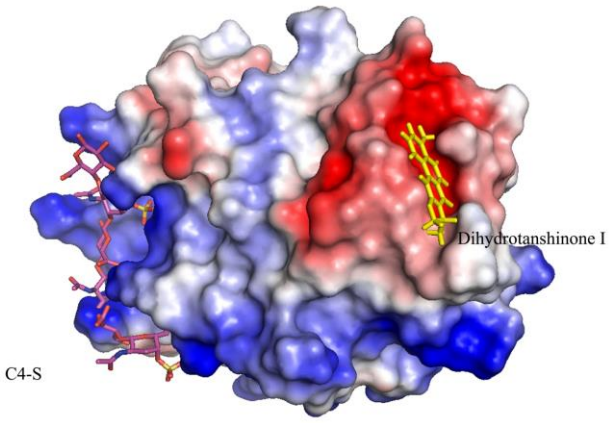


Figure 4.6. Degradation of Insoluble Type I Collagen by Cathepsin K with and without Dihydrotanshinone I (DHT) and Tamoxifen. A. Scanning electron micrograph of the untreated insoluble rail tail collagen fiber. B. Collagen fiber after treatment of 3 µM cathepsin K for 10 hours at 28°C. C. Collagen fiber after treatment of cathepsin K in presence of 50 µM dihydrotanshinone I (DHT). D. Collagen fiber after treatment of cathepsin K in the presence of 50 µM tamoxifen. Reaction buffer used for the assay contained 100 mM sodium acetate, 2.5 mM DTT and 2.5 mM EDTA at pH 5.5. Bars = 30 µm.

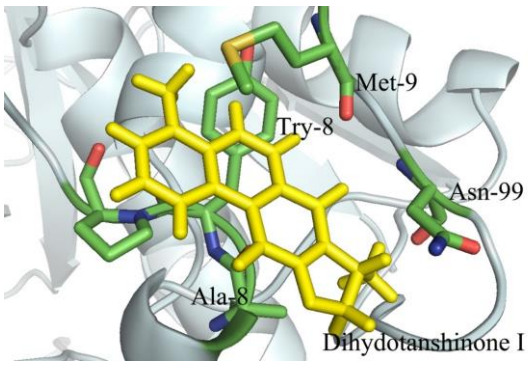
Molecular Docking of Exosite Inhibitors to Cathepsin K - To better understand the mechanisms of exosite inhibitors, dihydrotanshinone I and aurintricarboxylic acid were docked onto cathepsin K using the ArgusLab software and compared with the ineffective inhibitor

tanshinone IIA. Dihydrotanshinone I docks into a trench on cathepsin K surface formed by the backbone of A_{86}, Y_{87}, V_{90} and N_{99}, and the side-chains of Y_{87}, P_{88}, Met_{97} and N_{99} (Figure 4.7A, 4.7B and Table 4.5). The binding is stabilized by 3 H-bonds from the protein to the inhibitor and a docking energy of -9.99 kcal/mol was determined. No interference is found at the active site of cathepsin K, supporting the finding that this compound is not inhibiting the hydrolysis of Z-FR-MCA and gelatin. In contrast, tanshinone IIA, revealed multiple binding sites indicating a low affinity and non-specificity. The highest affinity was observed to a region comprising residues E_{84}, D_{85}, F_{28}, Y_{89}, K_{17}, Y_{89}, N_{18} (docking energy -9.90 kcal/mol) (Figure 4.7C and Table 4.5). This region does not participate in the complex formation. Although it may also dock near residues M_{97} – T_{101}, the binding strength is relatively weak as indicated by a smaller docking energy of -8.58 kcal/mol and less number of H-bond (Table 4.5). This weak binding is accordance with the finding of the collagenase assay where no inhibition up to a concentration of 260 µM was found (Table 4.2).

A



B



C

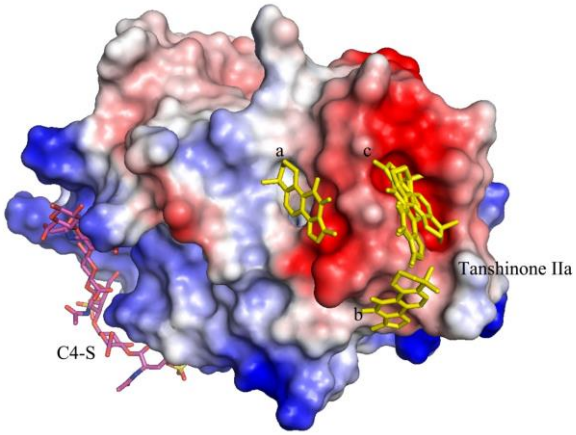


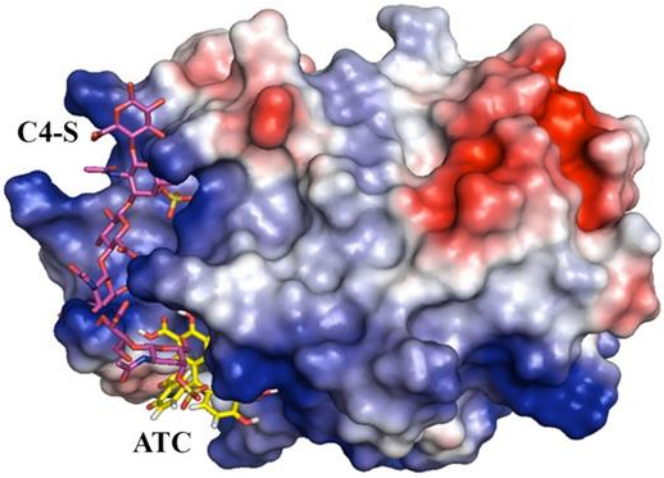
Figure 4.7. Putative Docking Models of Dihydrotanshinone I and Tanshinone IIA on Cathepsin K. A. Global view of dihydrotanshinone I binding on cathepsin K. B. Interactions of dihydrotanshinone I with the surrounding amino acid residues. C. Global view of tanshinone IIA binding on cathepsin K. C4-S is colored in purple, the inhibitors are colored in yellow, and amino acid residues of cathepsin K are colored in green. Labels a, b and c correspond to three different binding sites as listed in Table 4.5.

Compound	Hydrophobic Interaction	H-Bond	Docking Energy (kcal/mol)
Dihydrotanshinone I	Y_{87}, V_{90}, P_{88}, E_{94}, A_{86}, M_{97}, N_{99}	3	-9.99
Tanshinone IIA	[a] E_{84}, D_{85}, F_{28}, Y_{89}, K_{17}, Y_{89}, N_{18}	1	-9.90
	[b] S_{83}, A_{86}, G_{102}, T_{101}, D_{85}	2	-8.97
	[c] M_{97}, Y_{87}, E_{94}, P_{88}, N_{99}, M_{97}, A_{86}	1	-8.58

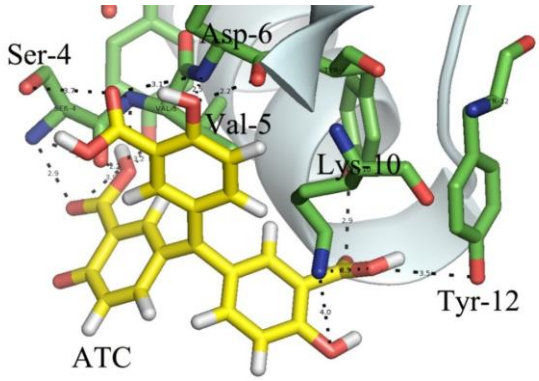
Table 4.5. Contacts Identified Between Dihydrotanshinone I and Cathepsin K on the Putative Docking Model. Docking was performed using ArgusLab (version 4.01, Mark A. Thompson Planaria Software LLC, Seattle, WA) with a rigid constrain on the ligand and without the presence of C4-S. Molecular fits giving the highest free energy were selected for study. Pymol was used to identify and analyze interactions between the ligand and cathepsin K. H-bonding was defined with a maximum distance of 4 Å. The original PDB file of cathepsin K (PDB ID: 3C9E) was obtained from Protein Data Bank and modified with PyMol Molecular Graphics System (version 1.3 Schrödinger, LLC) to remove the triazine ligand and water molecules. 3D ligand structures of the dihydrotanshinone I (CID 89406) and tanshinone IIA were obtained from PubChem (CID 164676). Labels a, b and c represent three different binding sites on cathepsin K as indicated in Figure 4.8.

Unlike dihydrotanshinone I, aurintricarboxylic acid docks at the back-side of cathepsin K, opposite from the active site. Analysis indicates multiple H-bonds between amino acid residues S_4, V_5, D_6, Y_7, K_{10}, Y_{169} and the compound (Figure 4.8A, & 4.8B, Table 4.6). Coincidently, S_4, D_6 and K_{10} are also the binding residues for C4-S (135). As a result, the presence of aurintricarboxylic acid may competitively impede the binding of C4-S to cathepsin K, which prevents or at least weakens the proper formation of a functional cathepsin K/C4-S complex for collagen degradation and thus explains its selective anti-collagenolytic activity (Figure 4.8C).

A



B



C

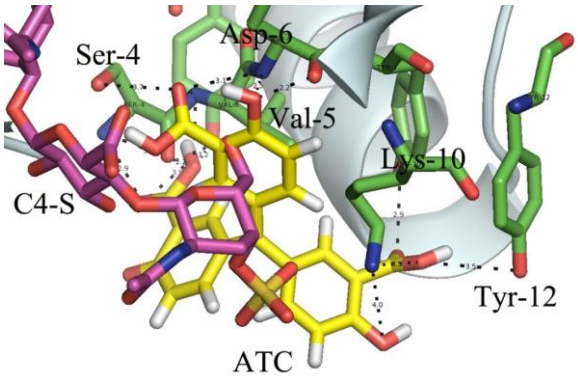


Figure 4.8. Molecular Docking of Aurintricarboxylic Acid on Cathepsin K. A. Global view of aurintricarboxylic acid binding on cathepsin K. B. Interaction of aurintricarboxylic acid with surrounding amino acid residues. C. Interaction of aurintricarboxylic acid with surrounding amino acid residues in the presence of C4-S. C4-S is colored in purple, the inhibitors are colored in green, and amino acid residues of cathepsin K are colored in green. Dash lines '- -' represents H-bonds.

