STRUCTURAL STUDIES OF CYSTEINE AND SERINE PROTEASE INHIBITORS TOWARDS THERAPEUTIC APPLICATIONS

RAJESH TULSIDAS SHENOY
(B.E)
A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES,
FACULTY OF SCIENCE,
NATIONAL UNIVERSITY OF SINGAPORE
June 2009

To my dear parents

Acknowledgements

I am grateful to A/P J Sivaraman who has been my mentor during four and a half years of my PhD course. He has always been approachable and extremely patient with me. He has provided me with a strong footing in protein crystallography and biological research in general. His vast experience in the field of cysteine proteases has nurtured my interest to work on human Cathepsin-L which is one of the important drug development targets. I am thankful to Prof. Ding Jeak Ling, for having provided me the opportunity to work on the exciting topic of serine protease inhibitors in the innate immunity of the horseshoe crab which has constantly fuelled my passion in my work. I thank Prof RM Kini for his support and encouragement. I would like to thank Prof. Enrico Purisima, Dr. Shafinaz Chowdhury, Dr. Adrian Velazquez for their invaluable contribution to the projects. I am grateful to my Lissa Joseph and Dr. Sundramurthy Kumar and who have worked in the projects and made my work enjoyable. I would like to thank Dr Anand Saxena and Dr J Seetharaman who helped me during my data collection at National Synchrotron Light Source, USA. I am grateful to Thangavelu, Pankaj Kumar Giri and Manjeet for helping me with my thesis proof reading and final experiments. I am grateful to A/P K Swaminathan and all members of his lab, especially Dileep Vasudevan, Shiva Kumar, Kuntal Pal, for their support. I am thankful to Dileep. G. Nair and Tzer Fong for their help and support. I would like to thank all my labmates for their help and support especially, Jobichen Chacko and Sunita. Finally I thank National University of Singapore for providing an intellectually stimulating environment and all the resources to make this work possible. I offer my special thanks to my parents, who constantly encouraged me

throughout my life in all my endeavors. Without their support this work would not have

been possible.

Table of Contents

Summary

Proteases play a very important role in a multitude of physiological reactions such as cell signaling, migration, immunological defense, wound healing and apoptosis and are crucial for disease propagation. Of the over 400 known human proteases, around 14% are under investigation as drug targets and the proportion is expected to increase considerably. The study of proteases and protease inhibitors are emerging with promising therapeutic uses. In this study we have selected the cysteine and serine protease inhibitor complexes to understand their inhibition mechanisms. Both proteases share similar catalytic triad (example: Subtilisin Asp32-His64-Ser221; Papain Cys25-His159-Asn175). Further, the nature of oxyanion hole found in cysteine proteases of papain super family is similar to that found in subtilisin. In addition, many of these proteases are secreted as inactive forms called zymogens and subsequently activated by proteolysis, thereby changing the architecture of the active site of the enzyme.

This PhD thesis consists of five chapters. Chapter I deals with the literature survey and general introduction for both cysteine and serine proteases and their inhibitors. Chapter II deals with the inhibitor complex studies with human cathepsin L. Cathepsin L plays a vital role in many pathophysiological conditions including rheumatoid arthritis, tumour invasion and metastasis, bone resorption and remodeling. In this chapter we report a series of noncovalent, reversible propeptide mimic inhibitors of cathepsin L that have been designed to explore additional binding interactions with the S' subsites. The design was based on the previously reported crystal structure that suggested the possibility of engineering increased interactions with the S' subsites. A few representatives of these new inhibitors have been co-crystallized with mature cathepsin L, and the structures have been

solved and refined at 2.2, 2.5, 1.8 and 2.5Å respectively. These four inhibitors were selected to help clarify and elucidate the binding mode of this class of inhibitors. Of particular interest was the disposition of the biphenyl groups in the S' subsites of the enzyme since the addition of a second biphenyl group to the inhibitor does not improve potency. These inhibitors described in this work extend farther into the S' subsites of cathepsins than any inhibitors reported in the literature thus far. These interactions appear to make use of a S3' subsite that can potentially be exploited for enhanced specificity and/or affinity. In collaboration with Prof Enrico McGill University, Canada, we have also carried out molecular dynamics simulations to supplement the information provided by the crystal structure studies and to explore the dynamical behavior of these inhibitors in the active site.

Chapter III reports the crystal structure of two covalent dipeptidyl inhibitors in complex with human Cathepsin-L. These two inhibitors have different groups in S1 subsite such as inhibitor **1** with α-keto-β-aldehyde and inhibitor **2** with diazomethylketone. The inhibitor **1** Z-Phe-Tyr (OBut)-COCHO has a Ki=0.6 nM. It is the most potent, synthetic peptidyl reversible inhibitor of cathepsin L reported to date. The co-crystal structure of this inhibitor with human Cathepsin-L was refined upto 2.2Å resolution. The inhibitor **2** Z-FY (*t*-Bu)-DMK, an irreversible inhibitor which inactivates cathepsin L at μM concentrations has been co-crystallized and the structure refined upto 1.7Å resolution. These inhibitors have a substrate-like interaction with the active site cysteine. These structural studies, combined with our previous complex structures of Cathepsin L reveal the structural basis for the potency and selectivity of these inhibitors. Our studies on the cathepsin inhibitor complexes

11

have the potential leading to further optimization of these inhibitors towards therapeutic intervention.

Chapter IV deals with serine protease and its inhibitor complex. Serine proteases play a crucial role in host-pathogen interactions. In the innate immune system of invertebrates, multidomain protease inhibitors are important in regulating host-pathogen interactions and antimicrobial activities. Serine protease inhibitors (CrSPI isoforms 1 and 2) of 9.3 kDa were identified from the hepatopancreas of horseshoe crab, *Carcinoscorpius rotundicauda*. The CrSPIs were found to be biochemically active, particularly, the CrSPI-1 which potently inhibits subtilisin (K_i=1.43nM). The CrSPI has been grouped with non-classical Kazal-type inhibitors due to its unusual cysteine distribution. Here we report the 2.6Å resolution crystal structure of CrSPI-1 in complex with subtilisin and its biophysical interaction studies. The CrSPI-1 molecule consists of two domains arranged in an extended conformation. These two domains act as two heads to independently interact with two separate subtilisin molecules resulting in the inhibition of subtilisin activity at a ratio of 1:2 (inhibitor: protease). Each subtilisin molecule interacts with a reactive site loop of CrSPI-1 from each domain through a standard canonical binding mode. We show how such multidomain inhibitors can simultaneously inhibit multiple enzymes within a single ternary complex. In addition we propose the substrate preference of the two domains of CrSPI-1. Domain-2 is more potent and specific towards the bacterial protease subtilisin. Domain-1 is likely to interact with the host protease, Furin, acting like an "on-off" switch in the regulation of host's and pathogen's proteases. The structure of CrSPI-1-subtilisin ternary complex will help to understand the innate immune system at a molecular level. Chapter V provides the overall conclusion and future directions of these projects.

List of Tables

List of figures

17

List of Abbreviations

Å	Angstrom (10^{-10}m)
AEI	Anemonia elastase inhibitor
BME	β mercaptoethanol
Boc	t-butoxycarbonyl
BPTI	Bovine Pancreatic Trypsin Inhibitor
Cbz	carboxybenzyl
CNS	Crystallography and NMR system
CrSPI-1	*Carcinoscorpius rotundicauda* serine protease
DLS	Dynamic Light Scattering
DMK	Diazomethylketone
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EC	Enzyme Commission
EDTA	Ethylenediamine tetraacetic acid
FPLC	Fast performance liquied chromatography gel electrophoresis.
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid inhibitor-1
IPTG	Isopropyl thio-galactoside

ITC   Isothermal Titration Calorimetry

kDa   kilo Dalton

LB   Luria-Bertani

LEKTI   lympho-epithelial Kazal-type inhibitor

MD   Molecular Dynamics

MES   2-(N-morpholino)ethanesulfonic acid

NCS   Non-crystallographic symmetry

Ni-NTA   Nickel nitrilo-triacetic acid.

NMR   Nuclear magnetic resonance

OD   Optical density

OMSVP3   Silver pheasant ovomucoid third domain

OMTKY3   Turkey ovomucoid third domain

PCR   Polymerase chain reaction

PDB   Protein Data Bank

PEG   polyethylene glycol

PSTI   Pancreatic Secretory Trypsin Inhibitor

Rmsd   root mean square deviation

SDS-PAGE   Sodium dodecyl sulphate polyacrilamide

AminoAcids

Gly (G)	glycine
Ala (A)	alanine
Val (V)	valine
Leu (L)	leucine
Ile (I)	isoleucine
Met (M)	methionine
Phe (F)	phenylalanine
Trp (W)	tryptophan
Pro (P)	proline
Ser (S)	serine
Thr (T)	threonine
Cys (C)	cysteine
Tyr (Y)	tyrosine
Asn (N)	asparagine
Gln (Q)	glutamine
Asp (D)	aspartic acid
Glu (E)	glutamic acid
Lys (K)	lysine
Arg (R)	arginine
His (H)	histidine

Publications

1. **Tulsidas, S.R.**, Thangamani, S., Ho, B., Sivaraman, J. and Ding, J.L. (2009). Crystallization of a nonclassical Kazal-type Carcinoscorpius rotundicauda serine protease inhibitor, CrSPI-1, complexed with subtilisin. *Acta Crystallogr Sect F Struct Biol Cryst Commun*., **65**, 533-5.

2. Chowdhury, S.F., Joseph, L., Kumar, S., **Tulsidas, S.R.**, Bhat, S., Ziomek, E., Nard, R.M., Sivaraman, J. and Purisima, E.O. (2008). Exploring inhibitor binding at the S' subsites of cathepsin L. *J Med Chem.,* **51**,1361-8.

3. **Shenoy, R.T.**, Chowdhury, S.F., Kumar, S., Joseph, L., Purisima, E.O., and Sivaraman, J. (2009). A combined crystallographic and molecular dynamics study of cathepsin L retro-binding inhibitors. *J. Med. Chem*., **52**, 6335–6346.

4. **Shenoy, R.T.** and Sivaraman, J. (2010). Structural basis for reversible and irreversible inhibition of human cathepsin L by their respective dipeptidyl glyoxal and diazomethylketone inhibitors. *J Struct Biol*., **173,**14-9.

5. **Shenoy, R.T.,** Thangamani, S., Velazquez-Campoy, A., Ho, B. , Ding, J.L. and Sivaraman, J. (2010). Structural Basis for Dual-inhibition Mechanism of a Non-classical Kazal-type Serine Protease Inhibitor from Horseshoe Crab in Complex with Subtilisin. *PLoS One* (accepted).

6. Husain, N., Tkaczuk, K.L., **Tulsidas, S.R.**, Kaminska, K.H., Cubrilo, S., Maravić-Vlahovicek, G., Bujnicki, J.M., Sivaraman, J. (2010). Structural basis for the methylation of G1405 in 16S rRNA by aminoglycoside resistance methyltransferase Sgm from an antibiotic producer: a diversity of active sites in m7G methyltransferases. *Nucleic Acids Res*., **38**, 4120-32.

Chapter I

General Introduction

Proteases have long been the centre of attention in biochemistry as models of enzyme catalysis, structure and also targets for drug development. They make up for approximately 2% of the gene content in most organisms. It has been estimated that more nearly 800 different proteases are expressed in the human genome. Their abundance in nature is only second to transcription factors. Proteases are involved in a multitude of physiological processes ranging from simple digestion of food to highly regulated cascades like the blood clotting cascade, apoptosis pathways, complement system, and the invertebrate prophenoloxidase activating cascade.

1.1 Classification and Nomenclature of Proteases

There are two systems of classifications for proteases – The IUBMB Nomenclature for Peptidases (EC 3.4) and the MEROPS system. The International Union of Biochemistry and Molecular Biology (IUBMB) have developed a classification for enzymes, the EC numbers. Each enzyme is described by a sequence of four numbers preceded by "EC". The first number broadly classifies the enzyme based on its mechanism. In this, enzymes have been classified as follows: Oxidoreductases (EC 1), Transferases (EC 2), Hydrolases (EC 3), Lyases (EC 4), Isomerases (EC 5) and Ligases (EC 6). Proteases have been assigned to the Hydrolases (EC 3) and further in EC 3.4 - as those acting on peptide bonds (peptidases). In the EC nomenclature, the classification is strictly based upon the type of reaction that an enzyme catalyses and this is the basis for the name of the enzyme. An enzyme is given a unique EC number which can be used as

an unambiguous reference. Additionally, EC also provides a unique *Recommended name* for a protease.

The MEROPS database is an information resource for proteases and proteinaceous inhibitors. The database is a hierarchical, structure-based classification of proteases. In this, each protease is assigned to a Family on the basis of similarities in amino acid sequence, and families that are thought to be homologous are grouped together in a Clan (group of structurally homologs families).

The MEROPS database classifies proteases based on structural similarity. Proteases can be roughly classified as Exopeptidases and Endopeptidases based on where they cleave the substrate proteins (Figure 1.1). Exopeptidases break peptide bonds of terminal amino acids, in contrast to endopeptidases, which break peptide bonds within the molecule. Further exopeptidases can be divided as Aminopeptidases (cleave at the N-terminal end), Carboxypeptidases (cleave at the C-terminal end), Peptidyl Dipeptidases (cleave 2 residues at C terminal), Dipeptidyl peptidases(cleave two residues at N-terminal), and Dipeptidases(cleave dipeptides). The EC system uses this system of classification.

Proteases

Exopeptidases

Aminopeptidases

Dipeptidyl peptidases

Tripeptidyl peptidases

Peptidyl dipeptidases

Carboxypeptidases

Dipeptidases

Endopeptidases

Oligopeptidases
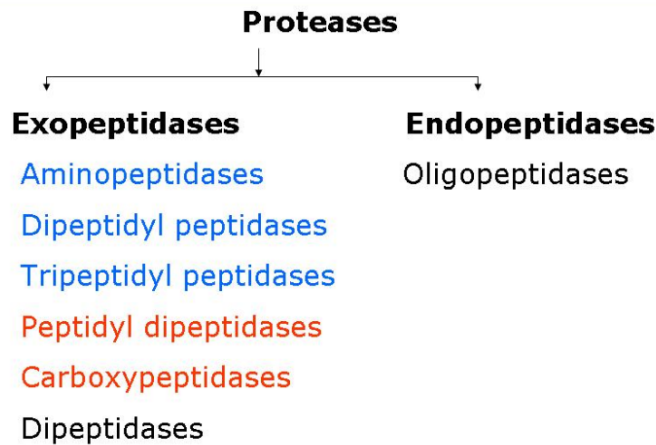
Figure 1.1. Classification of proteases based on cleavage specificity.

1. Catalytic types

| Protein-nucleophile: Ser + Thr + Cys | (Water nucleophile:) Asp | (Water nucleophile:) Met |

2. Structures

| Clans (C, P, S, T) | Clans (A) | Clans (M) |
| Families (C, S, T) | Families (A) | Families (M) |

3. Peptidases

Individual peptidases
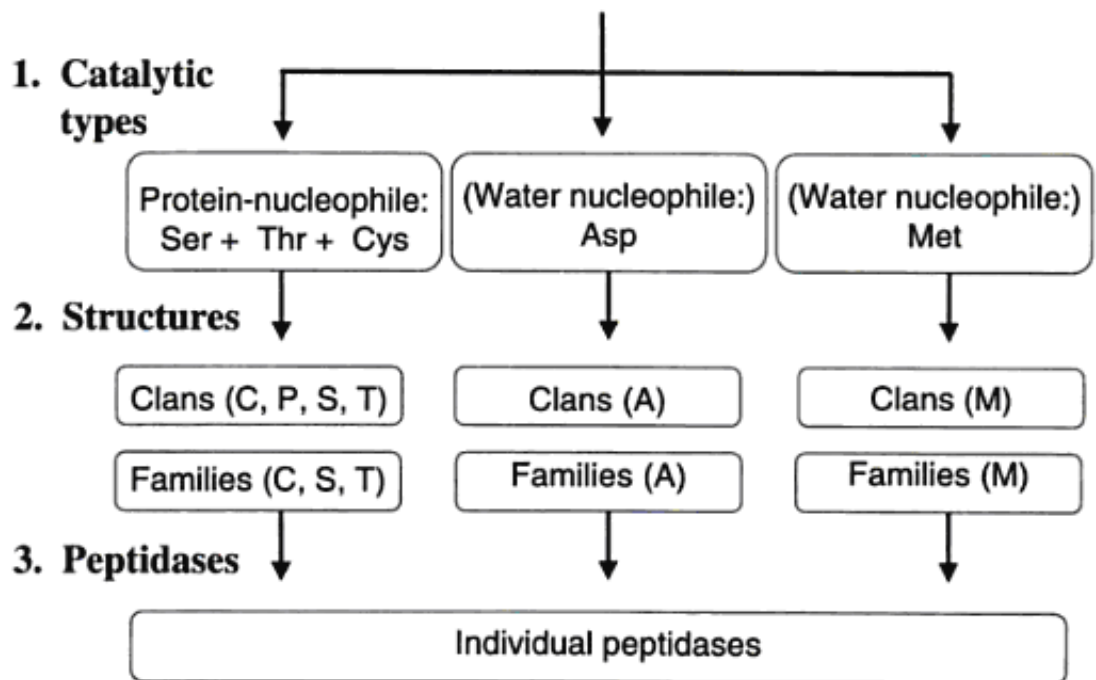
Figure 1.2. The three levels of classification of proteases (This figure was adapted from "Proteases New Perspectives" By Vito Turk, Birkhäuser, 1999)

1.2 Levels of Classification

In the EC and the MEROPS systems, there are three levels of classification for proteases (Fig 1.2). They are based on the (1) Catalytic types, (2) Molecular structures and (3) Peptidases. The following sections briefly discuss the details of each level.

1.2.1 Catalytic types

In both EC and MEROPS systems, the major classification is on the basis of the chemical groups responsible for catalysis i.e. the active site residue. They are **serine, cysteine, aspartic, metallo**, and the recently recognized **threonine** and **glutamic acid** proteases. The threonine and glutamic peptidases were not described until 1995 and 2004, respectively. In the serine, cysteine and threonine peptidases, the nucleophilic attack is carried out by an oxygen or sulphur atom which is part of the protease side chain. In this case, a covalent acyl enzyme intermediate is formed. This is usually hydrolyzed but the more N-terminal of the two products of peptide bond cleavage is transferred from the acyl enzyme to some acceptor other than water. In case of aspartic, metallo and glutamic peptidases, the attacking nucleophile is a water molecule, bound and activated in the catalytic site. The functional groups of the protease do not react directly with the substrate. Hence there are two major catalytic types recognized in the MEROPS database.   The serine, cysteine and threonine peptidases are now classified under the group – ***protein nucleophile*** and the aspartic, metallo and glutamic peptidases are classified under the group – ***water nucleophile***. In MEROPS, database, the name of each clan is represented by a single capital letter representing the catalytic type of the peptidase it contains (S for Serine, T for Threonine, C - Cysteine, A for Aspartic acid, M for Metallo and U where U stands for unknown type) followed by a letter. The name of a

family begins with S, T, C, A, M and U, and ends with a number. The letter P is used for a protein nucleophile clan that contains families of more than one catalytic type. This system gives an unambiguous reference to a particular family of proteases.

In the EC system, carboxypeptidases and endopeptidases have been subdivided into the catalytic types as given in Table 1.1. Carboxypeptidase (EC number 3.4.16 - 3.4.18) is an enzyme that hydrolyzes the carboxy-terminal (C-terminal) end of a peptide bond. Endopeptidases ((EC number 3.4.21 - 3.4.24) are proteases that break peptide bonds of nonterminal amino acids (i.e. within the molecule), in contrast to exopeptidases, which break peptide bonds from their terminal ends.

Sub-Subclass	Kind of Peptidase
EC 3.4.1	α-Amino-Acyl-Peptide Hydrolases
EC 3.4.2	Peptidyl-Amino-Acid Hydrolases
EC 3.4.3	Dipeptide Hydrolases
EC 3.4.4	Peptidyl Peptide Hydrolases
EC 3.4.11	Aminopeptidases
EC 3.4.12	Peptidylamino-Acid Hydrolases or Acylamino-Acid Hydrolases
EC 3.4.13	Dipeptidases
EC 3.4.14	Dipeptidyl-peptidases and tripeptidyl-peptidases
EC 3.4.15	Peptidyl-dipeptidases
EC 3.4.16	Serine-type carboxypeptidases
EC 3.4.17	Metallocarboxypeptidases
EC 3.4.18	Cysteine-type carboxypeptidases
EC 3.4.19	Omega peptidases
EC 3.4.21	Serine endopeptidases
EC 3.4.22	Cysteine endopeptidases
EC 3.4.23	Aspartic endopeptidases
EC 3.4.24	Metalloendopeptidases
EC 3.4.25	Threonine endopeptidases
EC 3.4.99	Endopeptidases of unknown catalytic mechanism

Table 1.1. Classification of proteases according to EC recommendations

1.2.2 Molecular Structures

In the level 2 of protease classification, the MEROPS database has classified peptidases into *Families* and *Clans* based on the amino acid sequence similarity and three-dimensional structure respectively. The database was started in 1993 with all the available sequences of proteases and subsequently the number of families increased. Only the sequence identity of portions of proteinases responsible for enzyme activity is considered for classification of families. Currently the database has 140 families of proteases. While the peptidase families were being established, it was found that it is not possible to strictly group proteases based on amino acid sequences as there was evidence by tertiary structure that some peptidases were related. In evolution, a certain fold is retained although the amino acid sequence may change. Such distantly related families of proteases having similar protein folds were termed as *Clans*. A clan is a group of families that are thought to have common ancestry. There are about 30 clans in the MEROPS database. The clans which consist of protein-nucleophile peptidases of more than one catalytic type are named with a P. The clan of aspartic peptidases is shown as an example in Fig 1.3.

CLAN OF ASPARTIC PEPTIDASES		
CLAN	FAMILY	TYPE PEPTIDASE
AA	A1	subfamily A1A unassigned peptidases (*Paramecium tetraurelia*)
	A2	Gypsy transposon peptidase (*Drosophila virilis*)
	A3	subfamily A3A unassigned peptidases ('Citrus yellow mosaic virus')
	A9	spumapepsin ('simian foamy virus')
	A11	Copia transposon peptidase (*Anopheles gambiae*)
AB	A6	nodavirus peptidase ('black beetle virus')
	A21	tetravirus peptidase ('Providence virus')
AC	A8	signal peptidase II (*Delftia acidovorans*)
AD	A22	subfamily A22B unassigned peptidases (*Paramecium tetraurelia*)
	A24	type 4 prepilin peptidase 1 (*Delftia acidovorans*)
AE	A25	family M63 non-peptidase homologues (*Carboxydothermus hydrogenoformans*)
	A31	HyaD peptidase (*Burkholderia phymatum*)
AF	A26	family A26 unassigned peptidases (*Agrobacterium tumefaciens*)
A-	A5	family A5 unassigned peptidases (*Picrophilus torridus*)

Figure 1.3. Clan of Aspartic Peptidases from the MEROPS database (This figure was adapted from http://merops.sanger.ac.uk/)

1.2.3 Individual Peptidase

At the lowest level of classification are given the criteria to distinguish between two different peptidases. Two distinct peptidases are expected to meet one of the following criteria (a) they have different specificities (b) they have different sensitivity with an inhibitor (c) they belong to different catalytic types or families or (d) they are encoded by different genes.

The EC system does not explain much about the classification of proteases and it divides all the proteases into 19 subclasses based on the reactions they catalyse. In contrast, MEROPS has provided 140 families of proteases grouped into 30 clans with unique code numbers. Although the EC system does not have a good sytem for the higher levels of classification, it assigns a unique number to each protease and provides a unique recommended name for each protease to reduce ambiguity. Hence the EC and MEROPS together contribute to a sound system of classification of proteases.

1.3.0 Role of the proteases in diseases

Proteases play important physiological roles and their dysfunction can lead to pathological states. Organisms use proteases in almost all metabolic processes. The importance of a few of the well known proteases and their related disorders are mentioned here. In the digestion of food, the gastric juice secreted in the stomach contains Pepsins, a class of aspartic proteases functioning in the pH range of 1.5 to 2. The appearance of pepsin in the oesophagus can lead to proteolytic damage leading to reflux oesophagitis (Roberts *et al*., 2006). In the intestine, further breakdown of food is carried out by serine proteases Trypsin and Chymotrypsin. An inherited autosomal recessive disorder cystic fibrosis is caused by the deficiency in the transport of Trypsin, a serine protease secreted by the pancreas. The inability to inhibit Trypsin by its inhibitor Pancreatic Secretory Trypsin Inhibitor (PSTI) can lead to pancreatitis (Noone *et al,* 2001). Incomplete protein digestion due to deficiency of chymotrypsin is linked with autism in children (Malhotra *et al*., 2002). In the circulatory system, mutations or disorders of Thrombin, a serine protease involved in many events in blood coagulation, can lead to Hyperprothrombinemia, cerebral ischemia and infarction (stroke) (Wolberg *et al*., 2007). Deficiency in Plasmin, a serine protease in the blood which degrades fibrin clots may lead to thrombosis. Neutrophil Elastase is an important serine protease involved in breakdown of bacterial cell walls, if expressed aberrantly causes the lung disease pulmonary emphysema (Takahashi *et al*., 1988). Cathepsins, a group consisting of 11 proteases expressed in all mammalian cells consisting mainly of cysteine proteases are involved in cellular turnover. Cathepsins have been implicated in Cancer, Stroke, Alzheimer's disease, Arthritis, parasite entry and lung diseases (McGrath *et al*., 1999).

Caspases a group of 11 cysteine proteases play essential roles in programmed cell death. Failure of apoptosis is one of the main causes of tumour development while unwanted apoptosis leads to Alzheimer's disease (Yuan *et al*., 2004). Furin, a serine protease belonging to the Subtilisin-like Proprotein Convertase Family is involved in activation of hormones. Furin is also utilized used by pathogens like HIV, Ebola, Dengue and Anthrax viruses for becoming fully functional and gain entry into the host cell (Nakayama *et al*., 1998). The ADAM family of metalloproteases is key regulators of cell surface events and play important functions *in vivo* for example the development of the heart and the brain (Hooper *et al*, 2005).

The projects reported in this thesis are mainly related to serine and cysteine proteases and their inhibitors. Serine proteases and Cysteine proteases have similar catalytic triad residues apart from the nucleophilic residue of cysteine or serine which are histidine and aspartic acid. The mechanisms of catalysis of these two proteases are similar. In the following section, more details about these two proteases are discussed.

1.3.1 Cysteine proteases

Cysteine proteases are found in all the kingdoms of life. The papain-like cysteine proteases form the largest subfamily among cysteine proteases. Papain is the archetype of this family (C1) which belongs to the clan CA Members of this clan has a catalytic triad composed of a histidine, asparagine or aspartic acid and a nucleophilic cysteine. Papain-like cysteine proteases are expressed as preproenzymes and are synthesized at the rough endoplasmic reticulum. Among all the cysteine proteases, the cytoplasmic calpains and the lysosomal cathepsins (Table 1.3 and Figure 1.4) play the most important

physiological roles and are implicated in a variety of diseases. Mammalian lysosomal cysteine proteases play important physiological roles such as in extracellular matrix turnover, antigen presentation, and processing events and that they may represent viable drug targets for major diseases such as osteoporosis, arthritis, immune-related diseases, atherosclerosis, cancer, and parasitic infections(Lecaille *et al*, 2002). Calpains are non-lysosomal cysteine proteases found in almost all mammalian and avian cells. Calpains perform limited proteolysis and are found to activate various kinases and also proteins of the cytoskeleton. Calpains are thought to be involved in muscle degradation, bone resorption, formation of cataracts and implicated in several other pathological processes (Table 1.4). Viral cysteine proteases like 3C proteinase (picornain) being very similar to chymotrypsin-like serine proteases, have been placed in the mixed category of proteases (clan PA) as they have evolved from a common ancestor (Otto and Schirmeister, 1997). Several parasitic protozoan species use cysteine proteases for invasion into host cells, for example Cruzain used by Trypanasoma cruzi to cause Chagas' disease. Papain and Caricain from the papaya plant and Actinidin (from kiwi fruit) are well known plant proteases.

Clan	Family	Example of cysteine proteases
CA	C1 (mammalian)	Cathepsins B, K, L, S, H
CA	C1 (parasite)	Falcipain, Cruzipain, Der p 1
CA	C2 (cytosolic)	m-Calpain, μ-Calpain, Calpain-8
PA(C)	C3 (viral)	Picornain 3C (*e.g.*, Hepatitis A virus, human rhinovirus, foot-and-mouth disease virus)
PA(C)	C30 (viral)	SARS virus 3C-like endopeptidase or severe acute respiratory syndrome coronavirus main protease or SARS-CoV M^{pro} or SARS-CoV 3CL^{pro}
CD	C14 (mammalian)	Caspase 1 (ICE), Caspase 3, Caspase 8
CD	C25 (bacterial)	Gingipain K, Gingipain R
CE	C5 (viral)	Adenovirus
CX	C99	Cancer procoagulant

Table 1.2. Classification of cysteine proteases (Leung-Toung *et al* 2006)

proteases AC number (Swiss-Prot)	signal peptide AA	prodomain length AA	prodomain size Da	prodomain pI	mature domain length AA	mature domain size Da	mature domain pI	prepro-enzyme length AA	prepro-enzyme size Da
				Human Cathepsins					
cathepsin L P07711	17	96	11724	9.0	220	24170	4.7	333	37546
cathepsin V (L2) O60911	17	96	11672	9.5	221	23999	8.6	334	37329
cathepsin S P25774	16	98	11830	9.3	217	23993	7.6	331	37479
cathepsin O P43234	23	84	9894	9.8	214	23460	5.8	321	35957
cathepsin F Q9UBX1	19	251	27846	9.4	214	23593	5.8	484	53365
cathepsin X (Z) Q9UBR2	23	38	4338	9.6	242	27149	5.5	303	33868
cathepsin B P07858	17	62	7245	9.6	260	28663	5.2	339	37807
cathepsin C P53634	24	206	23534	8.7	233	26032	5.6	463	51841
cathepsin K (O2) P43235	15	99	11833	6.3	215	23495	8.9	329	36966
cathepsin H P09668	22	75	9104	9.8	220	24187	5.7	335	37403
cathepsin W P56202	21	106	12060	4.8	249	28024	8.9	376	42099

Table 1.3 . Members of the Lysosomal Cathepsins. (Adapted from Chemical Reviews, 1997, Vol. 97, No. 1)
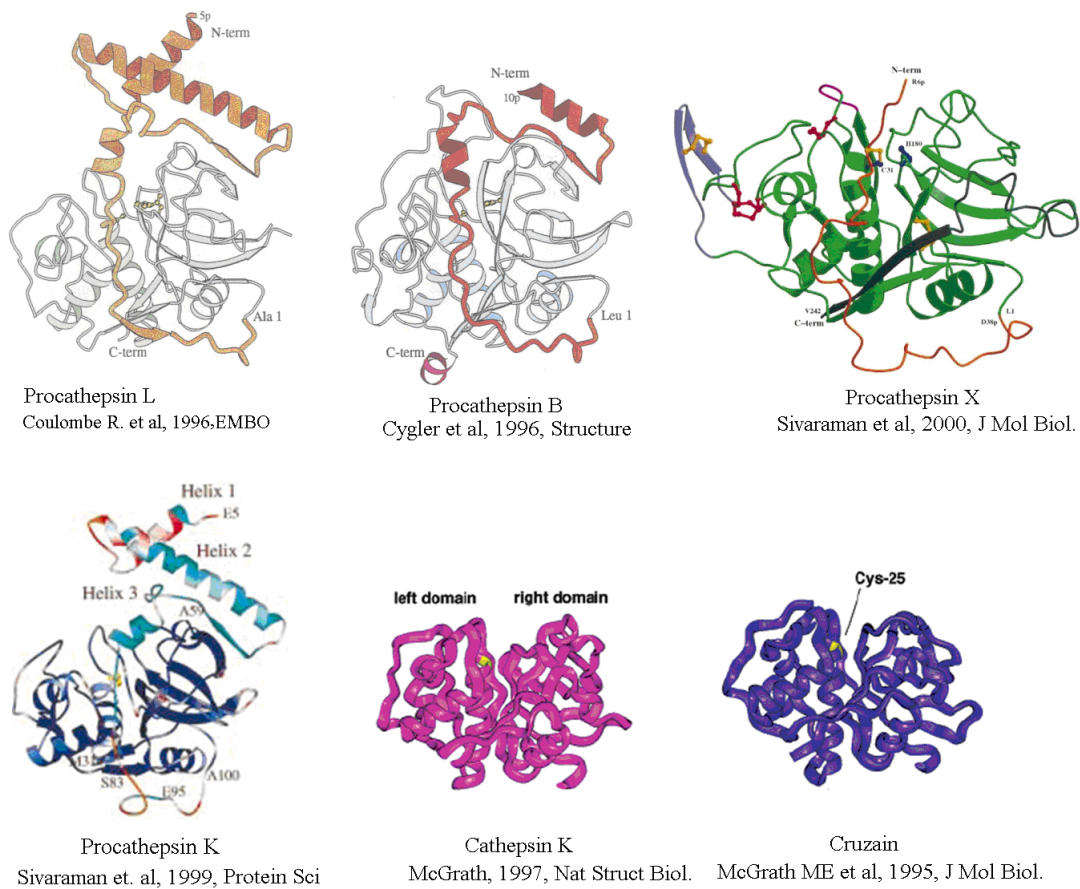
Figure 1.4. Crystal structures of selected proenzyme and mature forms of Cathepsins

disorder	proposed mechanism
stroke	degradation of cytoskeletal proteins leading to neuronal cell death and permanent tissue damage
brain trauma	similar mechanism to stroke
subarachnoid haemorrhage	activation of protein kinase C leading to sustained cerebral vasospasm
Alzheimer's disease	abnormal processing of amyloid precursor protein
spinal cord injury	degradation of myelin proteins
cardiac ischaemia	breakdown of myofibril proteins causing cardiac myocyte shrinkage, cell death, and tissue damage
muscular dystrophy	breakdown of myofibril proteins
cataract	lens protein (crystallins) breakdown leading to precipitation that causes lens opacity
thrombotic platelet aggregation	proteolysis of aggregin, promoting platelet aggregation
restenosis	renarrowing of blood vessels after angioplasty owing to calpain-mediated proliferation and migration of smooth muscle cells
arthritis	breakdown of cartilage and extracellular matrix component proteoglycan

Table 1.4. Calpain in Pathological Processes. (adapted from Chemical Reviews, 1997, Vol. 97, No. 1)

1.3.2 Catalytic mechanism of cysteine proteases

Fig. 1.5 shows the catalytic mechanism for cysteine proteases having the catalytic triad Cys-His-Asp/Asn. The catalytic reaction of Cysteine proteases takes place similar to serine proteases in two phases – (a) acylation and (b) deacylation. In cysteine proteases, the nucleophilic cysteine residue is already ionized prior to substrate binding. Hence cysteine proteases can be regarded as a priori activated enzymes. The acylation phase starts with the deprotonation of the thiol in the enzyme's active site by the adjacent histidine residue. In the next step, deprotonated cysteine's anionic sulfur carries out the nucleophilic attack on the substrate carbonyl carbon forming a tetrahedral transition state. In the transition state, the carbonyl double bond of the substrate is reformed thereby breaking the peptide bond between the carbon and nitrogen. The amino portion of the substrate accepts the hydrogen from the imidazole ring of His, breaking free one part of the peptide bond with an amino terminal. The histidine residue is released to its deprotonated form. A thioester intermediate linking the new carboxy-terminus of the substrate to the cysteine thiol is formed. The -OH portion of the water molecule bonds with the thioester intermediate forming a second tetrahedral transition state while the other hydrogen atom bonds with the imidazole ring of histidine. In the next step, the carbonyl bond of the substrate C-terminal portion is reformed and the bond with the Cys nucleophilic thiol is broken, thereby freeing the second part of the peptide chain. The hydrogen bonded to the His is now regained by the Cys thiol. A hydrogen bond is formed with the Cys and His and thus the active site is regenerated.
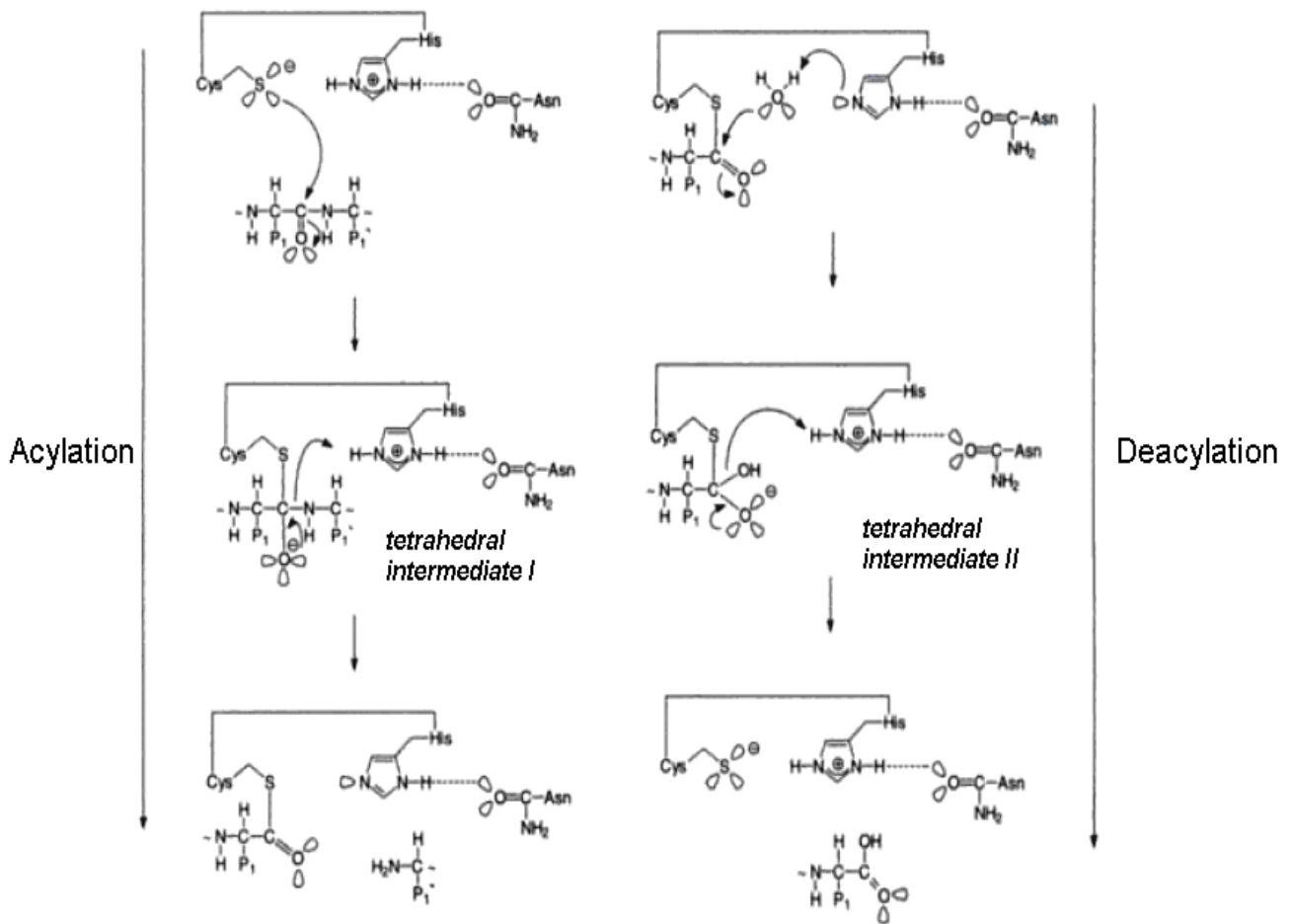
Figure 1.5. The catalytic mechanism of a cysteine protease (Fig adapted from "Enzymes and their inhibition" By H. J. Smith, Claire Simons, CRC Press, 2005 , ISBN 0415334020)

During the reaction, an oxyanion hole is formed by hydrogen bonding by the sidechain amide hydrogen atoms of Gln19 and Cys 25 backbone amide hydrogen. This serves to stabilize the developing negative charge on the carbonyl oxygen atom of the cleaved peptide. The charge relay involved in cysteine proteases and the oxyanion hole Gln19 are shown in Fig 1.6.
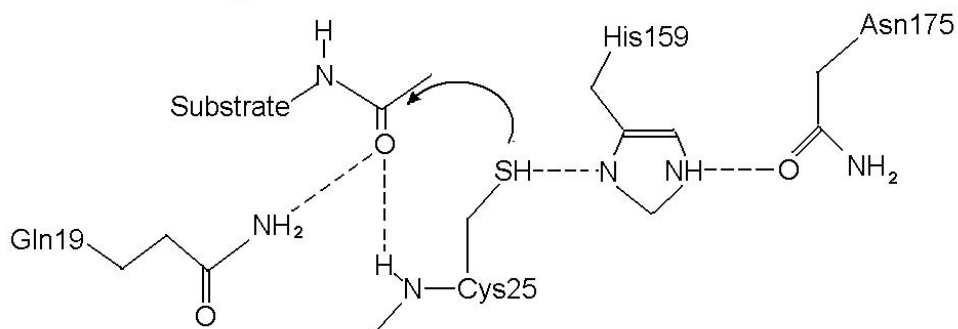
Figure 1.6. The catalytic triad of cysteine protease Papain showing the charge relay involved in making the active site thiol more electronegative.

1.3.3 Inhibitors of Cysteine proteases

The inhibitors of cysteine proteases can be endogenous inhibitors i.e. with a protein structure or small molecule synthetic inhibitors. The endogenous inhibitors inactivate cysteine proteases by competitive, noncovalent, reversible inhibition. Some of these inhibitors also inhibit serine proteases. A number of small molecule synthetic inhibitors of cysteine proteases are being developed as drugs for various pathological conditions. These inhibitors are discussed in the sections 1.3.4 and 1.3.5.

1.3.4 Endogenous inhibitors: Cystatin superfamily

The endogenous mammalian cysteine protease inhibitors are grouped under the cystatin superfamily (Otto *et al*., 1997). These inhibitors are found to be stable at high temperatures (100° C) and extreme pH and highly specific cysteine protease inhibitors. Cystatins have been implicated in pathogenic states related to immune responses, cancer and immune evasion by parasites. There are three families of inhibitor in the cystatin

superfamily based on sequence homology (a) the cystatins (b) the stefins and (c) the kininogens. The members of the Stefin family are typically of 11kDa molecular weight and lack disulphide bridges. Cystatin A and Cystatin B are examples of Stefin family CPIs. Cystatin A is found mainly in the epithelial cells and neutrophilic granulocytes whereas Cystatin B is found in all cells and tissues. The Cystatin family has proteins with typical molecular weight of 12 to 13 kDa. They have two disulphide bridges at the C-terminal.

Examples are Cystatins C, D, S, SN, and SA. Cystatin C is distributed in the extracellular space, cortical neurons, in pancreatic islet cells, in the thyroid glands and in parotid glands. Cystatin S is located in the salivary glands, tear fluid, serum, urine, gall, pancreas and bronchi. The kininongen family consists of two forms – high molecular weight kininogens of 120kDa and low molecular weight kininogens of 50-80 kDa. Kininogens are extracellular proteins. They are also precursors of vasodilatory peptides kallidin and bradykinin.

1.3.5 Synthetic inhibitors of Cysteine proteases

The low molecular weight inhibitors of proteases in general can be classified as active site directed or allosteric effectors. Active site directed inhibitors can be further divided according to the type of interaction into covalent/noncovalent and irreversible/reversible inhibitors (Table 1.5). Efforts are on to develop small molecule inhibitors of cysteine proteases for therapeutic purposes. There are several classes of synthetic inhibitors including peptide aldehydes, methyl ketones and nitriles as reversibly acting inhibitors. Diazomethanes, halomethyl ketones, acyloxymethyl ketones, O-acylhydroxamates, epoxysuccinyl derivatives and vinyl sulfones are irreversible inhibitors.

Inhibitor	Example
I. non-covalent interactions	
I.1. substrate analogs	malonate/succinate dehydrogenase
I.2. tight binding inhibitors	
- transition state analogs	Proscar/5-α-reductase
	phosphonyl peptides/metallo proteases
- multi domain inhibitors	phosphono acetyl-L-aspartate/carbamoyl transferase
II. covalent interactions	
II.1. mechanism based inhibitors	
- transition state analogs	peptidyl aldehydes/cysteine & serine proteases
- enzyme activated inhibitors	haloenollactones/serine proteases
	Sulbactam/ß-lactamases
- dead end inhibitors	peptidyl nitriles/cysteine proteases
- alternate substrate inhibitors	Physostigmine/acetylcholinesterase
II.2. affinity labels	
- chemical reactive affinity labels	chloromethylketones/cysteine & serine proteases
- quiescent affinity labels	acyloxymethylketones/cysteine proteases

Table 1.5. Classification of enzyme inhibitors (adapted from Otto and Schirmeister, 1997)

Aldehyde inhibitors are being developed for high selectivity towards Cathepsins. A drawback of these inhibitors is the high reactivity of the aldehyde moiety. Tripeptidyl leucine analogues inhibiting cathepsin K have been evaluated in rat adjuvant arthritis and PTH induced osteoporosis models (Votta *et al*., 1997). Peptidomimetic aldehyde inhibitors have been designed to be highly potent and selective to Cathepsin L and showed in vivo efficacy in a mice model of bone resorption (Yasuma *et al*., 1998). By incorporating an α-keto moiety in aldehyde inhibitors, a potent Cathepsin S inhibitor was designed which has shown efficacy in a murine model of Sjörgen syndrome (Saegusa *et al*., 2002). Aldehyde analogues of Vitamin B6 have been developed as potent inhibitors of Cathepsin B and Cathepsin K and has shown to be effective in an in vitro

bone resorption assay (Katunuma *et al.*, 1999). Double headed aldehyde inhibitors have been developed against cysteine and metalloproteases as they have been implicated in tumor invasion and metastasis (Yamamoto *et al.*, 2002). These have an aldehyde moiety specific to Cathepsin B and a chelating hydroxamic acid portion for inhibition of matrix metalloproteases (MMP).

