UNIVERSITY OF ALBERTA

STRUCTURE AND DESIGN OF PROTEINASES AND THEIR INHIBITORS

by 

ISABELLE BARRETTE-NG

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

FALL, 2003

Your file Votre référence
ISBN: 0-612-87935-6
Our file Notre référence
ISBN: 0-612-87935-6

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR:	Isabelle H. Barrette-Ng
TITLE OF THESIS:	Structure and design of proteinases and their inhibitors
DEGREE:	Doctor of Philosophy
YEAR THIS DEGREE GRANTED:	2003

Date: June 23, 2003

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled STRUCTURE AND DESIGN OF PROTEINASES AND THEIR INHIBITORS submitted by ISABELLE BARRETTE-NG in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

Dr. Michael N.G. James

Dr. Brian D. Sykes

Dr. Mark Glover

Dr. R. Chris Bleackley

Dr. David R. Bundle

Dr. Hazel M. Holden

Date: June 20, 2003

To my husband, Ken

ABSTRACT

Proteolytic enzymes perform many physiologically important functions by catalyzing the cleavage of peptide bonds. These enzymes can be classified according to catalytic type: aspartic proteinases, cysteine proteinases, metalloproteinases, serine proteinases, threonine proteinases and proteinases of unknown catalytic mechanism.

The serine proteinases have a nucleophilic serine residue at their active site and are widely distributed in vertebrates, bacteria and viruses. Nonstructural protein 4 (nsp4) from the *Equine Arteritis Virus* is a serine proteinase responsible for most of the proteolytic processing seen in the viral polyprotein. The three-dimensional structure of the 21-kDa nsp4 has been determined to 2.0-Å resolution. Nsp4 adopts the smallest known chymotrypsin-like fold, with a catalytic triad of Ser-120, His-39 and Asp-65, as well as a novel α/β C-terminal extension domain that may mediate protein-protein interactions. In four independent copies of nsp4, the oxyanion hole adopts either a collapsed inactive conformation or the standard active conformation, which may be a novel way of regulating proteolytic activity.

In many physiological processes, the activity of proteolytic enzymes is regulated by proteinaceous proteinase inhibitors (PIs). PIs have been isolated from many different organisms, including most commonly plant seeds and leaves as well as bird eggs and parasites. Interestingly, PIs isolated from plant seeds and leaves play a central role in plant defence mechanisms against attacking pests, whereas PIs isolated from parasites are involved in protecting these organisms against the host's proteinase defence mechanisms.

Many different proteinaceous serine proteinase inhibitors from plants have now been identified. The structure of the two-headed tomato inhibitor-II (TI-II) in complex

with two molecules of subtilisin Carlsberg has been solved to 2.5 Å resolution; it reveals how a multidomain inhibitor from the Potato II family can bind to and simultaneously inhibit two enzyme molecules. The two domains of TI-II adopt a similar fold and are arranged in an extended configuration that presents two reactive site loops at the opposite ends of the inhibitor. Each subtilisin molecule interacts with a reactive site loop through the standard, canonical binding mode. The structure of the unbound form of TI-II has also been solved to 2.15 Å resolution, revealing conformational flexibility in the absence of proteinase molecules. Four independent copies of unbound TI-II display a range of conformations when compared to the bound form, most strikingly in the orientations of the inhibitory domains and the conformations of the reactive site loops. The conformational flexibility of the reactive site loops suggests a mechanism by which the need for tight binding is balanced against the need for broad inhibitory specificity.

Contrary to proteinaceous serine proteinase inhibitors from plants, relatively few proteinaceous inhibitors have been identified for aspartic proteinases. One example is pepsin inhibitor-3 (PI-3) from the nematode *Ascaris*. A single-domain inhibitor was designed based on the structure of the two-domain PI-3 from *Ascaris suum*. A model for the single-domain inhibitor indicated that the sequence could be expected to form a stable structure in the absence of residues 11-69. A synthetic gene has been constructed and the one-domain inhibitor has been expressed at high levels in an insoluble form in *E. coli*. The insoluble inhibitor could be renatured, yielding an inhibitor of porcine pepsin with K_i of 0.5 μM. Inhibitory activity is retained in the absence of the N-terminal domain but the potency of the inhibitor appears to be dramatically reduced.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. Michael James, for his help, support and guidance, and for sharing his wealth of knowledge and enthusiasm about protein structure. The support of all of the members of the James laboratory is sincerely appreciated, most notably from Maia Cherney, Masao Fujinaga, Sheraz Khan and Brian Mark who taught me a lot about protein crystallization, data collection and refinement. I also thank Craig Garen, Roland Dargis and Dr. Smillie for their help and patience in helping me with protein purification.

I also wish to thank our collaborators, Danny van Aken and Dr. Eric Snijder from Leiden University Medical Center, and Gregory Pearce and Dr. Clarence A. Ryan from Washington State University, without whom some of this work would not have been possible.

I also want to thank my parents, André and Louise Barrette, who have always encouraged me to follow my dreams. Finally, I would also like to thank my husband, Ken, for his support, help and encouragement over the past three years.

The members of my supervisory committee, Drs. Mark Glover, Brian Sykes, Chris Bleackley and David Bundle, are gratefully acknowledged for providing constructive criticism and insight into the work presented in this thesis. I wish to particularly thank Dr. Hazel Holden for traveling from the University of Wiscosin at Madison to be a member of my thesis examining committee.

Data used to determine the structure of nonstructural protein 4 (NSP4) were collected at the Advanced Photon Source. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science.

Use of BioCARS Sector 14 was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707.

