Antipain and its Analogues, Natural Product Inhibitors of Cathepsin K Isolated from *Streptomyces*

by

Vincent Paul Lavallée

B.Sc., McGill University, 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)

April 2011

Abstract

In human bone, 90% of organic bone matrix is composed of type 1 collagen. Cathepsin K is a cysteine protease involved in osteoclast mediated bone absorption and has been identified as a major drug target for the treatment of osteoporosis. Numerous potent inhibitors of cathepsin K have already been identified from natural sources including epoxide inhibitors such as E-64 isolated from the fungi *Aspergillus japonicus* as well as various peptide aldehydes such as Leupeptin and alpha-MAPI purified from *Streptomyces*. 350 soil and lichen-associated bacterial strains collected in the rain forests of British Columbia were screened and 22 samples were identified containing significant cathepsin K inhibitory activity. From those active samples, L-91-3 was selected as one of the most potent samples for further characterization of their cathepsin inhibitor content. Three Antipain-related peptide inhibitors were identified from L-91-3 strain of *Streptomyces.*

Antipain (Ki 41nM +/- 37nM) and Vince 2 (cyclized P1 derivative Ki 295nM +/- 123nM) were isolated by traditional purification and subsequent NMR and Mass Spectrometry analysis. The cyclized compound, Vince 2 (phenylalanyl-ureido-arginyl-valinyl-cycloarginal), lacked the aldehyde function and resulted in a lower binding affinity towards cathepsin K. Using Cathepsin K as bait for active site directed inhibitors a third compound, named Lichostatinal was identified by x-ray crystallography where recombinant human cathepsin K was co-crystallized with the semi-crude fermentation broth resulting in 2.0 Å resolution crystal structure. Lichostatinal is a peptide-based

aldehyde with the amino acid composition (agmatinyl-ureido-serine-valinyl-arginal). The

P1-P4 substrate residues of Lichostatinal interact with the non primed S1-S4 subsites of

cathepsin K.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

Agm	Agmatine
AMC	7-aminomethylcoumarin
Å	Angstrom Unit (1 Å = 1×10^{-10} m)
13C	Carbon – 13
1H	Proton
ACN	Acetonitrile
BuOH	Butanol
C	Cytosine
COSY – 60	Correlation Spectroscopy with a Nuclear Spin Tilt of 60°
d	Doublet (in Relation to Nuclear Magnetic Resonance Signal)
DMSO – d6	Deuterated Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra Acetic Acid
EtOAc	Ethyl Acetate
FA	Formic Acid
G	Guanine
gFW	Grams Fresh Weight
HCl	Hydrochloric Acid
HMBC	Heteronuclear Multiple Bond Multiple Quantum Coherence
HMQC	Heteronuclear Multiple Quantum Coherence
HPLC	High Performance Liquid Chromatography
hPTH	Human Parathyroid Hormone
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)
M	Moles Per Liter

MAPI	Microbial Alkaline Protease Inhibitor
MCX	Mixed – Mode Strong Cation Exchange
MAX	Mixed – Mode Strong Anion Exchange
MeOH	Methanol
MHz	Megahertz
MMP	Matrix Metallo Protease
MOPS	(N-morpholino) Propanesulfonic Acid
NH_4OH	Ammonium Hydroxide
NMR	Nuclear Magnetic Resonance
N-Terminal	Amino Terminal
PEG	Polyethylene Glycol
q	Quartet (in Relation to NMR Signal)
RP	Reversed – Phase
rpm	Revolutions Per Minute
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 Rate of Reaction
WAX	Mixed – Mode Weak Anion Exchange
WCX	Mixed – Mode Weak Cation Exchange
Z-FR-MCA	Carbobenzoxy-phenylalanine-arginine-4-methylcoumarin-7-amide

1 Introduction

1.1 Proteases

Proteases are enzymes that cleave peptide bonds in proteins through a process known as proteolysis. Proteolytic cleavage has been characterized as the most common type of post-translational modification [1]. Many proteins are synthesized as inactive precursors and activated under certain conditions through limited proteolysis. A widely known example of this is the formation of active insulin from preproinsulin. Insulin is produced as preproinsulin in the pancreas were the signal peptide directs passage into the secretory vesicles and is then cleaved producing proinsulin. Inactive proinsulin is then stored until elevated glucose levels triggers an insulin response where the "C peptide" is cleaved leaving the active insulin A and B peptide connected through three disulfide bonds [1, 2]. Prior to the discovery of post-translation cleavage modifications it was known that proteins were cleaved during degradation in the cytosol or in lysosomes. Lysosomes have a high abundance of proteases and therefore it was assumed that their main function was in cellular house keeping activity. Estimates have put the potential number of proteases expressed in the human genome to be in the order of 800. More than 600 have been identified through protease related DNA sequence homology [3].

Proteases are mainly characterized through their substrate specificities and their mechanism of catalysis [4]. In addition, they can be characterized by where in the peptide they cleave. Endopeptidases (Endo) can cleave anywhere within a polypeptide chain, whereas exopeptidases (Exo) cleave at either the amino-terminal (aminopeptidase) or

1

carboxy-terminal (carboxypeptidase). Proteases are also categorized through their mechanism of peptide hydrolysis. The major mechanisms for protein hydrolysis involve one of the following amino acids: cysteine (23%), serine (32%), aspartic (3%), threonine (6%) residues forming the appropriate cysteine, serine, aspartate, and threonine protease classes. About 36 % of the proteases utilize a metallo ion for catalysis (metallo-proteases) [3].

1.2   Papain-Like Cysteine Proteases

Mammalian papain-like cysteine proteases are also known as thiol-dependent cathepsins. One of the largest families of cysteine proteases is the papain family, comprising enzymes from bacteria, plants, invertebrates and vertebrates [5]. Through analysis of the human genome database, 11 papain-like cathepsins have been identified (Figure 1.1). Mammalian cysteine proteases have been implicated in a large array of functions. These include simple protein degradation and turnover within the endosomal and lysosomal system, proenzyme activation [6], antigen processing [7], hormone maturation [8], and pathological processes such as bone resorption [9], muscular dystrophy [10], arthritis [11], invasion and metastasis [12] and Alzheimer's disease [13]. Due to the vast array of their functions and associated disorders, they have become regarded as potential therapeutic targets. All cathepsins are synthesized as inactive pre-proenzymes (Figure 1.2). The pre denoting the signal peptide, responsible for the translocation of the proenzyme into the endoplasmic reticulum where cathepsins undergo co-translational glycosylation [14]. The proenzyme refers to the inactive precursor peptide [4]. The

preproenzyme is transported across the rough endoplasmic reticulum (RER) membrane upon the cleavage of its signal peptide [15]. The propeptide is processed either through autoprocessing by cathepsin K or by other proteases such as cathepsin D in the acidified acidic lysosomal environment [16],[17],[18].

Papain-like cysteine protease
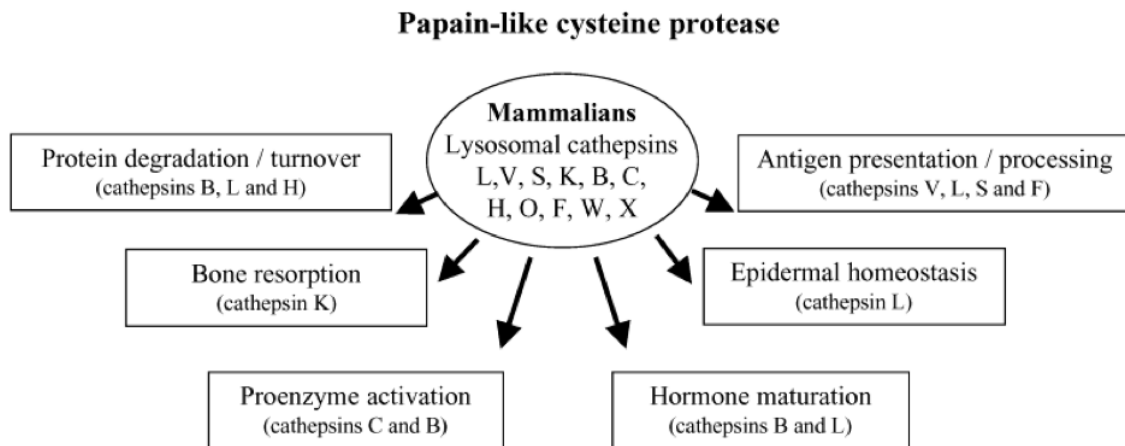
Figure 1.1 Cellular Distribution of the 11 Mammalian Cathepsins
Of the 11 mammalian cathepsins; B, H, O, X, C and L are ubiquitously expressed. The remaining five cathepsins, however, show significant selectivity in tissue expression providing distinct cellular functions. Figure taken from Lecaille et al 2002 **[4]**.

There are many functions of the cathepsin propeptide as determined through a series of mutation where amino acids were replaced with alanine in the prodomain [19]. First, it acts as a scaffold during protein folding of the catalytic domain and aiding in protein stability at neutral pH [11]. Additionally, the propeptide operates as a chaperone for the transport of the proenzyme to the endosomal-lysosomal compartment and finally, acts as a high-affinity reversible inhibitor preventing the premature activation of the catalytic domain before entering the lysosome [4]. The enzyme is fully active after cleavage and dissociation of the N-terminal domain.

Figure 1.2 Human Papain-Like Cysteine Proteases

Cathepsins are synthesized as zymogens, which are systematically cleaved into the mature protein containing the catalytic triad Cys_{25}, His_{159} and Asn_{175} (Papain numbering). Figure taken from Yasuda et al 2005 [20].

Cathepsins were once assumed to be ubiquitously expressed. However, the discovery of tissue specific cathepsins has shown that they function as highly specific metabolic or regulatory proteases [17]. While some cathepsins are ubiquitously expressed (Cathepsin B, H, X, O, C and L) and can therefore be assumed to be house keeping enzymes, Cathepsin K (osteoclast), V (thymus), S (antigen presenting cells), F (macrophages) and W (NK, CTL cells) are all selectively expressed in well-defined tissues or cell types (Figure 1.2) [4, 11]. The first papain-like cysteine protease identified to have tissue specific distribution was cathepsin S, which is selectively expressed in antigen-presenting cells such as dendritic cells and macrophages as well as in smooth muscle cells playing a critical role in antigen presentation [7].

The catalytic domains of cathepsins vary between 214 and 260 amino acids in length. The catalytic domains contain a highly conserved active site [4, 11]. Cathepsins consist of two domains in the mature enzyme; the L and R (left and right defined in accordance with the orientation used in the standard view) domain [17]. The active site cysteine residue is found in the L domain, while the histidine and asparagines are found in the R domain. Both, the L and R domain are stabilized by disulfide links, two in the L domain and one in the R domain [21]. The cysteine residue is found in a highly conserved peptide sequence, CGS$\underline{C}^{25}$WAFS with the active cysteine[25] being underlined (Cathepsin K numbering). The other two active site amino acids also show conserved sequences. The active site histidine[162] is neighbored by a small amino acid residue followed by four aliphatic hydrophobic residues and a glycine. The active site asparagine[182] is part of the Asn-Ser-Trp motif [4]. The catalytic triad (C^{25}, H^{162}, N^{182}) in the L and R domain are brought together forming a catalytic cleft between the two domains [21].

1.3   Cathepsin Mechanism of Proteolysis

The three active site residues are very important for the catalysis of peptide bond hydrolysis because they form a catalytic triad. In the active cathepsin, both the cysteine and histidine residues are ionized forming a thiolate-imidazolium pair [21, 22] that is stabilized by the nearby asparagine through hydrogen bonding [4]. In cysteine proteases, the cysteine is already ionized prior to substrate binding. For this reason cysteine proteases are regarded as "a priori" activated enzymes (Figure 1.3) [4].
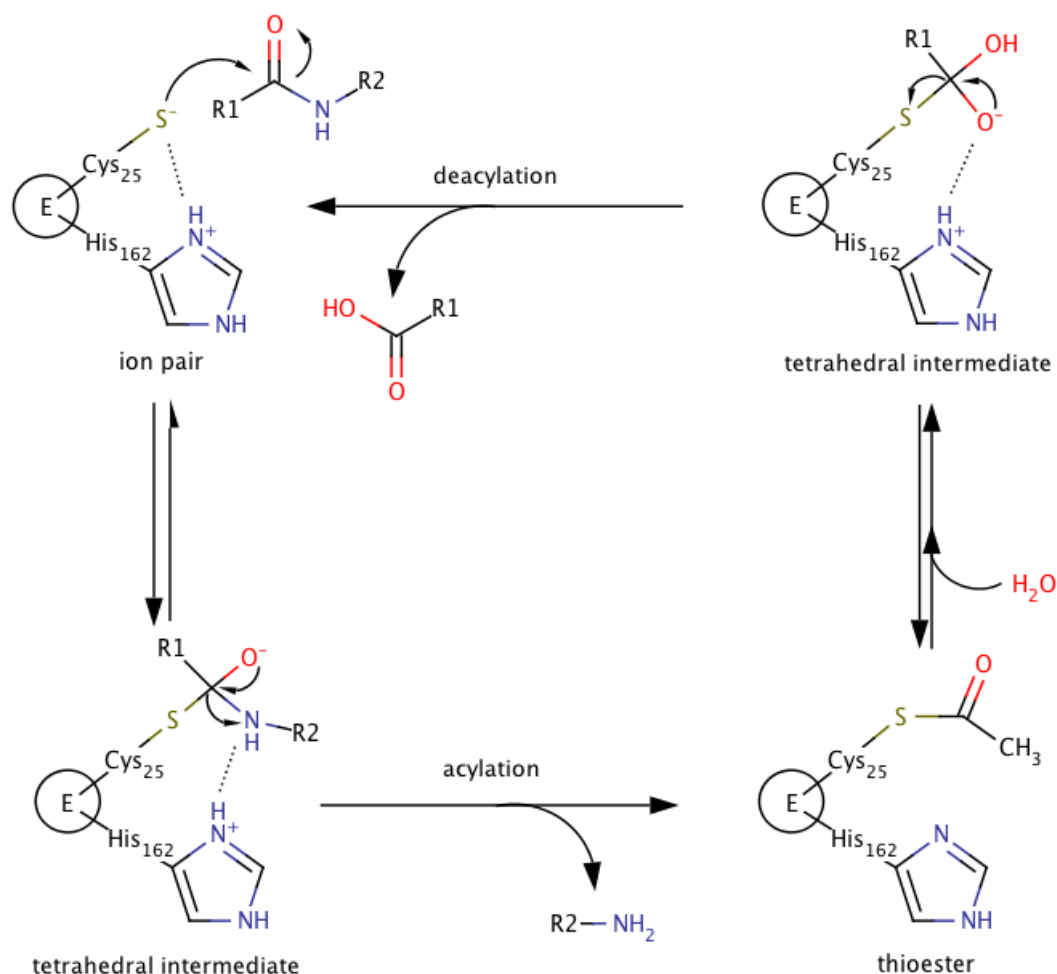
Figure 1.3 Mechanism of Substrate Hydrolysis by Cathepsins

The catalytic triad of cathepsin K is made up of Cys_{25}, His_{162} and stabilized by Asn_{182} (not shown). The nucleophilic cysteine attacks the substrate forming a tetrahedral intermediate. Cleavage of the peptide is achieved through an acylation reaction forming a thioester which is then hydrolysed with water producing another tetrahedral intermediate which through deacylation reverts the enzyme to its original state.

The cysteine thiolate is an excellent nucleophile which attacks the carbonyl carbon on the substrate peptide bond. The nucleophilic attack on the scissile bond forms a tetrahedral intermediate between the cysteine and substrate peptide. The tetrahedral intermediate is inherently unstable as a result of the negative charge on the carbonyl oxygen of the substrate [23]. This intermediate is stabilized as the carbonyl oxygen moves deeper into the active site occupying the oxyanion hole [1]. The oxyanion hole is an arrangement of

6

hydrogen bond donors which can stabilize the transition state by hydrogen bonding to the tetrahedral oxyanion [24]. Once the tetrahedral intermediate is formed, acylation takes place as the carboxyl-terminal end of the substrate peptide is released. The tetrahedral intermediate becomes transformed into an acyl enzyme. Once the scissile peptide has been cleaved, the acyl enzyme is hydrolyzed with water forming a second tetrahedral intermediate. The second tetrahedral intermediate undergoes deacylation that renews the free enzyme and releases the amino-terminal end of the substrate [4].

1.4 Cathepsin K

Cathepsin K is very closely related to Cathepsin S, but differs in that it is predominantly expressed in osteoclasts and in osteoclast-related multinucleated giant cells responsible for bone and foreign body material degradation [25]. The expression of cathepsin K is not exclusive to bone and does extend to other cell types, having been found expressed in synovial fibroblasts along with various epithelial cells [26]. Cathepsin K is capable of cleaving triple helical collagens at multiple sites within their helical domains [20]. By analyzing the cDNA library of osteoclast and comparing the homology to known sequences of Cathepsin K, 98% of the total cysteine protease expressed sequence tags belonged to cathepsin K [27].

The hypothesis that cathepsins could be potential therapeutic targets originated from the fact that E-64 along with leupeptin, both inhibitors of papain-like cysteine proteases, were capable of inhibiting bone resorption in osteoclasts [28]. The treated cells contained

vacuolar accumulations of non hydrolyzed collagen fibrils suggesting the involvement of cysteine proteases in intracellular collagen degradation [29]. This hypothesis was reinforced when the autosomal recessive bone disorder, pycnodysostosis, which is characterized by osteosclerotic bone formation, short stature and acroosteolysis of the distal phalanges [30] was linked to cathepsin K deficiency [5]. In addition, the generation of cathepsin K knock out mice produced a phenotype showing signs of having dense bones comparable to a rare human disorder osteopetrosis [30]. Upon further investigation, the osteoclasts of pycnodysostosis patients contained an accumulation of non-hydrolyzed collagen fibrils within their lysosomes. In contrast, overexpression of cathepsin K led to increased bone and cartilage degradation [31].

1.5   Substrates for Cathepsin K

Type I collagen is found in skin, artery walls, scar tissue, tendons, teeth and bone. It is a major protein in bone making up 90% of the organic bone matrix [32]. The human skeleton is under constant remodeling, with the mean age of bone close to the marrow ranging from days to 4 years old. As the distance from the surface remodeling increases, the bone becomes isolated from osteoclast and osteoblast and the bone age nears that of the individual [33]. During this process, collagen is continually degraded and synthesized to ensure structural integrity of the bone [33]. Structurally, type I collagen is made up of covalently cross-linked triple helices which contain two $\alpha1$ and one $\alpha2$ chains [34].

Other cathepsins and MMP's are capable of cleaving type I collagen, however, their cleavage products differ from that of cathepsin K. Whereas MMP's are only capable of selectively cleaving collagens through a single cut across all three chains, cathepsin K is capable of cleaving at multiple sites within the triple helical domain and completely hydrolyze collagen fibrils into small peptides [34]. The collagenase activity of cathepsin K is only comparable in nature to that of bacterial collagenases [11].

Synthetic substrates for cathepsin K also exist; carbobenzoxy-phenylalanie-arginine-methyl-7-aminocoumarin amide (Z-FR-MCA) (figure 1.4) is the most widely used substrate for evaluating the catalytic activity cysteine proteases [35]. Z-FR-MCA binds in the prime and non-prime region of the active site of cathepsins, with the MCA group occupying the S'$_1$ subsite of the protease [34] (subsite and residue notation as defined for Papain like cysteine proteases [36]). Provided there is no inhibitory activity, Z-FR-MCA is cleaved releasing the fluorogenic 7-aminomethylcoumarin (AMC) molecule providing a quantifiable assay for cathepsin K.
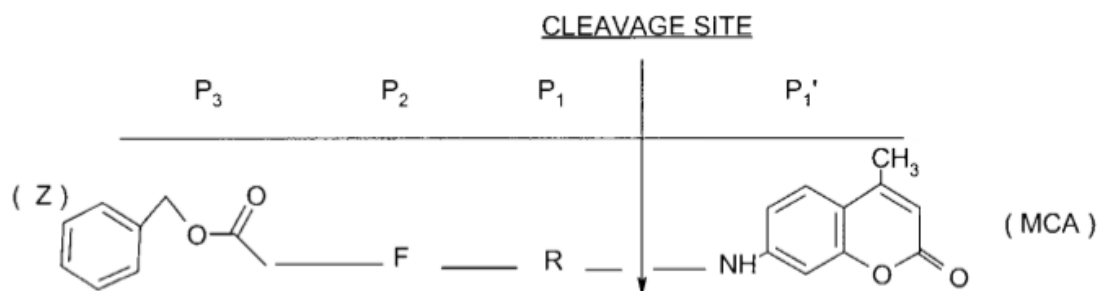
Figure 1.4 Z-FR-MCA Cleavage Site by Cathepsin K
Z-FR-MCA is hydrolyzed by cathepsin K with the MCA group binding in the P$_1$' site adjacent to the arginine amino acid binding in the P$_1$ site. Hydrolysis releases the

fluorogenic AMC molecule providing a means of quantification. Image taken from Robson M. et al 2001 [35]

1.6   Bone Remodeling

Bone resorption, a highly regulated process, which together with bone formation is responsible for giving bones their specific shape during growth along with maintaining their shape during adult life [37]. The equilibrium between resorption and formation also plays a critical role in calcium homeostasis as bones are the largest calcium store in the body containing 99% of total body calcium [38].

Bone remodeling is controlled through the interactions of two cell types, osteoclasts and osteoblasts (Figure 1.4). These cells are regulated by a vast array of autocrine (self stimulating cell capable of secreting a molecule which targets a receptor on the same cell leading to changes within the cell) and paracrine (cell signal where receptors are targeted by molecules of adjacent cells) factors [1, 2] [39]. Osteoblasts are responsible for the renewal of bone through the synthesis of the organic bone matrix components along with the deposition of the inorganic matrix. Osteoclasts on the other hand are responsible for the resorption of bone through generating an acidic and proteolytic environment which triggers the physiological and pathological degradation of the bone matrix [20].

Bone resorption begins with the expression of MMPs which are essential for the initiation of the osteoclastic resorption process by removing the collagenous layer from the bone

surface before the demineralization process can be initiated [38]. The main resorption

process is catalyzed by cathepsin K and occurs in specialized extra-cellular compartments

underneath the ruffled border of the osteoclasts. Bone resorption is characterized by two

distinct steps. First, protons are secreted onto the bone surface into the resorption lacunae

resulting in demineralization primarily of hydroxyapatite, which is the major mineral

constituent of bone [1].
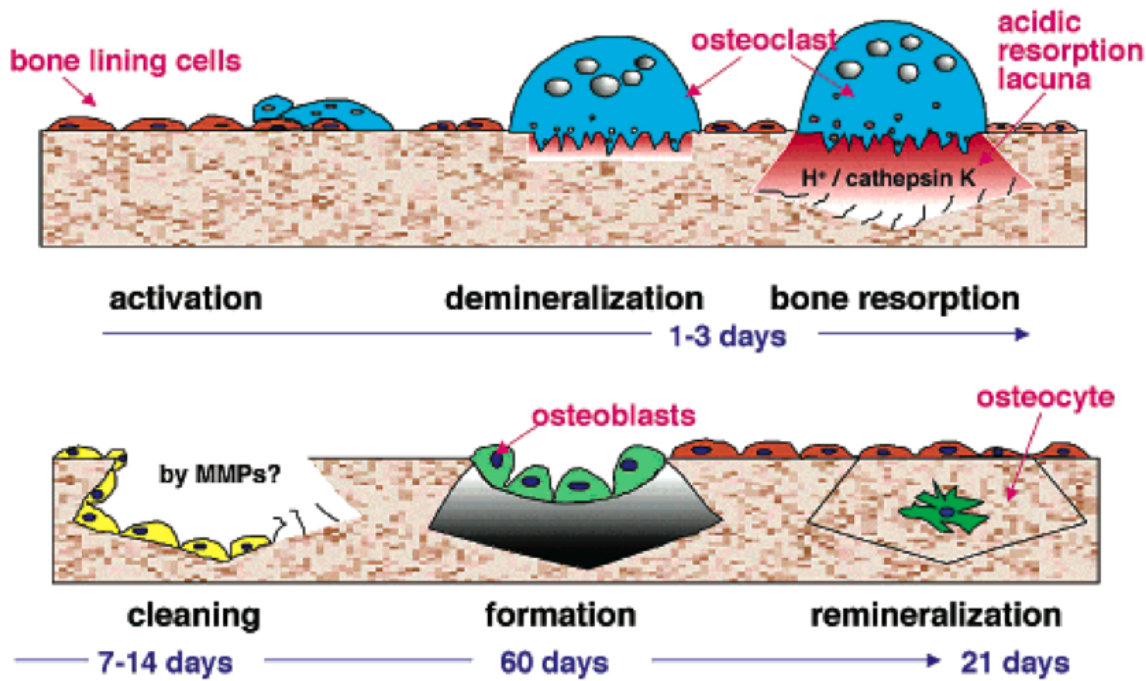
Figure 1.5 Bone Remodeling

Bone remodeling is initiated by the expression of MMP's prior to the demineralization
process. Osteoclasts are then recruited and protons are secreted into the resorption lacuna
resulting in demineralization. Cathepsin K is then released which cleaves type I collagen.
Once the bone has been excavated, osteoblast moves in and fills the excavation pit in a
process that nears completion after two months. Figure taken from Lecaille et al 2002 **[4]**.

This is followed by the enzymatic activity of cathepsin K which is responsible for the

proteolytic resorption of the exposed organic matrix made up primarily of type I collagen.

An acidic microenvironment is produced by osteoclasts and is required for both

processes, along with demineralization, lowering the pH favors the activation of cysteine

proteases [40]. Once collagen is cleaved by cathepsin K, the triple helix quickly unwinds which then becomes susceptible to any protease showing gelatinolytic activity [38].

Once the osteoclasts have fully excavated the site on the bones, the excavation pit is refilled during the "formation phase" [41] with fresh bone matrix. This is accomplished by osteoblasts using the degraded cartilage matrix as a scaffold [42] followed by the mineralization of bone at a rapid pace, with most mineralization occurring within the first two months [33].

1.7   Osteoporosis

Osteoporosis is defined as, "a disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and consequent increase in fracture risk." [43]. The decrease in bone mass is not a rare occurrence, it occurs universally with age [33]. The balance between bone formation and resorption is kept in homeostasis, and any imbalance in this equilibrium can lead to either decreased bone mass as seen in osteoporosis (Figure 1.5) or increased bone formation evident in osteopetrosis. Many factors are implicated in the genesis of osteoporosis, from genetic, environmental and lifestyle. The key to understanding osteoporosis and developing proper therapies is to establish the etiology of the disease, whether the primary defect is in over-stimulated osteoclasts or under-productive osteoblasts [33].