Propeptide mimetic inhibitors have been developed by us for Cathepsin L which mimics the mode of autoinhibition exhibited by the propeptide region of the proenzyme of Cathepsin L (Chowdhury *et al.*, 2002). The propeptide regions have been shown to be potent inhibitors of the mature form of the enzyme. By further optimization of the active site binding region of the propeptide (Chowdhury *et al.*, 2008) highly potent non covalent inhibitors have been developed. These inhibitors have a reverse-binding mode i.e. the backbone of the peptide bond is in the reverse direction as compared to a normal substrate which makes it resistant to cleavage by the protease. These inhibitors have been discussed in detail in Chapter II of this thesis.

Acyclic and cyclic ketones were developed in an attempt to reduce the reactivity of aldehyde inhibitors. In this class, Mercaptomethyl ketones have been designed as potent and selective inhibitors of cruzipain, a parasitic cysteine protease (Huang *et al.*, 2001). Conformationally constrained cyclic ketones have been developed as potent and selective inhibitors of Cathepsin K (Marquis *et al.*, 2001). The introduction of an azepanone moiety in the cyclic ketone inhibitor increased the potency and bioavailability and proved efficacious in a nonhuman primate model of osteoporosis (Marquis *et al.*, 2001). This was the first example of viability of Cathepsin K inhibitors as alternative treatment of osteoporosis (Stroup *et al.*, 2001).

Nonpeptidyl derivatives of nitriles having pyrrolidine or azetidine rings have been found to be potent inhibitors of cysteine proteases. Nitrile inhibitors with 1-cyanopyrrolidinyl ring were shown to be potent inhibitors of the degradation of denaturated collagen by cathepsin K and of bone resorption in an in vitro model (Falgueyret). Dipeptidyl Nitriles have been designed as potent and selective reversible inhibitors of Cathepsin B (Greenspan *et al.*, 2001).

Epoxysuccinyl analogues are synthetic analogues of the natural product (2S, 3S)-trans-epoxysuccinyl-L-leucyl-agmatine (E-64), which is a very powerful irreversible inhibitor of cysteine proteases. Additionally, the propeptide regions of the preproenzymes are potent inhibitors of cysteine proteases. To mimic the anti-substrate binding mode of the propeptide of Cathepsin B, an inhibitor was synthesized that contains the dipeptide moiety Leu-Pro-OH for the P1'-P2' substrate positions and the tripeptide moiety Leu-Gly-Gly-OMe for the P2-P4 positions in anti-substrate orientation. This inhibitor was found to be the most potent among its class (Schaschke *et al.*, 1998). A similar inhibitor designed for Cathepsin L was effective in reducing tumor-induced hypercalcemia in mice (Katunuma *et al.*, 1999).

Peptidyl vinyl sulfones are potent irreversible inhibitors of cathepsins. They have been shown to be effective in mice arthritis models by significantly reducing inflammation as well as bone and cartilage erosion. They have high potency against various parasitic cysteine proteases which makes them ideal drug candidates. A vinyl sulfone inhibitor was demonstrated to cure T. cruzi infections in a mouse model (Engel *et al.*, 1999). The same inhibitor was found to be orally effective against malaria in mouse

model. Preparations are on for the future human clinical trial for Chagas' disease for this inhibitor.

Peptidyl diazomethylketones were designed based on the antibiotic azaserine which inhibits cellular growth by alkylation of a thiol group on the amidotransferase involved in purine synthesis (Buchanan, 1973). Diazomethylketones were known to be intermediates in the synthesis of chloromethylketones and also that they did not inhibit serine proteases. Thus diazomethylketones could be developed as potent, irreversible and selective inhibitors of cysteine proteases (Shaw, 1984). The first such inhibitor was Z-Phe-CHN2, an inhibitor of papain (Leary *et al*, 1977). In chapter III of this thesis, the crystal structure of Cathepsin-L with a dipeptidyl diazomethylketone inhibitor Z-Phe-Tyr (t-Bu)-CHN2 is discussed.

Peptidyl α-keto-β-aldehydes or glyoxal inhibitors are potent inhibitors chymotrypsin, cathepsin B, cathepsin L, cathepsin S, and the proteasome (Lowther *et al*, 2002). These inhibitors can be synthesized by the oxidative cleavage of the diazogroup of peptidyl diazomethylketones. α-keto-β-aldehydes have two highly electrophilic carbonyl atoms as compared to chloromethylketones which have one electrophilic carbonyl and one alkylating group. They are inhibitors of both serine and cysteine proteases. However these inhibitors are 10 fold more potent towards cysteine proteases than peptide aldehydes (Walker *et al*, 2000). The crystal structure of Cathepsin-L with a dipeptidyl glyoxal inhibitor Z-Phe-Tyr (t-Bu)-COCHO is discussed in chapter III of this thesis.

β-lactams are well-known compounds as the antibiotic penicillin belongs to this group. As the drug resistance to antibiotics increased, several β-lactam analogues were synthesized. Recently reversible β-lactam inhibitors for cysteine proteases have been

developed. Single ring as well as a bicyclic ring β-lactam moieties have been evaluated as inhibitors of cysteine proteases. β-lactam analogue inhibitors for Cathepsins K, L, B and S have been developed(Brömme *et al*., 2002).

Other classes of inhibitors include Diacyl bis hydrazides which span both the S and S' subsites of Cathepsin K and Diamino Pyrrolidinones which are derived by incorporating a five membered rings in the Diacyl bis hydrazides at the P2 an P2' positions. Diacyl bis hydrazides have shown efficacy in human osteoclast resorption assays (Thompson *et al*., 1997).

1.4 Serine proteases

The serine proteases are named for the nucleophilic Ser residue in their active sites. The serine proteases have a catalytic triad or charge relay system comprising "Asp-His-Ser" in their active sites. Almost one third of all known proteases can be classified as serine proteases. The Asp-His-Ser triad can be found in at least four different clans (group of structurally homolgous families) of serine proteases - chymotrypsin, subtilisin, carboxypeptidase Y, and Clp protease which indicate that this catalytic machinery has evolved on at least four separate occasions (Dodson *et al*., 1998). The details of these clans are shown in Table 1.6. This is an example of convergent evolution. Recently, serine proteases with novel Ser-His-Glu, Ser-Lys/His, His-Ser-His, and N-terminal Ser catalytic triads and dyads have been discovered (Rawlings *et al*., 2000). The Chymotrypsin-like proteases are the most abundant proteases in nature. The Chymotrypsin family S1 includes intestinal digestive proteins trypsin, chymotrypsin and also chymase and tryptase found in mast cells, granzymes found in cytotoxic cells and

thrombin to name a few. Until the sequence and structure of subtilisin was discovered, Chymotrypsin-like proteases were thought to be the only type of serine proteases. The subtilisin family S8 is the second largest family of proteases. Most members of this family are found in bacteria, archaea and fungi and involved in nutrition and some are pathogenic. Furin, a protease is used by several viruses such as HIV, Anthrax and Ebola viruses for entry into the host. Subtilisin and Proteinase K are used in the pharmaceutical industry.

clan	members	example	catalytic residues	distribution
		"Classic" Catalytic Triad Serine Proteases		
PA(S)	301	chymotrypsin	His−Asp−Ser	B, Ar, F, Pl, An, V
SB	91	subtilisin	Asp−His−Ser	B, Ar, Pr, F,Pl,An,V
SC	64	carboxypeptidase Y	Ser−Asp−His	B, Ar, Pr, F, Pl, An,
SK	14	Clp protease	Ser−His−Asp	B, Ar, Pr, F, Pl,An,V
		"Novel" Serine Proteases		
SE	16	D-Ala-D-Ala carboxypeptidase A	Ser−Lys	B, Ar, Pl, An
SF	24	signal peptidase I	Ser−Lys/His	B, Ar, F, Pl, An, V
SH	7	cytomegalovirus assemblin	His−Ser−His	V
SM	7	C-terminal processing protease-1	Ser−Lys	B, Ar, Pl
SN	4	dipeptidase E	Ser−His−Glu	B, An
PB(S)	4	penicillin amidohydrolase precursor	N-terminal Ser	B, Ar, Pr

Table 1.6. Clans of Serine proteases classified based on structural similarity in the MEROPS database. (Table adapted from Chemical Reviews, 2002, Vol. 102, No. 12)

1.4.1 Catalytic mechanism of Serine proteases

The catalytic triad or charge relay system of most serine proteases consists of Asp, His and Ser. Examples are: Asp 102, His 57, Ser 195 in Chymotrypsin , Trypsin and Asp 32, His 64, Ser 221 in Subtilisin. The Ser side chain –OH group acts as a nucleophile which attacks the peptide bond. The His imidazole ring, having a pair of electrons, has the ability to accept hydrogen from Ser. The Aspartic acid carboxyl group forms a hydrogen bond with His, making the pair of electrons of the His Nitrogen more

electronegative and in turn making the nucleophilic Ser Oxygen more electronegative. The catalytic reaction of serine proteases takes place in two phases – (a) acylation and (b) deacylation (Carter, 1988; Voet, 1990). The acylation step starts with the Ser side chain nucleophilic Oxygen attacking the carbonyl group of the scissile peptide bond forming a tetrahedral transition state. The imidazole ring of His accepts the hydrogen from the Ser OH and thus acts as a proton acceptor. In the transition state, the carbonyl double bond of the substrate is reformed thereby breaking the peptide bond between the carbon and nitrogen. The amino portion of the substrate accepts the hydrogen from the imidazole ring of His, breaking free one part of the peptide bond. The acylation phase is shown in Figure 1.7.
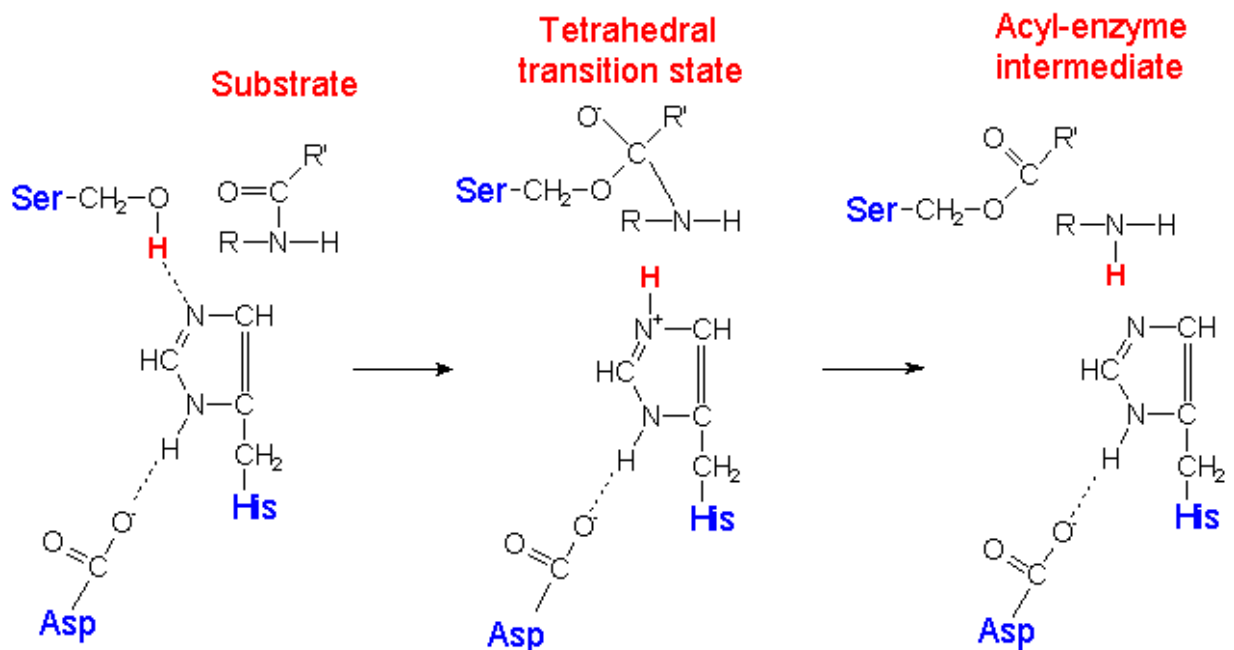
Figure 1.7. Acylation reaction in the catalytic mechanism of a serine protease
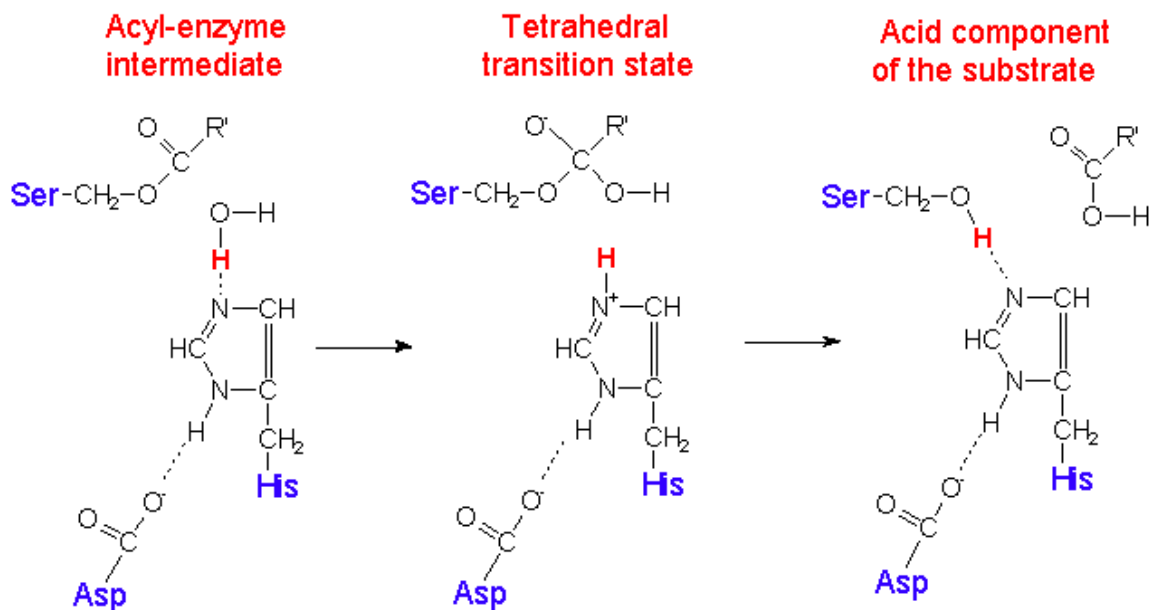
Figure 1.8. The deacylation reaction in the catalytic mechanism of a Serine protease (adaptedfrom http://www.cryst.bbk.ac.uk/pps97)

In the deacylation phase, hydrolysis of the acylated serine residue takes place by the reaction with a water molecule. The -OH portion of the water molecule bonds with the acylated serine residue forming a second tetrahedral transition state while the hydrogen atom bonds with the imidazole ring of histidine. Next, the carbonyl bond of the substrate C-terminal portion is reformed and the bond with the Ser nucleophilic oxygen is broken, thereby freeing the second part of the peptide chain. The hydrogen bonded to the His is now regained by the Ser oxygen and a hydrogen bond is formed with the Ser and His and thus the active site is regenerated. The deacylation phase is shown in Figure 1.8. In addition, during the reaction, an "*oxyanion hole*" is formed by hydrogen bonding by the backbone amide hydrogen atoms of Gly-193 and Ser-195(in case of Trypsin and

Chymotrypsin) (Asn 155 and Ser 221 in case of Subtilisin). This serves to stabilize the developing negative charge on the carbonyl oxygen atom of the cleaved peptide. During the formation of the tetrahedral transition state, the negative oxygen ion after accepting the electrons from the carbonyl double bond, perfectly fits into the oxyanion hole. This makes the tetrahedral transition state more favourable for serine proteases, and increases the efficiency of the catalytic reaction. The oxyanion hole of serine proteases is shown in Figure 1.9.
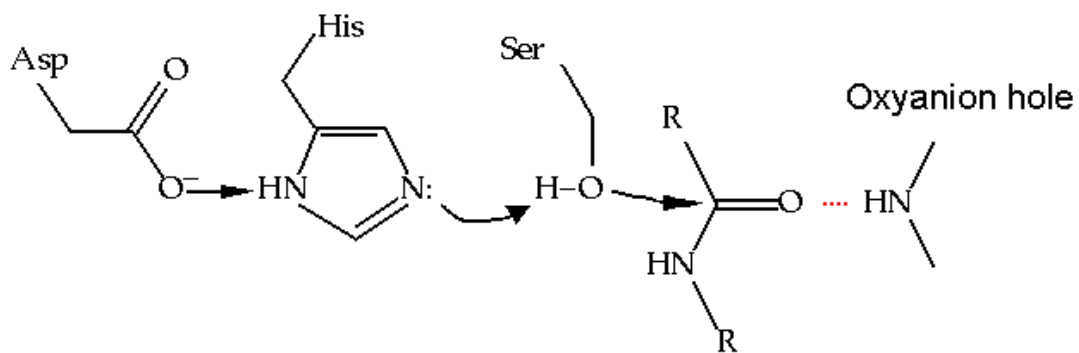
Figure 1.9 The catalytic triad of serine proteases showing the charge relay involved in making the Ser Oxygen more electronegative.

1.4.2 Serine protease inhibitors

The natural serine protease inhibitors can be classified into three groups – serpins, the canonical and non-canonical. The serine protease inhibitors have been grouped into atleast 18 different families (Wlodzimierz *et al*., 1998). The representative three dimensional structures of 13 of these families are known. Most of these serine protease

inhibitors inhibit their cognate enzymes according to a common substrate-like "standard" mechanism. This group of inhibitors is termed as the canonical inhibitors. They usually have low molecular weight with around 29 to 190 amino acids (Laskowski *et al*., 1980). They all possess an exposed binding loop of a characteristic conformation but are otherwise not structurally related. The serpins form a family of homologous proteins with larger molecular weight proteins comprising around 400 amino acid residues (Travis *et al*., 1983). They usually have a sugar molecule attached. Serpins also have an exposed binding loop which interacts with their cognate proteases.

1.4.3 Serpins

Members of the serpin superfamily consist of a core domain of three β-sheets and 8-9 α helices. Although this common fold is found in all serpins, the sequence identity can be as low as 25%. 16 clades designated A to P have been formed in the superfamily so far. Serpins are widely distributed among eukaryotes and viruses that infect them. They have not been found in fungi but a bacterial serpin was recently discovered (Irving *et al*, 2002). Serpins are abundantly found in the human plasma and most of them are inhibitors of thrombin. Others are inhibitors of coagulation proteins Factor Xa, Factor XIa or of fibrinolytic proteinases like plasmin. The gene knock-out of the serpin antithrombin has proven to be lethal in mice. Serpins are found in many highly regulated physiological processes, such as blood coagulation, fibrinolysis, and inflammation (Stein *et al.*, 1995). Some important serpins are α1-PI (Schulze *et al.*, 1994), α2-antiplasmin(Wiman *et al.*, 1978), C1 inhibitor(Bock *et al.*, 1986), and neuroserpin(Davis *et al*., 1999) . Serpins are found in four different conformations - metastable state, latent

state, cleaved state and the δ state. From the latent state, the serpins fold into a metastable

state, and subsequently undergo insertion of the
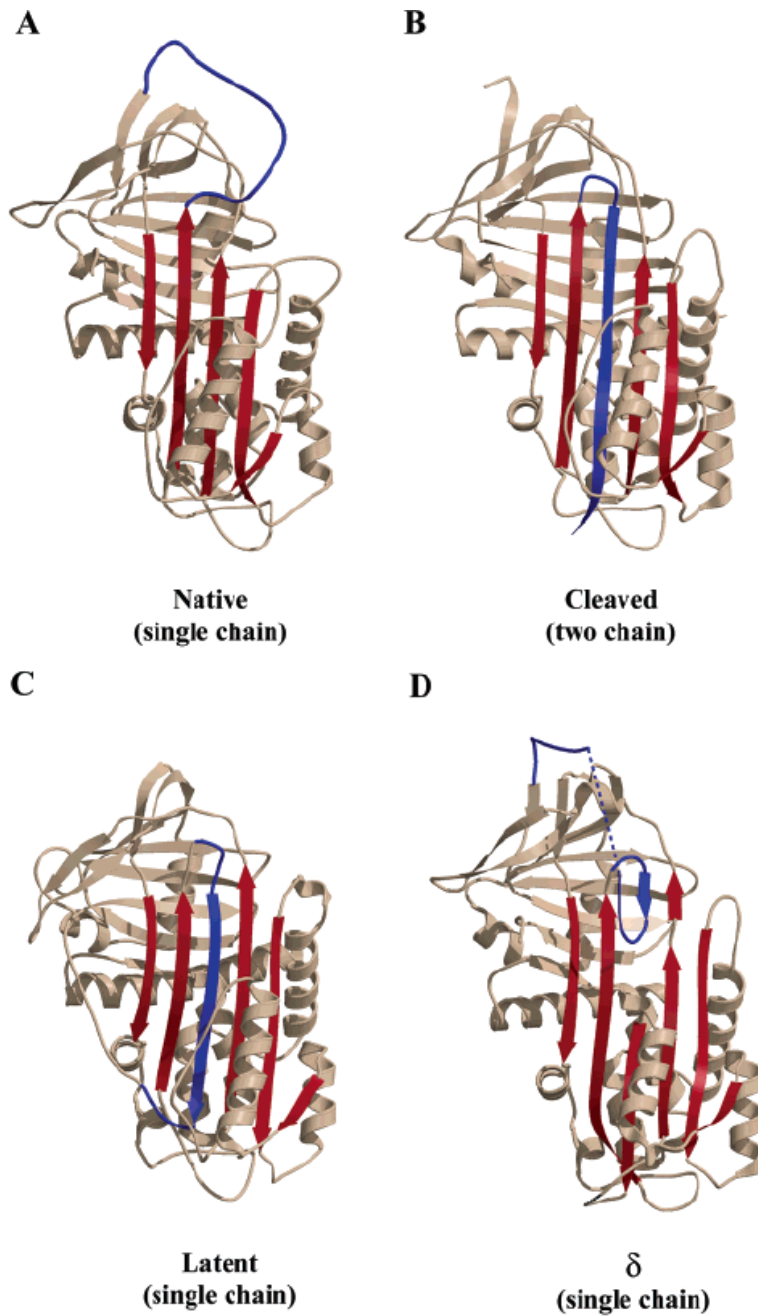
Figure 1.10. Comparison of the four different conformations of the serpin α1-proteinase

inhibitor (adapted from Gettins, 2002)

reactive loop into β-sheet A, and form a covalent serpin-proteinase complex. Thus serpins undergo dramatic change in their conformations in the reactive site loop (Figure 1.10) unlike the rigid mode of inhibition in canonical binding inhibitors discussed in the following section.

1.4.4 Canonical serine protease inhibitors

The binding loop or reactive site loop residues P3-P3' serve as a recognition motif which, depending on the sequence inhibits a particular protease (Lee *et al*, 2007). The P1 residue located in the center of the loop is responsible for most of the interactions with the protease and binds to the S1 specificity pocket of the protease. The binding loop mode of interaction with the protease is similar to a substrate. Usually a short antiparallel β-sheet is formed between the P3-P1 residues and the 214–216 (chymotrypsinogen numbering) segment of the enzyme. An important feature is a short contact (about 2.7Å) between the P1 carbonyl carbon and the catalytic serine residue and two hydrogen bonds between carbonyl oxygen of P1 and Gly193/Ser195 amides. The conformation of the binding loops of different families is similar, which is an example of convergent evolution. The canonical inhibitor families include BPTI (Kunitz) family, Kazal family, STI family(Soybean Trypsin Inhibitor or STI-Kunitz family), SSI family(Streptomyces Subtilisin Inhibitor), Potato inhibitor 1 family, Potato inhibitor 2 family, Chelonianin family, Bowman-Birk family and Squash seed inhibitors family. Selected members of the families are shown in Figure 1.11.

The crystal structure and the canonical binding mode of a Kazal inhibitor, CrSPI-1 with its cognate protease Subtilisin has been discussed in detail in Chapter III of this thesis.

Labels in figure: 15 ARG, 16 ALA, LET 45, ASP 46, ARG 5, ILE 6, Basic Pancreatic Trypsin Inhibitor(BPTI), Eglin C, Squash seed inhibitor-I, ARG 47, SER 48, LEU 18, GLU 19, LEU 72, MET 73, Mung bean trypsin inhibitor, Turkey ovomucoid third domain, Mucous proteinase inhibitor
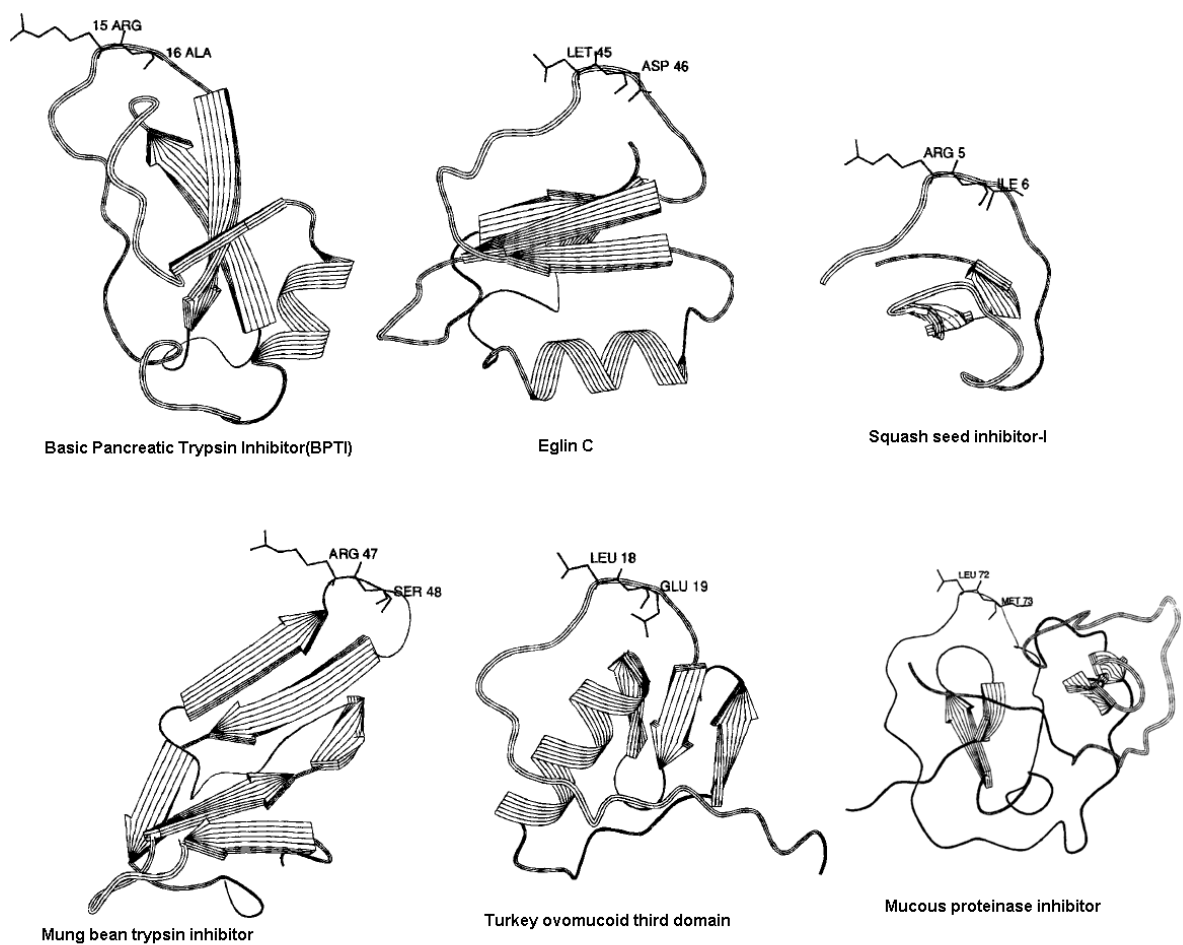
Figure 1.11. Representative structures of selected canonical serine protease inhibitor families. BPTI (BPTI family), Eglin C (Potato Inhibitor 1 family), Squash seed inhibitor 1(Squash seed inhibitor family), Mung bean trypsin inhibitor(Bowman-Birk family), Turkey ovomucoid third domain(Kazal family), Mucous proteinase inhibitor(Chelonianin family) Fig adapted from Bode and Huber, Eur J Biochem, 1991

1.4.5 Non canonical serine protease inhibitors

The Hirudin family of serine protease inhibitors bind in the noncanonical mode to the cognate protease (Folkers et al, 1989). Hirudin is an inhibitor of thrombin isolated from the leech and binds according to a very different mechanism. Hirudin's globular domain contacts surface patches adjacent to the thrombin active site. The extended C-terminal tail of hirudin runs in a long groove extending from the active sit cleft of thrombin to the positively charged putative fibrinogen binding exosite, thereby making several hydrophobic contacts and surface salt bridges. This non-canonical interaction may be used more frequently in serine proteinase inhibition, but has so far been defined only for hirudin.

1.4.6 Synthetic inhibitors of serine proteases

The irreversible inhibitors of serine proteases can be grouped as alkylating agents, acylating agents, phosphonylating agents, and sulfonylating agents based on the mechanism of inhibition (Powers *et al*., 2002) . Alkylating agents include fluoromethyl ketones, chloromethyl ketones, acyloxymethyl ketones, epoxides, aziridines, vinyl sulfones, and other Michael acceptors. Acylating agents include β-lactams, lactones, aza-peptides, and heterocyclic derivatives. Phosphonylating agents include peptide phosphonates and phosphonyl fluorides. Several fluorescent derivatives of phosphonate esters have been used for cellular localization of serine proteases. Phosphonyl fluorides are very potent inhibitors of serine proteases and esterases. Phosphonyl fluoride inhibitors such as diisopropyl phosphonofluoridate (DFP), isopropylmethyl phosphonofluoridate (Sarin), and 1,2,2-trimethylpropylmethyl phosphonofluoridate (Soman) are nonselective toward a particular serine protease. Sulfonylating agents include Sulfonyl fluorides.

Sulfonyl fluorides inhibit most serine proteases such as chymotrypsin, trypsin, elastase, complement, coagulation, and fibrinolytic serine proteases. Examples are PMSF and 4-(2-aminoethyl) benzenesulfonyl fluoride. Several irreversible inhibitors have been tested for the possible therapeutic application. For instance inhibitors such as β-lactams and saccharins are the inhibitors of neutrophil elastase for the treatment of emphysema. A number of other bioavailable or orally active inhibitors are now available including benzoxazinones and saccharins as inhibitors for elastase and viral proteases and vinyl sulfones. Several inhibitors are undergoing clinical trials, including a γ-lactam inhibitor that is being tested in trials for inflammatory disorders. A thrombin inhibitor is being developed for use as a surgical glue by an Atlanta biotechnology company. Thrombin can be acylated by trans-cinnamoyl active esters to give stable acyl enzyme derivatives.If the cinnamoyl group has an o-hydroxy group, then it can easily be deacylated photochemically regenerating active thrombin. This led to the concept of "surgical glue", which is being developed by Porter and co-workers. The acyl thrombin derivative can be applied to any desired surgical site and then activated with a laser. The tissue gets glued together when "fibrin" is generated by Thrombin.

The serine and cysteine proteases have similar catalytic mechanisms involving a charge with respect to the active site nucleophile, the histidine and the oxyanion hole. The following chapters of this thesis report the co crystal structures and mechanism of inhibition of Cathepsin-L and Subtilisin with their inhibitors. The major objective of this thesis is to elucidate the structure based mechanism of inhibition of a cysteine protease i.e. human Cathepsin-L and a serine protease Subtilisin using potent reversible and

irreversible inhibitors. X-ray crystallographic approach is used as a primary method throughout. Structure based drug design is the most powerful method for design of inhibitors for various targets involved in disease states. The X-ray crystallographic data shows at an atomic level how a drug or an inhibitor interacts with the target protein which will contribute immensely to the development of better inhibitors. Chapter II describes the inhibition of Cathepsin-L by a series of propeptide mimetic inhibitors. The aim of this work is to enhance the potency of retrobinding inhibitors by increasing the interactions in the S' subsites of Cathepsin-L. In this chapter, interactions made by four inhibitors from this series are described in detail. To supplement the crystal structure data, the dynamic behavior of these inhibitors and electrostatic preferences are also studied by using Molecular Dynamics simulations in collaborations with Prof Enrico, McGill University Canada. In chapter III the co-crystal structures of Cathepsin-L with two covalent inhibitors- a peptidyl diazomethyl ketone and a peptidyl glyoxal inhibitor are described. These inhibitors have identical peptide sequences and differ only in the reactive group which modifies the active site thiol. This study helps to explain the structural basis for the difference in potency of these two inhibitors. Finally, chapter IV outlines the crystal structure of the serine protease Subtilisin in complex with a Kazal-type inhibitor from the horseshoe crab, *Carcinoscorpius rotundicauda*. This double headed kazal inhibitor inhibits two molecules of subtilisin i.e. captures one molecule of subtilisin with each of its domains. Isothermal Titration Calorimetry has been used to supplement the dynamics of binding of this inhibitor to subtilisin. Synthetic peptides derived from the reactive site loop residues of the inhibitors have been tested for their inhibition. These studies will

further advance our knowledge on the innate immune defense mechanism of arthropods and also provide a basis for design of antibacterials and endotoxin detection.

Chapter II

Propeptide Mimetic Inhibitor Complexes of Human Cathepsin L

2.1 Introduction

The papain superfamily of cysteine proteases consists of numerous members among which Cathepsins form the largest group. In humans, 11 Cathepsins are known to-date (McGrath, 1999). The main function of these cathepsins is the terminal protein degradation in the lysosome. Although mainly being scavengers in the lysosome, they play vital physiological roles like in bone resorption and remodelling, T-cell maturation, prohormone processing and apoptosis (Li *et al*., 2009). Recent studies have shown the involvement of Cathepsins in several pathological conditions like tumour invasion and metastasis, osteoporosis, arthritis, atherosclerosis, emphysema, muscular dystrophy, parasitic infections and cancer (Sheahan *et al.*, 1989; Esser *et al*., 1994; Turk *et al.*, 2001; Joyce *et al.*, 2004). Cathepsin L is an important member of this group which has been implicated in tumour growth and invasion. Therefore Cathepsin L has been recognized as viable targets for therapeutics and cysteine proteases in general are attractive targets for drug design.
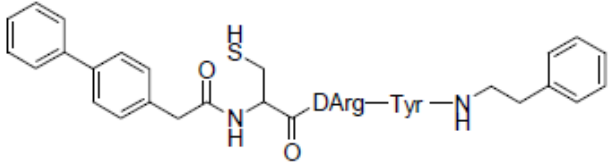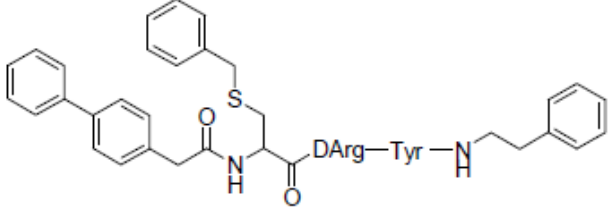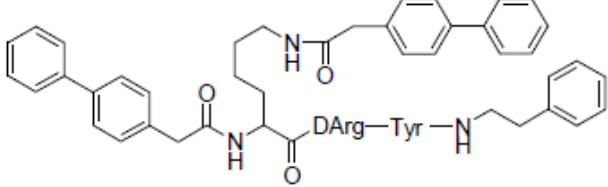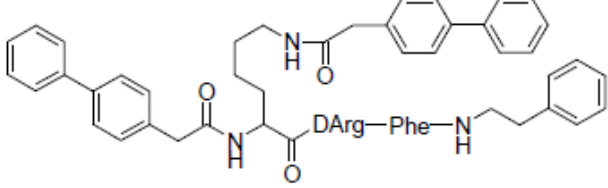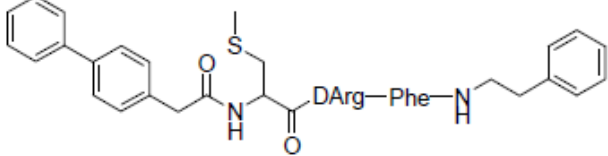
Cathepsin L is biosynthesized as an autoinhibited zymogen with a 96-residue propeptide segment that spans across the entire active site cleft completely blocking access to it. The direction of the propeptide backbone is reversed as compared to the substrate binding mode, conferring resistance to hydrolysis. We have reported a series of non-covalent retro-binding inhibitors designed to span the S–S' regions of the active site of Cathepsin L that mimic the propeptide binding mode (Chowdhury *et al.*, 2008; Chowdhury *et al.*, 2002). These inhibitors showed nanomolar potency and high selectivity for Cathepsin-L. The first crystal structure of such an inhibitor (pdb code 1mhw) confirmed the reverse-binding noncovalent nature of inhibition. However, this

inhibitor contained a Cys amino acid that, during co-crystallization, caused the inhibitor to dimerize through disulfide bond formation (Chowdhury *et al.*, 2002). The resulting bound conformation placed the biphenyl rings from each monomer in the S' subsites of the protein. This raised some uncertainty as to the nature of the protein-ligand interactions in the S' subsites for the monomeric inhibitor.

In this chapter we report the second crystal structure of a member of this class of inhibitors containing an S-benzyl Cys derivative that precluded dimer formation (pdb code 3bc3). The binding mode of the inhibitor at the S1 to S3 subsites was identical to that previously reported for the dimeric form. However, the interactions of the biphenyl groups in the S' subsite were significantly different between the two crystal structures. In addition this chapter reports the crystal structures of three more inhibitors that have been selected to help clarify and elucidate the binding mode of this class of inhibitors. Further, in collaboration with Prof Enrico, McGill University Canada, we have carried out molecular dynamics simulations to supplement the information provided by the crystal structure studies.

2.2 Experimental

Inhibitor synthesis and enzyme assay was performed in the collaborator's lab. The chemical diagram of the inhibitors and their binding affinities are shown in Table 2.1a.

Inhibitor ID	Structure	K_i (µM)
1		0.045
2		0.155
4		0.023
9		0.511
14		0.019

All K_i values were taken from Chowdhury et al.[7,8]

Table 2.1a. Inhibitor Structures co-crystallized with Cathepsin-L and their activities.

2.2.1 Co-crystallization and Data collection

Inhibitor 2: The mature cathepsin L was inhibited with inhibitor **2** by incubating the protein and inhibitor in the presence of 2 mM DTT at room temperature for 3 h. Protein was stored in a buffer of 20 mM sodium acetate at pH 5.7, 100 mM NaCl, and 1 mM EDTA. Because of the limited solubility of the inhibitor, approximately 0.01 mM concentration of protein and inhibitor were initially mixed to obtain a final ratio of 1:3 M

(protein: inhibitor). After incubation, the complex was gradually concentrated up to 14 mg/mL. Crystals of the complex were grown by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 17% (w/v) polyethylene glycol 8000 (PEG 8K) and 0.2 M ammonium sulfate.

Inhibitor 4: The complex of Cathepsin L with inhibitor **4** was prepared by incubating the protein with inhibitor in the presence of 2 mM DTT at room temperature for 3 hours. Protein was kept in 20 mM Sodium acetate pH 5.7, 100 mM NaCl, and 1mM EDTA. Due to the limited solubility of inhibitor 0.01M concentration of protein and inhibitor, that is dissolved in 25% DMSO, were used to prepare the initial mixture with a final ratio of 1:4 M (protein : inhibitor). After incubation the complex was subsequently concentrated up to 9 mg/ml. Crystals of the complex were grown after 10 days by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 19%(w/v) polyethylene glycol 8000 and 200mM Ammonium sulfate.

Inhibitor 9: The complex of Cathepsin L with inhibitor **9** was prepared by following the above mentioned procedure to a ratio of 1:5 M (protein: inhibitor) and concentrated up to 10 mg/ml. Crystals of the complex were grown after 3 weeks by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 25%(w/v) polyethylene glycol 8000 and 200mM Ammonium sulfate.

Inhibitor 14: The complex of Cathepsin L with inhibitor **14** was prepared by following the above mentioned procedure to a ratio of 1:5 M (protein: inhibitor) and concentrated up to 15 mg/ml. Crystals appeared after 2-3 weeks by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 30%(w/v) polyethylene glycol 8000 and 200mM Ammonium sulfate.

All the co-crystallization drops were composed of 1 μl of reservoir solution and 1μl of the complex. We have also soaked all the crystals to ensure the inhibition of Cathepsin L in the crystals by the addition of 1.5μM respective inhibitors solution into the drop. Diffraction data sets were collected on an R-axis IV + + area detector mounted on RU300 rotating anode detector and / or on a synchrotron beam line, BNL, with 25% glycerol as cryo-protectant.

2.2.2 Structure Solution and Refinement

Initial phases for Cathepsin L inhibitor complexes were obtained by molecular replacement method with Molrep and using the mature cathepsin L coordinated taken from Procathespin L (PDB code 1CS8) (Vagin and Teplyakov, 1997). The initial R-factor was 42% for all the complexes and subsequent refinement was carried out with CNS (Brünger *et al.*, 1998). Resulting model with the electron density map was examined and the model was fitted with the O program (Jones *et al.*, 1991). Omit maps were calculated for positioning the inhibitor molecules. All four model buildings and refinements were carried out using O and CNS programs with appropriate entries in their respective dictionaries. Overall geometry of final models was analyzed by PROCHECK (Laskowski *et al.*, 1993). The data collection and refinement statistics are provided in the Table 2.1b and Table 2.2.

2.3 Results and Discussion

2.3.1 Structure of inhibitor 2 and Cathepsin L complex

The crystal structure of mature human cathepsin L complexed with inhibitor **2** was determined at 2.2 Å resolution. Compound **2** inhibits mature cathepsin L with a K_i of 0.155 μM. The asymmetric unit consists of two inhibitor complex molecules. The structure was solved by the molecular replacement method using mature cathepsin L coordinates derived from procathepsin L (PDB code 1CS8). The model has been refined with good stereochemical parameters. Statistics for the Ramachandran plot from an analysis using PROCHECK[19] for the inhibitor complex model gave 84.5% of nonglycine residues in the most favored region (Table 2.1b). Electron density of the mature cathepsin L residues from Thr175 to Gly179 is not observed and is presumably disordered.

Figure 2.1 shows the simulated annealing Fo - Fc omit map for the substrate-binding region of the cathepsin L inhibitor **2** complex. The inhibitor molecule is well-defined in the electron density map. As predicted, similar to the proregion, the crystal structure of this inhibitor complex adopts noncovalent and reverse substrate-binding mode of inhibition. The electron density around the active-site Cys25 side chain indicates the possibility of oxidized sulfur. There are 10 direct hydrogenbonding contacts (<3.2 Å) between the bound inhibitor and the mature cathepsin L of the complex.

Table 2.1b. Crystallographic data and refinement statistics for inhibitor **2**.

Data collection	
Resolution range (Å)	50-2.2
Wavelength (Å)	1.5418
Observed *hkl*	86866
Unique *hkl*	21015
Completeness (%)	94.5
Overall $I/\sigma I$	21.1
[a]R_{sym}	0.048
Refinement and quality of the model	
*Resolution range	30.0-2.2
[b]Rwork (%) no. reflections	18.34 (18333)
[c]Rfree (%) no. reflections	23.41 (1558)
Root mean square deviation	
Bond length (Å)	0.006
Bond angle (º)	1.044
Ramachandran plot (%)	
Favored region	84.5
Allowed regions	15.5
Generously allowed region	0.0
Disallowed regions	0.0
[d]Average B-factors (Å^2)	
Main chain atoms	28.41
Side chain atoms	30.50
Overall protein atoms (no. atoms)	29.42 (3322)
Waters (no. atoms)	40.3 (395)
Ligand (no. atoms)	27.90 (12)

[a] Rsym = |Ii -<I>|/ |Ii| where Ii is the intensity of the ith measurement, and <I> is the mean intensity for that reflection.

[b] Rwork = | Fobs - Fcalc|/ |Fobs| where Fcalc and Fobs are the calculated and observed structure factor amplitudes, respectively.

[c] Rfree = as for Rwork, but for 8.5% of the total reflections chosen at random and omitted from refinement.

[d] Individual B-factor refinement were carried out.

