Novel Peptide-Derived Cathepsin K Inhibitors From *Streptomyces*

by

NATALYA KRUGLYAK

B.Sc., The University of Waterloo, 2007

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

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

November 2010

© Natalya Kruglyak, 2010
ABSTRACT

Human cathepsin K is a cysteine protease that is a member of the papain superfamily. It is selectively expressed in osteoclasts where it is involved in collagen type I degradation during bone resorption. As such, cathepsin K represents a potential drug target for the treatment of metabolic bone diseases such as osteoporosis.

In the search for novel inhibitors of cathepsin K, several *Streptomyces* strains have been screened. The strain designated IS2-4 was observed to secrete inhibitors of cathepsin K into its growth media. A bioassay-guided purification of the inhibitory activity resulted in the isolation of five compounds, 6-10. Although appearing to be derivatives of the known microbial cysteine protease inhibitor leupeptin, compounds 6-10 are structurally novel. Compounds 6 and 9 inhibited cathepsin K in a concentration dependent manner with $K_{i}$ values of 44 and 64 µM, respectively.

In addition, a 2.1 Å resolution crystal structure of cathepsin K in complex with 6 was determined. The structure revealed that compound 6 has been cleaved by cathepsin K into acetyl-leucyl-leucine and a pyridotriazine fragment, with the former interacting with the S1’ and S2’ subsites and the latter binding in the S2 subsite. These results suggest a unique mechanism for the inhibition of cathepsin K. Moreover, since cathepsin K normally prefers leucine residues at S2, the preferential binding of the pyridotriazine fragment of 6 over the acetyl-leucyl-leucine fragment at S2 is unusual as well.
# TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... ii
TABLE OF CONTENTS ........................................................................................................ iii
LIST OF TABLES .................................................................................................................. v
LIST OF FIGURES .............................................................................................................. vi
LIST OF ABBREVIATIONS ................................................................................................. viii
ACKNOWLEDGMENTS ........................................................................................................ x

1.0 INTRODUCTION ............................................................................................................. 1
1.1 The Human Cathepsin Family ......................................................................................... 1
   1.1.1 The Structure of Human Cathepsins ....................................................................... 2
   1.1.2 The Catalytic Activity of Human Cathepsins ........................................................ 4
1.2 Human Cathepsin K ...................................................................................................... 7
   1.2.1 Human Cathepsin K in Bone Remodeling and Osteoporosis ................................. 7
   1.2.2 Inhibition of Human Cathepsin K ......................................................................... 10
   1.2.3 Natural Products as a Source of Novel Cathepsin K Inhibitors ............................... 12
   1.2.4 *Streptomyces* sp. as a Source of Novel Cathepsin K Inhibitors ......................... 13
1.3. Thesis Objective .......................................................................................................... 14

2.0 MATERIALS AND METHODS .................................................................................... 16
2.1 The Isolations of Cathepsin K Inhibitors ...................................................................... 16
   2.1.1 Culturing and Harvesting *Streptomyces* sp. ......................................................... 16
   2.1.2 Isolation and Identification of the IS2-4 Inhibitors ............................................... 17
   2.1.3 Enzymatic Assay and Inhibitor Quantification ..................................................... 19
   2.1.4 Determination of Inhibitory Constants .................................................................. 21
2.2 X-ray Crystallography .................................................................................................. 21
   2.2.1 Protein Crystallization .......................................................................................... 22
   2.2.2 Data Collection and Processing ........................................................................... 22
   2.2.3 Obtaining Phases ................................................................................................. 23
   2.2.4 Model Refinement ............................................................................................... 24
2.3 Determination of Cathepsin K – Inhibitor Complexes by X-ray Crystallography .......... 25
2.3.1 Crystallization and Data Collection ................................................................. 25
2.3.2 Structure Determination ...................................................................................... 26
3.0 RESULTS .................................................................................................................. 28
3.1 Inhibitor Production During Fermentation ............................................................ 28
3.2 Inhibitor Purification ............................................................................................... 28
3.3 Inhibition Kinetics ................................................................................................... 36
3.4 Cathepsin K-Inhibitor Complex Structure Elucidation by
  Crystallography ........................................................................................................ 45
4.0 DISCUSSION .......................................................................................................... 53
  4.1 Production of Compounds 6-10 in Streptomyces IS2-4 ......................................... 53
  4.2 Usefulness of Compounds 6-10 as Cathepsin K Inhibitors ..................................... 56
  4.3 Analysis of the X-ray Structure of CatK-C6 Complex ........................................... 57
    4.3.1 Cleavage of Compound 6 by Cathepsin K ....................................................... 57
    4.3.2 Binding of Compound 6 in the Active Site of Cathepsin K ............................. 58
    4.3.3 Comparison of CatK-C6 to WT-CatK ............................................................. 61
    4.3.4 Comparison of CatK-C6 and CatK-CH Complexes ......................................... 67
    4.3.5 Inhibition of Cathepsin K by Compound 6 ...................................................... 74
    4.3.6 Formation of C25-Sulfenic Acid ...................................................................... 75
  4.4 Overview of Unsuccessful Approaches Used in Inhibitor Purification .......... 75
5.0 CONCLUDING REMARKS ..................................................................................... 80
REFERENCES ............................................................................................................. 82
LIST OF TABLES

Table 3.1  Compounds 6-8 NMR Data in DMSO-d6 ........................................37
Table 3.2  Compounds 9 and 10 NMR Data in DMSO-d6 ...............................41
Table 3.3  Data Collection and Refinement Statistics for
            Cathepsin K – Inhibitor Complex.........................................................46
Table 4.1  Data Collection and Refinement Statistics for WT-CatK ...............62
LIST OF FIGURES

Figure 1.1 Ribbon Diagram of the Polypeptide Chain Fold of Cathepsin K …… 3
Figure 1.2 The Catalytic Mechanism of Cathepsin K…………………………… 5
Figure 1.3 Subsite Nomenclature Used with Proteases…………………... 6
Figure 1.4 The Bone Remodeling Cycle……………………………. 8
Figure 1.5 Known Inhibitors of Cathepsin K ………………………………… 11
Figure 2.1 Inhibitors of Cathepsin K Isolated from the Streptomyces IS2-4
Strain……………………………………………………………………… 18
Figure 2.2 Isolation of Microbial Natural Products 6-10………………… 20
Figure 3.1 A Representative Fermentation Profile for Streptomyces Strain
IS2-4……………………………………………………………………… 29
Figure 3.2 A Representative Production Profile for Streptomyces Strain
IS2-4……………………………………………………………………… 30
Figure 3.3 First HPLC Run to Purify the IS2-4 Inhibitors………………… 32
Figure 3.4 The Separation of Fraction I by HPLC………………………… 33
Figure 3.5 The Separation of Compounds 7 and 8 by HPLC……………. 34
Figure 3.6 The Separation of Compounds 9 and 10 by HPLC………….. 35
Figure 3.7 ¹H (A) and ¹³C (B) NMR Spectrum of 6 in DMSO-d6 at
600 MHz…………………………………………………………………… 38
Figure 3.8 ¹H (A) and ¹³C (B) NMR Spectrum of 7 in DMSO-d6 at
600 MHz…………………………………………………………………… 39
Figure 3.9 ¹H Spectrum of 8 in DMSO-d6 at 600 MHz ....................... 40
Figure 3.10 ¹H (A) and ¹³C (B) NMR Spectrum of 9 in DMSO-d6 at
600 MHz…………………………………………………………………… 42
Figure 3.11 ¹H (A) and ¹³C (B) NMR Spectrum of 10 in DMSO-d6 at
600 MHz…………………………………………………………………… 43
Figure 3.12 Determination of Kᵢ Values of Compounds 6 and 9, and Z-IL…… 44
Figure 3.13 Ramachandran Plot of Cathepsin K-Inhibitor Complex………… 47
Figure 3.14 The Binding of Compound 6 in the Active site of Cathepsin K…… 49
Figure 3.15 Surface Diagrams of the Binding of the Cleaved Compound 6
in the Active Site of Cathepsin K……………………………………… 50
Figure 3.16 Hydrogen Bond Interactions of the Fragmented Compound 6 in the Active Site of Cathepsin K ..................................................51
Figure 3.17 Electron Density Maps for Fragmented Compound 6 ...................52
Figure 4.1 Possible Biosynthetic Pathways of Compounds 6-10 .......................55
Figure 4.2 Subsite Occupancy and Cleavage Site of Z-FR-MCA and Compounds 6 .........................................................59
Figure 4.3 Ramachandran Plot of WT-CatK .................................................63
Figure 4.4 Average R.M.S. Deviation (A) and Average Thermal Factors Comparison (B) Between WT-CatK and CatK-C6 ..........................64
Figure 4.5 Comparison of the S2’ Subsite Between WT-CatK and CatK-C6 ....66
Figure 4.6 Hydrogen Bond Interaction of CH in the Active Site of Cathepsin K ...........................................................................68
Figure 4.7 Comparison of the Binding Mode of CH and Compound 6 ..........69
Figure 4.8 Comparison of the S3 Subsite Between CatK-CH and CatK-C6 ......71
Figure 4.9 Comparison of the S2’ Subsite Between CatK-CH and CatK-C6 ......72
Figure 4.10 Average R.M.S. Deviation Between CatK-CH and CatK-C6 .......73
Figure 4.11 Isolation of Microbial Natural Products 6-10 ...............................79
LIST OF ABBREVIATIONS

Å Angstrom unit (1 Å = 1*10^{-10} m)

^{13}C Carbon-13

^{1}H Proton

a, b, c Unit cell axis lengths

Ac-LI Acetyl-L-leucyl-L-isoleucine

Ac-LL Acetyl-L-leucyl-L-leucine

Ac-LV Acetyl-L-leucyl-L-valine

ACN Acetonitrile

b Broad (in relation to NMR signal)

BuOH Butanol

cDNA Complementary deoxyribonucleic acid

catK-C6 Human cathepsin K in complex with cleaved compound 6

catK-CH Human cathepsin K in complex with carboxyhydrazide inhibitor

COSY-60 Correlation spectroscopy with a nuclear spin tilt of 60º

d Doublet (in relation to NMR signal)

DMSO-d6 Deuterated dimethyl sulfoxide

DNA Deoxyribonucleic acid

EtOAc Ethyl acetate

F_c and F_o Calculated and observed structure factors, respectively

FA Formic acid

h, k, l Miller indices of a reflection

HMBC Heteronuclear multiple bond multiple quantum coherence

HMQC Heteronuclear multiple quantum coherence

HPLC High performance liquid chromatography

HSQC Heteronuclear single quantum coherence
[I] Inhibitor concentration

$K_i$ Dissociation constant for enzyme-inhibitor complex

kDA Kilodalton

m Multiplet (in relation to NMR signal)

MAPI Microbial alkaline protease inhibitor

MeOH Methanol

MHz Megahertz

NMR Nuclear magnetic resonance

PEG Polyethylene glycol

PT Pyridotriazine

q Quartet (in relation to NMR signal)

r.m.s. Root mean squared

RP Reversed-phase

s Singlet (in relation to NMR signal)

sp. Species

t Triplet (in relation to NMR signal)

TFA Trifluoroacetic acid

tROESY Transverse rotating-frame Overhauser enhancement spectroscopy

v Enzymatic reaction rate

WCX Mixed-mode weak cation exchange

wt-catK Unliganded, wild-type human cathepsin K

x, y, z Positional parameters in the unit cell

Z-IL Carbobenzoxy-isoleucyl-leucine

Z-FR-MCA Carbobenzoxy-phenylalanine-arginine-4-methylcoumarin-7-amide

Amino acids are defined according to the IUPAC-IUB Commission on Biochemical Nomenclature (CBN), a one-letter notation of amino acid sequences [J Biol Chem 243: 3557-3559 (1968)]. The IUPAC convention is also used to describe the conformational torsion angles of the polypeptide chain [J Biol Chem 245: 6489-6497 (1970)].
ACKNOWLEDGMENTS

Numerous people have contributed to this work by providing guidance, advice and expertise. Much appreciation is expressed to Vincent Lavallee, Jadwiga Kaleta, Dr. Adeleke Aguda, Raymond Pan, Nham Nguyen, Dr. Sung-Hye Grieco, Dr. Robert Maurus, Dr. David Williams, Dr. Gary Brayer, Dr. Raymond Andersen, and Dr. Dieter Bromme.
CHAPTER 1 – Introduction

1.1 The Human Cathepsin Family

The term “cathepsins” refers to lysosomal proteolytic enzymes that can be eukaryotic or prokaryotic in origin and may belong to aspartic, cysteine or serine class of proteases. These enzymes are present in virtually all mammalian cells and tissues, although at different concentrations, where they extensively hydrolyze proteins and peptides and participate in various physiological processes (Agarwal, 1990). Cathepsins are synthesized by membrane-bound ribosomes as proenzymes that are cotranslationally translocated into the endoplasmic reticulum, and subsequently to the Golgi apparatus. In the Golgi apparatus, the proenzymes are glycosylated with mannose-6-phosphate, which is recognized by the mannose-6-phosphate receptors that target the proenzymes into the lysosome via the late endosome, where they are activated upon exposure to low pH (Agarwal, 1990; Alberts et al., 2002).