It is estimated that one out of four women, and one out of eight men over the age of 50 are living with osteoporosis with an even larger number at risk of developing the disease [44].

An estimated 1.9 billion dollars is spent annually in Canada on the treatment of osteoporosis and osteoporosis related bone fractures and this number is projected to increase as the general population of Canada ages [45]. Most cases of osteoporosis are caused by increases in osteoclast function, either through increased activity of individual osteoclasts or an increase in the overall number of osteoclasts rather than through deficient osteoblasts [46].
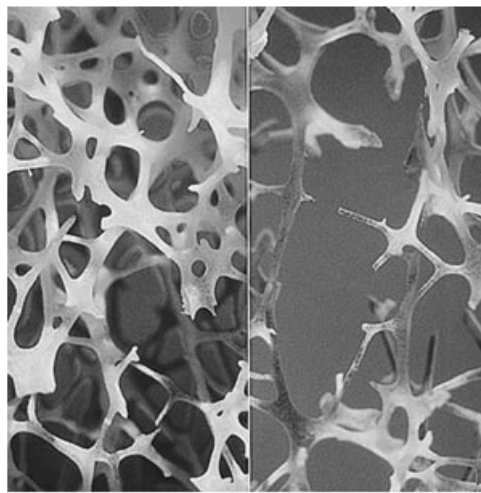
Figure 1.6 Healthy Bone vs Osteoporotic Bone
Osteoporotic bone becomes significantly more brittle and porous as bone density decreases due to disease.
Photo taken from Leo Pharmaceutical Products Sarath Ltd. website:
http://www.leo-sarath.com/index.asp?a_id=546 [47]

Decrease in bone mineralization in relation to bone strength is not a linear relationship. A modest 7% increase in bone mineralization has been found to be responsible for a 3 – fold increase in bone stiffness and doubles the bones resistance to fracture [33]. Although mineralization is the primary contributor to bone strength and fracture resistance, collagen and the organic matrix give bone the structural integrity that allows it to absorb shock [33].

1.8 Treatment of Osteoporosis

There are many treatments available to prevent bone loss that focuses on three principles. (1) Reducing the differentiation rate of osteoclasts to decrease the number of active resorbing cells, (2) inhibiting the resorptive ability of osteoclasts without affecting their formation, and (3) shortening osteoclast life span and reducing their total capacity for resorption [48]. Therapies range from hormone replacement therapy, bisphosphonates, calcitonin, calcium and vitamin D supplements [20]. All of the treatments with the exception of human parathyroid hormone (hPTH) therapy reduce loss of bone mineral by diminishing bone resorption [20, 48-53]. Although all of these therapies aid in the treatment of osteoporosis, only hPTH actively produces new bone, however, no therapy can completely prevent osteoporotic bone loss and fractures [50]. Additionally none are without side effects ranging from physical discomfort in the form of gastrointestinal disturbances, muscle and stomach cramps, constipation, nausea, headaches, bloating, hot flushes and skin tenderness. They are also associated with some more serious symptoms including deep vein thrombosis and high blood pressure [20]. Even more serious, estrogen replacement therapy along with estrogen receptor modulators (raloxifene) have been demonstrated to cause tumor growth [54]. Potential side effects such as tumor growth and cancer are of great concern when considering the long-term therapy that is required for the treatment and prevention individuals will undergo for osteoporosis.

The discovery of cathepsin K opens up a whole new approach for combating osteoporosis. Selective inhibitors for cathepsin K would result in a novel anti-resorptive drug which would be safer and more effective since cathepsins have fewer off site targets.

The active site of cathepsins is formed by a nucleophilic thiol residue which can easily be targeted by an electrophilic moiety placed into a peptide or other small molecular structure which is recognized by the substrate binding region of cathepsin K [20]. Inhibitors can come in two forms, reversible and irreversible, however, for the purpose of therapeutic use (long term treatment), reversible inhibitors are preferred [20, 55]. Highly potent irreversible inhibitors, despite their specificity, when used chronically as would be the case in the treatment of osteoporosis, could eventually react with other cysteine proteases causing toxic side effects [56].

1.9 Inhibitors of Cathepsin K

Papain – like cysteine proteases are regulated by numerous endogenous protein based inhibitors, the two major classes are the cystatins and serpins [4]. These macromolecules are too large for therapeutic use, and the focus for therapeutic cathepsin K inhibitors was redirected towards small natural products. The first cathepsin inhibitors isolated were small peptide based aldehydes of natural product origin, with one of the first isolated being Leupeptin identified in 1969 from a strain of *Streptomyces* exfoliates [57]. Leupeptin was followed by other small peptide aldehydes such as Chymostatin (1970)

[58], Antipain (1972) [59] and the microbial alkaline proteinase inhibitor (alpha-MAPI) (1979) [60]. Unfortunately, aldehyde inhibitors require a nucleophilic attack of cysteine proteases, which makes them a target for nucleophilic attack from other proteases such as serine proteases resulting in non selective inhibition [4]. Although Leupeptin, Chymostatin, Antipain and alpha-MAPI are all potent cathepsin K inhibitors, they are not selective enough at inhibiting only one protease and instead inhibit a whole array of cysteine and serine proteases rendering the inhibitors too toxic for therapeutic use [55]. Numerous selective inhibitors for cysteine proteases have been synthesized, diacyl bis hydrazides which are selective irreversible inhibitors of cathepsin K [61], irreversible epoxide compounds similat to E-64 which inhibit a wide range of cysteine proteases [62], and finally vinyl sulfones [63]. Unfortunately the irreversible behavior and wide spread inhibition towards cysteine proteases made these inhibitors inadequate therapeutic compounds for cathepsin K. The irreversible inhibition, when used chronically resulted in antigenic and immunologic complications by generating immunogenic haptens from covalently bound inhibitor-cathepsin adducts as well as the eventual inhibition of off target proteases [64].

As a result of the negative effects of irreversible inhibitors, isolation efforts were focused on reversible inhibitors, which include nitriles [65], acyclic and cyclic ketones [66, 67], amides [68], and peptidyl aldehydes [58]. With the identification of reversible cathepsin K inhibitors, in order for them to be therapeutically relevant they need to be selective for cathepsin K by having a higher affinity for the enzyme than the other closely related cathepsins (100 fold greater affinity) [64].

Currently no FDA approved cathepsin K inhibitors are available for the treatment of osteoporosis; however there are numerous drugs currently at different phases of testing in clinical trials. Balicatib (AAE581 Figure 1.7) a peptidic nitrile developed by Novartis has passed phase II clinical trials and has very high affinity for cathepsin K (K_i 0.7-1.4nM) [69, 70], with significantly greater affinity than towards the other closely related cathepsin (B, L and S with K_i 500-65000 fold greater). Unfortunately Balicatib accumulates in the lysosomes and has been linked to adverse effects due to nonselective off target interactions [71]. Another inhibitor in clinical trials, Relacatib from GlaxoSmithKline, is in phase I clinical trials. Relacatib has a very high affinity for cathepsin K (K_i 41 pM) however unfortunately is not very selective for it shows similar affinity towards cathepsin L and V as well [72]. And finally, Odanacatib developed by Merck Frosst/Celera is the furthest along in phase III clinical trials. It is a nitrile – based cathepsin K inhibitor which is very potent towards cathepsin K and shows high selectivity when compared to Balicatib and Relacatib. So far, odanacatib has shown no major side effects [73].

	IC$_{50}$ (nM)
Balicatib	cat K = 1.4
	cat B = 4800
	cat L = 503
	cat S = 65,000

	K$_{i,app}$ (nM)
Relacatib	cat K = 0.041
	cat B = 13
	cat L = 0.068
	cat S = 1.6
	cat V = 0.063

	IC$_{50}$ (nM)
Odanacatib	cat K = 0.2
	cat B = 1034
	cat L = 2995
	cat S = 60
	cat V = 762
	cat F = 795
	cat C, H and Z > 10,000
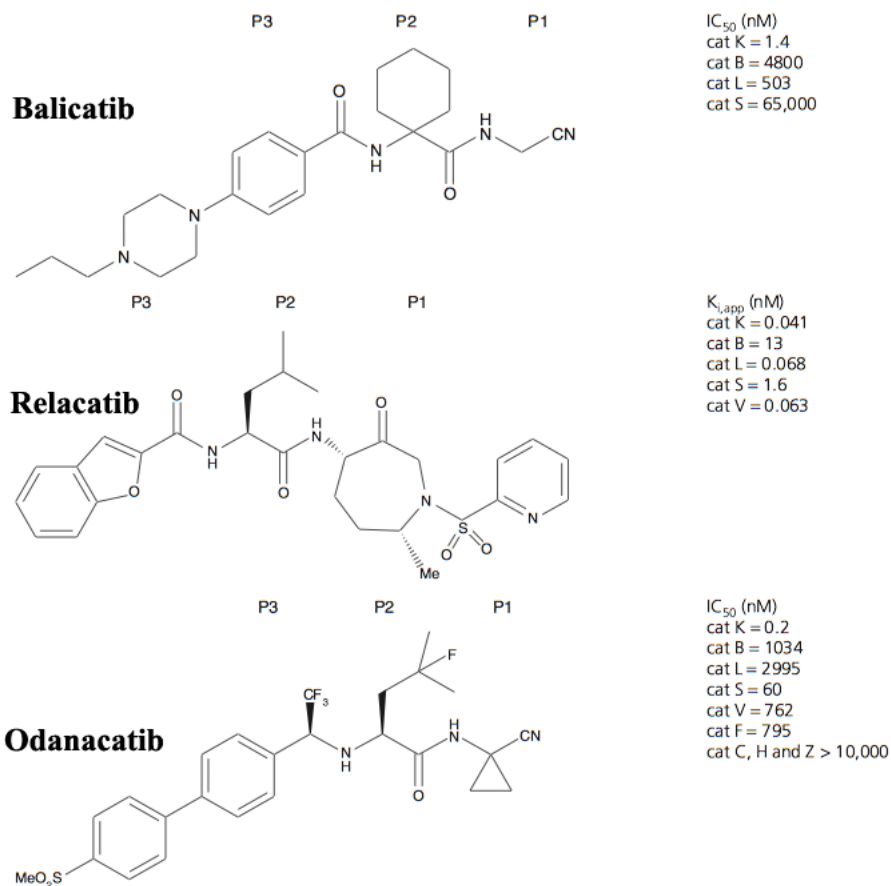
Figure 1.7 Inhibitors of Cathepsin K in Clinical Trials for the Treatment of Osteoporosis

Three inhibitors of cathepsin K are currently in clinical trials for the treatment of Osteoporosis; Balicatib (Novartis, Phase II), Relacatib (GlaxoSmithKline, Phase I) and Odanacatib (MerckFrosst/Celera, Phase III). Figure altered from Bromme et al 2009 [64].

1.10 Streptomyces as Natural Cathepsin Inhibitor Sources

Natural products play a highly significant role in drug discovery efforts. By the year 1990, 80% of the drugs on the market were of natural product origin, or an analogue whose structure was inspired by natural products. Natural products have revolutionized modern medicine with antibiotics (penicillin), antiparasitic (avermectin), antimalarials (antemisinin, quinine) cholesterol regulation (lovastatin), immunosuppresants

(cyclosporine) and anticancer drugs (taxol) [74]. *Streptomyces* are a leading producer of antibiotics and other classes of biologically active secondary metabolites. Secondary metabolites are low molecular weight organic molecules which are not essential for maintenance of cellular function or for normal growth of an organism [75]. Actinomycetes make up two-thirds of the known antibiotics that are produced by microorganisms, and amongst them nearly 80% are made by *Streptomyces*, the largest genus in the Actinomycetes order [76]. One example of a cathepsin K inhibitor isolated from Actinomyces is leupeptin as previously mentioned (section 1.9), which was first identified from *streptomyces* exfoliates, however has since been isolated in as many as 11 cultured bacteria filtrates [57, 77]. Other cathepsin K inhibitors have been isolated from *streptomyces*, alpha-MAPI isolated in no less than 24 independent strains [78]. Although none of these isolated natural products proved to be useful therapeutic agents towards the treatment of osteoporosis, they have proven to be valuable as protease inhibitors for laboratory research and are widely used in many inhibitor tablets by companies such as Sigma-Aldrich® and Roche [79]. In addition, even if a novel compound cannot be used as a therapeutic, there is the potential that the compound reveals some new insight into the substrate specificity of cathepsin K which could aid in the production of future drugs as was the case in the development of Odanacatib (Figure 1.7) [73].
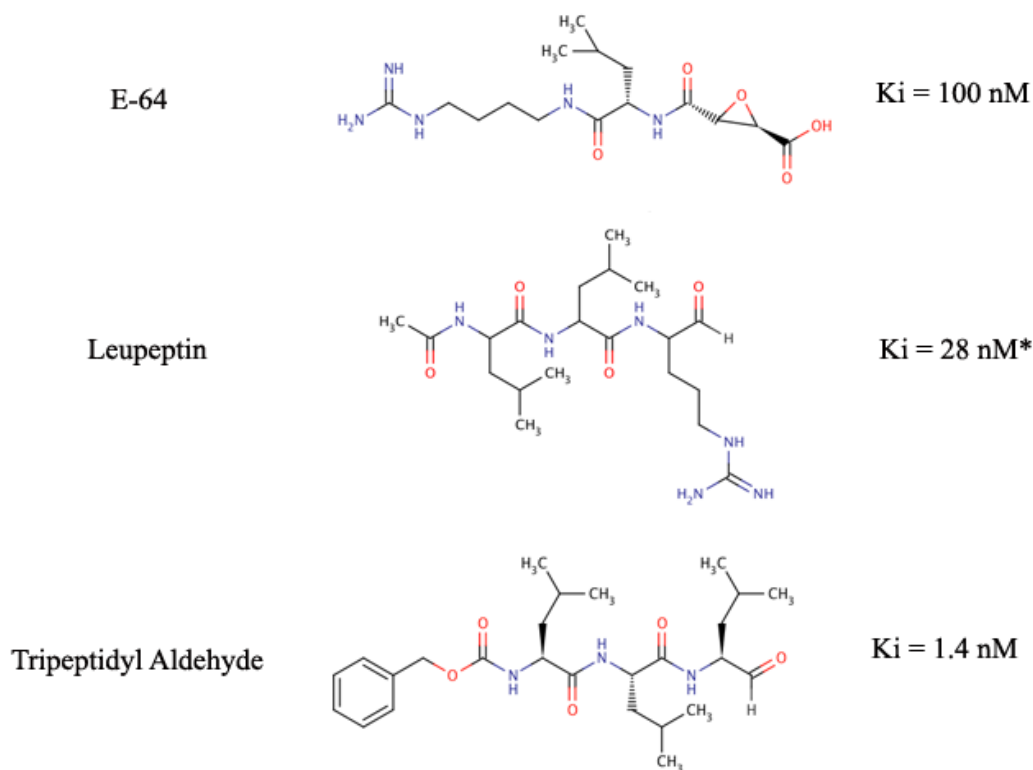
Figure 1.8 Structures of Known Inhibitors for Cathepsin K: E-64, Tripeptidyl Aldehydes and Leupeptin (*Ki Value Determined Experimentally).

The production of secondary metabolites is highly dependent on the growth phase of the *streptomyces*. In liquid cultures, the period of largest production is attained once the bacteria cultures enter the stationary phase. Physiological and environmental factors can also influence the production of secondary metabolites, as a result variations in temperature, nutrient supply, and physiological stresses can be effectors of antibiotic production [76].

Commercially available drugs, originating from natural product discoveries do not necessarily mean that the drug is an exact replica, it could also be semisynthetic derivative or fully synthetic compound based on natural products. The structural

variability of natural products is what makes them such likely targets, some structural features created by bacteria have yet to be replicated synthetically making the potential for a novel drug discovery[80]. By screening *streptomyces* in an enzymatic bioassay against cathepsin K, identifying novel inhibitors could lead to the next osteoporosis drug.

Due to the critical role cathepsin K plays in bone resorption, selective inhibitors of cathepsin K would be beneficial in the treatment of osteoporosis. The aim of this thesis is to isolate novel natural product based inhibitors against cathepsin K. In the following, I will describe the isolation of several novel cathepsin K inhibitors from the *streptomyces* strain, L-91-3, followed by their structural and kinetic characterization.

2 Materials and Methods

The following flowchart summarizes the fermentation and purification processes applied in this thesis:
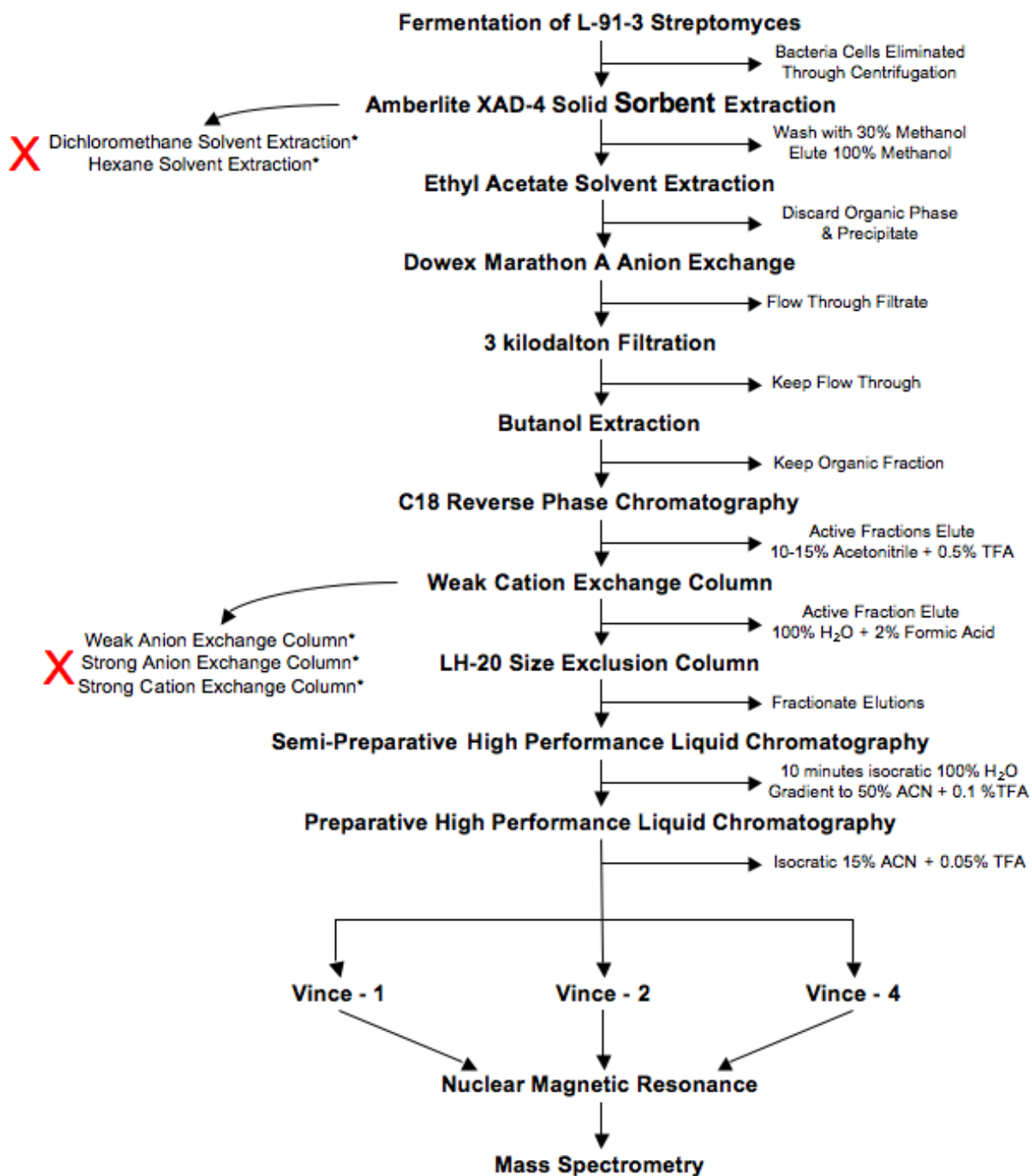
Figure 2.1 L-91-3 Active Compound Isolation Flow Chart

Stepwise purification of Active Compounds Vince 1, 2 and 4

2.1 Sample Screens

Soil bacteria *streptomyces* were collected by Dr. Julian Davies' laboratory in British Columbia, primarily from the lower mainland and Vancouver Island. In their bacteria library, Dr. Davies' laboratory has nearly 2000 strains of *streptomyces*. From this collection Jadwiga Kaleta performed a screening process, in which she pre-selected 384 strains based on growth characteristics as well as testing for inhibitory activity against cathepsin K using 96 well plates.

From the original 384, 15 strains showed strong inhibitory activity against cathepsin K. From this pool of 15 active strains, two were chosen; L-91-3 (to be purified by myself) and IS$_2$-4 (to be purified by Natasha Kruglyak). The human recombinant cathepsin K was produced by Raymond Pan, expressed in *Pichia pastoris* as previously described [25].

2.2 Enzymatic Activity Assay

Inhibitor activity was measured using the Spectra Max Gemini XS 96 well plate reader. The substrate Z-FR-MCA is incubated with cathepsin K at room temperature along with the potential inhibitors. Each reaction well received 5µl of 0.2mM Z-FR-MCA (BACHEM AMERICAS, Inc. Torrance, CA) (substrate) and 190µl sodium acetate buffer (2.5mM DTT, 2.5mM EDTA, 100mM Sodium Acetate, pH 5.5). 5µl of Cathepsin K 4nM (reaction concentration 0.1nM, K_m 7.4 +/- 0.9µM, K_{cat} 0.7 +/- 0.2s^{-1}) was added to each well making a total volume 200µl. The inhibitor assay was performed at room temperature.

Adding a sliding scale of dilutions of the inhibitor sample to each well, a master mix containing 5µl of 0.2mM Z-FR-MCA, 5µL of the inhibitor dilutions, 5µL of the enzyme and filled to a total volume of 200µL with sodium acetate buffer was added to each well. Fluorescence from the Z-FR-MCA cleavage was monitored at the excitation wavelength of 380nm and the emission wavelength of 460nm for 5 minutes [35]. Each concentration of inhibitor was loaded into two reaction wells in order to run duplicates.

In order to have a quantitative estimate of the inhibitor concentration, a 1:1 ratio of enzyme to inhibitor was assumed. Although this is highly unlikely for reversible inhibitors (typically a much larger ratio of inhibitor to enzyme), it allows for a determination of yield after each purification step.

2.3 Media Selection For L-91-3 Fermentation

All chemicals were purchased from Fisher Scientific (Fisher Scientific, Fair Lawn, NJ) unless otherwise stated. Due to complexity of the original fermentation media, there was too much material to effectively purify the active sample. Optimization of the starting media components was investigated trying to identify if a simpler media could be used in order to eliminate some of the contaminants from the source. The rich peptide media used for the production of L-91-3 contains 8 ingredients (Table 2.1), a few of which are necessary, while others are redundant nutrient sources. Experiments eliminating or decreasing non-essential media components were performed for L-91-3 production.

Glucose and Soytone (BECTON, DICKINSON AND COMPANY, Sparks, MD) could both be decreased in the broth, while (N-morpholino) propanesulfonic acid (MOPS), malt extract and yeast extract could be eliminated completely. L-91-3 strain of *streptomyces* were grown following the media recipes shown below and were tested for production of L-91-3 (in duplicate) as well as cell pellet production to identify the ideal starting conditions for future fermentations.

Table 2.1 Media Optimization Recipes

Media Component	A	B	C	D	E	F	G	H	I
Glucose (10g)			½	½				½	½
Glycerol (15ml)									
Soytone (15g)					X	½		½	½
NaCl (3g)									
Malt Extract (5g)		X	X					X	X
Yeast Extract (5g)	X		X					X	½
Tween (1ml)									
MOPS (20g)	X	X	X	X	X	X		X	X
Grams/Litre	33	33	23	33	23	31	58	15	20

2.4 Purification of L-91-3

2.4.1 Fermentation of L-91-3 Streptomyces

L-91-3 frozen cultures (in 20% glycerol) were thawed slowly in ice, streaked onto fresh BD Difco™ ISP Medium 4 (BECTON, DICKINSON AND COMPANY, Sparks, MD) plates then placed into a 30°C incubator. After 7 to 10 days, colonies had begun to grow on the plates. Individual colonies were picked using a metallic loop and placed into 300ml (in 1L preparatory flask) rich media previously autoclaved and cooled containing: 0.5% D Glucose, 1.5% Glycerol, 1.5% Soya Peptone, 0.3% NaCl, 0.1% Tween 20 and 2% (N-morpholino) propanesulfonic acid (MOPS).

When preparing the ISP-4 plates, special attention was taken to ensure that all the powder was completely dissolved in the solution prior to autoclaving. To accomplish this, water was first boiled and ISP-4 powder was added (37g/L) mixing continuously. If not fully dissolved, it was microwaved and repeated. Once the powder was fully dissolved it was autoclaved immediately and poured onto plates as soon as possible, with the addition of 25ml of ISP-4 medium to each Petri dish, then allowed to set.

Once the growth media had been seeded, it was placed in the 30°C room and mixed on a shaker at 200rpm. 1ml samples were taken using sterile technique, spun down in micro centrifuge and tested for inhibitory activity against cathepsin K. Once production of L-91-3 had reached a maximum, the flask was brought to Sung-Hye Grieco (Fermention suite/Center for Blood Research) and used to inoculate larger scale bio-reactors (5L or 7L) containing the same media as the preparatory flasks (Figure 2.3).

L-91-3 *streptomyces* is grown in a bio-reactor kept at a controlled environment of 30°C, pH 7 using HCl (2M) and NH_4OH (1M) and O_2 (20%) while between a range of 500 to 700 rpm depending on viscosity due to bacteria growth. The growth media samples were taken every few hours and tested for activity via cleavage inhibition of Z-FR-MCA by cathepsin K to identify maximum production of the inhibitor (typically between 38 to 48 hours) along with ensuring there was no contamination by viewing cell pellet under a microscope. Once the activity gross value of the inhibitor started to decline, the full contents of the bio-reactor were collected and the sample was harvested.