Data used to determine the structure of the bound and unbound forms of tomato inhibitor-II were collected at the Stanford Synchrotron Radiation Laboratory, national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

This research has been supported by the Canadian Institutes of Health Research, through a grant to Dr. Michael N.G. James. I have been supported as a graduate student by studentships from the Canadian Institutes of Health Research, the Alberta Heritage Foundation for Medical Research and the University of Alberta.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES OR ILLUSTRATIONS

LIST OF SYMBOLS, NOMENCLATURE, OR ABBREVIATIONS

Å	angstroms (10^{-10} metres)
CLP	chymotrypsin-like proteinase
EAV	equine arteritis virus
$\|F_c\|$	Calculated structure factor amplitude
$\|F_o\|$	Observed structure factor amplitude
HRV-2	Human Rhinovirus-2
IPTG	isopropyl β-D-galactoside
k_{cat}	First-order catalytic rate constant
K_m	Michaelis Menten constant
LB	Luria broth
M	moles/litre
MBP	maltose-binding protein
mPI3Δ1	single-domain mutant of PI-3 containing only the C-terminal domain
Ni-NTA	nickel-nitrilotriacetic acid agarose resin
NMR	nuclear magnetic resonance
nsp	nonstructural protein
OD	optical density
ORF	open reading frame
PEG MME	polyethylene glycol monomethyl ether
PI	proteinase inhibitor
PI-II	potato proteinase inhibitor-II
Pot II	Potato II inhibitor

PRRSV	porcine reproductive and respiratory syndrome virus
r.m.s.d	root mean squared difference
rpm	rotations per minute
SCP	*Sindbis virus* core protein
SDS	sodium dodecyl sulphate
SFCP	*Semliki forest virus* core protein
SGPE	*Streptomyces griseus* proteinase E
TFA	trifluoroacetic acid
TI-II	tomato proteinase inhibitor-II

CHAPTER 1:

INTRODUCTION

1.0 Proteinases and their inhibitors

Proteolytic enzymes are ubiquitous and comprise approximately 2% of the genes in organisms with sequenced genomes (1). They are divided into groups by catalytic type in the MEROPS database (2) as follows: aspartic proteinases, cysteine proteinases, metalloproteinases, serine proteinases, threonine proteinases and proteinases of unknown catalytic mechanism. These enzymes perform many different physiological functions of vital importance by catalyzing the cleavage of peptide bonds in proteins and peptide substrates (3,4). Apart from hydrolyzing proteins in the digestive systems of many organisms, some of the other physiological functions in which proteolytic enzymes play important roles include blood coagulation and fibrinolysis (4-6), activation of prohormones and zymogens (4), macromolecular assembly of viruses and virus replication (4), mammalian fertilization (7) and tumour invasion (8).

Inherent to their participation in such vital physiological functions is the need for tight regulation of proteolytic enzymes. Many serious diseases such as emphysema (9) and arthritis (10) have now been linked to defects in the regulation of proteolytic enzymes. Control of these enzymes is normally achieved through regulation of transcription and translation rates, sub-cellular targeting, activation of zymogens, degradation of mature enzymes, and inhibition of proteolytic activity by naturally-occurring proteinase inhibitors (PIs) (4).

Naturally-occurring PIs have received a great deal of attention in the past few decades to better understand their mechanism of action and use them as possible drug

candidates. With the exception of small non-proteinaceous inhibitors from some microorganisms, all naturally-occurring PIs are proteins and have been isolated from the cells, tissues and fluids of a variety of organisms, including most commonly plant seeds and leaves as well as bird eggs and parasites (4,11). Interestingly, PIs isolated from plant seeds and leaves have been implicated in plant defence mechanisms against attacking pests whereas PIs isolated from parasites are involved in protecting these organisms against the host's defence mechanisms (11-17).

As structural and biochemical studies have now revealed, PIs come in many different three-dimensional structures and can range in size from small peptides to proteins larger than their target proteinases. PIs also adopt quite different mechanisms of inhibition which include blocking the proteinase's active site in a substrate-like manner, binding to the proteinase in an area adjacent to the active or substrate-binding site, and allosteric inhibition of the proteinase through binding to distantly located exosites on the target proteinase (18) (Figure 1-1).

This thesis will concentrate on several common classes of proteinases and their inhibitors. Firstly, the class of serine proteinases will be discussed, especially as they relate to the proteinases involved in viral replication. Secondly, serine proteinase inhibitors will be examined with particular emphasis on PIs involved in plant defence mechanisms. Thirdly, the structure and mechanisms of PIs of aspartic proteinases will be described with emphasis on PIs implicated in parasitic defence mechanisms.

1.1 Serine proteinases

Approximately one-third of all currently known proteinases can be classified as serine proteinases that are characterized as having a nucleophilic serine residue at the

active site (19). The catalytic mechanism and three-dimensional structure of a serine proteinase were first determined for chymotrypsin (20). Although chymotrypsin-like serine proteinases comprise a major class of serine proteinases, several structurally distinct classes are also known. Remarkably, despite a wide variation in the overall folds of the chymotrypsin, subtilisin and serine carboxypeptidase classes of serine proteinases, the arrangement of catalytic serine and histidine residues at the active site as well as the catalytic mechanism are conserved. The chymotrypsin family is characterized by the presence of two "Greek-key"-barrel domains with the active site residues lying at the cleft formed by the junction of the two domains (20). In the subtilisin family, proteinases adopt an α/β-barrel fold consisting of nine α-helices packed against a seven-stranded parallel β-sheet, with the active site residues lying at the top of the barrel (21). The serine carboxypeptidases adopt a mixed α/β fold in which the active site residues lie at the base of a hemispherical pit formed at the junction of two domains (22).

Chymotrypsin-like proteinases are the most commonly found serine proteinases, with over 240 members found in vertebrates, bacteria and viruses (2). In most chymotrypsin-like proteinases, the active site serine residue is connected by hydrogen bonding to a histidine that is in turn hydrogen-bonded to an aspartic acid residue in a catalytic triad. However, several variations on the arrangement of active site residues have been found (23). In the viral 3C proteinases, for example, the active site serine residue is replaced by cysteine, and in some members, the aspartic acid residue is either missing or has been replaced by another residue (*e.g.*, glutamic acid) (24,25).

1.1.1 Catalytic mechanism of chymotrypsin-like serine proteinases

The catalytic mechanism of the chymotrypsin-like serine proteinases is one of the most thoroughly characterized enzymatic mechanisms (19). The mechanism can be broken down into a series of steps involving substrate binding, nucleophilic attack, the formation of an acyl-enzyme intermediate, deacylation and release of products (Figure 1-2).