Papain-like cysteine proteases form the largest group in the cysteine protease class and are present in eukaryotes as well as prokaryotes, where they function in protein degradation (Lecaille et al., 2002). All of the enzymes in this family exhibit a high degree of sequence homology and a similar polypeptide fold consisting of two domains of roughly equal size (Bromme and Kaleta, 2002; McGrath, 1999).

Humans express 11 papain-like cathepsins, all of which have been previously thought to be involved in random lysosomal degradation of proteins. It is now known that only cathepsin B, L, and H fulfill housekeeping functions due to their ubiquitous tissue distribution. The remaining cathepsins are tissue-specific and as such perform more sophisticated roles (Lecaille et al., 2002). For example, cathepsin S is selectively
expressed in antigen-presenting cells and thus regulates antigen presentation and
immunity (Riese et al., 1998), while cathepsin K plays a crucial role in bone turnover due
to its osteoclast-specific distribution (Bromme et al., 1996).

1.1.1 The Structure of Human Cathepsins

All human cathepsins are biosynthesized as inactive proenzymes, consisting of a
signal peptide, a propeptide, and a highly conserved catalytic domain. The signal peptide
is a 10-20 amino acid sequence that directs the proenzyme into the endoplasmic
reticulum during translation (Lecaille et al., 2002). The propeptide ranges between 36
and 315 amino acids in length, and serves a number of functions, including inhibition of
the mature enzyme in a pH-dependent manner (Jerala et al., 1998; Lecaille et al., 2002).
Specific physiological conditions (like low pH) trigger the cleavage of the propeptide,
either by other proteases or by autocatalysis, to release the active enzyme as a single
polypeptide chain, 220-260 amino acids in length (Erickson, 1989; Lecaille et al., 2002).

All mature cathepsins, with the exception of cathepsin C, are monomers with
molecular weights ranging from 20-30 kDa (Dolenc et al., 1995; McGrath, 1999). The
enzymes are organized into two distinct domains, as shown in Figure 1.1. The left
domain is made of three \( \alpha \)-helices, with the catalytic cysteine, C25, positioned at the N-
terminus of the longest of these helices. The main feature of the right domain is a five to
six stranded \( \beta \)-barrel. The right domain also houses the other two catalytic residues: H162
and N182. The domains are stabilized by disulfide links, two in the left domain and one
in the right domain (McGrath, 1999).
Figure 1.1 Ribbon Diagram of the Polypeptide Chain Fold of Cathepsin K

Shown are the left and the right structural domains of cathepsin K. Highlighted is the catalytic triad located in the cleft between the two domains: C25 (colored red), H162 (colored blue), and N182 (colored yellow).
1.1.2 The Catalytic Activity of Human Cathepsins

The catalytic triad, C25, H162 and N182, are brought together in a cleft separating the two domains of cathepsins (Figure 1.1). C25 and H162 form a thiolate-imidazolium ion pair, which is stabilized by N182 via a hydrogen bond. Proteolysis begins with a nucleophilic attack by C25 on the scissile peptide’s carbonyl group to form a tetrahedral intermediate, which is stabilized by the oxyanion hole. The tetrahedral intermediate then decomposes to the acyl-enzyme intermediate, with the concomitant release of the C-terminal portion of the substrate, and its replacement by a water molecule. Consequently, a second tetrahedral intermediate is formed, which eventually splits to release the N-terminal portion of the substrate and the free enzyme (Lecaille et al., 2002; Storer and Menard, 1994). The catalytic mechanism of cathepsins is outlined in Figure 1.2.

The active site of cathepsins contains several substrate-binding subsites, located in the interdomain cleft together with the catalytic triad (Figure 1.1). Since most human cathepsins are endopeptidases, the subsites flank the catalytic residues so that the incoming peptide can bind the active site on both sides of the scissile bond. In the case of exopeptidases, part of the substrate-binding channel is blocked, and cathepsin residues are positioned in such a way as to stabilize the C- or N-terminal portion of the substrate peptide (McGrath, 1999). According to the Schechter and Berger nomenclature (Schechter and Berger, 1967), subsites in the N-terminal direction are named S1, S2, S3, etc., and subsites in the C-terminal direction are named S1’, S2’, S3’, etc. Similarly, the residues of the bound substrate or inhibitor are designated as P1, P2, P3, etc. and P1’, P2’, P3’, etc (Figure 1.3).
Figure 1.2  The Catalytic Mechanism of Cathepsin K

The catalytic residues C25 and H162 are the nucleophile and the acid/base catalyst, respectively. N182 (not shown) stabilizes the thiolate-imidazolium ion pair via a hydrogen bond to H162. R1 and R2 represent amino acids towards the N- and C-terminals of the substrate polypeptide, respectively.
Figure 1.3  Subsite Nomenclature Used with Proteases

Residues on the N-terminal side of the scissile bond are designated P1, P2, and P3. Residues on the C-terminal side of the scissile bond are designated P1’, P2’, and P3’. The corresponding regions of the protease’s active site are designated S1, S2, S1’, etc. The scissile bond is indicated with an arrow.
1.2 Human Cathepsin K

Cathepsin K was first isolated by Tezuka and colleagues (Tezuka et al., 1994) from a rabbit osteoclast cDNA library. The human version of the enzyme was identified in human osteoclasts and characterized a few years later (Bromme et al., 1996; Drake et al., 1996). Cathepsin K shares the highest amino acid and DNA sequence homology with cathepsins S and L, and is localized in chromosome 1q21, in close proximity to cathepsin S (Rood et al., 1997).

1.2.1 Human Cathepsin K in Bone Remodeling and Osteoporosis

Drake and colleagues (Drake et al., 1996) showed that cathepsin K is abundantly and selectively expressed in osteoclasts, while cathepsins S, L, and B are not. Since osteoclasts are bone-resorbing cells, they speculated that cathepsin K plays an important role in bone resorption.

Bone resorption is part of a continuous process known as bone remodeling, which involves loss and formation of bone in the same spot on the skeleton to renew the bone and maintain its mechanical competence (Raisz, 1999). Bone is composed of approximately 65% mineral, 5% water, and 30% organic matrix. Hydroxyapatite crystals, Ca₁₀(PO₄)₆(OH)₂, constitute the mineral phase of bone, while the organic matrix of bone is composed of approximately 90% type I collagen (Kaplan et al., 1994). Bone remodeling begins when the bone resorbing cells, called osteoclasts, attach to bone and secrete hydrogen ions to demineralize the bone, followed by the secretion of cathepsin K, which requires acidic pH to degrade collagen. This leads to the formation of excavation cavities that are filled in with new bone matrix by the bone forming osteoblasts (Raisz, 1999). The bone remodeling process is schematically illustrated in Figure 1.4.
Figure 1.4 The Bone Remodeling Cycle

Shown are the three main events comprising the bone remodeling cycle:

1. Resorption – osteoclasts secrete acid and cathepsin K to demineralize and degrade the collagen content of bone, respectively, leading to the formation of a resorption pit
2. Bone formation – osteoblasts lay down new organic matrix into the resorption pit
3. Mineralization – The new bone tissue is mineralizes to complete the bone remodeling cycle. Osteoblasts that become trapped in the newly mineralized bone are called osteocytes

(Diagram adapted from http://www.ns.umich.edu/Releases/2005/Feb05/bone.html)
Cathepsin K is the only known mammalian protease able to cleave inside the native triple helical conformation of type I collagen at multiple sites and bring about its complete hydrolysis (Garnero et al., 1998). It is known that the collagenolytic activity of cathepsin K depends on the formation of a complex with chondroitin sulfate, a glycosaminoglycan found in bone (Li et al., 2002), however, the specific mechanism by which cathepsin K degrades type I collagen remains a mystery.

Clearly, cathepsin K plays an important role in bone resorption. This is further corroborated by the finding that cathepsin K deficiency leads to the accumulation of undigested collagen fibrils in the cytoplasmic vacuoles of osteoclasts taken from patients suffering from pycnodysostosis, an autosomal recessive condition characterized by a variety of skeletal abnormalities, including dwarfism and fragile bones (Everts et al., 1986; Gelb et al., 1996; Meredith et al., 1978). While pycnodysostosis results from insufficient activity of cathepsin K, excessive cathepsin K activity increases bone resorption and leads to bone loss and consequently, osteoporosis (Kiviranta et al., 2001; Meier et al., 2006). Osteoporosis is a disease that mainly affects postmenopausal women, and is characterized by low bone mass and an increased risk of fracture (Lecaille et al., 2002; Stoch and Wagner, 2008). Current treatment for osteoporosis, such as bisphosphonates, raloxifene (an estrogen receptor modulator) and estrogen replacement therapy, primarily work to reduce the rate of bone resorption, however, they do not completely eliminate skeletal fragility. In addition, all available therapies have unpleasant side effects, including gastrointestinal problems, bloating, leg cramps and high blood pressure (Cohen, 2004; Umland et al., 1999). Therefore, better therapeutic agents are needed for the treatment of osteoporosis, and recent efforts have been directed at
developing inhibitors of cathepsin K for the treatment of this disease.

1.2.2 Inhibition of Human Cathepsin K

Evidence that the inhibition of cathepsin K is beneficial for the treatment of osteoporosis has been demonstrated using in vitro resorption assays as well as in rodent and monkey animal models of osteoporosis (Binkley et al., 1998; James et al., 1999; Lark et al., 2002). Although there are no cathepsin K inhibitors currently on the market, several major pharmaceutical companies are developing such inhibitors for therapeutic use. One example is odanacatib (1 in Figure 1.5), a selective, reversible inhibitor of cathepsin K that has shown to decrease bone resorption and increase bone mass in postmenopausal women suffering from osteoporosis, who took a weekly odanacatib pill for three years (Eisman et al., 2010). However, since osteoporosis is a chronic condition, odanacatib’s efficacy and patient tolerance of taking the drug over much longer periods of time still need to be investigated.

The difficulty in identifying highly selective cathepsin K inhibitors lies in the fact that all human cathepsins share a similar polypeptide fold, active site residues and catalytic mechanism. The task is further complicated by the fact that cysteine protease inhibitors are also reactive towards other proteases, especially the serine proteases (Lecaille et al., 2002). Clearly, such cross-reactivity is undesirable since it can lead to serious pathological conditions. Therefore, for cathepsin K inhibitors to be clinically useful, they must exhibit high efficacy as well as selectivity. The growing availability of x-ray structures of cysteine proteases in complex with small molecule inhibitors has allowed for the identification of substrate-binding subsites and their specific binding determinants. It has become clear that subtle differences exist between the subsites of
Figure 1.5 Known Inhibitors of Cathepsin K

The structures of odanacatib 1, leupeptin 2, E-64 3, antipain 4, and α-MAPI 5.
human cathepsins, and it is these differences that must be taken into account when designing cathepsin inhibitors for therapeutic use (McGrath, 1999; Turk et al., 1998).

The most obvious difference in the substrate-binding site exists between endopeptidases and exopeptidases, as described in Section 1.1.2. Cathepsin K is an endopeptidase, and as such it binds its substrate on both sides of the scissile bond (McGrath, 1999). Of the seven apparent substrate-binding subsites (S3’ to S4) (Schechter and Berger, 1967), S2 is the only well-defined binding pocket and the predominant factor in substrate specificity between cathepsins. The S2 site is formed by hydrophobic amino acids, and variation in the shape and length of these residues determines the size of the S2 pocket and the binding preference of the enzyme (McGrath, 1999; Turk et al., 1998). S1’, S1 and S3 subsites impart additional substrate specificity to cathepsin K, although to a lesser degree than the S2 subsite (Choe et al., 2006; Turk et al., 1998).

1.2.3 Natural Products as a Source of Novel Cathepsin K Inhibitors

Natural products have been used in medicine for centuries, and almost half of the drugs that are currently available on the market have been derived from natural sources. These include various antibiotics (e.g. penicillin), antimalarials (e.g. quinine), immunosuppressants (e.g. rapamycin), and anticancer agents (e.g. taxol) (Li and Vederas, 2009).

Medically relevant natural products are usually secondary metabolites of low molecular weight (< 2 kDa) produced by microorganisms or plants. Compared to synthetic drugs, natural products offer a higher degree of structural diversity, potency and biochemical specificity. This is because natural products have been fine-tuned to interact with specific biological targets for specific purposes through many years of evolution
(Paterson and Anderson, 2005; Sarker et al., 2006). The biological targets of natural products are most often proteins, and the structural motifs of protein domains are often well conserved even in proteins with low sequence homology. Assuming that natural products have evolved to interact with these folds, a single natural product fulfilling a specific purpose in an organism may elicit a different but just as potent response in another setting (Anantharaman et al., 2003; Paterson and Anderson, 2005). Clearly, natural products are invaluable as sources of lead drug candidates. Some of these can be used as effective drugs in their unmodified state, while others can be derivatized to improve their potency and specificity (Sarker et al., 2006).