Project name:				Vessel size (L):	7.0
BROM	B	7	F7	Impeller:	3 x 6 blades
EXP file name:				Sparger:	Bacterial
Suite:	08/05/06/F7			Working volume (L):	5.0

		Temp (oC)			dO2						pH	
				Air	O2	rpm		N2	Base NH4OH			Acid HCl
Before induction	HJ	30.0	Water	0.500	2.800	600	>20%	0.000	1.0	7.0		2.0
Configuration	1	SP	4	-	5	-	SP	-	2	SP		3
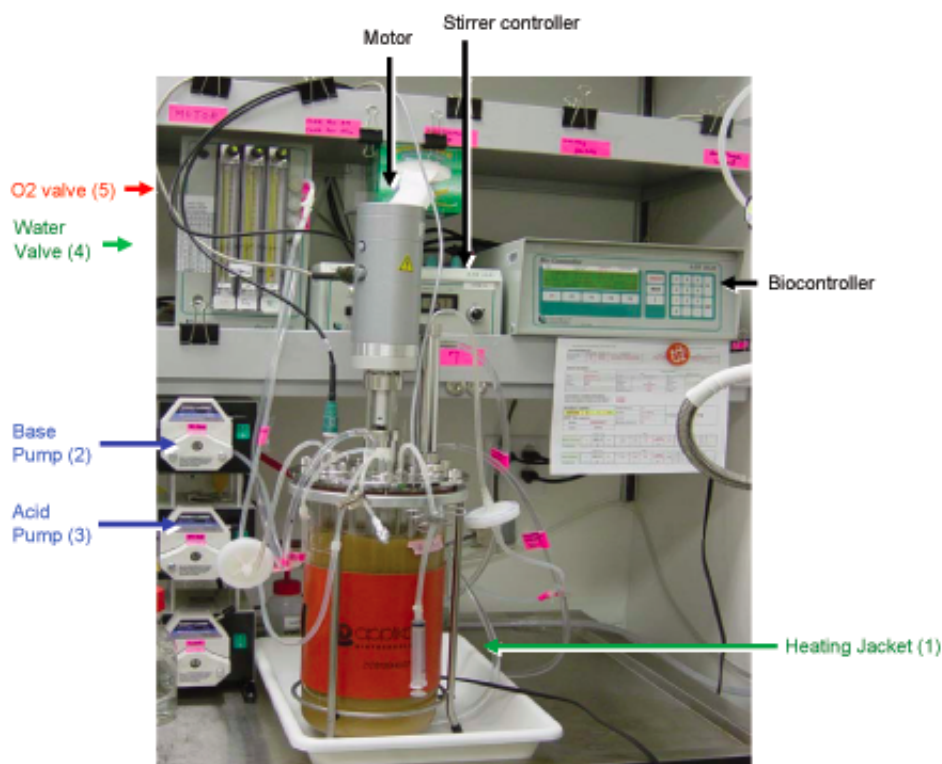
Figure 2.2 Fermentation Bioreactor Setup

The Bioreactor used to grow *Streptomyces* keeps a stable environment for the bacteria by controlling temperature, pH, and O_2 concentration while mixing.

2.4.2 Harvest of L-91-3 Media

Once the inhibitory activity of L-91-3 had reached its maximum, the media was harvested and the bacteria pellets were removed. To accomplish this, the broth was centrifuged at 18,600g for 10 minutes in the Beckman Coulter Anvanti® J-E Centrifuge (BECKMAN

COULTER, Brea, CA) using the JLA – 10.500 rotor. Once the bacteria were spun down, the media containing L-91-3 was decanted and filtered through a 11μm grade #1 Whatman® filter paper (Cat No 1001 110) (WHATMAN INTERNATIONAL LTD, Maidstone, England) by vacuum to remove any suspended bacteria.

Prior to discarding, bacteria pellets were tested for inhibitor activity using two different bacterial cell wall lysis methods. The first method washed bacteria cell pellet in an 80:20 Methanol/Water solution and the second method used an equal volume acetonitrile/water mixture. Upon testing the bacterial cell lysate, no substantial inhibitory activity was identified for either lysis method and it was concluded that the inhibitors in L-91-3 was released into the media rather than being contained in the bacteria cells.

2.4.3 Amberlite Solid Sorbent Extraction

Following filtration, the media was treated through solid phase adsorption in the first step of the purification. 50g of Amberlite® XAD4 (SIGMA-ALDRICH, St. Louis, MO), a nonionic macro-reticular resin that adsorbs and releases ionic species through hydrophobic and polar interactions, was added to each liter of media and mixed gently for 24 hours at 4°C.

After 24 hours, the activity in the supernatant was tested. If activity was still present an additional 30g per liter Amberlite® XAD4 was added and mixed for an additional 6 hours.

Once the L-91-3 inhibitors had been removed from the media by Amberlite® XAD4, the supernatant was decanted off and the beads were rinsed thoroughly with H_2O. The beads were then treated with the first wash step using 1/5 harvest volume with 30% methanol and mixed gently for two hours. This wash step was repeated 3 more times prior to sample elution using 100% methanol.

To elute with methanol, 2:1 (volume : weight) 100% methanol [ex 500ml methanol: 250g beads] was added to the beads and washed through gentle stirring for 2 hours. Methanol was filtered via vacuum filtration using Whatman® filter paper and beads were recovered for subsequent washes. The activity of the filtered wash was tested and beads were subsequently washed until no substantial L-91-3 inhibitory activity (less than 5% of starting activity) was recovered. [1]

In order to optimize the Amberlite® step, an attempt was made to create a stepwise gradient of methanol following the wash step. The Amberlite® resin after the 30% wash

---

[1] Alternatively, additional Amberlite® XAD resins (SIGMA-ALDRICH, St. Louis, MO) were tested, namely XAD7 and XAD16. XAD7 and 16 both have larger pore size than XAD4 and although they were both capable of binding L-91-3, each had their drawbacks. Elution off Amberlite® XAD7 proved difficult and resulted in a significant loss of L-91-3 sample when comparing to original harvest yield. Amberlite® XAD16 simply eluted too much during the wash steps. Both contaminants and L-91-3 eluted off the resin resulting in a less efficient clean up step than XAD4.

step was loaded into a large glass column. L-91-3 inhibitors were eluted from the resin using 10% steps in the gradient from 40% to 100%. As was expected, activity started to elute immediately. Unfortunately this didn't prove to be a useful technique because with each increase in methanol concentration, more inhibitors would elute resulting in a large smear of activity.

Due to the unsatisfactory results from the gradient elution, one single elution was performed with methanol. The 100% methanol elution was then roto-evaporated down at 40°C. Once volume reached 20% harvest volume, 10% (of harvest volume) H_2O was added. The remaining methanol could then be evaporated off without fear of precipitate forming. If methanol was concentrated too much (less than 10% harvest volume) without the addition of H_2O, a precipitate would form on the round bottom flask which was impossible to redissolve without destroying the activity of the L-91-3.

2.4.4 Ethyl Acetate Extraction

Once the L-91-3 sample had been concentrated down to 10% original harvest volume in H_2O the first liquid extractions were performed. The sample was mixed in a 1:1 ratio with ethyl acetate. The sample was stirred energetically for 2 hours in the cold room (4°C). After two hours the sample was loaded into a large separatory funnel and the two immiscible liquids separated leaving the aqueous phase containing the active compound on the bottom. Caution was taken to separate the two phases, as a precipitate forms at the interface which must also be discarded. Due to the fact that ethyl acetate has a small

solubility with H_2O (8.3g / 100ml), the ethyl acetate must be centrifuged in 50ml falcon tubes in the IEC Centra CL3R centrifuge at 3200g for 5 minutes to further separate the organic from aqueous phase.

After centrifugation, the aqueous phase was placed on a rotoevaporator (Heidolph Laborota 4001) to completely remove solubilized ethyl acetate. At this point the sample could also be concentrated down to 75% of starting concentration at which point the whole ethyl acetate extraction process can be repeated once again at a 1:1 ratio.

Following the 4 to 5 repetitions, there was no further extraction taking place and precipitate was no longer forming at the interface. Therefore the aqueous phase could be concentrated down to 5% harvest volume in preparation for Dowex Marathon A (SIGMA-ALDRICH. INC., St. Louis, MO) treatment[2].

2.4.5 Dowex Ion Exchange Resins

Once all ethyl acetate had been removed from the sample, it was then treated with an ion exchange resin. The strong base anion exchange resin DOWEX® marathon A (SIGMA-ALDRICH, St. Louis, MO) was added at 30g/L harvest. DOWEX® beads were first washed using H_2O. Due to the smell of the ammonia functional groups, all processes involving DOWEX® were performed in the fume hood. The L-91-3 sample was mixed

---

[2] Solvent extractions using both hexane and dichloromethane were tested, however, in the final purification scheme, neither technique were used due to their inability to purify the L-91-3 media sample any further.

with DOWEX® marathon A using a magnetic stirrer gently for 1 hour. After an hour, the solution was filtered and DOWEX Beads were washed with H_2O. The filtered sample containing L-91-3 inhibitors was then concentrated further to a final volume approaching 1% of the harvest volume.

DOWEX® Marathon C, a strong acid cation exchange was used for the purification process applying the same method described above prior to the concentration of the sample down to 1% of the harvest volume.

2.4.6 Filtration

After filtering through Whatman #1 grade filter paper, the sample was passed through a 3 kDa filter. The inhibitor passed through a Millipore Amicon® centrifugal filter leaving larger macromolecules on the filter top. The filters first must be washed with distilled water. Once the sample was loaded onto the filters, they were spun at 3200g using a IEC Centra CL3R centrifuge (Thermo IEC, NEEDHAM HEIGHTS, MA) until 70% of the solution has passed through. The sample was then diluted and spun again paying special attention not to allow the membrane to spin dry.

2.4.7 Butanol Extraction

Once the L-91-3 sample had been concentrated down post DOWEX® Marathon A treatment and filtered, another solvent extraction was performed. Due to the solubility of

n-butanol in water, along with the difficulty in evaporating it off, extensive extractions were not favorable. As a result only 3 extractions of the 2:1 (Sample : Butanol) were performed. Solutions were mixed using a stir bar, mixing vigorously for 90 minutes then separated using a separatory funnel. After each extraction, the butanol phase was spun at 3200g for 20 minutes in an attempt to separate any soluble H_2O which could be returned to the aqueous phase of the extractions.

Once butanol extractions were complete, the L-91-3 sample moved into the organic phase. To prepare the sample for the subsequent purification step, the sample is lyophilized to remove butanol and re dissolved in a 20% methanol.

2.4.8 Phenomenex C18 Reverse Phase Chromatography

Phenomenex® Sep Pak C18 cartridges (PHENOMENEX. INC., Torrance, CA) were the first stage of reverse phase chromatography. Available in 10g disposable cartridges, they do not have a large enough binding capacity for the entire sample. To accommodate for this, polycarbonate columns were used to pack multiple cartridges (1 cartridge / 2L Harvest) at once. The column was conditioned using HPLC grade acetonitrile (ACN) (50ml / cartridge) with 0.1% trifluoroacetic acid (TFA). Following conditioning, H_2O (50ml / cartridge) was run through the column to re-equilibrate the resin prior to loading of the sample.

The dry sample was dissolved in a minimum volume of 20% methanol (less than 3ml) and loaded onto the C18 column. The column was washed using 100% H_2O (50ml / cartridge) followed by an equal volume of H_2O with 0.1% TFA. The sample was eluted off the column using a stepwise gradient of ACN with 0.1% TFA. Using the same volume of eluant as was used for the washes (50ml / cartridge) starting with 5% ACN, the gradient was increased by 5% increments until all the L-91-3 sample had eluted. To optimize this procedure, a fraction collector was used to collect the sample in 13ml fractions which were later tested for activity discarding the non active fractions.

2.4.9 Strong and Weak Ion Exchange Columns

Another approach to purifying using ion exchange was through the use of columns. Four mixed - mode ion exchange columns were tested from Phenomenex, Weak Cation Exchange (WCX) (PHENOMENEX. INC., Torrance, CA), Strong Cation Exchange (MCX) (PHENOMENEX. INC., Torrance, CA), Weak Anion Exchange (WAX) (PHENOMENEX. INC., Torrance, CA), and Strong Anion Exchange (MAX) (PHENOMENEX. INC., Torrance, CA).

In weak ion exchanges it is the charge on the resin which changes during conditioning and elution steps, where as for strong ion exchanges it is the charge on the bound compound that are neutralized Similar to the Phenomenex® Sep Pak C18 columns, the ion exchange columns were broken up and placed into a larger column. All four types of ion exchange columns were available in 2g columns, using 1 column / 2L harvest.

For MAX / WCX columns, conditioning began using methanol (10ml / cartridge) followed by equilibration with an equal volume of H_2O prior to loading the sample (Figure 2.4). The sample was washed using an equal volume (10ml / cartridge) 5% Ammonia Hydroxide (NH_4OH) followed by a wash using 100% methanol. Elution was performed using a stepwise gradient of methanol (20% increments) containing 2% formic Acid.

The MCX / WAX columns were conditioned using methanol (10ml / cartridge) followed by equilibration with an equal volume of H_2O prior to loading the sample. The sample was washed using an equal volume (10ml / cartridge) 2% Formic Acid followed by a wash of 100% methanol. Elution was performed using a stepwise gradient of methanol (10% increments) containing 5% Ammonium Hydroxide. In the final, purification, only the WCX column was used since neither the WAX nor MAX column provided any cleanup, and the MCX column bound the inhibitor irreversibly. WCX provided the greatest purification of the L-91-3 sample when analyzed by mass and using HPLC.

MAX/WCX Column

Condition with Methanol
↓
Re-Equilibrate with Water
↓
Load Sample
↓
5% NH_4OH Wash
↓
Elute with 0-100% MeOH and 2% Formic Acid

MCX/WAX Column

Condition with Methanol
↓
Re-Equilibrate with Water
↓
Load Sample
↓
2% Formic Acid Wash
↓
Elute with 0-100% MeOH and 5% NH_4OH
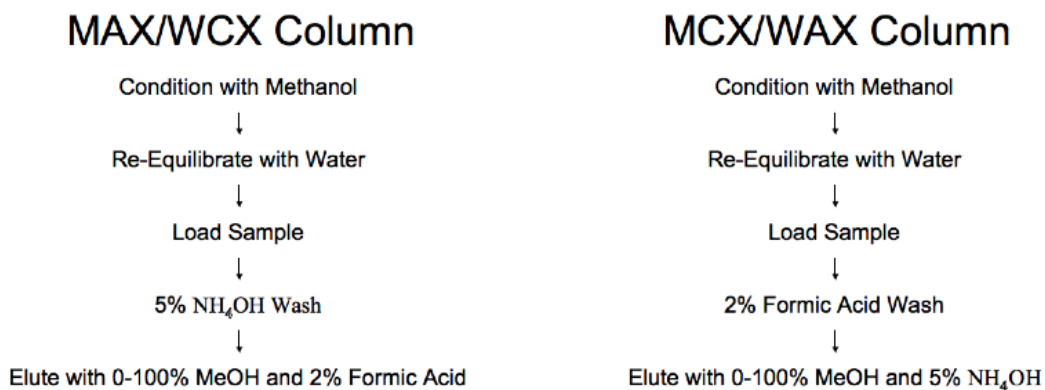
Figure 2.3 Anion and Cation Exchange Column Conditioning and Elution

2.4.10 Sephadex^{TM}LH-20 Size Exclusion Column

The final cleanup prior to HPLC purification utilized Sephadex^{TM}LH-20 resin (GE HEALTHCARE, PISCATAWAY, NJ). LH-20 was used as a size exclusion gel filtration for natural products. A critical phase for LH-20 column use is the preparation and packing steps. The resin was kept as slurry and left to swell in solvent for at least 24 hours to allow the resin to equilibrate with the solvent. The slurry was then poured in one continuous motion into a large column filled 1/5 with the solvent used to run the sample to avoid introduction of bubbles. To further avoid bubbles, it was necessary to allow the slurry to slide down a glass rod or the side of the column.

Once the column was loaded, it must be given adequate time to settle. The column was left to gravity for 48 hours to completely settle the beads, followed by running 5L of solvent through the column to complete the packing process. Due to the high solubility of gases in methanol, the column is loaded and flushed with degassed 50/50 methanol / water mixture.

The L-91-3 sample was loaded onto the LH-20 column in as small a volume as possible (1 ml per liter harvest). Acid can sometimes be added to the solvent to aid in the elution; however, this was not necessary for L-91-3 purification.

By means of gravity, the LH-20 column elution was collected using a fraction collector into 13ml fractions which were later tested for activity. Active fractions were collected, concentrated and lyophilized in preparation of HPLC purification.

2.4.11 Semi-Preparative High Performance Liquid Chromatography

The first round of High Performance Liquid Chromatography was performed on a C18 semi-preparative column from Phenomenex using a System Gold HPLC from Beckman (BECKMAN COULTER CANADA, Inc. Mississauga, ON). The column was equilibrated with 0.1% TFA in H_2O at a flow rate of 3ml/min. 5mg of the L-91-3 sample was loaded onto the column dissolved into 50μL of H_2O. The elution profile for purification began with 10 minutes isocratic flow at 100% H_2O. After 10 minutes, a steady gradient ran for 30 minutes from 0-50% ACN still with 0.1% TFA. Fractions were collected in 1 minute intervals, tested for activity and split up depending on HPLC spectra profile and activity. From here, the most active fraction (#28 figure 3.8) containing a grouping of peaks was collected and further processed by HPLC on a second column in Dr Raymond Andersen's laboratory with the help of Dr. David Williams. Additionally, a portion of the sample brought to Dr Raymond Anderseren's laboratory (Department of Earth and Oceanic Sciences) was taken for X-ray crystallography studies with the help of Dr. Adeleke Aguda in cooperation with Dr. Gary Brayer's laboratory (Department of Biochemistry and Molecular Biology).

2.5 Structural Analysis of L-91-3 Samples

2.5.1 Preparative High Performance Liquid Chromatography Purification

Following C18 Semi-preparative HPLC purification, the active sample was dried down and dissolved once again in 20% methanol in water and purified with the help of Dr David Williams (Dr. Raymond Andersen Laboratory). Using a different HPLC system, the 600E system matched with a 486 absorbance detector from Waters®, an Intersil C18 column (WATERS CORPORATION, Milford, MA) was used to purify Fraction #1 of L-91-3. The sample was loaded onto the column once again using 5mg per run.

Running at a flow rate of 2ml/min the active sample from the first round of HPLC (*2.15*) was run using an isocratic profile at 15% ACN and 0.05% TFA. Peaks reaching baseline were separately collected into round bottom flasks, concentrated to dryness and analyzed by nuclear magnetic resonance (NMR) and mass spectrometry.

2.5.2 Nuclear Magnetic Resonance and Mass Spectrometry

Once samples had been fully purified to individual compounds, their structures were determined by Dr. David Williams (Dr. Raymond Andersen Laboratory) through NMR analysis and correlating this with the mass spectrometry data. NMR analysis was performed using proton and carbon NMR using the Bruker Advance 600 MHz with CRYOPROBE while low and high resoluion ESI-QIT-MS were recorded on a Bruker-Hewlett Packard 1100 Esquire-LC system mass spectrometer.

2.5.3 X – Ray Crystallography

X-ray crystallography was performed by Dr. Adeleke Aguda. An L-91-3 sample purified from semi-preparative HPLC was supplied to Dr. Aguda. Cathepsin K and L-91-3 were mixed prior to crystallization and were successfully co-crystallized at room temperature using the sitting drop vapor diffusion method. The reservoir solution contained 24% PEG 8000, 0.2 M ammonium sulfate at pH 6.5 and the sitting drops consisted of 5 µl of protein ligand mix solution (4.6 mg/ml) mixed with 5 µl of reservoir solution. Diffraction quality crystals appeared over the period of a week.

The co-crystallized Cathepsin K - L-91-3 complex diffraction data was collected on an ADSC Quantum 315R imaging plate area detector system on beam line 7-1 at the Stanford Synchrotron Lightsource. A total of 150 data frames were collected at 0.8° oscillation and a wavelength of 0.9795 to 2.0 Å resolution. All data frames were exposed for 5 seconds.

2.5.4 Ultra High Performance Liquid Chromatography (CDRD)

The sample (L-91-3 sample purified from semi-preparative HPLC) submitted for X-ray crystallography was also run on Ultra High Performance Liquid Chromatography matched with mass spectrometry at the Centre for Drug Research and Development by Dr. Adam Galey with the help of Dr. Markus Heller. To accomplish this, 0.1mg of material in 5µL of 20% methanol with 0.1% TFA was loaded onto a Waters UPLC system equipped with 2767 Sample Manager from Waters. The column was conditioned

with 10% ACN and 0.1% TFA then the sample was loaded and run on a gradient to 25% ACN and 0.1% TFA over 20 minutes controlled by a 2545 Binary Gradient Module from Waters. Data was collected using Waters 2998 Photodiode Array Detector and a 3100 Mass Detector using positive electrospray ionization.

2.6 Sample Kinetics, Ki Determination

Kinetics experiments were performed using the Spectra Max Gemini XS plate reader. The HPLC purified and dried sample was dissolved in 100μL dimethyl sulfoxide (DMSO). From here dilutions were made using H_2O to identify a range of active compounds where activity readings ranged between 10% to 70% inhibition using the same plate reader procedure described in Section 2.2.

Once the inhibitor concentration range had been established, kinetic analysis was performed by screening the activity of the inhibitor at multiple substrate concentrations (2 to 8 μMol). Each substrate concentration was repeated three times.

By plotting results for each substrate concentration, 1/V (1/reaction rate) [Y- axis] vs. inhibitor concentration [X-axis] (Dixon plot), a linear line should result. Provided the compound reacts as a typical competitive inhibitor, the trend lines of each substrate concentration will intercept at X = -Ki. The Ki values were determined by taking the average of all of the intercepts from each plot (3 plots).

3 Results

3.1 L-91-3 Identification

L-91-3 along with a few other samples (IS$_2$-4, L-85-1, L-90-1) was identified as having strong inhibitory action for cathepsin K based on sample screens performed by Jadwiga Kaleta. L-91-3 was selected as the most potent cathepsin K inhibitor. After selecting L-91-3 strain of *streptomyces* to purify, 0.3-liter cultures were grown and used to inoculate two 5L bio-reactors containing a simplified peptide media.

3.2 Media Optimization

After repeated difficulties with purification, the decision was made to try and simplify the process from the source. The media had many complex ingredients so attempts were made to eliminate a few components hoping to diminish the contaminants from the source (Table 2.1).

Eliminating half of the glucose, as well as, all of the malt and yeast extract showed no decrease in inhibitor production (Figure 3.1). Additionally, since the fermentation bioreactor had pH control, when doing the large scale fermentation only half of the buffer MOPS was required to maintain a stable pH. By making these minor changes, the total inhibitor production was not changed yet it managed to eliminate 35 grams per liter of material.

Inhibition of Cathepsin K Activity at 1000X Dilution
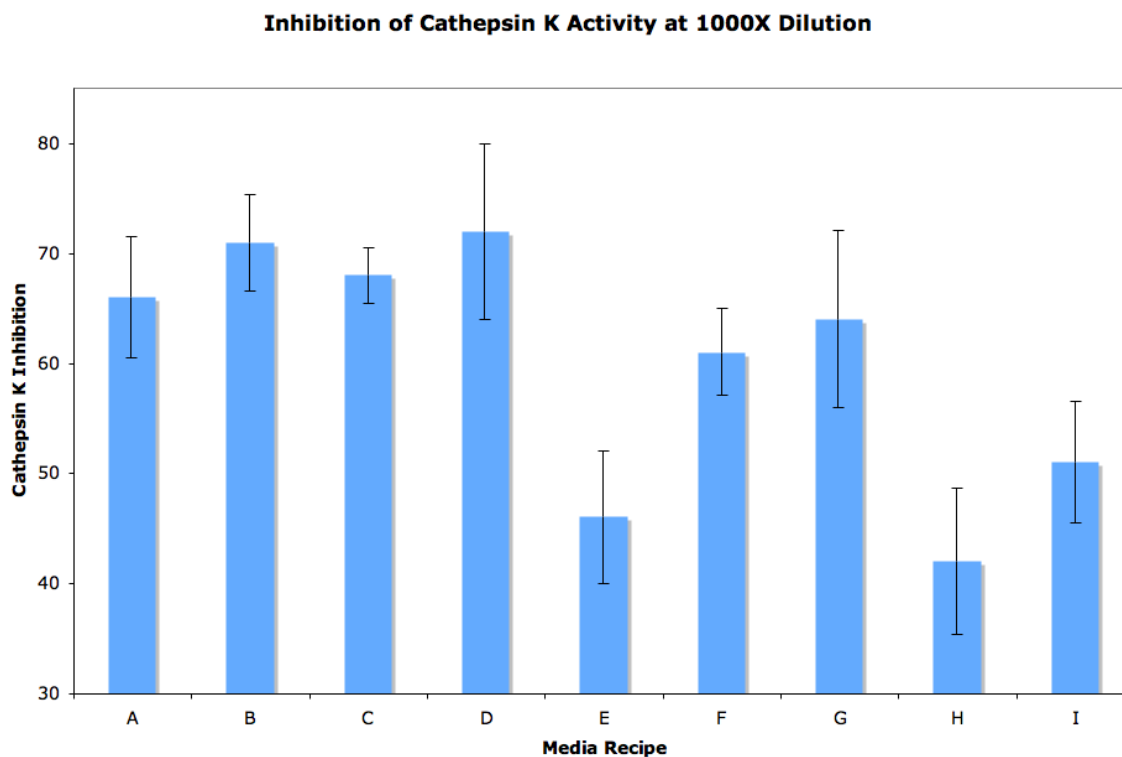
Figure 3.1 Inhibitory Activity of L-91-3 at 1000 Fold Dilution Based on Inoculation Media Recipe

By eliminating yeast and malt extract (A-D), there was no effect on inhibitor production, however, soytone proved to be an essential ingredient (E), letters corresponding to media recipes listed in table 2.1. Enzyme activity assay was run in duplicate.