* Reflections greater than I>cI where used in the refinement
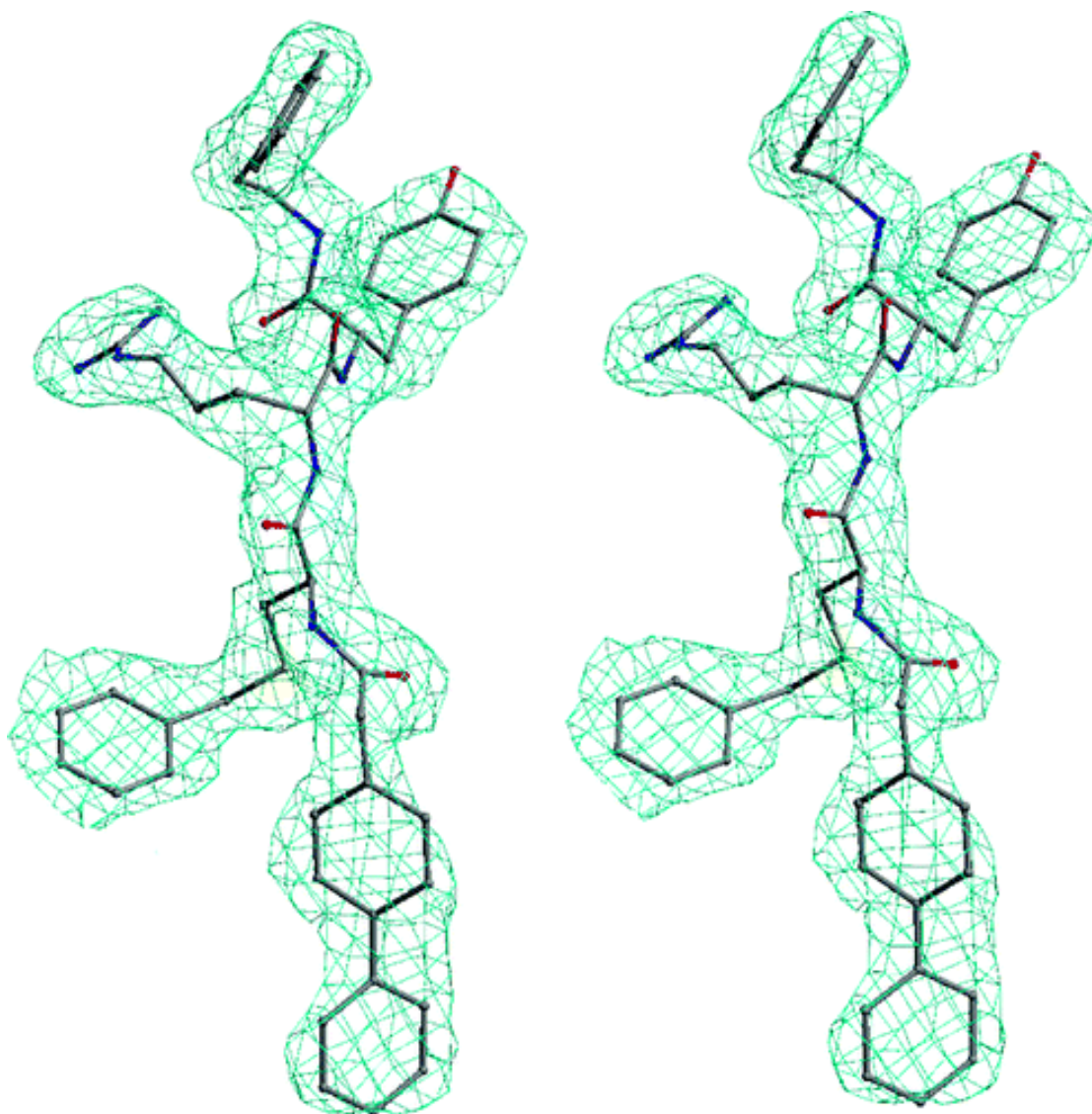
Figure 2.1. Simulated-annealing Fo - Fc omit map in the active-site region of the cathepsin L inhibitor complex. Inhibitor molecule and all atoms within 3 Å from the inhibitor molecule were omitted prior to refinement. The map is contoured at a level of 2σ. This figure was prepared using Bobscript (Esnouf, 1999).

The overall binding mode of inhibitor **2** is comparable to that of the dimeric form of inhibitor **1** with D-Arg anchored at the S1 subsite and the Tyr and phenylethyl groups at the S2 and S3 subsites, respectively, in essentially identical orientations.18 The major differences lie in the orientation of the biphenyl rings (Figure 2.2).
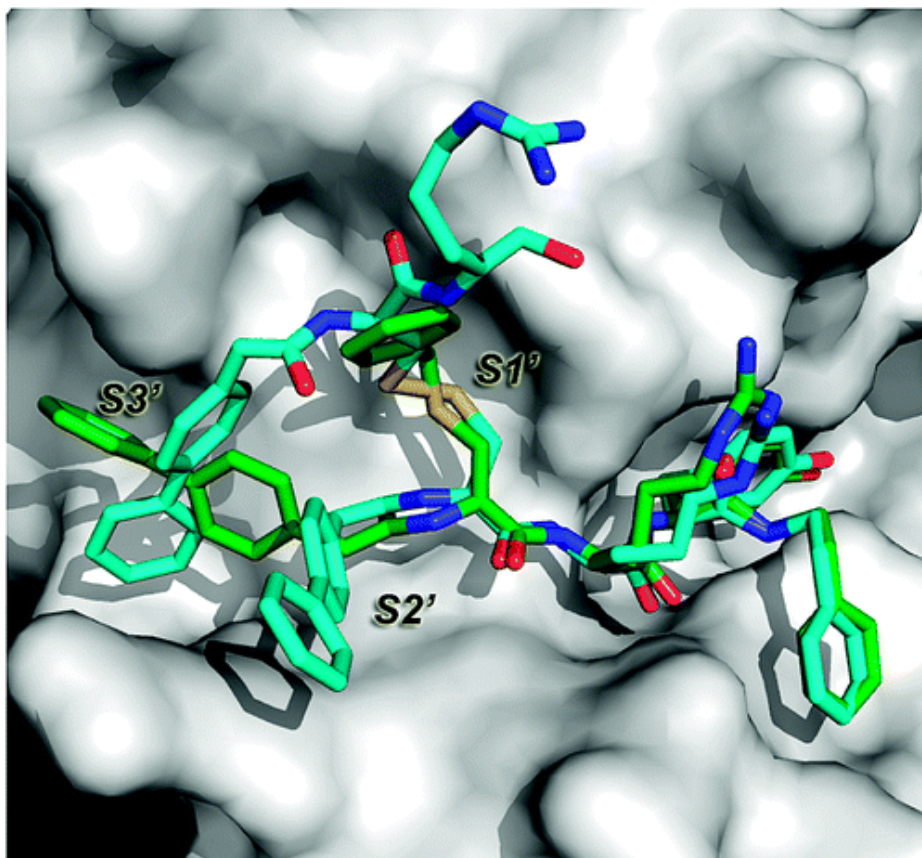
Figure 2.2. Overlay of the 1MHW crystal structure (cyan carbon chain) and the crystal structure of **2** (green carbon chain).

In the current structure, the biphenyl group starts out in S2′ but projects into S3′in an orientation perpendicular to the biphenyl in the dimeric part of inhibitor **1** (PDB code 1MHW). The orientation of the *S*-benzyl group follows the chain of the disulfide bond in 1MHW and positions the center of the benzene ring near the CR of Cys in the second

monomer. These results suggest that the positioning in S2′of the biphenyl rings of the first half of the inhibitor dimer in 1MHW was induced by the presence of the dimer and that the preferred position for the monomer is at S3'. These findings motivated us to determine the structures of remaining complexes, in particular inhibitor **4.** It is interesting to see the crystal structure of inhibitor **4**, which is a close mimic of the dimer, and once again displace the first biphenyl toward the S2'-binding site (please see the section on Inhibitor **4**).

The diaminoketone inhibitors of cathepsin K also have substantial interactions with the S'-binding sites. In Figure 2.3, we aligned the cathepsin K structure (PDB code 1AU0) with the cathepsin L structure in this work and display and overlay of the cathepsin K inhibitor and molecule **2** in the cathepsin L active site. The diaminoketone inhibitor has Cbz-Leu occupying the S′subsites. We see that the Cbz group roughly overlaps with the first half of the biphenyl group of **2** and that Leu side chain occupies part of the region used by the *S*-benzyl group.

Figure 2.4 shows a similar alignment for a vinylsulfone inhibitor (PDB code 1AU2).15 The phenyl rings of the inhibitor overlap with part of the biphenylacetyl group of **2**, and the sulfone S atom is positioned similarly to the *S*-benzyl sulfur of **2**. However, neither the diaminoketone nor vinylsulfone inhibitors reach as far as the second ring of the biphenyl group of **2**. These molecules represent three different classes of inhibitors that make significant interactions with the S′subsites of cathepsins. Of these classes, the inhibitors in this work make the most extensive use of the putative S3′subsite.
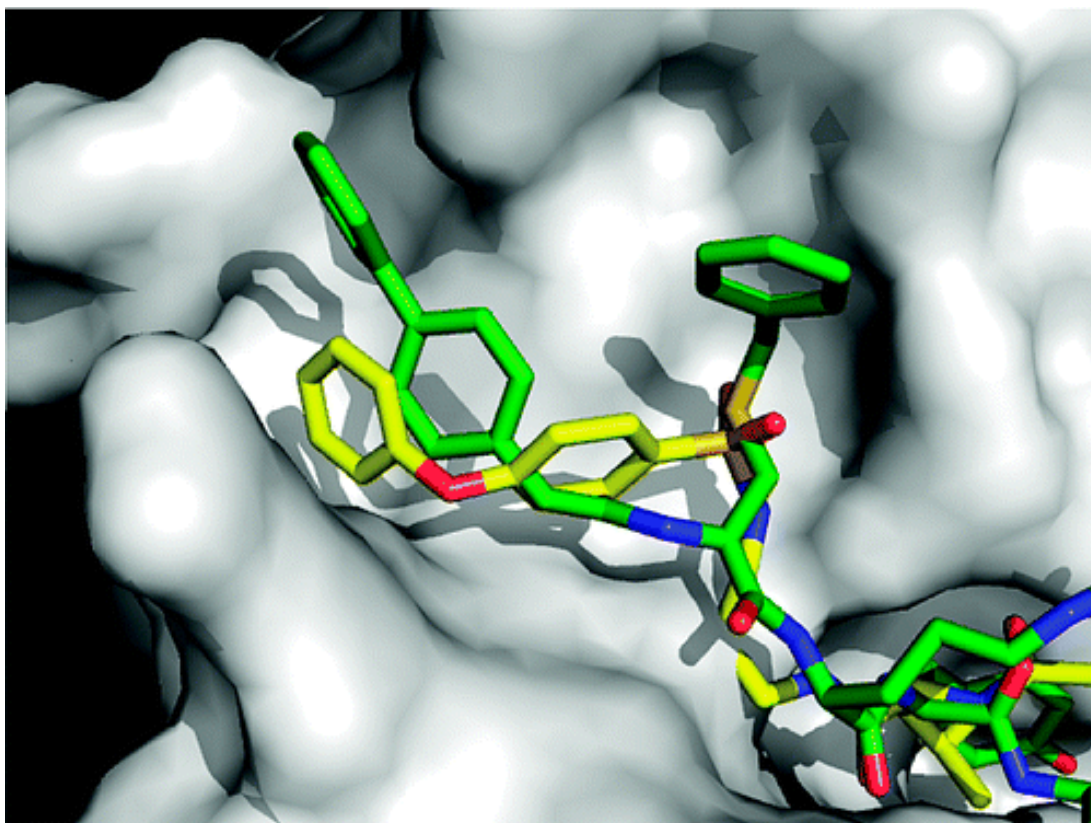
Figure 2.3. Overlay of **2** with a diaminoketone inhibitor (PDB code 1AU0).
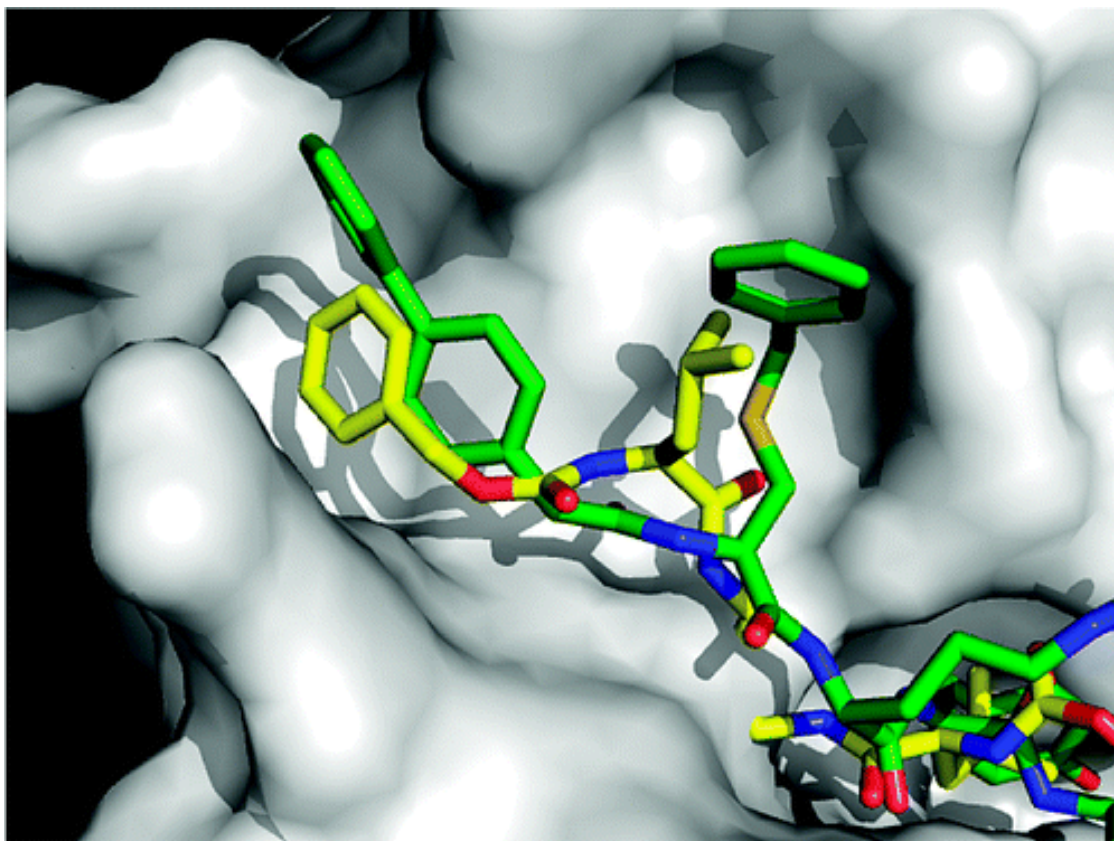
Figure 2.4. Overlay of **2** with a vinylsulfone inhibitor (PDB code 1AU2).

2.3.2 S3' Subsite

Many years ago, Schechter and Berger suggested the existence of seven subsites in papain spanning S4-S3' based on hydrolysis data on a series of polyalanine peptides. However, there is no direct structural evidence for the existence of all these subsites. More recently, Turk and co-workers have revised the definition of substrate-binding sites in cysteine proteases to S3-S2' based on the available crystal structures of inhibitors and substrate analogues. In their analysis, they see a S1'- and S2'-binding site but no evidence for S3'. The crystal structure in the present work reveals a binding site on cathepsin L that extends far beyond the S2' subsite. Of course, this is a crystal structure of an inhibitor,

and there is no evidence that a substrate would actually use the extended binding site. However, the proximity of the binding site to S2' makes it a reasonable binding site for a P3' side chain. Figure 2.5 shows a manually generated polyalanine peptide overlaid onto the crystal structure of inhibitor **2**. We see that there is the potential for the P3' side chain to make use of the extended subsite. Hence, the region around the second ring in the biphenyl of **2** plausibly suggests a putative S3' subsite.
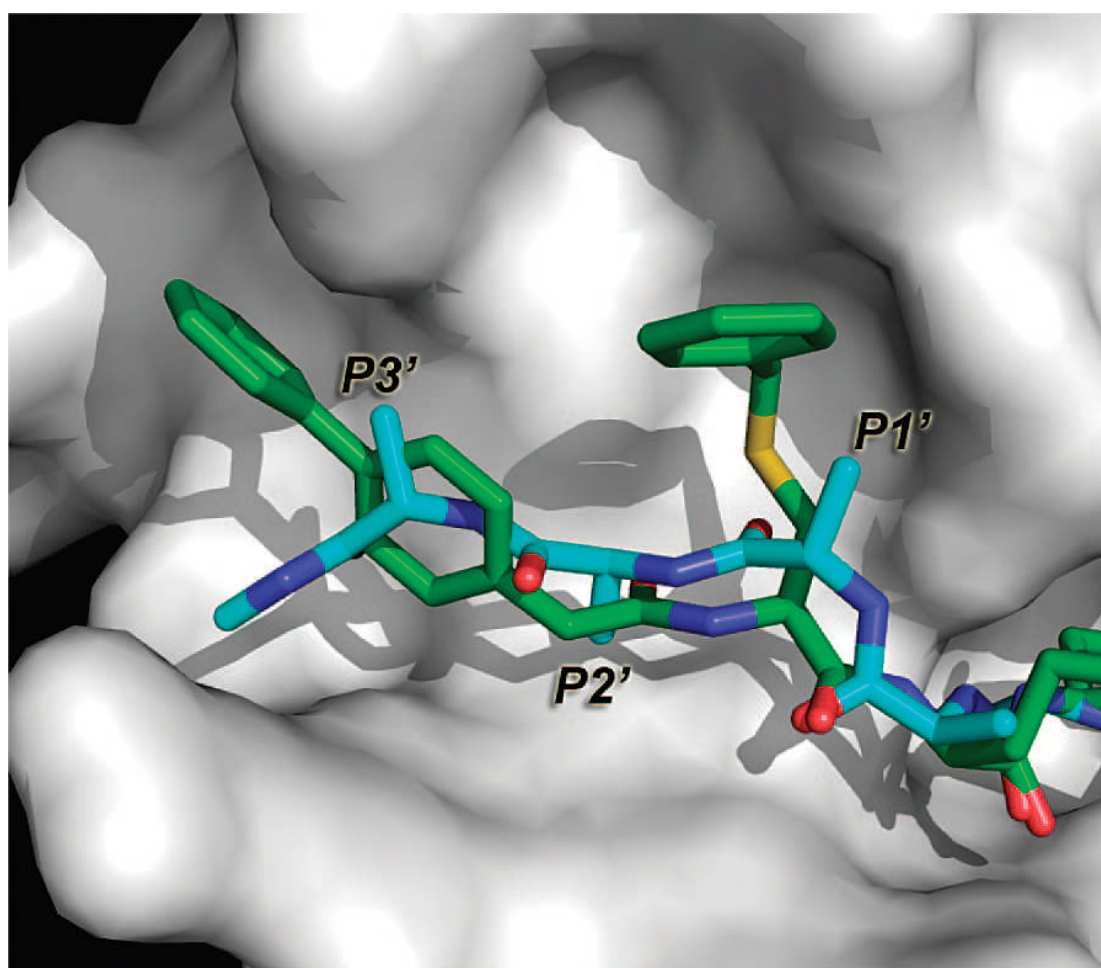
Figure 2.5. Overlay of **2** with a hypothetical polyalanine substrate.

2.3.3 Electrostatics of the S1' Subsite

In the vicinity of the S1' binding is a negatively charged Asp162. In particular, the distance between the sulfur atom of *S*-benzylcysteine and one of the carboyxlate oxygens

of Asp162 is 4.4 Å. This raised the possibility of designing an inhibitor with a positively charged moiety occupying S1' to interact with the carboxylate of Asp162. To explore the electrostatic preferences at the S1' subsite, three inhibitors (**10**, **11**, and **12**) containing neutral, positive, and negative groups for binding at S1' were synthesized and tested. In these inhibitors, the Cys in **1** was replaced by aminobutyric, reported in other charge selectivity calculations in the literature, this value is relatively small. This suggests that although the S1' subsite prefers a neutral group it does not have a strong preference for one.

2.3.4 Design of dimer-mimetic propeptide inhibitors

The previously described design and synthesis of inhibitors **4** and **9**, which were inspired by the crystal structure of inhibitor **1** (pdb code 1mhw) (Chowdhury *et al.*, 2008; Chowdhury *et al.*, 2002). In that structure, inhibitor **1** had inadvertently dimerized through a disulfide bond and positioned two biphenyl groups in the S' subsites of cathepsin L. Inhibitors **4** and **9** were designed to mimic the interactions seen in the inhibitor **1** crystal structure. These inhibitors contain a second biphenyl group without the additional bulk of a second monomer (Table 2.1a). This was accomplished by attaching a biphenylacetyl group to Nε of a Lys side chain in the inhibitor. However, neither **4** nor **9** exhibited improved binding affinity compared to **14**, the most potent compound in the congeneric series.8 In order to understand the lack of improved potency; we solved the crystal structures of **4**, **9** and **14**.

2.3.5 Structure of the dimer-mimetic propeptide inhibitor complexes

The structure of mature human cathepsin L complexed with inhibitors **4, 9** and **14** was solved by using rotating anode and / or synchrotron data sets, and refined up to 2.5, 1.8 and 2.5 Å resolution, respectively. The structures were determined by the molecular replacement method using mature cathepsin L coordinates taken from the procathepsin L pdb (pdb code 1cs8). Each of the models has been refined with good stereochemical parameters (Table 2.2). Statistics for the Ramachandran plot from an analysis using PROCHECK for the three models gave over 83% of non-glycine residues in the most favored regions, and no residues in the disallowed region. Electron density corresponding to the residues Thr175 to Gly179 of mature Cathepsin L is not observed and is presumably disordered. The inhibitor molecules are well defined in the electron density maps. The inhibitors occupy the active site in a noncovalent and reverse-binding mode of inhibition. The overall structure of mature cathepsin L resembles that of mature part of the proenzyme. Thus we will not provide any details of the mature cathepsin L structure.

2.3.6 Inhibitor 4

The binding mode of inhibitor **4** in the S1 to S3 subsites is similar to that seen in previous crystal structures of this class of inhibitors (Figures 2.6 and 2.7). The Arg residue of the inhibitor is located in S1 with its side chain solvent exposed and its backbone carbonyl making a hydrogen bond with the NH group of Gly68.

Table 2.2 Crystallographic data and refinement statistics

	Inhibitor 4	Inhibitor 9	Inhibitor 14
Space group	C2	C2	$P2_12_12_1$
Cell parameters			
Resolution range (Å)	50-2.89	50-1.78	50-2.5
Wavelength (Å)	1.000	1.000	1.5418
Observed *hkl*	62123	864300	120192
Unique *hkl*	24602	118469	16224
Completeness (%)	91.2	99.4	99.1
Overall $I/\sigma I$	14.9	15.4	17.8
[a]R_{sym}	0.091	0.154	0.067
Refinement and quality of the model			
*Resolution range	25-2.8(Å)	45.0 – 1.8(Å)	25.0-2.5(Å)
[b]Rwork (%) no. reflections	79.2(34041)	86.6 (100135)	84.1(13732)
[c]Rfree (%) no. reflections	7.0(3002)	7.6 (8770)	5.1(838)
Root mean square deviation			
Bond length (Å)	0.008	0.01	0.008
Bond angle (°)	1.3	1.3	1.13
Ramachandran plot (%)			
Favored region	84.2	84.7	82.5
Allowed regions	15.8	15.3	17.5
Generously allowed region	0.0	0.0	0.0
Disallowed regions	0.0	0.0	0.0
[d]Average B-factors (Å²)			
Main chain atoms	25.648	23.972	54.954
Side chain atoms	30.274	30.763	56.493
Overall protein atoms (no. atoms)	27.876(9954)	26.756(9953)	55.69(3322)
Waters (no. atoms)	26.094(226)	32.942(562)	60.045(129)
Ligand (no. atoms)	42.226(426)	42.297(420)	29.342(106)

[a] $R_{sym} = |I_i - <I>|/ |I_i|$ where I_i is the intensity of the ith measurement, and $<I>$ is the mean intensity for that reflection. [b] $R_{work} = | F_{obs} - F_{calc}|/ |F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. [c] R_{free} = as for R_{work}, but for 8.5% of the total reflections chosen at random and omitted from refinement. [d] Individual B-factor refinement were carried out.

* Reflections greater than $I>\sigma I$ where used in the refinement

In the S2 subsite, the inhibitor Tyr side chain has nonpolar interactions with Leu69, Met70 and Ala135 and its amide NH forms a hydrogen bond with the Gly68 carbonyl. The N-(2-6 phenylethyl)-amide group packs against the Gly68 main chain and the side chain of Leu69. Inhibitor **4** makes fairly extensive interactions with the S' subsites. The *N*ε-biphenylacetyl modified Lys side chain is in a fairly extended conformation in the S' region. The distance from the Lys Cα to the carbon at the tip of the biphenyl group is 15.6Å. The biphenyl rings of the *N*ε-biphenylacetyl-Lys side chain interact with the side chains of Glu141, Phe145 and somewhat more distantly with Leu144. The acetyl methylene group interacts with Cα of Gly139. The Lys side chain methylene groups pack against the protein surface. The amide group of the *N*ε-biphenylacetyl-Lys side chain forms a hydrogen bond with the carbonyl of Ala138. In the S2' subsite, the 4-biphenylacetyl rings makes hydrophobic interactions with the side chain of Leu144. The carbonyl oxygen of the biphenylacetyl group makes a hydrogen bond with the indole nitrogen of Trp189. Figure 2.8 shows the simulated annealing Fo-Fc omit map for inhibitor **4**.
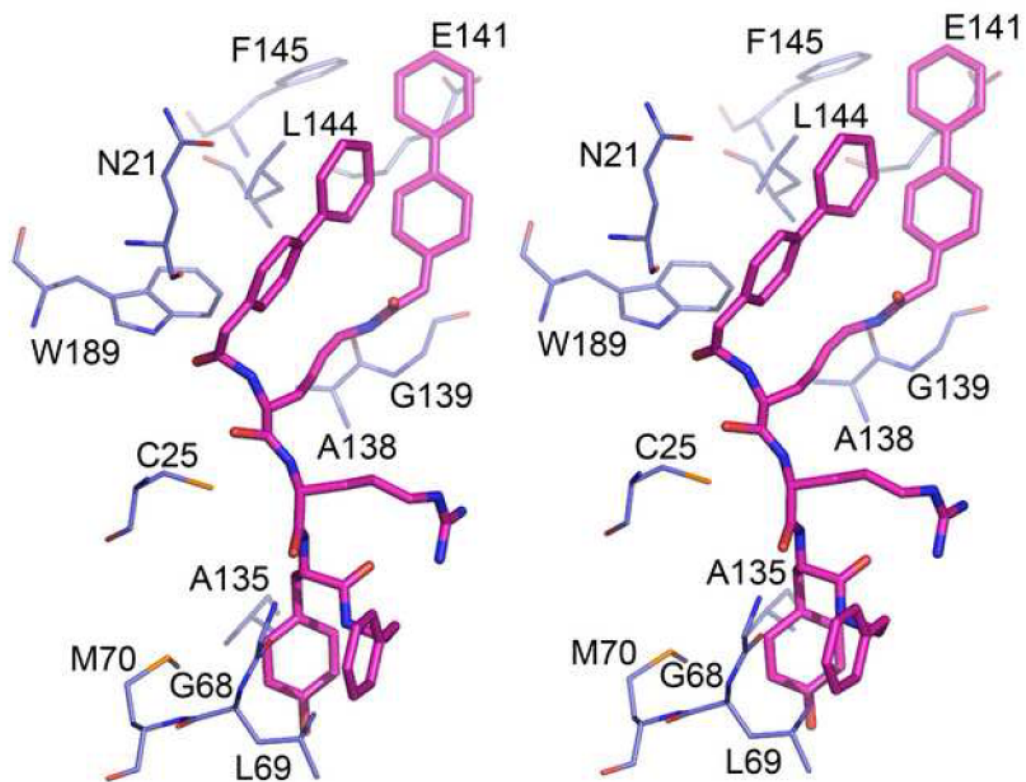
Figure 2.6. Stereo view of interactions between the inhibitor 4 and cathepsin L. Protein atoms are shown in thin lines whereas the inhibitors are shown in thick lines. The figure was prepared using PyMOL (DeLano Scientific, Palo Alto, CA)
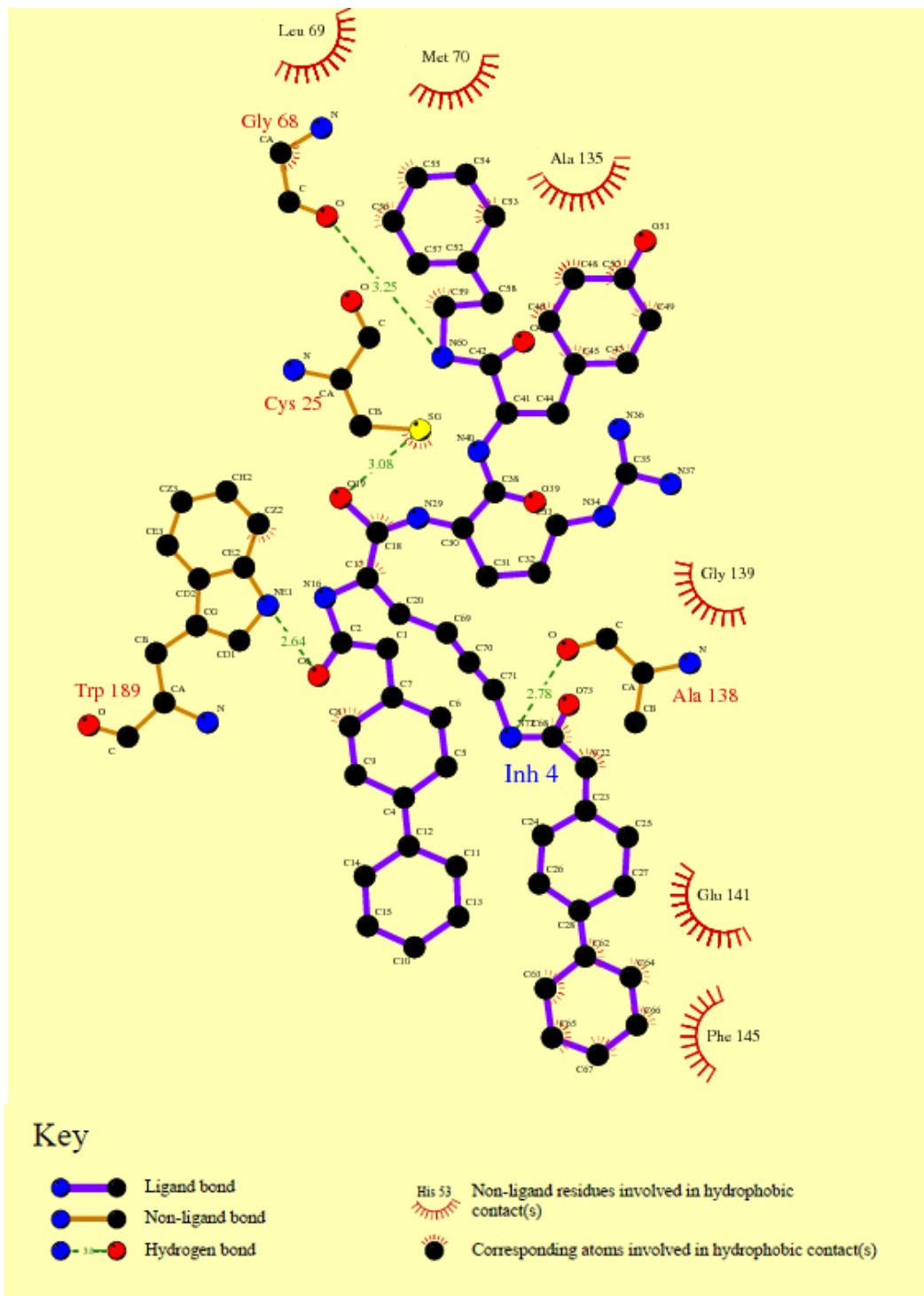
Figure 2.7. Schematic view of cathepsin L interactions with inhibitor **4**. The Figure was prepared using LIGPLOT (Wallace *et al*, 1995)
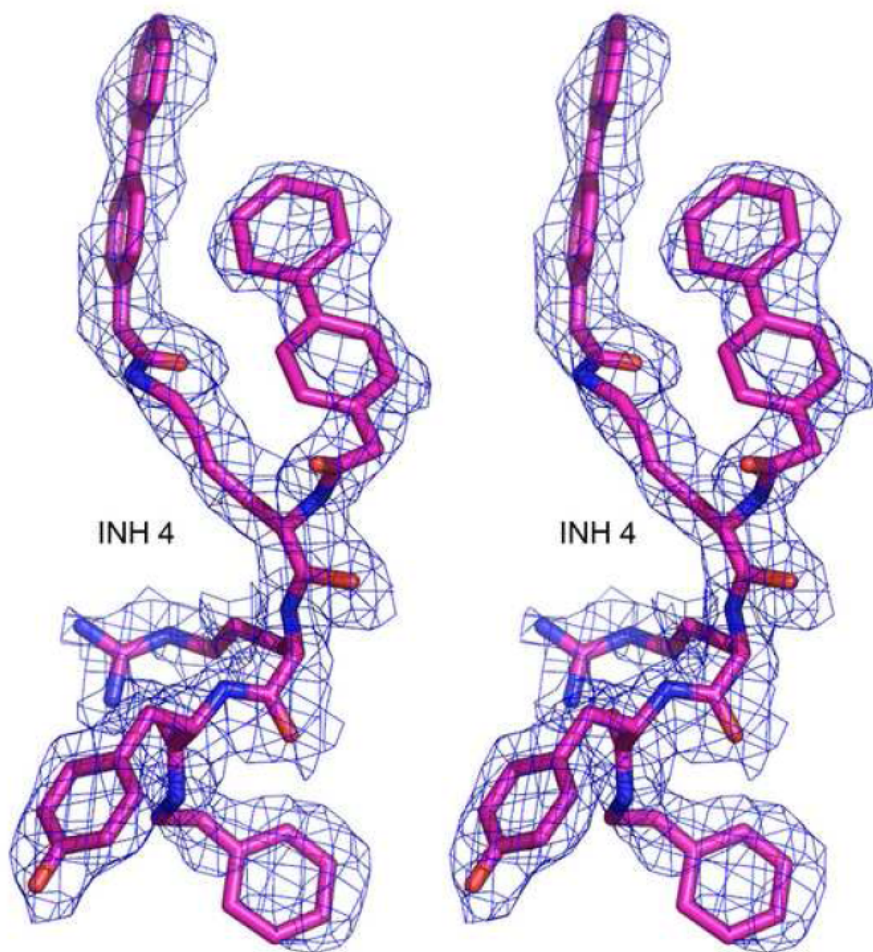
Figure 2.8. Stereo view of the simulated annealing Fo-Fc omit map in the active site region of cathepsin L. The bound inhibitor 4 and all atoms within 3Å of the inhibitor molecule were omitted prior to refinement. The map is contoured at a level of 2σ for inhibitor **4**. The figure was prepared by using PyMOL (DeLano Scientific, Palo Alto, CA).

Examination of the crystal packing contacts revealed that the S' regions of two adjacent protein molecules face each other and create a shared binding cavity at the interface that subsequently gets occupied by parts of the inhibitors bound to each protein unit. The biphenyl rings of the inhibitor bound to one protein interact with the biphenyl rings of the other inhibitor bound to the other protein (Fig 2.9). The Nε-biphenyl of one protein is packed between the backbone biphenyl and the Gln21 side chain of the other
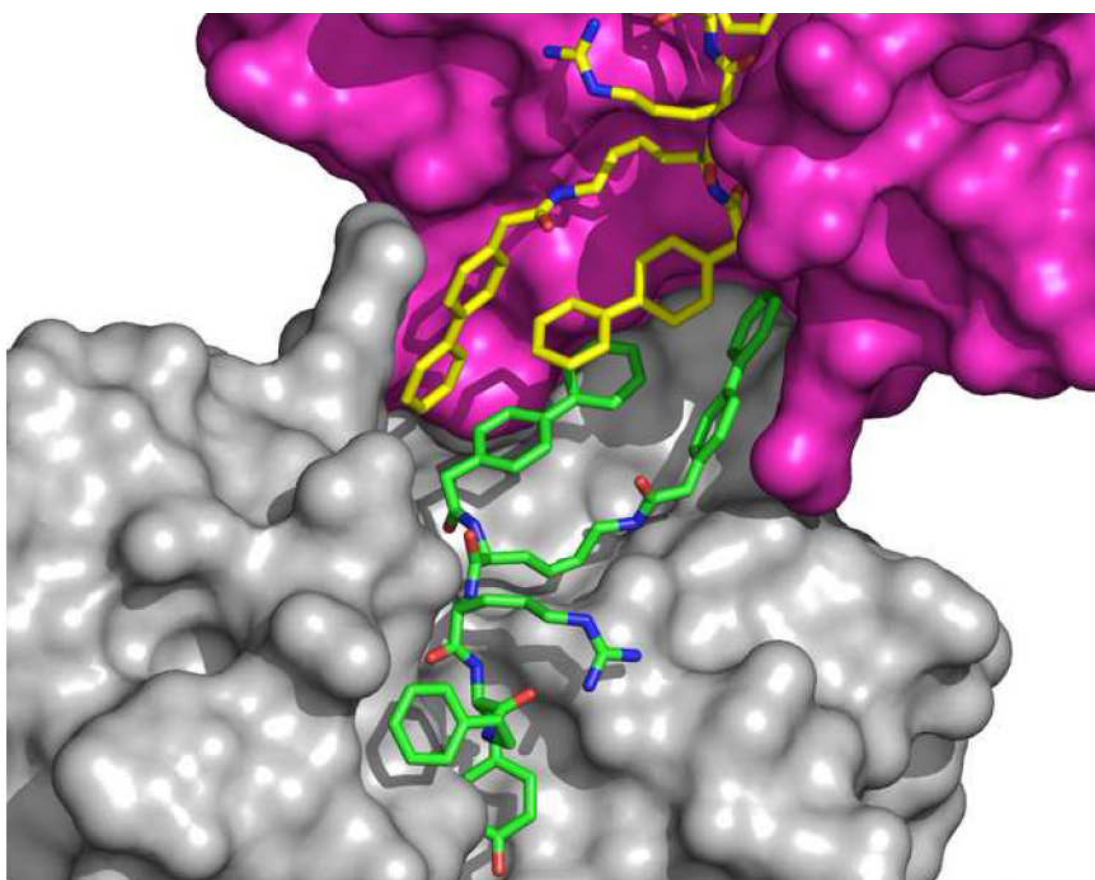
Figure 2.9. Crystal packing interactions of two adjacent protein-ligand complexes in the unit cell of Inhibitor **4** and cathepsin L complex

protein. It is thus possible that the bound conformation of the inhibitor in the S' subsites in this crystal structure may not be representative of that found in an isolated protein-

ligand complex in solution. Binding in the S subsites, however, does not seem to be affected by the crystal packing.

2.3.7 Inhibitor 9

The binding mode of inhibitor **9** in the S1 to S3 subsites is essentially the same as that seen for inhibitor **4** (Figures 2.10 and 2.11). D-Arg is found in the S1 subsite with the side chain guanidium group exposed to solvent and backbone carbonyl hydrogen bonded to the Gly68 amide. The Phe side chain makes nonpolar interactions with the side chains of Leu69, Met70 and Ala135 in the S2 subsite. In the S3 subsite, the NH group of the N-(2-phenylethyl)-amide makes a direct hydrogen bonding contact with the main chain carbonyl oxygen of Gly68. The phenyl ring of the N-(2-phenylethyl)-amide group packs
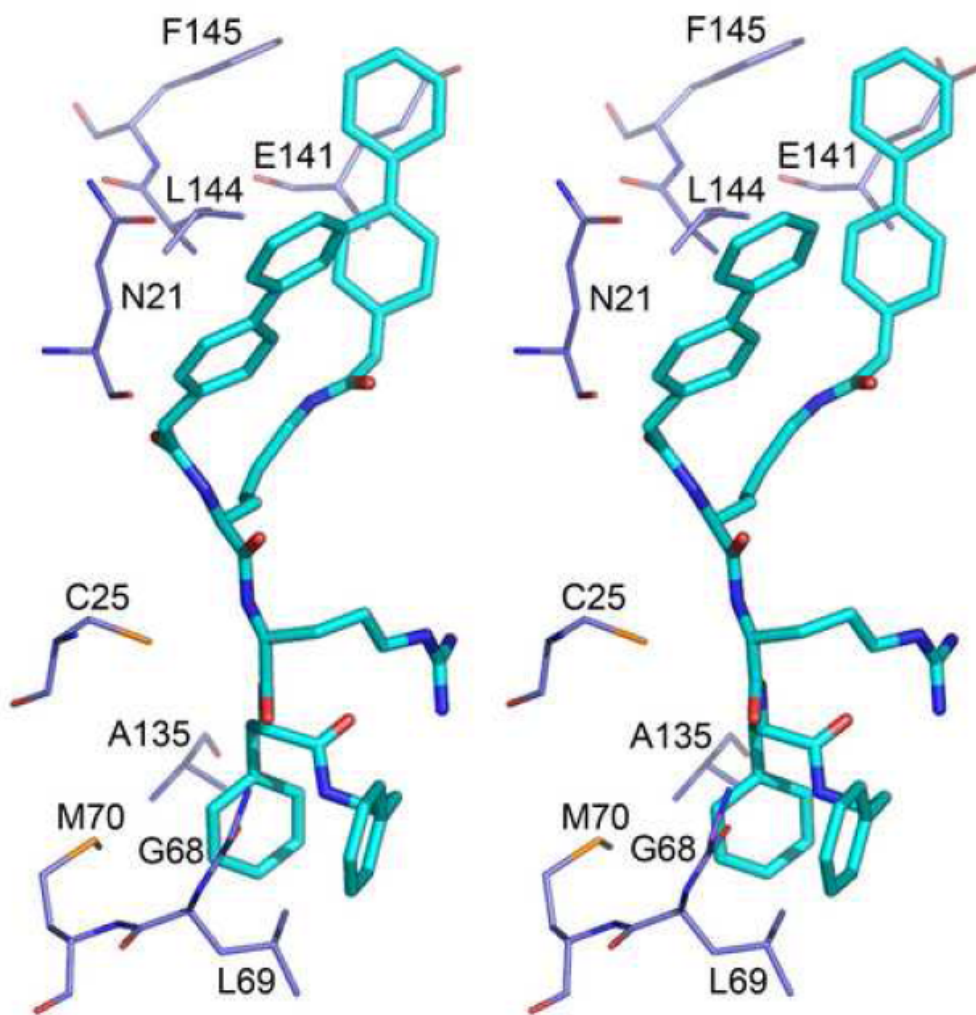
Figure 2.10. Stereo view of interactions between the inhibitor 9 and cathepsin L. Protein atoms are shown in thin lines whereas the inhibitors are shown in thick lines. The figure was prepared using PyMOL (DeLano Scientific, Palo Alto, CA)
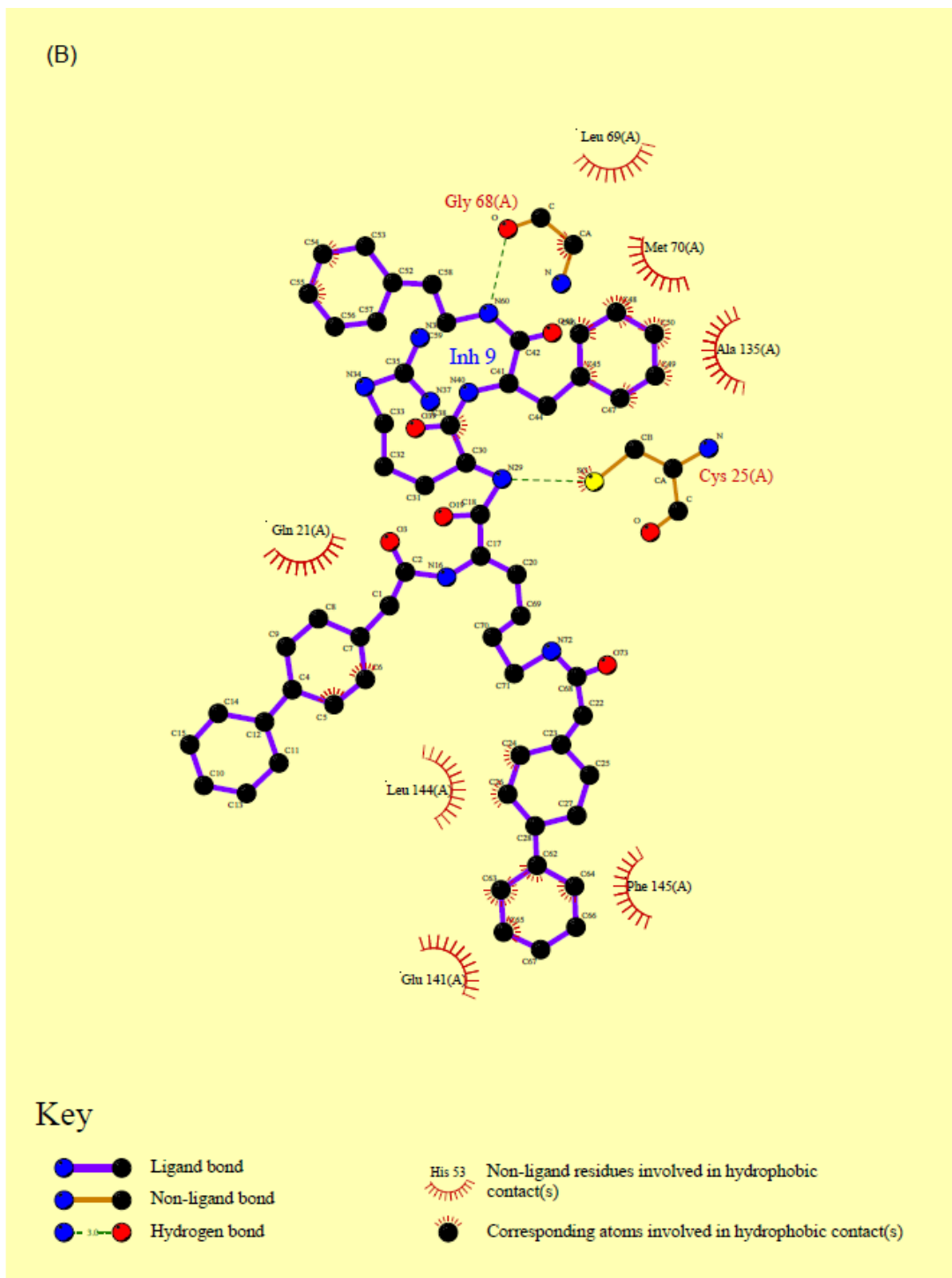
Figure 2.11. Schematic view of cathepsin L interactions with inhibitor **9**. This figure was prepared using LIGPLOT (Wallace *et al*, 1995)

against the backbone atoms of Gly68 and the side chain of Leu69. The orientation of the *N*ε-biphenylacetyl-Lys side chain in the S' subsites is very similar between inhibitors **4** and **9** but the position of the other biphenylacetyl group is significantly different in the two structures. As with inhibitor **4**, the rings of the *N*ε- biphenylacetyl-Lys side chain interacts with the side chains of Glu141, Leu144 and Phe145. The amide group of the *N*ε-biphenylacetyl-Lys makes a direct hydrogen bond with the carbonyl of Ala138. Unlike inhibitor **4**, the other biphenylacetyl group does not appear to have extensive interactions with the protein. Its main points of contact are nonpolar interactions with the backbone and side chains of Gln21.