Production of protease inhibitors, including cysteine proteases, by animals, plants and microorganism is well documented. Protease inhibitors from animals and plants are macromolecular peptides, called cystatins, that regulate the activity of proteases intracellularly. Microorganisms, on the other hand, produce high and low molecular weight protease inhibitors. The former function intracellularly, while the latter are released extracellularly (Umezawa, 1982). Several extracellular cathepsin K inhibitors of microbial origin have been previously isolated. One such inhibitor is the peptide aldehyde leupeptin (2 in Figure 1.5), which was first isolated from Streptomyces exfoliates, but later was found to be produced by at least 11 other Streptomyces strains (Aoyagi et al., 1969; Umezawa, 1982). Another example is the epoxide E-64 (3 in Figure 1.5), which was isolated from Aspergillus japonicus (Hanada et al., 1978).

1.2.4 Streptomyces sp. as a Source of Novel Cathepsin K Inhibitors

Streptomyces belong to the actinomycete genera, and they are Gram-positive, filamentous bacteria. Over 450 species of Streptomyces have been described, however,
some are still pending a detailed phylogenetic analysis based on DNA sequencing (Kim et al., 2004). These bacteria are found predominantly in soil, and while they are not pathogenic to humans, they can cause disease in some plants. Most importantly, Streptomyces are characterized by a complex secondary metabolism, and as such, they produce a vast number of biologically active secondary metabolites (for example, approximately 60% of all antibiotics are produced by Streptomyces) (Kieser et al., 2000). For this reason, Streptomyces are an attractive target for the discovery of novel cathepsin K inhibitors. As stated above, leupeptin is one example of a cathepsin K inhibitor that has been isolated from Streptomyces (Aoyagi et al., 1969; Umezawa, 1982). Antipain (4 in Figure 1.5) and various derivatives of the microbial alkaline proteinase inhibitors (MAPI) (α-MAPI, 5 in Figure 1.5) were also isolated from Streptomyces, and show broad protease specificity (Suda et al., 1972; Watanabe and Murao, 1979), including the inhibition of cathepsin K. Clearly, the Streptomyces species are a treasure chest of biologically active compounds, and a possible source of novel inhibitors of cathepsin K.

1.3 Thesis Objective

The research described in this thesis has two main objectives. The first objective was to isolate inhibitors for cathepsin K secreted by Streptomyces sp. Throughout the purification procedure, semi-crude aliquots showing inhibitory activity against cathepsin K were soaked into crystals of wild-type human cathepsin K and analyzed by x-ray crystallography. The assumption is that the inhibitor will have high affinity for the active site of cathepsin K, and thus will diffuse through the crystal and bind the enzyme’s active site. Therefore, the second objective of this work was to elucidate a putative structure of the target cathepsin K inhibitor using x-ray crystallography, and examine the interactions
the inhibitor makes with the enzyme’s active site.
CHAPTER 2 – Materials and Methods

The following chapter describes the techniques used to purify cathepsin K inhibitors from *Streptomyces* sp., followed by a brief overview of x-ray crystallography and a discussion of the methods by which the three dimensional x-ray structures of cathepsin K – inhibitor complexes were obtained.

2.1 The Isolation of Cathepsin K Inhibitors

2.1.1 Culturing and Harvesting *Streptomyces* sp. IS2-4

The inhibitory activity was obtained from a *Streptomyces* strain designated IS2-4, which was provided by collaborators in the laboratory of Dr. Julian Davies (Department of Microbiology and Immunology, University of British Columbia). Frozen spore stocks of the IS2-4 strain were defrosted, spread on ISP4 agar (BD, MD) plates, and incubated at 30°C for 4-6 days. Single colonies of IS2-4 from the plates were used to inoculate 1 L shake-flasks containing 300 ml seed media (1% D-glucose, 1.5% glycerol, 0.5% malt extract, 0.5% yeast extract, 1.5% soytone, 0.3% NaCl, 0.1% Tween 20, and 2% MOPS) at pH 7. The flasks were shaken at 200 rpm at 30°C until the inhibitory activity of a 100 times diluted culture broth was at least 90% (approximately 2 days). The entire 300 ml of the seed culture was then used to inoculate a 7 L fermenter (The Center for Blood Research Fermentation Suite, University of British Columbia) containing 5 L of media of the same composition as the seed media. The cultures were kept at 30°C with aeration and agitation, and the pH was maintained at 7. After approximately 24 hours, when the inhibitory activity of a 100 times diluted fermentation broth reached at least 90%, the fermentation broth containing the inhibitory activity was separated from the bacterial
mycelia by centrifugation at 10,000 rpm.

### 2.1.2 Isolation and Identification of the IS2-4 inhibitor

Following centrifugation, the fermentation broth was stirred with 30 g/L Amberlite XAD4 (Sigma-Aldrich, MO) at 4°C, until all the inhibitory activity was adsorbed to the resin. The resin was then washed with distilled water, followed by 20% methanol (MeOH), and the inhibitor eluted using 100% MeOH. The MeOH eluate was evaporated and replaced with 100 ml of distilled water, which was extracted three times with ethyl acetate (EtOAc) (1:1). The aqueous fraction was then decolorized by stirring with 20 g/L Dowex Marathon A anion-exchange resin (Sigma-Aldrich, MO) overnight at 4°C. Then, it was extracted three times with n-butanol (BuOH) (1.2:1). The BuOH fraction was lyophilized, dissolved in dimethyl sulfoxide, and applied to a 50 g reversed-phase (RP) C18 column (Phenomenex, CA), eluting with increasing concentration of acetonitrile (ACN) + 0.05% trifluoroacetic acid (TFA) (300 ml each fraction). The active fraction (20% ACN + 0.05% TFA) was concentrated and further purified using a 10 g weak cation exchange column that also incorporates reversed-phase selectivity (WCX) (Phenomenex, CA), preequilibrated with 5% ammonium hydroxide. Following sample application, the column was washed with 50% MeOH, and the inhibitor eluted with 50% MeOH + 2% formic acid (FA) (using 240 ml of each). The eluate was then purified by HPLC (System Gold, Beckman, CA) using a 250 x 10 mm RP C18 column (Phenomenex, CA) and isocratic 20% ACN + 0.1% TFA with a flow rate of 3 ml/min to afford two active fractions, fraction I and fraction II. Further purification of fraction I using isocratic 15% ACN + 0.05% TFA afforded 6 (Figure 2.1) as a pure compound, and a mixture of 7 and 8 (Figure 2.1). Compounds 7 and 8 were separated using a different
Figure 2.1  Inhibitors of Cathepsin K Isolated from the *Streptomyces* IS2-4 Strain

The structures were elucidated based on one- and two-dimensional NMR. Squiggly lines indicate ambiguous stereochemical assignment. Note that the only difference between the structure of 9 and 10 is the stereochemistry of the ring moiety.
HPLC system (600E system and 486 absorbance detector, Waters, MA) equipped with a 250 x 9.4 mm RP C18 column (Intersil, CA) and isocratic 13% ACN + 0.05% TFA with a flow rate of 2 ml/min. Fraction II was also purified by HPLC (600E system and 486 absorbance detector, Waters, MA) using a 250 x 9.4 mm RP C18 column (Intersil, CA) and isocratic 12% ACN + 0.05% TFA with a flow rate of 2 ml/min to afford 9 and 10 as pure compounds (Figure 2.1). The structures of all the isolated compounds were determined by Dr. David Williams (Department of Chemistry, University of British Columbia) on the basis of one-dimensional and two-dimensional NMR as well as mass spectrometric analyses. A flow diagram summarizing the purification steps is shown in Figure 2.2.

2.1.3 Enzymatic Assay and Inhibitor Quantification

The purification of cathepsin K inhibitors from the Streptomyces IS2-4 strain was bioassay guided. In other words, at each step of the purification, the resultant fractions were assayed for cathepsin K inhibition using the synthetic peptide substrate, carbobenzoxy-phenylalanine-arginine-4-methylcoumarin-7-amide (Z-FR-MCA) (Alexis, Switzerland), which upon cleavage by cathepsin K releases MCA, a fluorophore. Wild-type human cathepsin K used in the assay was expressed in Pichia pastoris and purified as previously described (Linnevers et al., 1997) by Raymond Pan and Linda Zhang.

The assay was performed in duplicates using a 96-well plate and a fluorescent microplate reader (SpectraMax Gemini XS, Molecular Devices, CA). The assay mixture consisted of 100 mM sodium acetate buffer, pH 5.5, containing 2.5 mM dithioerythritol and 2.5 mM ethylenediaminetetraacetic acid, as well as 5 μM Z-FR-MCA, 0.04 μM cathepsin K, and varying concentrations of the IS2-4 inhibitors to a total volume of
Fermentation broth of *Streptomyces* sp.

> Supernatant separated by centrifugation

**Adsorption resin**

*Amberlite XAD-4*

> Wash with water and 20% MeOH
> Elute with 100% MeOH

**EtOAc extraction**

> Aqueous phase

**Anion exchange resin**

*Dowex Marathon A*

> Filtrate

**BuOH extraction**

> Organic fraction

**RP C18 Column**

> Active fraction (20% ACN + 0.05% TFA)

**WCX Column**

> Active fraction (50% MeOH + 2% FA)

**RP C18 HPLC**

<table>
<thead>
<tr>
<th>15% ACN + 0.05% TFA isocratic</th>
<th>13% ACN + 0.05% TFA isocratic</th>
<th>12% ACN + 0.05% TFA isocratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound 6</td>
<td>Compound 7 and 8</td>
<td>Compound 9 and 10</td>
</tr>
</tbody>
</table>

**Figure 2.2** Isolation of Microbial Natural Products 6-10
200 µl. The rate of Z-FR-MCA hydrolysis was monitored at 25°C and fluorescence was measured continuously at excitation and emission wavelengths of 355 and 460 nm, respectively, for 5 minutes.

To estimate the amount of IS2-4 inhibitor (in moles) present at each purification step, it was assumed that cathepsin K and the inhibitor react in a 1:1 ratio when the assay shows 90% inhibition of the enzyme. The following equation was used:

\[(0.04 \text{ µM cathepsin K}) \times (\text{total volume of purified sample}) \times (\text{inhibitor dilution factor})\]

### 2.1.4 Determination of Inhibitory Constants

The inhibitory constants, \(K_i\), of the isolated compounds were determined using the assay described in section 2.1.3. The enzyme’s reaction rate (v) was measured at a series of Z-FR-MCA concentrations (4, 5, and 6 µM) and over a range of inhibitor concentrations [I] (typically 4-500 µM). The \(K_i\) value was then calculated using a Dixon plot, in which 1/v for each substrate concentration is plotted against the inhibitor concentrations. The resulting straight lines intersect at a point corresponding to \([I] = -K_i\).

The \(K_i\) of the commercial peptide carbobenzoxy-isoleucyl-leucine (Z-IL) (Sigma-Aldrich, MO) was determined in a similar manner using 4 and 5 µM of Z-FR-MCA. The concentration of Z-IL used in the assay ranged between 0.5-30 mM.

### 2.2 X-ray Crystallography

This section provides a brief overview of protein x-ray crystallography. For additional details, the interested reader may refer to (Rhodes, 2006; Sands, 1969).
2.2.1 Protein Crystallization

The first step in protein x-ray crystallography is to obtain diffraction-quality protein crystals. These crystals consist of a three-dimensional arrangement of protein molecules held together by noncovalent interactions. Protein crystals may also be considered as composed of individual building blocks, called the unit cells, that are repeated in a three-dimensional lattice. The unit cell itself is generated from an asymmetric unit through crystallographic symmetry. It may consist of one molecule for a monomeric protein or multiple molecules for a multimeric protein, and it is the goal in protein crystallography to determine the positions of all the atoms in the asymmetric unit.

Crystals are commonly grown using the hanging-drop method. In this set-up, a few µl of a protein solution is mixed with an equal amount of a buffer solution containing a precipitant. A drop of this mixture is suspended underneath a glass cover slip that is sealed onto a well containing the same buffer-precipitant solution to form a closed system. Through vapour diffusion from the hanging drop to the more concentrated solution in the well, the protein and precipitant concentrations in the drop slowly increase until equilibrium is reached. This brings the protein solution to supersaturation and induces crystal formation. Initial screens are set up to determine the optimal crystallization conditions, and these involve variations in the buffer pH, temperature, protein and precipitant concentrations among other factors.

2.2.2 Data Collection and Processing

For data collection, protein crystals must be mounted and placed in the x-ray beam. Consequently, they are susceptible to radiation damage and dehydration. To combat this problem, protein crystals are mounted in the presence of a cryoprotectant,
such as glycerol or polyethylene glycol (PEG). Then, they are scooped up with a fiber loop and flash frozen in liquid nitrogen. For analysis, the loop is placed on a goniometer head, and the crystal is irradiated with x-rays to yield a pattern of diffraction spots (also called reflections) that are collected using an area detector. To obtain all the necessary information to solve a protein structure, the crystal must be gradually rotated through 180° (or less, for high symmetry crystals), with a diffraction images recorded every 0.5-2°, as appropriate.

Notably, data collection results in many images of diffraction spots. First, these spots need to be indexed to determine the dimensions and symmetry of the unit cell. Then, these diffraction spots are merged and scaled. In other words, spots that are present in two or more images are identified and merged to obtain the final intensities. This is crucial because intensities contain structural information about the molecules in the unit cell.