The second factor in selecting the media was the ability to maintain production of the cell mass. The malt and yeast extract proved unimportant in the fermentation, as eliminating these two ingredients had little effect on grams fresh weight (gFW) of cell pellet (Figure 3.2). All four samples A to D (for recipe see Table 2.1) produced roughly the same mass of cell pellet to the control media G with none exceeding the controls.
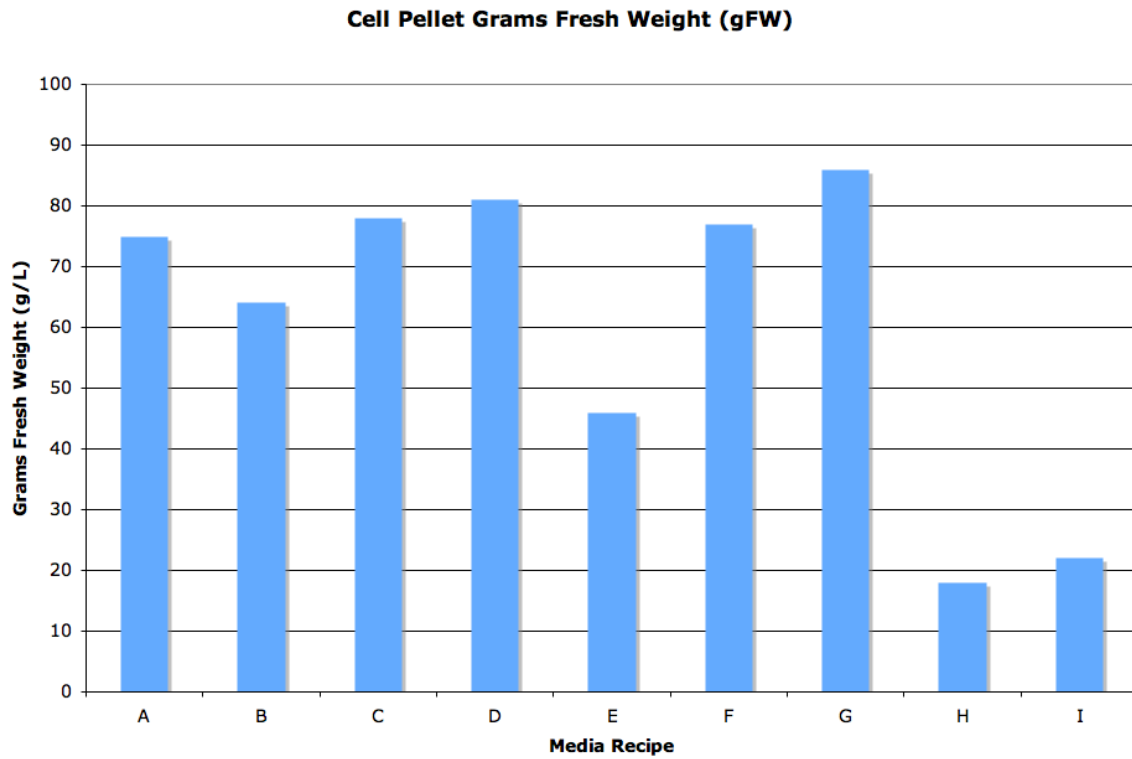
Figure 3.2 Grams Fresh Weight of L-91-3 Bacteria Cell Pellet

Letters corresponding to media recipes on Table 2.1.

3.3 Fermentation
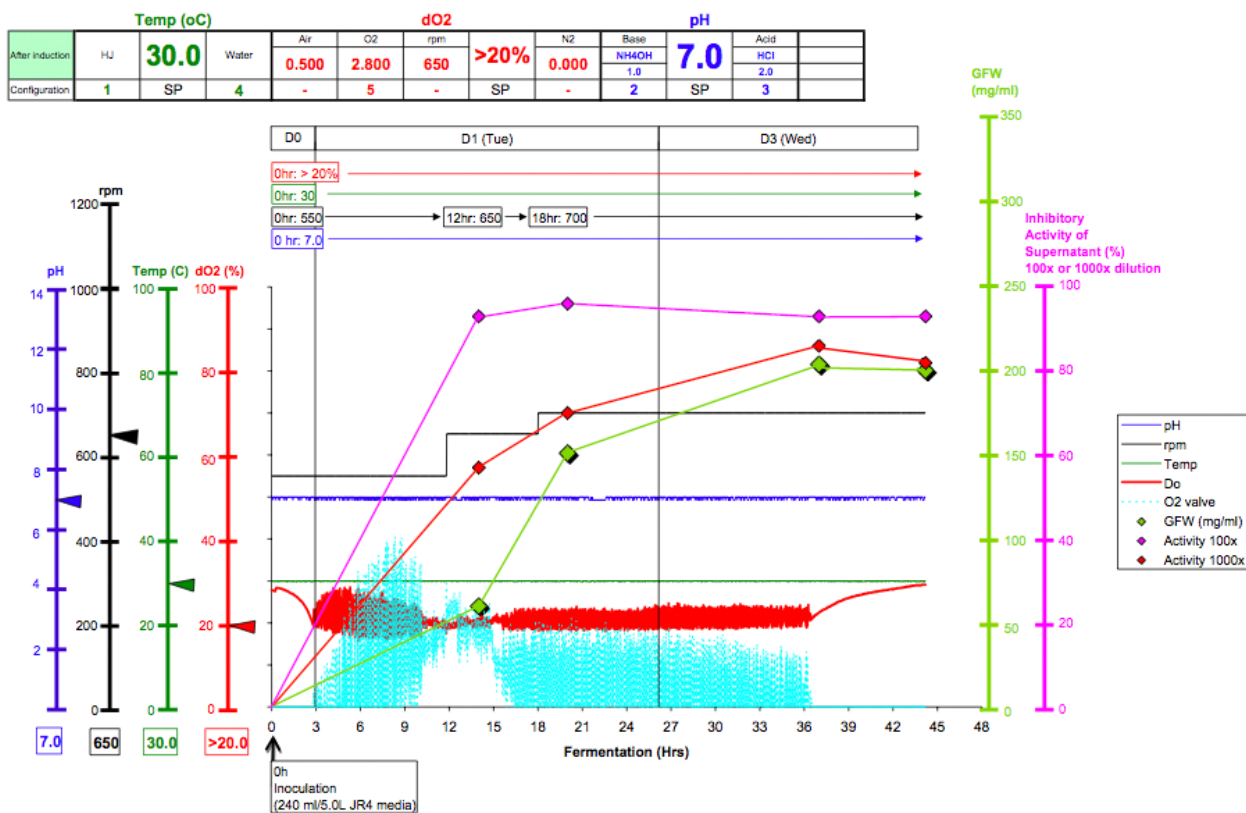
Figure 3.3 Fermentation Profile for L-91-3 Strain of *Streptomyces*

7-liter bioreactors containing 5 liters of media recipe C were used to grow the strain of *streptomyces* L-91-3. Cathepsin K activity inhibition at 100 and 1000 times dilution as well as grams fresh weight were measured at specific time points while temperature (30°C green), oxygen (20% dO2, red) and pH (7, blue) were kept constant while being aerated by mixing (600-700 rpm black). Inhibitor production peaked at 39 hours and was harvested at 45 hours. Figure prepared by Dr. Sung-Hye Grieco, The Center for Blood Research Fermentation Suite, University of British Columbia.

Maximum cell growth occurred after 39 hours of incubation reaching maximum inhibitory activity (80% inhibition at 1000 fold dilution) (Figure 3.3) at that time. Harvest of media yielded 20 µMol of inhibitor from L-91-3 contained in 84g of solid material.

3.4 Purification of L-91-3

3.4.1 Amberlite® Treatment

During the gradient elution with methanol of Amberlite XAD4 the majority of the inhibitory activity was found in the 40% and 50% methanol fractions (Figure 3.4). All four fractions from 40 to 70% when tested in duplicate for cathepsin K inhibitory activity contained some activity and as such each fraction was independently further purified. The initial results looked promising as the 50% methanol elution had a cleaner HPLC profile to that of the batch 100% methanol bulk elution. However, in subsequent purification steps (ethyl acetate extraction and Dowex Marathon A) not nearly as much contaminant was eliminated. Following the ethyl acetate extraction, both the 100% and 50% elutions looked similar on HPLC suggesting the gradient elution was no better for purifying the sample since it resulted in significantly lower yield.

Column Elution of L-91-3 from Amberlite Resin
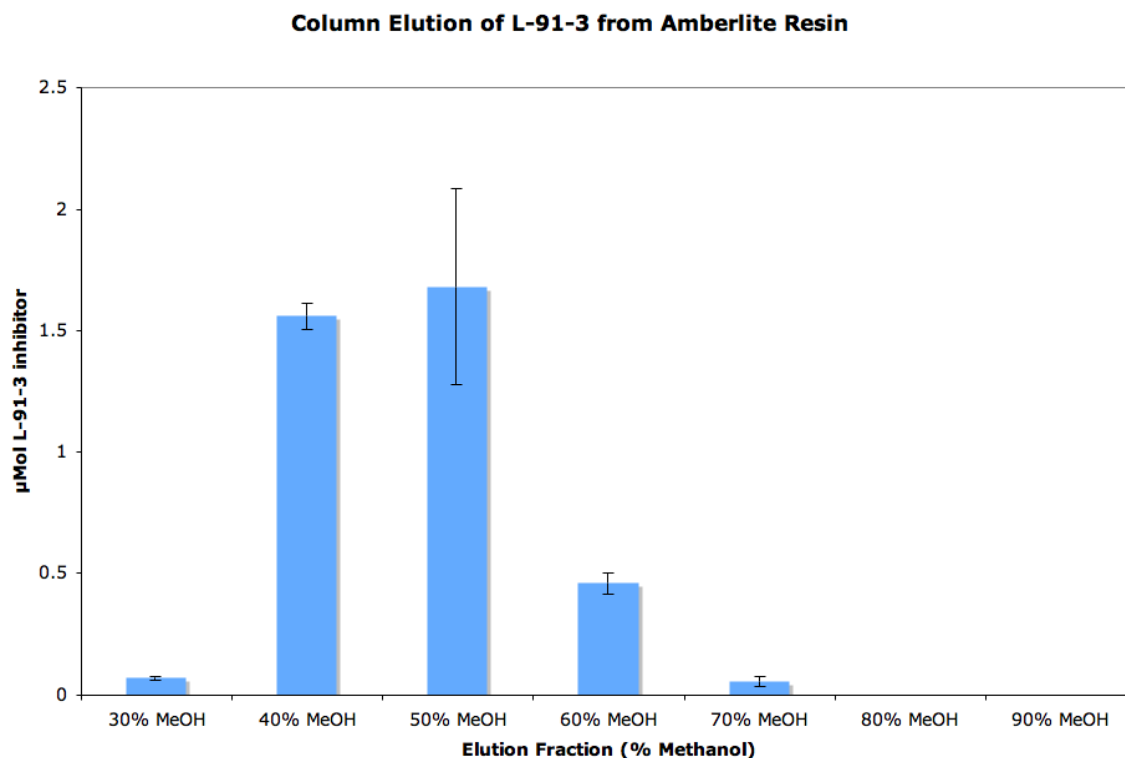
Figure 3.4 Gradient Elution of Amberlite XAD - 4 with Methanol

Based on the results from the gradient elution off Amberlite®, the sample was then eluted using a bulk elution with 100% methanol. Following elution, the sample contained 8.8 grams of material, having eliminated nearly 90% of contaminants, and the full 20 µMol of inhibitors from L-91-3 was recovered (cathepsin K inhibitory activity assay performed in duplicate).

3.4.2 Organic Solvent Extractions

Following vigorous mixing of ethyl acetate with the aqueous sample, a white precipitate formed at the interface of the two solvents. Upon testing, the precipitate as well as the ethyl acetate layer were found to be inactive and were discarded. A further 7.3 grams of contaminants were eliminated leaving 1.5 grams of material while retaining 15µMol of L-91-3. Sample showed significant improvement on HPLC (Figure 3.5).
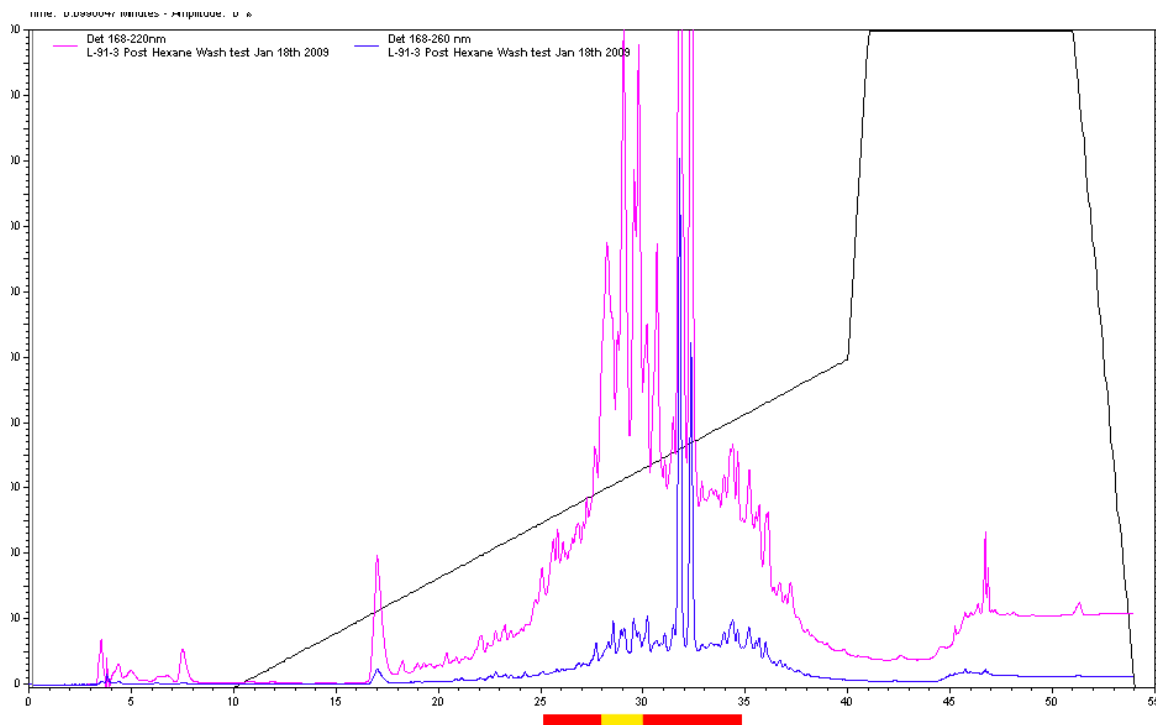
Figure 3.5 Analytical High Performance Liquid Chromatography - Post Solvent Extraction

Following the first round of solvent extractions (Ethyl Acetate, Hexane, Dichloromethane) L-91-3 sample was run on analytical HPLC at an isocratic elution of H_2O with 0.1% TFA for 10 minutes followed by a 30 minute gradient to 50% ACN with 0.1% TFA. Absorbance was measured at two wavelengths, 220 nm (blue) and 260nm (pink). Active fractions against cathepsin K were found between 25-35 minutes (red, with the most active fractions highlighted in yellow).

3.4.3 Dowex Marathon® A Ion Exchange Resin

Dowex Marathon® A ion exchange resin bound the sample if left mixing overnight. As a result, short treatments were necessary. After 1 hour of mixing much of the dark brown color of the L-91-3 sample solution had been removed. While not a significant amount of impurities were eliminated when analyzing by weight or HPLC analysis, the sample became much easier to work with following Dowex treatment. In total, 0.2g of material

was eliminated and 5 μMol of inhibitor from L-91-3 retained. Although this resulted in a net loss of material compared to the purification efficiency, Dowex treatment was found to be essential in making the sample manageable for column chromatography downstream in the purification.

3.4.4 Filtration

When L-91-3 sample was placed on the Amicon 3K filter, the active compound passed through the membrane completely leaving a minor residue of impurities on the filter top.

3.4.5 Butanol Extraction

Three butanol extractions were performed each with 50% sample transfer per extraction into the butanol phase. By the 3rd extraction, no active compound remained in the aqueous phase, having eliminated another 0.3g of contaminant and retaining the entire active compound.

3.4.6 Reverse Phase C18

L-91-3 was collected using a fraction collector following elution from the Phenomenex C18 column. All of the active compounds in L-91-3 bound to the column and were subsequently eluted using 5% increases in ACN concentration with 0.5% TFA. The sample started eluting at 10%, however, the bulk of the material eluted once the column

conditions reached 15% ACN. A marked improvement of purity could be seen on analytical HPLC (Figure 3.6).
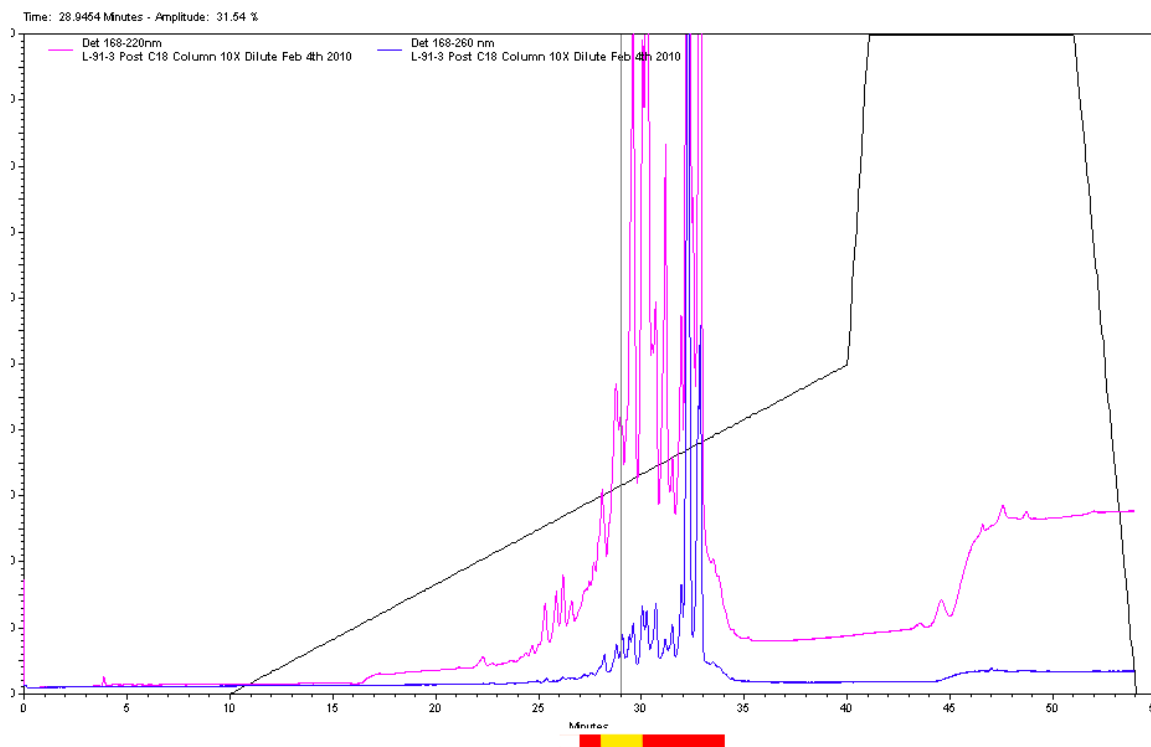
Figure 3.6 Analytical High Performance Liquid Chromatography - Post Reverse Phase Column

Following Phenomenex C18 reverse phase column chromatography L-91-3 sample was run on analytical HPLC at an isocratic elution of H_2O with 0.1% TFA for 10 minutes followed by a 30 minute gradient to 50% ACN with 0.1% TFA. Absorbance was measured at two wavelengths, 220 nm (blue) and 260nm (pink). Active fractions against cathepsin K were found between 27-34 minutes (the yellow highlighted section containing the most activity, and inhibitory activity measured across the red).

3.4.7 Strong and Weak Ion Exchange Columns

Four ion exchange columns were tested for use in purifying L-91-3. Two cation exchanges; weak cation exchange (WCX) and strong cation exchange (MCX), along with

two anion exchange columns; weak anion exchange (WAX) and strong anion exchange (MAX).

In both the weak and strong anion exchange column, the entire sample simply flowed through the resin without having any retention on the column nor was any of the sample cleaned when analyzed on HPLC. This was expected since the sample had already passed through the Dowex Marathon® A resin. The cation exchanges on the other hand proved quite useful. Looking at the elution profiles, it appeared that the strong cation exchange would be the ideal candidate. However, when reviewing the results quantitatively a large proportion of the sample did not elute from the column and was lost. WCX on the other hand, being a weak cation exchange did not irreversibly bind the sample. The only negative to this resin was that it had a lower load capacity requiring larger columns to be used for the purification. Following the loading step and the two washes (H_2O followed by 5% NH_4OH), the sample eluted immediately with the subsequent addition of H_2O containing 2% Formic Acid (Figure 3.7). Following elution from the weak cation exchange column, all 3μMol of inhibitors from L-91-3 was recovered.
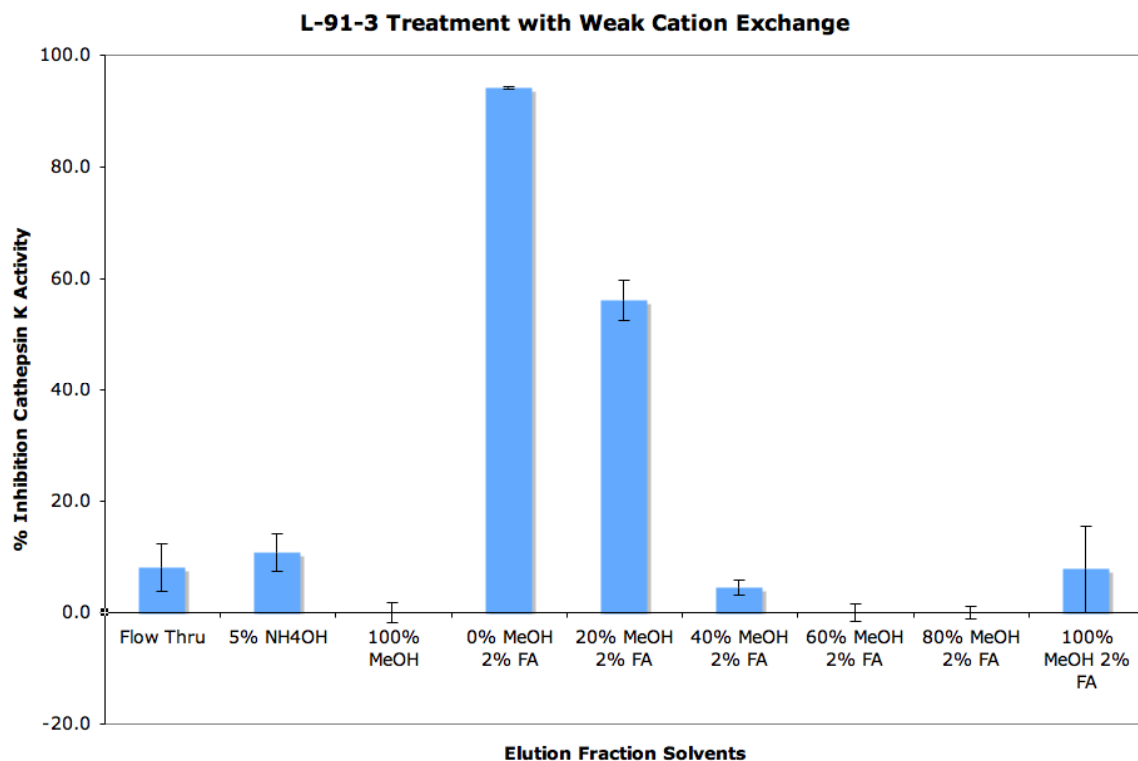
Figure 3.7 Elution of L-91-3 from Weak Cation Exchange

L-91-3 loaded onto the weak cation exchange column in H_2O was washed with 5% NH_4OH followed by 100% MeOH. Elution began by the addition of 2% Formic Acid (FA) at which point all of the active compound came out in the first two elution concentrations (0% MeOH - 2% FA and 20% MeOH - 2% FA). Inhibition of cathepsin K activity tested in duplicate.

3.4.8 LH-20 Size Exclusion Column

The sample, dissolved in 50% methanol prior to loading onto the LH-20 column, was collected into 12ml fraction. The sample began to elute after 252 minutes with the active fractions beginning to elute at 372 minutes. Once the active peak was collected and concentrated to dryness, all of the active sample was recovered (3μMol). Running the sample on analytical HPLC showed distinct improvement in the absorbance profile (Figure 3.8).
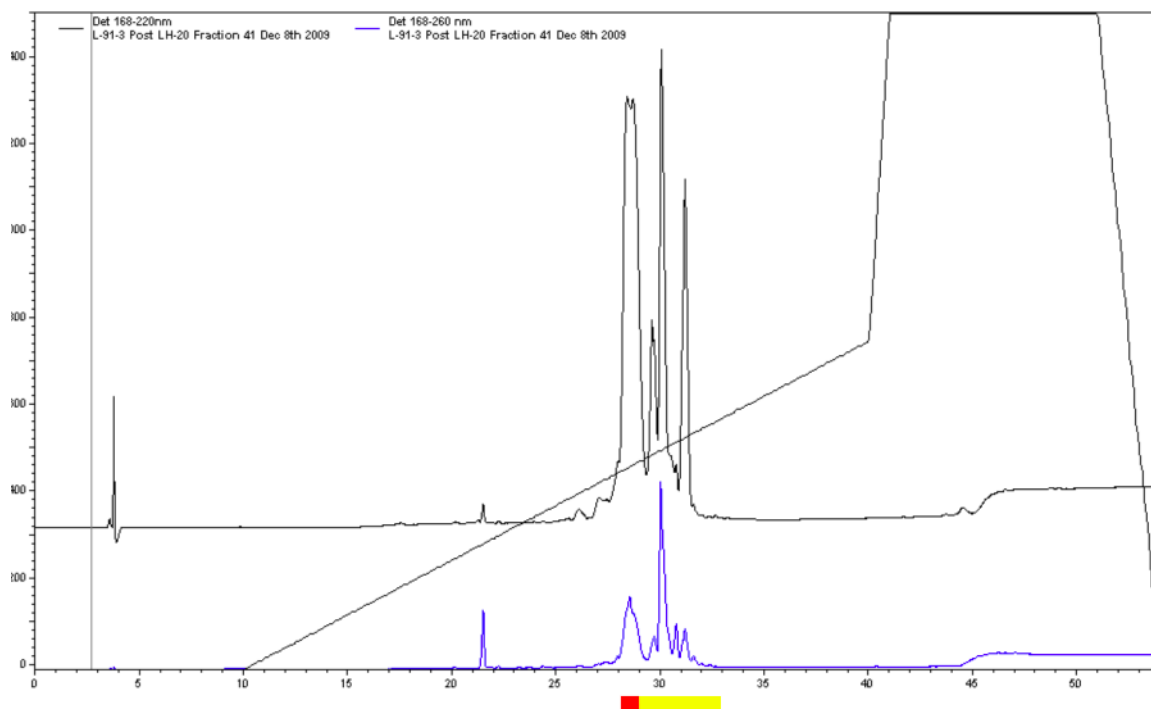
Figure 3.8 Semi Preparative High Performance Liquid Chromatography

Following the LH-20 size exclusion L-91-3 sample was run on semi-preparative HPLC at an isocratic elution of H_2O with 0.1% TFA for 10 minutes followed by a 30 minute gradient to 50% ACN with 0.1% TFA. Absorbance was measured at two wavelengths, 220 nm (blue) and 260nm (black). Active fractions against cathepsin K were found between 28-33 minutes. Fraction # 28 (Red) was isolated and further purified using RP C18 preparative HPLC.