First, an enzyme-substrate complex is formed in which the polypeptide backbone of the substrate forms segments of antiparallel β-sheet with the polypeptide backbone of the enzyme and in which substrate side chains interact with specificity pockets on the enzyme. According to the nomenclature of Schechter and Berger (26), where P1 is the residue N-terminal to the scissile bond and P1' is the residue C-terminal to the scissile bond, residues P4 to P2' of the substrate almost always interact with binding pockets S4 to S2' on the proteinase (Figure 1-3).

Although multiple specificity pockets contribute to binding specificity, the most important determinant of specificity appears to be interactions between the P1 residue and the S1 pocket. Trypsin-like proteinases will interact with substrates that have Lys or Arg at the P1 position, whereas chymotrypsin-like proteinases will recognize Tyr, Phe, Leu or Met at this position. Trypsin's preference for basic residues can be attributed to the formation of a salt bridge between the basic P1 residue of the inhibitor with the acidic Asp189 at the base of the S1 pocket. In chymotrypsin, Ser189 lies at the base of the S1 pocket and is implicated in this proteinase's preference for aromatic residues. The phi-psi backbone angles at residue 216 display characteristic values among trypsins (φ = 176°, ψ = -153°) and chymotrypsins (φ = 173°, ψ = 167°), and shape their S1 pockets

(27). As observed in high-resolution X-ray crystallographic studies of these proteinases (28,29), the size of the S1 pocket is narrower in trypsin than in chymotrypsin and is in part due to the deletion of residue 218 which results in a conformational change of residues 216 to 226.

The extensive set of interactions between proteinases and their substrates not only provides a high degree of binding specificity, but also positions the substrate in the appropriate orientation for catalysis. For catalysis to occur, the carbonyl carbon of the scissile peptide bond must be located near the hydroxyl group of the active site serine (Ser195 in chymotrypsin) to allow for nucleophilic attack. In addition, the backbone amide groups of Ser195 and Gly193 (oxyanion hole) form hydrogen bonds with the carbonyl oxygen of the P1 residue and promote the polarization of the C=O double bond.

As nucleophilic attack occurs, a proton is transferred from Ser195 to His57, and the resulting positively charged histidine side chain is stabilized by a salt bridge with the carboxylate of Asp102. In addition, the buildup of negative charge on the carbonyl oxygen of the P1 residue is further stabilized by the hydrogen-bond donors of the oxyanion hole as the C=O double bond becomes a single bond. Once nucleophilic attack has occurred, a covalent bond is formed between Ser195 $O\gamma$ and the carbonyl carbon of the P1 residue, thus forming a tetrahedral intermediate.

Following the formation of the tetrahedral intermediate, the scissile bond is cleaved as His57 donates a proton to the leaving group nitrogen of the P1' residue and the C=O double bond of the P1 residue is reformed. This results in the formation of an acyl-enzyme intermediate and the release of the polypeptide segment C-terminal to the scissile bond (30).

After the release of the leaving group, a water molecule enters the active site and forms a hydrogen bond to His57. This water molecule donates a proton to His57 and the negatively charged hydroxide ion attacks the carbonyl carbon of the P1 residue to form the second tetrahedral intermediate (31). As in the first tetrahedral intermediate, the buildup of negative charge on the carbonyl oxygen of the P1 residue is stabilized by the oxyanion hole. The tetrahedral intermediate breaks down as a proton is donated from His57 to Ser195 and the C=O double bond of the P1 residue is reformed.

Although the reaction mechanism of serine proteinases has been most thoroughly characterized in chymotrypsin-like enzymes, similarities in the structures of the active sites of other classes of serine proteinases suggest that the basic catalytic mechanism is highly conserved. For example, even though the folds of subtilisin and chymotrypsin differ dramatically, the binding of substrates and the geometry of active site residues are highly similar. Moreover, despite wide variations in the detailed structures of bacterial, viral, and vertebrate chymotrypsin-like enzymes, the structures of substrate binding pockets and catalytic residues are remarkably conserved.

1.1.2 *Equine arteritis virus*

The arteriviruses *Porcine Reproductive and Respiratory Syndrome Virus* (*PRRSV*) (32) and *Equine Arteritis Virus* (*EAV*) (33) are major economic concerns for the swine- and horse-breeding industries worldwide. Both viruses are widespread in the population, can establish persistent infections, and are easily transmitted via both respiratory and venereal routes. *PRRSV* is currently considered to be a major swine pathogen, causing reproductive failure and severe pneumonia in neonates. Horses infected with *EAV* are often asymptomatic, but persistently infected stallions ("shedding stallions") can infect

mares through semen, regularly leading to spontaneous abortion of the fetus (34). Apart from vaccinations that are costly and – given the use of live attenuated vaccines – not without risk (35), there are currently no effective methods to combat arteriviral diseases.

The family *Arteriviridae* (order *Nidovirales*) comprises enveloped, mammalian RNA viruses with a 12 to 16-kilobase positive-stranded genome (36-38). In addition to the prototype *EAV* and *PRRSV*, the family includes the *lactate dehydrogenase-elevating virus* of mice, and *simian hemorrhagic fever virus*. The arterivirus virion consists of a membrane encapsulated sphere of 40-60 nm in diameter and an icosahedral nucleocapsid comprising the viral genomic RNA and protein N (Figure 1-4) (36).

Nidovirus genomes are polycistronic, containing a large 5'-terminal replicase gene, which is expressed from the viral genome RNA, and a downstream set of (largely) structural protein genes. The latter are expressed via the transcription of a nested set of subgenomic mRNAs from the 3'-terminal region of the genome (39). The replicase gene is translated into two multidomain precursor proteins, that are cleaved into mature nonstructural proteins by multiple viral proteinases, a key regulatory mechanism in the nidovirus life cycle (38).