As in inhibitor **4**, examination of the symmetry related contacts revealed that the S' regions of two adjacent proteins face each other. The backbone biphenyl from one inhibitor stacks against the side chain biphenyl from the other inhibitor (Fig 2.12). It is possible that the binding mode in the S' sites is not representative of that in the isolated complex in solution. The binding in the S subsites appears unaffected by the crystal packing interactions. Figure 2.12 shows the simulated annealing Fo-Fc omit map for the inhibitor **9**. In inhibitor **4** the electron density of the inhibitor backbone biphenyl is somewhat weak and even more so in inhibitor **9**. This suggests that this group is highly mobile and may be fluctuating between the two positions represented in the two crystal structures.
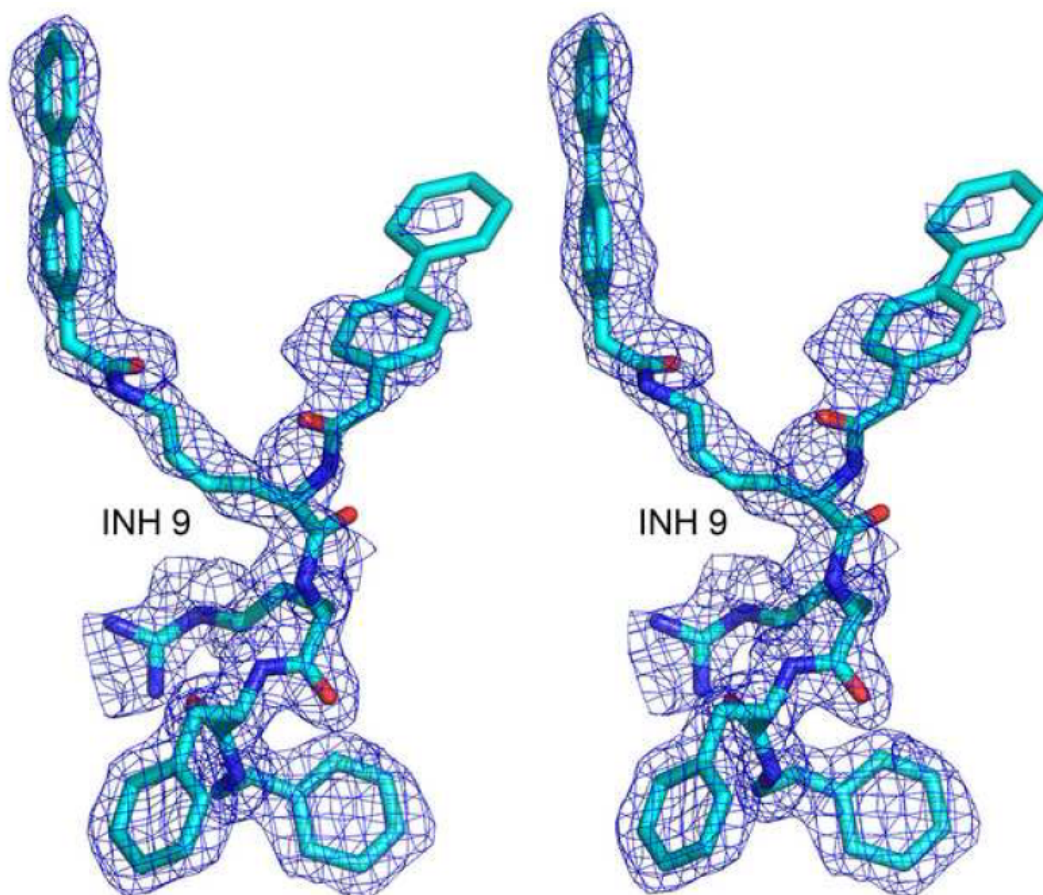
Figure 2.12. Stereo view of the simulated annealing Fo-Fc omit map in the active site region of cathepsin L. The bound inhibitor 9 and all atoms within 3Å of the inhibitor molecule were omitted prior to refinement. The map is contoured at a level of 2.5σ for inhibitor **4**. The figure was prepared by using PyMOL (DeLano Scientific, Palo Alto, CA).
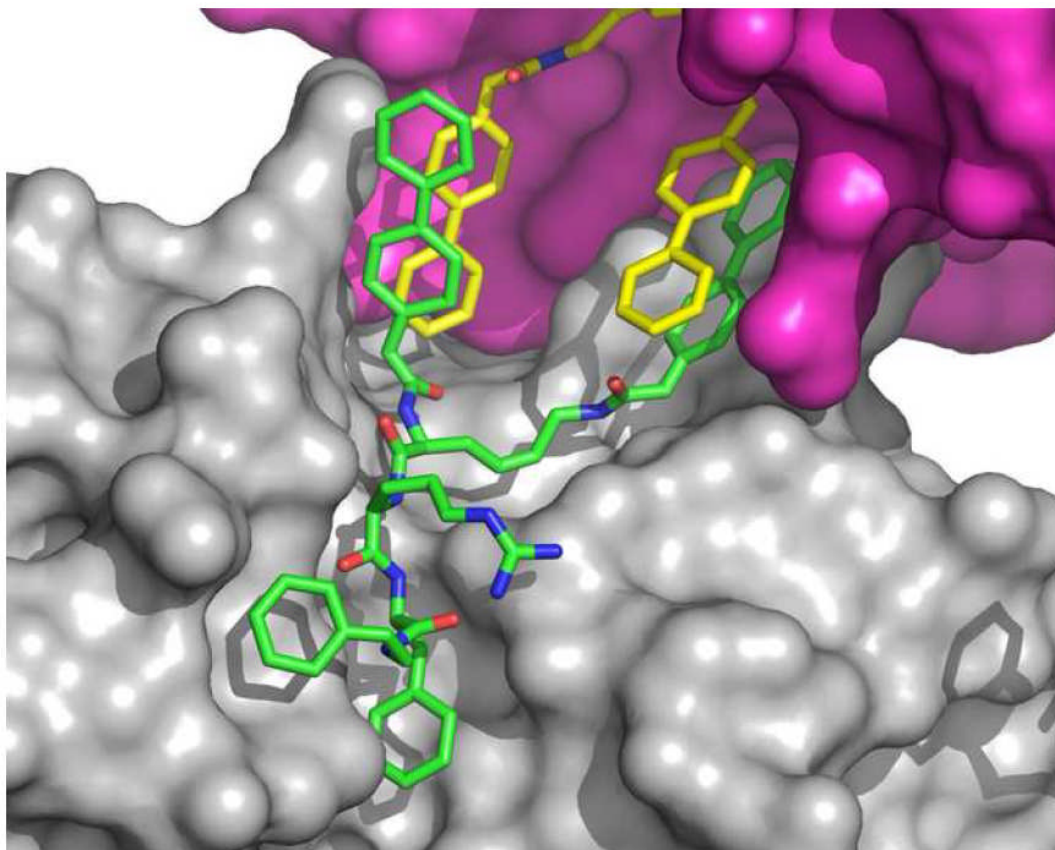
Figure 2.13. Crystal packing interactions of two adjacent protein-ligand complexes in the unit cell of Inhibitor **9** and cathepsin L complex.

2.3.8 Inhibitor 14

Inhibitor **14** is a close analog of inhibitor **1**, the first compound in this class that was crystallized (pdb code 1mhw). The main difference is the replacement of Cys by S methylCys (or MCys, for short). This precludes the formation of a disulfide-bridged dimer as occurred in the crystallization of inhibitor **1**. This inhibitor represents the original design of this class of compounds. The binding mode of inhibitor **14** at the S1–S3 subsites is essentially the same as that of inhibitors **4** and **9** and will not be discussed further (Figures 2.14 and 2.15). In the S1' subsite, the MeCys makes nonpolar

interactions with the side chain of Ala138. The 4-biphenylacetyl group is found to be in a conformation different from inhibitors **4** and **9** and is in a very extended conformation reaching towards putative S3' subsite (Fig 2.14) discussed previously. The biphenyl group stacks against Trp189 and interacts with Gln21 on the other face of the rings. At the farthest tip of the biphenyl group there is some contact with the Cα of Trp193. The simulated annealing *Fo-Fc* omit map for inhibitor **14** shows that the position of the biphenyl rings is reasonably well-defined (Fig 2.16). Unlike in the crystal structures of inhibitors **4** and **9**, examination of crystal packing interactions in the structure of inhibitor **14** do not show significant interactions between inhibitors from different symmetry related protein molecules. However, there is a symmetry-related interaction between the side chains of corresponding Leu144 residues that modifies the binding surface near the inhibitor biphenyl rings (see below),
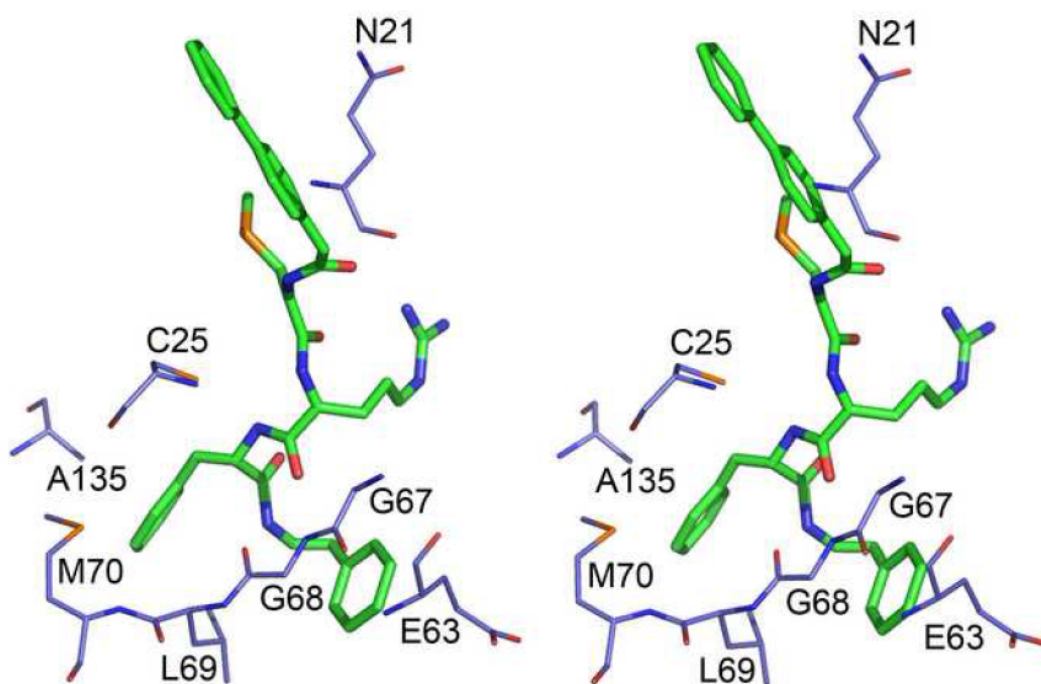
Figure 2.14. Stereo view of interactions between the inhibitor **14** and cathepsin L. Protein atoms are shown in thin lines whereas the inhibitors are shown in thick lines. This figure was prepared using PyMOL (DeLano Scientific, Palo Alto, CA)
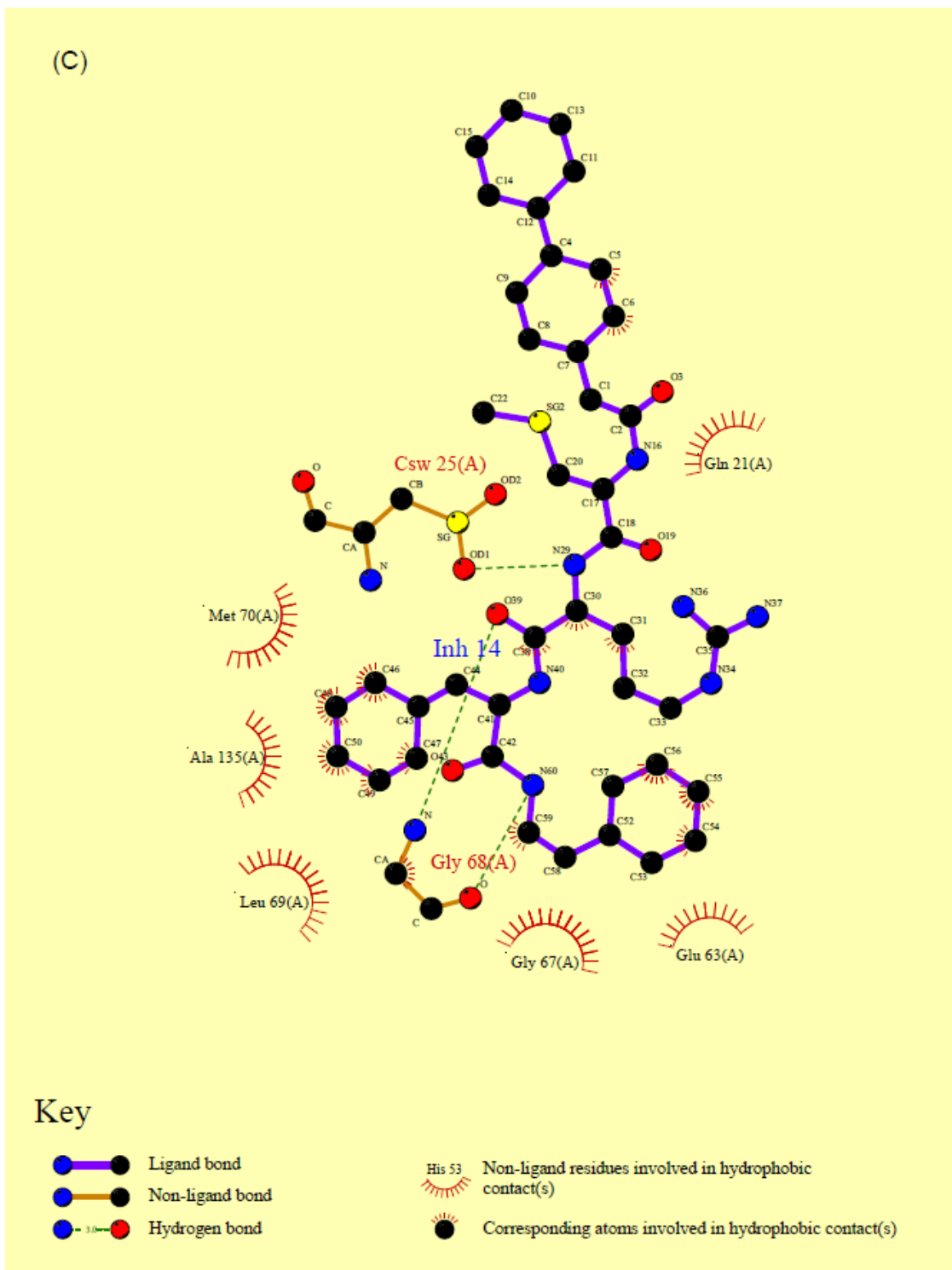
Figure 2.15. Schematic view of cathepsin L interactions with inhibitor **9**. The figure was prepared using LIGPLOT (Wallace *et al*, 1995)
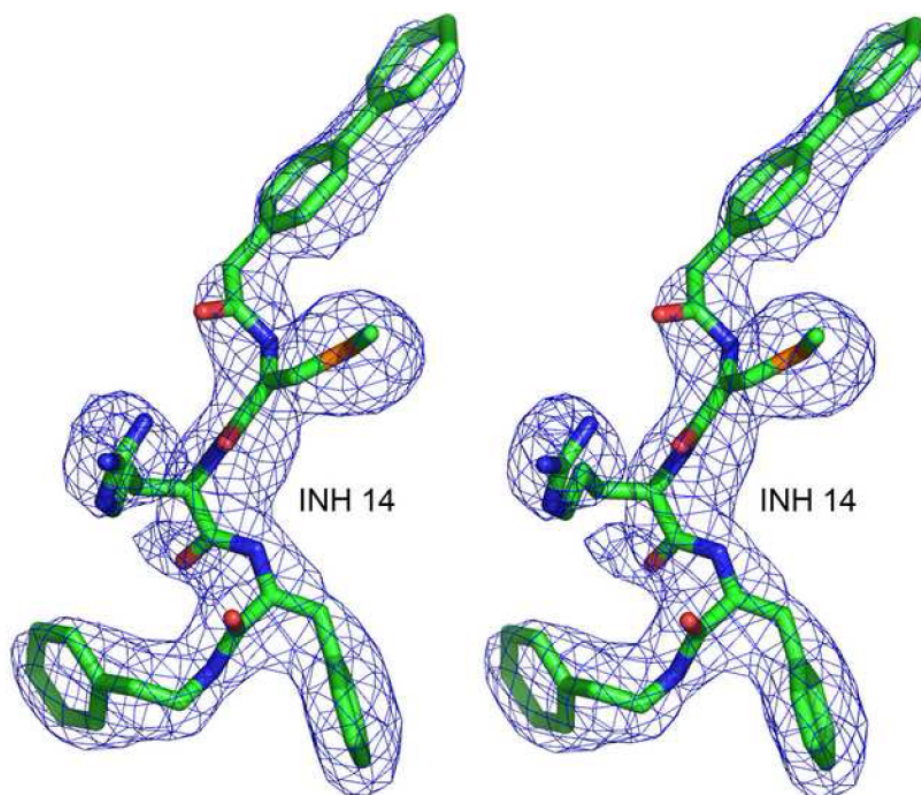
Figure 2.16. Stereo view of the simulated annealing Fo-Fc omit map in the active site region of cathepsin L. The bound inhibitor **14** and all atoms within 3Å of the inhibitor molecule were omitted prior to refinement. The map is contoured at a level of 2σ for inhibitor **4**. The figure was prepared by using PyMOL (DeLano Scientific, Palo Alto, CA).

In the section 2.3.1 of this chapter we have discussed the crystal structure of inhibitor **2** (pdb code 3bc3), a close congener of **14**. A comparison of the bound conformations of the two inhibitors shows a similar binding mode with the biphenyl rings extended towards the putative S3' binding site and the biphenyl rings stacked against the Trp189 rings. However, in the crystal structure of **2** the biphenyl ring overlaps more with the Trp189 rings and comes close to the Phe145 side chain. In the crystal structure of **14** a

90

symmetry-related interaction between side chains of corresponding Leu144 residues occludes the region between Trp189 and Phe145, altering the binding mode of the biphenyl group relative to that seen in the crystal structure of **2**. In order to explore how the binding modes might change in the absence of crystal packing interactions, in collaboration with Prof Enrico, McGill University, Canada, the molecular dynamics simulations were run as described in the next section.

2.3.9 Molecular Dynamics

Molecular dynamics simulations of cathepsin L protein complexed with **4, 9**, and **14** were carried out in order to further clarify the binding mode of the inhibitors. Trajectories of 12 ns were generated for each of the inhibitors. Figure 2.17 shows plots of the time series of selected distances describing the positions of S'-binding groups of **4**. The S'-binding moieties of **4** were highly mobile during the first half of the simulation before finally settling down to a reasonably stable conformation. The time series of the distance between C4 of the backbone biphenyl (see Figure 2.7 for ligand atom numbering) and the CG atom of Trp184 is shown in Figure 2.17a. The backbone biphenyl group fluctuates significantly during the first 6 ns and then ends up packing against the Trp184 indole ring. The side chain biphenyl is even more mobile during the first 6 ns before finally packing against the backbone biphenyl. The side chain Nε-biphenyl C62 carbon and the backbone biphenyl C4 ranged from 10-12 Å apart for much of the first 6 ns before coming together to a distance of about 5 Å (Figure 2.17b). The lysine side chain Nε of **4** is initially hydrogen bonded to the Ala138 carbonyl O and remains so for the first half of the simulation (Figure 2.17c).
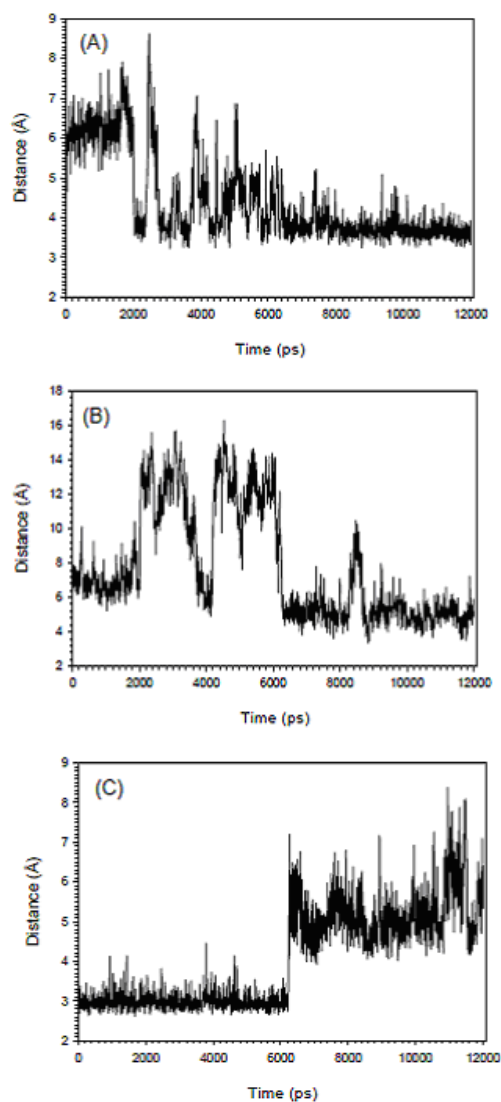
Figure 2.17. Interatomic distance fluctuations involving inhibitor **4**. *(a)* Trp184 Cγ – Inhibitor C4, *(b)* Inhibitor C4 – inhibitor C62 (biphenyl-biphenyl interaction), *(c)* Ala138 carbonyl O – Inhibitor Lys Nε. See Figure 2.7 for atom numbering of inhibitor **4.**

The hydrogen bond is eventually broken as the two biphenyl groups come together. The Nε remains in a fairly stable position for the remainder of the simulation. Figure 2.18 shows a similar time series for selected distances describing the positions of S'-binding

groups of inhibitor **9**. The evolution of the distance between the C4 of the backbone biphenyl (see Figure 2.11 for ligand atom numbering) and the CG atom of Trp184 is shown in Figure 2.18a. The backbone biphenyl group quickly settles down after less than 100 ps and then ends up packing against the Trp184 indole ring. The side chain Nε-biphenyl group is much more mobile than its counterpart in **4** and does not end up packing against the other biphenyl group (Figure 2.18b). At the end of the 12 ns simulation, it hasn't really settled down to a defined binding site. This is consistent with the lower binding affinity observed for **9** as compared to **4** and **14**. Also unlike in **4**, the lysine side chain Nε of **9** manages to maintain hydrogen bond to the Ala138 carbonyl O for most the 12 ns trajectory (Figure 2.18).

In contrast to **4** and **9**, the binding mode of **14** was quite stable throughout the 12-ns simulation. Figure 2.19 shows time series of the distance between the C4 carbon of the backbone biphenyl and the CG atom of Trp184. It remains stable at around 4.5 Å throughout the 12 ns. The position of the MCys side chain also remains stable throughout. The MD-refined structure of **4** suggests a possible explanation for its lack of improved potency compared to **14** despite the addition of an extra biphenyl group to interact with the S' subsites of cathepsin L. Figure 2.20 shows a snapshot towards the end of the trajectory that is representative of the final ensemble of conformations. The side chain biphenyl seems to form largely intramolecular interactions with the backbone biphenyl.
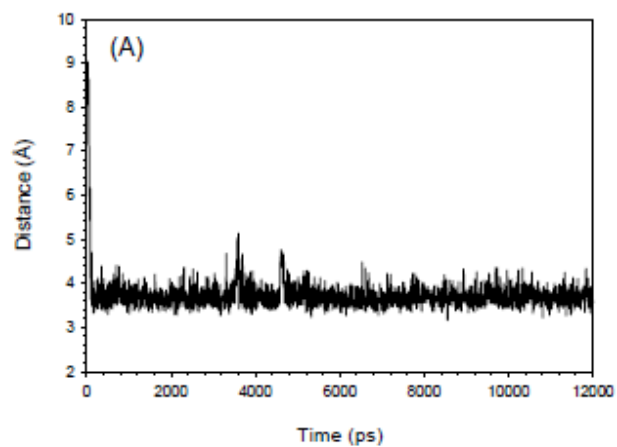
Figure 2.18a. Interatomic distance fluctuations involving inhibitor **9**. *(a)* Trp184 Cγ –
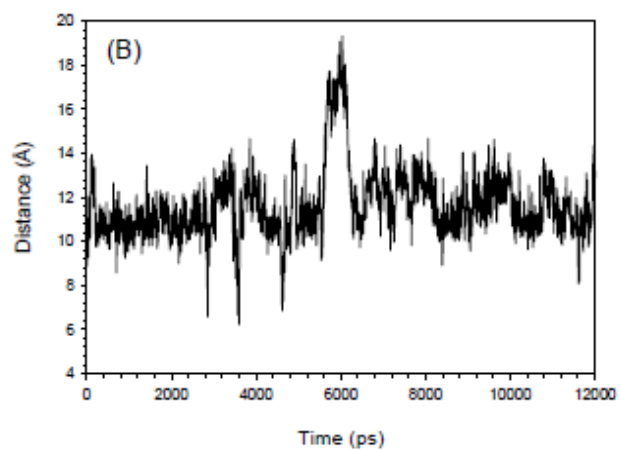
Inhibitor C4,

Figure 2.18b Interatomic distance fluctuations involving inhibitor **9**. Inhibitor C4 –

inhibitor C62
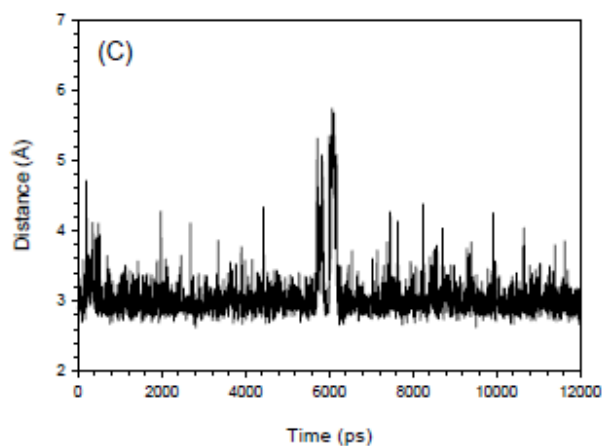
Figure 2.18c. Interatomic distance fluctuations involving inhibitor **9**. *(c)* Ala138 carbonyl O – Inhibitor Lys Nε. See Figure 2.11 for atom numbering of inhibitor **4.**
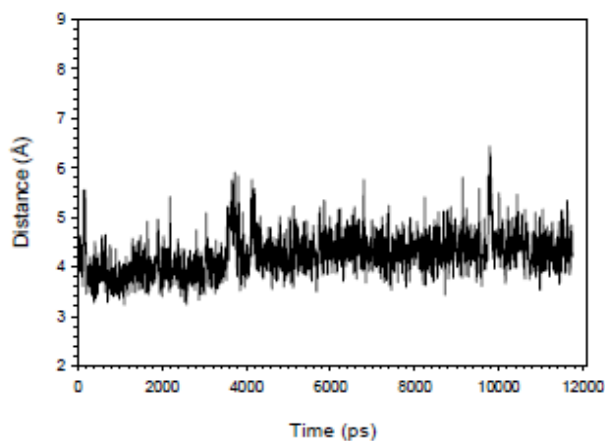
Figure 2.19. Interatomic distance fluctuations involving inhibitor **14**. Trp184 Cγ – Inhibitor C4. See Figure 2.15 for atom numbering of inhibitor **14.**
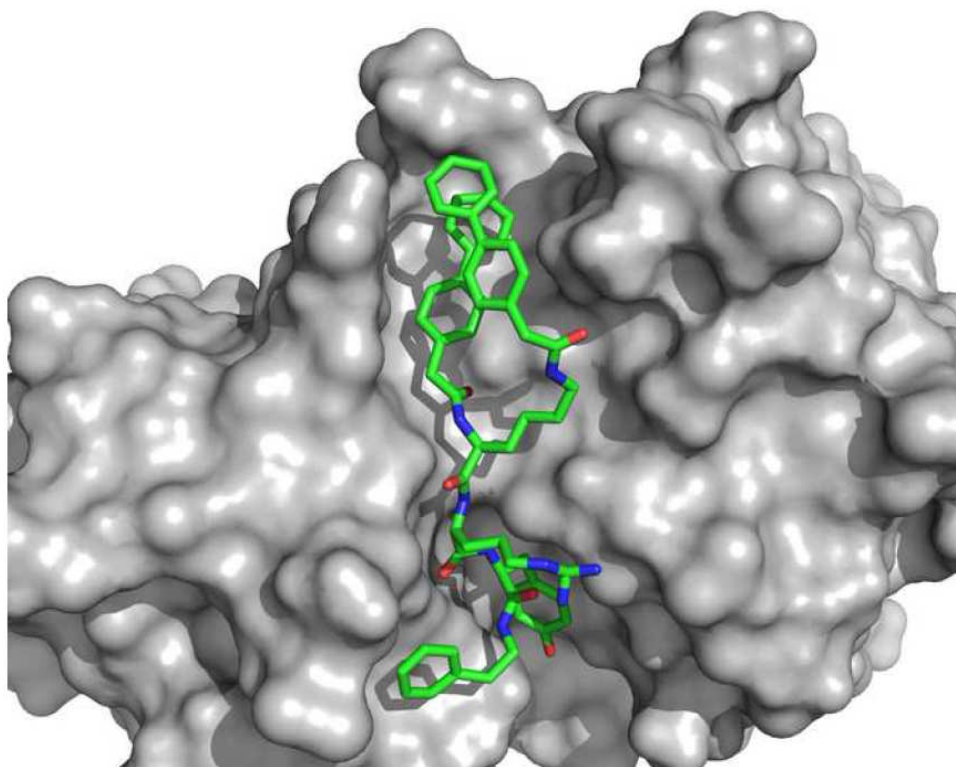
Figure 2.20. Snapshot of MD-refined structure of inhibitor **4**.

The diminished intermolecular interactions together with the larger entropic cost of restricting many rotatable bonds of the Nε- biphenylacetyl-lysine group leads to no net gain in binding affinity compared to **14**. Perhaps forming a macrocyclic structure instead of two independent biphenyls would reduce the entropic cost and result in improved potency.

2.4 Conclusion

In this chapter we have reported the crystal structures of four inhibitors in complex with mature cathepsin L. Through this work, we have established the binding conformation of the monomeric form of our retro-binding peptidomimetics. The conformation at the S1-S3 subsites was essentially identical to that previously reported

for the dimeric form. However, the disposition of the biphenyl group at the S' subsites was significantly altered. We explored extending the inhibitor to increase interactions with the S' subsites. Although these extensions appear to be well-accommodated, they do not confer enhanced binding affinity, perhaps because of increased entropic costs upon binding. The electrostatic preferences at the S1' subsite were also analyzed. A neutral group is slightly preferred at S1'. Further to explore the S' subsite, we have selected three more inhibitors to help clarify and elucidate the binding mode of this class of inhibitors. Of particular interest was the disposition of the biphenyl groups in the S' subsites of the enzyme since the addition of a second biphenyl group to the inhibitor does not improve potency. Due to some question about the effects of symmetry-related interactions in the crystal, we have also carried out molecular dynamics simulations to supplement the information provided by the crystal structure studies and to explore the dynamical behavior of these inhibitors in the active site. After the MD simulations, all three inhibitors have one biphenyl group packing against the Trp189 rings. In inhibitor **4**, the second biphenyl packs against the first one. In inhibitor **9**, the second biphenyl group is highly mobile and does not appear to have a highly preferred binding site. The behavior of these inhibitors in the MD simulations is qualitatively consistent with the experimentally observed potencies of these inhibitors. Overall the inhibitors (especially inhibitor **2**) described in this work have the most extensive interactions with the S' subsites of cathepsins reported in the literature thus far. These interactions are highly suggestive of the existence of a S3' subsite that can potentially be used for enhanced specificity and affinity.

Chapter III

Crystal Structures of Human Cathepsin L Complexed with a Peptidyl Glyoxal Inhibitor and a Diazomethylketone Inhibitor

3.1 Introduction

Cathepsin L a member of the cysteine proteases of the papain superfamily is responsible for the the breakdown of proteins in the lysosomes and plays an important role in antigen processing, tumour invasion and metastasis (Tomssen *et al*, 1995) and bone resorption (Kakegawa *et al*, 1993).   The catalytic triad of Cathepsin-L consists of Cys25, His163 and Asn187. The substrate binding cleft of Cathepsin L has six binding pockets and denoted as S3, S2, S1, S1', S2', and a putative S3' identified in Schechter and Berger annotation (Schechteri & Berger, 1967). These subsites are located along the catalytic cleft between the left (L) and right (R) domains of Cathepsin L. The unprimed subsites S1, S2 and S3 bind the N terminal of the substrate or inhibitor whereas the primed subsites accommodate the C-terminal part of the substrate. The substrate specificity of Cathepsin L is determined by the presence of an amino acid with a large hydrophobic side chain at the P2 position (Shaw *et al*, 1993). Cathepsin L cleaves substrates near bulky aromatic residues which bind to subsites S2 and S3 (Kirschke *et al*, 1988).

Synthetic inhibitors of cysteine proteases are composed of two parts. The first part is the peptidyl fragment which interacts with the binding pockets. This peptidyl fragment has a protecting group like Carboxybenzyl, tosyl (T) or t-butoxycarbonyl (Boc) (Janowski *et al*, 2004). The sequence of the peptidyl group is usually derived from natural protease inhibitors like human cystatins and E64, a natural irreversible inhibitor of cysteine proteases (Grubb *et al*, 1990). The second part of the synthetic inhibitors of cysteine proteases consists of the reactive group to interact with the active site. This

reactive site group includes aldehydes, nitriles, azapeptides, halomethylketones, diazomethylketones, and their intermediates like α-keto-β-aldehydes. There are several classes of compounds are known to inhibit cathepsin L, including diazomethyl ketones, acyloxymethyl ketones, epoxysuccinyl derivatives (such as E-64) and peptidyl aldehydes, such as leupeptin and peptide α-keto β-aldehydes (Powers *et al*, 2002).

Diazomethylketones have been developed for cysteine proteases such as papain, cathepsins B, C, H, L, and S, calpain, streptopain, and clostripain (Schroder et al, 1993). Peptidyl diazomethylketones have been shown to be effective against lysosomal cysteine proteases due to their ability to penetrate cells of various types. Leupeptin and, a peptide aldehyde and epoxide inhibitors have been used for the inhibition of cysteine proteases. However they lack selectivity and inhibit a wide range of cysteine proteases. The inhibitor Z-Phe-Tyr (t-Bu)-CHN2 inhibits Cathepsin-L selectively and irreversibly with an inhibition constant in the range of 10^{-8} M. This inhibitor is over 10^4 times more effective for Cathepsin L than Cathepsin S (Shaw *et al*, 1993). It inhibits Cathepsin L with a rate 2.5 x 10^4 times greater than that for cathepsin B.

Similarly peptidyl α-keto-β-aldehydes or glyoxal inhibitors are potent inhibitors of chymotrypsin, cathepsin B, cathepsin L, cathepsin S, and the proteasome (Lowther *et al*, 2002). These inhibitors can be synthesized by the oxidative cleavage of the diazogroup of peptidyl diazomethylketones. α-keto-β-aldehydes have two highly electrophilic carbonyl atoms as compared to chloromethylketones which have one electrophilic carbonyl and one alkylating group. They are inhibitors of both serine and cysteine proteases. However these inhibitors are 10 fold more potent towards cysteine

proteases than peptide aldehydes (Walker et al, 2000). The Z-Phe-Tyr (OBut)-COCHO is a slow, tight-binding reversible inhibitor of cathepsin L with a Ki=0.6 nM and is the most potent Cathepsin-L inhibitor reported to date (Lynas *et al*, 2000).

The structure of Cathepsin L in complex with E-64, a natural irreversible inhibitor from Aspergillus *japonicus* has been previously reported (Fujishima *et al*, 1997). Cathepsin L in complex with propeptide-mimetic reverse binding non-covalent inhibitors has been studied extensively (Chowdhury *et al*, 2002). The design of these inhibitors is based on the propeptide region of Procathepsin-L, which is a potent inhibitor of its mature form. So far, there has been no crystal structure reported for any glyoxal inhibitor. As a continuation of our studies on human Cathepsin L, in this chapter we report the crystal structure of Cathepsin L in complex with a diazomethyl ketone inhibitor and a glyoxal inhibitor. This study has the potential to develop potent and selective inhibitors of cathepsin L towards therapeutic intervention.

3.2 MATERIALS AND METHODS

3.2.1 Crystallization and data collection

Z-Phe-Tyr (OBut)-COCHO complex: The human Cathepsin L was expressed in the methylotrophic yeast Pichia *pastoris* and purified as described previously (Coulombe *et al*, 1996). Z-FY (*t*-Bu)-DMK and Z-Phe-Tyr(OBut)-COCHO were purchased from Calbiochem. Cathepsin L was kept in a buffer consists of 20 mM Sodium acetate pH 5.1, 100 mM NaCl, and 1mM EDTA. The inhibitor was dissolved in 25% DMSO. Due to the limited solubility of inhibitor, initial concentration of protein and inhibitor were kept at 0.01M. The complex of mature Cathepsin L with inhibitor Z-Phe-Tyr (OBut)-COCHO

was prepared by incubating the protein with inhibitor (1:4 molar ratio) in the presence of 2 mM DTT at room temperature for 3 hours. After incubation the complex was subsequently concentrated up to 15 mg/ml. Initial crystallization condition was obtained from the Hampton Research Screens (Cudney *et al*, 1994). Diffraction quality crystals of the complex were grown after 2 days by the hanging drop vapor diffusion method at room temperature with a reservoir solution consists of 15 % w/v PEG 8000 and 200 mM Ammonium Sulfate. A hexagonal shaped crystal was obtained with a space group $P6_5$, a=85.55, b=85.55, c= 50.20 Å. These crystals diffracted to a maximum resolution of 2.2 Å. A complete data set was collected using the frozen complex crystal at the Beamline X29A, National Synchrotron Light Source, Brookhaven National Laboratories, USA (Shi *et al*, 2006).

Z-FY (t-Bu)-DMK: Cathepsin L complex: The Cathepsin L : Z-FY(t-Bu)-DMK inhibitor complex was prepared following the same procedure as above and concentrated up to 15 mg/ml. The initial crystallization condition was derived from Hampton Research Screens. The diffraction quality crystals were appeared after 2 days by the hanging drop vapor diffusion method at room temperature with a reservoir solution consists of 30%(w/v) polyethylene glycol 8000 and 200mM Ammonium sulfate. These crystals were belongs to hexagonal form and diffracted up to 1.7Å with space group $P6_5$ and unit cell parameters a=85.55, b=85.55, c= 50.20 Å. X-ray diffraction data were recorded at 100 K at X29A beamline at the Brookhaven National Laboratories, Upton, New York, USA.

3.2.2 Structure Solution and Refinement

The structure of Cathepsin L inhibitor complexes were solved by molecular replacement method using Molrep (Vagin & Teplyakov, 1997). Mature cathepsin L coordinates from Procathepsin L (PDB code 1CS8) was taken as a search model. The initial R-factor was 41.1% and 44.5% for inhibitor 1 and inhibitor 2 complexes respectively. Subsequent refinement was carried out with Refmac (Murshudov *et al*, 1997), which reduced the R-factor to 22% and 24% respectively. At this stage the calculated difference electron density map was examined and the inhibitors were modeled using the O program (Jones *et al*, 1991). Further refitting and alternated with refinements were carried out using O and CNS (Brunger et al, 1998) programs with appropriate entries in their respective dictionaries. Overall geometry of final model of both inhibitor complexes were analyzed by PROCHECK (Laskowski *et al*, 1993). Crystallographic statistics and refinement details are provided in Table 3.1.

3.3 Results and Discussion

The structure of human Cathepsin L in complexed with (1) Z-Phe-Tyr(OBut)-COCHO (hereafter referred as inhibitor **1**) and (2) Z-FY(t-Bu)-DMK (hereafter referred as inhibitor **2**) were determined by the molecular replacement method from synchrotron data sets. Both complexes were refined to final R-factors of 19.9% and 20.4% (R_{free}=23.1% and 24.9%) at 1.7 and 2.2Å resolutions respectively. The refined models have good stereochemical parameters (Table 3.1). There is one complex molecule per asymmetric unit of both crystals. The mature Cathepsin L of both inhibitor complexes is very similar and they superimpose with an rmsd of less than 1Å for all Cα atoms. Further the mature cathepsin molecule is identical to the previous complexes as well as the

mature portion of the procathepsin L crystal structure (pdb code 1cs8). Thus we will not discuss the details of the mature Cathepsin L structure. In both complex structures the inhibitors are bound to the active site Cys25 by a covalent bond. In addition the inhibitor molecules are tightly held in the substrate binding pockets through several hydrogen bonding and hydrophobic interactions.

3.3.1 Z-Phe-Tyr (OBut)-COCHO : Cathepsin-L complex: This structure reveals that the β-aldehyde forms a tetrahedral thiohemiacetal(Fig 3.1a) with the α-ketone oxygen atom facing towards the Cathepsin L oxyanion hole. The inhibitor sidechains fit well in the S subsites as opposed to S' subsites of Cathepsin L. The aldehyde moiety of Z-Phe-Tyr (OBut)-COCHO is covalently attached to the Cys25 thiol group forming a thiohemiacetal. The Tyr (O-But) group occupies the S1 pocket of Cathepsin-L, which is relatively wide and unrestricted. The phenyl and carboxybenzyl group of the inhibitor occupy hydrophobic pockets on opposite faces of the substrate binding cleft in the S2 and S3 subsites respectively (Figs 3.1b, 3.1c).
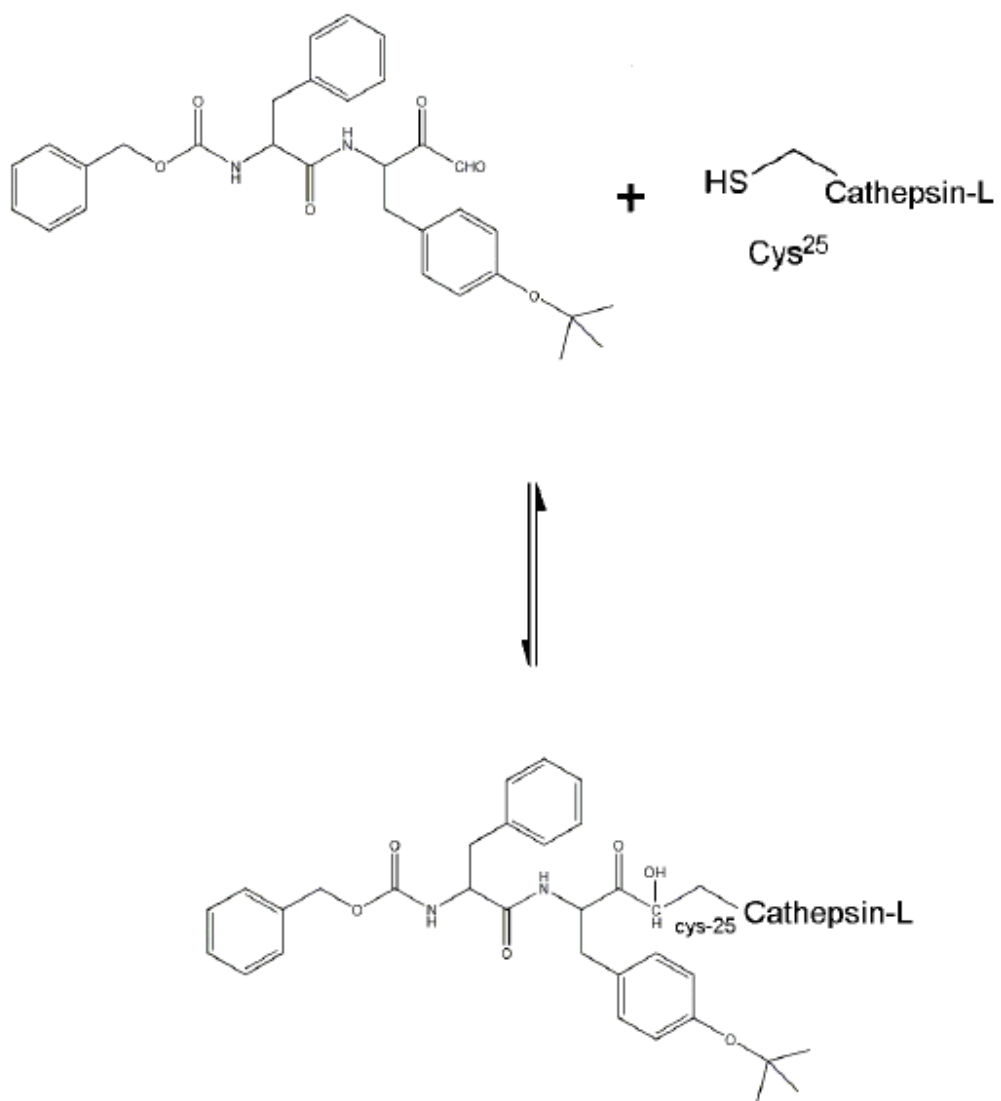
Figure 3.1a Chemical reaction between Z-FY(t-Bu)-COCHO(inhibitor **1**) and Cathepsin L.