3.4.9 High Performance Liquid Chromatography

Multiple runs of HPLC were performed with 5mg of the L-91-3 sample loaded onto the semi-preparative C18 HPLC column. At this point in the purification, the sample was split into 2 fractions; the first fraction, #28 (red, figure 3.8) containing the most active peak at 28 minutes, and the second fraction which contained minutes 29 to 33 (yellow, figure 3.8), which were significantly less active however kept nonetheless. Although fraction #28 contained the majority of the active compound, from looking at the chromatogram it was clear that it still contained many overlapping peaks that was

confirmed through NMR analysis. A portion of fraction #28 was then taken for crystallography studies, while the majority was subsequently purified using a C18 preparative HPLC column with the help of Dr. David Williams (Dr. Raymond Andersen Laboratory, Department of Chemistry, UBC).

3.4.10 High Performance Liquid Chromatography (Raymond Andersen Laboratory)

Sample #1 (fraction #28) was loaded 5mg at a time onto the C18 column in Dr. Raymond Andersen's laboratory resolving 4 major peaks. The 4 peaks (named Vince 1-4) (Figure 3.11) were collected and analyzed by NMR and mass spectrometry. Based on the NMR data, peak Vince-3 was simply a mixture of peak Vince-2 and Vince-4.

Vince – 4 was the predominant compound purified from HPLC with 6.4 mg recovered while Vince – 1 and Vince – 2 recovered 5.8 mg and 4.4 mg respectively. Additionally there was a large grouping of peaks which eluted early, however due to the lack of baseline resolution; these fractions were not analyzed by NMR and Mass spectrometry.
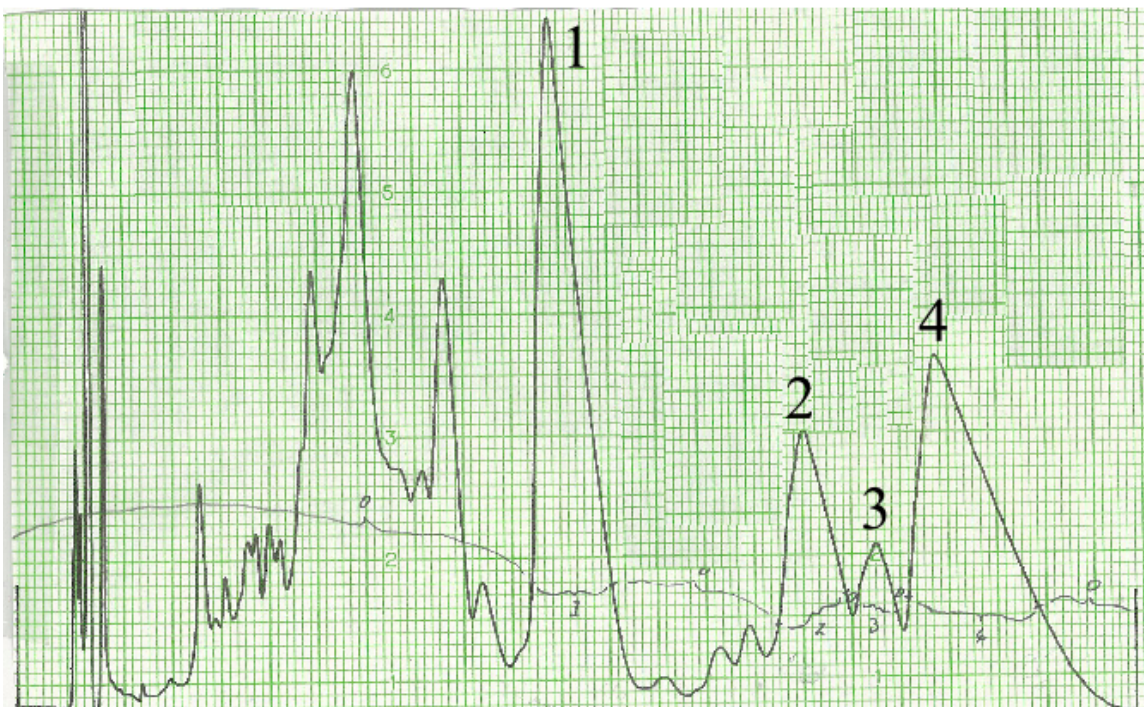
Figure 3.9 Separation of Compounds Vince 1 - 4

Fraction #28 (Figure 3.8) was loaded onto a reverse phase C18 preparatory column with an elution gradient of 15% ACN with 0.05% TFA. The profile was monitored at 204 nm with peaks Vince 1 – 4 separated and analyzed further by mass spectrometry and NMR.

3.5 Structural Analysis

3.5.1 Mass Spectrometry and Nuclear Magnetic Resonance

The mass spectrometry gave two molecular weights, 586 grams per mole for Vince-2 and 604 grams per mole for both Vince-1 and Vince-4. All three compounds were found active, inhibiting cathepsin K collagenase activity. These weights were used as a reference for the structural determinations through NMR. The NMR data (Figure 3.13-3.14 Appendix A2) for the three compounds purified out of L-91-3 (Vince – 1, 2 and 4) confirmed the presence of small peptide inhibitors (Chart 3.1-3.2). Unfortunately no conclusive structure of Vince – 1 was attained. Vince – 4 was identified as the

cycloarginal tautomer of antipain. Using the NMR data, assignment of the structure was possible through comparison with the literature. While Vince – 2 proved to be a very similar structural derivative. The molecular weight of Vince – 2 differs from Vince – 4 by the loss of H_2O and requiring an additional site of unsaturation. The difference occurs at the terminal cycloarginal residue of antipain. Dehydration across the C-1/C-2 bond would account for the loss of H_2O (OH eliminated from the loss of the aldehyde and loss of a hydrogen atom from the formation of double bond).

As mentioned above, two molecular weights, 586 grams / mole and 604 grams / mole were identified for the compounds Vince – 2 and Vince – 4, which matches perfectly with the NMR results. However this does not match the structure identified through crystallography.
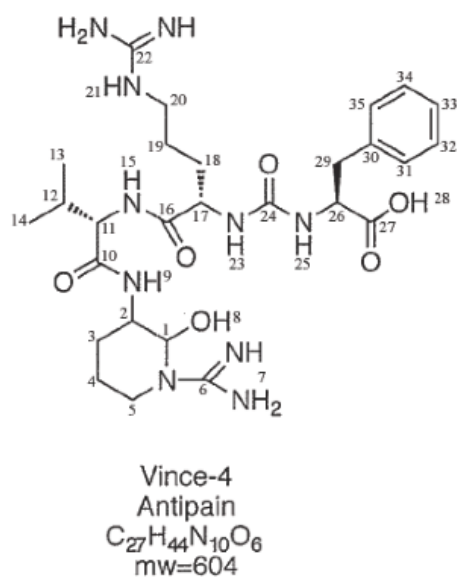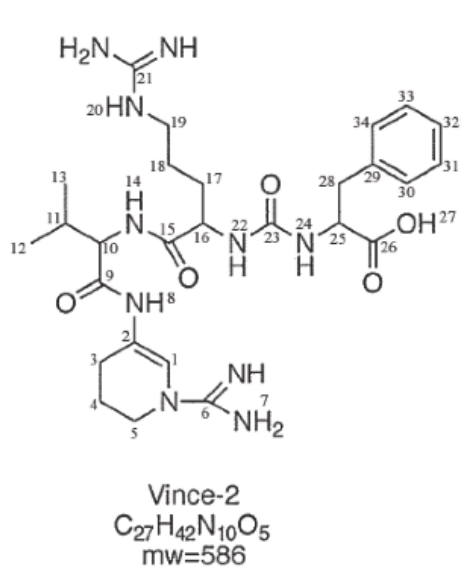
Figure 3.10 Structure of Compounds Vince - 2 and Vince - 4 Isolated from Reverse Phase HPLC

Carbon assignment correlates to NMR signals in Tables 3.1 and 3.2.

Table 3.1 Compound Vince – 2 (see Figure 3.12) NMR data in DMSO-d6

^{1}H and ^{13}C NMR at 600 MHz, Correlations were assigned based on COSY-60, HSQC, HMQC, HMBC, and tROESY, all ran at 600MHz.

Atom Number	Proton	Carbon
1	7.34s	113.24
2		122.7
3	2.21m	23.92
4	1.83m	20.64
5	3.34m, 3.39m	43.29
6		154.05
7	7.62bs	
8	9.24s	
9		169.71
10	4.2ds	57.84
11	1.93s, 6.58	30.78
12	0.84d, 7.14	17.98
13	0.85d, 6.84	19.13
14	7.82d, **8.52**	
15		172.12
16	4.2bs	52.17
17	1.41b,1.58m	30.14
18	1.42-1.47m	24.95
19	3.07m, 3.10m	40.41
20	5.76, 7.57bt	
21		156.60
22	6.50d, 7.98	
23		157.14
24	6.36d, 7.92	
25	4.33dd, (12.66, 12.9, 7.14, 7.38)	53.83
26		173.57
27	12.70bs	
28	2.86dd, 13.74, 7.74 3.00dd, 13.44, 4.92, 13.98, 4.38	37.37
29		137.21
30	7.16d 7.38	129.14
31	7.26dstt	128.11
32	7.19t, 6.84	126.37
33	7.26dstt, 7.29	128.11
34	7.16d, 7.38	129.19

Table 3.2 Compound Vince – 4 (Figure 3.12) NMR data in DMSO-d6

^{1}H and ^{13}C NMR at 600 MHz, Correlations were assigned based on COSY-60, HSQC, HMQC, HMBC, and tROESY, all ran at 600MHz.

Atom Number	Proton	Carbon
1	5.62bs	76.34
2	3.77m	48.74
3	1.51m, 1.65ds	23.78
4	1.48d, 1.68dsm	23.26
5	3.14td, 12.6, 3.45bd, 11.52	39.3
6		156.58
7	7.57bs	
8	6.56bs	
9	7.84d, 8.42	
10		170.26
11	4.28dd, 6.6, 9.1	59.92
12	1.92sixtet, 6.6	31.09
13	0.79d, 6.9	17.69
14	0.81d, 6.6	19.08
15	9.1, 7.68d	
16		171.86
17	4.17m	52.43
18	1.42ds, 1.61m	30.01
19	1.43	25.02
20	3.06dsm, 3.10ds	40.43
21	7.61t, 5.61	
22		156.66
23	6.34d, 8.52	
24		157.16
25	6.35d, 7.98	
26	4.34dd, 13.2, 13.14, 7.38, 7.11	53.87
27		173.57
28	12.70bs	
29	2.87dd, 13.68, 6.84, 13.74, 7.13 2.99dd, 13.68, 4.98, 13.74, 4.92	37.36
30		137.20
31	7.17d, 7.08	129.25
32	7.26dstt, 7.41orm	128.15
33	7.20m, 6.87	126.41
34	7.26distt, 7.41orm	128.13
35	7.17d, 7.08	129.25
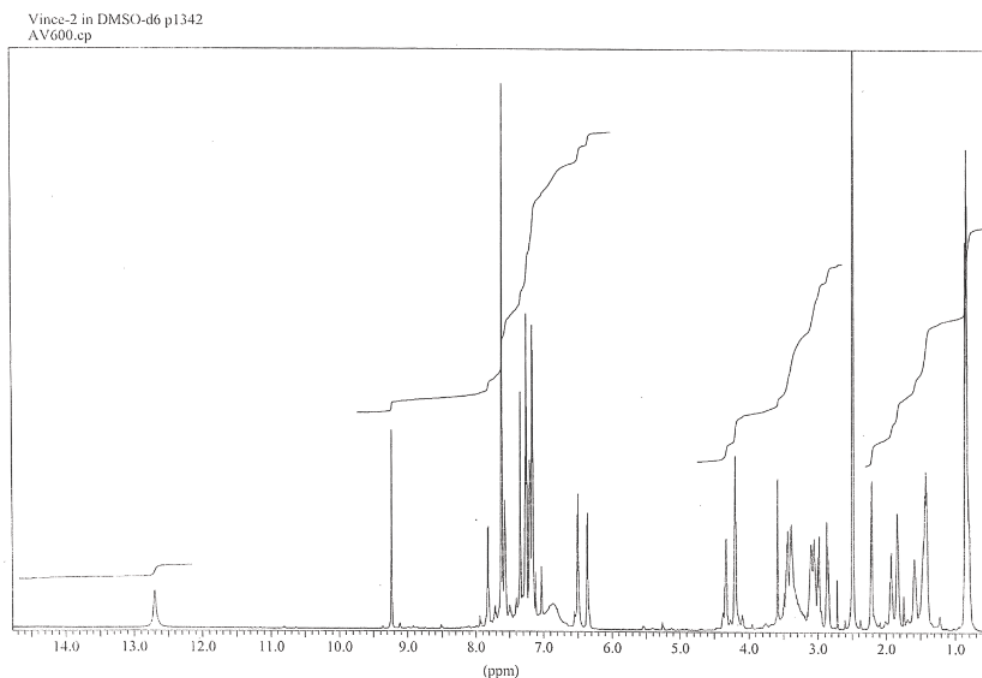
Figure 3.11 Vince – 2 1H NMR Spectrum in DMSO-d6 at 600MHz
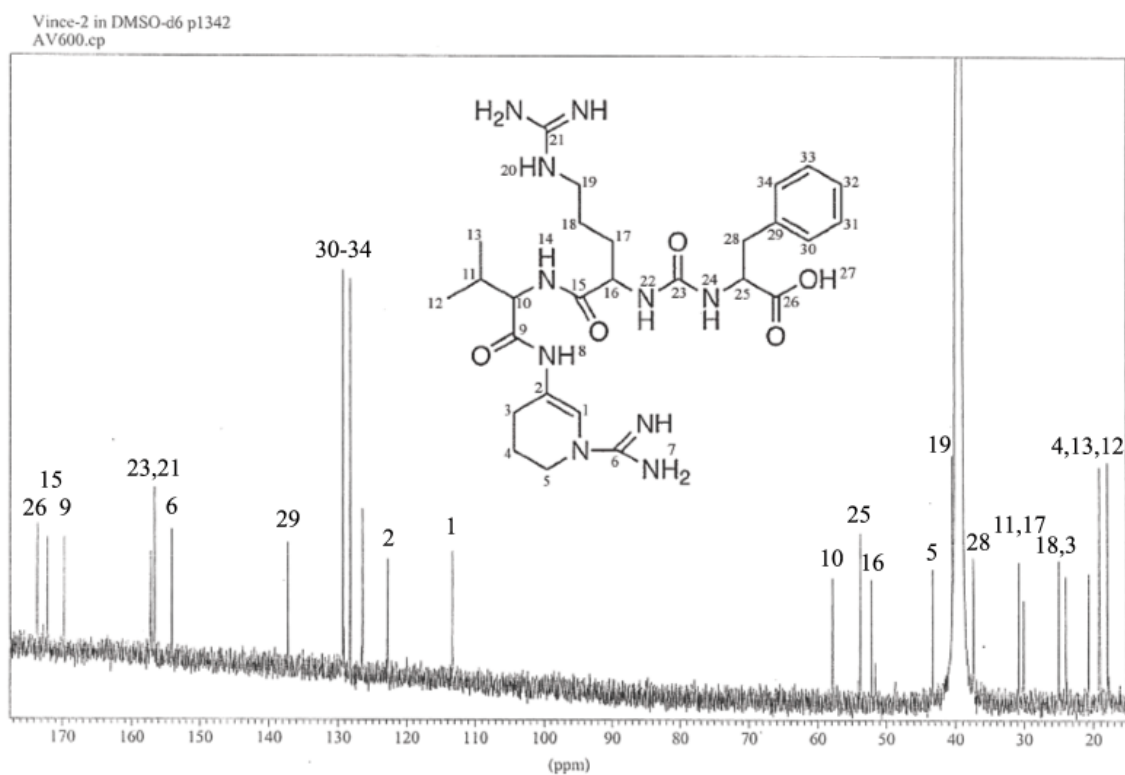
Figure 3.12 Vince - 2 13C NMR Spectrum in DMSO-d6 600MHz

59
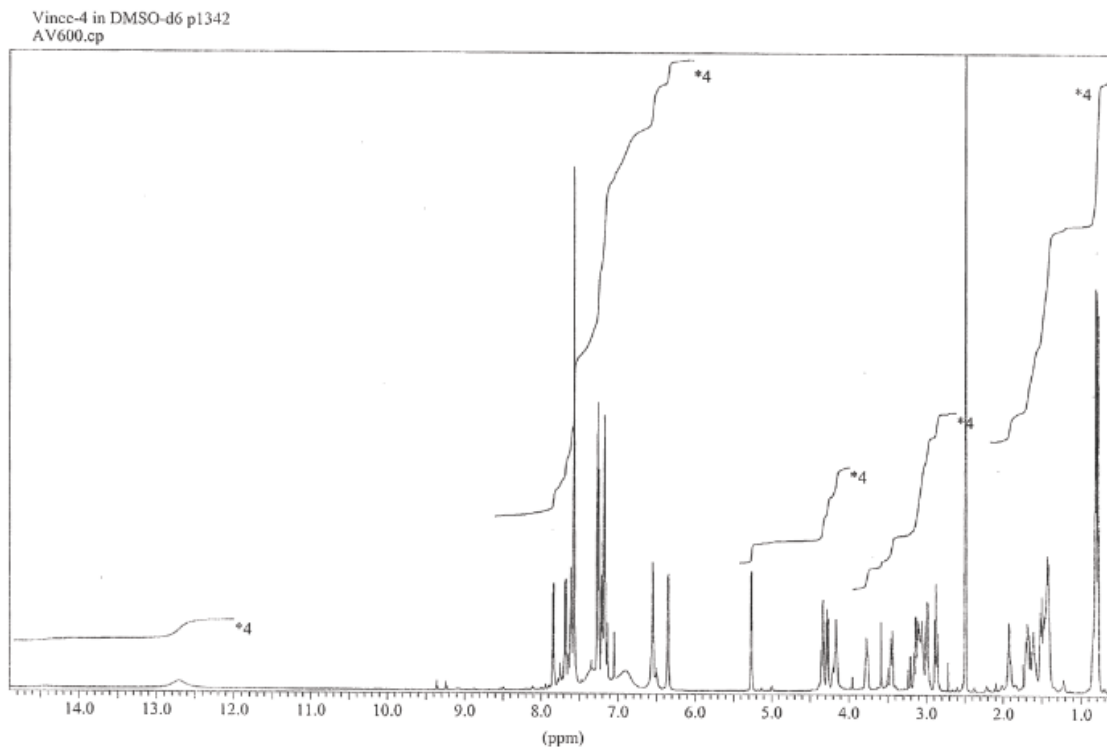
Figure 3.13 Vince - 4 1H NMR Spectrum in DMSO-d6 at 600MHz
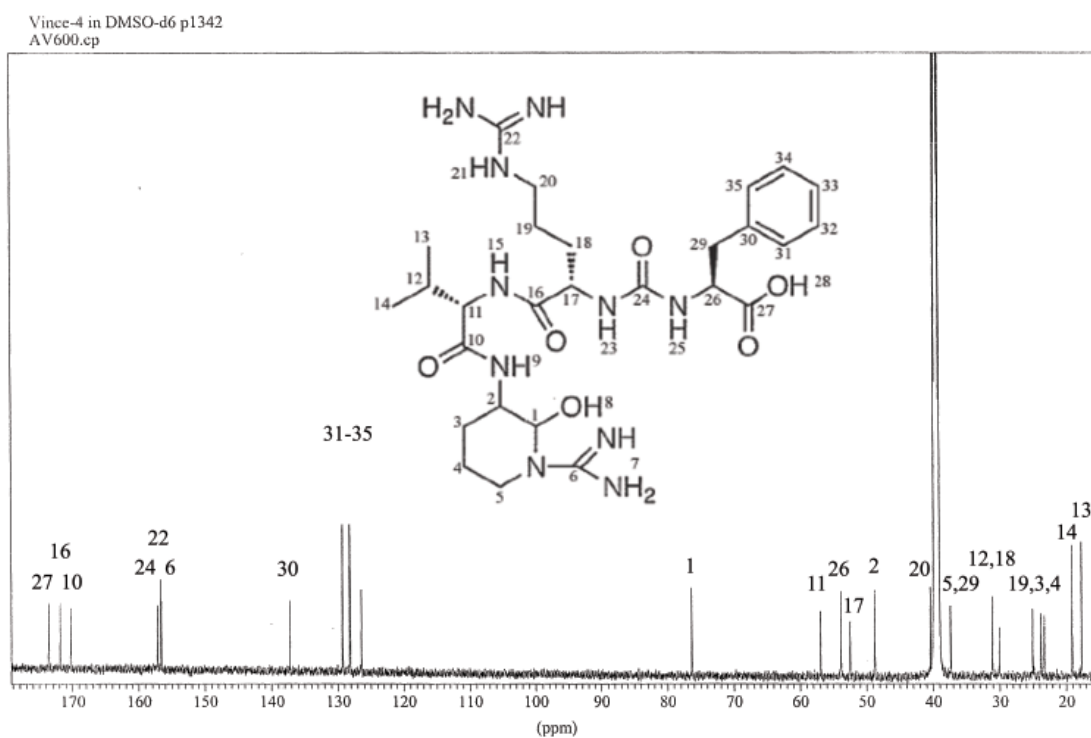
Figure 3.14 Vince – 4 13C NMR Spectrum in DMSO-d6 at 600MHz

60

3.5.2 Crystallography

Following the first round of HPLC, fraction #28 was collected and a sample was provided to Dr. Adeleke Aguda for crystallization and analysis. Using the co-crystallizing method, Dr. Aguda managed to obtain crystals for a L-91-3 compound bound to the active site of cathepsin K by mixing the ligand (from L-91-3) with the protein, and then suspending them over the crystallizing agent. This generated a crystal which diffracted giving a structure that was subsequently refined to a resolution of 2.0 Å and a final R-factor of 19.2% with an R_{free} of 23%.

The crystal structure (Figure 3.9-3.10) revealed a small peptide inhibitor made up of three amino acids and the decarboxylation product of arginine, agmatine (Agm). The three amino acids are arginine, valine and serine followed by agmatine linked through a ureido derivative to the valine. The structure identified through crystallography also contained an aldehyde covalently bound to cysteine[25] of cathepsin K, with the remaining structure of L-91-3 extending into and beyond the P1 site of cathepsin K. The hydrogen bonding interactions occurring between the inhibitor ligand and cathepsin K are summarized in Figure 3.17 and Table 3.3. We named this structure, Lichostatinal due to being isolated from lichens
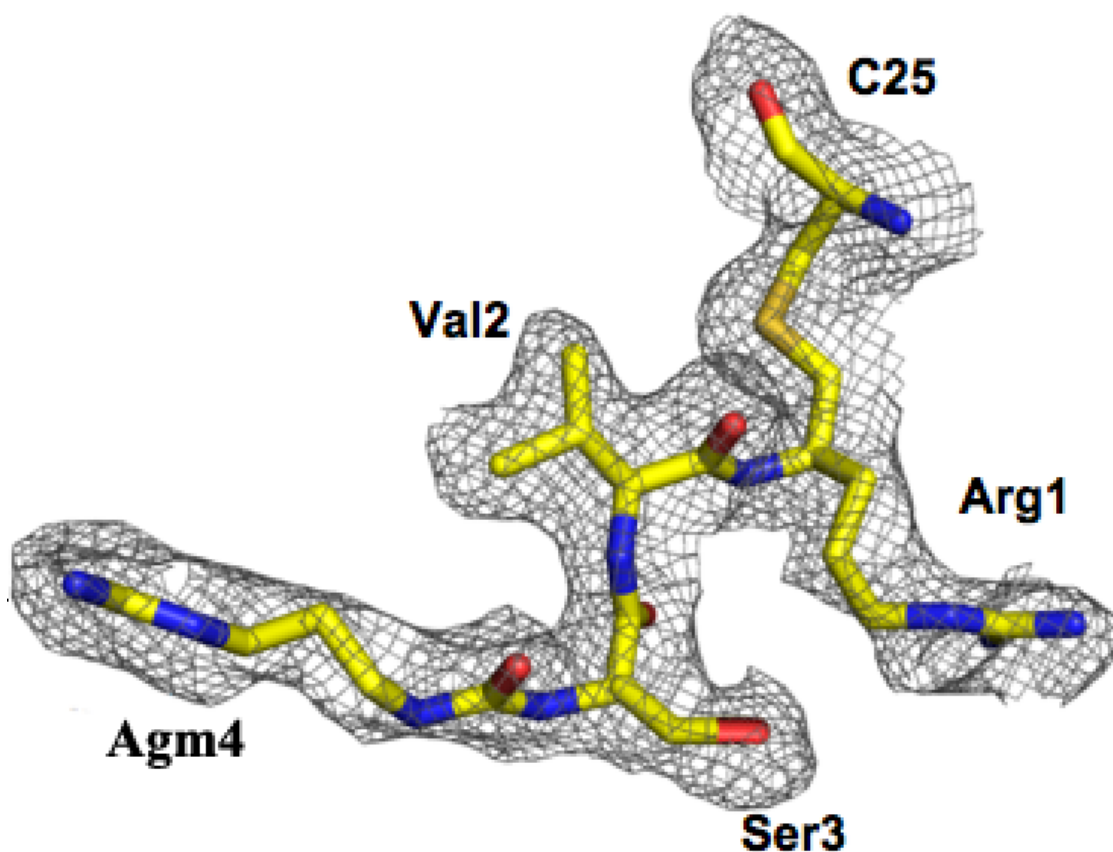
Figure 3.15 Electron Density Omit Map (Fo – Fc) Contoured to 2.0 Sigma of Lichostatinal (a Compound Found in L-91-3) as Found Bound in the Active Site of Cathepsin K