In the arterivirus prototype *EAV*, the replication cycle takes place entirely in the cytoplasm of the infected cell (36) (Figure 1-5). The replicase gene is translated into open reading frame 1a (ORF1a) and ORF1ab polyproteins of 1,727 and 3,175 amino acids, respectively, the latter product resulting from a ribosomal frameshift that can occur just prior to termination of ORF1a translation. These polyproteins are then proteolytically processed by three *EAV* proteinases, of which the main proteinase is nonstructural protein 4 (nsp4) (40).

1.1.3 *EAV* nsp4 chymotrypsin-like serine proteinase

The 21-kDa nsp4 is the *EAV* main proteinase and it has been proposed, based on sequence analysis (38,40), that it adopts a chymotrypsin-like fold and is a member of the 3C-like serine proteinases (40). 3C-like serine proteinases contain the catalytic triad of classic chymotrypsin-like proteinases and the substrate specificity found in the picornavirus 3C cysteine proteinases (36,40). Based on sequence analysis (41) and site-directed mutagenesis results (40), it was predicted that the catalytic triad of nsp4 would contain His39, Asp65 and Ser120. Mutational analyses of cleavage sites reveal that nsp4 possesses a cleavage specificity for (Glu/Gln)-(Gly/Ser) (40,42,43).

As illustrated in Figure 1-6, the *EAV* ORF1a and ORFab proteins are cleaved by three different ORF1a-encoded proteinases: papain-like cysteine proteinases located in nonstructural protein 1 (nsp1) and nsp2, and the chymotrypsin-like serine proteinase nsp4 (41). Following the rapid autocatalytic release of nsp1 and nsp2, the remainder of the polyproteins (nsp3-8 and nsp3-12) is processed by nsp4 (44). Probably, internal hydrophobic domains target these proteins to intracellular membranes, which are modified to accommodate viral RNA synthesis (45).

Nsp4 processes the remaining eight cleavage sites, five in the ORF1a protein (40,43) and three in ORF1b-encoded part of the ORF1ab protein (42), which encodes (among other functions) the viral RNA-dependent RNA polymerase and helicase (41). Alternative processing pathways are used during cleavage of the nsp3-8 intermediate (43). In the major pathway, nsp2 associates with nsp3-8 as a cofactor and triggers nsp4 to autocatalytically cleave its C terminus to yield the nsp3-4 and nsp5-8 intermediates. The relatively slow processing of the nsp3|4 and nsp7|8 sites ensues (44), but the nsp5|6

and nsp6|7 bonds are not cleaved, and cleaved nsp6 and nsp7 are not produced. In the minor pathway, nsp2 does not associate with nsp3-8, and the nsp4|5 junction is not cleaved (43). However, nsp4 does cleave the nsp3|4, nsp5|6, nsp6|7, and nsp7|8 sites of the nsp3-8 intermediate, indicating that the nsp4 proteinase is fully functional.

To provide a structural framework for understanding the complexities of proteolysis in arteriviruses, the crystallographic structure of *EAV* nsp4 was determined at 2.0 Å resolution (Chapter 2). The structure reveals a molecule having the smallest chymotrypsin-like fold known as well as a novel C-terminal α/β extension domain. This first structure of an arteriviral proteinase opens new avenues for understanding replicase maturation and establishes a basis for rationally developing antiviral agents.

1.2 Standard mechanism, canonical proteinase inhibitors of serine proteinases

The most commonly found and studied PIs bind to target proteinases in a substrate-like manner and are known as standard mechanism (12), canonical (46) serine PIs. To date, 18 different families of this class of inhibitor have been identified (47). On the surface of these PIs lies a 'reactive site loop' (48) which binds to the active site of a target proteinase in a substrate-like manner (12). According to the nomenclature of Schechter and Berger (26), where P1 is the residue N-terminal to the scissile bond and P1' is the residue C-terminal to the scissile bond, residues P4 to P2' of the reactive site loop of each inhibitor almost always interact with binding pockets S4 to S2' on the proteinase (Figure 1-3). The reactive site loop adopts an extended conformation and forms a distorted antiparallel β-strand pair with residues near the active site of the proteinase (49). This is contrary to another class of PIs, the non-canonical PIs, which

also bind to proteinases in a substrate-like manner, but form a parallel β-strand pair with residues near the active site of the proteinase (4,18,46).

The conformation of the reactive site loop is remarkably well-conserved in all PIs even though each inhibitor family has its own unique global three-dimensional structure (or 'scaffold' of the inhibitor) (46). The reactive site loop is fairly rigid, is often flanked by one or two disulfide bridges, and is involved in hydrogen bonding to the inhibitor scaffold (46). The rigid conformation of the reactive site loop is thought to prevent loss of conformational freedom and lead to tight binding (50,51). In fact, kinetic studies of the interaction between inhibitor and proteinase have shown that the k_{cat}/K_m value for peptide hydrolysis is higher than in enzyme:substrate interactions, while k_{cat} and K_m values are unusually low, thus resulting in tight binding and slow hydrolysis (12,52).

The interaction between serine proteinases and canonical, standard mechanism serine proteinase inhibitors can be summarized in its simplest form as follows (46,47,51,52):

$$E + I \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} E \cdot I \underset{k_{on}^*}{\overset{k_{off}^*}{\rightleftharpoons}} E + I^*$$

where E is the enzyme, I is the virgin inhibitor, I* is the inhibitor with its scissile bond hydrolyzed, and E·I is the stable enzyme-inhibitor complex. The amount of stable enzyme-inhibitor complex (E·I) at equilibrium can be calculated as follows (51,52):

$$K_{assoc} = [E \cdot I]/([E][I + I^*])$$

A good serine proteinase inhibitor will possess a high value of K_{assoc} so that the concentration of free enzyme is minimized (51). The equilibrium constant describing the hydrolysis of the inhibitor can be calculated as follows (51,52):

$$K_{hyd} = [I^*]/[I]$$

This equilibrium constant has been shown to be on the order of unity at physiological pH for most enzyme:inhibitor complexes and is thought to result from rigidity in the reactive site loop which prevents relaxation upon hydrolysis of the scissile bond (4,12,46,51-53). Finally, the equilibrium constant for the association of inhibitor with proteinase can be expressed as (51,52):

$$K_a = [E{\cdot}I]/([E][I])$$

As for the other equilibrium constants, a large value for K_a describes a good inhibitor.