Table 3.1. Crystallographic data and refinement statistics

	Z-FY(*t*-Bu)-DMK	Z-Phe-Tyr(OBut)-COCHO
Space group	P6$_5$	P6$_5$
Cell parameters (Å, º)	a = 85.56, b = 85.56, c = 50.20 α= 90 β=90 γ=120	a = 85.56, b = 85.56, c = 50.20 α= 90 β=90 γ=120
Resolution range (Å)	50-2.2	50-1.8
Wavelength (Å)	1.000	1.000
Observed *hkl*	208669	164173
Unique *hkl*	22014	10156
Completeness (%)	95 (55.4)	95.1 (75.7)
Overall *I/σI*	9.9	5.5
[a]R$_{sym}$	0.057(0.55)	0.117(0.52)
Refinement and quality of the model		
*Resolution range	50-1.70 (1.76-1.7)	50-2.2 (2.28-2.2)
[b]Rwork (%) no. reflections	19.9 (16998)	20.4 (7382)
[c]Rfree (%) no. reflections	23.1 (1434)	24.9 (678)
Root mean square deviation		
Bond length (Å)	0.107	0.009
Bond angle (º)	1.54	1.34
Ramachandran plot (%)		
Favored region	85.3	83.2
Allowed regions	14.1	16.3
Generously allowed region	0.5	0.5
Disallowed regions	0	0
[d]Average B-factors (Å^2)		
Main chain atoms	23.37	33.82
Side chain atoms	26.29	36.11
Overall protein atoms (no. atoms)	24.774 (1706)	34.9 (1706)
Waters (no. atoms)	52.1 (190)	54.9 (136)
Ligand (no. atoms)	23.90 (38)	39 (39.2)

[a] R_{sym} = |Ii -<I>|/ |Ii| where Ii is the intensity of the ith measurement, and <I> is the mean intensity for that reflection. [b] R_{work} = | Fobs - Fcalc|/ |Fobs| where Fcalc and Fobs are the calculated and observed structure factor amplitudes, respectively. [c] R_{free} = as for R_{work}, but for 8.5% of the total reflections chosen at random and omitted from refinement. [d] Individual B-factor refinement were carried out.

* Reflections greater than I>σI where used in the refinement

There are six hydrogen bonding contacts between inhibitor **1** and the substrate binding cleft of Cathepsin L. In the S1 pocket, the backbone amide of Tyr (OBut) group makes a hydrogen bond with the carbonyl oxygen of Asp162 (3.12Å). The hydroxyl group of the β-aldehyde oxygen donates a proton to the $N^{\delta 1}$ imidazole ring of His163 which is a part of the catalytic triad of Cathepsin-L. There are two contacts in the S1 subsite made by the α-keto carbonyl oxygen – two with the backbone amides of Ser24 and Cys25 (2.73 and 2.54Å respectively). The Carboxybenzyl group fits well into the S3 pocket. This group makes two hydrogen bonding contacts with the S3 subsite. The backbone carbonyl oxygen of the inhibitor involves in a hydrogen bonding contact with the amide of Gly68 (2.87Å). Moreover the N-terminal amide of the inhibitor forms a hydrogen bond with the carbonyl oxygen of Gly68 (3.0Å). The interactions made by the inhibitor are shown in Figures 3.2a and 3.2b.
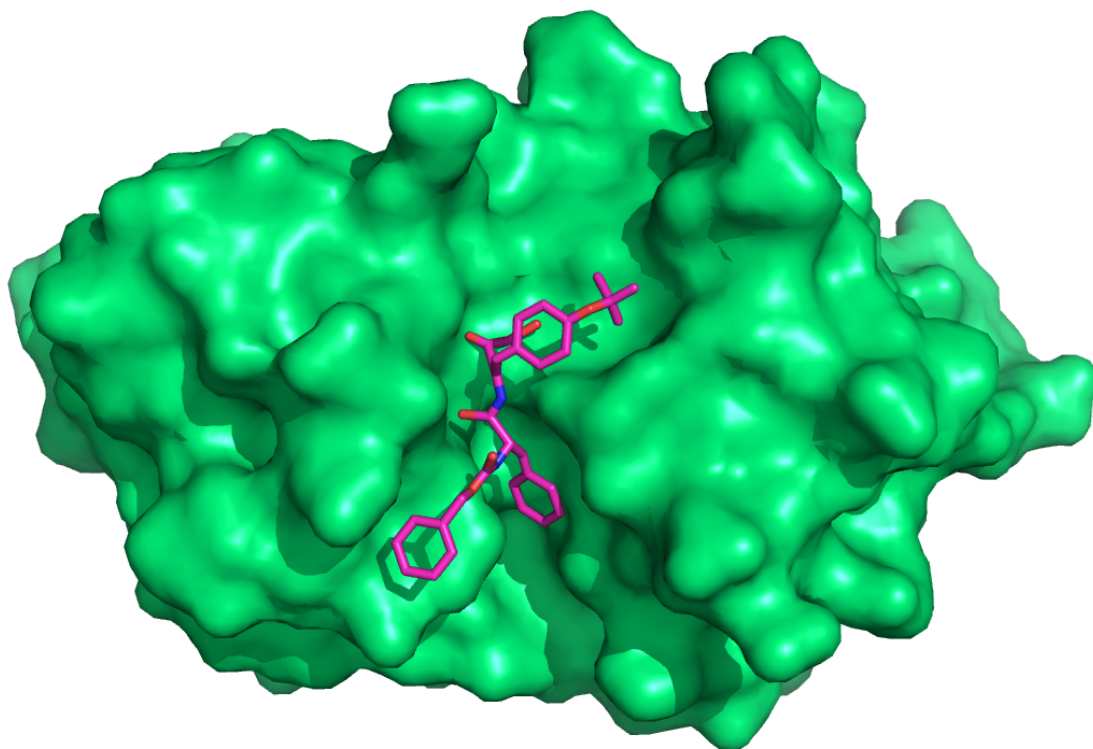
Figure 3.1b. Crystal structure of the molecule showing inhibitor **1** in the active site of Cathepsin L displayed in surface representation. This figure and the following figures of this chapter were prepared using PyMol (DeLano Scientific, Palo Alto, CA)
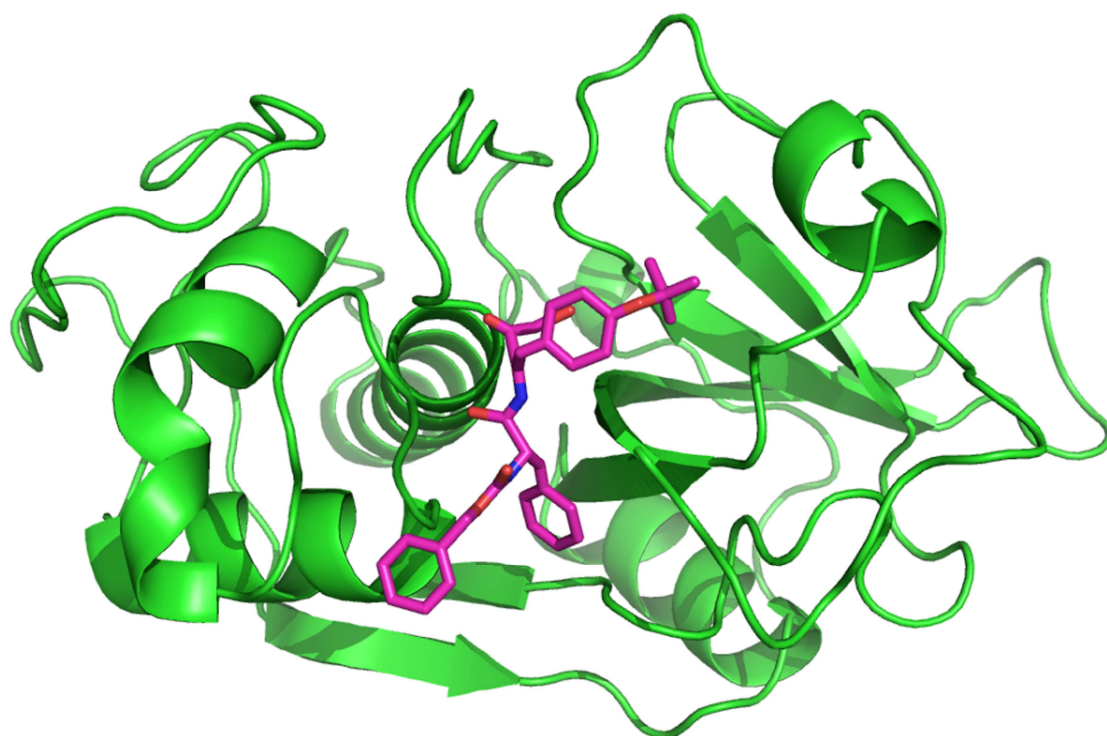
Figure 3.1c. Crystal structure of the molecule showing inhibitor **1** in the active site of Cathepsin L displayed in cartoon representation.
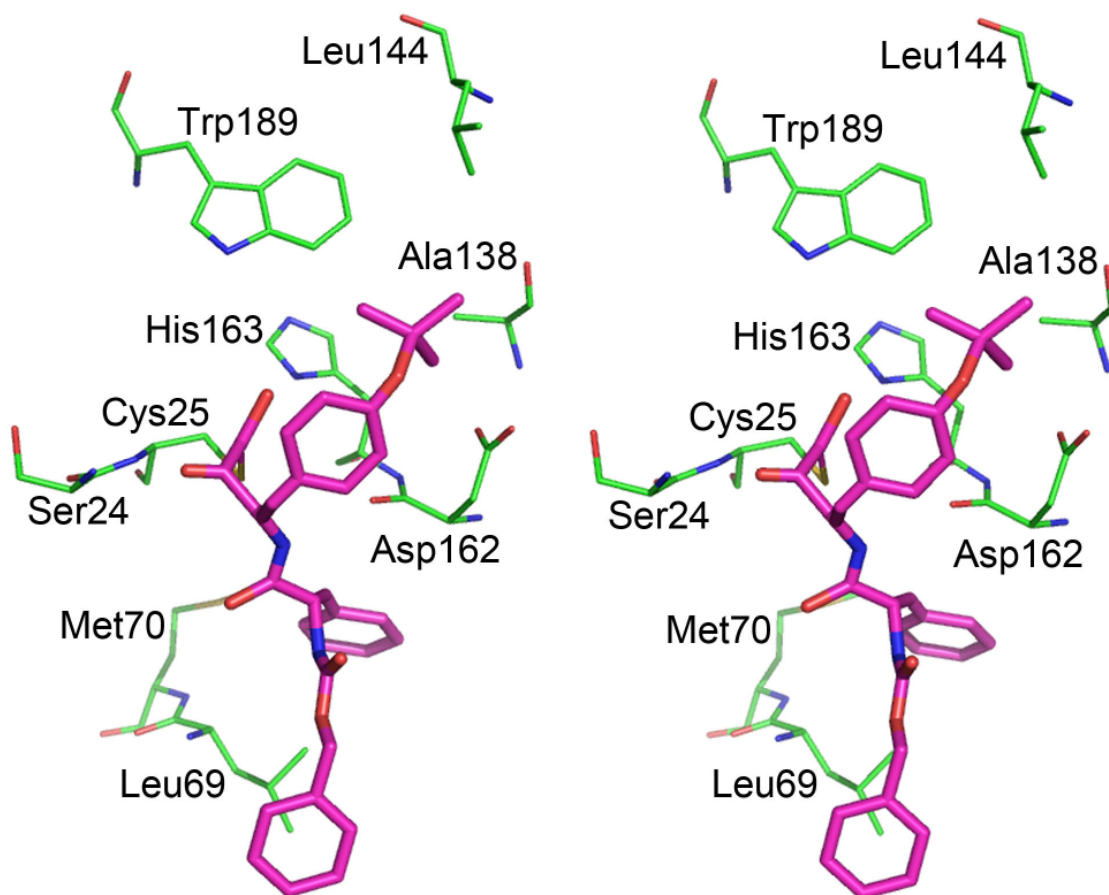
Figure 3.2a. Interactions made by inhibitor 1 in the catalytic cleft of Cathepsin L. Inhibitor 1 is shown in magenta and Cathepsin L residues are shown in green. This figure was prepared using PyMol.
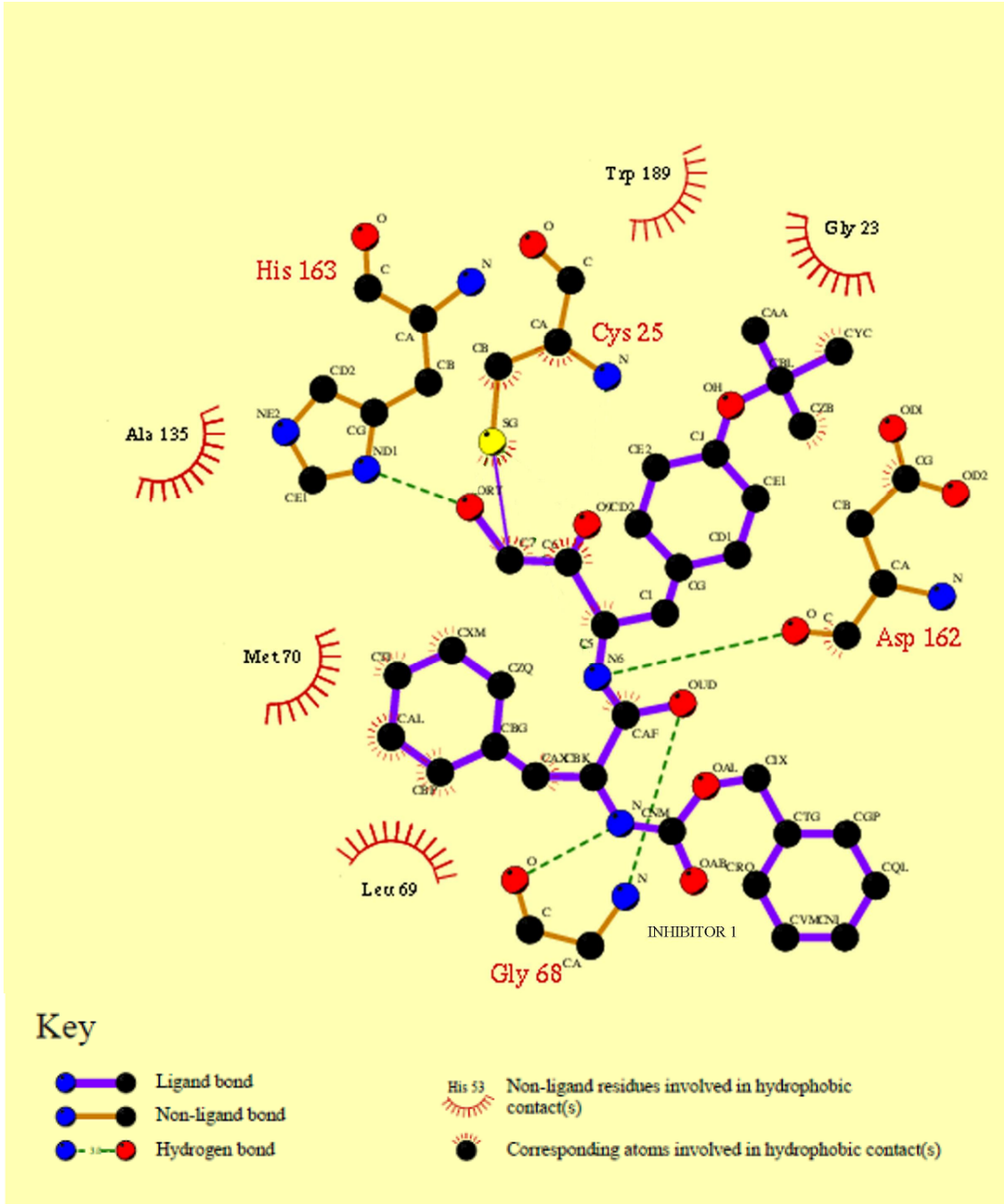
Figure 3.2b. Schematic view of cathepsin L interactions with inhibitor **1**. The Figure was prepared using LIGPLOT (Wallace *et al*, 1995)

It was previously proposed that inhibitor **1** (Z-Phe-Tyr (OBut)-COCHO) forms a transition state thiohemiketal tetrahedral adduct similar to the transition state tetrahedral adduct formed during substrate catalysis in Cathepsin L (Lynas *et al*, 2000). It has been shown by NMR experiments (Lowther *et al*, 2002) that the high potency of the glyoxal inhibitor Z-Phe-Ala-glyoxal against papain is due to the high reactivity of the β-aldehyde group and not due to the formation of a thiohemiketal tetrahedral adduct. The complex structure presented here is the first crystallographic evidence which proves this concept. Figure 3.3 shows the final electron density map for the substrate binding site region of this inhibitor complex which clearly shows that the inhibitor Z-Phe-Tyr (OBut)-COCHO forms a thiohemiacetal with active site Cys25 of Cathepsin L. Moreover the ketone carbonyl group of the inhibitor molecule is not in a position to engage a nucleophilic attack with the thiol group of Cys25.
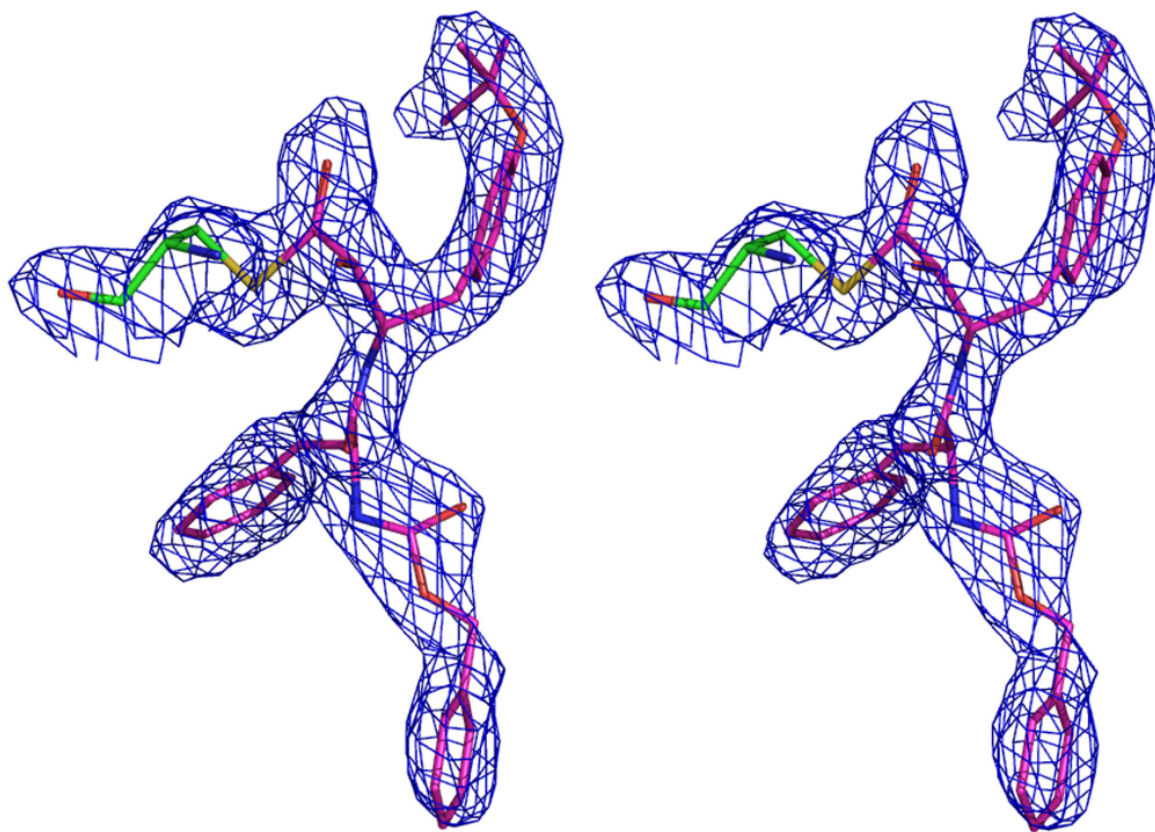
Figure 3.3. Final 2Fo-Fc electron density map for the inhibitor 1 and its covalent attachment to Cys25 of Cathepsin L. The figure was prepared using Pymol.

In the S1' subsite of the substrate binding cleft, the tert-Butyl group involves in hydrophobic interactions with the sidechains from Leu144, Trp189, Ala138 and Gly139. Notably, Trp189 forms a part of the aromatic cluster of Cathepsin L along with Trp193, and Phe143. Similarly in the S2 subsite, the phenyl sidechain has non-polar interactions with the sidechains of Leu69 and Met70. The tert-butyl group is a bulky substituent usually used for kinetic stabilization and can cause a reaction rate acceleration by a factor of 240 compared to hydrogen as a substituent (Cauwberghs *et al*, 1988). In addition the P3 sidechain Carboxybenzyl has hydrophobic interactions with the aliphatic part of

113

Glu63 sidechain. The selected hydrogen bonds and hydrophobic interactions are listed in Table 3.2a and 3.2b.

Table 3.2a: Hydrogen bonds(Å) formed by Inhibitor 1 in the catalytic cleft of Cathepsin L

Inhibitor 1 atom	Cathepsin L atom	Distance(Å)
Phe O15(O)	Gly 68 N	2.87
Phe N33 (N)	Gly 68 O	3.00
Tyr N18(N)	Asp 162 O	3.12
Tyr O22(O)	Ser 24 N	2.73
Tyr O22 (O)	Cys 25 N	2.54
Tyr O23(O)	His 163 Nδ1	2.71

Table 3.2b: Hydrophobic contacts between Inhibitor 1 and Cathepsin L binding pockets

Inhibitor 1 atom	Cathepsin L atom	Distance(Å)
t-Butyl C26(C)	Leu 144 Cδ1	3.86
t-Butyl C26(C)	Gly 139 Cα	3.84
t-Butyl C6(C)	Ala 138 Cβ	3.73
t-Butyl C7(C)	Trp 189 Cζ2	3.65
Phe C10(C)	Leu 69 Cδ2	3.84
Phe C11(C)	Leu 69 Cδ2	3.68
Phe C11(C)	Met 70 Cε	3.76
Phe C11(C)	Ala 135 β	3.85

3.3.2 Z-Phe-Tyr (t-Bu)-DMK : **Cathepsin L complex:** The inhibitor molecule Z-Phe-Tyr (*t*-Bu)-DMK is covalently bound to the Cathepsin L active site Cys25 by a thioester bond i.e. the methylene of the diazomethane group is covalently attached to the S^γ atom of Cys25(Fig 3.4a). As seen in the glyoxal inhibitor (inhibitor **1**) complex structure, the Tyr (*t*-Bu) group, phenyl group and the Carboxybenzyl group of the inhibitor occupies the S1, S2 and S3 subsites (Fig 3.4b and 3.4c). A total of five hydrogen bonding contacts between the inhibitor molecule and the substrate binding cleft of Cathepsin L are

observed. The carbonyl oxygen of the inhibitor is pointing towards the oxyanion hole of

Cathepsin L comprising
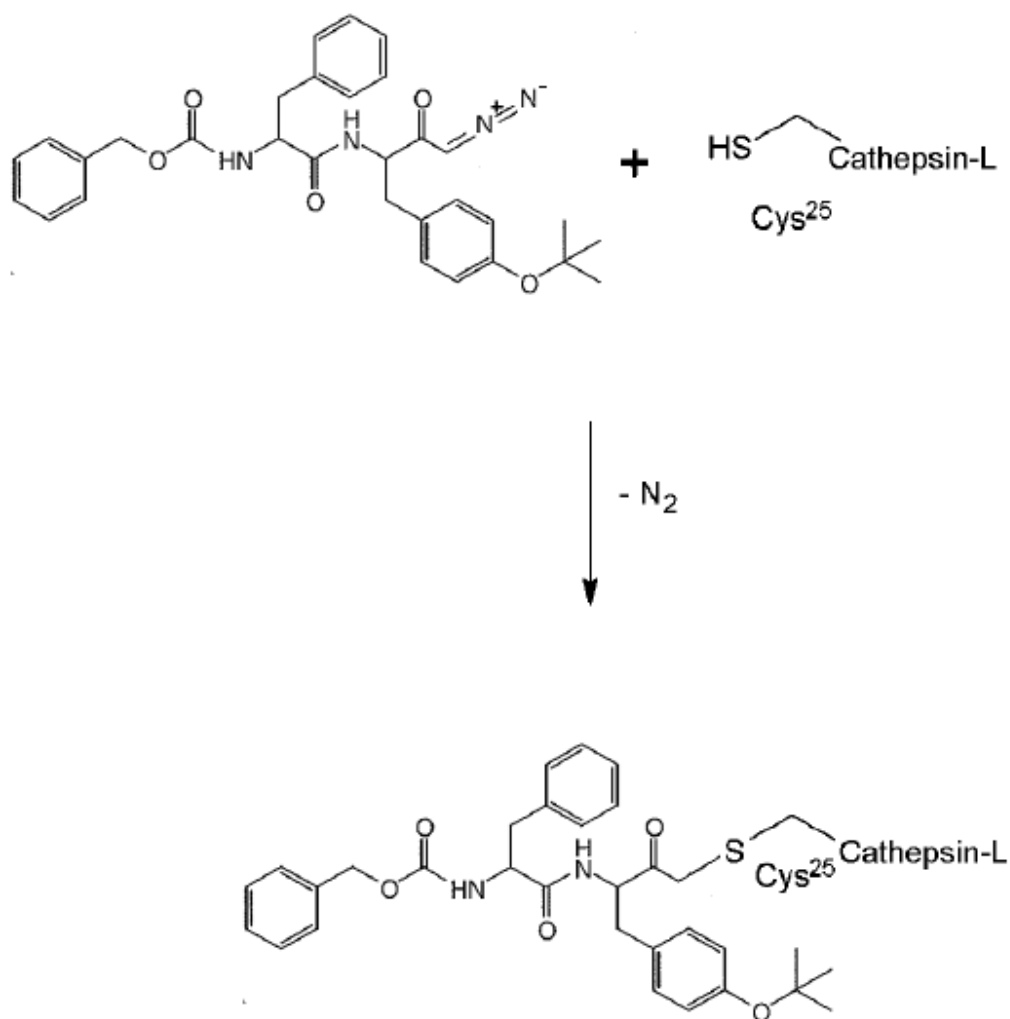
Figure 3.4a Chemical reaction between Z-FY(t-Bu)-diazomethylketone(inhibitor **2**) and

Cathepsin L.

of Gln19 and Cys25. Further this forms hydrogen bonding contacts with the sidechain

amide of Gln19 and the backbone amide of Cys25. In the S1 subsite, the backbone amide

115

of the Tyr (*t*-Bu) group makes a hydrogen bond with the carbonyl oxygen of Asp162. It is noteworthy that in the S3 pocket, there are two hydrogen bonding contacts between the backbones of the inhibitor Phenyl group and the Gly68 of Cathepsin-L, which introduces an anti-parallel β-sheet. Previously similar antiparallel β-sheet formation has been reported for papain diazomethylketone complex (Janowski *et al*), papain chloromethylketone complex (Drenth et al, 1976), and papain leupeptin complex (Schroder et al, 1993). The selected hydrogen bonding contacts are listed in Table 3.3a.

The *t*-Butyl group of the tyrosine sidechain of the inhibitor forms a part of the hydrophobic cluster which consists of the sidechains from Trp189, Leu144 and Ala138 in the S1' subsite. Similarly in the S2 subsite, the Phenyl group sidechain of the inhibitor molecule has non-polar interactions with the sidechain of Leu69 and Ala135. Further in the S3 subsite, the Carboxybenzyl group has hydrophobic interactions with the aliphatic part of the sidechain Glu63. The selected hydrophobic interactions are listed in the Table 3.3b. The interactions made by inhibitor 2 are shown in Figures 3.5a and 3.5b and the final 2Fo-Fc electron density map are shown in Figure 3.6.
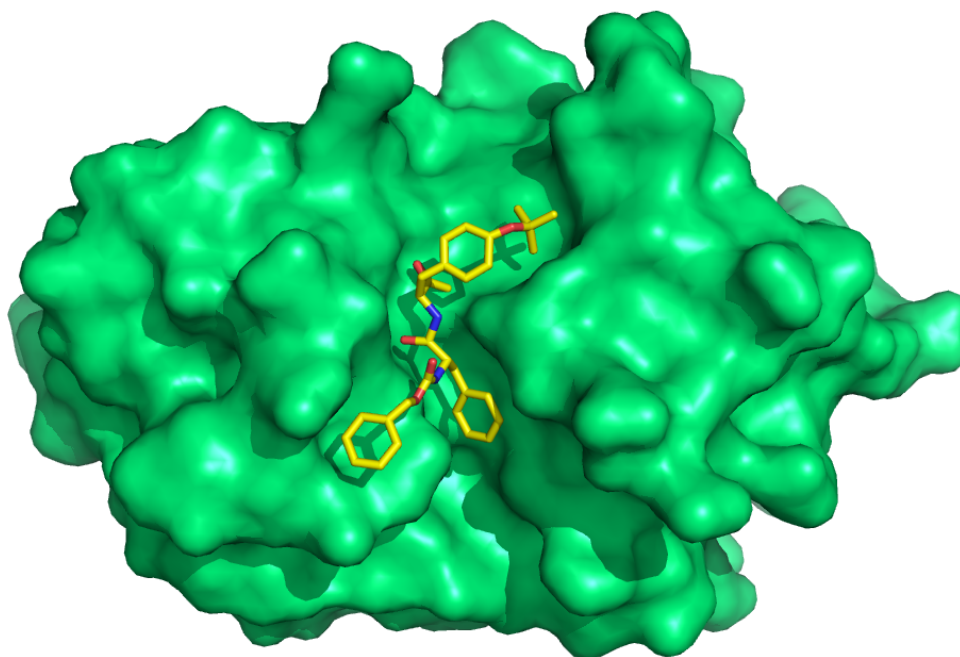
Figure 3.4b. Crystal structure of the molecule showing inhibitor **2** in the active site of Cathepsin L displayed in surface representation. The figure was prepared using Pymol.
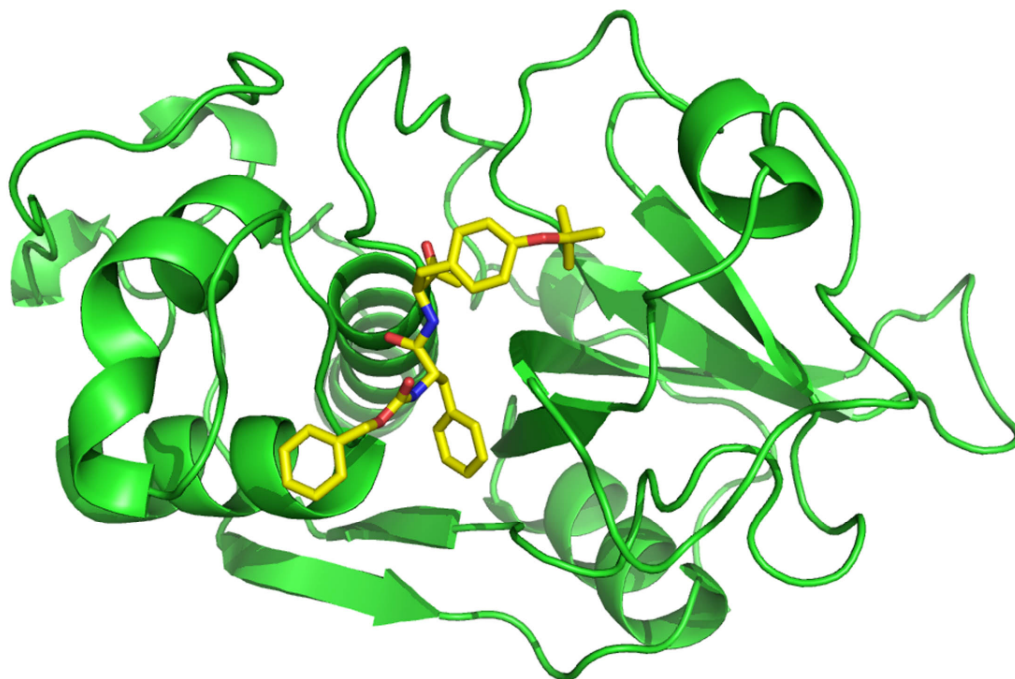
**Figure 3.4c.** Crystal structure of the molecule showing inhibitor **2** in the active site of Cathepsin L displayed in cartoon representation. The figure was prepared using PyMol.
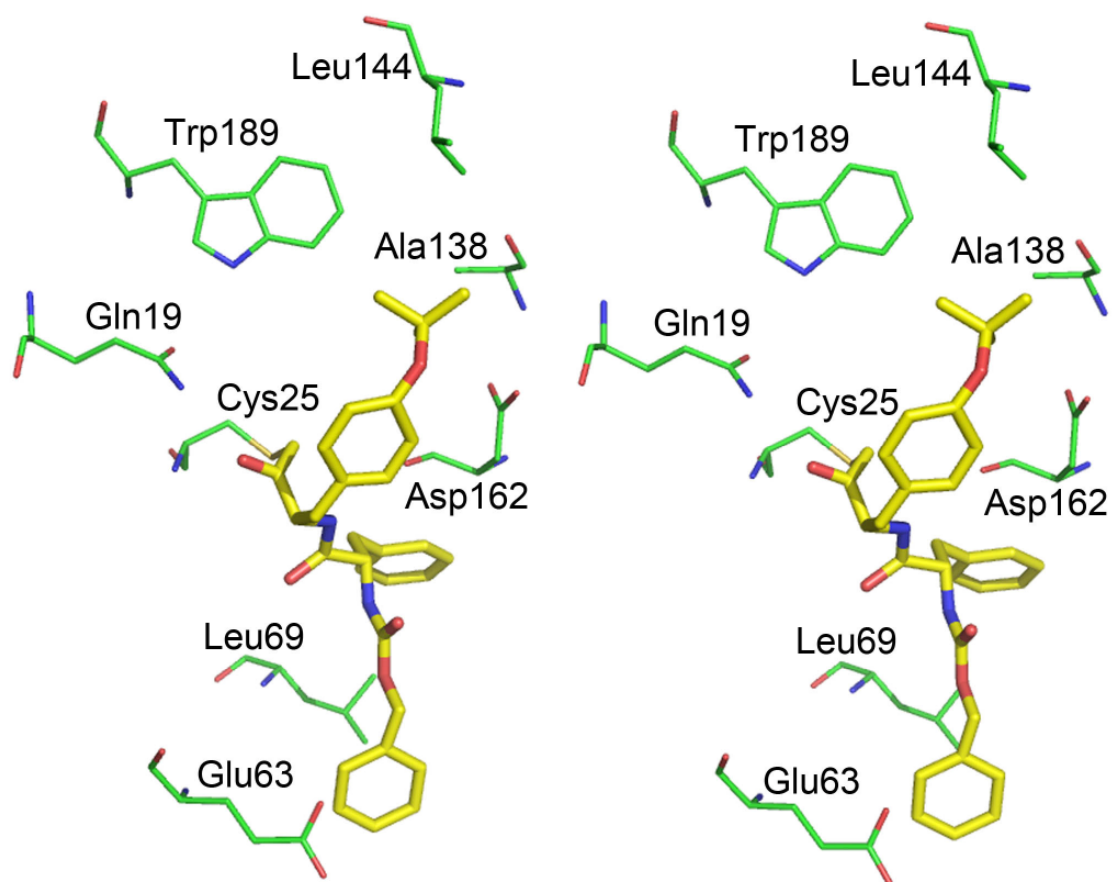
Figure 3.5a. Interactions made by inhibitor 2 in the catalytic cleft of Cathepsin L. Inhibitor **2** is shown in yellow and Cathepsin L residues are shown in green. The figure was prepared using Pymol.
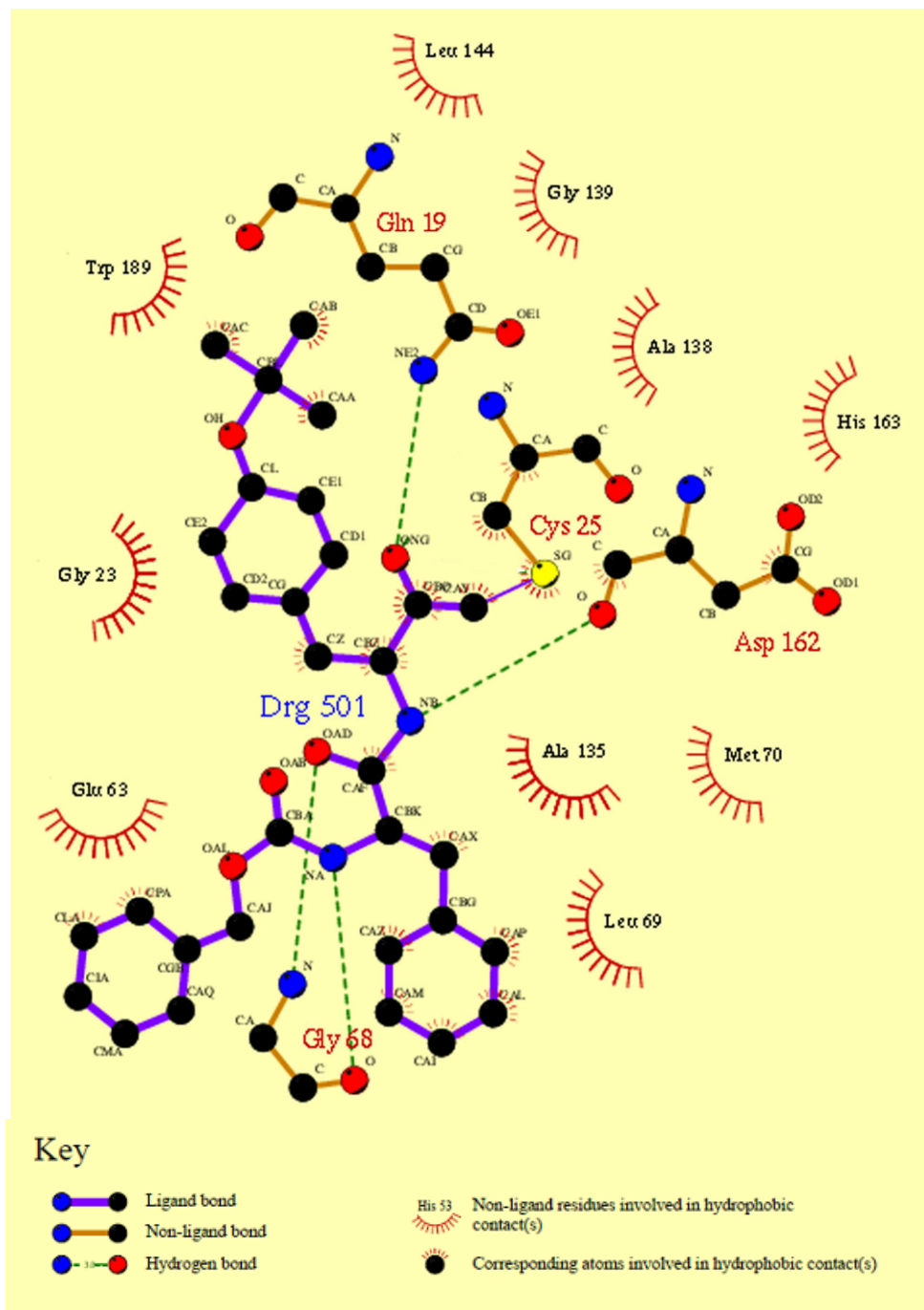
Figure 3.5b. Schematic view of cathepsin L interactions with inhibitor **2**. The Figure

was prepared using LIGPLOT (Wallace *et al*, 1995)
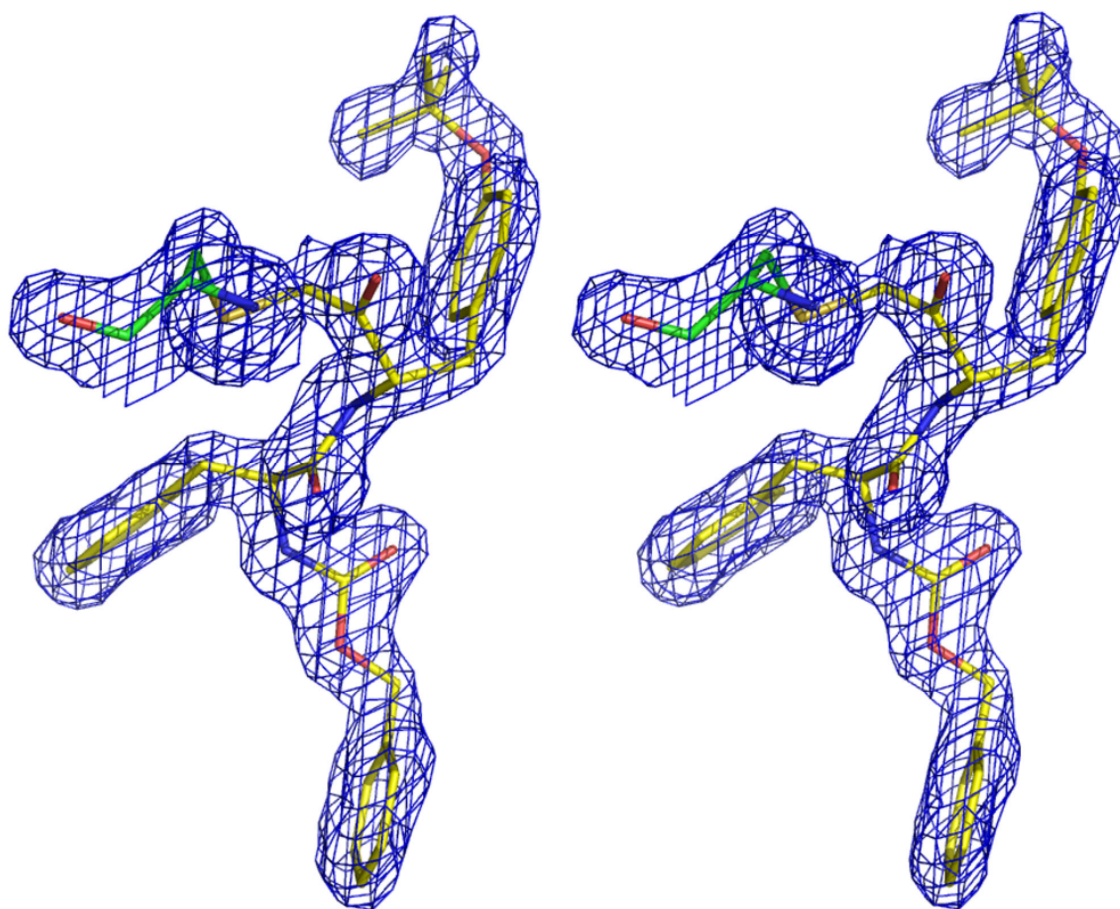
Figure 3.6. Final 2Fo-Fc electron density map for the inhibitor **2** and its covalent attachment to Cys25 of Cathepsin L. The figure was prepared using Pymol.

Table 3.3a: Hydrogen bonds (Å) formed by Inhibitor **2** in the catalytic cleft of Cathepsin L.