H-Bonding			Docking Energy (kcal/mol)
Cathepsin K	**Aurintricarboxylic Acid**	**Distance (Å)**	
S_4 NH	COOH (benzene ring 1)	3.3	-8.98
S_4 NH	C= O (benzene ring 1)	2.9	
S_4 O	COOH (benzene ring 1)	2.2	
S_4 OG	C=O (benzene ring 2)	3.7	
V_5 NH	COOH (benzene ring 1)	3.0	
V_5 NH	C=O (benzene ring 2)	3.4	
D_6 NH	COH (benzene ring 2)	2.5	
D_6 O	COH (benzene ring 2)	2.2	
Y_7 OH	C=O (benzene ring 3)	2.9	
K_{10} NZ	COOH (benzene ring 3)	3.5	
K_{10} NZ	COH (benzene ring 3)	4.0	
Y_{12} OH	COOH (benzene ring 3)	3.5	
Y_{169} OH	COOH (benzene ring 1)	3.2	
Y_{169} OH	C=O (benzene ring 1)	3.9	

Table 4.6. Contacts Identified Between Aurintricarboxylic Acid and Cathepsin K on the Putative Docking Model. Docking was performed using ArgusLab (version 4.01, Mark A. Thompson Planaria Software LLC, Seattle, WA) with a rigid constrain on the ligand and without the presence of C4-S. Molecular fits giving the highest free energy were selected for study. Pymol was used to identify and analyze interactions between the ligand and cathepsin K. H-bonding was defined with a maximum distance of 4 Å. The original PDB file of cathepsin K (PDB ID: 3C9E) was obtained from Protein Data Bank and modified with PyMol Molecular Graphics System (version 1.3 Schrödinger, LLC) to remove the triazine ligand and water molecules. 3D ligand structures of the aurintricarboxylic acid (CID 2259) were obtained from PubChem.

Discussion

The collagenolytic activity of cathepsin K depends on the presence of GAGs such as C4-S (65). This implies the formation of an oligomeric complex between cathepsin K and GAGs. Considering that the current active-site-targeting inhibitor approach may lead to various adverse side effects, this study aims to identify exosite inhibitors of cathepsin K as specific collagenase blockers. A total of 1280 compounds were screened of which 15 positive hits were identified with an inhibition of greater than 40%. Of these selected candidates and a number of structurally similar compounds were further evaluated in collagenase and gelatinase assays to verify the exosite targeting of the compounds without interfering with the active site.

Of the 15 compounds identified by FPA, suramin is the only polyanionic compound that prevents collagen degradation by cathepsin K. It has been describes as a potent anti-cancer agent used to suppress various types of cancers such as metastatic renal cell carcinoma, lymphoma, and metastatic prostatic cancer (244). It has also been shown to inhibit growth factor-induced bone resorption in both neonatal mouse calvariae and nude mice bearing a human squamous tumor (242,245,246). The mechanism of its actions in bone resorption remains unclear, but it is mainly inferred as a blocker of osteoclastogenesis (245,246). Our results offer an alternative explanation for the role of suramin in preventing bone resorption. Its relatively low IC_{50} value of 4.8 ± 0.9 μM makes it a potent exosite inhibitor of the collagenolytic activity of cathepsin K. An examination on the structure of suramin leads to the speculation that it may compete with the binding site of C4-S. Both, suramin and C4-S have a chain-like structure composed of polycarbon rings with negatively charged sulfate groups. C4-S may exert stronger binding preferences to cathepsin K due to its longer chain length of repetitive polysaccharides and the greater number of negative charges. Nonetheless, at higher concentrations, the presence of suramin in solution may effectively

challenge the binding of C4-S to cathepsin K. This results in the disruption of the proper formation of the cathepsin K/C4-S complex and thus weakens its collagenolytic activity. AFM-based observation supported such proposition as only single monomers of cathepsin K were found on the mica surface for suramin. Questions may be raised on the 34% inhibition of the Z-FR-MCA hydrolysis by suramin as shown in Table 4.4. Although this reflects a significant interference on the enzyme active site, it is important to point out that such inhibition is measured at a $2.7x10^5$-fold molar excess of the inhibitor. No inhibition was observed on Z-FR-MCA at 3.0 µM (600-fold excess) or on gelatin hydrolysis at 12.5 µM ($2.5x10^3$-fold excess) (Figure 4.4 and Table 4.4).

The majority of the inhibitors identified by FPA belong to polyaromatic compounds such as ellipticine and sanguinarine. In addition, a total of 20 structurally related compounds to those identified in the FPA with known effects on bone resorption were also evaluated in this study. For instance, dihydrotanshinone I, one of the major constituent from the Chinese herbal medical plant, Danshen, prevents bone loss in an estrogen deficient-induced osteoporosis rat model (247). It is also identified to be as an effective inhibitor of RANKL mediated osteoclastogenesis at a concentration of 3.6 µM (248). This concentration is in a similar range to the IC_{50} value of 11.0 ± 3.0 µM for cathepsin K (Table 4.3). SEM analysis revealed that 50 µM dihydrotanshinone I is sufficient to inhibit 3 µM cathepsin K from digesting type I insoluble collagen fibers (Figure 4.6). Its effect on cathepsin K is not the result of its interaction with the enzyme active site as no inhibition was found for the hydrolysis of Z-FR-MCA and gelatin (Figure 4.4 & Table 4.4). Its neutral charge also implies that it may not be a competitor for C4-S. Because the formation of the cathepsin K and C4-S complex involves protein-protein interaction between individual cathepsin K molecules (see chapter 3), it is likely that dihydrotanshinone I disrupts protein-protein interactions in the complex via polar or hydrophobic interactions. This is also supported by the

AFM analysis, which revealed a distortion of the ring-shaped cathepsin K/C4-S structure (Figure 4.5IV). Using molecular docking, dihydrotanshinone I was found to interact with the interface I site for the formation of tetrameric and dimeric cathepsin K complexes. Site-directed mutagenesis demonstrated that residues Met_{97} – Thr_{101} were involved in the protein-protein interaction holding the complexes together (see chapter 3). Dihydrotanshinone I preferentially interacts to Met_{97} and Asn_{99}, which can function like a steric blocker to the complex formation. Although the accuracy of docking using Arguslab remains indeterminate, it provides a potential explanation to the inhibitory mechanism of dihydrotanshinone I on the hydrolysis of soluble and insoluble collagen.

The IC_{50} values of other polyaromatic compounds range from 9.5 – 186.3 μM (Table 4.3). Among them, aurintricarboxylic acid has been reported to be an inhibitor of topoisomerase (249). Several other functions of this compound have also been identified including: a direct inhibitor of neuraminidase to fight against influenza (250); an inhibitory agent to nitric oxide production and thus protects macrophages from LPS-mediated cell death (251). Prior to this study, no linkage between aurintricarboxylic acid and bone resorption had been reported. While its exact mechanism on the collagenolytic activity of cathepsin K remains elucidative, docking has suggested that it might be an effective competitor for C4-S, which could lead to disruption of the active cathepsin K/C4-S complex. Imaging results from AFM evidenced such disruption as no proper ring-formation was observed for a mixture of cathepsin K/C4-S at 2:1 molar ratio in the presence of aurintricarboxylic acid. Its lack of blocking the hydrolysis of Z-FR-MCA and gelatin, aids in its novel role as an exosite site inhibitor of cathepsin K. Interestingly, the structure of aurintricarboxylic acid shares some resemblance to compounds known as selective estrogen-receptor modulators (SERMs). Studies have shown that it could act as an estrogen receptor antagonist (252,253).

SERMs are a diverse group of compounds that binds to estrogen receptors to up regulate or down regulate estrogen-mediated activities depending on type and location of the estrogen receptors (254). Their actions appears paradoxical as certain SERMs can act as an agonist to activate estrogen receptors in some tissues and can also serve as an antagonist in peripheral tissues such as bone. For example, while the first anticancer SERM, tamoxifen, is effective in treating breast cancer patients as an anti-estrogen to block breast cancer cell proliferation, it has also been demonstrated to mimic the actions of estrogen to act as an anti-resorptive (255,256). Depending on the chemical structure, they are classified into six groups comprising of triphenylethylenes (e.g. tamoxifen and clomiphene), benzothiophenes (e.g. reloxifene), indoles (e.g. bazedoxifene), napthalenols (e.g. lasofoxifene), dihydronapthylenes (nafoxidine), and benzopyrans (e.g. acolbifene) (257). Tamoxifen and clomiphene, which have been approved by the FDA for treating other estrogen-related diseases such as breast cancer, have also been shown to reduce bone resorption in osteoporosis (258,259). Although the mechanism of SERMs in bone remodeling has not yet been fully understood, our data also indicate a direct inhibitory effect on the collagenolytic activity of cathepsin K. As shown in Table 4.2, not all SERMs are effective. Among the effective ones, clomiphene citrate, nafoxidine hydrochloride and tamoxifen demonstrated IC_{50} values of approximately 30.3 ± 4.2 μM, 25.6 ± 3.8 μM, and 48.7 ± 10.3 μM whereas that of raloxifene was about 175.5 ± 35.1 μM (Table 4.3). Raloxifene was the first anti-resorptive SERM approved in North America (260-262). Clinical trials indicated that patients up-taking 60 mg/day increased their lumbar spine bone density by 2.5% and femoral neck density by 2% compared to the control after a 2-year therapy, and women with osteoporosis decreased their risk of incident vertebral fractures by 30-50% after a 3-year treatment (261,262). The effective dosage of raloxifene in treating postmenopausal women is comparable to that of odanacatib, which is a reversible drug

selectively targeting the active site of cathepsin K (50 mg/day) (236). Nonetheless, a significantly lower IC_{50} value for odanacatib (~0.2 μM, data no published) against tropocollagen in the presence of 200 nM cathepsin K was found *in vitro* (585-fold less than raloxifene). This makes cathepsin K an effective target for raloxifene *in vivo* rather questionable. Odanacatib resulted greater increases in lumbar spine density (5.5%, 1-year) and total hip bone mineral density (3.2%, 1-year) after a 2-year clinical trial (236).