To provide a molecular structural explanation for the tight binding and high specificity of proteinase inhibitors binding to cognate proteinases, many crystallographic and kinetic studies have been undertaken (4,12,46,47,52), revealing various mechanisms in different proteinase inhibitor families. This thesis presents, in Chapters 3 and 4, structural studies of the two-headed tomato inhibitor-II of the Potato II family of inhibitors both in the free and complexed forms to provide a molecular structural understanding of multidomain Pot II inhibitors.

1.2.1 Potato II family of proteinase inhibitors

Proteinaceous serine proteinase inhibitors (PIs) from plants were first identified nearly 65 years ago (54) and are now known to be major constituents of seeds, tubers and leaves of members of the *Solanaceae* and *Leguminosae* families (5-15% of the total protein) (55-57). These PIs are an integral part of the constitutive and inducible defensive mechanisms that protect plants from attacking pests (bacteria, fungi and insects) (13,58,59). These defensive mechanisms involve the systemic synthesis of serine PIs that accumulate in distal tissue and can inhibit the digestive trypsin- and chymotrypsin-like enzymes of insects and other related serine proteinases of plant

pathogens (60,61). The inhibitory properties towards serine proteinases of these PIs have already been exploited with varying degrees of success for the production of transgenic plants overexpressing the PIs in an attempt to control pests (61-65). However, a greater understanding of the molecular mechanism of inhibition of these PIs with pest proteinases is required at the structural level to fully harness the potential benefits of these natural PIs to crop protection.

PIs of the Potato II (Pot II) inhibitor family have been isolated from wounded tomato and tobacco leaves (66,67), green tomatoes (68), potato tubers (69,70), eggplant fruits (71), paprika seeds (72) and ornamental tobacco flower stigma (73). Pot II PIs can inhibit trypsin, chymotrypsin, subtilisin, oryzin and elastase (66,74), and accumulate systemically in plant tissue as a result of wounding or pest attack. The systemic response to attack in the *Solanaceae* family has been attributed to a complex signaling cascade that is initiated by the binding of systemin (75) to a cell-surface receptor (76) and leads to the release of linolenic acid which is then converted to 12 oxophytodienoic acid and jasmonic acid (77). The release of jasmonic acid leads to the activation of several signaling pathways that in turn lead to the production of more jasmonic acid, H_2O_2 and the synthesis of PIs (78). Within 48 hours of insect attack or wounding, PIs can accumulate to levels of 2% or more of the total soluble protein in the leaves of tomato, potato and alfalfa plants (56,79) and are thought to have adverse effects on the digestive physiology of insects (80). The wide distribution and inducible expression of Pot II PIs in plants strongly suggest the fundamental importance of these proteins to the pest defence strategies of many commercially important crops.

Sequence analyses of various Pot II family PIs have revealed that most are composed of two (62,66,69,81,82), three (83,84), four or six (73) repeats. Tomato and potato proteinase inhibitor II (TI-II and PI-II, respectively) each contain two copies of a 54 amino acid repeat, whereas the *N. alata* PI is a 40.3-kDa precursor consisting of six repeated elements that are proteolytically processed to yield six different PIs (C1 and C2 which are chymotrypsin inhibitors, and T1-T4 which are trypsin and chymotrypsin inhibitors) (73,85,86). High sequence identity exists among the Pot II PI sequences, although interesting variations (mostly in the reactive site loops) occur among different species and among different isoforms of the same species (Figure 1-7). These differences in sequence can give rise to a wide range of differences in cognate proteinase specificities. The expression of multiple isoforms of Pot II PIs with differing target proteinase specificities may be particularly important for protecting plants against predators with an arsenal of proteinases having a wide range of substrate specificity. Understanding the structural basis of PI inhibitory specificity is clearly crucial for understanding how variations in PI sequence give rise to different inhibitory specificities.

1.2.2 Structural and functional characterization of Pot II family

Few three-dimensional structures are currently available for members of the Potato II inhibitor family and there is currently only one structure of a complex with a bound proteinase. In 1989, the structure of one of the two domains of PI-II, the chymotrypsin-binding domain (known as PCI-1), was solved to 2.1 Å by X-ray crystallography (87) in complex with *Streptomyces griseus* proteinase B. Remarkably, this structure revealed that the sequence of PCI-I corresponds to a region that lies across the two repeats of inhibitor II instead of being contained within one repeat. As in other

inhibitors, the reactive site loop in PCI-I adopts the canonical conformation (46) and is constrained by disulphide bridges Cys3-Cys40 and Cys7-Cys36 which are invariant in the Pot II family (Figure 1-7). Based on the structure of PCI-1, a model for the arrangement of the two domains was proposed, in which a pseudo 2-fold rotation axis relates the two domains across a continuous six-stranded antiparallel β-sheet (87).

The NMR structures of single domains of *N. alata* have also been determined (86,88,89). In these single-domain inhibitors, the scaffold of the protein is similar to that seen in PCI-1 and the reactive site loops adopt the same canonical conformation with cysteines involved in disulphide bridge formation at P3 and P2'.

Structural information on other families of PIs is mostly restricted to single-domain inhibitors, even though many natural inhibitors have multiple domains. Structures for a multi-domain PI bound to proteinases have only been determined for the two-headed Bowman-Birk inhibitors from mung bean (90), soybean (91) and wheat germ (92) bound to trypsin. The Bowman-Birk inhibitors consist of two domains of roughly 60 amino acids that are related by a nearly perfect two-fold rotational symmetry axis. Each inhibitory domain binds to and inhibits a separate proteinase molecule in a very similar manner through a reactive-site loop in the "standard, canonical" mechanism.