### 2.2.3 Obtaining Phases

As described above, the only information that can be obtained from diffraction images are the positions and intensities of the recorded diffraction spots. The position of a diffraction spot in the three-dimensional diffraction spot lattice is described with the Miller indices $h$, $k$, and $l$. Each recorded diffraction spot intensity $hkl$ is the sum of x-ray scattering by all atoms in the unit cell. This sum is known as the structure factor, $F_{hkl}$, and its amplitude $|F_{hkl}|$ is directly proportional to the square root of the observed intensity of diffraction spot $hkl$. Structure factors are necessary for the calculation of electron density as follows:
\[ \rho(x,y,z) = \frac{1}{V} \sum_h \sum_{k} \sum_{l} |F_{hkl}| e^{i\alpha_{hkl}} e^{-2\pi i (hx + ky + lz)} \] (2-1)

where \( \rho(x, y, z) \) is the electron density at a point in the unit cell, \( V \) is the volume of the unit cell, and \( \alpha_{hkl} \) is the phase of structure factor \( |F_{hkl}| \). From equation (2-1) it becomes clear that in order to calculate an electron density map and from this build a model of a protein, \( |F_{hkl}| \) and \( \alpha_{hkl} \) both need to be known. This poses a problem since \( \alpha_{hkl} \), unlike \( |F_{hkl}| \), cannot be directly recorded in a diffraction experiment.

The resulting phase problem can be resolved by obtaining initial phase estimates using a variety of approaches. The method used in this study was molecular replacement, and it relies on a previously solved protein structure that is isomorphous (same space group and similar unit cell dimensions) and, in this case, homologous to the new unknown structure. Thus, by placing the previously solved homologue in the unit cell of the unknown, but related new protein, estimates for structure factors, including phases, could be calculated as follows:

\[ F(h,k,l) = c \sum_{j} f_{j} \left( s_{hkl} \right) e^{-B_{j}(s_{hkl})^{2}} e^{2\pi i (hx + ky + lz)} \] (2-2)

where the constant \( c \) is a scale factor, \( f_{j} \) is the scattering factor of atom \( j \), \( s_{hkl} \) is equal to \( \sin \theta / \lambda \) for the reflection \( hkl \), \( B_{j} \) is the thermal factor that is related to the atom \( j \)'s displacement about its average position in the model, which is specified by \( (x_{j}, y_{j}, z_{j}) \).

### 2.2.4 Model Refinement

The initial phase estimates obtained are used to build an initial model of the new protein structure. This model and its associated electron density can be improved through repeated cycles of computational least-squares fitting of atomic coordinates to the observed diffraction data and cycles of temperature factor refinement, which adjust the
motional factors of the atoms in the model. The model is also adjusted manually based on
electron density maps, where necessary. The correlation between the calculated structure
factor amplitudes $|F_c|$ and the observed structure factor amplitudes $|F_o|$ is monitored
throughout structural refinement as follows:

$$R - \text{factor} = \frac{\sum_{kl} |F_{o, kl} - F_{c, kl}|}{\sum_{kl} |F_{o, kl}|} \times 100 \quad (2-3)$$

A low R-factor indicates a better correlation between the observed data and the protein
model. In general, the final R-factor should be below 20% for a model built from good
quality diffraction data. Another parameter used to monitor model refinement is the R-
free. It is calculated from equation (2-3) using only a random subset (~5%) of the
measured diffraction spots that have not been used during refinement. The R-free is only
used in cross-validation of model quality, and it is thought to give a less biased
assessment of the refinement progress.

The R-factor and R-free are statistical parameters that are used to judge the
quality of the final model. Its structural parameters, such as bond lengths, $\alpha$-carbon
chirality and peptide bond planarity, are monitored as well to ensure that the model is
chemically, stereochemically and conformationally valid.

2.3 Determination of Cathepsin K – Inhibitor Complexes by X-ray Crystallography

2.3.1 Crystallization and Data Collection

Wild-type human cathepsin K was expressed in *P. pastoris* and purified according
to previously established protocols (Linnevers *et al.*, 1997) by Raymond Pan, Dr.
Adeleke Aguda, and the author. The purified enzyme was crystallized using the hanging
drop method by Nham Nguyen and Dr. Adeleke Aguda. For the structure described in this thesis, 2 µl of a 4.75 mg/ml cathepsin K solution (50 mM sodium acetate, 0.5 mM dithioerythreitol, 2.5 mM ethylenediaminetetraacetic acid, pH 5.0) was mixed with 1 µl of a semi-purified inhibitor solution (HPLC Fraction I described in Section 2.1.2, at a concentration of ~10 µM, calculated as described in Section 2.1.3, dissolved in MeOH) and 1 µl of reservoir solution (0.2 M ammonium sulfate, 30% PEG 8000, pH 6.5), and then equilibrated against reservoir solution until diffraction quality crystals were obtained. Diffraction data was collected by Dr. Gary Brayer at the Stanford Synchrotron Radiation Lightsource beamline 7-1 using an ADSC Quantum-315R CCD detector.

### 2.3.2 Structure Determination

Following data collection, the inhibited cathepsin K dataset was indexed and integrated using MOSFLM (Leslie, 1992) and then scaled using SCALA (Evans, 2005). The initial phases for the structure were determined by molecular replacement using MOLREP (Vagin and Teplyakov, 1997) and the isomorphously crystallized human cathepsin K in complex with a purine nitrile inhibitor (Altmann et al., 2004; PDB # 1U9X), with the inhibitor and the waters removed. Least squares and thermal factor refinement of the initial model was carried out in REFMAC5 (Murshudov et al., 1997), and manual electron density fitting was performed using COOT (Emsley and Cowtan, 2004). Once the polypeptide chain of cathepsin K was refined, a sulfate ion and the inhibitor were positioned based on F_o-F_c and 2F_o-F_c difference electron density maps. The initial models of the sulfate ion and the inhibitor were obtained using PRODRG (Schuettelkopf and van Aalten, 2004). Solvent molecules were added using ARP/wARP (Zwart et al., 2004), and included in the final model based on hydrogen bonding potential
to protein atoms and the refinement of a thermal factor to below 75 Å$^2$. The stereochemical quality of the final model was examined using PROCHECK (Laskowski et al., 1993). The programs SCALA, MOLREP, REFMAC5, ARP/wARP, and PROCHECK are all part of the CCP4 suite of programs (Collaborative Computational Project, 1994).
CHAPTER 3 – Results

3.1 Inhibitor Production During Fermentation

Following inoculation of shake-flasks with *Streptomyces* strain IS2-4, production of the inhibitor reached a maximum after ~2 days. At this point, typical inhibitory activity against cathepsin K of a 100 and 1000 times diluted fermentation broth was 90-100%, and 70-90%, respectively. Assuming that all the inhibitory activity belongs to a single inhibitor, there was 1.2-12 µmol inhibitor per flask (see section 2.1.3 for calculation details). In a large scale 7 L fermenter, inhibitor production usually reached a maximum 22-24 hours following inoculation, at which point typical inhibitory activity against cathepsin K of a 100 and 1000 times diluted fermentation broth was 90-100%, and 60-90%, respectively, resulting in the production of at least 20 µmol inhibitor. A representative large-scale fermentation and inhibitor production profiles are presented in Figure 3.1 and Figure 3.2, respectively.

3.2 Inhibitor Purification

The results described in this section are for a representative purification, starting with a 20 L fermentation broth.

Overall, 700 µmol inhibitor was harvested from a 20 L fermentation broth. Next, the fermentation broth containing the inhibitor was treated sequentially with Amberlite XAD4, EtOAc extraction, Dowex Marathon A, and BuOH extraction. During these steps, no loss of inhibitory activity was observed. Unfortunately, when the sample was applied onto a RP C18 column, only 2 µmol inhibitor was recovered from the column, resulting in over 99% loss of the inhibitory activity. No further loss occurred through the use of a
Figure 3.1  A Representative Fermentation Profile for *Streptomyces* Strain IS2-4

The strain was grown in a 7 L fermenter containing 5 L of defined media with aeration, oxygenation (in red), and agitation (in black). The pH was maintained at 7 (in blue) and the temperature at 30ºC (in green). Also indicated is the inhibitory activity of 100 times and 1000 times diluted fermentation broth, with pink and red diamonds respectively, at 0, 16, and 22 hours following inoculation. Note that inhibitor production is associated with increased consumption of oxygen by the bacteria (as indicated by low dissolved oxygen levels, in red), and thus bacterial growth. The figure was prepared by Dr. Sung-Hye Grieco, The Center for Blood Research Fermentation Suite, University of British Columbia.
Figure 3.2 A Representative Production Profile for *Streptomyces* Strain IS2-4

The production profile is for the fermentation presented in Figure 3.1. Aliquots of the fermentation broth were diluted 100 and 1000 times, and tested against cathepsin K activity at 0, 16, and 22 hours post inoculation. The fermentation was halted at 22 hours post inoculation.
WCX column for further purification, as all of the remaining 2 µmol inhibitor were eluted from the column.

The final steps in the purification involved RP C18 HPLC. The first HPLC purification still involved a very complex sample, and the activity against cathepsin K smeared for the entire HPLC profile. However, it was possible to identify a region that showed the highest activity against cathepsin K (Figure 3.3). This region was separated into two fractions, fraction I (16-18 minutes) and fraction II (19-22 minutes). Fraction I contained 40 nmol inhibitor and had a total weight of 18 mg. Fraction II contained 600 nmol inhibitor and had a total weight of 44 mg.

When fraction I and II were reapplied onto a RP C18 HPLC column for further purification, it became clear that they consisted of several very closely related compounds. The inhibitory activity in fraction I seemed to spread between two peaks, and these were isolated to yield pure compound 6 (2.3 mg) and a mixture of two compounds (Figure 3.4). Since these compounds could not be separated on the present HPLC column, a different column that afforded better resolution was utilized, resulting in the isolation of compounds 7 (0.4 mg) and 8 (0.2 mg) (Figure 3.4 and 3.5). Fraction II consisted of at least six compounds that showed some activity against cathepsin K (Figure 3.6) as well as other inactive contaminants. Only two of the active compounds, 9 (3.4 mg) and 10 (0.6 mg) were pure enough for structural analysis. The other four, although appearing as single peaks by HPLC, were determined to still be mixtures by NMR. It is unclear whether they are indeed mixtures of several compounds, or whether the NMR data is complicated due to the presence of structural isomers. At this point, their purification was not pursued further. The structures of compounds 6-10 were elucidated
Figure 3.3  First HPLC Run to Purify the IS2-4 Inhibitor

The active WCX eluate was applied onto a RP C18 semi-prep HPLC column and eluted with an ACN gradient (indicated in black) + 0.1% TFA. The resulting profile was monitored at 220 nm (in red) and 260 nm (in blue). The activity against cathepsin K was concentrated between 16-22 minutes. This active fraction was divided into fraction I (16-18 minutes) and fraction II (19-22 minutes), which were pooled for further purification.
Figure 3.4  The Separation of Fraction I by HPLC

The active HPLC fraction I (Figure 3.3) was reapplied onto the RP C18 semi-prep column and eluted with an ACN gradient (indicated in black) + 0.05% TFA. The resulting profile was monitored at 204 nm (in blue). Pure compound 6, and an inseparable mixture of compounds 7 and 8 were isolated from this run.
Figure 3.5 The Separation of Compounds 7 and 8 by HPLC

The inseparable mixture of compounds 7 and 8 (Figure 3.4) was applied onto a RP C18 semi-prep column and eluted with isocratic 13% ACN + 0.05% TFA. The resulting profile was monitored at 204 nm. HPLC peaks belonging to compounds 7 and 8 are indicated.
Figure 3.6 The Separation of Compounds 9 and 10 by HPLC

The active HPLC fraction II (Figure 3.3) was applied onto a RP C18 semi-prep column and eluted with isocratic 12% ACN + 0.05% TFA. The resulting profile was monitored at 204 nm. HPLC peaks belonging to compounds 9 and 10 are indicated. The HPLC peaks indicated with asterisks were also collected, but were deemed to be mixtures by NMR and it was not possible to determine their structures.
by Dr. David Williams (Department of Chemistry, University of British Columbia) using mass spectrometry and NMR techniques. The molecular weights of the isolated compounds as determined by mass spectrometry were 450 g/mol for 6, 436 g/mol for 7, 470 g/mol for 8, and 425 g/mol for 9 and 10. The NMR assignments for 6-10 can be found in Table 3.1 and 3.2 and the corresponding NMR spectra in Figures 3.7-3.11. Note that some of the hydrogen atoms in the cyclic moiety of 6 and 7 and the carbon and hydrogen atoms in the biguanide of 8 do not show up in the NMR spectra. These were assigned based on the mass spectrometry results, and their confirmation requires further structural investigation, which is dependant on the production of 6-8 in larger quantities. Also note that there was not enough material to obtain a $^{13}$C spectrum of 8, and the carbon chemical shifts listed in Table 3.1 come from the HSQC and HMBC experiments.