Co-crystallization revealed that a small peptide inhibitor, Lichostatinal, containing an aldehyde which was covalently bound to Cysteine 25 of cathepsin K. This tetra-peptide molecule contained an arginine at the carboxy terminal end, followed by a valine, serine and agmatine connected through a ureido derivative.
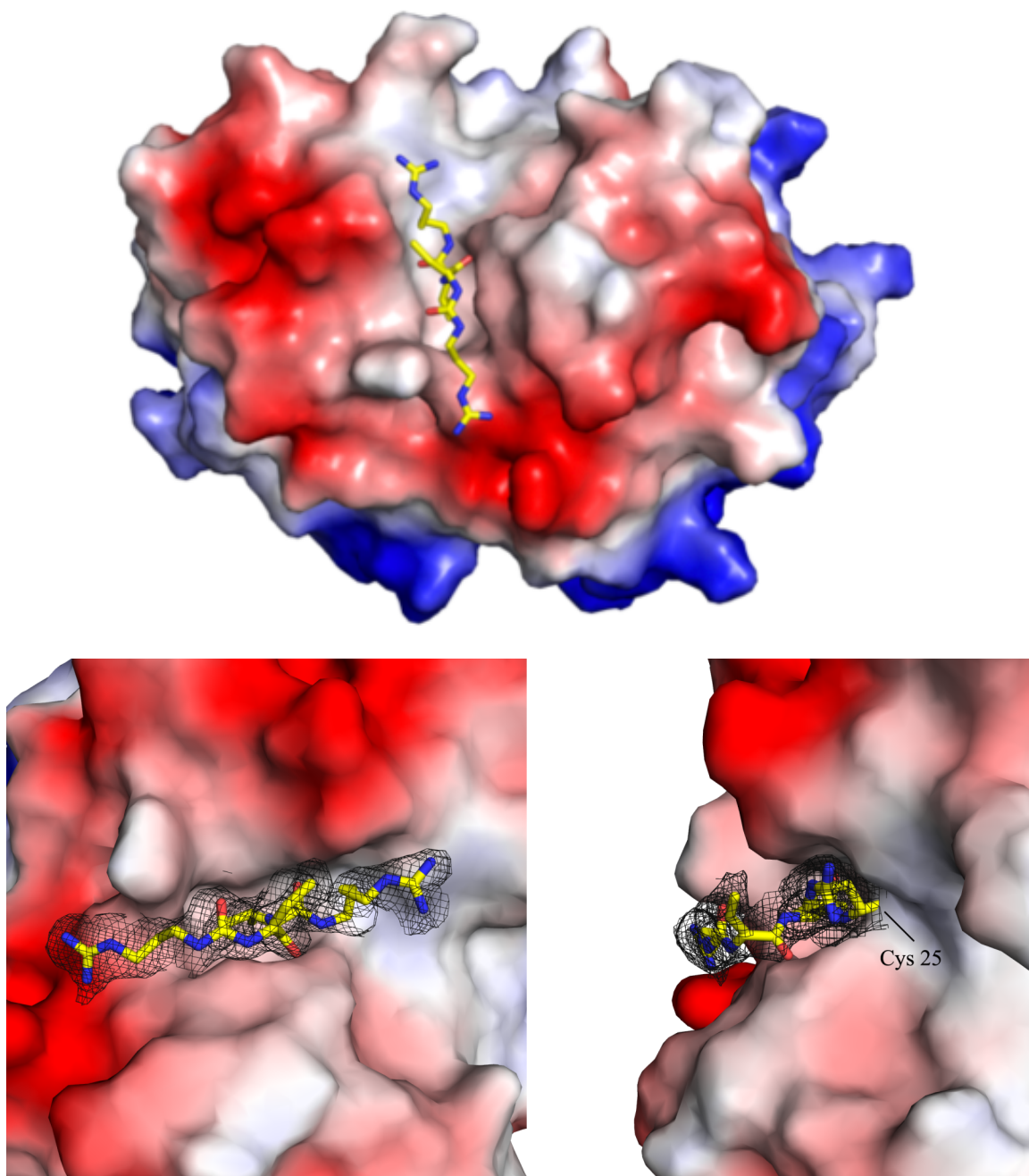
Figure 3.16 Three-Dimensional Structure of Cathepsin K Lichostatinal Covalently Bound to Active Site as Determined by X-ray Diffraction Methods

Space filling, surface charge representation of cathepsin K (blue positive, red negative, white neutral) in the region where Lichostatinal is covalently bound to the active site Cysteine 25 with the peptide inhibitor backbone extending into and beyond the P1 site of the protein. The electron density map is contoured at the same level as figure 3.15.
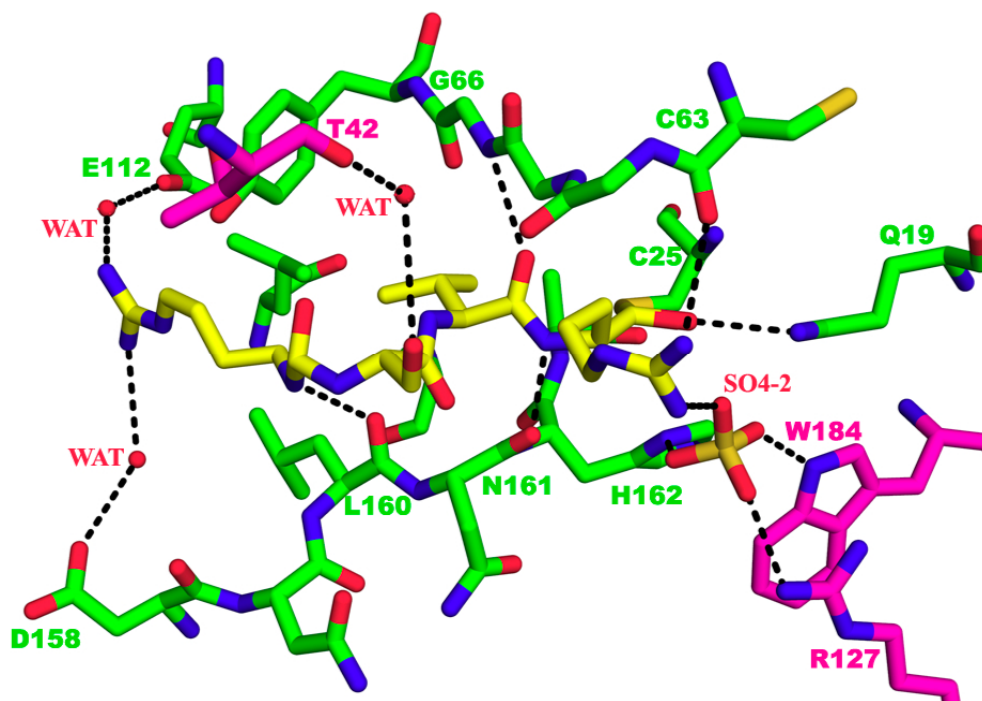
Figure 3.17 Lichostatinal Hydrogen Bonding Interactions When Bound in the Active Site of Cathepsin K

To identify interacting residues, searched for residues within a 4 Å radius of the Lichostatinal using structure analysis program in the CCP4 program suite. Lichostatinal is highlighted with a yellow backbone in the active site of cathepsin K highlighted in green. Lichostatinal also interacts with the symmetry related cathepsin K molecule in the crystal packing highlighted in magenta. Hydrogen bonding interactions are indicated through dashed lines.

Table 3.3 Hydrogen Bonding Interactions of Lichostatinal in the Binding Subsites of Cathepsin K

P residue	Ligand Atom	Interacting Atom on Cathepsin K	Interaction Mediated Through	Distance in Å
P1	Arg NH1	Arg 127 Nη1	SO_4^{-2}	3.92*
P1	Arg NH1	Trp184 Nε1	SO_4^{-2}	3.92*
P1	Arg NH1	His162	SO_4^{-2}	3.92*
P1	Arg NH2	Cys63 Carbonyl Oxygen		3.87
P1	Arg O	Gln19 Nε2		3.44
P1	Arg N	Asn161 Carbonyl Oxygen		2.99
P2	Val O	Gly66 Amine		2.80
P3	Ser OG	Thr42 Oγ1	H_2O	2.67*
P4	Agm N	Leu160 Carbonyl Oxygen		3.86
P4	Agm NH1	Glu112 Oε1	H_2O	3.19*
P4	Agm NH2	Asp158 Oδ1	H_2O	3.49*

* Indicates distance to molecule SO_4^{-2} or H_2O molecule.

The sample used in the co-crystallization experiment was also analyzed by Ultra Performance Liquid Chromatography matched with Mass Spectrometry, searching for a molecular weight of 500.6 g/Mol as calculated for the compound found in the crystal structure. Although 500.6 g/Mol was not identified, a compound with a mass charge ratio of 251 m/z corresponding to the inhibitor being double charged (positive charge on the guanidinium of the arginine and agmatine) was identified giving a distinct peak. A mass charge ratio of 251 m/z eluted early from the UPLC at minutes 6 to 7 (Figure 3.18). To ensure that it was not simply noise, 250 m/z was searched and did not show any relevant peaks. Additionally a mass charge of 229 m/z showed peaks, however they did not overlap with those of 251 m/z ruling out the potential of a 229 g/Mol compound with a sodium adduct.
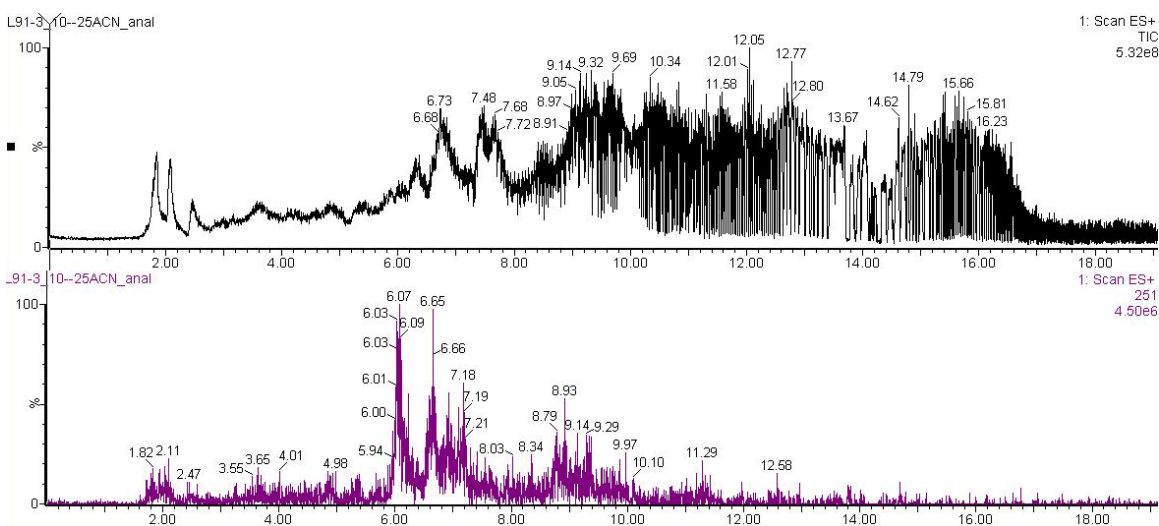
Figure 3.18 UPLC Elution of L-91-3 Minute 28 Fraction

Fraction #28 from the semi preparative HPLC was run on UPLC matched with mass spectrometry of the elutions at the CDRD with the help of Dr. Adam Galey and Dr. Markus Heller. A compound giving a mass charge ratio of 251 m/z was identified eluting early (purple profile, minutes 6-7) when compared with the total elution array.

3.6 Kinetics

The K_i (dissociation constants) values were determined for the 4 samples, Vince 1, 2, 4 and Antipain using Dixon plots. Using the synthetic cathepsin K substrate Z-F-R-MCA, Dixon plot analysis was performed. Cathepsin K inhibitory activity was tested in triplicate. Results suggested that all three compounds were competitive inhibitors of cathepsin K with K_i values of 258nM (+/- 68nM), 295nM (+/- 123nM) and 125nM (+/- 89nM) for Vince 1, 2 and 4 respectively (Figure 3.17).

Vince – 4 was found to have the same structure as Antipain, the papain-like protease inhibitor. Antipain is commercially available, and it was found to have a K_i for cathepsin K_i of 41nM (+/-37nM) using the same method described above. Dixon plots were generated for each enzyme assay, and the x intercept values of $-K_i$ were averaged.

Dixon plots generated below are a representative of 1 data set. Each inhibitor was tested three times with the intercepts of the plots averaged out to generate the K_i values.
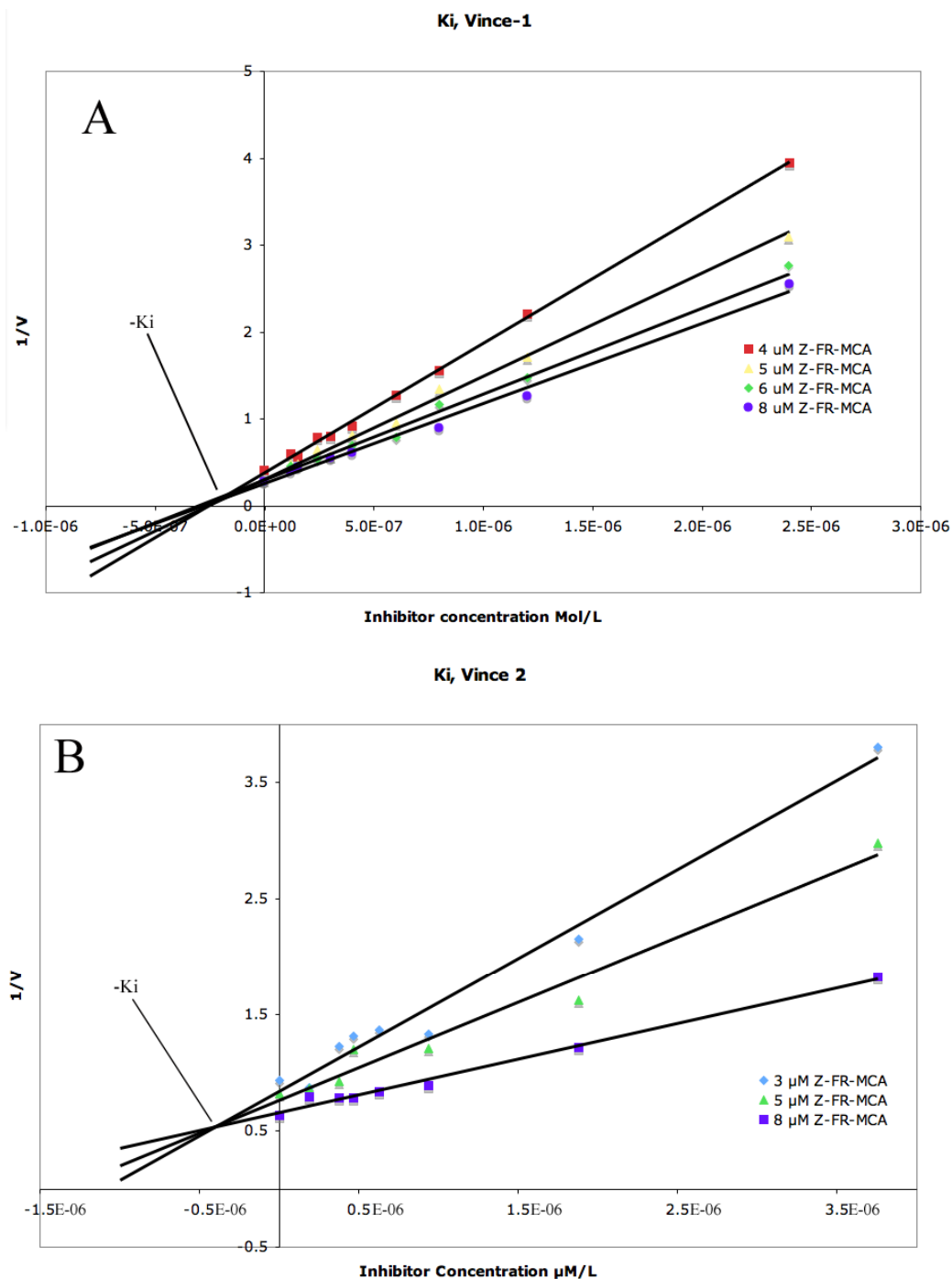
Figure 3.19 Ki Value Determination of Vince 1, 2

K_i values for the inhibition of the hydrolysis of Z-F-R-MCA by cathepsin K by the inhibitors Vince – 1 (A), Vince – 2 (B) were determined using Dixon plots. The concentration of Z-F-R-MCA ranged between 3μM to 8μM per reaction well while the concentration of the inhibitors ranged from 200nM to 8μM. Both inhibitors were found to inhibit cathepsin K in a competitive manor with K_i vales of 258 nM for Vince – 1 (A), 295 nM for Vince – 2 (B).
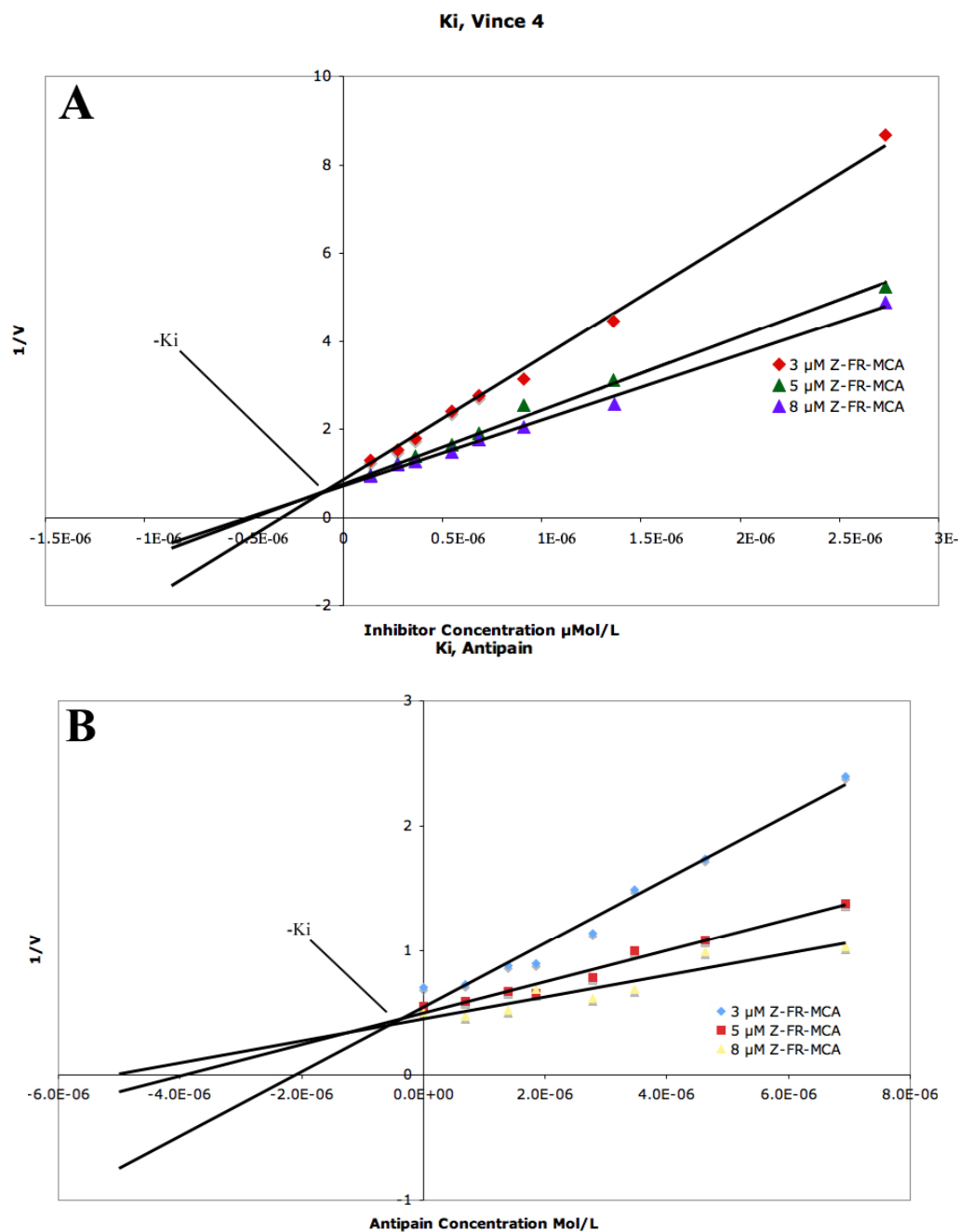
Figure 3.20 Ki Value Determination of Vince 4 and Antipain

K_i values for the inhibition of the hydrolysis of Z-F-R-MCA by cathepsin K by the inhibitors Vince – 4 (A) and the commercially available Antipain (B) were determined using Dixon plots. The concentration of Z-F-R-MCA ranged between 3µM to 8µM per reaction well while the concentration of the inhibitors ranged from 200nM to 8µM. Both inhibitors were found to inhibit cathepsin K in a competitive manor with k_i vales of 125 nM for Vince – 4 (A) and 41nM for Antipain (B).

4 Discussion

4.1 Changes to Media

Difficulties separating individual peaks during the purification of the L-91-3 sample required that a different approach be taken. As previously mentioned, the growth of *streptomyces* along with the production of secondary metabolites is highly dependent on physiological and environmental factors such as nutrient supply and temperature [76, 81]. As such the recipe for the *streptomyces* media was altered in order to establish the best starting conditions for L-91-3 production.

Considerations were taken based on not simply which media recipe produced the largest quantity of inhibitor, but also which one had the lowest net media weight. To calculate the net media weight, two factors were measured; (1) the initial weight of the media (combination of all ingredients added (Table 2.1)) along with (2) the subtraction of the cell pellet weight. Since the cell pellet was found to contain no active compound, this was a very easy source of contaminants that were readily eliminated simply through filtration. Evidence has shown that secondary metabolite production is decreased when bacteria cells are cultured in a rich, balanced nutrient medium. It was also demonstrated that they only reach their full production potential when growth is restricted [82].

By eliminating two of the amino acid sources, malt and yeast extract, 10 grams of media contaminants were eliminated per liter prior to the start of fermentation. An additional 5 grams of glucose and half (10 grams) of the 3-(N-morpholino) propanesulfonic acid

(MOPS) could be eliminated without any major change to the inhibitor production (Figure 3.1). As a result the next consideration consisted of finding which media recipe produced the largest cell pellet while having comparable bioactivity.

As would be expected, limiting the nutrients resulted in an increase in inhibitor production up to a point, at which time there were simply insufficient building blocks for *streptomyces* growth and secondary metabolite production. The same could not however be said for bacteria cell mass. The more the bacteria cells were limited in nutrient supply, the less biomass was created (Figure 3.2). As such, media recipes A, B, C, and D (see Table 2.1) produced comparable cell pellets to the original stock recipe. Having produced comparable inhibitor quantities to the control, as well as cell growth as measured through grams fresh weight, media recipe C was selected for all future fermentations since it resulted in a decrease in media contaminants from the start of purification.

The final environmental factor investigated was the fermentation temperature. Historically fermentation was always performed at 30°C, but with the availability of a 37°C room, fermentation at this temperature was also tested. Despite using the same media recipe, and allowing the same amount of time for growth, fermentation at 37°C yielded very poor results. Although cell pellet grew, very little inhibitor activity was detected. Studies suggested that perhaps decreasing the temperature of fermentation rather than increasing it would have been advisable. A study investigating the environmental factors on *streptomyces* secondary metabolite production showed that the greatest secondary metabolite productivity occurred depending on the strain and

metabolite in question between 17 - 32°C [81]. When considering the source of the bacteria (soil bacteria from coastal BC along with Vancouver Island), the natural growth environment should be comparable to that found in British Columbia which would be far more temperate than 30-37°C assuming the compound has a physiological role.

4.2 Purification Analysis

Prior to NMR, mass spectrometry and X-ray crystallography, many characteristics of L-91-3 inhibitors were already known thanks to the experiments performed during the purification process. The first information obtained came from the numerous solvent extractions performed with ethyl acetate, dichloromethane and hexane. In all three cases, the active compound remained in the aqueous phase suggesting that the compounds in L-91-3 were polar.

Due to the fact that Amberlite XAD4 bound L-91-3 with greater efficiency than did Amberlite XAD16, this suggests that L-91-3 compounds have a relatively small molecular weight.

The polarity of L-91-3 inhibitor mix was reconfirmed by the presence of charged compounds while performing washes with the ion exchange columns. The lack of binding of L-91-3 to Dowex Marathon A (anionic exchange resin) suggested that the structurally characterized compounds (Vince 1, 2, 4 and Lichostatinal) did not contain a negative charge, however, since there was a significant loss of activity (Table 4.1), the

presence of negatively charged inhibitors is likely. The complete irreversible binding of L-91-3 inhibitors to Dowex Marathon C, a cation exchange gave strong evidence that L-91-3 inhibitors must contain a positively charged moiety. This fact was reconfirmed once the ion exchange columns were used. Once again in both the strong and weak anion exchange columns, the active compound in L-91-3 did not bind.

Table 4.1 Purification Outline for L-91-3 *Streptomyces* Bacteria Media

Purification Step	mg Material	Purification Factor	µmol L-91-3 inhibitor	Activity (µMol/g)
Harvest	84000	1	20	0.24
Amberlite XAD4®	8800	9.5	20	2.3
Ethyl Acetate Extraction	1500	56	15	10
Dowex Marathon A	1300	64.6	5	3.8
3kD Filtration	1300	64.6	5	3.8
Butanol Extraction	1000	84	5	5
C 18 RP Column	800	105	5	6.3
WCX Column	350	240	3	8.6
LH-20	300	280	3	10
SP HPLC	150	560	0.4	2.7
Vince – 1	5.8	14482.8	0.0003	0.052
Vince – 2	4.2	20000.0	0.0004	0.095
Vince – 4	6.4	13125.0	0.008	1.25

The cation exchange columns offered a little more insight into the compounds. When using a strong ion exchange resin, it is the compound of interest on which you are changing the ionic property while the resin remains charged. In contrast, for a weak ion exchange resin, it is the resin that changes charge and not the compound of interest. The weak cation exchange resin reconfirmed the positive charge on L-91-3 compounds however the irreversible binding to the strong cation exchange gave proof that the positively charged molecule had a high pKa value and is not easily neutralized.

From filtering the L-91-3 sample through the Amicon 3K filter, we were able to determine that the active compound was in fact less than 3000 grams per mole since full activity was retrieved from the flow through. This was expected since *streptomyces* are known for producing a large variety of small molecular weight secondary metabolites [80].

4.3 NMR and Mass Spectrometry

Following purification of the L-91-3 bacteria media samples, 4 active peaks were isolated through C18 HPLC with the help of Dr. David Williams in Dr. Raymond Andersen's laboratory. Named Vince 1 – 4, the four peaks (Figure 3.11) were analyzed through NMR and mass spectrometry. The results showed that of the four peaks, Vince – 1, 2 and 4 were pure compounds with Vince-3 being a mixture of Vince – 2 and 4.

Mass spectrometry along with NMR provided a structure for the compounds; one isolation being the cysteine protease inhibitor Antipain (Vince – 4), with Vince – 2 being a very similar structural derivative. Vince – 1 was found to have the same molecular weight as Vince – 4, with the NMR data suggesting that it is the same compound. However, this could not be confirmed through NMR analysis alone.