1.2.3 Tomato inhibitor-II

TI-II is a 13.5-kDa potent inhibitor of subtilisin (average K_i = 9 nM) that possesses a remarkable dual specificity towards chymotrypsin (average K_i = 30 nM) and trypsin (average K_i = 80 nM) (66). This inhibitor and other Pot II PIs belong to the class of "standard mechanism, canonical" proteinaceous inhibitors of serine proteinases (47). There is high sequence homology between TI-II and PI-II (86% sequence identity) and

both inhibitors contain two repeats, each containing one reactive site loop located near the N-terminus. Depending on the species, the amino-terminal repeat contains the trypsin-specific region, and the carboxy-terminal repeat contains the chymotrypsin-specific region. In the TI-II isoform (from *Lycopersicon esculentum* cv. Bonny Best) used for our structural studies, the reactive site loop of the first repeat inhibits trypsin (residues P2-P1-P1' are TRE) whereas that of the second repeat inhibits chymotrypsin (residues P2-P1-P1' are TFN). This thesis presents, in Chapter 3, the three-dimensional X-ray crystallographic structure of a ternary complex of the full-length two-headed inhibitor TI-II, bound to two molecules of subtilisin Carlsberg which was solved to 2.5 Å resolution. The novel three-dimensional structure of TI-II in a ternary complex with proteinases opens avenues of research previously difficult to undertake due to the lack of structural information on the mode of inhibition of multidomain Pot II inhibitors (see Chapter 6). In Chapter 4, the three-dimensional X-ray crystallographic structure of unbound TI-II is presented. The presence of four copies of unbound TI-II in the asymmetric unit presents a unique opportunity to study the conformational flexibility of this inhibitor.

1.3 Aspartic Proteinases: Structure and Mechanism

Aspartic proteinases are found in a wide range of organisms, including viruses, fungi, plants and animals (93). Many aspartic proteinases are involved in important physiological processes, including the processing of the polyprotein in the human immunodeficiency virus (HIV), digestion of proteins in the stomachs of animals and cleavage of renin in the regulation of blood pressure (94). In addition, aspartic

proteinases have been exploited for centuries in the production of cheese and soy sauce (95). As a result, aspartic proteinases and their inhibitors have been intensively studied.

Structural and enzymological studies of aspartic proteinases reveal that diverse members of the aspartic proteinase family adopt a common structure and enzymatic mechanism (93,95). The structure consists of two β-barrel domains connected by a central six-stranded β-sheet, giving rise to a kidney-shaped overall structure with a substrate-binding cleft in the middle (Figure 1-8). A variation on this overall structure is shown in retroviral aspartic proteinases which consist of a homodimer of single domain β-barrels (95,96). Substrate binding pockets are formed by loops contributed by both β-barrel domains. A minimum of six amino acids in peptide substrates, including four residues N-terminal to the scissile bond and two residues C-terminal to the scissile bond, are typically bound in an extended β-strand conformation by aspartic proteinases (95). Many aspartic proteinases show a preference for hydrophobic residues flanking the scissile bond at the P1 and P1' positions (97). An additional feature important to substrate binding and catalysis in aspartic proteinases is a β-hairpin loop near the active site known as the 'flap' (95). The 'flap' appears to be flexible in the unbound state of aspartic proteinases but it becomes ordered and interacts with substrates when they are bound to the enzyme (95,98).

Two aspartic acid residues (Asp32 and Asp215 in pepsin), one donated by each of the two β-barrel domains, are perfectly conserved at the active site cleft and coordinate a catalytic water molecule between their carboxylate side chains. The mechanism of peptide hydrolysis in aspartic proteinases is generally believed to proceed by a non-covalent general acid-base mechanism (Figure 1-9) (93,95,99). Following the binding of

a peptide substrate, Asp215 acts as a general base and accepts a proton from the catalytic water molecule. The resulting nucleophilic hydroxide ion then attacks the carbonyl carbon of the scissile bond, thus forming a tetrahedral intermediate in which the C=O double bond of the carbonyl group becomes a single bond. The buildup of negative charge on the carbonyl oxygen atom of the tetrahedral intermediate is stabilized by Asp32, which acts as a general acid in donating a proton to the negatively charged oxyanion.

The breakdown of the tetrahedral intermediate and cleavage of the scissile bond proceeds by a reversal of the steps involved in the formation of the tetrahedral intermediate. Asp32 acts as a general base to accept a proton from the oxyanion and the C=O double bond of the carbonyl group re-forms. Asp215 acts as a general acid and donates a proton to the amide nitrogen of the scissile bond. The protonated amide nitrogen now becomes a good leaving group and the scissile bond is broken.

1.3.1 Aspartic Proteinase Inhibitors

Aspartic proteinase inhibitors have been intensively studied in the past because of their potential for therapeutic applications in human diseases (94). The most notable application of aspartic proteinase inhibitors in the treatment of human disease has been the treatment of Acquired Immunodeficiency Syndrome (AIDS) using inhibitors of HIV proteinase. Currently, there is great interest in the development of novel inhibitors of aspartic proteinases for the treatment of fungal diseases, cancer, hypertension and Alzheimer's disease (94,100,101).

Most inhibitors of aspartic proteinases are small peptides or peptide-like compounds that have been isolated from natural sources or chemically synthesized (93).

These compounds typically mimic the binding of natural peptide substrates to aspartic proteinases but have modifications at the bond corresponding to the scissile bond of peptide substrates that prevent hydrolysis. The prototypical small molecule inhibitor of aspartic proteinases is pepstatin, which is a hexapeptide in which a central statine residue ((3S,4S)-4-amino-3-hydroxy-6-methylheptanoic acid) results in the replacement of the scissile peptide bond with a non-hydrolyzable hydroxymethylene isostere of the tetrahedral amide hydrate intermediate normally produced during hydrolysis (95,102).

In addition to small molecule inhibitors, four proteinaceous inhibitors of aspartic proteinases are known (93). This number is quite small compared to the large number of proteinaceous inhibitors known for serine, cysteine and metalloproteinases. The proteinaceous inhibitors of aspartic proteinases are the pepsin inhibitor PI-3 from the intestinal nematode parasite *Ascaris* (14,103), IA-3 from yeast (104), equistatin from sea anemone (105) and a protein from squash phloem (106). None of these proteins shares any sequence similarity, and it is likely that all inhibit aspartic proteinases by differing mechanisms. The mechanism for inhibition has been determined only for PI-3 and IA-3, where crystal structures have been determined for both inhibitors in complex with aspartic proteinases (107,108).