3.3 Inhibition Kinetics

The dissociation constants $K_i$ were determined for compounds 6 and 9 using Dixon plots. The hydrolysis of the substrate Z-FR-MCA decreased with increasing concentration of the inhibitors. Accordingly, compounds 6 and 9 inhibit cathepsin K with $K_i$ values of 44 and 64 μM, respectively, and as indicated by the Dixon plots, both act in a competitive manner (Figure 3.12A, B). Unfortunately, due to their limited amounts, the $K_i$ of compounds 7, 8, and 10 could not be determined. Nevertheless, considering the similarities in the structures of the isolated compounds, and the fact that none of them possess groups that would allow for covalent inhibition of the enzyme, it is likely that 7, 8, and 10 have $K_i$ values in the same range as 6 and 9. It should be remembered, however, that the differences in the stereochemical configurations of the isolated compounds might dramatically influence their biological activities.
Table 3.1  Compounds 6-8 NMR Data in DMSO-d6

<table>
<thead>
<tr>
<th>Atom Number</th>
<th>Proton 1</th>
<th>Carbon 6</th>
<th>Proton 2</th>
<th>Carbon 7</th>
<th>Proton 3</th>
<th>Carbon 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.81 s</td>
<td>22.46</td>
<td>1.84 s</td>
<td>22.48</td>
<td>1.83 s</td>
<td>23.03</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>169.07</td>
<td></td>
<td>169.5</td>
<td></td>
<td>169.44</td>
</tr>
<tr>
<td>3</td>
<td>7.97 d</td>
<td></td>
<td>8.06 d</td>
<td></td>
<td>8.00 d</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.27 q</td>
<td>50.88</td>
<td>4.28 q</td>
<td>51.25</td>
<td>4.21 q</td>
<td>51.62</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>171.97</td>
<td></td>
<td>172.28</td>
<td></td>
<td>171.66</td>
</tr>
<tr>
<td>6</td>
<td>7.83 d</td>
<td></td>
<td>7.74 d</td>
<td></td>
<td>7.81 d</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.41 q</td>
<td>50.74</td>
<td>4.08 q</td>
<td>57.81</td>
<td>4.21 q</td>
<td>51.62</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>172.67</td>
<td></td>
<td>171.21</td>
<td></td>
<td>171.99</td>
</tr>
<tr>
<td>9</td>
<td>7.93 d</td>
<td></td>
<td>7.96 d</td>
<td></td>
<td>7.45 d</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4.07 m</td>
<td>46.53</td>
<td>3.64 bm</td>
<td>49.66</td>
<td>3.65 m</td>
<td>50.23</td>
</tr>
<tr>
<td>11</td>
<td>1.71 bm</td>
<td>26.59</td>
<td>1.51 bm</td>
<td>29.50</td>
<td>1.24 bm</td>
<td>28.61</td>
</tr>
<tr>
<td>12</td>
<td>1.47 bm</td>
<td>18.52</td>
<td>1.63 bm</td>
<td>22.43</td>
<td>1.40 m</td>
<td>25.00</td>
</tr>
<tr>
<td>13</td>
<td>3.93 bm</td>
<td>44.28</td>
<td>2.98 bm</td>
<td>45.15</td>
<td>3.04 bm</td>
<td>41.46</td>
</tr>
<tr>
<td>14</td>
<td>4.79 bd</td>
<td>66.80</td>
<td>4.48 d</td>
<td>67.16</td>
<td>1.40 m</td>
<td>41.16</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>157.35</td>
<td></td>
<td>157.69</td>
<td>1.56 m</td>
<td>24.72</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>156.59</td>
<td></td>
<td>156.63</td>
<td>0.82 d</td>
<td>22.13</td>
</tr>
<tr>
<td>17</td>
<td>7.71 s</td>
<td></td>
<td>7.83 d</td>
<td></td>
<td>0.87 d</td>
<td>23.62</td>
</tr>
<tr>
<td>18</td>
<td>1.38 t</td>
<td>40.77</td>
<td>1.41 t</td>
<td>40.43</td>
<td>1.44 m</td>
<td>41.16</td>
</tr>
<tr>
<td>19</td>
<td>1.56 m</td>
<td>24.16</td>
<td>1.58 m</td>
<td>24.16</td>
<td>1.56 m</td>
<td>24.72</td>
</tr>
<tr>
<td>20</td>
<td>0.86 d</td>
<td>22.91</td>
<td>0.82 d</td>
<td>21.63</td>
<td>0.87 d</td>
<td>23.62</td>
</tr>
<tr>
<td>21</td>
<td>0.81 d</td>
<td>21.97</td>
<td>0.87 d</td>
<td>23.00</td>
<td>0.82 d</td>
<td>22.13</td>
</tr>
<tr>
<td>22</td>
<td>1.41 t</td>
<td>40.92</td>
<td>2.01 m</td>
<td>30.14</td>
<td>3.20 bm</td>
<td>63.68</td>
</tr>
<tr>
<td>23</td>
<td>1.56 m</td>
<td>24.16</td>
<td>0.81 d</td>
<td>17.90</td>
<td>3.31 s</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.83 d</td>
<td>21.67</td>
<td>0.83 d</td>
<td>19.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.86 d</td>
<td>22.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹H and ¹³C NMR at 600 MHz; Correlations were assigned based on COSY-60, HSQC, HMQC, HMBC, and tROESY, all ran at 600 MHz. Carbon assignments for 8 come from HSQC and HMBC experiments.
Figure 3.7 $^1$H (A) and $^{13}$C (B) NMR Spectrum of 6 in DMSO-d6 at 600 MHz
Figure 3.8 $^1$H (A) and $^{13}$C (B) NMR Spectrum of 7 in DMSO-d$_6$ at 600 MHz
Figure 3.9 $^1$H NMR Spectrum of 8 in DMSO-d6 at 600 MHz
Table 3.2  Compounds 9 and 10 NMR Data in DMSO-d6

<table>
<thead>
<tr>
<th>Atom Number</th>
<th>Proton 9</th>
<th>Carbon 9</th>
<th>Proton 10</th>
<th>Carbon 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.82 s</td>
<td>22.46</td>
<td>1.82 s</td>
<td>22.46</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>168.16</td>
<td>-</td>
<td>169.20</td>
</tr>
<tr>
<td>3</td>
<td>8.00 d</td>
<td>-</td>
<td>8.00 d</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>4.29 q</td>
<td>50.95</td>
<td>4.28 q</td>
<td>50.98</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>172.08</td>
<td>-</td>
<td>172.05</td>
</tr>
<tr>
<td>6</td>
<td>7.77 d</td>
<td>-</td>
<td>7.74 d</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>4.05 t</td>
<td>56.78</td>
<td>4.07 t</td>
<td>56.64</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>171.79</td>
<td>-</td>
<td>171.46</td>
</tr>
<tr>
<td>9</td>
<td>8.60 s</td>
<td>-</td>
<td>8.61 s</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>86.69</td>
<td>-</td>
<td>86.82</td>
</tr>
<tr>
<td>11</td>
<td>2.06 m</td>
<td>29.22</td>
<td>2.08 m</td>
<td>29.02</td>
</tr>
<tr>
<td>12</td>
<td>1.84 m</td>
<td>1.89 m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2.12 bm</td>
<td>25.72</td>
<td>2.08 bm</td>
<td>25.99</td>
</tr>
<tr>
<td>14</td>
<td>3.27 bm</td>
<td>46.44</td>
<td>3.26 bm</td>
<td>46.16</td>
</tr>
<tr>
<td>15</td>
<td>3.28 bm</td>
<td>3.32 bm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>158.46</td>
<td>-</td>
<td>158.21</td>
</tr>
<tr>
<td>17</td>
<td>8.29 s</td>
<td>-</td>
<td>8.34 bs</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>9.24 s</td>
<td>-</td>
<td>9.19 s</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>5.15 s</td>
<td>81.74</td>
<td>5.09 d</td>
<td>82.14</td>
</tr>
<tr>
<td>20</td>
<td>6.81 bs</td>
<td>-</td>
<td>6.88 d</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>1.39 t</td>
<td>40.63</td>
<td>1.39 t</td>
<td>40.57</td>
</tr>
<tr>
<td>22</td>
<td>1.55 m</td>
<td>24.21</td>
<td>1.55 m</td>
<td>24.21</td>
</tr>
<tr>
<td>23</td>
<td>0.82 d</td>
<td>21.71</td>
<td>0.82 d</td>
<td>21.70</td>
</tr>
<tr>
<td>24</td>
<td>0.86 d</td>
<td>22.99</td>
<td>0.86 d</td>
<td>22.99</td>
</tr>
<tr>
<td>25</td>
<td>1.65 m</td>
<td>36.31</td>
<td>1.68 m</td>
<td>36.36</td>
</tr>
<tr>
<td>26</td>
<td>0.78 d</td>
<td>15.08</td>
<td>0.80 d</td>
<td>15.18</td>
</tr>
<tr>
<td>27</td>
<td>1.41 s</td>
<td>24.48</td>
<td>1.40 s</td>
<td>24.32</td>
</tr>
<tr>
<td>28</td>
<td>1.05 m</td>
<td>1.05 m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>0.80 t</td>
<td>10.77</td>
<td>0.81 t</td>
<td>10.81</td>
</tr>
</tbody>
</table>

$^1$H and $^{13}$C NMR at 600 MHz; Correlations were assigned based on COSY-60, HSQC, HMQC, HMBC, and tROESY, all ran at 600 MHz.
Figure 3.10 $^1$H (A) and $^{13}$C (B) NMR Spectrum of 9 in DMSO-d$_6$ at 600 MHz
Figure 3.11 $^1$H (A) and $^{13}$C (B) NMR Spectrum of 10 in DMSO-d6 at 600 MHz
Figure 3.12  Determination of $K_i$ Values of Compounds 6 and 9, and Z-IL

Dixon plots of the inhibition of the hydrolysis of Z-FR-MCA by cathepsin K in the presence of the isolated compounds 6 (A) and 9 (B), as well as the commercially available peptide Z-IL (C, D). The concentrations of Z-FR-MCA used in the assay are indicated. The concentrations of compound 6 and 9 were 4-260 and 6-400 µM, respectively. 6 and 9 inhibit cathepsin K in a competitive manner with respective $K_i$ values of 44 and 64 µM. The concentration of Z-IL used in the assay was 0.5-30 mM, and it competitively inhibits cathepsin K with a $K_i$ of 18 mM.
The synthetic peptide Z-IL was a much weaker inhibitor of cathepsin K, with a $K_i$ value of 18 mM (Figure 3.12C, D).

### 3.4 Cathepsin K-Inhibitor Complex Structure Elucidation by Crystallography

Human cathepsin K was cocrystallized with fraction I obtained from the first round of HPLC purification (Figure 3.3), which contained compounds 6, 7, and 8. The structure was refined to a resolution of 2.1 Å and a final R-factor of 17.5%. The complete data collection and refinement statistics for this model are presented in Table 3.3. All the $\phi$ and $\psi$ torsion angles of cathepsin K are found in the allowable region of the Ramachandran plot (Ramachandran and Sasisekharan, 1968), as shown in Figure 3.13.

$2F_o-F_c$ difference maps indicated that an inhibitor is bound in the active site. However, it was not immediately clear which compound, 6, 7, or 8, is bound in the active site since the electron density corresponding to the inhibitor appeared in two discontinuous parts: one part concentrated in the S2 subsite and the other in the S’ subsites. Upon further examination of the electron density maps, it became apparent that the density in the S2 subsite correlates with a cyclic moiety. As such, the inhibitor found bound in the active site of the crystal structure must belong to either compound 6 or 7. These two compounds share identical planar structures, except that one of the leucines in 6 is substituted by a valine in 7. When the electron density in the S’ sites was modeled with the inhibitor, it appeared that the longer chain of leucine fit better than the shorter valine. Thus, the obtained crystal structure is of cathepsin K in complex with compound 6 (CatK-C6). As alluded to earlier, the inhibitor appears to have been cleaved, with its pyridotriazine (PT) fragment bound to the S2 subsite (composed of residues 66, 68, 134, 160, 163, and 209), and its acetyl-leucyl-leucine (Ac-LL) fragment bound in the S’ subsite.
### Table 3.3  Data Collection and Refinement Statistics for Cathepsin K – Inhibitor Complex