The peptide analogue antipain was first discovered as a protease inhibitor by the Institute of Microbial Chemistry in Japan in 1972 [59]. Antipain was not unique to a single strain of actinomyces. When it was first discovered, the same compound was identified in

numerous actinomyces cultures with two strains of primary focus (MB-561-C2 and MC-829-AS1) [59] while the first structural data was identified using strain KC 84-AG 13 [83]. It is fairly common that a natural product produced by bacteria is found in numerous different strains, as has already been illustrated with the microbial inhibitor Leupeptin [57, 77].

Antipain along with Vince – 2 share very similar structural characteristics. Based on NMR and mass spectrometry data they are all tetrapeptides made up of arginine – valine – arginine – phenylalanine with an ureido derivative linking the arginine with the phenylalanine. The chemical name of antipain (Vince – 4) is (s) – 1 – carboxy – 2-phenylethyl] carbamoyl – L arginyl – L – valyl – argininal [83] with the derivatives (Vince – 1 and 2) having minor changes in the P1 subsite argininal amino acid. Due to this ureido derivative, it can be assumed that the final synthesis of these compounds occurs post-translationally, outside of the ribosome.
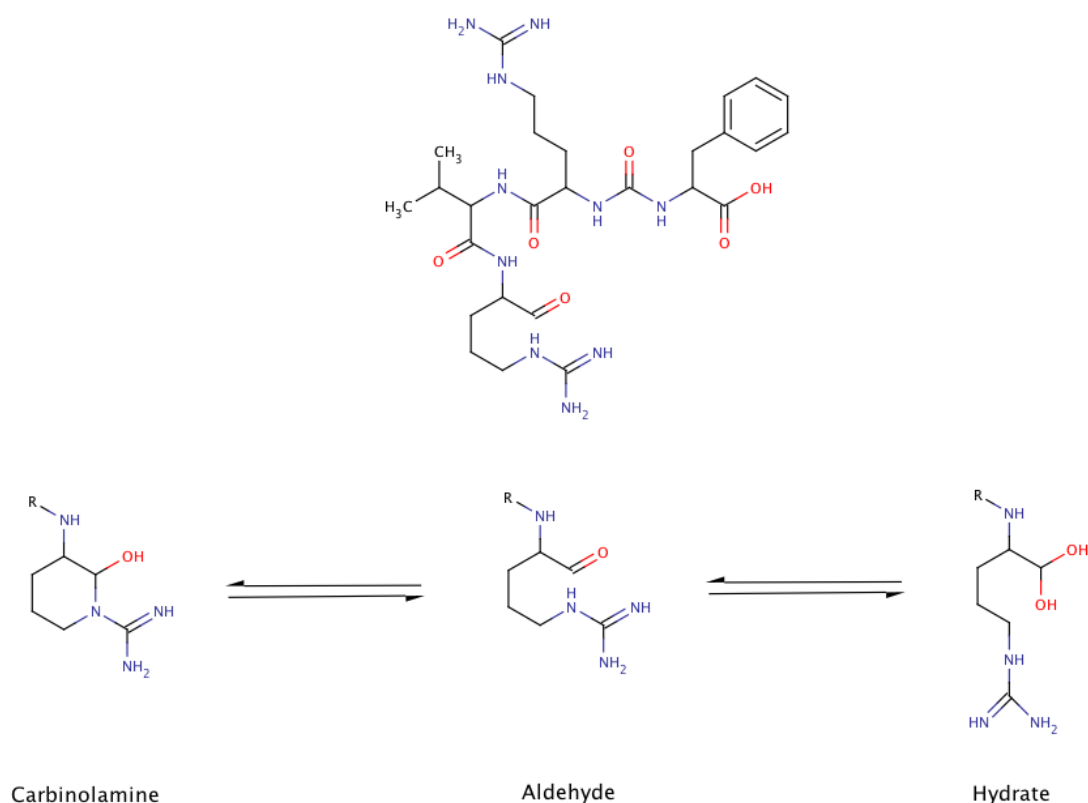
Figure 4.1 Antipain Equilibrium

Antipain can co-exist in three different forms. The aldehyde form although the least abundant is the most active inhibitor of cathepsin K due to its ability for covalent binding to the active site cysteine 25. However the carbinolamine and hydrate form represent the majority of the equilibrium.

Vince – 4 was identified with a cyclized ring on the terminal arginine (Figure 4.1). Although not typically represented this way, literature shows that the carbinolamine and hydrate forms of antipain are actually the most common naturally occurring form with the aldehyde being a minor constituent of the equilibrium despite being the only active form of inhibitor of the three. Similar structural components are seen in the protease inhibitor leupeptin where the inhibitor is found to exist in equilibrium between three inter converting forms in an aqueous solution. The carbinolamine form (Figure 4.1) represented 56%, the hydrate accounted for 42% with the free aldehyde only composing 2% of the equilibrium solution [84].

4.4 X- Ray Crystallography

The crystal structure of cathepsin K co-crystallized with the peptide ligand purified from L-91-3 *streptomyces* showed a novel structure different from those identified through NMR and Mass Spectrometry. Crystallography revealed a small aldehyde inhibitor, named Lichostatinal, covalently bound to Cysteine 25 of cathepsin K containing three amino acids, arginine – valine – serine connected with a ureido derivative to decarboxylated arginine, agmatine.
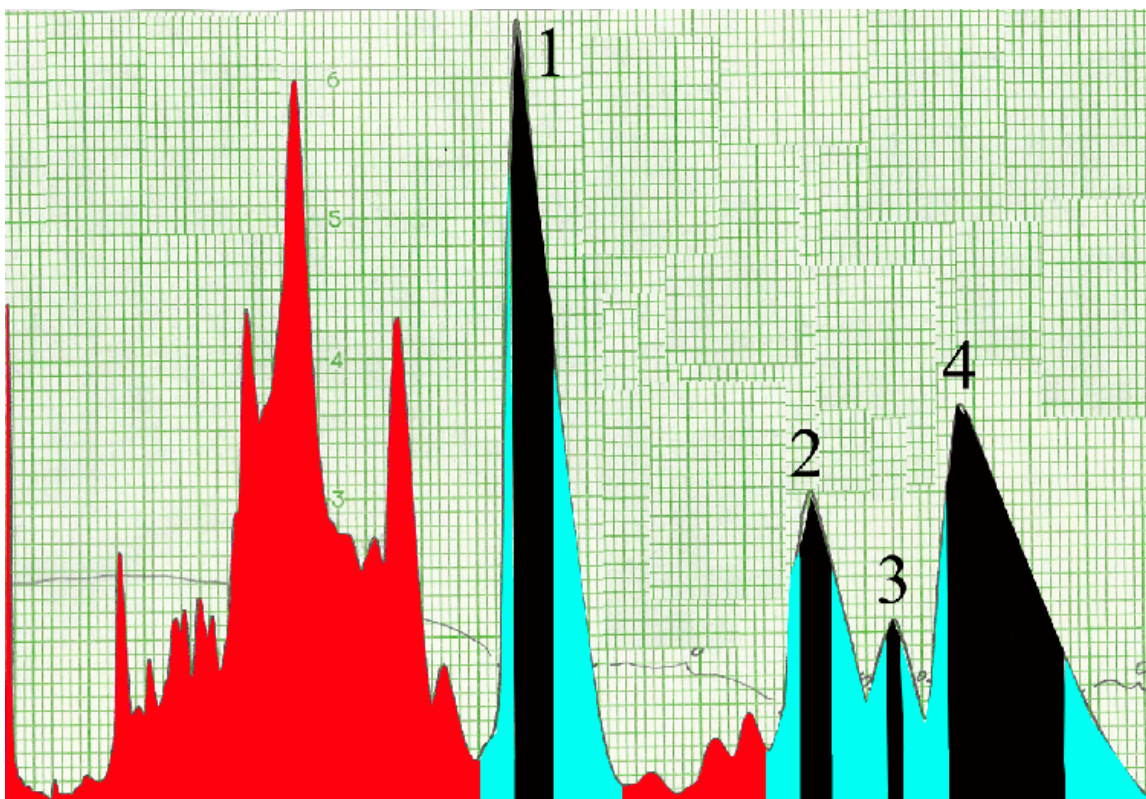
Figure 4.2 High Performance Liquid Chromatography of Fraction #28

Only the black highlighted reagions of peaks 1-4 were collected and analyzed by NMR and Mass spectrometry, while the teal and red were discarded.

4.5 Structural Comparison

All three of the structures identified through crystallography and NMR are potent inhibitors for cathepsin K. It has been previously determined that cathepsin K has a preference of binding arginine in the S1 subsite. Arginine is a common structural characteristic of all three identified compounds in their P1 positions.

One explanation for why the structure found in the crystallography approach was not identified in the samples characterized by NMR/Mass spectroscopy is due to the fact that the sample preparation that went to crystallography was crude and contained several active compounds at various concentrations. Fraction #28 from the first round of semi-preparative HPLC was separated into four purified peaks (black highlights, figure 4.2) from a significant amount of other compounds (red highlights, figure 4.2). In order to ensure purity of the four peaks only the center of each peak was collected, with the edges of each peak (blue highlights, figure 4.2) being discarded. When quantifying the inhibitor before and after the separation, a significant quantity was lost (Table 4.1). It is likely that the leading and tail edges of peaks 1-4 (blue area) contained activity. However, it is more likely that the early eluting peaks (red), which were not collected due to insufficient baseline resolution, could have contained Lichostatinal identified through X-ray crystallography. The initial peak area was collected and tested for cathepsin K inhibitory activity, but did not contain significant anti-cathepsin K activity. This could be due to a low concentration of this inhibitor species. The activity assay quantifies competitive inhibitors, which must compete with the synthetic substrate, Z-FR-MCA, for the binding

pocket of cathepsin K. In competitive inhibitor assays, the substrate concentration is important since the reaction occurs quickly. In a crystal structure however, the reaction rate can be considered infinite which allows for more interactions to take place during binding resulting in an increase in binding stability. The inhibitor has weeks to co-crystallize with cathepsin K. Due to the extended period of time, the inhibitor concentration is not as critical and instead it is simply the inhibitor with the greatest affinity for the active site that should bind. It is most likely, based on the UPLC mass spectrometry data (Figure 3.18), that the inhibitor identified through X-ray crystallography was present in the unresolved peaks at the beginning of the elution. This is supported by our finding that a compound of the molecular weight of 500.6 g/Mol, equivalent for Lichostatinal, was eluted from the column in this peak area (Figure 3.18). Comparing the structure of the two compounds, Vince – 4 purified through HPLC and Lichostatinal, an explanation can be made why Lichostatinal was found in the crystal structure. Looking at the S1-S4 subsite specificity of cathepsin K, both inhibitors begin with an arginine in the P1 residue, followed by a valine in the P2 position with the differences occurring at the P3 and P4 positions of the inhibitors. Looking purely at the amino acids in the P3 position, Vince – 4 would make a slightly better inhibitor for cathepsin K than the structure identified through crystallography due to the fact that the arginine in the P3 position is preferable for cathepsin K inhibition over serine [34]. When peptide substrate libraries were tested for cathepsin K, arginine in the P3 position allowed for a 25% better hydrolysis when compared to a serine residue in the same position (figure 4.3). The serine residue in Lichostatinal interacts with the oxygen on threonine 42 mediated through hydrogen bonding with water (table 3.3). Substituting the serine with

arginine in the S3 subsite as seen in Antipain would result in different interactions. The arginine would not fit into the subsite pocket with the same orientation as serine. This could lead to a more favorable orientation. Additionally the nitrogens in the P3 arginine would not need the water molecule to mediate the interaction with threonine 42. The guanidinium side chain of arginine would also extend into a negatively charged environment on the surface of cathepsin K, resulting in favorable interactions for binding. More important for the high affinity of Lichostatinal, however, is the S4P4 interaction. The P4 residue in Lichostatinal is a positively charged agmatine residue (decarboxylated arginine) in contrast to the the aromatic and hydrophobic phenylalanine residue in Vince-4. Based on the substrate library results, an arginine is about 70% more effective than a phenylalanine residue (Figure 4.3). Placing agmatine as the P4 residue allows for interactions between glutamic acid 112 and aspartic acid 158 with the terminal nitrogens in agmatine, both mediated through hydrogen bonding with a water molecule. Replacing agmatine with phenylalanine however would not allow for these interactions. Phenylalanine would not reach as far as agmatine eliminating the potential interactions. In addition, the bulkiness of the phenylalanine residue would not fit as well into the S4 subsite as well as the lack of charged groups eliminating the potential of hydrogen bonding interactions. Finally, the guanidinium group in agmatine would be extended out into a negatively charged environment, one which would not be possible with phenylalanine. Using these conclusions, Lichostatinal identified through X-ray crystallography would form a more stable complex with the protein.
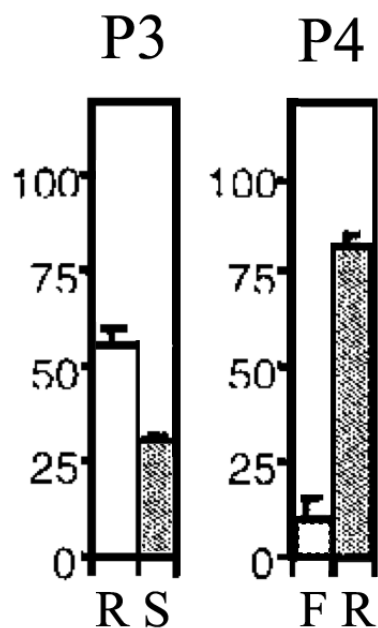
Figure 4.3 Substrate Specificity of Cathepsin K S3 and S4 Subsite

Subsite specificity of cathepsin in the S3 and S4 subsite determined through screening of synthetic substrate library hydrolysis [34].

4.6 Agmatine

The fourth residue in Lichostatinal, agmatine (4-(aminobutyl)guanidine) is the decarboxylation product of L-arginine. This reaction is likely performed by the enzyme, arginine decarboxylase [85]. Arginine decarboxylase (ADC) is highly conserved in nature, found in bacteria, plants and invertebrates [86]. In mammals, agmatine is synthesized in the brain where it is stored in synaptic vesicles. Agmatine has been shown to bind α_2-adrenergic receptors and imidazoline binding sites in addition to having affinity for nicotinic receptors. It can also block NMDA receptors and induce the release of some peptide hormones (luteinizing hormone) [87]. In bacteria, ADC is expressed cytosolically and its activity is dependant on pyridoxal phosphate and magnesium^{2+} [86].
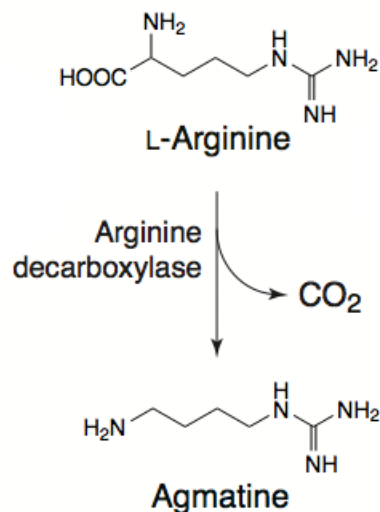
Figure 4.4 Formation of Arginine Decarboxylase Product, Agmatine

Agmatine, the arginine derivative is produced ubiquitously throughout nature by the action of the enzyme arginine decarboxylase resulting in the release of CO_2.

4.7 Kinetics

By looking at the structures of Vince - 2 and 4, it is obvious that Vince – 4 would be a better competitive inhibitor due to the presence of the aldehyde in Vince - 4 allowing for covalent inhibition. This was confirmed from the Dixon plots with K_i values of 125nM (Vince -4) versus K_i value of 295nM (Vince – 2) and 258nM (Vince – 1). It is surprising that although Vince – 4 is structurally identical to Antipain (K_i 41nM) they did not measure the same K_i values. One explanation for this comes from compound stability. Having run through multiple steps of purification along with experimental techniques to establish a successful purification, the integrity of the structure of Vince – 4 would be more questionable than that of commercially available Antipain. Over time the aldehyde in the structure of Vince – 4 would be oxidized producing the carboxylic acid derivative or reduced to an alcohol, resulting in a significantly less potent cathepsin K inhibitor.

Without the presence of an aldehyde, the K_i values of Vince – 1 and Vince – 2 are expected to be much larger. Another explanation could be that although sample was lyophilized, there could be some contaminating salts or humidity, which makes the assumed concentration of inhibitors higher than it is.

As far as potential therapeutic agents for the inhibition of cathepsin K, all compounds listed above proved to be only moderate inhibitors when compared to other previously identified inhibitors. For example reversible tripeptidyl aldehydes have been found to inhibit cathepsins with K_i values of 1.4nM and irreversible inhibitors have been measured to high picomolar values [4].

L-91-3 purified compounds along with Antipain were found to be non-specific inhibitors of cysteine proteases as shown through their ability to inhibit papain and trypsin. This is unfavorable because non-specific inhibitors can lead to numerous cellular disorders if a wide range of cathepsins along with other proteases within the cell are rendered inactive. This is particularly important since off target proteases are found in the lysosome, and lysosomal accumulation of non-specific inhibitors would be in contact with all of them [73]. For example, antipain has been shown to inhibit calpain I and calpain II (calcium dependant cysteine proteases) with K_i values of 1.41 and 1.45 µM respectively [88]. Due to the aldehyde in antipain, it becomes a target for any protease requiring a nucleophilic attack resulting in inhibition through covalent binding (figure 4.3). Unfortunately this means that Antipain will inhibit all cysteine proteases as well as serine and threonine proteases provided it fits into the binding cleft [88-90]. Most inhibitors that have been

isolated or developed tend to be two to four amino acids in length, or at least fit that volume and interact with the non-prime subsites of cathepsin K. To mimic the mechanism of amide bond hydrolysis, most inhibitors incorporate an electrophilic isostere (aldehyde, α-ketoamide, ketone or vinyl sulfones) to block the enzymes from cleavage [91]. Other cathepsin K inhibitors have been tested in clinical trials, Balicatib reached phase II however has since been discontinued because side effects involving skin rash and morphae like skin changes were occurring in test patients suggesting that the compound was affecting more targets than only cathepsin K in osteoclast [73].
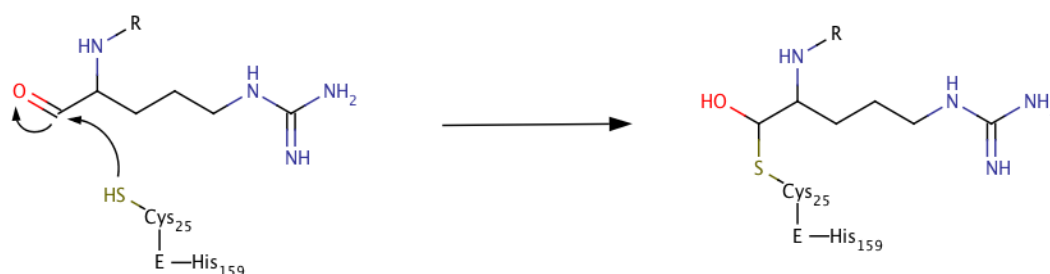
Figure 4.5 Mechanism of Inhibition of Cathepsin K by Antipain

The nucleophilic cysteine 25 of Cathepsin K attacks Antipain forming a tetrahedral intermediate, however, while deacylation occurs when binding to a peptide bond, the aldehyde of Antipain prevents this reaction rendering cathepsin K inhibited.

4.8 Necessity of Protease Inhibitors in Science

Although Antipain proved not to be a useful compound in the treatment to osteoporosis, the identification of novel protease inhibitors has its value. As was demonstrated with Balicatib, discovering a novel inhibitor which is not selective can still provide some insight into potential therapeutics. Following the discovery of Balicatib, Odanacatib

(figure 1.6) which is structurally similar Balicatib was synthesized with a substituted P1 site and modified residue in the P2 site, resulting in a inhibitor with similar potency, however is more selective which has allowed it to reach phase III clinical trials [73].

Protease inhibitors are used for research in the lab on a daily basis. One barrier scientist face is the instability of proteins. Proteases are ubiquitous in all cells, and when purifying proteins, cells are lysed releasing all the proteases which had been compartmentalized into organelles and quickly start to degrade protein. Protease inhibitor cocktails, containing antipain, leupeptin, E-64, pepstatin or a number of other inhibitors are routinely used to eliminate unwanted proteolysis of proteins during purifications [92]. Therefore although a therapeutic use may not be derived from the isolation of novel inhibitors, they can still serve a purpose in research.

5 Conclusion and Further Studies

Three compounds were structurally characterized from L-91-3 *streptomyces* bacteria. The first compound, Vince – 2, is an Antipain-related compound where the P1 arginine residues formed a ring structure containing one unsaturated double bond. The second compound, Vince-4, is the known Antipain, and third, Lichostatinal, is a novel peptide aldehyde inhibitor. Similar to Antipain, Lichostatinal is a tetrapeptide derivative but is differentiated from Antipain by its P3 and P4 residues. Vince 2 and 4 were structurally characterized by NMR whereas the structure of Lichostatinal was solved by X-ray crystallography as a bound inhibitor in the active site of cathepsin K. Vince – 1 was isolated, however, the NMR data were inconclusive. Vince 2 and 4 experimentally had Ki values of 295nM and 125nM respectively compared to 41nM for commercially bought Antipain.

Having purified the two compounds Vince – 2 and Antipain (Vince – 4) from *streptomyces* strain L-91-3, the next objective would be to use these pure samples in X-ray crystallography studies attempting to co-crystallize the pure inhibitors with cathepsin K to identify the binding pocket interactions that are taking place.

Finally, the origin of Lichostatinal identified through X-ray crystallography is still unknown. It is hypothesized that the compound eluted from HPLC early was hidden beneath the unresolved peaks. Future efforts should be focused on purifying larger amounts of L-91-3 inhibitors in order to have sufficient material to expand on the

purification in order to resolve the initial peaks on HPLC to confirm the structure identified from the crystal structure and to perform kinetics on the compound to identify its mode of inhibition as well as Ki value. Alternatively, efforts should be made to synthesize Lichostatinal and analyse its inhibitor kinetics in more detail.

References

1. Voet, D. and J.G. Voet, *Biochemistry*. 2nd ed. ed. 1995, New York ; Chichester: J. Wiley & Sons. xvii,1361p.
2. David L. Nelson, M.M.C., *Principles of Biochmistry 4th Edition*. 4 ed. Lehninger. 2004: W. H. Freeman. 1100.
3. Southan, C., *A genomic perspective on human proteases as drug targets*. Drug Discov Today, 2001. **6**(13): p. 681-688.
4. Lecaille, F., J. Kaleta, and D. Bromme, *Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design*. Chem Rev, 2002. **102**(12): p. 4459-88.
5. Buhling, F., et al., *Review: novel cysteine proteases of the papain family*. Adv Exp Med Biol, 2000. **477**: p. 241-54.
6. Samarel, A.M., et al., *Effects of cysteine protease inhibitors on rabbit cathepsin D maturation*. Am J Physiol, 1989. **257**(6 Pt 1): p. C1069-79.
7. Riese, R.J., et al., *Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading*. Immunity, 1996. **4**(4): p. 357-66.
8. Docherty, K., R.J. Carroll, and D.F. Steiner, *Conversion of proinsulin to insulin: involvement of a 31,500 molecular weight thiol protease*. Proc Natl Acad Sci U S A, 1982. **79**(15): p. 4613-7.
9. Gelb, B.D., et al., *Cathepsin K: isolation and characterization of the murine cDNA and genomic sequence, the homologue of the human pycnodysostosis gene*. Biochem Mol Med, 1996. **59**(2): p. 200-6.
10. Katunuma, N. and E. Kominami, *Abnormal expression of lysosomal cysteine proteinases in muscle wasting diseases*. Rev Physiol Biochem Pharmacol, 1987. **108**: p. 1-20.
11. Bromme, D. and J. Kaleta, *Thiol-dependent cathepsins: pathophysiological implications and recent advances in inhibitor design*. Curr Pharm Des, 2002. **8**(18): p. 1639-58.
12. Sloane, B.F. and K.V. Honn, *Cysteine proteinases and metastasis*. Cancer Metastasis Rev, 1984. **3**(3): p. 249-63.
13. Lemere, C.A., et al., *The lysosomal cysteine protease, cathepsin S, is increased in Alzheimer's disease and Down syndrome brain. An immunocytochemical study*. Am J Pathol, 1995. **146**(4): p. 848-60.
14. Rieman, D.J., et al., *Biosynthesis and processing of cathepsin K in cultured human osteoclasts*. Bone, 2001. **28**(3): p. 282-9.
15. Kornfeld, S., *Trafficking of lysosomal enzymes*. FASEB J, 1987. **1**(6): p. 462-8.
16. McQueney, M.S., et al., *Autocatalytic activation of human cathepsin K*. J Biol Chem, 1997. **272**(21): p. 13955-60.
17. Lecaille, F., D. Bromme, and G. Lalmanach, *Biochemical properties and regulation of cathepsin K activity*. Biochimie, 2008. **90**(2): p. 208-26.
18. Gonzalez-Noriega, A., et al., *Chloroquine inhibits lysosomal enzyme pinocytosis and enhances lysosomal enzyme secretion by impairing receptor recycling*. J Cell Biol, 1980. **85**(3): p. 839-52.

19.   Wiederanders, B., *The function of propeptide domains of cysteine proteinases.* Adv Exp Med Biol, 2000. **477**: p. 261-70.

20.   Yasuda, Y., J. Kaleta, and D. Bromme, *The role of cathepsins in osteoporosis and arthritis: rationale for the design of new therapeutics.* Adv Drug Deliv Rev, 2005. **57**(7): p. 973-93.

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

22.   Turk, B., D. Turk, and V. Turk, *Lysosomal cysteine proteases: more than scavengers.* Biochim Biophys Acta, 2000. **1477**(1-2): p. 98-111.

23.   Menard, R., et al., *Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain.* Biochemistry, 1991. **30**(37): p. 8924-8.

24.   Menard, R. and A.C. Storer, *Oxyanion hole interactions in serine and cysteine proteases.* Biol Chem Hoppe Seyler, 1992. **373**(7): p. 393-400.

25.   Linnevers, C.J., et al., *Expression of human cathepsin K in Pichia pastoris and preliminary crystallographic studies of an inhibitor complex.* Protein Sci, 1997. **6**(4): p. 919-21.

26.   Hummel, K.M., et al., *Cysteine proteinase cathepsin K mRNA is expressed in synovium of patients with rheumatoid arthritis and is detected at sites of synovial bone destruction.* J Rheumatol, 1998. **25**(10): p. 1887-94.