IA-3 is a 68-residue polypeptide that is a very tight-binding inhibitor of yeast proteinase A (K_i = 1.1 nM) (104). When free in solution, it possesses little ordered secondary structure, but it forms a long amphipathic α-helix when bound to yeast proteinase A (108). The inhibitor binds in a manner unlike that of peptide substrates, but the binding occurs in a manner that occludes the substrate-binding cleft from access by potential substrates.

PI-3 is a tight-binding inhibitor of several vertebrate aspartic proteinases involved in digestion (K_i = 1 nM for pepsin, 2 nM for gastricsin), and is believed to be produced by the intestinal parasite *Ascaris* to protect the nematode from being digested as it passes through the stomach to the intestine (14,103). The crystal structure of the pepsin:PI-3 complex (107) reveals that the amino-terminus of the inhibitor binds to pepsin in a manner nearly identical to the portion of a substrate carboxy-terminal to the scissile bond (Figure 1-10). The mechanism of inhibition thus appears to be steric blockage of the peptide binding cleft. Further stabilization of the inhibitor-proteinase complex is contributed by the formation of an antiparallel β-strand with the β-hairpin 'flap' in pepsin adjacent to the substrate-binding cleft, and the packing of a polyproline II helix in the inhibitor with a surface loop also adjacent to the substrate-binding cleft of pepsin.

The structure of PI-3 reveals the presence of two domains, remarkably with only one domain interacting with pepsin. Furthermore, the two domains of the extended inhibitor make few interactions with each other, and the structure of the inhibitor suggests that a single-domain inhibitor may be both active and stable in the absence of the second domain. The design, construction and testing of a single-domain inhibitor based on the structure of PI-3 bound to pepsin is described in Chapter Five of this thesis.

Figure 1-1: Possible mechanisms of proteinase inhibition. **(A)** Direct blockage of the proteinase's active site in a substrate-like manner, as seen in standard mechanism, canonical proteinase inhibitors of serine proteinases. **(B)** Indirect blockage of the proteinase's active site, as seen in cystatin inhibition of papain-like cysteine proteinases. **(C)** Binding of inhibitors to sites adjacent to the active site or to exosites, as seen in thrombin inhibitors. **(D)** Allosteric inhibition of the proteinase through binding to distantly located exosites, as seen in propeptides of serine proteinase zymogens. Adapted from (18). The active site is shown as a black rectangle. The proteinase is coloured in blue and the inhibitor in green.

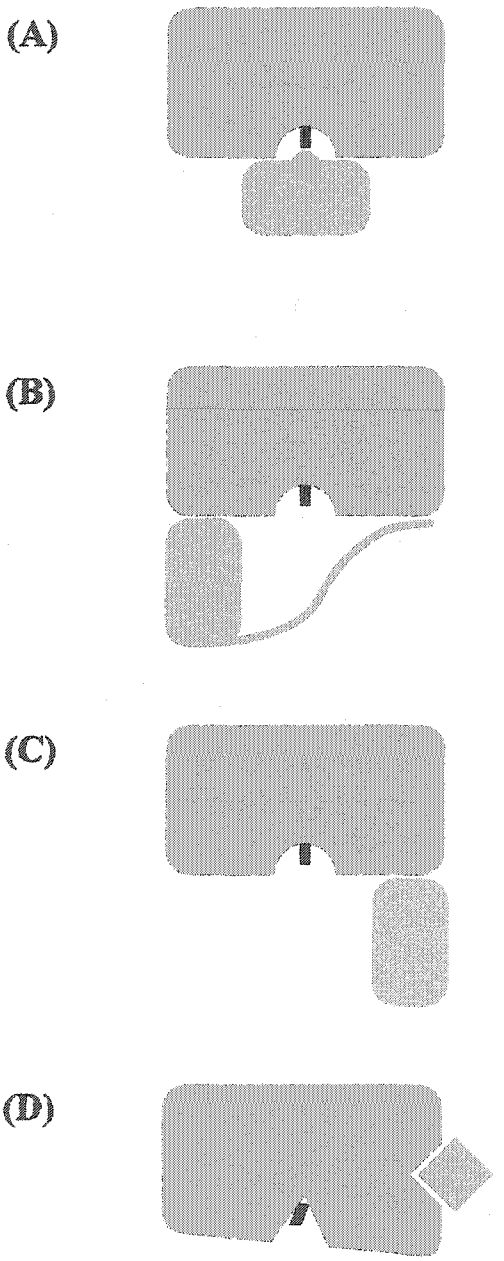


Figure 1-2: Catalytic mechanism of serine proteinases. **(A)** Formation of Michaelis Menten complex and nucleophilic attack. **(B)** Tetrahedral intermediate and formation of an acyl enzyme. **(C)** Acyl enzyme and formation of a second tetrahedral intermediate. **(D)** Second tetrahedral intermediate and deacylation. **(E)** Release of products. Adapted from (109) and (19)