<table>
<thead>
<tr>
<th>DATA COLLECTION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P4$_3$2$_1$2</td>
</tr>
<tr>
<td>Unit Cell Dimensions (Å)</td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>56.8</td>
</tr>
<tr>
<td>$b$</td>
<td>56.8</td>
</tr>
<tr>
<td>$c$</td>
<td>130.3</td>
</tr>
<tr>
<td>Number of Unique Reflections</td>
<td>12456</td>
</tr>
<tr>
<td>Mean $I/\sigma I$</td>
<td>22.2 (7.4)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>8.4 (8.6)</td>
</tr>
<tr>
<td>Merging R-factor (%)</td>
<td>8.6 (27.6)</td>
</tr>
<tr>
<td>Maximum Resolution (Å)</td>
<td>2.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REFINEMENT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution Range (Å)</td>
<td>42.8-2.1</td>
</tr>
<tr>
<td>Completeness within Range (%)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Average Thermal Factors (Å$^2$)</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>31.6</td>
</tr>
<tr>
<td>Protein Atoms</td>
<td>27.7</td>
</tr>
<tr>
<td>Pyridotriazine moiety of 6</td>
<td>78.8</td>
</tr>
<tr>
<td>Acetyl-leucyl-leucine moiety of 6</td>
<td>65.2</td>
</tr>
<tr>
<td>Water Molecules</td>
<td>48.7</td>
</tr>
<tr>
<td>Final R-factor (%)</td>
<td>17.5</td>
</tr>
<tr>
<td>Final R-free (%)</td>
<td>24.1</td>
</tr>
<tr>
<td>RMS Bond Deviation (Å)</td>
<td>0.01</td>
</tr>
<tr>
<td>RMS Angle Deviation (°)</td>
<td>1.4</td>
</tr>
</tbody>
</table>

*a Values in parenthesis refer to the highest resolution shell (2.10-2.16 Å)

*b 5% of the diffraction data was set aside for the R-free
Figure 3.13  Ramachandran Plot of Cathepsin K-Inhibitor Complex

The Ramachandran plot (Ramachandran and Sasisekharan, 1968) of the cathepsin K-inhibitor structure was created using PROCHECK (Laskowski et al., 1993). Most favorable regions are colored red, additional allowed regions are colored bright yellow, generously allowed regions are colored light yellow, and disallowed regions are white. Black triangles represent glycine residues, whereas black squares represent the remaining amino acids. Excluding glycines, all the main chain torsion angles of cathepsin K lie within the allowed regions of the plot.
subsites. More specifically, the carboxyl group of the Ac-LL fragment occupies the entrance to the S1’ site (composed of residues 137, 138, 143, 162, and 184), while the following leucine residue extends out of the active site into the solvent. The other leucine lies in the S2’ subsite (composed of residues 19-24, and 184). The catalytic C25 is oxidized to a sulfenic acid, and does not form a covalent bond with either portion of the inhibitor (Figure 3.14 and 3.15). The PT fragment of 6 is held in the S2 subsite by hydrogen bond interactions with the main chain atoms of G66, the sulfenic acid oxygen of C25, a symmetry-related K44, and a water molecule (Figure 3.16). The Ac-LL fragment of the inhibitor hydrogen bonds to G20 and Q19, which shape the S2’ subsite, W184 and H162, which comprise the S1’ subsite, and a water molecule (Figure 3.16). PT and Ac-LL are shown in their respective omit F_o-F_c electron density maps in Figure 3.17.
Figure 3.14  The Binding of Compound 6 in the Active Site of Cathepsin K

Figure (A) shows the cleavage of compound 6 to generate the two fragments that are found bound in the active site of cathepsin K. Figure (B) is a surface diagram of cathepsin K in complex with the two fragments of 6: PT bound in S2 subsite (shown in red) and Ac-LL bound in the S1’ and S2’ subsites (shown in magenta and pink, respectively). Figures (C) and (D) are snap shots of the PT fragment of 6 bound in the S2 subsite, and the Ac-LL fragment of 6 bound in the S1’ and S2’ subsites, respectively. The inhibitors are shown in thick lines with respect to the residues making up the substrate binding sites. Atoms are colored according to type: carbons in light grey, nitrogens in blue, carbons in red, and sulfurs in yellow. The figure was prepared using PyMOL (DeLano, 2002).
Figure 3.15  Surface Diagrams of the Binding of the Cleaved Compound 6 in the Active Site of Cathepsin K

Cathepsin K is shown in a surface representation in grey, and the PT and the Ac-LL fragments of 6 are represented as sticks, with atoms colored according to type: carbons in light grey, nitrogens in blue, and carbons in red. The residues making up subsites S2, S1’ and S2’ are shown. The figure was prepared using PyMOL (DeLano, 2002).
Figure 3.16 Hydrogen Bond Interactions of the Fragmented Compound 6 in the Active Site of Cathepsin K

The PT (A) and the Ac-LL (B) fragments of 6 are shown. Hydrogen bonds are indicated in dashed lines and the distances are shown in Å. Interacting amino acids are designated with their one letter code and the main or side chain atoms to which interactions are made are specified. Interacting water molecules are designated as H2O.
Figure 3.17  Electron Density Maps for Fragmented Compound 6

The PT (A) and the Ac-LL (B) fragments of 6 are shown in their respective omit F_o-F_c electron density maps contoured at 1.5 σ. Atoms are colored according to type: carbons in light grey, nitrogens in blue, and carbons in red. The figure was prepared using PyMOL (DeLano, 2002).
CHAPTER 4 – Discussion

4.1 Production of Compounds 6-10 in Streptomyces IS2-4

All the compounds isolated from the Streptomyces strain IS2-4 are peptide analogues resembling leupeptin. The most abundant form of leupeptin produced by Streptomyces is acetyl-L-leucyl-L-leucyl-DL-arginal (Figure 1.5), but derivatives containing L-isoleucine and L-valine instead of one or both of the leucines have also been found (Aoyagi et al., 1969; Hori et al., 1978). The isolated compounds 6 and 8 have an acetyl-L-leucyl-L-leucyl (Ac-LL) attached to a pyridotriazine (PT) and a butylbiguanide, respectively, instead of the arginal. Compound 7 has an acetyl-L-leucyl-L-valine (Ac-LV) attached to a PT, and compounds 9 and 10 have an acetyl-L-leucyl-L-isoleucine (Ac-LI) attached to a pyrrole imidazole instead of the arginal. As such, 6-10 represent structurally novel compounds. Still, it is easy to see their relationship to leupeptin, suggesting that 6-10 are either biosynthesized from leupeptin, or that leupeptin is biosynthesized from 6-10.

In the late 1970s and the early 1980s, a Japanese group has extensively studied the biosynthesis of leupeptin in Streptomyces roseus MA839-A1. They found that the biosynthesis of leupeptin is a nonribosomal process, utilizing three different enzymes. First, leucyl acyltransferase catalyzes the reaction between acetyl-CoA and L-leucine to yield acetyl-L-leucine (Suzukake et al., 1980). Then, leupeptin acid synthetase sequentially adds L-leucine and L- or D-arginine to acetyl-L-leucine to generate leupeptin acid (Suzukake et al., 1979). Finally, leupeptin acid reductase reduces the carboxyl of the terminal arginine in leupeptin acid to an aldehyde to produce leupeptin (Suzukake et al., 1981). Considering that the biosynthetic pathway for leupeptin has been
elucidated, it is likely that leupeptin is the precursor of compounds 6-10 and not the other way around.

The proposed biosynthesis of compounds 6 and 7 would begin with leupeptin acid and proceeds via the addition of a free guanidine to the guanidine moiety of leupeptin’s arginine, followed by a cyclization of the resulting butylbiguanide (Figure 4.1). The only two differences between compound 6 and 7 are the substitution of the middle leucine in 6 for a valine in 7, and the stereochemistry of the PT moiety. As stated above, leupeptin acid synthetase is capable of incorporating valine residues in the molecule. Moreover, leupeptin acid synthetase while capable of incorporating D-amino acids, is unable to extend chains containing D-amino acids (Suzukake et al., 1979). Therefore, since the arginine residue in leupeptin is terminal, it can be in either D or L configuration, and 6 and 7 arise from one of these possibilities. Since the stereochemistry of the PT group has not yet been elucidated, it is impossible to know which of the stereo configurations is the precursor for which compound.

It is unknown whether the formation of the PT moiety is an enzyme-mediated process. Still, it seems likely that the reaction between the free guanidine and leupeptin’s arginine is catalyzed by a Streptomyces enzyme since the reactants need to be brought together in the active site. The subsequent cyclization is probably spontaneous. Given the unique nature of the cyclic moieties of compounds 6, 7, 9 and 10, the elucidation of their biosynthetic pathway might be an interesting future area of investigation.

Compound 8 closely resembles the butylbiguanide intermediate in the formation of 6 and 7 from leupeptin acid (Figure 4.1). Compound 8, however, does not cyclize, and this is because it contains an alcohol instead of a carbonyl group. The alcohol group
The proposed biosynthetic pathways of compounds 6-10 begins with leupeptin or its acid derivative, followed by various modifications of their arginine residues. R1 designates Ac-LL for 6 and 8, Ac-LV for 7, and Ac-LI for 9 and 10. Compound 8 resembles the butylbiguanide intermediate in the biosynthesis of 6 and 7. Note that the isolated compound 8 has an alcohol in place of a carboxyl group, which prevents the cyclization of 8 to form 6 and 7.
could come from a reduced aldehyde of leupeptin or a reduced carboxyl of leupeptin acid. While 9 and 10 are also likely to be derived from leupeptin, it is more difficult to postulate how they are biosynthesized. Nevertheless, a possible outline of their biosynthesis is shown in Figure 4.1, where the guanidine moiety of arginine reacts with the terminal aldehyde to form an eight-member ring, which is then divided into two five-member rings through a reaction between CA and ND atoms of leupeptin’s arginine residue. The only difference between compounds 9 and 10 is the stereochemistry around the pyrrole imidazole moiety, which as discussed above, arises from the incorporation of a D- or L-arginine into leupeptin.

**4.2 Usefulness of Compounds 6-10 as Cathepsin K Inhibitors**

The $K_i$ of compounds 6 and 9 were determined and found to be 44 and 65 $\mu$M, respectively. The $K_i$ for compounds 7, 8, and 10 could not be determined due to insufficient material. Nevertheless, it is likely that their $K_i$ values are in the same range as those of 6 and 9 due to the high structural similarity between the compounds. Therefore, the isolated compounds are only moderate inhibitors of cathepsin K. In comparison, other known potent cathepsin K inhibitors have $K_i$ values in the low nM range (for example see Duffy et al., 1999; Thompson et al., 1998; Votta et al., 1997; Yamashita et al., 1997). It has been previously reported that the aldehyde group of leupeptin plays a crucial role for activity, and when it is oxidized to a carboxylic acid or reduced to an alcohol, the activity is greatly diminished (Aoyagi et al., 1969). This is because in the absence of an aldehyde group, leupeptin can no longer covalently modify the target enzyme, thus losing efficacy. Consequently, the moderate inhibition of cathepsin K observed for the isolated compounds 6-10 is probably due to the fact that they lack a group that would allow them
to covalently attach to the enzyme.

Leupeptin exhibits broad range specificity among proteases. It efficiently and reversibly inhibits serine and cysteine proteases such as plasmin, trypsin as well as cathepsins B, L, S, and K (Aoyagi et al., 1969; Bossard et al., 1996; Yamashita et al., 1997). The broad specificity of leupeptin is probably enhanced by the high reactivity of the aldehyde group, which compounds 6-10 do not contain. Nevertheless, leucine, isoleucine and valine can be accommodated in the active sites of many serine and cysteine proteases (Alves et al., 2003; Kurinov and Harrison, 1996; Melo et al., 2001; Storer and Menard, 1996), and it is not yet known if the PT, the butylbiguanide, and the pyrrole imidazole moieties in 6-10 will impart any selectivity to the inhibitors. Selectivity tests should be performed in the future once 6-10 are obtained in larger quantities.

4.3 Analysis of the X-ray Structure of CatK-C6 Complex

A crystal structure of cathepsin K in complex with compound 6 (CatK-C6) was obtained by cocrystallization and refined to 2.1 Å and an R-factor of 17.5% (Table 3.3). Compound 6 used for the experiment was part of a mixture containing 7 and 8 (HPLC Fraction I, Figure 3.3 and 3.4). In the structure, 6 appears to have been cleaved by cathepsin K into Ac-LL and PT moieties, with Ac-LL bound in the prime subsites, and PT occupying the nonprime S2 subsite (Figure 3.14 and 3.15).

4.3.1 Cleavage of Compound 6 by Cathepsin K

In order to be cleaved by cathepsin K, 6 must bind in the active site in such a way that positions PT in S1’ subsite, and the two leucines in Ac-LL in the S1 and S2 subsites. The S2 subsite is the main determinant of specificity in cathepsin K, and it preferentially
accommodates the aliphatic amino acids leucine, isoleucine, and valine (Alves et al., 2003). While the S1 subsite has a preference for basic amino acids, it can still accommodate leucine residues, especially if they are directed into the S1 subsite by a leucine in the P2 position (Alves et al., 2003; Lecaille et al., 2002). The S1’ subsite has fairly broad specificity, accepting both hydrophobic and hydrophilic amino acids (Alves et al., 2003). It is also spacious (McGrath, 1999), and thus should be able to accommodate the bulky PT moiety. The MCA group of the cysteine protease substrate Z-FR-MCA is a bicyclic bulky group similar to PT (Figure 4.2). The fact that MCA interacts with the S1’ subsite (Melo et al., 2001) supports the assumption that PT can fit in this site as well. Similar to Z-FR-MCA, in which the peptide bond joining MCA (in S1’ subsite) and arginine (in S1 subsite) is cleaved, the peptide bond joining PT and the leucine in the S1 subsite in 6 is cleaved by cathepsin K to yield free PT and Ac-LL through its normal catalytic mechanism (Figure 1.2 and 4.2).