(-)-Epigallocatechin gallate, which is extracted from green tea and shares some SERM-like structural features, is also effective against cathepsin K. According to Devine *et al.*, non-tea drinkers lose hip bone mineral density nearly 2.5 times faster than tea drinkers over a 4-year analysis (263). Interestingly, AFM studies implied a potential mechanistic difference between (-)-epigallocatechin gallate and clomiphene. While both compounds disrupted the formation of cathepsin K/C4-S complex formation, (-)-epigallocatechin gallate did so by weakening the protein-protein interaction as both monomers and deformed rings were present simultaneously on mica, whereas clomiphene tended to induce non-specific protein-protein interaction as uneven size aggregates were observed.

In conclusion, a novel class of exosite inhibitors of cathepsin K has been identified in this study, which selectively targets the collagenase activity of the protease. They may overcome the intrinsic off-site effects by the active site-directed inhibitors presently in development. These chemicals can also be used as starting templates for future drug design and modifications.

Chapter 5. Conclusions and Suggestions for Future Work

Conclusions

It has become increasingly obvious that the catalytic activities of many proteases are not solely dependent on their active sites. Surface structures that are nearby or distal from the active sites may also play important roles. These surface structures, typically termed exosites, can carry out various functions including substrate recognition, positioning, stabilization and modification. Exosites can also be utilized by proteases to form oligomeric complexes with either the same or different types of proteins. Exosites of this kind, although not directly interacting with substrates, can be crucial for proteases to breakdown of certain specific substrates. This thesis provided examples of exosites in cathepsins that are required for their elastolytic and collagenolytic activities.

In chapter 2, two exosites were identified in cathepsin V that are crucial for the degradation of insoluble elastin. Both exosites are distant from the active site of the protease. Their contribution to the potent elastolytic activity distinguishes cathepsin V from the structurally highly homologous cathepsin L, which shares a 78% amino acid sequence identify but lacks elastase activity. Replacement of exosite 1 (V_{92}AVDEICKYRPEN$_{104}$) or exosite 2 (T_{113}VVAPGK$_{119}$) with analogous amino acid residues from cathepsin L led to a 75% and a 43% loss in the elastolytic activity, respectively. Replacement of both exosites completely abolished the elastolytic activity. Although the exact mechanism of the exosites is yet to be elucidated, the finding that the double exosite variant failed to bind to insoluble elastin implies that these exosites are involved in substrate recognition.

In chapter 3, the role of exosites in conjunction with the collagenase activity of cathepsin K and their implication in protease/GAG complex formation was studied. Monomeric cathepsin K is not capable of digesting collagen as its active cleft (5Å) is not sufficient to accommodate the diameter of a tropocollagen molecule (15Å). Its collagenase activity is largely dependent on the interactions with bone and cartilage resident GAGs such chondroitin and dermatan sulfates. To elucidate this reliance, two putative models including a cathepsin K/GAG tetramer and a dimer complex were proposed based on two available crystal structures. Both models, although displaying different modes of GAG binding, share a number of important amino acid residues in their protein-protein interactions. GAG-dependent collagen degradation showed the highest collagenase activity within a narrow range of a cathepsin K and GAGs molar ratio of approximately ~2:1 which supports both models. Atomic force microscopy revealed ring-like structures in mixtures of 2:1 molar ratio of cathepsin K and C4-S that are reminiscent to the tetramer model. Mutational analysis of protein-protein interactions in the tetramer and dimer models supported the existence and critical roles of tetramer and dimer models in the degradation of soluble and insoluble collagens, respectively. These protein interaction sites can be considered as exosites. In contrast to the elastase-related exosites, these sites do not directly interact with the substrate (here collagen), but participate in either the interaction with GAGs or protein-protein interaction with other cathepsin K molecules. Their presence makes cathepsin K the predominant collagenase in osteoclasts.

In chapter 4, a library of 1280 known drug derivatives was screened using a fluorometric polarization assay to identify potential exosite inhibitors that prevent the formation of cathepsin K/GAG active complexes. Two groups of compounds were identified: 1) polyanionic and 2) polyaromatic compounds. Of particular interest are the polyaromatic compounds. IC_{50} values for

the inhibition of soluble collagen degradation were between 4.7 – 186.3 μM. Some of the identified inhibitors were known anti-resorptive drugs. Although the mechanism of these compounds has been primarily described as transcriptional modulators or compounds interfering with various signal transduction pathways, they were herein demonstrated to be also direct inhibitors of the collagenolytic activity of cathepsin K and might have the advantage of overcoming the off-site effects of active site-directed inhibitors presently in development.

In conclusion, the work presented in this dissertation set an example of using the understanding of enzymatic mechanisms to explore and achieve selective proteolytic control. It diversified our scope from the traditional protein active site to other secondary binding sites on the protein surfaces known as exosites. Using cathepsins V and K as representative enzymes, significant achievements have been made in defining the exosites that participate in two major ECM remodeling events – elastolysis and collagenolysis, respectively. Identification of these exosites as structural requirements of cathepsins to carry out specific substrate hydrolysis also allowed deductions of their proteolytic mechanisms: cathepsin V may utilize the two hydrophobic exosites for elastin absorption prior to hydrolysis; cathepsin K employs exosites to form dimeric and tetrameric oligomers with GAGs for unwinding and degrading collagen. These mechanistic insights further served as foundation for the finding of alternative inhibitors, exosite inhibitors, for the regulation of proteolytic activities. Exosite inhibitors have the advantage over active site inhibitors that they allow substrate specific inhibition without perturbing the enzyme's other biological functions (Figure 5.1).

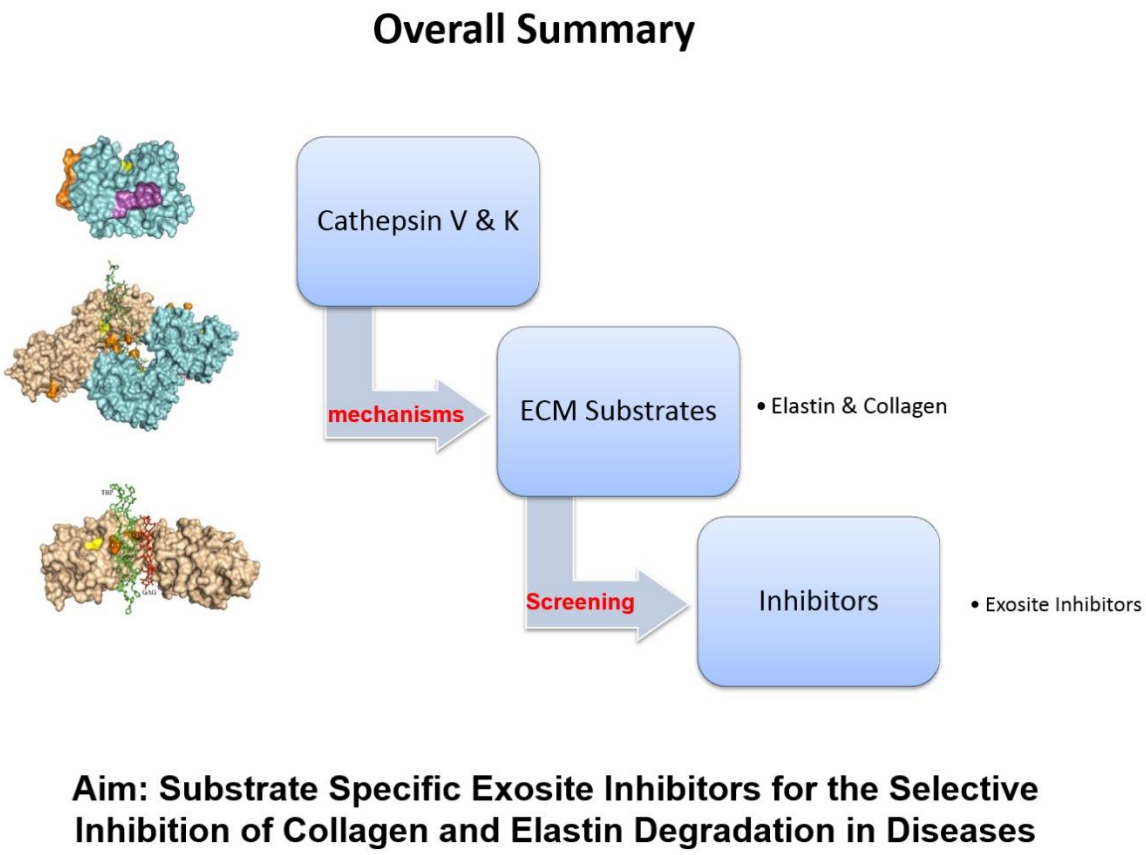


Figure 5.1. Schematic Summary of All Projects. Elucidation of the structural requirements and mechanisms for the elastolytic and collagenolytic activities of cathepsins may allow identification of effective exosite inhibitors.

Suggestions for Future Work

This thesis focused on the identification of exosites as complementary structures to the active sites in cathepsins to achieve their elastolytic and collagenolytic activities, and on the discovery of selective anticollagenase inhibitors. The obtained results inspired further studies:

Chapter 2:

1. Having identified two elastin binding exosites in cathepsin V, it would be interesting to see if similar exosites exist in cathepsins K and S, which are also

potent elastases. Since the 3D structures of cathepsins are highly conserved, the presence of similar elastin binding exosites among elastases in the cathepsin family might be exploited for the development of specific exosite inhibitors.

2. As cathepsin V has also other physiological functions such as the generation of antigen-presentable CLIP complexes in the immune system or releasing endostatin from collagen XVIII and thus effecting angiogenesis (114,116,125), it would be beneficial to identify and develop selective inhibitors that only block elastin degradation of this protease without impairing its active site. Preliminary screening efforts already identified potent anti-elastase inhibitors.

Chapter 3:

1. Although the putatively collagenolytically active tetramer and dimer complex models were derived from available crystal structures, no structures of cathepsin K/GAG complexes interacting with collagen exist. Future work will try to crystallize inactive cathepsin K in the presence of GAGs and a synthetic triple helical peptide. This would provide detailed information about the mechanism of helical peptide unwinding and subsequent cleavage.

2. Scanning electron microscopy and atomic force microscopy analysis of cathepsin K associated with insoluble collagen fibers would provide new information about preferred binding sites and the nature of cathepsin K complexes involved in collagen degradation.