27.   Fred H. Drake, R.A.D., Ian E. James, Janice R. Connor, Christine Debouck, Susan Richardson, Elizabeth Lee-Rykaczewski, Lindsay Coleman, David Rieman, Ray Barthlow, Gregg Hastings, and Maxine Gowen, *Cathepsin K, but Not Cathepsins B, L, or S, Is Abundantly Expressed in Human Osteoclasts.* THE JOURNAL OF BIOLOGICAL CHEMISTRY, 1996. **271**(No 21): p. 12511-12516.

28.   Bossard, M.J., et al., *Proteolytic activity of human osteoclast cathepsin K. Expression, purification, activation, and substrate identification.* J Biol Chem, 1996. **271**(21): p. 12517-24.

29.   Everts, V., W. Beertsen, and R. Schroder, *Effects of the proteinase inhibitors leupeptin and E-64 on osteoclastic bone resorption.* Calcif Tissue Int, 1988. **43**(3): p. 172-8.

30.   Saftig, P., et al., *Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice.* Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13453-8.

31.   Kiviranta, R., et al., *Accelerated turnover of metaphyseal trabecular bone in mice overexpressing cathepsin K.* J Bone Miner Res, 2001. **16**(8): p. 1444-52.

32.   Polewski, M.D., et al., *Inorganic pyrophosphatase induces type I collagen in osteoblasts.* Bone, 2010. **46**(1): p. 81-90.

33.   Marcus, R., *Osteoporosis.* 3rd ed. / edited by Robert Marcus ... [et al.]. ed. 2008, Amsterdam ; London: Elsevier. 2 v. (xxiv, 1941 p., [18] p. of plates).

34.   Lecaille, F., et al., *Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity.* Biochemistry, 2002. **41**(26): p. 8447-54.

35.   Melo, R.L., et al., *Synthesis and hydrolysis by cysteine and serine proteases of short internally quenched fluorogenic peptides.* Analytical Biochemistry, 2001. **293**(1): p. 71-7.

36.   Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain.* Biochem Biophys Res Commun, 1967. **27**(2): p. 157-62.

37.   Delaisse, J.M., et al., *Proteinases in bone resorption: obvious and less obvious roles.* Clin Chim Acta, 2000. **291**(2): p. 223-34.

38.   Cashman, K.D., *Calcium intake, calcium bioavailability and bone health.* Br J Nutr, 2002. **87 Suppl 2**: p. S169-77.

39.   Lazner, F., et al., *Osteopetrosis and osteoporosis: two sides of the same coin.* Hum Mol Genet, 1999. **8**(10): p. 1839-46.

40.   Stoch, S.A. and J.A. Wagner, *Cathepsin K inhibitors: a novel target for osteoporosis therapy.* Clin Pharmacol Ther, 2008. **83**(1): p. 172-6.

41.   Raisz, L.G., *Physiology and pathophysiology of bone remodeling.* Clin Chem, 1999. **45**(8 Pt 2): p. 1353-8.

42.   de Crombrugghe, B., V. Lefebvre, and K. Nakashima, *Regulatory mechanisms in the pathways of cartilage and bone formation.* Curr Opin Cell Biol, 2001. **13**(6): p. 721-7.

43.   *Consensus development conference: prophylaxis and treatment of osteoporosis.* Am J Med, 1991. **90**(1): p. 107-10.

44.   Ministry-of-Health. *Seniors and Aging - Osteoporosis*. 2007; Available from: http://www.hc-sc.gc.ca/hl-vs/iyh-vsv/diseases-maladies/seniors-aines-ost-eng.php.

45.   Osteoporosis-Canada. *Osteoporosis at a Glance*. 2010 [cited 2010; Available from: http://www.osteoporosis.ca/index.php/ci_id/5526/la_id/1.htm.

46.   Teitelbaum, S.L. and F.P. Ross, *Genetic regulation of osteoclast development and function.* Nat Rev Genet, 2003. **4**(8): p. 638-49.

47.   LEO-Pharmaceutical-Product-Sarath. *What is Osteoporosis*. 2010 [cited 2010 September 27th 2010]; Available from: http://www.leo-sarath.com/index.asp?a_id=546.

48.   Vaananen, K., *Mechanism of osteoclast mediated bone resorption--rationale for the design of new therapeutics.* Adv Drug Deliv Rev, 2005. **57**(7): p. 959-71.

49.   Adami, S., et al., *Vitamin D status and response to treatment in post-menopausal osteoporosis.* Osteoporos Int, 2009. **20**(2): p. 239-44.

50.   Cranney, A., et al., *Parathyroid hormone for the treatment of osteoporosis: a systematic review.* CMAJ, 2006. **175**(1): p. 52-9.

51.   Krum, S.A. and M. Brown, *Unraveling estrogen action in osteoporosis.* Cell Cycle, 2008. **7**(10): p. 1348-52.

52.   Lindsay, R. and J.F. Tohme, *Estrogen treatment of patients with established postmenopausal osteoporosis.* Obstet Gynecol, 1990. **76**(2): p. 290-5.

53.   Riggs, B.L. and L.C. Hartmann, *Selective estrogen-receptor modulators -- mechanisms of action and application to clinical practice.* N Engl J Med, 2003. **348**(7): p. 618-29.

54.   Balaburski, G.M., et al., *Raloxifene-stimulated experimental breast cancer with the paradoxical actions of estrogen to promote or prevent tumor growth: a unifying concept in anti-hormone resistance.* Int J Oncol, 2010. **37**(2): p. 387-98.

55.   Marquis, R.W., *Inhibition of the cysteine protease cathepsin K*, in *Annual Reports in Medicinal Chemistry*. 2004, GlaxoSmithKline: Collegeville PA. p. 79-98.

56.   Visentin, L., et al., *A selective inhibitor of the osteoclastic V-H(+)-ATPase prevents bone loss in both thyroparathyroidectomized and ovariectomized rats.* J Clin Invest, 2000. **106**(2): p. 309-18.

57.   Aoyagi, T., et al., *Leupeptins, new protease inhibitors from Actinomycetes.* J Antibiot (Tokyo), 1969. **22**(6): p. 283-6.

58.   Umezawa, H., et al., *Chymostatin, a new chymotrypsin inhibitor produced by actinomycetes.* J Antibiot (Tokyo), 1970. **23**(8): p. 425-7.

59.   Suda, H., et al., *Antipain, a new protease inhibitor isolated from actinomycetes.* J Antibiot (Tokyo), 1972. **25**(4): p. 263-6.

60.   T.SHIN-WATANABE, S.M., *Purification and characterization of crystalline microbial alkaline proteinase inhibitors (MAPI), produced by Streptomyces nigrescens WT-27.* Agricultural and Biological Chemistry, 1979. **43**: p. 243-250.

61.   Thompson, S.K., et al., *Design of potent and selective human cathepsin K inhibitors that span the active site.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14249-54.

62.   Hashida, S., E. Kominami, and N. Katunuma, *Inhibitions of cathepsin B and cathepsin L by E-64 in vivo. II. Incorporation of [3H]E-64 into rat liver lysosomes in vivo.* J Biochem, 1982. **91**(4): p. 1373-80.

63.   Bromme, D., et al., *Peptidyl vinyl sulphones: a new class of potent and selective cysteine protease inhibitors: S2P2 specificity of human cathepsin O2 in comparison with cathepsins S and L.* Biochemical Journal, 1996. **315 (Pt 1)**: p. 85-9.

64.   Bromme, D. and F. Lecaille, *Cathepsin K inhibitors for osteoporosis and potential off-target effects.* Expert Opin Investig Drugs, 2009. **18**(5): p. 585-600.

65.   Falgueyret, J.P., et al., *Novel, nonpeptidic cyanamides as potent and reversible inhibitors of human cathepsins K and L.* J Med Chem, 2001. **44**(1): p. 94-104.

66.   Marquis, R.W., et al., *Cyclic ketone inhibitors of the cysteine protease cathepsin K.* J Med Chem, 2001. **44**(5): p. 725-36.

67.   Marquis, R.W., et al., *Conformationally constrained 1,3-diamino ketones: a series of potent inhibitors of the cysteine protease cathepsin K.* J Med Chem, 1998. **41**(19): p. 3563-7.

68.   Barrett, D.G., et al., *Novel, potent P2-P3 pyrrolidine derivatives of ketoamide-based cathepsin K inhibitors.* Bioorg Med Chem Lett, 2006. **16**(6): p. 1735-9.

69.   Chappard, D., et al., *The cathepsin K inhibitor AAE581 induces morphological changes in osteoclasts of treated patients.* Microsc Res Tech, 2010. **73**(7): p. 726-32.

70.   Turk, B., *Targeting proteases: successes, failures and future prospects.* Nat Rev Drug Discov, 2006. **5**(9): p. 785-99.

71.   Falgueyret, J.P., et al., *Lysosomotropism of basic cathepsin K inhibitors contributes to increased cellular potencies against off-target cathepsins and reduced functional selectivity.* J Med Chem, 2005. **48**(24): p. 7535-43.

72.   Kumar, S., et al., *A highly potent inhibitor of cathepsin K (relacatib) reduces biomarkers of bone resorption both in vitro and in an acute model of elevated bone turnover in vivo in monkeys.* Bone, 2007. **40**(1): p. 122-31.

73.   Gauthier, J.Y., et al., *The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K.* Bioorg Med Chem Lett, 2008. **18**(3): p. 923-8.

74.   Li, J.W. and J.C. Vederas, *Drug discovery and natural products: end of an era or an endless frontier?* Science, 2009. **325**(5937): p. 161-5.

75.   Michael G. Surette, J.E.D., *A New Look A Secondary Metabolites*, in *Chemical Communication Among Bacteria*, S.C.W.a.B.L. Bassler, Editor. 2008, American Society Microbiolgy. p. 483.

76.   Kieser, T., *Practical streptomyces genetics*. 2000, Norwich: John Innes Foundation. 613 p.

77.   Kondo, S.I., et al., *Isolation and characterization of leupeptins produced by Actinomycetes.* Chem Pharm Bull (Tokyo), 1969. **17**(9): p. 1896-901.

78.   Stella, S., et al., *Isolation of alpha-MAPI from fermentation broths during a screening program for HIV-1 protease inhibitors*. J Antibiot (Tokyo), 1991. **44**(9): p. 1019-22.

79.   Sigma-Aldrich. *SigmaFast Protease Inhibitor Tablets*. 2010 12/06/10]; Available from: http://www.sigmaaldrich.com/catalog/ProductDetail.do?N4=S8820%7CSIGMA&N5=SEARCH_CONCAT_PNO%7CBRAND_KEY&F=SPEC.

80.   Newman, D.J., G.M. Cragg, and K.M. Snader, *The influence of natural products upon drug discovery.* Nat Prod Rep, 2000. **17**(3): p. 215-34.

81.   Rao, S.R. and G.A. Ravishankar, *Plant cell cultures: Chemical factories of secondary metabolites.* Biotechnol Adv, 2002. **20**(2): p. 101-53.

82.   Vining, L.C., *Functions of secondary metabolites.* Annu Rev Microbiol, 1990. **44**: p. 395-427.

83.   Umezawa, S., et al., *Structure of antipain, a new Sakaguchi-positive product of streptomyces.* J Antibiot (Tokyo), 1972. **25**(4): p. 267-70.

84.   Ortiz, C., et al., *Diastereotopic covalent binding of the natural inhibitor leupeptin to trypsin: detection of two interconverting hemiacetals by solution and solid-state NMR spectroscopy.* Biochemistry, 1991. **30**(41): p. 10026-34.

85.   Galea, E., et al., *Inhibition of mammalian nitric oxide synthases by agmatine, an endogenous polyamine formed by decarboxylation of arginine.* Biochemical Journal, 1996. **316 (Pt 1)**: p. 247-9.

86.   Reis, D.J. and S. Regunathan, *Is agmatine a novel neurotransmitter in brain?* Trends Pharmacol Sci, 2000. **21**(5): p. 187-93.

87.   Piletz, J.E., H. Zhu, and D.N. Chikkala, *Comparison of ligand binding affinities at human I1-imidazoline binding sites and the high affinity state of alpha-2 adrenoceptor subtypes.* J Pharmacol Exp Ther, 1996. **279**(2): p. 694-702.

88.   Sasaki, T., et al., *Comparative specificity and kinetic studies on porcine calpain I and calpain II with naturally occurring peptides and synthetic fluorogenic substrates.* J Biol Chem, 1984. **259**(20): p. 12489-94.

89.   McGowan, E.B., S.A. Shafiq, and A. Stracher, *Delayed degeneration of dystrophic and normal muscle cell cultures treated with pepstatin, leupeptin, and antipain.* Exp Neurol, 1976. **50**(3): p. 649-57.

90.   Baird, T.T., Jr., W.D. Wright, and C.S. Craik, *Conversion of trypsin to a functional threonine protease.* Protein Sci, 2006. **15**(6): p. 1229-38.

91.   Abbenante, G. and D.P. Fairlie, *Protease inhibitors in the clinic.* Med Chem, 2005. **1**(1): p. 71-104.

92.   Lockwood, B.C., et al., *The use of a highly sensitive electrophoretic method to compare the proteinases of trichomonads.* Mol Biochem Parasitol, 1987. **24**(1): p. 89-95.

Appendix

## *Appendix A Unsuccessful Approach Used in Inhibitor Purification*

The investigation into creating a potential affinity column with papain began by assuring that papain was inhibited by L-91-3. L-91-3 having roughly 2/3rds the inhibitory potency towards papain as it does for cathepsin K identified through an enzymatic assay (as described in section 2.2) made papain an ideal candidate. The premise was to incubate L-91-3 with papain followed by filtration through a 10 kD filter thus eliminating impurities, however it was discovered that L-91-3 bound papain reversibly and the compounds of interest simply flowed through the filter. This was surprising since structurally it would be assumed that the compounds would bind cathepsin K covalently.

Appendix B Additional NMR Spectrometry Data

Structural determination of Vince – 2 and Antipain was accomplished through the use of NMR data presented in figures 3.11 – 3.14 in to that presented in figures A1-A6 and outlined in Table 3.1 and 3.1. Structural determination was performed by Dr. David Williams (Dr. Raymond Andersen Laboratory).
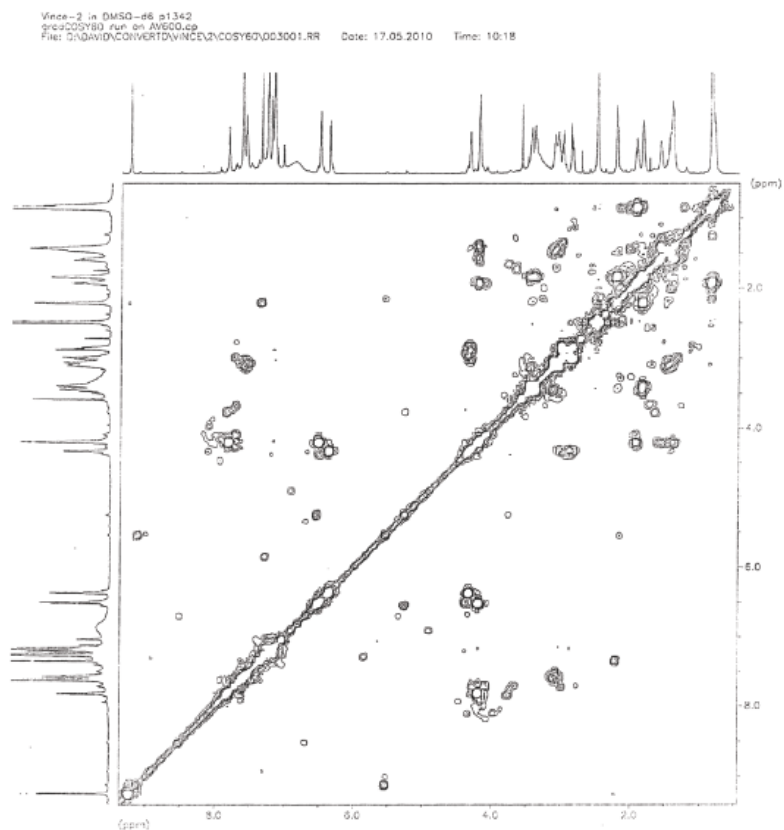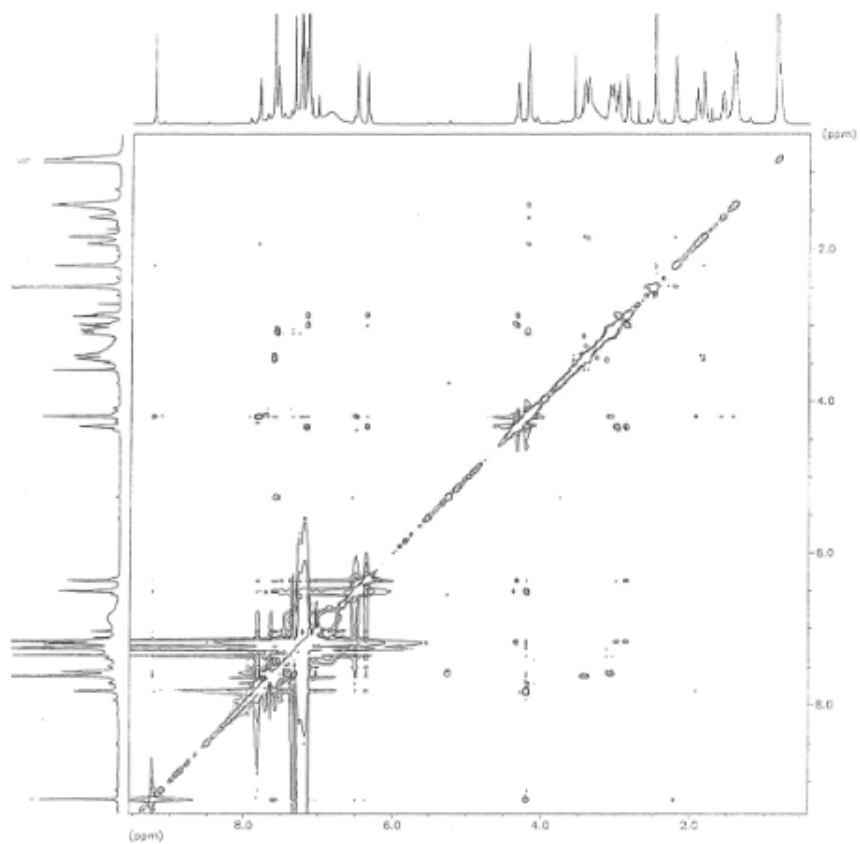
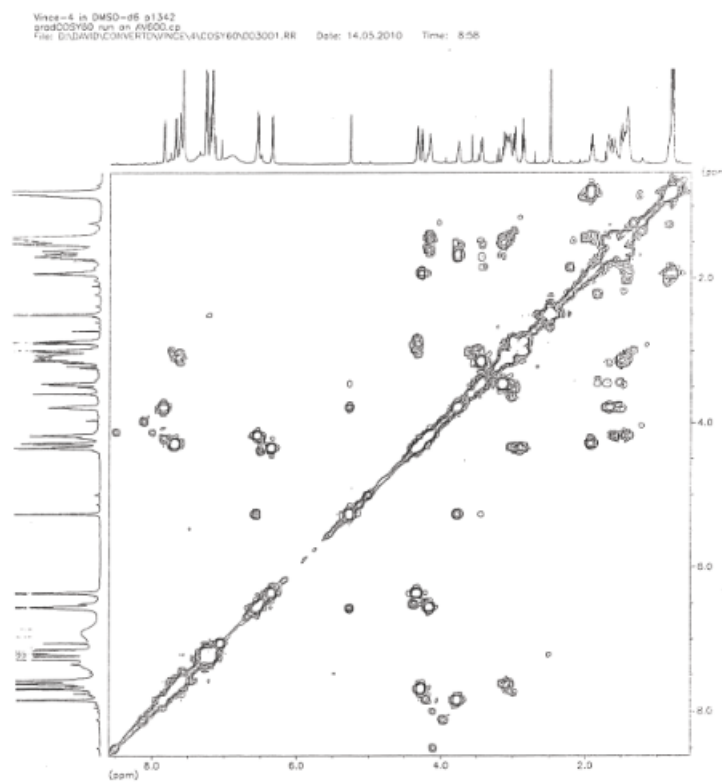**Figure A.1 COSY – 60 of Vince – 2 in DMSO-d6 at 600MHz**
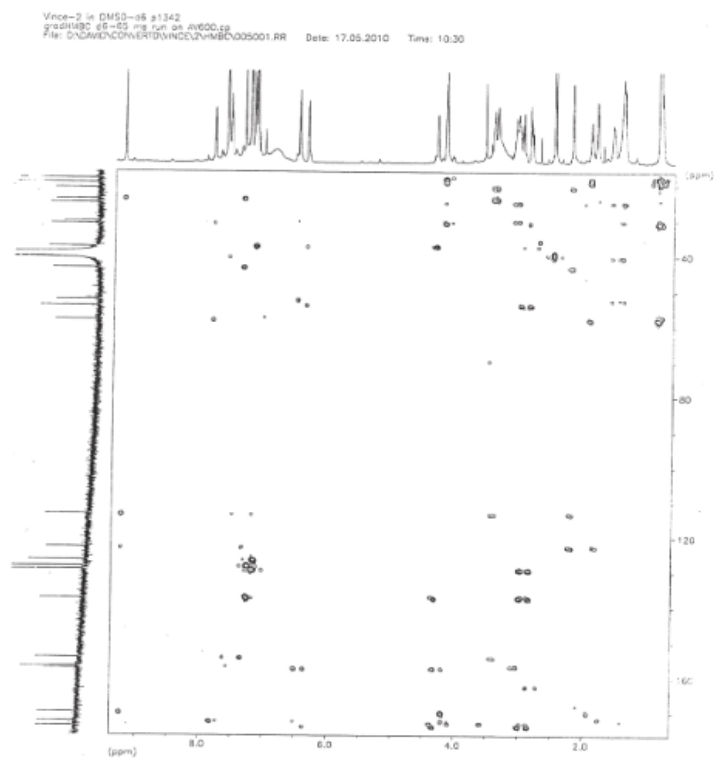
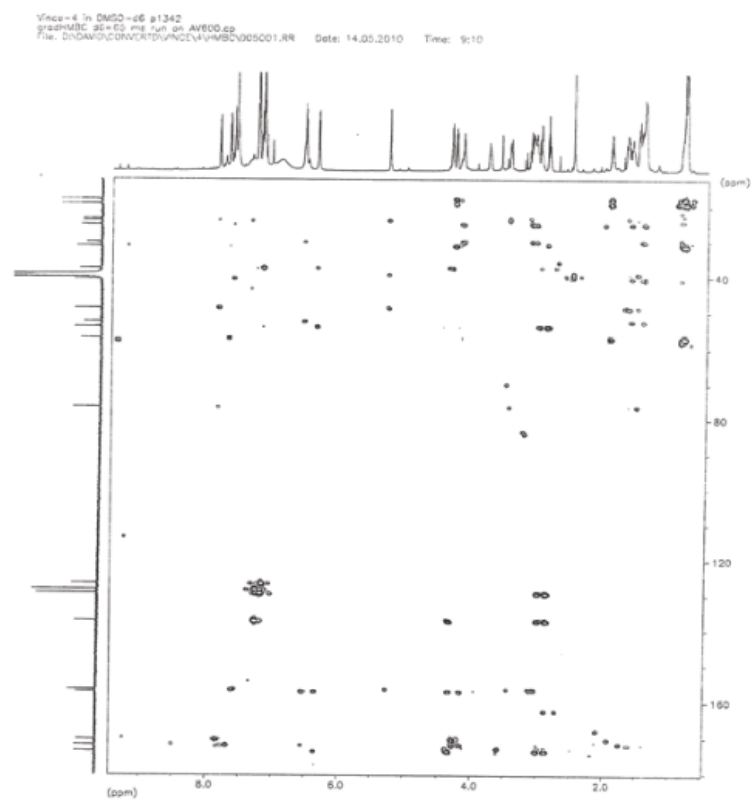**Figure A.2 HMBC of Vince – 2 in DMSO-d6 at 600MHz**
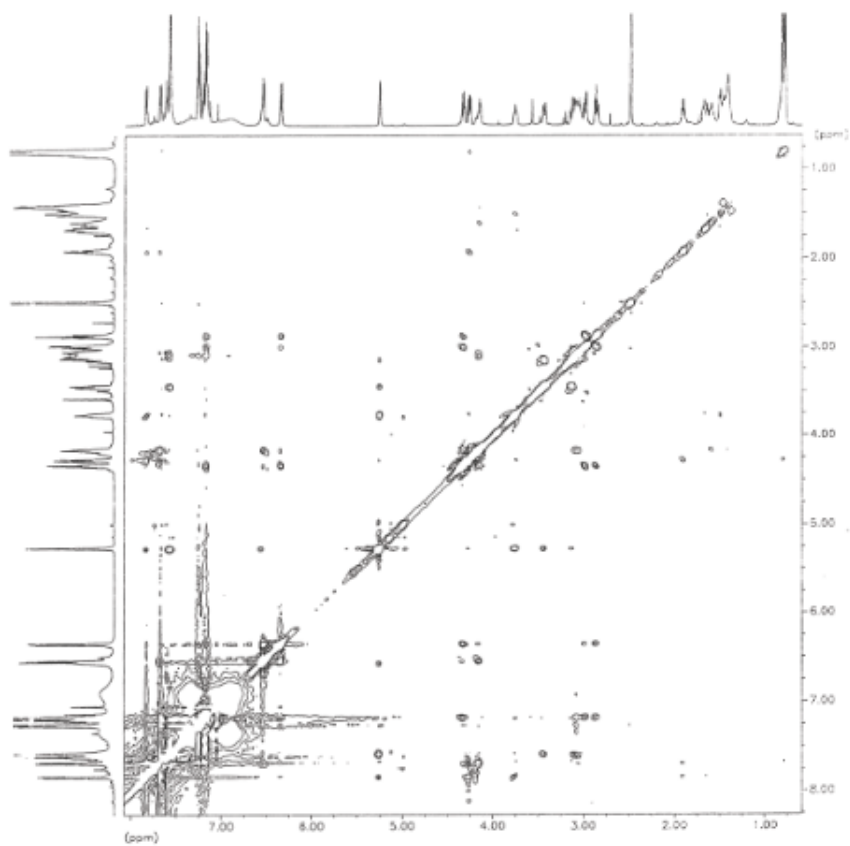
Figure A.3 tROESY of Vince – 2 in DMSO-d6 at 600MHz

Figure A.4 COSY-60 of Vince – 4 in DMSO-d6 at 600MHz

Figure A.5 HMBC of Vince – 4 in DMSO-d6 at 600MHZ

Figure A.6 tROESY of Vince – 4 in DMSO-d6 at 600MHz