Inhibitor 2 atom	Cathepsin L atom	Distance(Å)
Tyr O3(O)	19 Gln (Nε2)	2.97
Tyr O3(O)	25 Cys (N)	3.08
Phe O28 (O)	68 Gly (N)	2.92
Phe N36 (N)	68 Gly (O)	2.97
Tyr N38 (N)	162 Asp (O)	3.16

Table 3.3b: Hydrophobic contacts between Inhibitor **2** and Cathepsin L binding pockets

Inhibitor **2** atom	Cathepsin L atom	Distance(Å)
t-Butyl C4(C)	Trp 189 Cη2	3.64
t-Butyl C7(C)	Leu 144 Cδ1	3.65
t-Butyl C7(C)	Ala 138 C	3.97
t-Butyl C7(C)	Gly 139 Cα	3.77
Cbz C16(C)	Glu 63 Cδ	3.82
Phe C23(C)	Leu 69 Cδ2	3.62
Phe C24(C)	Ala 135 Cβ	3.70

The superposition of the crystal structures (Figure 3.7) of the two inhibitor complexes shows that the overall conformations of the two inhibitors are similar. There is high congruity between the Carboxybenzyl, Phenyl and Tyr (t-Bu) groups. The only difference is found in the reactive groups which form the covalent bond with the active site Cys25 of the Cathepsin L. The geometry of the covalent link between the inhibitor molecule and the cathepsin L is different in the two complexes. In both complex structures, the tertiary butanol group makes similar hydrophobic interactions with sidechains of the residues in the S' site of Cathepsin-L. The antiparallel β-sheet formed between the Gly68 and the inhibitor Phenyl group is a feature seen in both the inhibitor structures as well as in other Cathepsin-L complexes like chagasin-cathepsin complex and cruzipain complex (Ljunggren *et al*, 2007). This β-sheet formation is possible due to the flexibility of Gly68 in this region of Cathepsin L. Although the hydrophobic interactions between the two inhibitor molecules are similar, the glyoxal inhibitor (inhibitor **1**) makes 2 additional hydrogen bonds compared to the diazomethylketone inhibitor (inhibitor **2**). It was previously speculated that P' hydrogen-bonding interactions between active site residues and the hydrated β-aldehyde hydroxyl group of the α-keto-β-

aldehyde group are crucial for the effectiveness of this inhibitor (Lynas *et al*, 2000). From our crystal structure we found that there are no P' hydrogen bonding interactions between the inhibitor molecule and the hydrated β-aldehyde hydroxyl group.
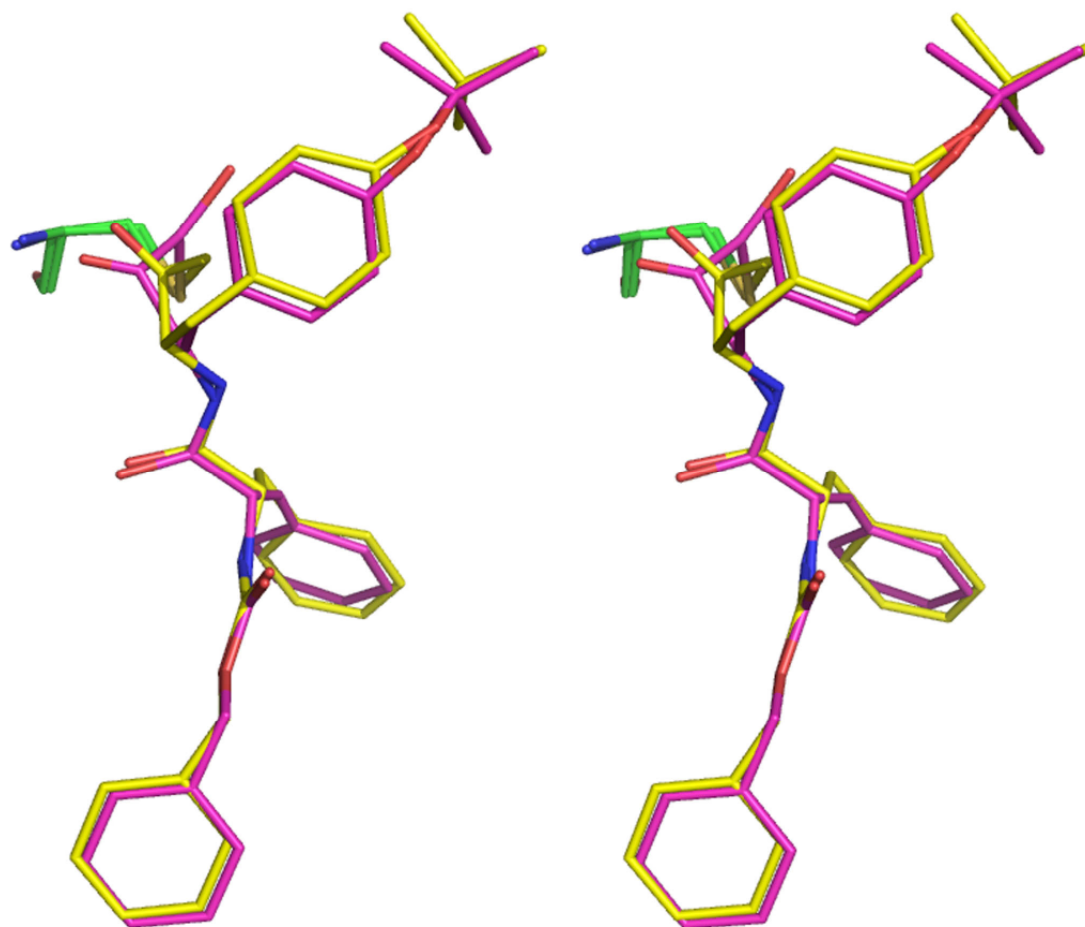
Figure 3.7. Overlay of crystal structures of Cathepsin L complexes with inhibitors 1 and 2. Inhibitor **1** is shown in magenta and inhibitor **2** is shown in yellow. The figure was prepared using PyMol.

In fact, it forms a hydrogen bond with the catalytic triad residue His163 in the S1 pocket. Moreover the carbonyl oxygen of the α-keto group is involved in two hydrogen bonds i.e. with Ser24 and Cys25 whereas in case of the diazomethylketone inhibitor, the carbonyl oxygen of the inhibitor makes two hydrogen bonding interactions i.e. with the oxyanion hole residues Gln19 and Cys25. These unique interactions in inhibitor **1** may be crucial in explaining the higher potency of the glyoxal inhibitor.

Gln19 is conserved among all known cysteine proteases and is an important feature of the catalytic mechanism of cysteine proteases. The mutation of Gln19 leads to a 60-600 fold decrease in the activity of papain (Ménard *et al*, 1995). When papain is inhibited by an aldehyde inhibitor, a tetrahedral transition state similar to that found in substrate catalysis is formed (Menard *et al*, 1991). But in this inhibitor complex, the oxyanion hole Gln19 is not utilized as it forms a neutral tetrahedral transition state. In the case of a substrate, the Gln19 is utilized because there is a charge transfer between Cys25 and carbonyl oxygen of the substrate. Hence the catalytic machinery used for charge stabilization will not be used in all compounds reacting with the enzyme. In our present study of the α-keto β-aldehyde inhibitor, Gln19 is not utilized because the tetrahedral hemithioacetal is a neutral group in which there is no charge transfer. But in the case of diazomethyl ketone inhibitor (inhibitor **2**), there is a hydrogen bonding contact with the Gln19 which implies that the oxyanion hole is utilized as the formation of the thioester with Cys25 involves the removal of the electrophilic diazogroup. It is evident that different types of inhibitors utilize the catalytic machinery of Cathepsin L differently as has been observed in the case of papain (Dufour *et al*, 1995). Aldehydes are potent inhibitors of cysteine proteases. The potency of these inhibitors is suggested to be due to

124

the formation of covalent adducts similar to the tetrahedral transition state of substrate catalysis. In the case of inhibitor **2**, the diazomethylketone inhibitor complex, the carbon atom of the thioester bond has sp2 hybridization and adopts a trigonal geometry (Figure 3.8). This is a comparatively slower reaction compared with inhibitor **1**. The structure of the complex formed with inhibitor **2** resembles the acyl enzyme intermediate than the tetrahedral hemithioacetals formed between cysteine proteases and their transition state intermediates. In inhibitor **1**, similar to a transition-state intermediate, Cathepsin-L forms a tetrahedral hemithioacetal (sp^3 hybridization) with the β-aldehyde group of the glyoxal inhibitor.
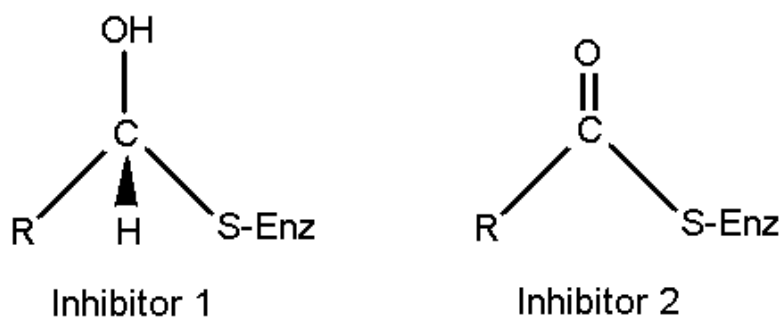
Figure 3.8. Comparison of the geometries of hemithioacetal formed in inhibitor **1** with the thioester formed with inhibitor **2**.

Several inhibitor complexes of papain with peptide aldehyde inhibitors have been previously studied by crystallography and NMR. In these complexes, the hemiacetal complexed with peptide aldehyde inhibitors in complex with papain can exist in two conformations – (1) the oxygen atom of the hemiacetal is in the oxyanion hole or (2) is in hydrogen bonding distance of the catalytic site Histidine residue. This has also been

observed in the case of chymostatin-Streptomyces griseus protease A complex (Delbaere & and Brayer, 1985) and leupeptin-trypsin complex (Ortiz *et al*, 1991). In the present study, we find that the hydrated β-aldehyde is in hydrogen bonding distance of His163 instead of the oxyanion hole. However the presence of an alternative conformation is yet to be verified.

3.4 Conclusion: We have solved the crystal structures of two covalent inhibitors in complex with mature Cathepsin L up to resolutions of 2.2 and 1.7Å respectively. These two classes of inhibitors have identical peptide sequences but different reactive groups. These were chosen to explore the binding mode of these inhibitors. Both the inhibitors showed identical binding conformations with respect to S2 and S3 pockets whereas notable differences observed in the S1 subsite. The glyoxal inhibitor formed a tetrahedral hemithioketal whereas the diazaomethyl ketone inhibitor formed a thioester bond with trigonal geometry. The carbonyl oxygen of the β aldehyde group forms a hydrogen bond with the catalytic His163 while the diazomethylketone is found to have its carbonyl oxygen facing the oxyanion hole, forming a hydrogen bond with Gln19. The large differences in the binding mode of the two inhibitors could be attributed to the reactive groups of these inhibitors. The glyoxal inhibitor is reversible due to its hydrolytically labile binding whereas the diazomethylketone is an irreversible inhibitor which has a comparitively slower reaction with the active site thiol. Cathepsin L displays a typical papain-like architecture and closely resembles Cathepsins S and K.

Chapter IV

Structural basis for a non-classical Kazal-type serine protease inhibitor in regulating host-pathogen interaction via a dual-inhibition mechanism

4.1 Introduction

Serine proteases play an important immunomodulatory role in host-pathogen interactions. Invertebrates lack an adaptive immune system which recognizes and remembers specific pathogens (Jiravanichpaisal *et al,* 2006). As an evolutionarily ancient defense strategy, the innate immune system responds instantaneously to invading pathogens in a non-specific manner. The innate immune system in the horseshoe crab comprises of serine protease cascades like the blood coagulation system, melanization and complement system (Theopold *et al*, 2004). Horseshoe crab hemocytes contain granules filled with several serine protease zymogens. Upon mechanical injury or pathogen invasion, the granules are released into the extracellular milieu by exocytosis, and clotting enzymes in their precursor forms are activated by a serine protease cascade triggered by bacterial endotoxin (Armstrong, 2001; Ding *et al*, 1993). Several serine protease zymogens including procolotting enzymes, Factor B and Factor C are associated with the hemolymph coagulation system. The subsequent formation of the coagulation plug prevents further entry of pathogens (Cerenius and Söderhäll, 2004; Ding & Ho, 2001).

A large number of Kazal-type serine protease inhibitors are expressed in hemocytes with different specificities against different proteases that target different microorganisms (Iwanaga & Kawabata, 1998). They are potent inhibitors of serine proteases like trypsin, α-chymotrypsin, elastase, subtilisin and other bacterial and fungal proteases, which belong to α2-macroglobulins. These low molecular weight inhibitors belonging to the Kunitz family of proteinase inhibitors use a lock-and-key mechanism with their flexible reactive site loops, to mechanistically bind the active-sites of the target

proteases (Kanost, 1999). It is speculated that these inhibitors might participate in the regulation of the hemolymph coagulation cascade (Jiang *et al*, 2009; Ding *et al*, 2004)

During our recent work on subtractive ESTs from horseshoe crab infected with *Pseudomonas aeruginosa*, two isoforms of Kazal-type inhibitor, CrSPI, each of 9.3 kDa were discovered (Jiang *et al*, 2009). The CrSPI-1 and CrSPI-2 (GenBank Accession numbers: DQ090491 and DQ090492, respectively) isoforms share 97% sequence identity. Both isoforms are biochemically active. CrSPI-1 inhibits microbial proteases such as, subtilisin A and protease K. However it is more specific towards subtilisin with $K_i = 1.4nM$.

Serine protease inhibitors in plasma are known to perform dual functions of (i) regulating the activity of endogenous serine proteases, and (ii) inhibiting microbial serine proteases as a form of host immune defense against the invading pathogen (Hiemstra *et al*, 2002). For instance, the human lympho-epithelial Kazal-type inhibitor (LEKTI) plays an important role in the epithelial tissue homeostasis by regulation of trypsin, and dysfunction of LEKTI is implicated in Netherton syndrome (Chavanas *et al*, 2000). Further, the horseshoe crab Limulus serpins LICI-1, LICI-2 and LICI-3 are known to regulate serine proteases in the pathogen-induced coagulation cascade (Muta and Iwanaga, 1996).

The Kazal family is one among 18 families of serine protease inhibitors. Since the 1980's structures of several members of this family have been reported (Bode and Huber, 1992). The Kazal family is mainly divided into two groups- the classical and the nonclassical inhibitors. Positions of the cysteine residues that form the disulphide bonds

differ among classical and non-classical Kazal-type inhibitors. The classical group is composed of pancreatic secretory trypsin inhibitor, and the ovomucoids.

Both domains 1 and 2 of CrSPI-1 share sequence homology with the non-classical group I Kazal inhibitors. The characteristic feature of this group is the formation of the disulfide bond between the first and fifth cysteine residues which are shifted towards the C-terminus in comparison to the respective residues of a classical Kazal inhibitor like OMTKY3(Turkey ovomucoid third domain) or OMSVP3 from the silver pheasant(Fig 4.1). Other representatives of this non-classical group are *Anemonia Elastase* inhibitor, Crayfish inhibitor(Johansson *et al*, 1994), Ciona trypsin inhibitor(Odum et al, 1999) and Galleria trypsin inhibitor from the greater wax moth Galleria *mellonella* (Nirmala et al, 2001).
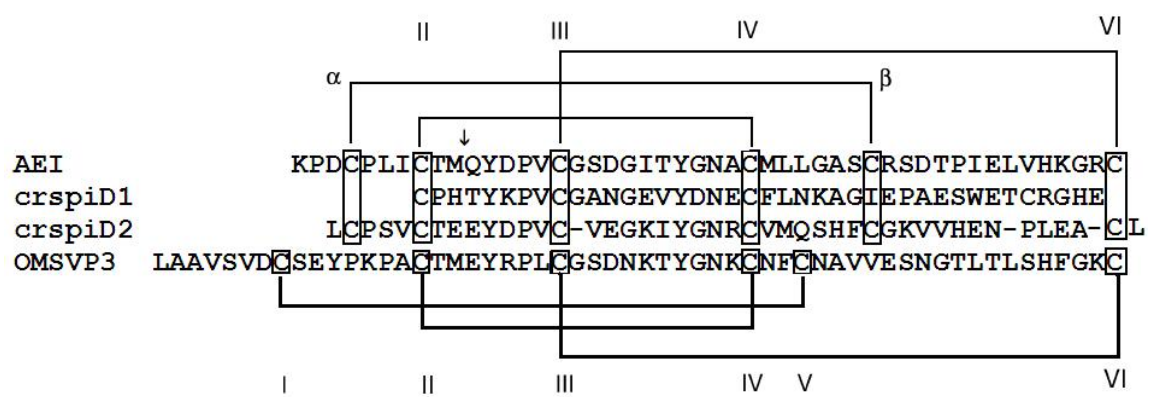
Figure 4.1. Alignment of amino acid sequences of non-classical group I Kazal-type inhibitors AEI, CrSPI-1 domain I, domain II and a selected classical Kazal-type inhibitor OMSVP3. The sequences were aligned using CLUSTAL-W. The reactive site is denoted with an arrow. Disulfide bonds are linked as follows: α-β, II-IV, and III-VI for the non classical group I inhibitors and I-V, II-IV, and III-VI for the classical inhibitors. In nonclassical group II inhibitor family, there is an additional disulphide bridge between α and β half cystines.

Hitherto, no crystal structure has been reported for a protease inhibitor from the blood of arthropods. Furthermore, no crystal structures are available for the non-classical group I Kazal-type inhibitors. The only available three-dimensional structure in this group is the solution structure of *Anemonia elastase* inhibitor (Hemmi *et al*, 2005). In this chapter, we report the crystal structure of a double headed non-classical Kazal-type group I inhibitor CrSPI-1 in complex with its cognate protease, subtilisin, at a ratio of 1:2 and refined upto 2.6Å resolution. One CrSPI-1 molecule binds two subtilisin molecules. The reactive site loops of both domains of CrSPI-1 occupy the substrate binding pockets of subtilisin. Further, based on our structural and biophysical interaction studies, we propose that domain-2 of CrSPI-1 is a more specific and potent inhibitor for subtilisin, whereas domain-1 is likely to interact with Furin, a subtilisin homolog, from the host. Thus CrSPI-1 may act as a regulatory ON/OFF switch in modulating antimicrobial activities while maintaining homeostasis of host proteases. The present studies on CrSPI-1 provide new insight in the area of host-pathogen interactions in the innate immune system of the living fossil, *C. rotundicauda*.

4.2 Experimental

4.2. 1 *Expression, purification, crystallization and structure determination*

The CrSPI-1 gene was cloned and expressed with pET32-EkLIC system (Novagen) using the following primers -Forward:5' ACG GAC GAC GAC AAG ATG TGT CCT CAT ACT TAC AAA 3', Reverse:5' ACG GAG GAG AAG CCC GGT TTA CAA GCA AGC TTC TAG TGG 3' and expressed in *E. coli*. The expressed protein contained a thioredoxin tag, a His-tag and an enterokinase cleavage site. The recombinant CrSPI, henceforth referred to as rCrSPI-1 fusion protein, was overexpressed at 37 °C

from a single colony picked from an agar plate. The culture was induced with 300 mM isopropyl 1-thio-D-galactopyranoside for 4 h to an OD_{600nm} of 0.6. The cells were then harvested and sonicated. The protein was purified using Nickel-NTA affinity beads (Qiagen) with PBS (Phosphate Buffered Saline), pH 7.4, 10 mM β-mercaptoethanol, (BME). The protein was eluted in 300 mM imidazole. The thioredoxin tag was cleaved by 2-h incubation with enterokinase (Sigma). The complex was prepared by mixing rCrSPI-1 with subtilisin (obtained from Sigma) in a molar ratio of 1.2:1 (inhibitor: enzyme) and incubated for 1 hour at 37° C. The complex was then passed through a Superdex-200 gel filtration column, and fractions were pooled and concentrated upto 20 mg/ml. The crystallization screens were carried out using the hanging drop vapor diffusion method with Hampton Research Crystal Screen kits I and II and JB Screens (Jena BioScience, Germany) at room temperature. The initial crystallization conditions were further optimized using a grid screen by varying the concentration of precipitant. Plate-like diffraction quality crystals were obtained from 11% (w/v) Polyethylene Glycol 20000 in 0.1 M 2-(N-morpholino) ethanesulfonic acid (MES) at pH 6.5. The crystallization solution was supplemented with 25% glycerol which acts as a cryo-protectant. A complete data set was collected in the X29-A synchrotron beamline at Brookhaven National Laboratories, USA. The data collection and refinement statistics are provided in Table 4.1.

4.2.2 *Structure Solution and Refinement*

The structure of rCrSPI-1 - subtilisin complex was solved by molecular replacement method with Molrep (Vagin & Teplyakov, 1997) using the subtilisin Carlsberg coordinates (PDB code 1SCA). The initial R-factor was 47% and subsequent

refinement was performed with Refmac (Murshudov *et al*, 1997) and CNS (Brunger et al, 1998). Strict noncrystallographic symmetry (NCS) was used for subtilisin molecules during the refinement in CNS. At the stage the R-factors were close to 40, the calculated electron density map allowed us to build the rCrSPI-1 molecule. Several cycles of refinements in CNS alternated with model building for the inhibitor complex was carried out using the program O (Jones *et al*, 1991) and Coot (Emsley & Cowtan, 2004). The overall geometry of final model was analyzed by PROCHECK (Laskowski *et al*, 1993) (Table 4.1). Figures 4.2a and 4.2b show the final *2Fo-Fc* electron density map of the reactive site loop region (RSL) of both domains.
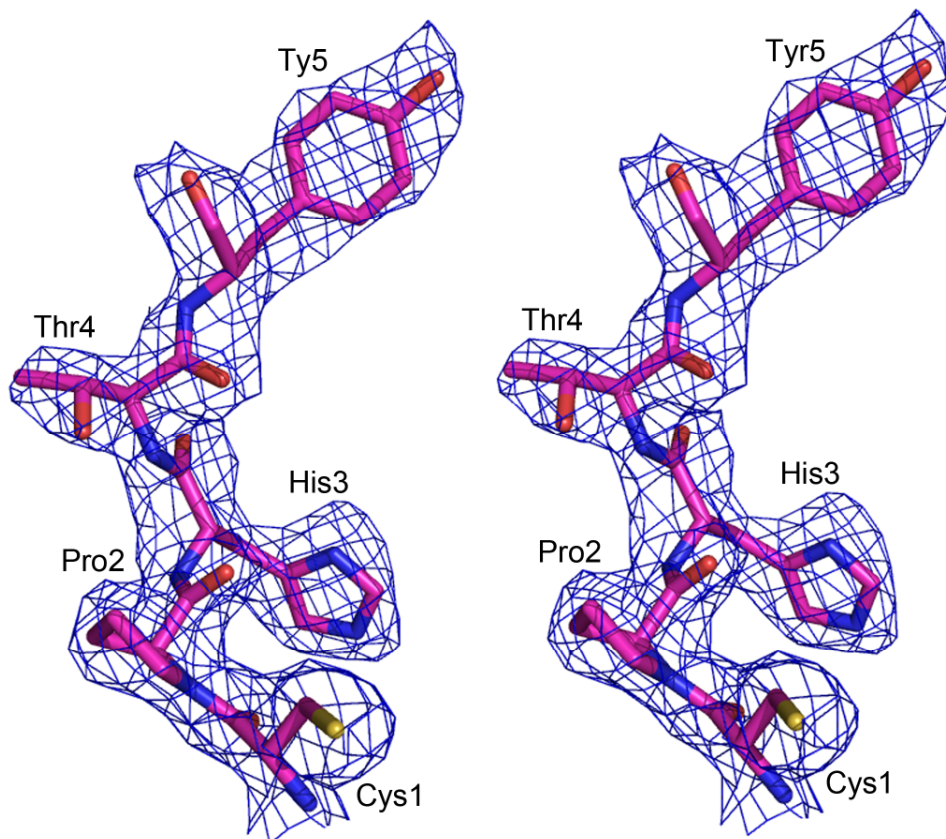
Figure 4.2a. Stereo view of *2Fo-Fc* map for the reactive site loop (RSL) region of domain-1 and of rCrSPI-1 bound to subtilisin contoured at a level of 1σ. The figure was prepared by using PyMol (DeLano Scientific, Palo Alto, CA).
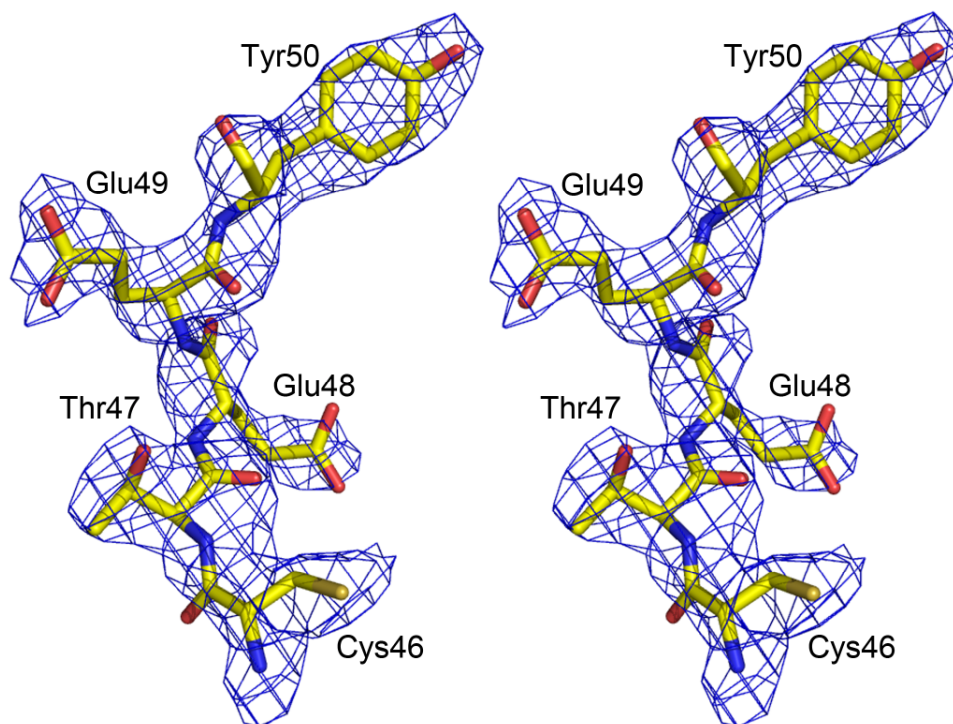
Figure 4.2b. Stereo view of *2Fo-Fc* map for the reactive site loop (RSL) region of domain-2 of rCrSPI-1 bound to subtilisin contoured at a level of 1σ. The figure was prepared by using PyMol (DeLano Scientific, Palo Alto, CA).

4.2.3 *Isothermal Titration Calorimetry (ITC)*

ITC studies were performed using a MicroCal VP-ITC instrument. The rCrSPI-1 at a concentration of 0.2 mM in PBS pH 7.4, 10 mM BME and subtilisin in the sample cell at a concentration of 0.012 mM. All experiments were performed at 37°C. rCrSPI-1 was titrated into subtilisin in 18 injections; each injection consisting of 2 μL of ligand

solution. Samples were degassed prior to use. Due to dilution across the needle during equilibration, the initial peak was discarded. Data analysis was performed using the MicroCal Origin software. A model considering a stoichiometry 1:2 CrSPI-1: Subtilisin was considered.

4.2.4 *Inhibition of Furin by CrSPI-1*

rCrSPI-1 at various concentrations (0.9, 6.3, 8.9, 11.1 and 13.5 nM) was incubated with 0.02 Unit of Furin (obtained from Sigma) for 60 min at 37°C in 100 µl of reaction buffer (0.1 M Hepes pH 7.5, 1mM BME and 0.1% Triton X-100). After incubation, (100 µM) Pyroglutamic acid-Arg-Thr-Lys-Arg-AMC, (Calbiochem), a Furin-specific fluorogenic substrate was added to the incubated samples following the protocol by Dufour *et al.,* 1998. Fluorescence readings were measured at after 8, 3, 5, 9 and 24hrs by taking 10 µl from the reaction mixture and diluting it to 2ml in a cuvette with the buffer in a Perkin Elmer spectrometer at 437nm wavelength while excitation was at 380nm. The control consisted of Furin and substrate without the inhibitor. The fluorescence emitted by the fluorogenic substrate and rCrSPI-1 was also measured simultaneously and the background fluorescence was subtracted later.

4.3 Results and Discussion

4.3.1 *Overall structure*

The structure of recombinant CrSPI-1 (rCrSPI-1) in complex with subtilisin was solved by the molecular replacement method from a synchrotron data set. The model was refined to a final R-factor of 0.21 (R_{free}=0.24) at 2.6Å resolution with good

stereochemical parameters (Table 4.1). The final refined model consists of residues from Cys1 to Val73 of rCrSPI-1 and Ala1 to Gln274 of subtilisin. The last 10 residues from the C-terminal of rCrSPI-1 had no interpretable electron density map and were not modeled. There are three subtilisin and one rCrSPI-1 molecules in the asymmetric unit. Each rCrSPI-1 molecule interacts with two subtilisin molecules, i.e. domain-1 and domain-2 of rCrSPI-1 interacts with two independent subtilisin molecules respectively (Figures 4.3a, 4.3b, 4.3c). The third subtilisin molecule of the asymmetric unit is not in complex with rCrSPI-1.

Table 4.1. Crystallographic data and refinement statistics.

Data collection	
Data collection	
Unit cell parameters(Å, °)	a = 73.84, b = 65.07, c = 111.90 α = 90, β = 95.44, γ = 90
Space group	P2$_1$
Resolution range	50-2.6
Wavelength	0.9600
Observed *hkl*	64562
Unique *hkl*	34045
Completeness (%)	93.8
Redundancy	3.4
Overall *I/6I*	4.5
[a]R$_{sym}$ (%)	0.11
Refinement and quality of model	
*Resolution range (Å)	15-2.6
[b]R$_{work}$ (%) no. reflections	0.21(24818)
[c]R$_{free}$ (%) no. reflections	0.24(1533)
Root mean square deviation	
Bond length (Å)	0.009
Bond angle (º)	1.5
Ramachandran plot	
Favored region (%)	83.4
Allowed regions (%)	15.3
Generously allowed region (%)	1.4
Disallowed regions (%)	0.0
[d]Average B-factors (Å^2)	
Main chain atoms	37.396
Side chain atoms	37.314
Overall protein atoms (no. atoms)	37.361(6275)
Waters (no. atoms)	35.474 (168)

[a] Rsym = |Ii -<I>|/ |Ii| where Ii is the intensity of the ith measurement, and <I> is the mean intensity for that reflection.
[b] Rwork = | Fobs - Fcalc|/ |Fobs| where Fcalc and Fobs are the calculated and observed structure factor amplitudes, respectively.
[c] Rfree = as for Rwork, but for 8.5% of the total reflections chosen at random and omitted from refinement.
[d] Individual B-factor refinement were carried out.

Figure 4.3a. Structure of the rCrSPI-1-(Subtilisin) $_2$ complex. Subtilisin molecules are drawn in blue and green color and two domains of rCrSPI-1 are in magenta and yellow color. This figure was prepared by using Molscript(Kraulis, 1991)
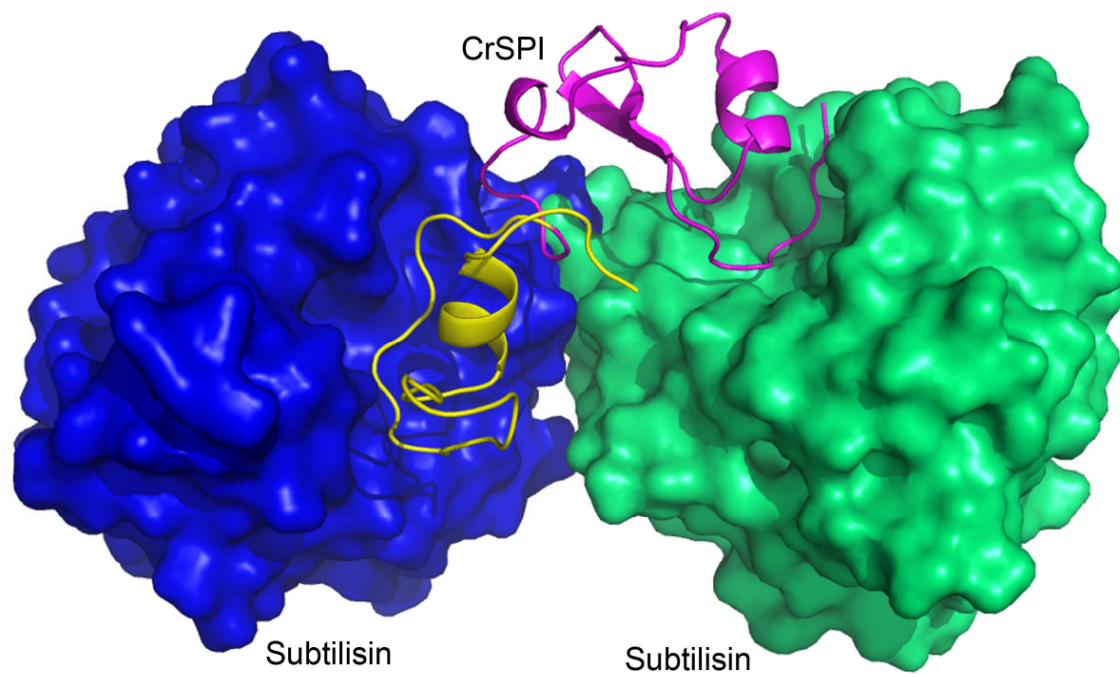
Figure 4.3b. Shows the CrSPI-1: Subtilisin complex. CrSPI-1 is in ribbon representation and subtilisin is in surface representation.
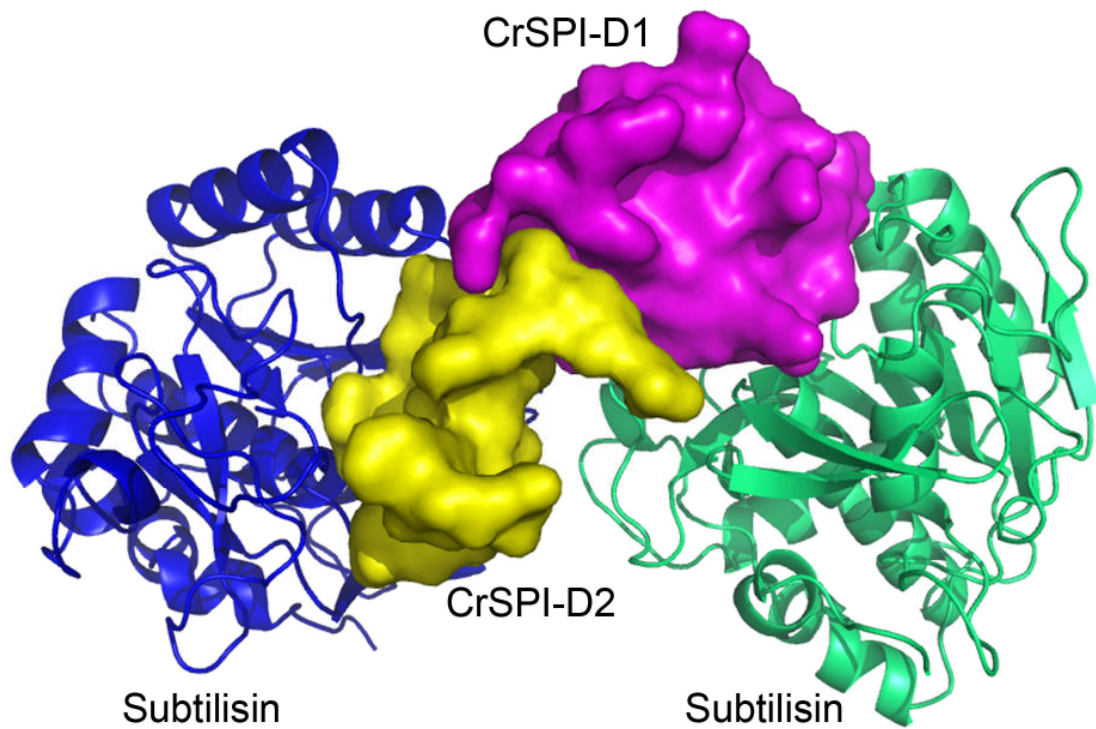
Figure 4.3c. Shows the CrSPI-1: Subtilisin complex. CrSPI-1 in surface representation and subtilisin is in ribbon representation. This figure was prepared by using PyMol.

The conformation of subtilisin molecules in the complex is similar to that seen in the apo subtilisin Carlsberg (EC 3.4.21.62) structure and / or in complex with small ligands (Stoll et al, 1998, Bode *et al*, 1987; Schmitke et al, 1998), and therefore we will not describe the structure in detail but provide only a general overview. The structure consists of a central seven-stranded parallel β-sheet with two α-helices on one side and, a group of four α -helices on the other side of the central β-sheet. The catalytic triad Ser220, His64 and Asp32 are located in the substrate binding cleft. The three subtilisin molecules of the asymmetric unit are identical with an rmsd of 0.1Å in a pairwise superposition of 274 Cα atoms.

Structure of CrSPI-1

rCrSPI-1 was found to exist as a mixture of both monomers and dimers in solution as determined by gel filtration, native-PAGE and Dynamic Light Scattering experiments (DLS). rCrSPI-1 molecule comprises of two domains: domain-1 from Cys1 to Glu40 and domain-2 from Leu41 to Leu83. Both domains adopt similar secondary structures (domain-1 β1↑β2↓α1β3↑ and domain-2 β4↑β5↓α2). The presence of a central α-helix α1 (Glu18–Ala24) in the domain-1 and the α2 (Arg63-Ser68) in the domain-2, and an antiparallel β-sheet in each domain is characteristic of the classical Kazal-type inhibitors. However, it also shows features of a non-classical Kazal-type serine protease inhibitor in that rCrSPI-1 harbors an unusual pattern of conserved cysteines. There are two intra-domain disulphide bridges in domain-1 (Cys1-Cys19 and Cys8-Cys35) and three in domain-2 (Cys41-Cys70, Cys45-Cys64 and Cys53-Cys82) (Figure 4.3d).
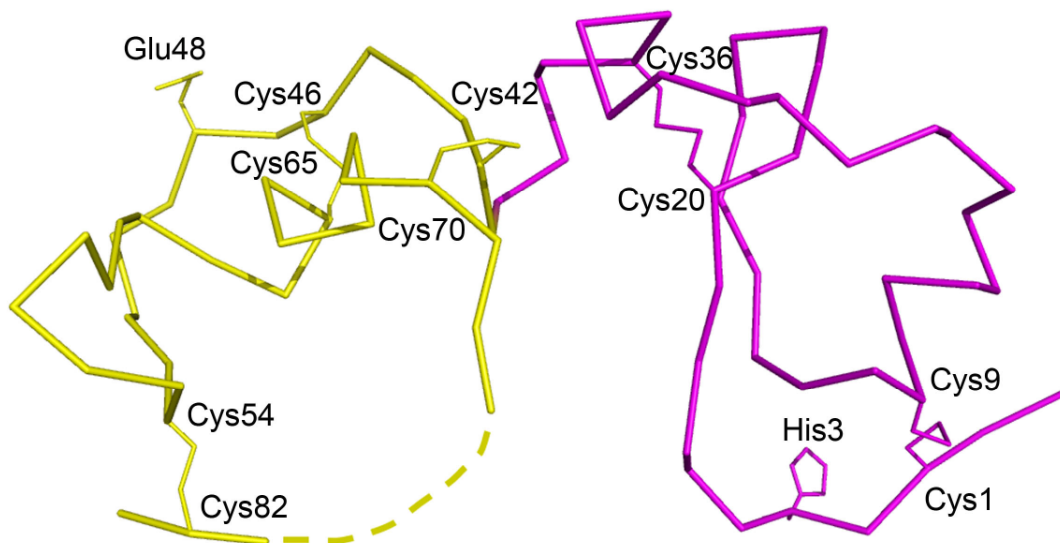
Figure 4.3d. Cα trace of rCrSPI-1. The disulphide bonds are shown in thick lines along with the residue numbers. Domain-1 is in magenta and domain-2 is in yellow color. The approximate location of the missing residues (residues 74 to 81) is shown as doted line. These figures were prepared by using PyMol. Using BLAST search, the sequence identities between rCrSPI-1 and the most homologous member of the Kazal family of inhibitors were analyzed. The observed pattern of S-S bridges in rCrSPI-1 was similar to the non-classical Kazal-type group 1 inhibitors from sea anemone and crayfish than the mammalian and avian inhibitors (Figure 4.4).
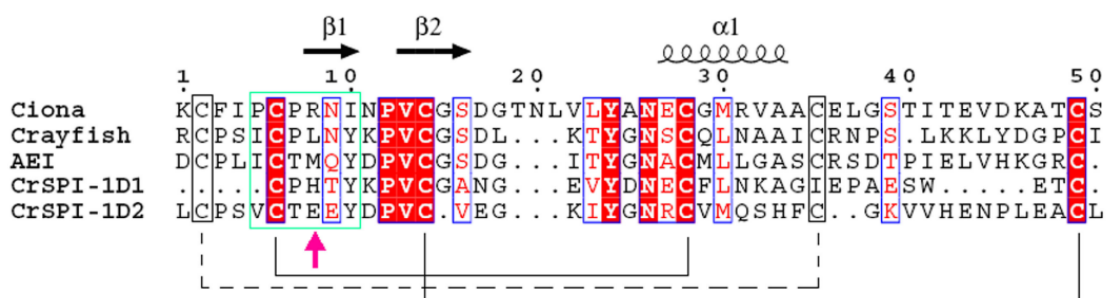
Figure 4.4. Multiple sequence alignment for the representative members of Kazal-type Non classical group I proteinase inhibitors. Secondary structural elements are shown

above the sequences. Conserved residues are shaded in red and yellow boxes. The reactive site loop residues of P4-P2' are represented in a green box. The reactive site residue P1 is shown using an upward pointing magenta arrow. This figure was prepared by the program Espript (Gouet, 1999)

A search for topologically similar proteins within the PDB database was performed with the DALI program (Holm & Sander, 1991). The full length rCrSPI-1 did not find any structural homologs. However the DALI search with individual domains shows significant similarity with several Kazal-type inhibitors. The DALI search for domain-1 (residues Cys1 – Glu40) against the pdb database revealed the highest structural similarities with leech-derived tryptase inhibitor (pdb code 1ldt) yielding an rmsd of 2.0Å for 38 Cα atoms with 30% sequence identity. This is followed by dipetalin a Kazal-type thrombin inhibitor (rmsd=2.0Å for 37 Cα atoms; 27% identity; pdb code 1kma) and the insect derived double domain Kazal inhibitor rhodniin, a highly specific thrombin inhibitor (rmsd=1.5Å for 36 Cα atoms; 35% identity; pdb code 1tbq). Similarly the structural homology search for domain-2 (Leu41-Leu83) shows several hits with Kazal-type inhibitors. Domain-2 shares highest structural homolog with turkey ovomucoid third domain (OMTKY3), a subtilisin inhibitor (pdb code 1yu6), gave an rmsd of 2.3Å for 32 Cα atoms with 25% sequence identity. This is followed by the thrombin inhibitor rhodniin (rmsd=2.3Å for 31 Cα atoms; 8% identity; pdb code 1tbq) and leech-derived tryptase inhibitor (pdb code 1an1) yielding an rmsd of 2.4Å for 29 Cα atoms with 20% sequence identity. To our knowledge most of the structurally known Kazal-type inhibitors contains single domain except for rhodniin. Rhodniin is the only

structurally characterized Kazal-type inhibitor with two domains in complex with thrombin (1:1 ratio). Both rCrSPI-1 domains show significant structural similarities with the two domains of rhodniin. However the full length rCrSPI-1 could not be superimposed with full length rhodniin. This indicates the difference in relative orientation of the two domains.

The superposition of rCrSPI-1 domain-1 and -2 reveals that their core regions are similar with 24 Cα atoms having an rmsd of 3.9Å (Figure 4.3). The sequence alignment shows 27% identical (42% similar) residues between the two domains which includes the highly conserved cysteines (Figure 4.5a). The preservation of the highly conserved S-S bridges maintains the architecture of these domains. The overall fold similarity of these two domains together with the structural homology with other Kazal-type inhibitors suggests that all Kazal-type inhibitors evolved from a common ancestral gene via duplication while maintaining diverge amino acid sequences. A study by Merckel *et al.,* (2005; Mol Cell 18; 161-170) showed that the similarities between the tertiary structures are the indicators of gene duplication.
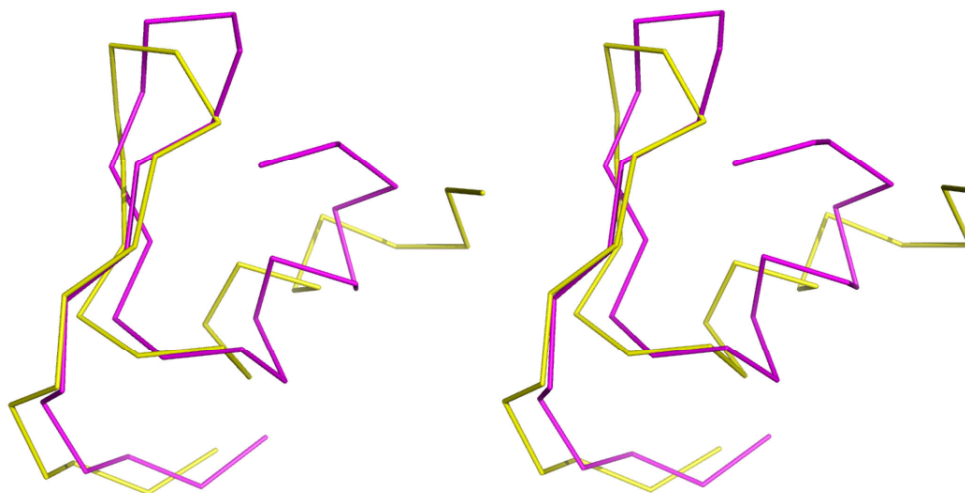
Figure 4.5a. Stereo view of the Cα superposition of domain-1 (magenta) and domain-2 (yellow) of rCrSPI-1. These two domains superimpose with an rmsd of 3.9Å for 24 Cα atoms. The superposition was carried out using DALI (Holm & Sander, 1991). This figure was prepared by using PyMol.