Chapter 4:

1. A novel class of exosite inhibitors was identified in this chapter. Some of these inhibitors could block the collagenase activity of cathepsin K without interfering its active site function. Nonetheless, the exact mechanism of action and the binding sites of these inhibitors are yet to be elucidated. Crystallization studies of cathepsin K/GAG/inhibitor complexes will provide exact information about the binding sites and will allow subsequent and detailed structure-activity-studies (SAR).

2. Compounds like dihydrotanshinone I need to be characterized for their antiresorptive activity in osteoclasts and animal models of osteoporosis. It may also serve as a lead compound for medicinal chemistry efforts to improve the potency of the compound.

3. Professional docking software such as ICM, GLIDE and Surflex can be used for the identification of novel exosite inhibitor scaffolds. Hits identified *in silico* will be verified by fluorescent polarization assay and *in vitro* collagenase and gelatinase assays. Information gathered from modelling studies can also contribute to the understanding of complex-inhibitor interactions and the subsequent drug development.

4. A number of SERMs such as clomiphene, nafoxidine and (-)-epigallocatechin gallate also exhibited inhibitory effects on the collagenase activity of cathepsin K. Modifications of these compounds to reduce their IC_{50} values and to increase their specificity towards cathepsin K would be beneficial.

References

1. Frantz, C., Stewart, K. M., and Weaver, V. M. (2010) *Journal of cell science* **123**, 4195-4200

2. Pizzo, A. M., Kokini, K., Vaughn, L. C., Waisner, B. Z., and Voytik-Harbin, S. L. (2005) *Journal of applied physiology* **98**, 1909-1921

3. Larsen, M., Artym, V. V., Green, J. A., and Yamada, K. M. (2006) *Current opinion in cell biology* **18**, 463-471

4. Harburger, D. S., and Calderwood, D. A. (2009) *Journal of cell science* **122**, 159-163

5. Xian, X., Gopal, S., and Couchman, J. R. (2010) *Cell and tissue research* **339**, 31-46

6. Jarvelainen, H., Sainio, A., Koulu, M., Wight, T. N., and Penttinen, R. (2009) *Pharmacological reviews* **61**, 198-223

7. Schaefer, L., and Schaefer, R. M. (2010) *Cell and tissue research* **339**, 237-246

8. Di Lullo, G. A., Sweeney, S. M., Korkko, J., Ala-Kokko, L., and San Antonio, J. D. (2002) *The Journal of biological chemistry* **277**, 4223-4231

9. Hulmes, D. J. (1992) *Essays in biochemistry* **27**, 49-67

10. Hulmes, D. J. (2002) *Journal of structural biology* **137**, 2-10

11. Bruckner, P., and van der Rest, M. (1994) *Microscopy research and technique* **28**, 378-384

12. Gelse, K., Poschl, E., and Aigner, T. (2003) *Advanced drug delivery reviews* **55**, 1531-1546

13. Viguet-Carrin, S., Garnero, P., and Delmas, P. D. (2006) *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* **17**, 319-336

14. Bachinger, H. P. (1987) *The Journal of biological chemistry* **262**, 17144-17148

15. Galat, A., and Metcalfe, S. M. (1995) *Progress in biophysics and molecular biology* **63**, 67-118

16. Kivirikko, K. I., and Myllyharju, J. (1998) *Matrix biology : journal of the International Society for Matrix Biology* **16**, 357-368

17. Myllyharju, J. (2003) *Matrix biology : journal of the International Society for Matrix Biology* **22**, 15-24

18. Nagata, K. (1998) *Matrix biology : journal of the International Society for Matrix Biology* **16**, 379-386

19. Nagata, K. (2003) *Seminars in cell & developmental biology* **14**, 275-282

20. Engel, J., and Prockop, D. J. (1991) *Annual review of biophysics and biophysical chemistry* **20**, 137-152

21. Knott, L., and Bailey, A. J. (1998) *Bone* **22**, 181-187

22. De Wever, O., Demetter, P., Mareel, M., and Bracke, M. (2008) *International journal of cancer. Journal international du cancer* **123**, 2229-2238

23. Dzamba, B. J., Wu, H., Jaenisch, R., and Peters, D. M. (1993) *The Journal of cell biology* **121**, 1165-1172

24. Hedbom, E., and Heinegard, D. (1989) *The Journal of biological chemistry* **264**, 6898-6905

25. Koyama, Y., Norose-Toyoda, K., Hirano, S., Kobayashi, M., Ebihara, T., Someki, I., Fujisaki, H., and Irie, S. (2000) *Archives of histology and cytology* **63**, 71-79

26. Mizuno, M., Fujisawa, R., and Kuboki, Y. (2000) *Journal of cellular physiology* **184**, 207-213

27. Zhang, G., Ezura, Y., Chervoneva, I., Robinson, P. S., Beason, D. P., Carine, E. T., Soslowsky, L. J., Iozzo, R. V., and Birk, D. E. (2006) *Journal of cellular biochemistry* **98**, 1436-1449

28. Bonadio, J., and Byers, P. H. (1985) *Nature* **316**, 363-366

29. Kamoun-Goldrat, A., Martinovic, J., Saada, J., Sonigo-Cohen, P., Razavi, F., Munnich, A., and Le Merrer, M. (2008) *American journal of medical genetics. Part A* **146A**, 1820-1824

30. Nuytinck, L., Tukel, T., Kayserili, H., Apak, M. Y., and De Paepe, A. (2000) *Journal of medical genetics* **37**, 371-375

31. Forlino, A., Cabral, W. A., Barnes, A. M., and Marini, J. C. (2011) *Nature reviews. Endocrinology* **7**, 540-557

32. Beighton, P., De Paepe, A., Steinmann, B., Tsipouras, P., and Wenstrup, R. J. (1998) *American journal of medical genetics* **77**, 31-37

33. Alexakis, C., Maxwell, P., and Bou-Gharios, G. (2006) *Nephron. Experimental nephrology* **102**, e71-75

34. Katsuda, S., Okada, Y., Minamoto, T., Oda, Y., Matsui, Y., and Nakanishi, I. (1992) *Arteriosclerosis and thrombosis : a journal of vascular biology / American Heart Association* **12**, 494-502

35. Lopez, B., Gonzalez, A., Varo, N., Laviades, C., Querejeta, R., and Diez, J. (2001) *Hypertension* **38**, 1222-1226

36. Sandberg, L. B., Soskel, N. T., and Leslie, J. G. (1981) *The New England journal of medicine* **304**, 566-579

37. Lemos, M., Pozo, R. M., Montes, G. S., and Saldiva, P. H. (1997) *Annals of anatomy Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft* **179**, 447-452

38. Wise, S. G., and Weiss, A. S. (2009) *The international journal of biochemistry & cell biology* **41**, 494-497

39. Rosenbloom, J. (1984) *Laboratory investigation; a journal of technical methods and pathology* **51**, 605-623

40. Vrhovski, B., Jensen, S., and Weiss, A. S. (1997) *European journal of biochemistry / FEBS* **250**, 92-98

41. Hinek, A. (1995) *Ciba Foundation symposium* **192**, 185-191; discussion 191-186

42. Sakai, L. Y., Keene, D. R., and Engvall, E. (1986) *The Journal of cell biology* **103**, 2499-2509

43. Sandberg, L. B., Weissman, N., and Gray, W. R. (1971) *Biochemistry* **10**, 52-56

44. Roten, S. V., Bhat, S., and Bhawan, J. (1996) *Journal of cutaneous pathology* **23**, 37-42

45. Starcher, B. (2001) *Analytical biochemistry* **292**, 125-129

46. Glagov, S., Vito, R., Giddens, D. P., and Zarins, C. K. (1992) *Journal of hypertension. Supplement : official journal of the International Society of Hypertension* **10**, S101-104

47. Wilson, B. D., Gibson, C. C., Sorensen, L. K., Guilhermier, M. Y., Clinger, M., Kelley, L. L., Shiu, Y. T., and Li, D. Y. (2011) *Annals of biomedical engineering* **39**, 337-346

48. Wise, S. G., Byrom, M. J., Waterhouse, A., Bannon, P. G., Weiss, A. S., and Ng, M. K. (2011) *Acta biomaterialia* **7**, 295-303

49. Li, D. Y., Brooke, B., Davis, E. C., Mecham, R. P., Sorensen, L. K., Boak, B. B., Eichwald, E., and Keating, M. T. (1998) *Nature* **393**, 276-280

50. Kamisato, S., Uemura, Y., Takami, N., and Okamoto, K. (1997) *Journal of biochemistry* **121**, 862-867

51. Jung, S., Rutka, J. T., and Hinek, A. (1998) *Journal of neuropathology and experimental neurology* **57**, 439-448

52. Mochizuki, S., Brassart, B., and Hinek, A. (2002) *The Journal of biological chemistry* **277**, 44854-44863

53. Urban, Z., Zhang, J., Davis, E. C., Maeda, G. K., Kumar, A., Stalker, H., Belmont, J. W., Boyd, C. D., and Wallace, M. R. (2001) *Human genetics* **109**, 512-520

54. Hucthagowder, V., Morava, E., Kornak, U., Lefeber, D. J., Fischer, B., Dimopoulou, A., Aldinger, A., Choi, J., Davis, E. C., Abuelo, D. N., Adamowicz, M., Al-Aama, J., Basel-Vanagaite, L., Fernandez, B., Greally, M. T., Gillessen-Kaesbach, G., Kayserili, H., Lemyre, E., Tekin, M., Turkmen, S., Tuysuz, B., Yuksel-Konuk, B., Mundlos, S., Van Maldergem, L., Wevers, R. A., and Urban, Z. (2009) *Human molecular genetics* **18**, 2149-2165

55. Palazzuoli, A., Gallotta, M., Guerrieri, G., Quatrini, I., Franci, B., Campagna, M. S., Neri, E., Benvenuti, A., Sassi, C., and Nuti, R. (2008) *Vascular health and risk management* **4**, 877-883

56. Snider, G. L., Ciccolella, D. E., Morris, S. M., Stone, P. J., and Lucey, E. C. (1991) *Annals of the New York Academy of Sciences* **624**, 45-59

57. Martyn, C. N., and Greenwald, S. E. (1997) *Lancet* **350**, 953-955

58. Lamoureux, F., Baud'huin, M., Duplomb, L., Heymann, D., and Redini, F. (2007) *BioEssays : news and reviews in molecular, cellular and developmental biology* **29**, 758-771