4.3.2 Binding of Compound 6 in the Active Site of Cathepsin K

As mention above, the PT group of the cleaved compound 6 binds in the S2 subsite, and its Ac-LL moiety interacts with S1’ and S2’ subsites. This implies that upon cleavage, PT and Ac-LL dissociate from their respective subsites and switch places with each other (Figure 3.14, 3.15 and 4.2).

The primary specificity of cathepsin K is for binding aliphatic amino acids in the S2 subsite, namely leucine, isoleucine, and valine (Alves et al., 2003. McGrath, 1999), and many known cathepsin K inhibitors place a leucine residue into the S2 subsite (Duffy et al., 1999; McGrath et al., 1997; Thompson et al., 1997; Thompson et al., 1998; Zhao et al., 1997). Nevertheless, there are inhibitors in which leucines are the P’ residues
Figure 4.2  Subsite Occupancy and Cleavage Site of Z-FR-MCA and Compound 6

Amino acids in Z-FR-MCA and 6 are indicated by their one letter code. The cleavage site and the subsite each moiety occupies are indicated. Figure adapted from (Melo et al., 2001).
(Thompson et al., 1997; Yamashita et al., 1997), and in the x-ray structure of the closely related cysteine protease papain complexed with the inhibitors carbobenzyloxy-leucinyl-leucinyl-leucinal and carbobenzyloxy-leucinyl-leucinyl methoxymethyl ketone, it was observed that both of the inhibitors occupied the S’ subsites (LaLonde et al., 1998). It should be noted, however, that in contrast to cathepsin K, papain favors aromatic residues over aliphatic residues as P2 (Portaro et al., 2000). Nevertheless, the prime subsites of cathepsin K seem to be able to accommodate Ac-LL, although, theoretically, dropping one of its leucine residues in the S2 subsites should be more favorable.

Therefore, the preferential binding of PT over Ac-LL in the S2 subsite is surprising. PT binds in the S2 subsite in such a way that positions its triazine group along the entrance to the subsite with only its aminopiperidine group protruding into the subsite. Since the aminopiperidine group of PT is less hydrophilic than its triazine group, and considering the preference of the S2 pocket for hydrophobic groups, this is the best orientation that the PT fragment can assume in order to bind within the S2 subsite.

Both PT and Ac-LL are weakly bound within their respective subsites, as judged by refined high average thermal factors (79 Å² for PT and 65 Å² for Ac-LL). Thus, despite the ability of the S1’-S2’ and S2 subsites to accommodate Ac-LL and PT, respectively, they do not appear to form any strong interactions within the active site. In the crystal form obtained for the CatK-C6 complex, even though some amino acids from symmetry related proteins protrude into the active site, except for a single hydrogen bond to the PT fragment, they do not make any other contacts with Ac-LL and PT. This suggests that such interactions do not play a major role in the binding of Ac-LL and PT in the active site.
In summary, the structural data from x-ray diffraction studies suggests that compound 6 must first bind in the active site by positioning the PT group near the S1’ subsite, with the two leucine residues occupying the S1 and S2 subsites. This positions the carbonyl carbon of the scissile peptide bond between the PT and Ac-LL moieties near the nucleophilic C25, which allows cathepsin K to cleave 6. Then, the generated PT and Ac-LL fragments diffuse out of their respective subsites, and bind back in the active site. This time, PT binds in the S2 subsite, and Ac-LL occupies the S1’ and S2’ subsites. The rearrangement of the cleaved fragments within the active site is unusual, and the reasoning behind this is unclear.

4.3.3 Comparison of CatK-C6 to WT-CatK

The overall fold of cathepsin K in catK-C6 complex is closely similar to the structure of wild-type, unliganded cathepsin K (wt-catK) (coordinates provided by Dr. Adeleke Aguda; data collection and refinements statistics for wt-catK are presented in Table 4.1 and the Ramachandran plot in Figure 4.3). The mean r.m.s. (root-mean-squared) deviation in the main chain atoms between the two structures is 0.35 Å (Figure 4.4A). Most of the differences in main chain positions occur on the surface of the protein and away from the active site. These residues also have high thermal factors (Figure 4.4B). Nonetheless, three polypeptide segments located near the substrate binding sites of cathepsin K exhibit significant changes in their thermal factors or are shifted considerably in catK-C6 compared to wt-catK. These segments include residues 21-22, 64-67, and 160-162.

Q21 and C22 are part of the polypeptide loop comprising the S2’ subsite. The nearby Q19 and G20 form hydrogen bond interactions with O and N of P2’ of Ac-LL
Table 4.1 Data Collection and Refinement Statistics for WT-CatK

<table>
<thead>
<tr>
<th>DATA COLLECTION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Unit Cell Dimensions (Å)</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>32.1</td>
</tr>
<tr>
<td>b</td>
<td>66.5</td>
</tr>
<tr>
<td>c</td>
<td>93.3</td>
</tr>
<tr>
<td>Number of Unique Reflections</td>
<td>8333</td>
</tr>
<tr>
<td>Mean I/σᵣᵃ</td>
<td>15.7 (10.0)</td>
</tr>
<tr>
<td>Multiplicityᵃ</td>
<td>4.7 (4.8)</td>
</tr>
<tr>
<td>Merging R-factor (%)ᵃ</td>
<td>9.5 (14.3)</td>
</tr>
<tr>
<td>Maximum Resolution (Å)</td>
<td>2.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REFINEMENT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution Range (Å)</td>
<td>24.6-2.4</td>
</tr>
<tr>
<td>Completeness within Range (%)ᵃ</td>
<td>99.5 (99.8)</td>
</tr>
<tr>
<td>Average Thermal Factors (Å²)</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>13.4</td>
</tr>
<tr>
<td>Protein Atoms</td>
<td>11.9</td>
</tr>
<tr>
<td>Water Molecules</td>
<td>22.4</td>
</tr>
<tr>
<td>Final R-factor (%)</td>
<td>16.8</td>
</tr>
<tr>
<td>Final R-free (%)ᵇ</td>
<td>25.9</td>
</tr>
<tr>
<td>RMS Bond Deviation (Å)</td>
<td>0.01</td>
</tr>
<tr>
<td>RMS Angle Deviation (°)</td>
<td>1.3</td>
</tr>
</tbody>
</table>

ᵃ Values in parenthesis refer to the highest resolution shell (2.40-2.53 Å)
b 5% of the diffraction data was set aside for the R-free
* WT-catK was crystallized using the hanging drop method. 5 µl of 4.75 mg/ml enzyme solution (50 mM sodium acetate, 0.5 mM dithioerythreitol, 2.5 mM ethylenediaminetetraacetic acid, pH 5.0) was mixed with 5 µl reservoir solution (0.2 M ammonium sulfate, 24% PEG 8000, pH 5.5), and equilibrated against reservoir solution until crystals were obtained.
Figure 4.3 Ramachandran Plot of WT-CatK

The Ramachandran plot (Ramachandran and Sasisekharan, 1968) of the wt-catK structure was created using PROCHECK (Laskowski et al., 1993). Most favorable regions are colored red, additional allowed regions are colored bright yellow, generously allowed regions are colored light yellow, and disallowed regions are white. Black triangles represent glycine residues, whereas black squares represent the remaining amino acids. Excluding glycines, all the main chain torsion angles of cathepsin K lie within the allowed regions of the plot.
Figure 4.4  Average R.M.S. Deviation (A) and Average Thermal Factors Comparison (B) Between WT-CatK and CatK-C6

Figure (A) represents the average r.m.s. deviation of main chain atoms between the structures of wt-catK and catK-C6. The red horizontal line represents the overall mean r.m.s. deviation observed for these atoms of 0.35 Å. Figure (B) represents the average thermal factors of main chain atoms in catK-C6 (red line) and wt-catK (black line).
(Figure 3.16). While Q19 does not significantly adjust its position in response to the binding of Ac-LL, the main chain O of G20 in wt-catK comes within 2.13 Å from the carbonyl of the acetyl group in Ac-LL, and as a result is pushed back by 0.88 Å. In addition, the main chain O of Q21 in catK-C6 is rotated away relative to its position in the uncomplexed enzyme to avoid steric clashed with P2’ (Figure 4.5). Residues 19-23 are part of a surface loop that includes two glycine residues, which might impart a degree of flexibility on this segment of the enzyme, and consequently, the S2’ subsite.

The polypeptide chain that includes residues 64-67 is located near the S2 subsite and the PT moiety bound to it. While no significant movement in the residues of this segment occurs upon ligand binding, there is a slight (~3 Å²) decrease in its thermal motion (Figure 4.4A, B). This stabilization could be the result of the two hydrogen bonds the main chain atoms of G66 form with the triazine moiety of PT (Figure 3.16). Nearby residues 59-62 also exhibit a decrease in their thermal factors, ranging from 2-9 Å². They are part of a surface loop and make interactions with a symmetry related enzyme in both catK-C6 and wt-catK. However, these interactions differ between the two enzymes because they are crystallized in different space groups (Table 3.3 and 4.1). It might be that catK-C6 makes stronger symmetry related interactions than wt-catK, and this plays a role in the stabilization of residues 64-67 near the S2 subsite.

Residues 160-163 help shape the S2 and S1’ subsites, and contain the catalytic H162. The positions of the atoms in this segment are nearly identical between catK-C6 and wt-catK, but a thermal motion decrease by ~9 Å² upon inhibitor binding is observed (Figure 4.4A, B). The only residue to form a hydrogen bond with the bound ligands is H162, which interacts with the carboxyl of Ac-LL. Additional stabilization might come
Figure 4.5  Comparison of the S2’ Subsite Between WT-CatK and CatK-C6
Figure (A) is an overlay of wt-catK (dark grey) and catK-C6 (light grey). Residues 19-22, which are part of a loop that defines the S2’ subsite, are shown. These residues shift position in response to Ac-LL binding. Close contacts are indicated with dashed lines and the distance is shown in Å. Figures (B) and (C) show the Ac-LL fragment of catK-C6 and wt-catK, respectively, in the same orientation as in Figure (A). Atoms are colored according to their type: carbons in light grey, nitrogens in blue, oxygens in red, and sulfurs in yellow. Ac-LL is shown in thick lines with respect to the residues making up the S2’ subsite. Figure prepared using PyMOL (DeLano, 2002).
from van der Waals contacts between PT and the nearby L160, N161, and A163.

4.3.4 Comparison of CatK-C6 and CatK-CH Complexes

Since the discovery that cathepsin K may represent a viable drug target for the treatment of osteoporosis, numerous active site directed inhibitors of the enzyme have been developed. Most make a covalent bond with C25 and occupy several substrate-binding sites. Some inhibitors bind only in the nonprime direction, while others extend across both the S and S’ subsites. The specificity and activity of these inhibitors depends on how well they fit in the subsites they occupy, and the strength of the interactions they make with nearby protein amino acids.

The carbohydrazide (CH) compound shown in Figure 4.6 is a covalent and potent inhibitor of cathepsin K, having a \( K_i \) value of 22 nM against the enzyme (Thompson et al., 1997; Yamashita et al., 1997). The inhibitor spans the S3’-S3 subsites of the active site. Specifically, its two phenyl rings interact with S3’ and S3 subsites and form aromatic interactions with W184 and Y67, respectively, and one of its leucine residues is buried in the S2 subsite, while the other interacts with elements of the S2’ site. In comparing the positions of the P2’ leucines in Ac-LL and CH, it becomes apparent that they are shifted relative to each other, which results in the different hydrogen bonding interactions they make with the active site (Figure 3.16, 4.6, and 4.7). In Ac-LL, the O and N of the P2’ leucine hydrogen bond to Q19 and G20, respectively. The only hydrogen bond that the P2’ leucine in CH makes is between its main chain O and W184. On the other hand, the binding of the P2 leucine of CH correlates well with the binding of PT in the S2 subsite (Figure 4.7), and both moieties make interactions with the main chain atoms of G66 (Figure 3.16 and Figure 4.6).
Figure 4.6  Hydrogen Bond Interactions of CH in the Active Site of Cathepsin K

Hydrogen bonds are indicated by dashed lines. Interacting amino acids are designated with their one letter code, and the side chain atoms to which interactions are made are specified. Hydrogen bond distances were taken to be $\leq 3.4$ Å. CH binds in both prime and nonprime substrate binding sites of cathepsin K.
Figure 4.7  Comparison of the Binding Modes of CH and Compound 6

Figure (A) is an overlay of CH and 6 bound in the active site of cathepsin K, with CH shown in dark grey and 6 in light grey. The binding subsites are labeled. Figure (B) is also an overlay of CH and 6, but with the atoms colored according to type. Carbons are in light grey for 6 and green for CH, nitrogens are in blue, and oxygens are in red. Figure prepared using PyMOL (DeLano, 2002).
Examination of amino acids comprising the substrate-binding sites revealed some differences in the conformations of the main chain and side chain atoms (Figure 4.8). In the structure of cathepsin K in complex with CH (catK-CH), Y67 adjusts itself to allow stacking with the P3 phenyl ring of CH. D61 also orients itself so that it faces the S3 subsite. The S3 subsite in the catK-C6 structure is occupied by a symmetry related K44, which hydrogen bonds to the PT moiety.