4.3.2 *rCrSPI-1: subtilisin complex*

rCrSPI-1: subtilisin complex is a heterotrimer in solution, confirmed by gel filtration, non-reducing SDS-PAGE and Dynamic Light Scattering (DLS) experiments (Figures 4.5b and 4.5c). This is consistent with the observation of a heterotrimer (1:2 ratio of rCrSPI-1: subtilisin) in the asymmetric unit of the crystal (Figure 4.1). Both domains of rCrSPI-1 act as two heads to bind with two individual subtilisin molecules. Our earlier work showed that a truncated domain-1 construct of rCrSPI, was not reactive to subtilisin, whereas domain-2 showed high affinity for subtilisin, at a Ki of 2.6 nM (Jiang *et al*, 2009). However, the structural studies revealed that although both domains indeed bind to subtilisin, only the domain-2 shows tight interactions with subtilisin. The

147
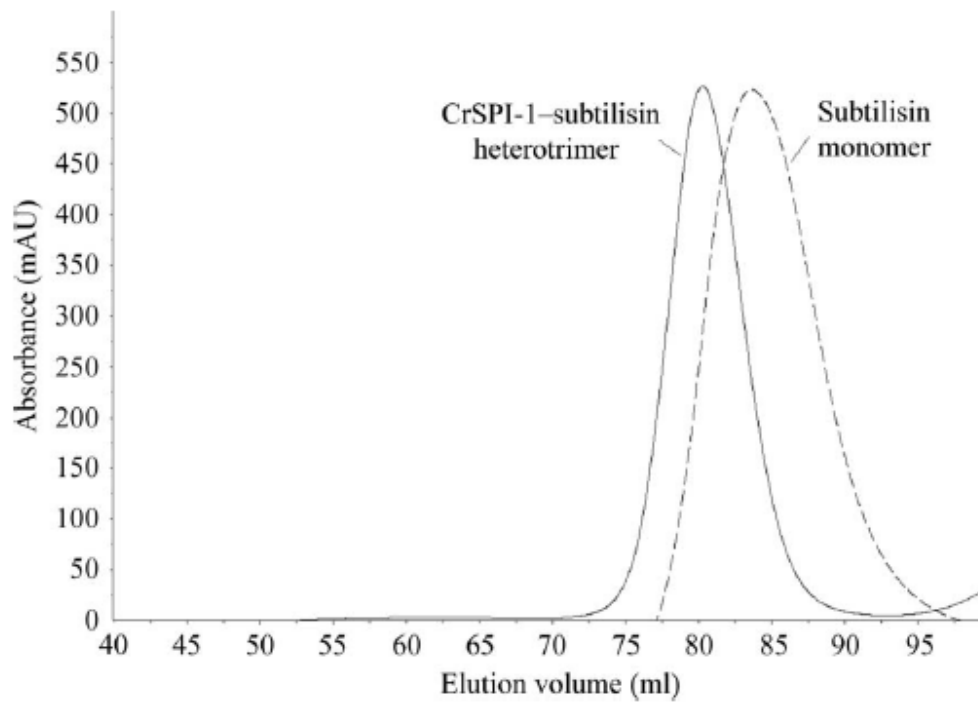
Figure 4.5b. Gel filtration profile of the CrSPI-1 Subtilisin complex together with subtilisin as a control run on a Superdex 75 column.
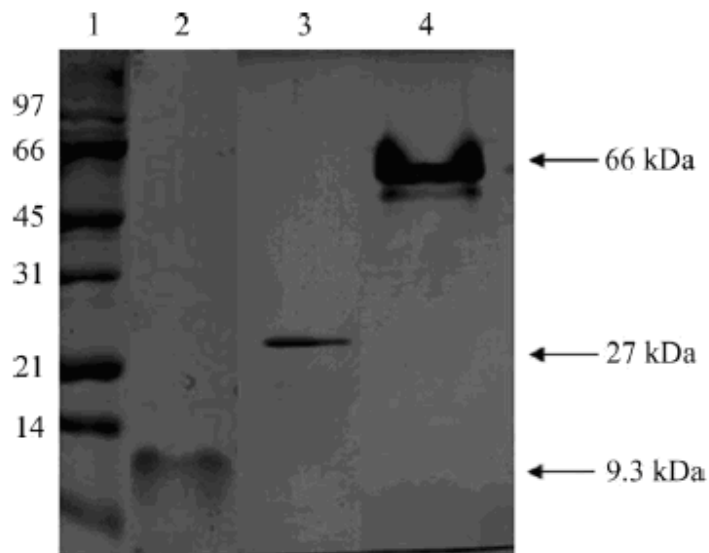
Figure4.5c.

Nonreducing SDS gel of the CrSPI-1–subtilisin complex. Lane 1, protein markers (kDa); lane 2, CrSPI-1; lane 3, subtilisin; lane 4, CrSPI-1–subtilisin complex with a molecular weight corresponding to a 1:2 ratio of CrSPI-1:subtilisin. Notably, CrSPI-1 runs slow on this gel. The actual size of this recombinant protein is 9.3 kDa and has been verified by mass spectrometry.

two subtilisin molecules are bound at opposite ends of the elongated inhibitor molecule forming a ternary complex. Apo subtilisin is a monomer in solution; upon formation of a complex with rCrSPI-1, it becomes a heterotrimer. Similar complex formation has been documented for rhodniin, a double domain non classical group II Kazal-type inhibitor from the blood sucking insect *Rhodnius prolixus,* which inhibits thrombin via one domain binding at the active site and the other domain binding the exosite (van de Locht *et al*, 1995). Similarly a tomato inhibitor belonging to the potato II inhibitor family, a completely different family of serine protease inhibitors; specifically inhibits (Ki=9nM) two subtilisin molecules (Barrette-Ng et al, 2003). The present study on rCrSPI-1: subtilisin heterotrimer complex is the first of this kind for the Kazal-type inhibitors which employs a dual-inhibition mechanism to engage two protease molecules with different specificity.

4.3.3 *CrSPI-1 RSL interactions with subtilisin*

The reactive-site loop (RSL) (Cys1 to Lys6) of domain-1 of rCrSPI-1 binds at the active site region of subtilisin from P3 (Cys1) to P3' (Lys6) in a substrate-like manner (Figure 4.4a). The reactive site residues are well defined in the electron density map (Figure 4.4c). There are 11 hydrogen bonding contacts between domain-1 RSL and subtilisin. Of these 5 of them are sidechains mediated hydrogen bonding contacts. The mainchain amide group of P1 residue His3 is involved in hydrogen bonding contact with O^{γ} of the catalytic Ser220 (2.63Å). The carbonyl O atom of the P1 His3 interacts with backbone amide Nitrogen of active site Ser220 and $N^{\gamma 2}$ of Asn154 of subtilisin (Table 4.2).

In Kazal-type inhibitors, the P1 residue is the most exposed residue and makes approximately 50% of the hydrogen bonding interactions with the protease active site residues. (Kleanthous, 2000; Maynes *et al*, 2005). In the subtilisin-rCrSPI-1 complex, domain-1 P1

Table 4.2. Selected hydrogen bonding contacts between rCrSPI-1 domain-1 and subtilisin

Site	CrSPI1	Subtilisin	Distance (Å)
P3	Cys D2 N	Gly A126 O	3.14
	Cys D2 O	Gly A126 N	2.88
P1	His D4 N	Ser A220 O^{γ}	2.63
	His D4 $N^{\delta 1}$	Asn A154 $N^{\delta 2}$	3.03
	His D4 O	Asn A154 $N^{\delta 2}$	2.66
	His D4 O	Ser A220 N	3.28
	His D4 O	Ser A220 O^{γ}	2.92
P1'	Thr D5 N	Ser A 220 O^{γ}	3.34
P2'	Tyr D6 N	Asn A217 O	2.65
	Asp D18 $O^{\delta 1}$	Asn A154 O	2.86
	Asp D18 $O^{\delta 2}$	Asn A154 O	2.54

residue His3 contributes 5 hydrogen bonding contacts (or 45% of the total hydrogen bonding interactions) with subtilisin (Figure 4.6a). Pro2 occupies the P2 position and having hydrophobic interactions with subtilisin. In most of the canonical serine proteinase inhibitors the P3 residue is engaged in a disulphide bond. In rCrSPI-1 Cys1 is in P3 position is engaged in a disulphide bond with Cys19 and part of the hydrophobic core which consists of Leu125 (subtilisin) and Pro2 (rCrSPI-1).   Thr4 is in the P1' position along with Tyr5 and Lys6 are in P2' and P3' positions respectively. P2' residue Tyr5 maintains stacking interactions with Phe188 of subtilisin.
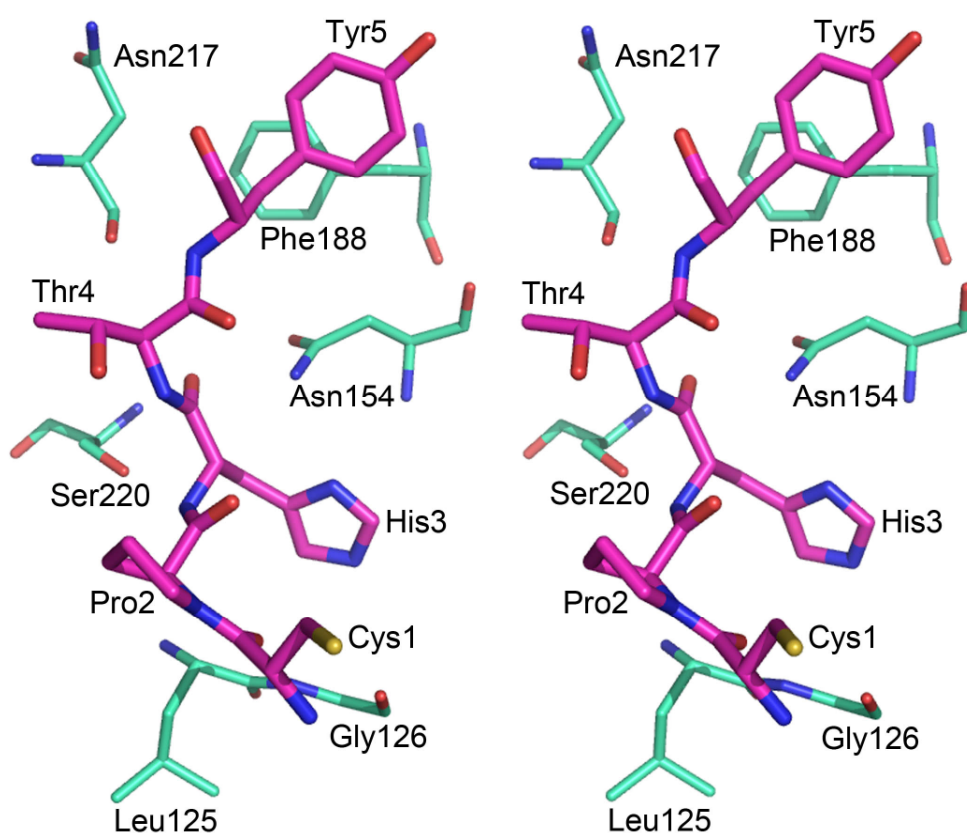
Figure 4.6a. Stereo view of the interactions between subtilisin (cyan) and the reactive site loop of domain-1 (magenta). In domain-2 of rCrSPI-1 Glu48 takes the position of His3 in the P1 pocket of domain-1 and makes 10 out of 13 total hydrogen bonding contacts with subtilisin (Fig 4.6b). The remaining hydrogen bonding contacts with subtilisin are from P1' Glu49, P2' Tyr50 and P4 Val45 of domain-2 respectively (Table 4.3). Similar to the P2' residue Tyr5 of domain-1, Tyr50 takes the P2' position in domain-2 and maintains stacking interactions with Phe188 of subtilisin. In addition P3 residue Cys46 mediates hydrophobic

Table 4.3. Selected interactions between rCrSPI-1 domain-2 and subtilisin

Site	CrSPI-1	Subtilisin	Distance (Å)
P4	Val D46 O	Gly C126 N	3.29
P1	Glu D49 N	Ser C220 O$^\gamma$	2.75
	Glu D49 O$^{\varepsilon 1}$	Asn C154 N$^{\delta 2}$	2.84
	Glu D49 O$^{\varepsilon 1}$	Asn C154 N	3.06
	Glu D49 O	Thr C219 N	3.36
	Glu D49 O	Ser C220 N	3.17
	Glu D49 O	Asn C154 N$^{\delta 2}$	2.60
	Glu D49 O	Ser C220 O$^\gamma$	2.84
	Glu D49 O	Asn C154 O$^{\delta 1}$	3.32
	Glu D49 C$^\alpha$	Ser C220 O$^\gamma$	2.79
	Glu D49 C	Ser C220 O$^\gamma$	2.64
P1'	Glu D50 N	Ser C220 O$^\gamma$	3.22
P2'	Tyr D51 N	Asn C217 O	2.92

interactions with the side chain Leu125 of subtilisin. Overall RSL interactions in domain-2 clearly indicate that the P1 residue Glu48 is the primary mediator of interactions with subtilisin. These observation combined with the binding affinities of both domains with subtilisin suggests that domain-2 is more preferred to interact tighter with subtilisin than domain-1.
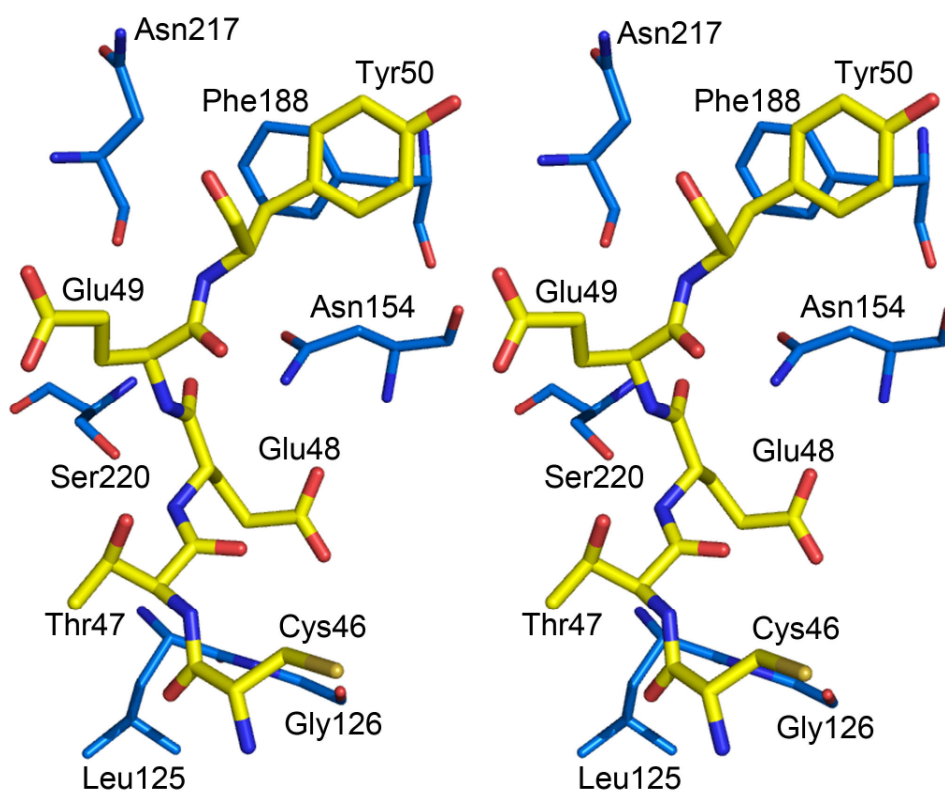
Figure 4.6b. Stereo view of the interactions between subtilisin and the reactive site loop of domain-2.

Table 4.4 show the P3 to P3' residues of RSLs of various substrates-like binding serine protease inhibitors such as for subtilisin, thrombin and Furin. It is evident that subtilisin prefers Glu in the P1 position of the Kazal-type inhibitors and the Furin inhibitors hold a basic residue. In addition to the highly conserved P3 position Cys, P2'

154

position is taken by a hydrophobic sidechain. Further the P1 position is relatively variable and it determines the specificity of the proteases. Specific interactions between the RSLs of rCrSPI-1 and subtilisin are schematically illustrated in supplementary Figure 4.3.

Table 4.4. Reactive site loop regions from P3 to P3' position of selected serine protease inhibitors

Inhibitors	**P3**	**P2**	**P1**	**P1'**	**P2'**	**P3'**
CrSPI-1 domain-1	Cys1	Pro2	His3	Thr4	Tyr5	Lys6
CrSPI-1 domain-2	Cys46	Thr47	Glu48	Glu49	Tyr50	Asp51
OMTKY3	Cys16	Thr17	Leu18	Glu19	Tyr20	Arg21
Eglin C	Asn117	Gly118	Met119	Asp120	Val121	Ile122
Tomato inhibitor II dom-1	Cys3	Thr4	Arg5	Glu6	Cys7	Gly8
Tomato inhibitor II dom-2	Cys60	Thr61	Phe62	Asn63	Cys64	Asp65
Rhodniin domain-1	Cys8	Pro9	His10	Ala11	Leu12	Arg13
Rhodniin Cys60 domain-2	Asp61	Gly62	Asp63	Glu64	Tyr65	Lys66
Spn4A (Furin inhibitor from Drosophila)	Arg350	Lys351	Arg352	Ala353	Ile354	Met355
Human PI8(Furin Inhibitor)	Asn337	Ser338	Arg339	Cys340	Ser341	Arg342

Note: <u>Subtilisin inhibitors</u>: CrSPI-1 domain-2; Tomato inhibitor II (domain-1 and domain-2); OMTKY3 and Eglin C. <u>Thrombin inhibitor</u>: Rhodniin inhibitor. <u>Furin inhibitor</u>: Spn4A from Drosophila melanogaster, Human Protease Inhibitor 8 and CrSPI-1 domain-1 (annotated).

4.3.4 *Rigidity of the RSL*

The conformation adopted by the reactive site loops (RSLs) of both domains of rCrSPI-1 is similar. They superimposed with an rmsd of 0.80Å for the residues from position P3 to P3' (Figure 4.7a). The RSL conformation in several families of serine proteinase inhibitors, in many complexes and in different crystal environments adopts a similar conformation. Similar to other members of the Kazal family of inhibitors, the disulfide bonds formed by cysteine residues at the P3 and P5' positions (Cys1 and Cys8 in domain-1 and Cys46 and Cys54 in domain-2) of CrSPI-1 may hold the RSL in a relatively rigid conformation.

The RSLs backbone torsion angles (ψ/φ angles) of rCrSPI-1 is similar to each other and compares with other proteinase inhibitors such as OMTKY3 and Eglin-C complex with subtilisin (Table 4.5). In addition, the RSLs backbone torsion angles of rCrSPI-1 are similar to several structurally unrelated serine protease inhibitors complexes and confirm the canonical binding mode for rCrSPI-1 (Supplementary Table 4.4; Figure 4.7b).

Table 4.5. Main chain torsion angles of the reactive site loops of serine protease inhibitors complexed with subtilisin.

	CrSPI-1 Domain I torsion angles (Φ/ψ)	CrSPI-1 Domain II torsion angles (Φ/ψ)	OMKTY3 torsion angles (Φ/ψ)	Eglin C complex torsion angles (Φ/ψ)
P4	-	-171/-63	-101/140	-71/140
P3	-126/163	61/167	-132/156	-138/168
P2	-76/156	-131/167	-62/161	-62/143
P1	-100/46	-131/40	-106/36	-115/44
P1'	-82/146	-91/133	-77/134	-96/168
P2'	-101/88	-82/100	-96/106	-117/109
P3'	-122/71	-138/68	-142/74	-121/112
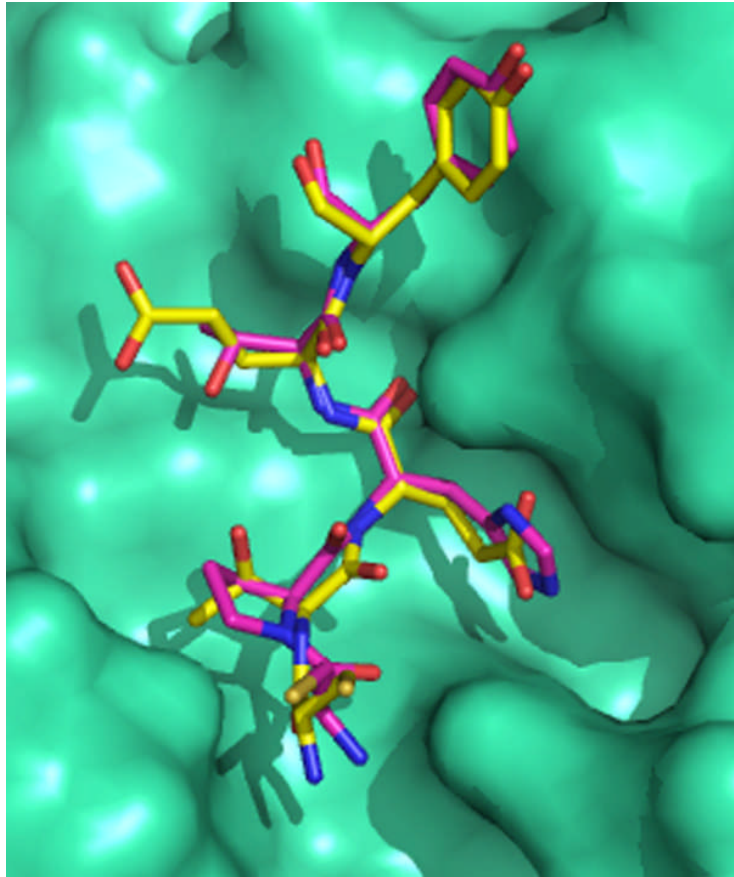
Figure 4.7a. Conformations of the reactive site loop (RSL) of domain-1 (magenta) and domain-2 (yellow) bound to subtilisin
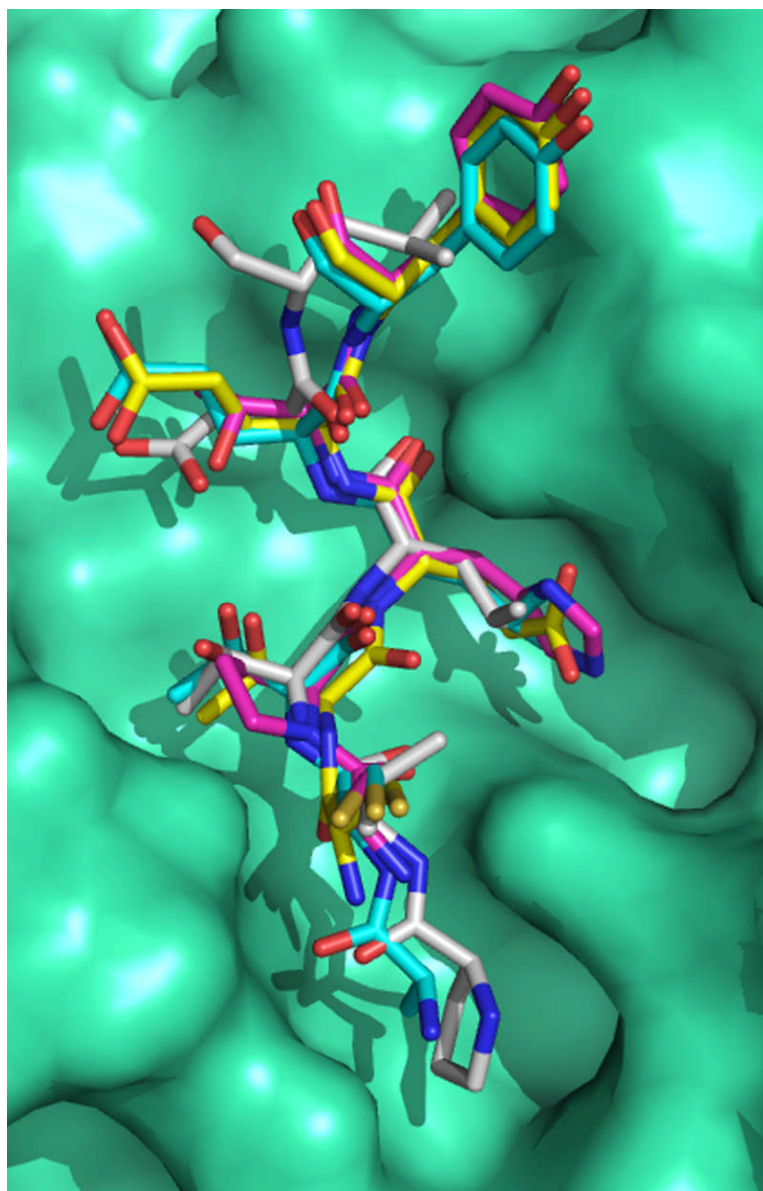
Figure 4.7b. Superposition of the reactive site loops of domain-1, domain-2, Eglin C (gray) and OMTKY3 (cyan). The RSLs are shown in stick representation whereas the substrate binding site of subtilisin is shown in surface representation. These figures were generated by using PyMol.

The superposition of the active site region of subtilisin complexes such as OMTKY3 and Eglin C with rCrSPI-1: subtilisin heterotrimer complex reveals that different inhibitor loop sequences can be accommodated in the substrate binding site of subtilisin with minimal sidechain rearrangement (Figure 4.7b). Further it indicates the rigid conformation of the RSL of these inhibitors. This is further supported by the fact that both the mainchain and sidechain groups of subtilisin remain relatively in the same conformation in several complexes and in apo form. In domain-2, out of three S-S bridges two of them (Cys41-Cys70 and Cys45-Cys64) are connected with the central α-helix region, which maintains a rigid conformation for the reactive site loop (RSL) of this domain than the RSL in domain-1. Notably, the presence of three disulphide bonds in domain-2 probably explains for its rigidity and makes it specific and potent towards its particular cognate protease subtilisin. It is worth mentioning here that in the case of OMSVP3 in which introducing an additional disulphide bridge made it specific to only one protease α-chymotrypsin and lost its inhibition for other proteases (Hemmi *et al*, 2003).

In addition to the S-S bridges, there are three important internal hydrogen bonds which aid maintaining the rigidity of the RSL in CrSPI-1. These hydrogen bonds in domain-1 are (1) P2-P1' hydrogen bond between the carbonyl oxygen of Pro2 and amide nitrogen of Thr4 of the reactive site loop, (2) hydrogen bonding between Asn18 and Phe21, and (3) the $N^{\delta 2}$ of Asn18 interacts with the mainchain carbonyl atoms of Pro2 and Thr4 at the P2 and P1' positions of the RSL. Similarly, in domain-2, the interactions which involved the sidechains from Asn62, Thr47 and Glu49 may aid to maintain the rigidity of the RSL. The interactions between Thr47 and Glu49 are similar to the

interactions of P2-P1' of domain-1. Similar interactions are also observed in the OMTKY3 complexes with several proteases (Maynes *et al*, 2005;).

In order to reduce the entropic cost of binding the reactive site loops (RSLs) of Kazal-type inhibitors and in particular, CrSPI-1 is firmly held in a preferred conformation that is complementary to the substrate binding site of specific proteases. Further a rigid conformation of the RSL is likely to prevent proteolytic cleavage of the inhibitor upon interaction with proteases.

4.3.5 *Specificity of CrSPI-1 domains*

Kazal type serine protease inhibitors usually occur as single or multi-domain proteins with each domain having a different specificity towards a particular protease (Somprasong *et al*, 2006). Although the sequence of the reactive site loop tends to be highly variable, the specificity for a protease is dictated by the P1 residue. Furin, a homolog of subtilisin, has specificity towards paired basic residues for cleavage (Thomas, 2002). Furin cleaves a wide range of proproteins at the consensus sequence -Arg-Xaa-Lys/Arg-Arg-↓ and the minimal consensus sequence for Furin is -Arg-Xaa-Xaa-Arg-↓ (Jean, Francois *et al*). The CrSPI-1 domain-1 RSL consists of two basic residues (His3 and Lys6). Furin might not cleave this RSL due to its rigid conformation and thus it may act as an inhibitor for Furin. Notably the Kazal protease inhibitors recognize the proteases in a substrate-like manner. Several proteins have been engineered to provide a Furin consensus motif in their reactive site loops and thereby inhibit Furin. For instance, the Turkey ovomucoid third domain RSL sequence was mutated from Ala-Cys-Thr-Leu18 to Arg-Cys-Lys-Arg18(Lu *et al*, 1993), α1- antitrypsin Portland RSL mutated

from Ala355-Ile-Pro-Met358- to -Arg355-Ile-Pro-Arg358 and both have successfully inhibited Furin (Jean *et al*, 1998). These inhibitors mimic highly specific substrates. Due to the tight binding and specific rigid conformation of the RSL residues, these inhibitors are able to arrest the enzymatic reaction at the intermediate stage of hydrolysis of the peptide bond (Radisky *et al*).

Glu48 is the P1 residue in the RSL of domain-2 and the presence of Glu in P1 position is more specific for subtilisin inhibitors (Somprasong *et al*, 2006; Odum *et al*, 1999). Glu48 is well buried in the S1 pocket with 10 hydrogen bonding contacts with subtilisin compared to His3, the P1 residue of domain-1 which makes only 5 hydrogen bonding contacts. This suggests that Glu48 in the P1 pocket of domain-2 is more specific and maintains a tight interaction with subtilisin. In addition the P2 residue of domain-2 makes an additional hydrogen bond with the active site His63 sidechain as compared to domain-1.

586Å^2 surface area of domain-1 of CrSPI-1 is buried upon complex formation with subtilisin, whereas 592Å^2 for domain-2. In particular the reactive site loop residues of domains-1 and -2 contribute 445Å^2 and 529Å^2 of buried surface area respectively. This account for 76% and 89% of each domain's binding interface with subtilisin. Thus the RSL of domain-2 contributes a greater buried area compared to the RSL of domain-1. Notably domain-2 inhibits subtilisin with Ki =2.6nM, whereas no reactivity observed between domain-1 and subtilisin. Further the ITC experiments reveals that the integration curve fitted well considering a model in which one CrSPI-1 molecule binds two Subtilisin molecules (stoichiometry 1:2), with binding sites in CrSPI-1 considered non-identical and independent with affinities 0.37 and 670 μM (Fig 4.8a). This suggests that

domain-2 of CrSPI-1 most likely binds first to a subtilisin molecule following with domain-1 binds with a second subtilisin at a lower affinity.

4.4 Isothermal Titration Calorimetry (ITC) binding studies with CrSPI-1

ITC binding studies with tagged CrSPI-1 were carried out using VP-ITC and ITC-200 Calorimeters. The results are in fair agreement. The interaction between thioredoxin tagged CrSPI-1 and subtilisin is endothermic, therefore, entropically driven. On the other hand, the interaction of untagged CrSPI-1 is exothermic and with unfavourable entropic contribution. There is a difference in the thioredoxin tagged and untagged protein binding kinetics. There is a reduced conformational entropy loss upon binding with the tagged protein probably, due to the stability provided by the tag CrSPI-1. Thioredoxin tag makes the enthalpy positive and without tag, the enthalpy is found to be negative. This maybe because there is a conformational change coupled to binding, and the tag helps in promoting the binding competent conformation). There are two binding sites for the CrSPI-1: Subtilisin interaction (1:2). The Hill coefficient for the binding is 0.3-0.4. The two binding sites in CrSPI-1 could presumably be non-identical and independent, identical with negative cooperativity or non-identical with negative cooperativity. Considering that CrSPI-1 is not a symmetric molecule, the simplest and more plausible explanation is that two binding sites are non-identical and independent. The two binding sites show significantly different affinities: $k1 = 2x10^6 \ M^{-1}$, $k2 = 2x10^4 \ M^{-1}$ and binding enthalpies: dh1 = 40 kcal/mol, 6 kcal/mol. The model used for the analysis is a general model with two binding sites in CrSPI-1. The titrations correspond to reverse titrations, where the macromolecule with two binding sites is placed in the injection syringe. This is

the reason for the unusual look of the titrations: during the first injections CrSPI-1 is in an excess of Subtilisin inside the calorimetric cell and, therefore, both binding sites are occupied; as the titration proceeds, CrSPI-1 increases concentration and free Subtilisin decreases in concentration and Subtilisin dissociates from low affinity binding sites (accompanied by a negative enthalpic contribution).

β_1	2.4×10^6 M^{-1}
ΔH_1	40.6 kcal/mol
β_2	4.1×10^{10} M^{-2}
ΔH_2	48.1 kcal/mol
k_1	2.4×10^6 M^{-1}
Δh_1	40.8 kcal/mol
k_2	1.7×10^4 M^{-1}
Δh_2	7.3 kcal/mol
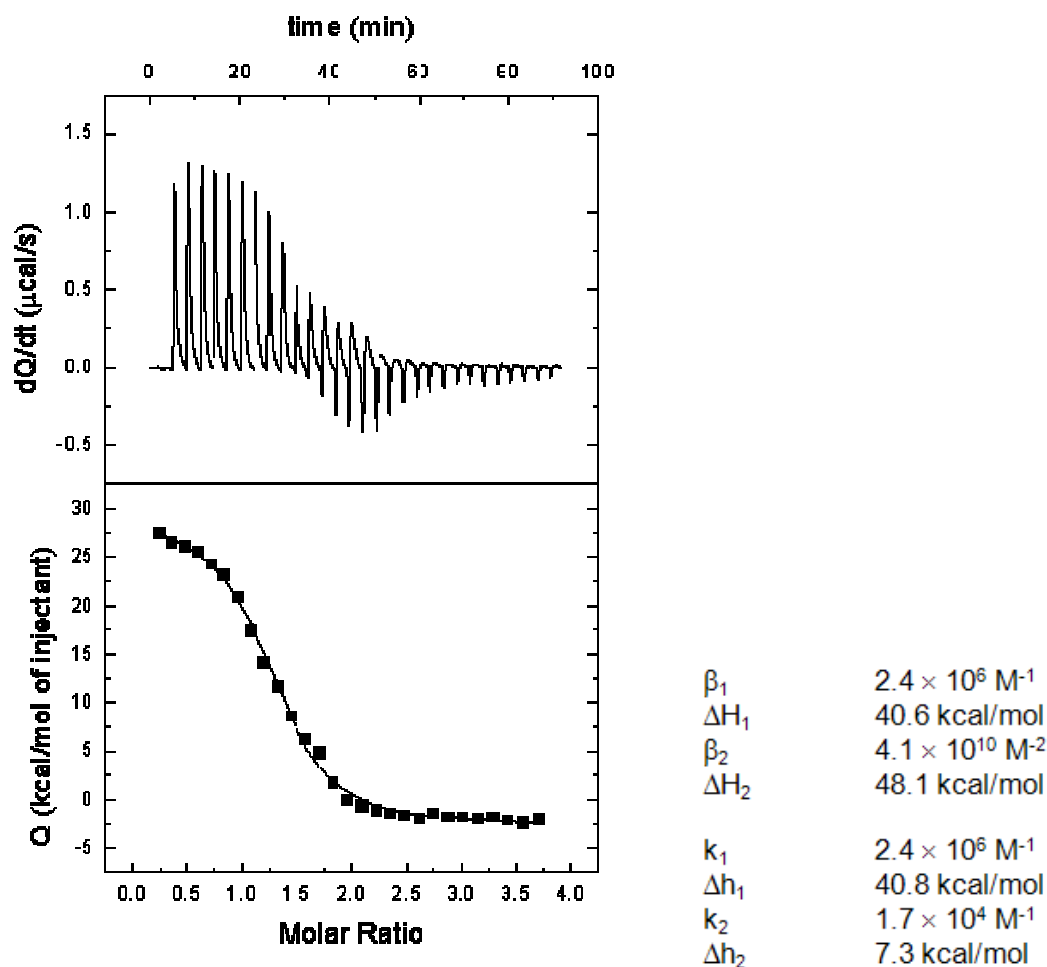
Figure 4.8a. Isothermal Titration Calorimetric (ITC) curve for rCrSPI-1 titrated against subtilisin at 37°C. Each peak represents the injection of rCrSPI-1 0.2 mM into the ITC cell containing subtilisin 0.012 mM, in buffer PBS pH 7.4, 10 mM BME. A sequence of 18 injections, each injection consisting of 2 μL of ligand solution, was performed. The experimental data were fitted considering a model in which CrSPI-1 binds two Subtilisin molecules, and the two binding sites in CrSPI-1 are non-identical and independent. Binding association constants of 2.7×10^6 M^{-1} and 1×10^3 M^{-1} were obtained from non-linear regression analysis, corresponding to dissociation constants of 0.37 and 670 μM, respectively.

β_1	1.6×10^6 M^{-1}
ΔH_1	39.0 kcal/mol
β_2	4.1×10^{10} M^{-2}
ΔH_2	44.6 kcal/mol
k_1	1.6×10^6 M^{-1}
Δh_1	39.6 kcal/mol
k_2	2.6×10^4 M^{-1}
Δh_2	5.0 kcal/mol
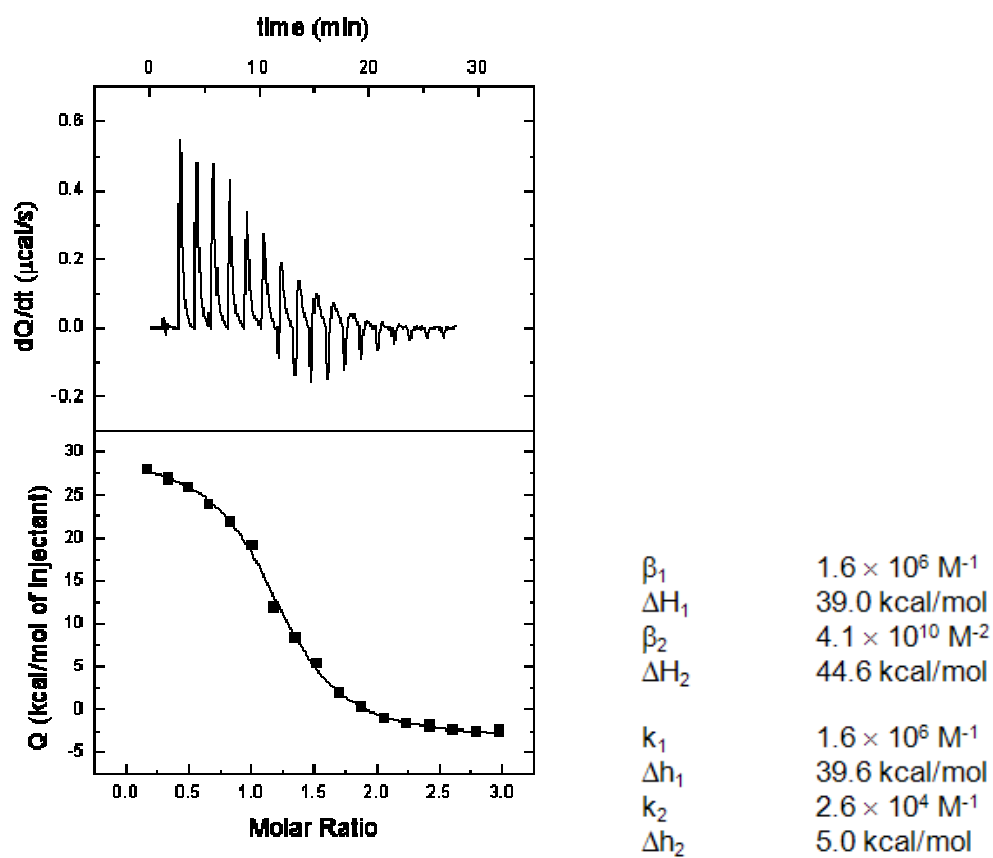
Figure 4.8b. Isothermal Titration Calorimetric (ITC) curve and binding parameters for rCrSPI-1 titrated against subtilisin. The experiment was carried out on a Microcal iTC-200 calorimeter.

ITC studies with peptide derived from CrSPI-1 domain 2

We have studied the interactions of various peptides derived from RSL regions of both domains of rCrSPI-1 using ITC experiments (Figure 4.9a, 4.9b). The interaction between the peptide VCTEEY derived from domain 2 reactive site loop and Subtilisin shows a 1:1 stoichiometry (one binding site). The binding affinity is low (k = 1.5×10^3 M^{-1}), which is lower than that of the low affinity binding site of CrSPI-1 domain 1 of 2×10^4 M^{-1}.

| K | 1.1×10^3 M^{-1} |
| ΔH | -21.2 kcal/mol |
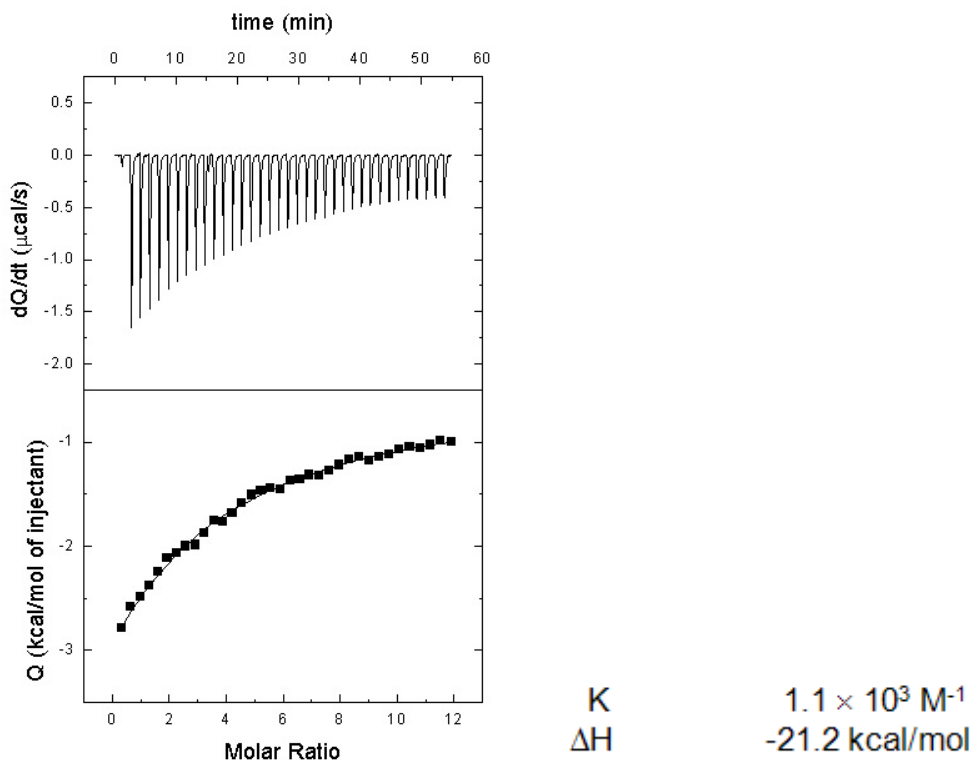
Figure 4.9a. Isothermal Titration Calorimetric (ITC) curve and binding parameters for VCTEEY titrated against subtilisin. The experiment was carried out on a Microcal iTC-200 calorimeter.
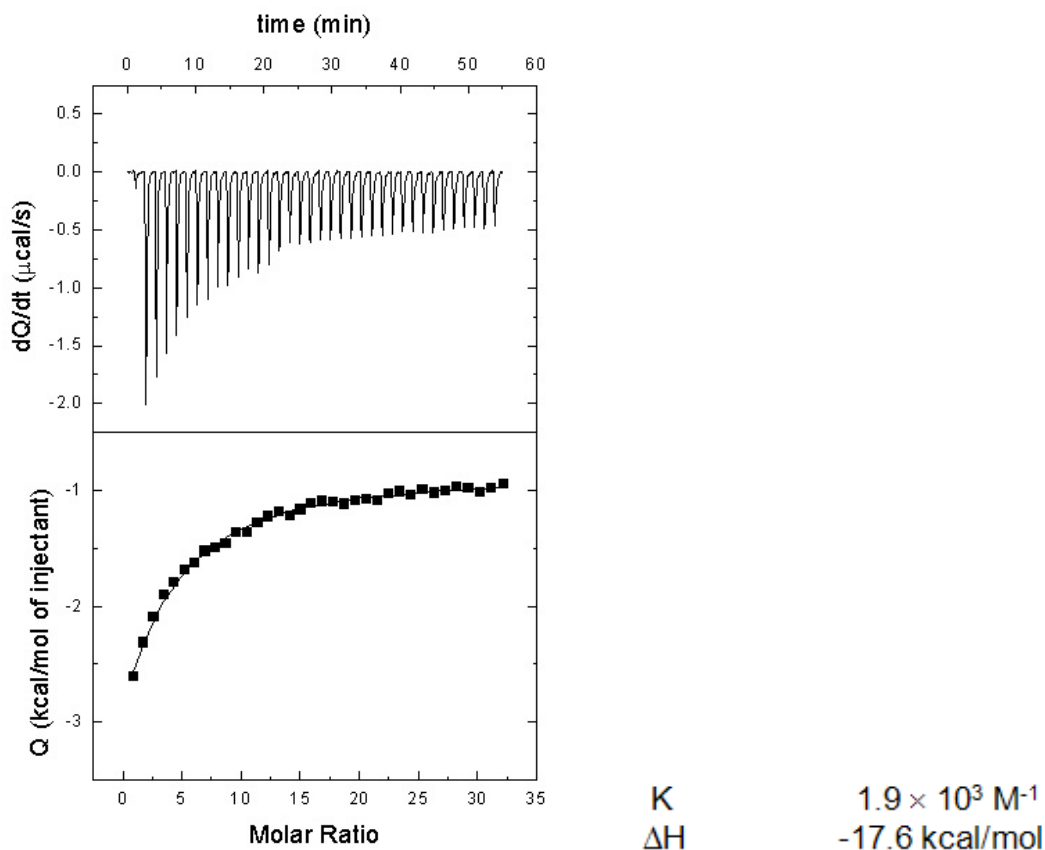
Figure 4.9b. Isothermal Titration Calorimetric (ITC) curve and binding parameters for VCTEEY titrated against subtilisin. The experiment was carried out on a Microcal iTC-200 calorimeter.

Several Kazal-type inhibitors with Glu in the P1 site are known to inhibit subtilisin. For instance, the five domain shrimp Kazal inhibitor SPIPm2 – with two subtilisin inhibiting domains having P1 Glu residues (Somprasong *et al,* 2006), EPI1, a Kazal-Like Protease Inhibitor from Phytophthora infestans (Miaoying Tian *et al,* 2005) has P1 Glu residue. But so far, no Kazal inhibitor of subtilisin has been known to have a His residue in the P1 site. Further His residue in P1 site is very rare among all the Kazal

inhibitors discovered so far. This leads us to believe that domain-1 may not be a subtilisin specific inhibitor and instead be responsible for Furin inhibition. The presence of two basic residues in the RSL of domain-1 is most likely the key residues for inhibition of Furin in a substrate-like manner. It must be noted that this model needs to be further tested for validation. These two domains of CrSPI-1 could have been a product of gene duplication to generate a dual specificity inhibitor with one domain functioning as an inhibitor of host proprotein converting subtilisin-like enzyme CrFurin and the other domain functioning as a pathogen specific protease, subtilisin inhibitor.

4.5 *Implications for the possible dual functions of CrSPI-1*

The preference of CrSPI for microbial protease, subtilisin suggests that this isoform targets serine proteases of the invading microbes. In fact, subtilisin is the virulence factor used by various pathogens or parasites to gain entry into the host cells during infection. Besides direct suppression of the microbial proteases, CrSPI-1 was shown to immunomodulate the host through its interaction with the host target, CrFurin (Jiang *et al,* 2009). CrFurin, a host serine protease which is a homolog of subtilisin-type serine protease involved in processing multiple immune proproteins. By an ex vivo inhibition assay using recombinant CrSPI-1 we have demonstrated that CrFurin is the endogenous cognate protease of CrSPI (Jiang *et al.,* 2009). It was observed that the transcription profile and the protein activities of CrFurin and CrSPI-1 showed a reciprocal relationship.