59. Bernfield, M., Gotte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., and Zako, M. (1999) *Annual review of biochemistry* **68**, 729-777

60. Ruoslahti, E. (1988) *Annual review of cell biology* **4**, 229-255

61. Velasco, C. R., Colliec-Jouault, S., Redini, F., Heymann, D., and Padrines, M. (2010) *Drug discovery today* **15**, 553-560

62. Asimakopoulou, A. P., Theocharis, A. D., Tzanakakis, G. N., and Karamanos, N. K. (2008) *In vivo* **22**, 385-389

63. Afratis, N., Gialeli, C., Nikitovic, D., Tsegenidis, T., Karousou, E., Theocharis, A. D., Pavao, M. S., Tzanakakis, G. N., and Karamanos, N. K. (2012) *The FEBS journal* **279**, 1177-1197

64. Kwok, J. C., Dick, G., Wang, D., and Fawcett, J. W. (2011) *Developmental neurobiology* **71**, 1073-1089

65. Li, Z., Hou, W. S., Escalante-Torres, C. R., Gelb, B. D., and Bromme, D. (2002) *The Journal of biological chemistry* **277**, 28669-28676

66. Theocharis, A. D., Seidel, C., Borset, M., Dobra, K., Baykov, V., Labropoulou, V., Kanakis, I., Dalas, E., Karamanos, N. K., Sundan, A., and Hjerpe, A. (2006) *The Journal of biological chemistry* **281**, 35116-35128

67. ten Dam, G. B., van de Westerlo, E. M., Purushothaman, A., Stan, R. V., Bulten, J., Sweep, F. C., Massuger, L. F., Sugahara, K., and van Kuppevelt, T. H. (2007) *Am J Pathol* **171**, 1324-1333

68. Skliris, A., Happonen, K. E., Terpos, E., Labropoulou, V., Borset, M., Heinegard, D., Blom, A. M., and Theocharis, A. D. (2011) *European journal of immunology* **41**, 437-449

69. Pumphrey, C. Y., Theus, A. M., Li, S., Parrish, R. S., and Sanderson, R. D. (2002) *Cancer research* **62**, 3722-3728

70. Lee, C. M., Tanaka, T., Murai, T., Kondo, M., Kimura, J., Su, W., Kitagawa, T., Ito, T., Matsuda, H., and Miyasaka, M. (2002) *Cancer research* **62**, 4282-4288

71. Miller, K. L., and Clegg, D. O. (2011) *Rheumatic diseases clinics of North America* **37**, 103-118

72. Rawlings, N. D., Morton, F. R., and Barrett, A. J. (2006) *Nucleic acids research* **34**, D270-272

73. Stoka, V., Turk, B., and Turk, V. (2005) *IUBMB life* **57**, 347-353

74. Lopez-Otin, C., and Overall, C. M. (2002) *Nature reviews* **3**, 509-519

75. Lecaille, F., Kaleta, J., and Bromme, D. (2002) *Chemical reviews* **102**, 4459-4488

76. McGrath, M. E. (1999) *Annual review of biophysics and biomolecular structure* **28**, 181-204

77. Turk, D., and Guncar, G. (2003) *Acta crystallographica* **59**, 203-213

78. Chapman, R. L., Kane, S. E., and Erickson, A. H. (1997) *The Journal of biological chemistry* **272**, 8808-8816

79. Nissler, K., Kreusch, S., Rommerskirch, W., Strubel, W., Weber, E., and Wiederanders, B. (1998) *Biological chemistry* **379**, 219-224

80. Schilling, K., Pietschmann, S., Fehn, M., Wenz, I., and Wiederanders, B. (2001) *Biological chemistry* **382**, 859-865

81. Yamamoto, Y., Watabe, S., Kageyama, T., and Takahashi, S. Y. (1999) *Archives of insect biochemistry and physiology* **42**, 167-178

82. Billington, C. J., Mason, P., Magny, M. C., and Mort, J. S. (2000) *Biochemical and biophysical research communications* **276**, 924-929

83. Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. S., and Menard, R. (1996) *Biochemistry* **35**, 8149-8157

84. Fox, T., de Miguel, E., Mort, J. S., and Storer, A. C. (1992) *Biochemistry* **31**, 12571-12576

85. Guo, Y. L., Kurz, U., Schultz, J. E., Lim, C. C., Wiederanders, B., and Schilling, K. (2000) *FEBS letters* **469**, 203-207

86. Dickinson, D. P. (2002) *Crit Rev Oral Biol Med* **13**, 238-275

87. Guncar, G., Pungercic, G., Klemencic, I., Turk, V., and Turk, D. (1999) *The EMBO journal* **18**, 793-803

88. Schechter, I., and Berger, A. (1967) *Biochemical and biophysical research communications* **27**, 157-162

89. Lecaille, F., Bromme, D., and Lalmanach, G. (2008) *Biochimie* **90**, 208-226

90. Scott, J. E. (1988) *The Biochemical journal* **252**, 313-323

91. Tiffany, M. L., and Krimm, S. (1972) *Biopolymers* **11**, 2309-2316

92. Turk, B., Turk, D., and Turk, V. (2000) *Biochimica et biophysica acta* **1477**, 98-111

93. Kirschke, H., Schmidt, I., and Wiederanders, B. (1986) *The Biochemical journal* **240**, 455-459

94. Bromme, D., Bonneau, P. R., Lachance, P., Wiederanders, B., Kirschke, H., Peters, C., Thomas, D. Y., Storer, A. C., and Vernet, T. (1993) *The Journal of biological chemistry* **268**, 4832-4838

95. Xia, L., Kilb, J., Wex, H., Li, Z., Lipyansky, A., Breuil, V., Stein, L., Palmer, J. T., Dempster, D. W., and Bromme, D. (1999) *Biological chemistry* **380**, 679-687

96. Bromme, D., Okamoto, K., Wang, B. B., and Biroc, S. (1996) *The Journal of biological chemistry* **271**, 2126-2132

97. Brömme, D., and Wilson, S. (2011) Role of Cysteine Cathepsins in Extracellular Proteolysis. in *Extracellular Matrix Degradation* (Parks, W. C., and Mecham, R. P. eds.), Springer Berlin Heidelberg. pp 23-51

98. Godat, E., Herve-Grvepinet, V., Veillard, F., Lecaille, F., Belghazi, M., Bromme, D., and Lalmanach, G. (2008) *Biological chemistry* **389**, 1123-1126

99. Jordans, S., Jenko-Kokalj, S., Kuhl, N. M., Tedelind, S., Sendt, W., Bromme, D., Turk, D., and Brix, K. (2009) *BMC biochemistry* **10**, 23

100. Linnevers, C., Smeekens, S. P., and Bromme, D. (1997) *FEBS letters* **405**, 253-259

101. Wex, T., Buhling, F., Wex, H., Gunther, D., Malfertheiner, P., Weber, E., and Bromme, D. (2001) *J Immunol* **167**, 2172-2178

102. Hou, W. S., Li, Z., Buttner, F. H., Bartnik, E., and Bromme, D. (2003) *Biological chemistry* **384**, 891-897

103. Buhling, F., Waldburg, N., Reisenauer, A., Heimburg, A., Golpon, H., and Welte, T. (2004) *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **23**, 620-628

104. Zheng, T., Zhu, Z., Wang, Z., Homer, R. J., Ma, B., Riese, R. J., Jr., Chapman, H. A., Jr., Shapiro, S. D., and Elias, J. A. (2000) *J Clin Invest* **106**, 1081-1093

105. Roshy, S., Sloane, B. F., and Moin, K. (2003) *Cancer metastasis reviews* **22**, 271-286

106. Mohamed, M. M., and Sloane, B. F. (2006) *Nature reviews. Cancer* **6**, 764-775

107. Coulibaly, S., Schwihla, H., Abrahamson, M., Albini, A., Cerni, C., Clark, J. L., Ng, K. M., Katunuma, N., Schlappack, O., Glossl, J., and Mach, L. (1999) *International journal of cancer. Journal international du cancer* **83**, 526-531

108. Santamaria, I., Velasco, G., Cazorla, M., Fueyo, A., Campo, E., and Lopez-Otin, C. (1998) *Cancer research* **58**, 1624-1630

109. Bromme, D., Li, Z., Barnes, M., and Mehler, E. (1999) *Biochemistry* **38**, 2377-2385

110. Itoh, R., Kawamoto, S., Adachi, W., Kinoshita, S., and Okubo, K. (1999) *DNA research : an international journal for rapid publication of reports on genes and genomes* **6**, 137-140

111. Niwa, Y., Suzuki, T., Dohmae, N., Umezawa, K., and Simizu, S. (2012) *FEBS letters*

112. Adachi, W., Kawamoto, S., Ohno, I., Nishida, K., Kinoshita, S., Matsubara, K., and Okubo, K. (1998) *Investigative ophthalmology & visual science* **39**, 1789-1796

113. Funkelstein, L., Lu, W. D., Koch, B., Mosier, C., Toneff, T., Taupenot, L., O'Connor, D. T., Reinheckel, T., Peters, C., and Hook, V. (2012) *The Journal of biological chemistry* **287**, 15232-15241

114. Ma, D. H., Yao, J. Y., Kuo, M. T., See, L. C., Lin, K. Y., Chen, S. C., Chen, J. K., Chao, A. S., Wang, S. F., and Lin, K. K. (2007) *Investigative ophthalmology & visual science* **48**, 644-651

115. Sevenich, L., Hagemann, S., Stoeckle, C., Tolosa, E., Peters, C., and Reinheckel, T. (2010) *Biochimie* **92**, 1674-1680

116. Tolosa, E., Li, W., Yasuda, Y., Wienhold, W., Denzin, L. K., Lautwein, A., Driessen, C., Schnorrer, P., Weber, E., Stevanovic, S., Kurek, R., Melms, A., and Bromme, D. (2003) *J Clin Invest* **112**, 517-526

117. Munglani, R., Hudspith, M. J., and Hunt, S. P. (1996) *Drugs* **52**, 371-389

118. Baltatzi, M., Hatzitolios, A., Tziomalos, K., Iliadis, F., and Zamboulis, C. (2008) *International journal of clinical practice* **62**, 1432-1440