Amino acids making up the S2 subsite do not exhibit any significant differences between the two structures. The PT of Ac-LL and the P2 leucine of CH protrude equally deep into the subsite. The fact that no amino acids had to adjust in order to accept the bulkier cyclic PT moiety suggests that such groups are well tolerated in the S2 subsite.

As stated above, both Ac-LL and CH have a leucine residue occupying the S2’ subsite, but the P2’ leucine of CH shifts ~1.2 Å towards the nonprime sites of the enzyme relative to the P2’ leucine of Ac-LL (Figure 4.7). This difference is most likely the result of the interactions CH makes in the active site, namely, its covalent bond to C25 as well as the π-stacking between its phenyl rings and W184 and Y67. Residue 20-22, which help define the S2’ subsite are also shifted by ~1.2 Å relative to one another in the two structures (Figure 4.9), to maintain favorable interactions with the P2’ residues. A similar observation was made when catK-C6 was compared to wt-catK (Figure 4.5), which supports the assumption that the S2’ subsite is flexible and able to adjust itself to better accommodate the bound P2’ residue.

The overall structure of cathepsin K in catK-C6 and catK-CH complexes is also very similar. The overall mean r.m.s. deviation in the main chain atoms between the two structures is 0.36 Å (Figure 4.10). With the exception of the polypeptide segments
Figure 4.8  Comparison of the S3 Subsite Between CatK-CH and CatK-C6

Figure (A) is an overlay of catK-CH (dark grey) and catK-C6 (light grey). D61 and Y67, which define the S3 subsite, are shown. These residues shift position in response to CH binding. Figures (B) and (C) show the PT fragment of catK-C6 and catK-CH, respectively, in the same orientation as in Figure (A). Atoms are colored according to their type: carbons in light grey, nitrogens in blue, and oxygens in red. In all Figures, the ligands are shown in thick lines with respect to the residues making up the S3 subsite. Figure prepared using PyMOL (DeLano, 2002).
Figure 4.9  **Comparison of the S2’ Subsite Between CatK-CH and CatK-C6**

Figure (A) is an overlay of catK-CH (dark grey) and catK-C6 (light grey). Residues 20-22, which define the S2’ subsite, are shown. These residues shift position in response to CH and Ac-LL binding. Figures (B) and (C) show the Ac-LL fragment of catK-C6 and catK-CH, respectively, in the same orientation as in Figure (A). Atoms are colored according to their type: carbons in light grey, nitrogens in blue, and oxygens in red. In all Figures, the ligands are shown in thick lines with respect to the residues making up the S2’ subsite. Figure prepared using PyMOL (DeLano, 2002).
Figure 4.10  Average R.M.S. Deviation Between CatK-CH and CatK-C6

The average r.m.s. deviation of main chain atoms between the structures of catK-CH and catK-C6 is shown. The red horizontal line represents the overall mean r.m.s. deviation observed for these atoms of 0.36 Å.
described above, all regions exhibiting large r.m.s. deviations are surface loops that are remote from the substrate binding site. Unfortunately, it was not possible to compare thermal factors between the two structures. The thermal factors of all protein and inhibitor atoms in the catK-CH structure appear to have been fixed during refinement at 15 Å².

4.3.5 Inhibition of Cathepsin K by Compound 6

One question that arises from the above discussion is in which form does compound 6 inhibit cathepsin K, the cleaved one or the uncleaved one?

The synthetic peptide Z-IL (Figure 3.12D) was used to simulate cathepsin K’s inhibition by the Ac-LL fragment of 6. Assuming that Z-IL binds in the active site in the exact same way as Ac-LL, it should position its leucine residue in the S1’ subsite, and the isoleucine residue in the S2’ subsite. This should allow the aromatic Z-group to pie stack with W184. The additional aromatic interaction that Z-IL is able to make within the active site should hold it tighter in the active site compared to Ac-LL, and thus, should make Z-IL a more potent inhibitor of cathepsin K. However, Z-IL was found to be a poor inhibitor of the enzyme, having a $K_i$ value of 18 mM. This suggests that Ac-LL is an even poorer inhibitor of cathepsin K.

Kinetic data for the inhibition of cathepsin K by PT or a compound that is similar to PT is yet to be determined. However, the crystallographic data suggest that it is also likely to be a poor inhibitor of cathepsin K. PT appears to be loosely bound in the active site of cathepsin K as evidenced by the high value of its thermal factors.

Consequently, it would seem cathepsin K is initially inhibited by the uncleaved compound 6. This is followed by a slow hydrolysis of 6 to yield PT and Ac-LL, which
appear to be poorer inhibitors of cathepsin K than the uncleaved 6.

4.3.6 Formation of C25-Sulfenic Acid

In the crystal structure of CatK-C6, the catalytic C25 appears to have been oxidized to a sulfenic acid. Protein sulfenic acids are normally unstable, and are readily and irreversibly oxidized by molecular oxygen to sulfinic and sulfonic acid, or reduced to form disulfides. Nevertheless, several factors can contribute to the stabilization of cysteine-sulfenic acids in proteins. These include the absence of nearby thiol groups, limited solvent access, an apolar environment and the presence of hydrogen bond acceptors (Allison, 1976; Clairborne et al., 1993). Three of the requirements for a stable cysteine-sulfenic acid are met in cathepsin K that is crystallized with 6. There are no thiol groups close enough to C25 to be able to form a disulfide bridge with it. Its solvent accessibility is limited by nearby symmetry related proteins, and it hydrogen bonds to the PT moiety of 6. In fact, stable cysteine-sulfenic acids have been previously observed in cysteine proteases, including cathepsin K (Lin et al., 1975; McGrath et al., 2003).

4.4 Overview of Unsuccessful Approaches Used in Inhibitor Purification

Since this thesis might be used in the Bromme lab as a reference for the future purification of natural products, a discussion of the methodologies that were not effective in the purification of compounds 6-10 is warranted.

First, much effort has been put into purifying the inhibitors using the size-exclusion resin, Sephadex LH-20 (GE Healthcare, NJ). Unfortunately, once the inhibitory activity started to elute from the column, it smeared over numerous fractions, resulting in poor resolution and sample recovery. Initially, it was thought that this was due to ionic or
cyclic groups in the active compounds, since these tend to adsorb to the resin. However, changes in the elution solvent to remedy the problem, such as adding 1% acetic acid, eluting with different percentages of MeOH, ACN or isopropanol did not help. Eventually, it was discovered that the “smearing problem” mainly stemmed from incorrect packing of the column. More specifically, the column was repacked before each use, and once the slurry was poured into the column, it was not given sufficient time to set appropriately before the column was used again. In the future, once the column is packed, it should be left to set for several days, while packing solvent is passed through the column at a slow flow rate. At the end, Sephadex LH-20 was not used in the purification of compounds 6-10 because by the time the problems were resolved, the purification of the compounds was optimized using RP C18 and WCX columns, and the use of the LH-20 column did not yield better results.

Second, the media used to grow Streptomyces is very complex, containing ingredients such as malt and yeast extracts as well as large amounts of carbohydrates and peptones. Considering that the active compounds are secreted into the media, trials were set up to simplify the media while retaining the inhibitory activity. These included various combinations of reduction or complete elimination of certain media components, particularly malt and yeast extract since they contain many different amino acids, peptides, and sugars. However, a dramatic reduction in the production of the inhibitory activity was observed with any of the simplified media tested when compared to the original complex media, suggesting that all of the ingredients present in the original media are required by the bacteria to produce the active compounds.
Third, the final step in the purification of the active compounds was their separation by RP C18 HPLC. As discussed in section 3.1, the *Streptomyces* strain IS2-4 produced at least nine compounds that had the ability to inhibit cathepsin K, only five of which were successfully isolated. All of the active compounds are chemically closely related, and are extremely difficult to separate by HPLC. The Phenomenex brand column was only successful in the resolution of compound 6 after extensive method development, which included different gradients of ACN and MeOH, with various concentrations of TFA and at a range of flow rates. Compounds 7 and 8 were consistently clustered as a single peak. The same was true for 9, 10, and the remaining four inhibitors that were not successfully purified. The Intersil brand semi-prep RP C18 HPLC column that was subsequently used had a better resolving power, and it effectively isolated 7, 8, 9, and 10. Nevertheless, four additional active compounds still remain impure. Reapplying them onto the Intersil brand column might have yielded better results, however, they were present in such low amounts (> 1 mg) that it was deemed better to produce them in larger quantities before continuing with their purification.

Finally, attempts were made at purifying the active compounds by batch affinity chromatography using papain. Papain is a plant cysteine protease, and it was found that the active compounds produced by *Streptomyces* IS2-4 inhibit papain almost as well as they inhibit cathepsin K. Papain was used over cathepsin K because, unlike cathepsin K, papain is readily available commercially. In the experiment, 10 µM papain was mixed with 4 nmol of the semi-purified mixture of active compounds. The resulting papain-inhibitor mixture was assayed for papain activity (as described in Section 2.1.3) to ensure that all of the papain was effectively inhibited. Then, the papain-inhibitor mixture was
filtered using a centrifuge 10K filter, which should trap papain (and consequently, the inhibitor bound to it), but allow all the low molecular weight impurities to pass through. The inhibitor could then be released by denaturing papain. Unfortunately, most of the inhibitory activity was recovered in the filtrate. This is not surprising since such purification requires the inhibitors to bind covalently to the enzyme, and once compounds 6-10 were isolated it became apparent that they are not covalent inhibitors.

A flow chart outlining the purification of compounds 6-10, including the “dead-end” steps described above is presented in Figure 4.11.
Figure 4.11  Isolation of Microbial Natural Products 6-10

Outlined in red, are the “dead-end” or “troublesome” steps in the purification

Fermentation broth of *Streptomyces* sp.

- Supernatant separated by centrifugation
- Adsorption resin
  - Amberlite XAD-4
  - Wash with water and 20% MeOH
  - Elute with 100% MeOH
- EtOAc extraction
  - Aqueous phase
- Anion exchange resin
  - Dowex Marathon A
  - Filtrate
  - BuOH extraction
    - Organic fraction
    - RP C18 Column
      - Active fraction
        (20% ACN + 0.05% TFA)
      - WCX Column
        - Active fraction
          (50% MeOH + 2% FA)
      - RP C18 HPLC
        - Compound 6
        - Compound 7 and 8
        - Compound 9 and 10

Growth media simplification

Size-exclusion column LH-20

Affinity purification using papain

Extensive method development and two different HPLC systems
In the search for naturally derived cathepsin K inhibitors, five structurally novel compounds 6-10 were isolated from the *Streptomyces* strain IS2-4. These compounds appear to be derivatives of the protease inhibitor leupeptin, although this still needs to be experimentally confirmed. Compounds 6 and 9 are only moderate inhibitors of cathepsin K. Although the activity of 7, 8, and 10 against cathepsin K was not tested, they are expected to show inhibitory activity similar to 6 and 9. Future investigation should concentrate on producing compounds 6-10 in larger amounts (at least 5 mg) in order to answer several unresolved questions. In particular, the stereochemistry of 6-10 still needs to be verified, the inhibitory potency of 7, 8, and 10 still needs to be determined, and selectivity studies of all the isolated compounds against serine and cysteine proteases need to be preformed.

Furthermore, a crystal structure of human cathepsin K in complex with compound 6 was obtained. The structure revealed that 6 was cleaved into two fragments: Ac-LL and PT, and these fragments show a unique binding pattern in the active site. Specifically, Ac-LL is expected to position one of its leucine residues in the S2 subsite. Instead, PT binds in the S2 subsite, whereas Ac-LL is found in the S’ subsites. PT and Ac-LL do not appear to bind tightly in the active site, and it is still not clear what determines their new apparent specificity. It would be beneficial to recrystallize 6 with cathepsin K in a space group that does not obstruct the active site to examine whether the nearby symmetry related residues in the current space group play a significant role in the binding of Ac-LL and PT. In addition, crystallization of cathepsin K with an uncleaved 6 would verify its proposed binding in the active site (PT at S1’, and Ac-LL in S1-S2) and the possibility of
cleavage by cathepsin K. Finally, crystallization of compounds 7-10 with cathepsin K might also lead to interesting findings. For example, revealing whether 9 and 10 will be found cleaved in the crystal structure, and determining the effect of stereochemistry on ligand binding in the active site.

Overall, the goal of this investigation was to discover novel inhibitors of cathepsin K for potential therapeutic use against osteoporosis. While structurally novel compounds have been found, they are not potent inhibitors of the enzyme. In addition, they are expected to exhibit broad range specificity against proteases. Although this still needs to be verified experimentally, it is known that in addition to cathepsin K, 6-10 also inhibit papain. Nonetheless, this investigation has revealed that *Streptomyces* sp. offer some promise for the discovery of novel cathepsin K inhibitors. Since many *Streptomyces* strains exhibit activity against cathepsin K, there is hope that future investigations might reveal novel compounds that also display better pharmacological potential.
References