The Furin binding mode of CrSPI-1 was predicted by a comparison with the complex crystal structure of rCrSPI-1: subtilisin and Furin (Pdb code 1p8j). Furin is a
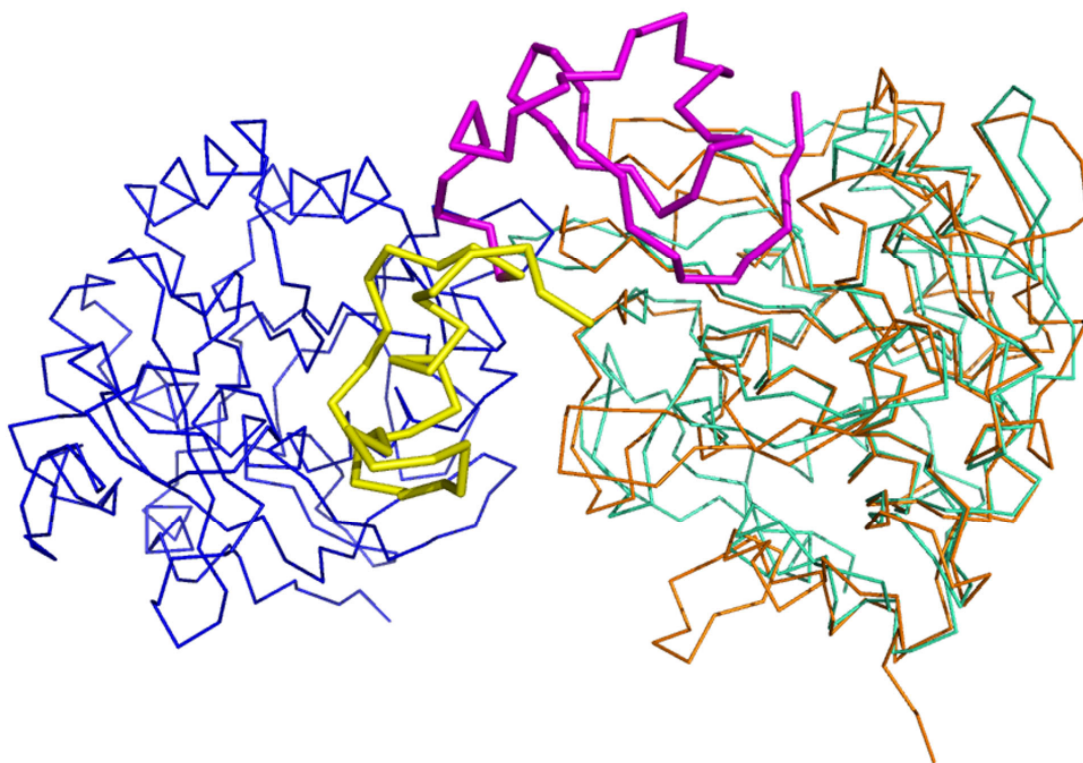
Figure 4.10a. Model of Furin-rCrSPI-Subtilisin heterotrimer complex shows Cα representation of superimposed Furin (in orange color) on subtilisin-rCrSPI complex. Furin and subtilisin share a sequence identity of 23%. The Furin: CrSPI-1 complex model was generated by superimposing domain-1 of CrSPI-1: subtilisin complex onto the structure of Furin (pdb code 1p8j) to yield an rmsd of 2Å for 268 Cα out of 274 Cα atoms of Subtilisin. The Furin-rCrSPI-1-Subtilisin heterotrimer complex was generated using the modeled Furin-rCrSPI-1-domain-1 and subtilisin-rCrSPI-1-domain-2 complex crystal structure.

homolog of subtilisin and shares a sequence identity of 23% (~40% similarity). Moreover the structure of subtilisin superimposes with subtilisin-like domain of Furin with an rmsd of 2Å for 268 Cα atoms out of 274 Cα atoms of subtilisin. The domain-1 CrSPI-1: subtilisin complex was superimposed onto the structure of Furin (pdb code 1p8j) and copied the coordinates of CrSPI-1 to Furin (Figure 4.8). The RSL region from P3 (Cys1) to P3' (Lys6) occupies the substrate binging site of Furin. In addition compared to the binding of subtilisin with CrSPI-1-domain-1, the Furin-CrSPI-1 domain-1 reveals several additional interactions between Furin and domain-1 (Figure 4.8). The reactive site loop residues of domain 1 of CrSPI-1 spanning P3 to P3' are accommodated well in the active site cleft of Furin. Similar to a typical Furin inhibitor Arg in the P1 position, the P1 residue His 3 side chain extends into the S1 pocket of the active site and is in close proximity of the nucleophilic Ser368 of Furin. The second basic residue P3' Lys is found in the active site cleft close to the S2' subsite of Furin. The presence of these two basic residues in the negatively charged cleft of Furin could explain the potential inhibitory activity of CrSPI-1 against Furin. Further, the superimposition of Subtilisin and Furin reveals that the catalytic triads of Subtilisin (Ser 220 – His 64 – Asp 32) and Furin (Ser 368 - His 194 - Asp 153) are found to be similar and can be superimposed very well. The subtilisin-like domain of Furin is very similar to Subtilisin but there are some differences in the residues forming the active site which may explain the difference in the surface charge of the active sites of the two proteases (Henrich *et al*, 2003). Furin has a much higher negative charge in the active site cleft and hence prefers basic residues.
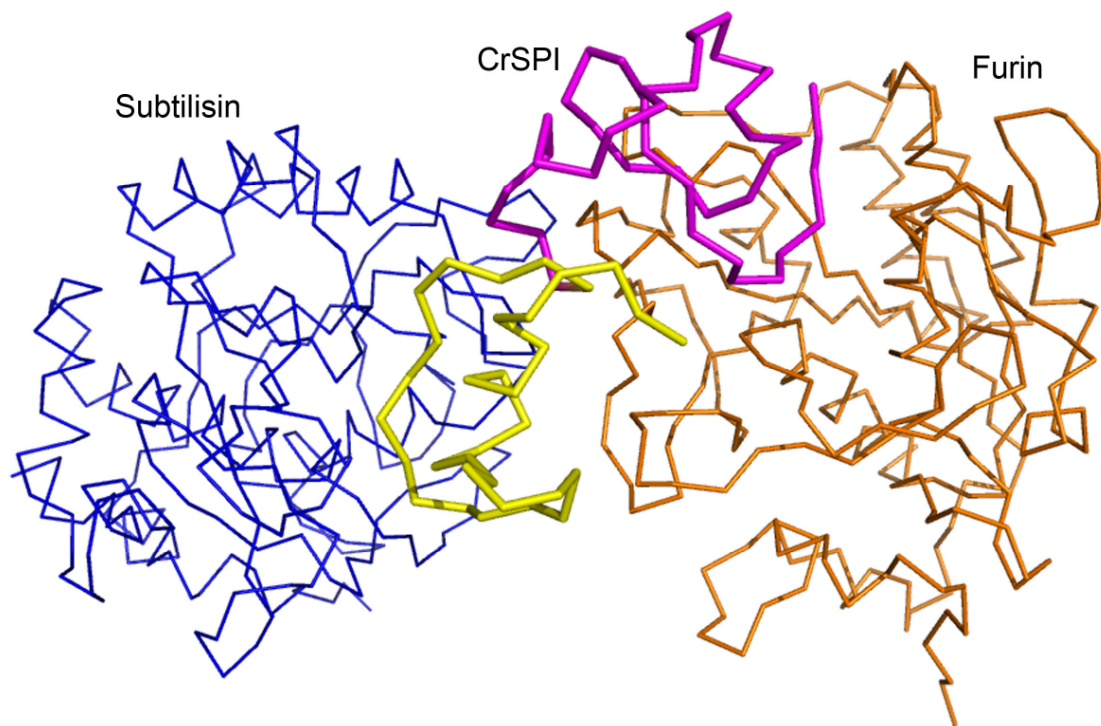
Figure 4.10b. Cα trace for the heterotrimer Furin-CrSPI-Subtilisin complex proposed model.
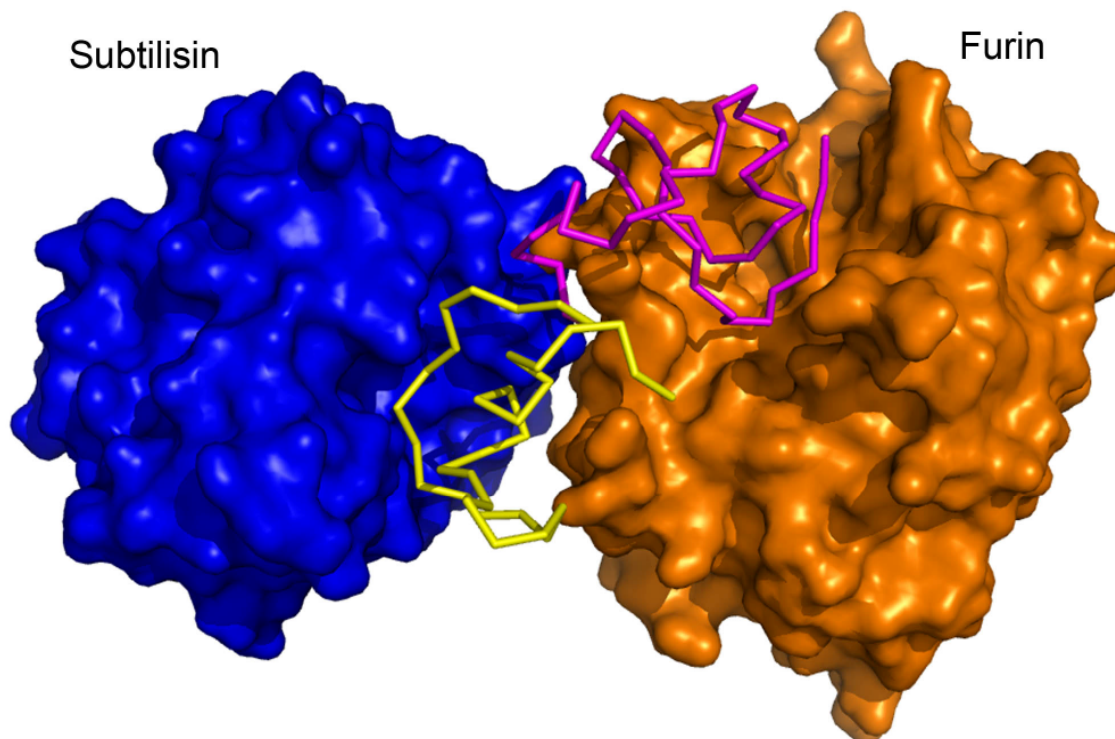
Figure 4.10c. Surface representation for Furin and Subtilisin, and ribbon representation for CrSPI-1 of the heterotrimer model.

Our studies revealed that in comparison to domain-1, the domain-2 comprising of three S-S bridges enhances the rigidity of the RSL to achieve tight interactions with subtilisin. Besides, domain-2 posses the most preferred residue Glu in the P1 position of the Kazal type inhibitors to specifically inhibit subtilisin. While domain-1 has a pair of basic residues to act as a potential substrate-like binding inhibitor for Furin. It is important to note that both RSLs are rigid. This rigid conformation is a prerequisite to resist the proteolytic cleavage of the inhibitor upon interaction with the protease. Thus, CrSPI regulates serine protease driven antimicrobial defense at the acute phase of infection while maintaining homeostasis under native condition (Jiang et al, 2009). CrSPI-1 has been found to inhibit its cognate protease from the host, CrFurin in a dose

dependent inhibitory assay. It interacts with microbial subtilisins as well as a host endogenous C3 complement component, p50 which is required for ingestion and destruction of pathogens. All three proteases are involved in important roles in immune responses and homeostasis. These studies demonstrate that CrSPI-1 serves several immunomodulatory functions by interacting with host and microbial proteases with its independent domains.

Chapter V

Conclusions and Future Directions

5.1 Conclusions

The crystal structures of human cathepsin L complexed with six inhibitors have been thoroughly described in this thesis. Of these four of them are propeptide mimetic and two of them are synthetic dipeptidyl complexes. In addition, their inhibition mechanisms of action were proposed based on these complex crystal structures combined with molecular dynamics analyses (in the case of propeptide mimite inhibitors). These propeptide mimetic inhibitors are very potent inhibitors similar to the full length 96aa propeptide. Similar to propeptides, these propeptide mimetic inhibitors binds in the reverse direction relative to a substrate and span both the S' and S subsites of the cathepsin L active site. These inhibitors have inhibitory constants (Ki) between 20nM and 500nM. This tight binding facilitated the co-crystallization with the cathepsin L.

Further the mode of inhibition of these series of inhibitors with the cathepsin L will provide a general strategy for inhibiting other cathepsins as well as other proteases. Besides, the structural information resulted from this study allowed us to identify the putative specificity determinants of cathepsin L. This site is specifically located on the S' subsite which was not explored before. These vital structural implications will be used for developing highly specific inhibitors of cathepsin L towards drug design for the treatment of human diseases involving cathepsin L.

As a comparison with the propeptide mimetic reverse binding mode inhibitors this thesis also reported two crystal structure complexes of cathepsin L covalently bound inhibitors. These two covalent bound inhibitors bind in the active site in the same direction as substrates. The two inhibitors have identical peptide sequences but different reactive groups. These were chosen to explore the binding mode of these inhibitors. Both

the inhibitors showed identical binding conformations with respect to S2 and S3 pockets whereas there were some differences in the S1 subsite. The glyoxal inhibitor formed a tetrahedral hemithioketal whereas the diazaomethyl ketone inhibitor formed a thioester bond with trigonal geometry. The carbonyl oxygen of the beta aldehyde group forms a hydrogen bond with the catalytic His 163 while the diazomethylketone is found to have its carbonyl oxygen facing the oxyanion hole, forming a hydrogen bond with Gln19. The large differences in the binding mode of the two inhibitors could be attributed to the reactive groups of these inhibitors. The glyoxal inhibitor is reversible due to its hydrolytically labile binding whereas the diazomethylketone is an irreversible inhibitor which has a comparitively slower reaction with the active site thiol. Cathepsin L displays a typical papain-like architecture and closely resembles Cathepsins S and K. The structure of human Cathepsin L in complex with these two inhibitors will be important in the structure-based drug design. These results will lay groundwork for designing the therapeutically useful inhibitors for Cathepsin L as well as Cathepsins S and K which has slightly modified S2 and S3 binding pockets.

Similarly in the case of subtilisin: CrSPI-1 complex, interesting structural and functional features were revealed by the 2.6Å resolution crystal structure. This study showed how such multidomain inhibitors can simultaneously inhibit multiple enzymes within a single ternary complex. In addition based on our structural and biophysical studies we proposed the substrate preference of the two domains of CrSPI-1. I.e. Domain-2 is more potent and specific towards the bacterial protease subtilisin. Domain-1 is likely to interact with the host protease, Furin, acting like an "on-off" switch in the regulation of host's and pathogen's proteases. The structure of CrSPI-1-subtilisin ternary complex will

help to understand the innate immune system at a molecular level. This study on rCrSPI-1: subtilisin heterotrimer complex is the first of this kind for the Kazal-type inhibitors which employs a dual-inhibition mechanism to engage with two protease molecules with different specificity, a hither to unknown mechanism.

5.2 Future directions

1) Cathepsin L: The various Cathepsin structures demonstrate that the difference in function is attributed to the specific structural features. Crystal structures of Cathepsins with sophisticated small-molecule inhibitors will be extremely beneficial in understanding the specificity differences of these enzymes. Detailed structural information of the target enzyme and the binding mode of the inhibitor have been shown to yield more potent and selective inhibitors than combinatorial chemistry techniques. By effectively utilizing the structural information generated through the present study, a series of structure optimized next generation inhibitors will be designed towards drug development. Further the knowledge gained through this study will be expanded to other cathepsin such as cathepsin S and cathepsin K which share similarities in overall structure as well as binding pockets S2 and S3 which are slightly modified.

2) CrSPI-1: As a continuation of our studies in subtilisin : CrSPI-1, the Furin CrSPI-1 complex, particularly the activity of domain-1 of CrSPI-1 with Furin will be investigated to completely elucidate the ON/OFF swtich mechanism involving the heterotrimer. In addition CrSPI-2, the isoform of CrSPI-1, will be structurally characterized. Since CrSPI-2 has specificity for Trypsin, complex formation will be carried out. The differences in the mode of inhibition of

CrSPI-1 and CrSPI-2 will be compared in order to understand the differences in their specificities. Further, along with their cognate protease complexes, the potential role of domain1 in Furin inhibition will be investigated for CrSPI-2. These studies will be beneficial for not only understanding enzyme inhibition but will provide a foundation for developing antipathogenic agents. The involvement of Furin in the entry of several viral pathogens makes it an important target for treatment of several infectious diseases like Anthrax and Pseudomonal infections. The potential role of CrSPI-1 domain 1 in the inhibition of Furin is currently under progress.

In summary, the inhibition mechanism of cysteine and serine protease such as Cathepsin L and subtilisin were studied. Both proteases adopts a similar catalytic triad (example: Subtilisin Asp32-His64-Ser221; Papain Cys25-His159-Asn175) and a similar oxyanion hole. Moreover many of these proteases are secreted as inactive forms called zymogens and subsequently activated by proteolysis. The propeptide mimic inhibitors of Cathepsin L bind non covalently in the active site in a reverse direction. Further, the study on covalent dipeptidyl inhibitors in complex with Cathepsin L uncover how they inhibit both reversibly or irreversibly based on the reactive group of the inhibitor. In addition, the studies on serine protease subtilisin and its natural proteinaceous inhibitor that binds non-covalently , reveal the inhibition by a substrate-like conformation of the inhibitor's reactive site loop. The results presented in this thesis will lead to the structure based optimization of next generation inhibitors for therapeutic purposes.

References

Apostoluk, W. and Otlewski, J. (1998). Variability of the canonical loop conformations in serine proteinases inhibitors and other proteins. *Proteins*, **32**, 459–474.

Armstrong, P.B. (2001). The contribution of proteinase inhibitors to immune defense. *Trends Immunol.* **22**, 47-52.

Barrette-Ng, I.H., Ng, K.K., Cherney, M.M., Pearce, G., Ryan, C.A. and James, M.N. (2003). Structural basis of inhibition revealed by a 1:2 complex of the two-headed tomato inhibitor-II and subtilisin Carlsberg. *J Biol Chem.* **278**, 24062-71.

Basak, A., Ernst, B., Brewer, D., Seidah, N.G., Munzer, J.S., Lazure, C., and Lajoie, G.A. (1997). Histidine-rich human salivary peptides are inhibitors of proprotein convertases furin and PC7 but act as substrates for PC1. *J Pept Res.* **49**, 596-603.

Bock, S. C., Skriver, K., Nielsen, E., Thøgersen, H. C., Wiman, B., Donaldson, V. H., Eddy, R. L., Marrinan, J., Radziejewska, E., Huber, R., Shows, T. B., and Magnusson, S. (1986). Human C1 inhibitor: primary structure, cDNA cloning, and chromosomal localization. *Biochemistry*, **25**, 4292-4301.

Bode, W. and Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem*., **204**,433-451.

Bode, W., Papamokos, E., and Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech Hirudo medicinalis. Structural analysis, subtilisin structure and interface geometry. *Eur.J.Biochem.* **166**, 673-692.

Brunger, A.T., Adams, P.D., Clore, G.M., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L (1998). Crystallography & NMR System (CNS), A new software suite for macromolecular structure determination, *Acta Cryst,***D54**, 905-921.

Buchanan, J.M. (1973). The amidotransferases. Adv *Enzymol Relat Areas Mol Biol.***39**, 91-183.

Carter, P. and Wells, J.A. (1988). Dissecting the catalytic triad of a serine protease. *Nature***, 332**, 564-568.

Cauwberghs, S., De Clercq, P.J., Tinant, P. and Declercq, J.P. (1988). Factors affecting ease of ring formation. The effect of anchoring substitution on the rate of an intramolecular diels-alder reaction with furan-diene. *Tetrahedron Letters,* **29**, 2493-2496.

Cerenius L, Söderhäll K (2004). The prophenoloxidase-activating system in invertebrates. *Immunol Rev.,***198**,116-26.

Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M., Irvine, A.D., Bonafé, J.L., Wilkinson, J., Taïeb, A., Barrandon, Y., Harper, J.I., de Prost, Y. and Hovnanian, A. (2000). Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. *Nat Genet.,* **25**,141-2.

Chowdhury, S.F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Ménard, R., Lefebvre, J., Konishi, Y., Cygler, M., Sulea, T. and Purisima, E.O. (2002). Design of Non-covalent Inhibitors of Human Cathepsin L. From the 96-residue Proregion to Optimized Tripeptides. *J. Med. Chem.,* **45**, 5321-5329.

Chowdhury, S.F., Joseph, L., Kumar, S., Tulsidas, S.R., Bhat, S., Ziomek, E., Ménard, R., Sivaraman, J. and Purisima, E.O. (2008). Exploring inhibitor binding at the S' subsites of cathepsin L. *J. Med. Chem.,* **51**, 1361-1368.

Coulombe, R., Li, Y., Takebe, S., Menard, R., Mason, P., Mort, J.S. and M. Cygler. (1996). Crystallization and preliminary X-ray diffraction studies of human procathepsin L. *Proteins*, **25**,,398–400.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994). Screening and optimization strategies for macromolecular crystal growth . Acta Cryst. **D50**, 414-423.

Davis, R. L. Shrimpton, A. E. Holohan, P. D. Bradshaw, C. Feiglin, D. Collins, G. H. Sonderegger, P. Kinter, J. Becker, L. M. Lacbawan, F. Krasnewich, D. Muenke, M.Lawrence, D. A. Yerby, M. S. Shaw, C. M. Gooptu, B. Elliott, P. R. Finch, J. T., Carrell, R. W. and Lomas, D. A. (1999). Familial dementia caused by polymerization of mutant neuroserpin. *Nature*, **401**, 376-379.

DeLano, W.L. The PyMOL Molecular Graphics System (2002). DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

Delbaere, L.T., and Brayer, G.D. (1985). The 1.8 A structure of the complex between chymostatin and Streptomyces griseus protease A. A model for serine protease catalytic tetrahedral intermediates. *J Mol Biol*. **183**, 89-103.

Ding, J.L. and Ho, B. (2001). A new era in pyrogen testing. *Trends Biotechnol.,* **19**, 277-81.

Ding, J.L., Navas, M.A. and Ho, B. (1993). Two forms of factor C from the amebocytes of Carcinoscorpius rotundicauda: Purification and characterization. *Biochim Biophys Acta*. **1202**, 149-156

Ding, J.L., Wang, L.H. and Ho, B. (2004). Current genome-wide analysis on serine proteases in innate immunity. Current Genomics, **5**, 147-155.

Dodson, G. and Wlodawr, A. (1998). Catalytic triads and their relatives. *Trends Biochem. Sci.*, **23**, 347-352.

Drenth, J., Kalk, K.H. and Swen, H.M. (1976). Binding of chloromethyl ketone substrate analogues to crystalline papain. *Biochemistry,* **15**, 3731–3738.

Dufour, E., Storer, A.C. and Ménard, R. (1995). Peptide aldehydes and nitriles as transition state analog inhibitors of cysteine proteases. *Biochemistry.*, **34**, 9136-43.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst.*. **D60**, 2126-2132.

Engel, J.C., Doyle, P.S. and McKerrow, J.H. (1999) Trypanocidal effect of cysteine protease inhibitors in vitro and in vivo in experimental chagas disease. *Medicina (B Aires),* **59**,171-175.

Esnouf, R. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta *Crystallogr., Sect. D: Biol. Crystallogr.* , **55**, 938–940.

Esser, R.E., Angelo, R.A., Murphey, M.D., Watts, L.M., Thornburg, L.P., Palmer, J.T., Talhouk, J.W., Smith, R.E. (1994). Cysteine proteinase inhibitors decrease articular cartilage and bone destruction in chronic inflammatory arthritis. *Arthritis Rheum,*. **37**, 236-247.

Falgueyret, J.P., Oballa, R.M., Okamoto, O., Wesolowski, G., Aubin, Y., Rydzewski, R.M., Prasit, P., Riendeau, D., Rodan, S.B. and Percival, M.D. (2001). Novel nonpeptidic cyanamides as potent and reversible inhibitors of human cathepsins K and L. *J. Med. Chem.,* **44**, 94-104.

Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Kohler, S. and Gronenborn, A. M. (1989). Solution structure of recombinant hirudin and the Lys-47----Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study. *Biochemistry*, **28**, 2601 -2617.

Fujishima, A., Imai, Y., Nomura, T., Fujisawa, Y., Yamamoto, Y. and Sugawara, T. (1997). The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett,*. **407**, 47-50.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784-3788.

Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*. **15,** 305-8.

Greenspan, P.D., Clark, K.L., Tommasi, R.A., Cowen, S.D., McQuire, L.W., Farley, D.L., van Duzer, J.H., Goldberg, R.L., Zhou, H., Du, Z., Fitt, J.J., Coppa, D.E., Fang, Z., Macchia, W., Zhu, L., Capparelli, M.P., Goldstein, R., Wigg, A.M., Doughty, J.R., Bohacek, R.S. and Knap, A.K. (2001). Identification of dipeptidyl nitriles as potent and selective inhibitors of cathepsin B through structure-based drug design. *J. Med. Chem,* **44**, 4524-4534.

Grubb, A., Abrahamson, M., Olafsson, I., Trojnar, J., Kasprzykowska, R.,Kasprzykowski, F. and Grzonka, Z. (1990). Synthesis of cysteine proteinase inhibitors structurally based on the proteinase interacting N-terminal region of human cystatin C. *Biol. Chem. Hoppe Seyler.,* **371**, 137–144.

Hemmi, H., Kumazaki, T., Yoshizawa-Kumagaye, K., Nishiuchi, Y., Yoshida, T., Ohkubo, T. and Kobayashi, Y. (2005). Structural and Functional Study of an Anemonia Elastase Inhibitor, a "Nonclassical" Kazal-Type Inhibitor from Anemonia sulcata. *Biochemistry,* **44**, 9626-36.

Henrich, S., Cameron, A., Bourenkov, G.P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W. andThan, M.E. (2003). The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat Struct Biol.*, **10**, 520-6.

Hiemstra, P.S. (2002). Novel roles of protease inhibitors in infection and inflammation. Biochem Soc Trans. **30**, 116–120.

Holm, L. and Sander C. (1991). Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors. *J Mol Biol*. **218**, 183-94.

Hooper, N.M. and Lendeckel, U. (2005). The Adam Family of Proteases. Springer ISBN: 0387251499.

Huang, L., Lee, A. and Ellman, J.A. (2002). Identification of potent and selective mechanism-based inhibitors of the cysteine protease cruzain using solid-phase parallel synthesis. *J. Med. Chem*, **45**, 676-684.

Irving, J. A., Steenbakkers, P. J. M., Lesk, A. M., Op den Camp, H. J. M., Pike, R. N. and Whisstock, J. C. (2002). Serpins in prokaryotes, *Mol. Biol. Evol*. **19**, 1881-1890.

Iwanaga, S. and Kawabata, S. (1998). Evolution and phylogeny of defense molecules associated with innate immunity in horseshoe crab. *Front Biosci.* **3**, D973-84.

Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z. and Jaskólski M. (2004). Two polymorphs of a covalent complex between papain and a diazomethylketone inhibitor. *J Pept Res*., **64**,141-50.

Jean, F., Stella, K., Thomas, L., Liu, G., Xiang, Y., Reason, A.J. and Thomas, G. (1998). alpha1-Antitrypsin Portland, a bioengineered serpin highly selective for furin: application as an antipathogenic agent. *Proc Natl Acad Sci U S A*. **95**, 7293-8.

Jiang, N., Thangamani, S., Chor, C.F., Wang, S.Y., Winarsih, I., Du R.J., Sivaraman, J., Ho, B., Ding, J.L. (2009). A Novel Serine Protease Inhibitor Acts as an Immunomodulatory Switch while Maintaining Homeostasis. *J Innate Immun.* **1**, 465-479.

Jiravanichpaisal, P., Lee, B.L. and Söderhäll, K. (2006). Cell-mediated immunity in arthropods: hematopoiesis, coagulation, melanization and opsonization. *Immunobiology*. **211**, 213-36.

Johansson, M.W., Keyser, P. and Söderhäll, K. (1994). Purification and cDNA cloning of a four-domain Kazal proteinase inhibitor from crayfish blood cells. *Eur J Biochem.*, **223**, 389-94.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjelgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.*, **A47**, 110-119.

Joyce, J.A., Baruch, A., Chehade, K., Meyer-Morse, N., Giraudo, E., Tsai, F.Y., Greenbaum, D.C., Hager, J.H., Bogyo, M. and Hanahan, D. (2004). Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell,* **5**, 443-453.

Kakegawa, H., Nikawa, T., Tagami, K., Kamioka, H., Sumitani, K., Kawata, T., Drobnic-Kosorok, M., Lenarcic, B., Turk, V. and N. Katunuma. (1993). *FEBS Lett.* **321**, 247–250.

Kanost, M.R. (1999). Serine proteinase inhibitors in arthropod immunity. Dev *Comp Immunol.* **23**, 291-301.

Katunuma, N., Matsui, A., Inubushi, T., Murata, E., Kakegawa, H., Ohba, Y., Turk, D., Turk, V., Tada, Y. and Asao, T. (2000). Structure-based development of pyridoxal propionate derivatives as specific inhibitors of cathepsin K in vitro and in vivoT, *Biochem. Biophys. Res. Commun*., **267**, 850-854.

Kirschke, H., Wikstrom, P., and Elliott Shaw (1988). Active center differences between cathepsins L and B: The S1 binding region. *FEBS Letters*, **228**, 128-130.

Kleanthous, C. (2000). Protein-protein recognition; Oxford University Press, ISBN 0199637601, 9780199637607.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst*. **24**, 946-950.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291.

Laskowski, M. Jr and Kato, I. (1980). Protein inhibitors of proteinases. *Annu. Rev. Biochem.,* **49**, 593-626.

Leary, R., Larsen, D., Watanabe, H. and Shaw, E. (1977). Diazomethyl ketone substrate derivatives as active-site-directed inhibitors of thiol proteases. Papain. *Biochemistry*., **16**, 5857-61.

Lecaille, F., Kaleta, J., and Brömme, D. (2002). Human and Parasitic Papain-Like Cysteine Proteases: Their Role in Physiology and Pathology and Recent Developments in Inhibitor Design. *Chem. Rev*., **102**, 4459-4488.

Lee, T.W., Qasim, M.A., Laskowski, M. Jr.., James, M.N. (2007). Structural insights into the non-additivity effects in the sequence-to-reactivity algorithm for serine peptidases and their inhibitors. *J Mol Biol*., **23**, 527-546.

Leung-Toung, R., Zhao, Y., Li, W., Tam, T.F., Karimian, K. and Spino M. (2006). Thiol proteases: inhibitors and potential therapeutic targets. *Curr Med Chem*.,**13**,547-81.

Li, W., Kornmark, L., Jonasson, L., Forssell, C. and Yuan, X.M. (2009). Cathepsin L is significantly associated with apoptosis and plaque destabilization in human atherosclerosis. *Atherosclerosis*, **202**, 92-102.

Ljunggren, A., Redzynia, I., Alvarez-Fernandez, M., Abrahamson, M., Mort, J.S., Krupa, J.C., Jaskolski, M., Bujacz, G. (2007). Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease *J.Mol.Biol*., **371**, 137-153.

Lowther, J., Djurdjevic-Pahl, A., Hewage, C., Malthouse, J.P. (2002). A 13C-NMR study of the inhibition of papain by a dipeptide-glyoxal inhibitor. *Biochem J.* **366**, 983-7.

Lu, W., Zhang, W., Molloy, S.S., Thomas, G., Ryan, K., Chiang, Y., Anderson, S. and Laskowski, M. Jr. (1993). Arg15-Lys17-Arg18 turkey ovomucoid third domain inhibits human furin. *J. Biol. Chem.*, **268**, 14583-14585.

Lynas, J.F., Hawthorne, S.J. and Walker, B. (2000). Development of peptidyl alpha-keto-beta-aldehydes as new inhibitors of cathepsin L- comparisons of potency and selectivity profiles with cathepsin B. *Bioorg Med Chem Lett.* **10**, 1771-3.

Mackenzie, N. E., Grant, S. K., Scott, A. I. and Malthouse, J. P. (1986). 13C NMR study of the stereospecificity of the thiohemiacetals formed on inhibition of papain by specific enantiomeric aldehydes. *Biochemistry*, **25**, 2293-2298.

Magert, H.J., Kreutzmann, P., Standker, L., Walden, M., Drogemuller, K. and Forssmann, W.G. (2002) LEKTI: a multidomain serine proteinase inhibitor with pathophysiological relevance. *Int J Biochem Cell Biol*, **34**, 573-576.

Maiti, R., Gary, H., Domselaar, V., Zhang, H. and Wishart, D.S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res*. **32** (Web Server issue): W590W594

Malhotra, S. and Gupta, N. (2002). Childhood disintegrative disorder. Re-examination of the current concept. *J Am Acad Child Adolesc Psych*, **41**, 1239–45.

Marquis, R.W., Ru, Y., LoCastro, S.M., Zeng, J., Yamashita, D.S., Oh, H.J., Erhard, K.F., Davis, L.D., Tomaszek, T.A., Tew, D., Salyers, K., Proksch, J., Ward, K., Smith, B., Levy, M., Cummings, M.D., Haltiwanger, R.C., Trescher, G., Wang, B., Hemling, M.E., Quinn, C.J., Cheng, H.Y., Lin, F., Smith, W.W., Janson, C.A., Zhao, B., McQueney, M.S., D'Alessio, K., Lee, C.P., Marzulli, A., Dodds, R.A., Blake, S.H., Wang, S.M., James, I.E., Gress, C.J., Bradley, B.R., Lark, M.W., Gowen, M. and Veber, D.F. (2001). Azepanone-based inhibitors of human and rat cathepsin K. *J. Med. Chem*, **44**,1380-1395.

Marquis, R.W., Ru, Y., Zeng, J., Trout, R.E., LoCastro, S.M., Gribble, A.D., Witherington, J., Fenwick, A.E., Garnier, B., Tomaszek, T., Tew, D., Hemling, M.E., Quinn, C.J., Smith, W.W., Zhao, B., McQueney, M.S., Janson, C.A., D'Alessio, K., Veber, D.F. (2001). Cyclic ketone inhibitors of the cysteine protease cathepsin K. *J. Med. Chem,* **44**,725-736.

Maynes, J.T., Cherney, M.M., Qasim, M.A., Laskowski, M. Jr. and James, M.N. (2005). Structure of the subtilisin Carlsberg-OMTKY3 complex reveals two different ovomucoid conformations. *Acta Crystallogr D Biol Crystallogr.* **61**, 580-8.

McGrath, M.E. (1999). The Lysosomal Cysteine Proteases. *Annu. Rev. Biophys. Biomol. Structure,* **28**, 181-204.

Ménard, R., Carrière, J., Laflamme, P., Plouffe, C., Khouri, H.E., Vernet, T., Tessier, D.C., Thomas, D.Y. and Storer, A.C. (1991). Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain. *Biochemistry*. **30**, 8924-8928.

Ménard, R., Plouffe, C., Laflamme, P., Vernet, T., Tessier, D.C., Thomas, D.Y. and Storer, A.C. (1995). Modification of the electrostatic environment is tolerated in the oxyanion hole of the cysteine protease papain. *Biochemistry*, **34,** 464-471.

Muta, T. and Iwanaga, S. (1996). The role of hemolymph coagulation in innate immunity. *Curr Opin Immunol*. **8**, 41-7.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst*. **D53**, 240-255.

Nakayama, K. (1998). Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem. J*., **327**, 625–35.

Nirmala, X., Kodrík, D., Zurovec, M. and Sehnal, F. (2001). Insect silk contains both a Kunitz-type and a unique Kazal-type proteinase inhibitor. *Eur J Biochem*. **268**, 2064-73.

Noone, P., Zhou, Z., Silverman, L., Jowell, P., Knowles, M. and Cohn, J. (2001). Cystic fibrosis gene mutations and pancreatitis risk: Relation to epithelial ion transport and trypsin inhibitor gene mutations. *Gastroenterology*, **121**, 1310-1319.

Odum, L., Bundgaard, J. R. and Johnsen, A. H. (1999). A Kazaltype trypsin inhibitor from the protochordate Ciona intestinalis. *Eur. J. Biochem*., **259**, 872-876.

Ortiz, C, Tellier, C., Williams, H., Stolowich, N.J. and Scott, A.I. (1991). Diastereotopic covalent binding of the natural inhibitor leupeptin to trypsin: detection of two interconverting hemiacetals by solution and solid-state NMR spectroscopy. *Biochemistry*, **30**, 10026-34.

Otto, H.H. and Schirmeister, T. (1997). Cysteine Proteases and Their Inhibitors, *Chem. Rev*, **97**, 133-171.

Powers, J.C., Asgian, J.L., Ekici, O.D. and James, K.E. (2002). Irreversible inhibitors of serine, cysteine, and threonine Proteases. *Chem. Rev*, **102**, 4639-4750.

Radisky, E.S. and Koshland, D.E. Jr. (2002). A clogged gutter mechanism for protease inhibitors. *Proc Natl Acad Sci U S A,* **99**, 10316-21.

Rawlings, N.D., Morton, F.R. and Barrett, A.J. (2006). MEROPS: the peptidase database. *Nucleic Acids Res*. **34**, D270-272.

Rawlings, N. D. and Barrett, A. J. (2000). MEROPS: the peptidase database. *Nucleic Acids Res*, **28**, 323-325.

Rawlings, N.D., Morton, F.R., Kok, C.Y., Kong, J. and Barrett, A.J. (2008). MEROPS: the peptidase database. *Nucleic Acids Res*, **36**, 320-325.

RECK--a newly discovered inhibitor of metastasis with prognostic significance in multiple forms of cancer. (2007). *Cancer Metastasis Rev*. **26**, 675-83.

Roberts, N.B. (2006). Human pepsins – their multiplicity, function and role in reflux disease. Aliment Pharmacol Ther., **24**, 2-9.

Saegusa, K., Ishimaru, N., Yanagi, K., Arakaki, R., Ogawa, K., Saito, I., Katunuma ,N. and Hayashi, Y. (2002). Cathepsin S inhibitor prevents autoantigen presentation and autoimmunity. *J. Clin. Invest,* **110**, 361-369.

Schaschke, N., Assfalg-Machleidt, I., Machleidt, W. and Moroder, L. (1998). Substrate/propeptide-derived endo-epoxysuccinyl peptides as highly potent and selective cathepsin B inhibitors. *FEBS Lett*,**421**, 80-82.

Schechteri, I. and Berger, A. (1967). On the size of the active site in proteases. *Biochem.Biophys. Res. Commun*., **27**, 157–162.

Schmitke, J.L., Stern, L.J. and Klibanov, A.M. (1998). Comparison of x-ray crystal structures of an acyl-enzyme intermediate of subtilisin Carlsberg formed in anhydrous acetonitrile and in water. *Proc.Natl.Acad.Sci.USA*., **95**, 12918-12923.

Schroder, E., Phillips, C., Garman, E., Harlos, K. and Crawford, C. (1993). X-ray crystallographic structure of a papain leupeptin complex. *FEBS Lett*., **315**, 38–42.

Schulze, A. J., Huber, R., Bode, W., and Engh, R. A. (1994). Structural aspects of serpin inhibition, *FEBS Lett*, **344**, 117-124.

Shaw, E. (1984). The selective inactivation of thiol proteases in vitro and in vivo. *J. Protein Chem*. **3**, 109−120.

Shaw, E., Mohanty, S., Colic, A., Stoka, V. and Turk, V. (1993). The affinity-labelling of cathepsin S with peptidyl diazomethyl ketones: Comparison with the inhibition of cathepsin L and calpain. *FEBS Lett,* **334**, 340-342.

Sheahan, K., Shuja, S. and Murnane, M.J. (1989). Cysteine protease activities and tumor development in human colorectal carcinoma. *Cancer Res.,* **49**, 3809-3814.

Shi, W., Robinson, H., Sullivan, M., Abel, D., Toomey, J., Berman, L.E.,Lynch, D., Rosenbaum, G., Rakowsky, G., Rock, L., Nolan, B., Shea-McCarthy, G., Schneider, D., Johnson, E., Sweet, R.M., and Chance, M. R. (2006). Beamline X29: a novel undulator source for X-ray crystallography. *J. Synchrotron Rad*., **13**, 365-372.

Somprasong, N., Rimphanitchayakit, V. and Tassanakajon, A. (2006). A five-domain Kazal-type serine proteinase inhibitor from black tiger shrimp Penaeus monodon and its inhibitory activities. *Dev Comp Immunol*, **30**, 998–1008.

Stein, P. E., and Carrell, R. W. (1995). What do dysfunctional serpins tell us about molecular mobility and disease? *Nat. Struct. Biol*., **2**, 96-113.

Stoll, V.S., Eger, B.T., Hynes, R.C., Martichonok, V., Jones, J.B. and Pai, E.F. (1998). Differences in binding modes of enantiomers of 1-acetamido boronic acid based protease inhibitors: crystal structures of gamma-chymotrypsin and subtilisin Carlsberg complexes. *Biochemistry* , **37**, 451-462.

Stroup, G.B., Lark, M.W., Veber, D.F., Bhattacharyya, A., Blake, S., Dare, L.C., Erhard, K.F., Hoffman, S.J., James, I.E., Marquis, R.W., Ru, Y., Vasko-Moser, J.A., Smith, B.R., Tomaszek, T. and Gowen, M. (2001). Potent and selective inhibition of human cathepsin K leads to inhibition of bone resorption in vivo in a nonhuman primate. *J. Bone Miner. Res,* **16**,1739-1746.

Takahashi, H., Nukiwa, T., Yoshimura, K., Quick, C.D., States, D.J., Holmes, M.D., Whang-Peng, J., Knutsen, T. and Crystal, R.G. (1988). Structure of the human neutrophil elastase gene. *J. Biol. Chem*, **263**, 14739–47.

Theopold, U., Schmidt, O., Söderhäll, K. and Dushay, M.S. (2004). Coagulation in arthropods: defence, wound closure and healing. Trends Immunol., **25**, 289-94.

Thomas, G. (2002). Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat Rev Mol Cell Biol.* **3**, 753-66.

Thompson, S.K., Halbert, S.M., Bossard, M.J., Tomaszek, T.A., Levy, M.A., Zhao, B., Smith, W.W., Abdel-Meguid, S.S., Janson, C.A., D¢Alessio, K.J., McQueney, M.S., Amegadzie, B.Y., Hanning, C.R., DesJarlais, R.L., Briand, J., Sarkar, S.K., Huddleston, M.J., Ijames, C.F., Carr, S.A., Garnes, K.T., Shu, A., Heys, J.R., Bradbeer, J., Zembryski, D. and Lee-Rykaczewski, L. (1997). Design of potent and selective human cathepsin K inhibitors that span the active site. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14249-14254.

Thomssen, C., Schmitt, M., Goretzki, L., Oppelt, P., Pache, L., Dettmar, P., Janicke, F. and Graeff, H., (1995). Prognostic value of the cysteine proteases cathepsins B and cathepsin L in human breast cancer. *Clin. Cancer Res.*, **1**, 741–746.

Travis, J. and Salvesen, G.S. (1983). Molecular cloning of human neutrophil elastase. *Annu. Rev. Biochem*, **52**,655-709.

Turk, V., Turk, B. and Turk, D. (2001). Lysosomal cysteine proteases: facts and opportunities. *EMBO J*., **20**, 4629-4633.

Vagin, A. and Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Cryst*., **30**, 1022-1025.

van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kröger, B., Höffken, W. and Bode, W. (1995). Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J*. **14,**5149-57.

Voet, D. and Voet, J.G. (1990). Biochemistry. John Wiley & Sons Inc. (Canada): 373-382.

Votta, B.J., Levy, M.A., Badger, A., Bradbeer, J., Dodds, R.A., James, I.E., Thompson, S., Bosshard, M.J, Carr, T., Connor, J.R., Tomaszek, T.A., Szewczuk, L., Drake, F.H., Veber, D.F. and Gowen, M. (1997) Peptide aldehyde inhibitors of cathepsin K inhibit bone resorption both in vitro and in vivo. *J. Bone Miner. Res*, **12**,1396-1406.

Walker, B., Lynas, J.F., Meighan, M.A. and Brömme, D. (2000). Evaluation of dipeptide alpha-keto-beta-aldehydes as new inhibitors of cathepsin S. *Biochem Biophys Res Commun.* **275**, 401-5.

Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.

Wiman, B.; Collen, D. (1978). On the kinetics of the reaction between human antiplasmin and plasmin. *Eur. J. Biochem*, **84**, 573.

Wolberg, A.S. (2007). Thrombin generation and fibrin clot structure. *Blood Rev*, **21**, 131–142.

Yamamoto, M., Ikeda, S., Kondo, H. and Inoue, S., (2002). Design and synthesis of dual inhibitors for matrix metalloproteinase and cathepsin. *Bioorg, Med. Chem. Lett*. **12**, 375-378.

Yasuma, T., Oi, S., Choh, N., Nomura, T., Furuyama, N., Nishimura, A., Fujisawa, Y. and Sohda, T. (1998). Synthesis of peptide aldehyde derivatives as selective inhibitors of human cathepsin L and their inhibitory effect on bone resorption. *J. Med. Chem*., **41**, 4301-4308.

Yuan, J., and Horvitz, H. R. (2004). A First Insight into the Molecular Mechanisms of Apoptosis. *Cell*, **116**, 53–56.