119. Sato, N., Ogino, Y., Mashiko, S., and Ando, M. (2009) *Expert opinion on therapeutic patents* **19**, 1401-1415

120. Kieffer, B. L., and Gaveriaux-Ruff, C. (2002) *Progress in neurobiology* **66**, 285-306

121. Holden, J. E., Jeong, Y., and Forrest, J. M. (2005) *AACN clinical issues* **16**, 291-301

122. Somoza, J. R., Zhan, H., Bowman, K. K., Yu, L., Mortara, K. D., Palmer, J. T., Clark, J. M., and McGrath, M. E. (2000) *Biochemistry* **39**, 12543-12551

123. Berti, P. J., and Storer, A. C. (1995) *J Mol Biol* **246**, 273-283

124. Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Bromme, D., Ellman, J. A., and Craik, C. S. (2006) *The Journal of biological chemistry* **281**, 12824-12832

125. Yasuda, Y., Li, Z., Greenbaum, D., Bogyo, M., Weber, E., and Bromme, D. (2004) *The Journal of biological chemistry* **279**, 36761-36770

126. Rood, J. A., Van Horn, S., Drake, F. H., Gowen, M., and Debouck, C. (1997) *Genomics* **41**, 169-176

127. Gelb, B. D., Shi, G. P., Heller, M., Weremowicz, S., Morton, C., Desnick, R. J., and Chapman, H. A. (1997) *Genomics* **41**, 258-262

128. Ling, H., Vamvakas, S., Busch, G., Dammrich, J., Schramm, L., Lang, F., and Heidland, A. (1995) *The American journal of physiology* **269**, F911-917

129. Lugering, N., Kucharzik, T., Gockel, H., Sorg, C., Stoll, R., and Domschke, W. (1998) *Clinical and experimental immunology* **114**, 377-384

130. Sukhova, G. K., Shi, G. P., Simon, D. I., Chapman, H. A., and Libby, P. (1998) *J Clin Invest* **102**, 576-583

131. Bromme, D., and Okamoto, K. (1995) *Biol Chem Hoppe Seyler* **376**, 379-384

132. van den Brule, S., Misson, P., Buhling, F., Lison, D., and Huaux, F. (2005) *Respir Res* **6**, 84

133. Hou, W. S., Bromme, D., Zhao, Y., Mehler, E., Dushey, C., Weinstein, H., Miranda, C. S., Fraga, C., Greig, F., Carey, J., Rimoin, D. L., Desnick, R. J., and Gelb, B. D. (1999) *J Clin Invest* **103**, 731-738

134. Haeckel, C., Krueger, S., Buehling, F., Broemme, D., Franke, K., Schuetze, A., Roese, I., and Roessner, A. (1999) *Dev Dyn* **216**, 89-95

135. Li, Z., Kienetz, M., Cherney, M. M., James, M. N., and Bromme, D. (2008) *J Mol Biol* **383**, 78-91

136. Blair, H. C., Teitelbaum, S. L., Ghiselli, R., and Gluck, S. (1989) *Science (New York, N.Y* **245**, 855-857

137. Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J. L., Fields, G. B., Visse, R., and Nagase, H. (2004) *The EMBO journal* **23**, 3020-3030

138. Garnero, P., Borel, O., Byrjalsen, I., Ferreras, M., Drake, F. H., McQueney, M. S., Foged, N. T., Delmas, P. D., and Delaisse, J. M. (1998) *The Journal of biological chemistry* **273**, 32347-32352

139. Kafienah, W., Bromme, D., Buttle, D. J., Croucher, L. J., and Hollander, A. P. (1998) *The Biochemical journal* **331 (Pt 3)**, 727-732

140. Meier, C., Meinhardt, U., Greenfield, J. R., De Winter, J., Nguyen, T. V., Dunstan, C. R., and Seibel, M. J. (2006) *Clinical laboratory* **52**, 1-10

141. Yasuda, Y., Kaleta, J., and Bromme, D. (2005) *Advanced drug delivery reviews* **57**, 973-993

142. Stroup, G. B., Lark, M. W., Veber, D. F., Bhattacharyya, A., Blake, S., Dare, L. C., Erhard, K. F., Hoffman, S. J., James, I. E., Marquis, R. W., Ru, Y., Vasko-Moser, J. A., Smith, B. R., Tomaszek, T., and Gowen, M. (2001) *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **16**, 1739-1746

143. Donnarumma, M., Regis, S., Tappino, B., Rosano, C., Assereto, S., Corsolini, F., Di Rocco, M., and Filocamo, M. (2007) *Human mutation* **28**, 524

144. Gelb, B. D., Shi, G. P., Chapman, H. A., and Desnick, R. J. (1996) *Science (New York, N.Y* **273**, 1236-1238

145. Schilling, A. F., Mulhausen, C., Lehmann, W., Santer, R., Schinke, T., Rueger, J. M., and Amling, M. (2007) *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* **18**, 659-669

146. Motyckova, G., and Fisher, D. E. (2002) *Current molecular medicine* **2**, 407-421

147. Lutgens, E., Lutgens, S. P., Faber, B. C., Heeneman, S., Gijbels, M. M., de Winther, M. P., Frederik, P., van der Made, I., Daugherty, A., Sijbers, A. M., Fisher, A., Long, C. J., Saftig, P., Black, D., Daemen, M. J., and Cleutjens, K. B. (2006) *Circulation* **113**, 98-107

148. Van den Wyngaert, T., Huizing, M. T., and Vermorken, J. B. (2007) *Current opinion in oncology* **19**, 315-322

149. Woo, S. B., Hellstein, J. W., and Kalmar, J. R. (2006) *Annals of internal medicine* **144**, 753-761

150. Kavanagh, K. L., Guo, K., Dunford, J. E., Wu, X., Knapp, S., Ebetino, F. H., Rogers, M. J., Russell, R. G., and Oppermann, U. (2006) *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7829-7834

151. Stephenson, J. (2003) *JAMA : the journal of the American Medical Association* **289**, 537-538

152. Bromme, D., and Lecaille, F. (2009) *Expert opinion on investigational drugs* **18**, 585-600

153. Yamashita, D. S., and Dodds, R. A. (2000) *Current pharmaceutical design* **6**, 1-24

154. Kim, T. S., and Tasker, A. S. (2006) *Current topics in medicinal chemistry* **6**, 355-360

155. Peroni, A., Zini, A., Braga, V., Colato, C., Adami, S., and Girolomoni, G. (2008) *Journal of the American Academy of Dermatology* **59**, 125-129

156. Eisman, J. A., Bone, H. G., Hosking, D. J., McClung, M. R., Reid, I. R., Rizzoli, R., Resch, H., Verbruggen, N., Hustad, C. M., DaSilva, C., Petrovic, R., Santora, A. C., Ince, B. A.,

and Lombardi, A. (2011) *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **26**, 242-251

157. Falgueyret, J. P., Desmarais, S., Oballa, R., Black, W. C., Cromlish, W., Khougaz, K., Lamontagne, S., Masse, F., Riendeau, D., Toulmond, S., and Percival, M. D. (2005) *Journal of medicinal chemistry* **48**, 7535-7543

158. Ng, K. W. (2012) *Clinical interventions in aging* **7**, 235-247

159. Desmarais, S., Black, W. C., Oballa, R., Lamontagne, S., Riendeau, D., Tawa, P., Duong le, T., Pickarski, M., and Percival, M. D. (2008) *Molecular pharmacology* **73**, 147-156

160. Obrink, B. (1973) *European journal of biochemistry / FEBS* **33**, 387-400

161. Yang, Y. L., Sun, C., Wilhelm, M. E., Fox, L. J., Zhu, J., and Kaufman, L. J. (2011) *Biomaterials* **32**, 7932-7940

162. Choy, A. T., Leong, K. W., and Chan, B. P. (2012) *Acta biomaterialia*

163. Parfitt, G. J., Pinali, C., Akama, T. O., Young, R. D., Nishida, K., Quantock, A. J., and Knupp, C. (2011) *Journal of structural biology* **174**, 536-541

164. Harley, B. A., Leung, J. H., Silva, E. C., and Gibson, L. J. (2007) *Acta biomaterialia* **3**, 463-474

165. Wilson, S., Hashamiyan, S., Clarke, L., Saftig, P., Mort, J., Dejica, V. M., and Bromme, D. (2009) *Am J Pathol* **175**, 2053-2062

166. Lonky, S. A., and Wohl, H. (1983) *Biochemistry* **22**, 3714-3720

167. Heinz, A., Jung, M. C., Jahreis, G., Rusciani, A., Duca, L., Debelle, L., Weiss, A. S., Neubert, R. H., and Schmelzer, C. E. (2012) *Biochimie* **94**, 192-202

168. Novinec, M., Grass, R. N., Stark, W. J., Turk, V., Baici, A., and Lenarcic, B. (2007) *The Journal of biological chemistry* **282**, 7893-7902

169. Ying, Q. L., and Simon, S. R. (2002) *American journal of respiratory cell and molecular biology* **26**, 356-361

170. Fulcher, Y. G., and Van Doren, S. R. (2011) *Biochemistry* **50**, 9488-9499

171. Palmier, M. O., Fulcher, Y. G., Bhaskaran, R., Duong, V. Q., Fields, G. B., and Van Doren, S. R. (2010) *The Journal of biological chemistry* **285**, 30918-30930

172. Pilcher, B. K., Dumin, J. A., Sudbeck, B. D., Krane, S. M., Welgus, H. G., and Parks, W. C. (1997) *The Journal of cell biology* **137**, 1445-1457

173. Buhling, F., Rocken, C., Brasch, F., Hartig, R., Yasuda, Y., Saftig, P., Bromme, D., and Welte, T. (2004) *Am J Pathol* **164**, 2203-2216

174. Egeblad, M., and Werb, Z. (2002) *Nature reviews. Cancer* **2**, 161-174

175. Sternlicht, M. D., and Werb, Z. (2001) *Annual review of cell and developmental biology* **17**, 463-516

176. Bertini, I., Fragai, M., Luchinat, C., Melikian, M., Toccafondi, M., Lauer, J. L., and Fields, G. B. (2012) *Journal of the American Chemical Society* **134**, 2100-2110

177. Lauer-Fields, J. L., Chalmers, M. J., Busby, S. A., Minond, D., Griffin, P. R., and Fields, G. B. (2009) *The Journal of biological chemistry* **284**, 24017-24024

178. Chung, L., Shimokawa, K., Dinakarpandian, D., Grams, F., Fields, G. B., and Nagase, H. (2000) *The Journal of biological chemistry* **275**, 29610-29617

179. Knauper, V., Patterson, M. L., Gomis-Ruth, F. X., Smith, B., Lyons, A., Docherty, A. J., and Murphy, G. (2001) *European journal of biochemistry / FEBS* **268**, 1888-1896

180. Li, Z., Hou, W. S., and Bromme, D. (2000) *Biochemistry* **39**, 529-536

181. Debelle, L., and Tamburro, A. M. (1999) *The international journal of biochemistry & cell biology* **31**, 261-272

182. Shapiro, S. D., Endicott, S. K., Province, M. A., Pierce, J. A., and Campbell, E. J. (1991) *J Clin Invest* **87**, 1828-1834

183. Shi, G. P., Sukhova, G. K., Kuzuya, M., Ye, Q., Du, J., Zhang, Y., Pan, J. H., Lu, M. L., Cheng, X. W., Iguchi, A., Perrey, S., Lee, A. M., Chapman, H. A., and Libby, P. (2003) *Circulation research* **92**, 493-500

184. Knauper, V., Lopez-Otin, C., Smith, B., Knight, G., and Murphy, G. (1996) *The Journal of biological chemistry* **271**, 1544-1550

185. Dauth, S., Sirbulescu, R. F., Jordans, S., Rehders, M., Avena, L., Oswald, J., Lerchl, A., Saftig, P., and Brix, K. (2011) *BMC neuroscience* **12**, 74

186. Asagiri, M., Hirai, T., Kunigami, T., Kamano, S., Gober, H. J., Okamoto, K., Nishikawa, K., Latz, E., Golenbock, D. T., Aoki, K., Ohya, K., Imai, Y., Morishita, Y., Miyazono, K., Kato, S., Saftig, P., and Takayanagi, H. (2008) *Science (New York, N.Y* **319**, 624-627

187. Herroon, M. K., Rajagurubandara, E., Rudy, D. L., Chalasani, A., Hardaway, A. L., and Podgorski, I. (2012) *Oncogene*

188. Podgorski, I., Linebaugh, B. E., Koblinski, J. E., Rudy, D. L., Herroon, M. K., Olive, M. B., and Sloane, B. F. (2009) *Am J Pathol* **175**, 1255-1269

189. Tamburro, A. M., Pepe, A., and Bochicchio, B. (2006) *Biochemistry* **45**, 9518-9530
190. Robert, L., Robert, A. M., and Fulop, T. (2008) *Biogerontology* **9**, 119-133

191. Harman, D. (2001) *Annals of the New York Academy of Sciences* **928**, 1-21

192. Makrantonaki, E., and Zouboulis, C. C. (2007) *Dermatology* **214**, 352-360

193. Robert, L., Robert, A. M., and Jacotot, B. (1998) *Atherosclerosis* **140**, 281-295

194. Liu, J., Sukhova, G. K., Sun, J. S., Xu, W. H., Libby, P., and Shi, G. P. (2004) *Arterioscler Thromb Vasc Biol* **24**, 1359-1366

195. Shapiro, S. D. (1998) *Current opinion in cell biology* **10**, 602-608

196. Wiedow, O., Schroder, J. M., Gregory, H., Young, J. A., and Christophers, E. (1990) *The Journal of biological chemistry* **265**, 14791-14795

197. Hornebeck, W., and Emonard, H. (2011) *Frontiers in bioscience : a journal and virtual library* **16**, 707-722

198. Matsumoto, S., Kobayashi, T., Katoh, M., Saito, S., Ikeda, Y., Kobori, M., Masuho, Y., and Watanabe, T. (1998) *Am J Pathol* **153**, 109-119

199. Puzer, L., Cotrin, S. S., Alves, M. F., Egborge, T., Araujo, M. S., Juliano, M. A., Juliano, L., Bromme, D., and Carmona, A. K. (2004) *Archives of biochemistry and biophysics* **430**, 274-283

200. Bashir, M. M., Indik, Z., Yeh, H., Ornstein-Goldstein, N., Rosenbloom, J. C., Abrams, W., Fazio, M., Uitto, J., and Rosenbloom, J. (1989) *The Journal of biological chemistry* **264**, 8887-8891

201. He, D., Chung, M., Chan, E., Alleyne, T., Ha, K. C., Miao, M., Stahl, R. J., Keeley, F. W., and Parkinson, J. (2007) *Matrix biology : journal of the International Society for Matrix Biology* **26**, 524-540

202. Li, D. Y., Toland, A. E., Boak, B. B., Atkinson, D. L., Ensing, G. J., Morris, C. A., and Keating, M. T. (1997) *Human molecular genetics* **6**, 1021-1028

203. Raju, K., and Anwar, R. A. (1987) *The Journal of biological chemistry* **262**, 5755-5762
204. Schmelzer, C. E., Getie, M., and Neubert, R. H. (2005) *Journal of chromatography. A* **1083**, 120-126

205. Yeh, H., Anderson, N., Ornstein-Goldstein, N., Bashir, M. M., Rosenbloom, J. C., Abrams, W., Indik, Z., Yoon, K., Parks, W., Mecham, R., and et al. (1989) *Biochemistry* **28**, 2365-2370

206. Brown-Augsburger, P., Broekelmann, T., Mecham, L., Mercer, R., Gibson, M. A., Cleary, E. G., Abrams, W. R., Rosenbloom, J., and Mecham, R. P. (1994) *The Journal of biological chemistry* **269**, 28443-28449

207. Foster, J. A., Bruenger, E., Rubin, L., Imberman, M., Kagan, H., Mecham, R., and Franzblau, C. (1976) *Biopolymers* **15**, 833-841

208. Gallop, P. M., Blumenfeld, O. O., and Seifter, S. (1972) *Annual review of biochemistry* **41**, 617-672

209. Gerber, G. E., and Anwar, R. A. (1975) *The Biochemical journal* **149**, 685-695

210. Reiser, K., McCormick, R. J., and Rucker, R. B. (1992) *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **6**, 2439-2449

211. Luisetti, M., Ma, S., Iadarola, P., Stone, P. J., Viglio, S., Casado, B., Lin, Y. Y., Snider, G. L., and Turino, G. M. (2008) *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **32**, 1146-1157

212. Hardegger, L. A., Kuhn, B., Spinnler, B., Anselm, L., Ecabert, R., Stihle, M., Gsell, B., Thoma, R., Diez, J., Benz, J., Plancher, J. M., Hartmann, G., Banner, D. W., Haap, W., and Diederich, F. (2011) *Angewandte Chemie* **50**, 314-318

213. Kyte, J., and Doolittle, R. F. (1982) *J Mol Biol* **157**, 105-132

214. Le Brun, A. P., Chow, J., Bax, D. V., Nelson, A., Weiss, A. S., and James, M. (2012) *Biomacromolecules* **13**, 379-386

215. Alves, M. F., Puzer, L., Cotrin, S. S., Juliano, M. A., Juliano, L., Bromme, D., and Carmona, A. K. (2003) *The Biochemical journal* **373**, 981-986

216. Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S., and Bromme, D. (2002) *Biochemistry* **41**, 8447-8454

217. Zhang, D., Leung, N., Weber, E., Saftig, P., and Bromme, D. (2011) *Respir Res* **12**, 72

218. Venyaminov, S., Baikalov, I. A., Shen, Z. M., Wu, C. S., and Yang, J. T. (1993) *Analytical biochemistry* **214**, 17-24

219. Venyaminov, S., Baikalov, I. A., Wu, C. S., and Yang, J. T. (1991) *Analytical biochemistry* **198**, 250-255

220. Manavalan, P., and Johnson, W. C., Jr. (1987) *Analytical biochemistry* **167**, 76-85

221. LeBaron, R. G., Hook, A., Esko, J. D., Gay, S., and Hook, M. (1989) *The Journal of biological chemistry* **264**, 7950-7956

222. Munakata, H., Takagaki, K., Majima, M., and Endo, M. (1999) *Glycobiology* **9**, 1023-1027

223. Raspanti, M., Viola, M., Forlino, A., Tenni, R., Gruppi, C., and Tira, M. E. (2008) *Journal of structural biology* **164**, 134-139

224. Yaoi, Y., Hashimoto, K., Koitabashi, H., Takahara, K., Ito, M., and Kato, I. (1990) *Biochimica et biophysica acta* **1035**, 139-145

225. Cherney, M. M., Lecaille, F., Kienitz, M., Nallaseth, F. S., Li, Z., James, M. N., and Bromme, D. (2011) *The Journal of biological chemistry* **286**, 8988-8998

226. Li, Z., Yasuda, Y., Li, W., Bogyo, M., Katz, N., Gordon, R. E., Fields, G. B., and Bromme, D. (2004) *The Journal of biological chemistry* **279**, 5470-5479

227. McGrath, M. E., Klaus, J. L., Barnes, M. G., and Bromme, D. (1997) *Nature structural biology* **4**, 105-109

228. Selent, J., Kaleta, J., Li, Z., Lalmanach, G., and Bromme, D. (2007) *The Journal of biological chemistry* **282**, 16492-16501

229. Robichaud, J., Oballa, R., Prasit, P., Falgueyret, J. P., Percival, M. D., Wesolowski, G., Rodan, S. B., Kimmel, D., Johnson, C., Bryant, C., Venkatraman, S., Setti, E., Mendonca, R., and Palmer, J. T. (2003) *Journal of medicinal chemistry* **46**, 3709-3727

230. Stoch, S. A., and Wagner, J. A. (2008) *Clinical pharmacology and therapeutics* **83**, 172-176

231. Khosla, S., Burr, D., Cauley, J., Dempster, D. W., Ebeling, P. R., Felsenberg, D., Gagel, R. F., Gilsanz, V., Guise, T., Koka, S., McCauley, L. K., McGowan, J., McKee, M. D., Mohla, S., Pendrys, D. G., Raisz, L. G., Ruggiero, S. L., Shafer, D. M., Shum, L., Silverman, S. L., Van Poznak, C. H., Watts, N., Woo, S. B., Shane, E., American Society for, B., and Mineral, R. (2007) *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **22**, 1479-1491

232. Lenart, B. A., Lorich, D. G., and Lane, J. M. (2008) *The New England journal of medicine* **358**, 1304-1306

233. Odvina, C. V., Zerwekh, J. E., Rao, D. S., Maalouf, N., Gottschalk, F. A., and Pak, C. Y. (2005) *The Journal of clinical endocrinology and metabolism* **90**, 1294-1301

234. Shane, E., Burr, D., Ebeling, P. R., Abrahamsen, B., Adler, R. A., Brown, T. D., Cheung, A. M., Cosman, F., Curtis, J. R., Dell, R., Dempster, D., Einhorn, T. A., Genant, H. K., Geusens, P., Klaushofer, K., Koval, K., Lane, J. M., McKiernan, F., McKinney, R., Ng, A., Nieves, J., O'Keefe, R., Papapoulos, S., Sen, H. T., van der Meulen, M. C., Weinstein, R. S., Whyte, M., American Society for, B., and Mineral, R. (2010) *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **25**, 2267-2294

235. Bone, H. (2012) *Endocrinology and metabolism clinics of North America* **41**, 655-661

236. Bone, H. G., McClung, M. R., Roux, C., Recker, R. R., Eisman, J. A., Verbruggen, N., Hustad, C. M., DaSilva, C., Santora, A. C., and Ince, B. A. (2010) *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **25**, 937-947

237. Linnevers, C. J., McGrath, M. E., Armstrong, R., Mistry, F. R., Barnes, M. G., Klaus, J. L., Palmer, J. T., Katz, B. A., and Bromme, D. (1997) *Protein Sci* **6**, 919-921

238. Rankovic, Z., Cai, J., Kerr, J., Fradera, X., Robinson, J., Mistry, A., Hamilton, E., McGarry, G., Andrews, F., Caulfield, W., Cumming, I., Dempster, M., Waller, J., Scullion, P., Martin, I., Mitchell, A., Long, C., Baugh, M., Westwood, P., Kinghorn, E., Bruin, J., Hamilton, W., Uitdehaag, J., van Zeeland, M., Potin, D., Saniere, L., Fouquet, A., Chevallier, F., Deronzier, H., Dorleans, C., and Nicolai, E. (2010) *Bioorganic & medicinal chemistry letters* **20**, 1524-1527

239. Ingham, K. C., Brew, S. A., Migliorini, M. M., and Busby, T. F. (1993) *Biochemistry* **32**, 12548-12553

240. Jameson, D. M., and Seifried, S. E. (1999) *Methods* **19**, 222-233

241. Sakamoto, S. (1986) *The Compendium of continuing education in dentistry* **Suppl 7**, S221-223, S226

242. Yoneda, T., Williams, P., Rhine, C., Boyce, B. F., Dunstan, C., and Mundy, G. R. (1995) *Cancer research* **55**, 1989-1993

243. Panwar, P., Du, X., Sharma, V., Lamour, G., Castro, M., Li, H., and Bromme, D. (2013) *The Journal of biological chemistry*

244. Stein, C. A. (1993) *Cancer research* **53**, 2239-2248

245. Farsoudi, K. H., Pietschmann, P., Cross, H. S., and Peterlik, M. (1993) *The Journal of pharmacology and experimental therapeutics* **264**, 579-583

246. Walther, M. M., Kragel, P. J., Trahan, E., Venzon, D., Blair, H. C., Schlesinger, P. H., Jamai-Dow, C., Ewing, M. W., Myers, C. E., and Linehan, W. M. (1992) *Endocrinology* **131**, 2263-2270

247. Chae, H. J., Chae, S. W., Yun, D. H., Keum, K. S., Yoo, S. K., and Kim, H. R. (2004) *Immunopharmacology and immunotoxicology* **26**, 135-144

248. Kim, H. K., Woo, E. R., Lee, H. W., Park, H. R., Kim, H. N., Jung, Y. K., Choi, J. Y., Chae, S. W., Kim, H. R., and Chae, H. J. (2008) *Immunopharmacology and immunotoxicology* **30**, 347-364

249. Benchokroun, Y., Couprie, J., and Larsen, A. K. (1995) *Biochemical pharmacology* **49**, 305-313

250. Hashem, A. M., Flaman, A. S., Farnsworth, A., Brown, E. G., Van Domselaar, G., He, R., and Li, X. (2013) *PloS one* **8**

251. Tsi, C. J., Chao, Y., Chen, C. W., and Lin, W. W. (2002) *Molecular pharmacology* **62**, 90-101

252. Moudgil, V. K., and Eessalu, T. E. (1980) *Life sciences* **27**, 1159-1167

253. Moudgil, V. K., and Weekes, G. A. (1978) *FEBS letters* **94**, 324-326

254. Burge, R., Dawson-Hughes, B., Solomon, D. H., Wong, J. B., King, A., and Tosteson, A. (2007) *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **22**, 465-475

255. Fisher, B., Costantino, J. P., Redmond, C. K., Fisher, E. R., Wickerham, D. L., and Cronin, W. M. (1994) *Journal of the National Cancer Institute* **86**, 527-537

256. Wickerham, D. L., Fisher, B., Wolmark, N., Bryant, J., Costantino, J., Bernstein, L., and Runowicz, C. D. (2002) *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 2758-2760

257. Bolognese, M. A. (2010) *Reviews in endocrine & metabolic disorders* **11**, 253-259

258. Lufkin, E. G., Whitaker, M. D., Nickelsen, T., Argueta, R., Caplan, R. H., Knickerbocker, R. K., and Riggs, B. L. (1998) *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **13**, 1747-1754

259. Uyar, Y., Koltan, S. O., Pogun, S., Vatansever, S., and Caglar, H. (2008) *Archives of gynecology and obstetrics* **278**, 107-114

260. Cummings, S. R., Eckert, S., Krueger, K. A., Grady, D., Powles, T. J., Cauley, J. A., Norton, L., Nickelsen, T., Bjarnason, N. H., Morrow, M., Lippman, M. E., Black, D., Glusman, J.

E., Costa, A., and Jordan, V. C. (1999) *JAMA : the journal of the American Medical Association* **281**, 2189-2197

261. Delmas, P. D., Ensrud, K. E., Adachi, J. D., Harper, K. D., Sarkar, S., Gennari, C., Reginster, J. Y., Pols, H. A., Recker, R. R., Harris, S. T., Wu, W., Genant, H. K., Black, D. M., Eastell, R., and Mulitple Outcomes of Raloxifene Evaluation, I. (2002) *The Journal of clinical endocrinology and metabolism* **87**, 3609-3617

262. Ettinger, B., Black, D. M., Mitlak, B. H., Knickerbocker, R. K., Nickelsen, T., Genant, H. K., Christiansen, C., Delmas, P. D., Zanchetta, J. R., Stakkestad, J., Gluer, C. C., Krueger, K., Cohen, F. J., Eckert, S., Ensrud, K. E., Avioli, L. V., Lips, P., and Cummings, S. R. (1999) *JAMA : the journal of the American Medical Association* **282**, 637-645

263. Devine, A., Hodgson, J. M., Dick, I. M., and Prince, R. L. (2007) *The American journal of clinical nutrition* **86**, 1243-1247

Appendix

Supplemental Data

Standard Protein	Measured AFM Height (nm)	Average Crystal Height (nm)	Dehydration Ratio
cathepsin L	0.67 ± 0.08	4.4 ± 1.1	6.6
carbonic anhydrase	1.23 ± 0.15	4.8 ± 6.3	3.9
BSA	2.59 ± 0.23	9.7 ± 4.4	3.7

Table A1. List of Standard Proteins for Constructing Calibration Curve. Crystal height is determined by averaging the three dimensional sizes of the particle from crystal structure. Dehydration factor is calculated using the average size divides by the measured AFM height. Detailed method was described by Kang R. Cho. *et al*. (A Multistage Pathway for Huamn Prion Protein Aggregation in Vitro: From Multimeric Seeds to β-Oligomers and Nonfibrillar Structures, J.Am. Chem. Soc 2011, 133: 8586-8593).

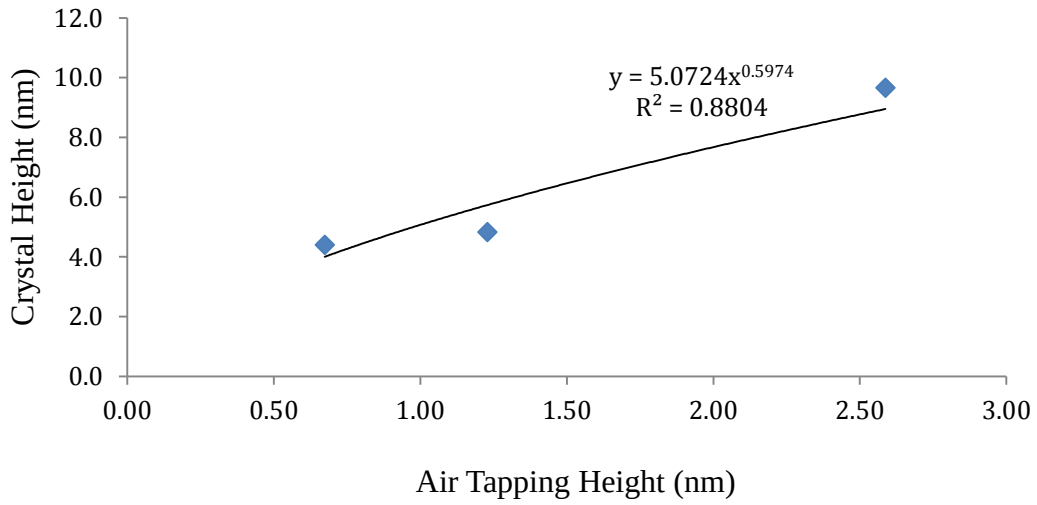


Figure A1.Tapping Mode AFM Size Correction Curve. Detailed method was described by Kang R. Cho. *et al*. (A Multistage Pathway for Huamn Prion Protein Aggregation in Vitro: From Multimeric Seeds to β-Oligomers and Nonfibrillar Structures, J.Am. Chem. Soc 2011, 133: 8586-8593).