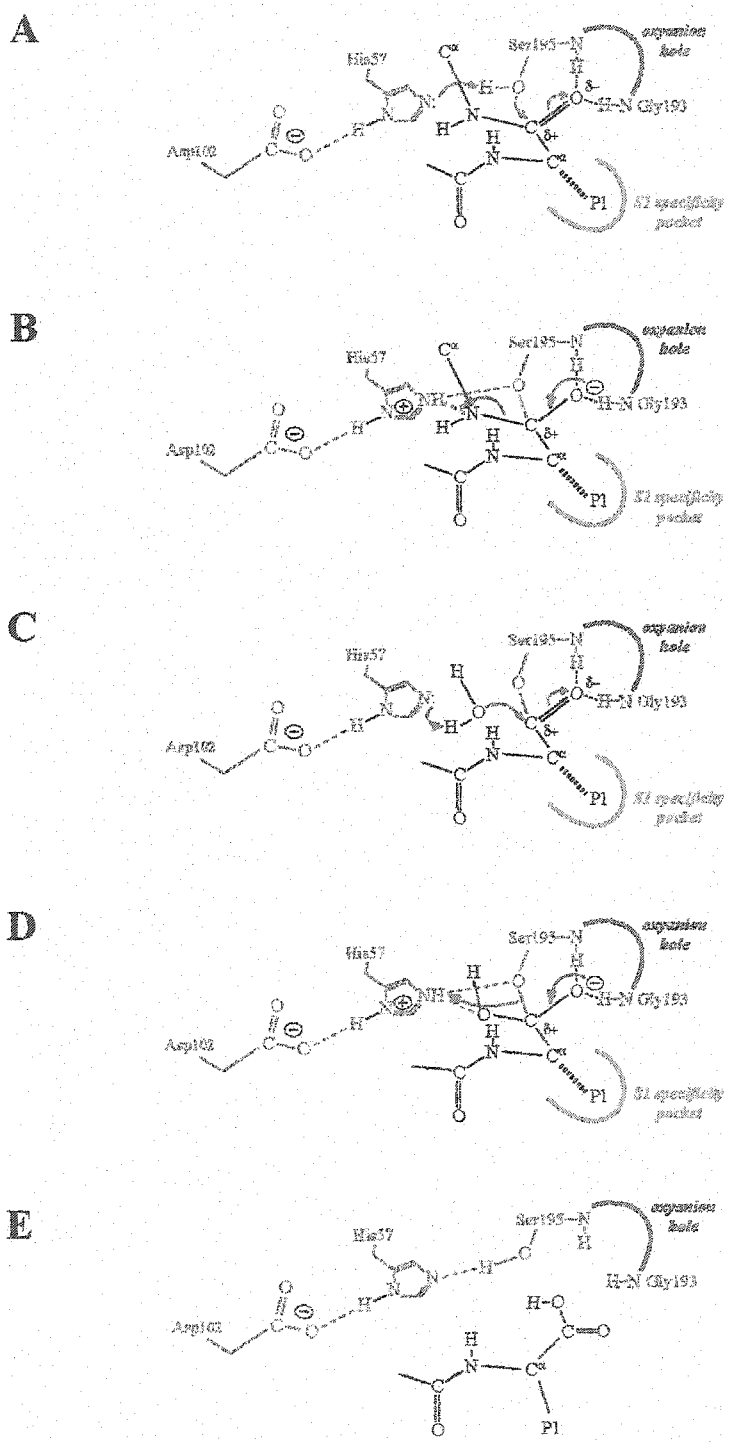


Figure 1-3: Schematic diagram based on the Schechter and Berger nomenclature for proteinases (26). Subsites (S) and peptide positions (P) are related to the position of the peptide substrate in the proteinase active site. The scissile bond is shown as the bond broken by a red star.

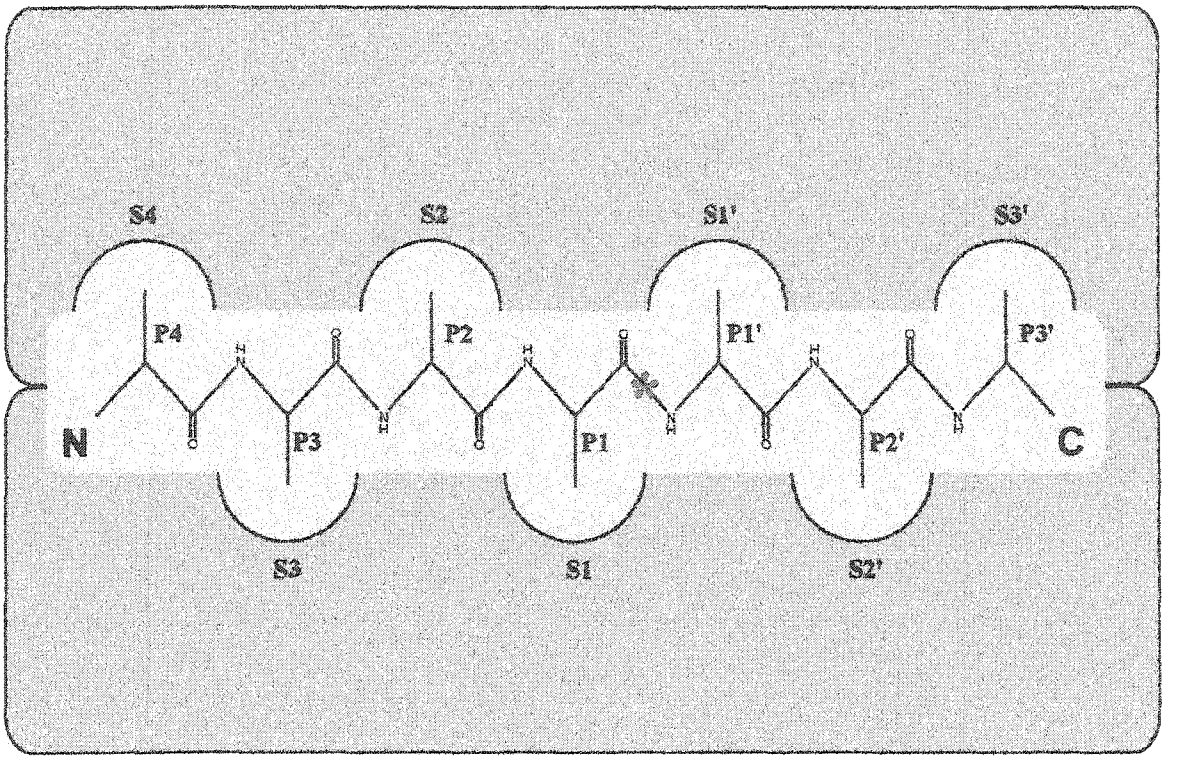


Figure 1-4: Electron micrograph of *EAV* particle budding from smooth intracellular membranes (bar is 25 nm). Reproduced from (36).

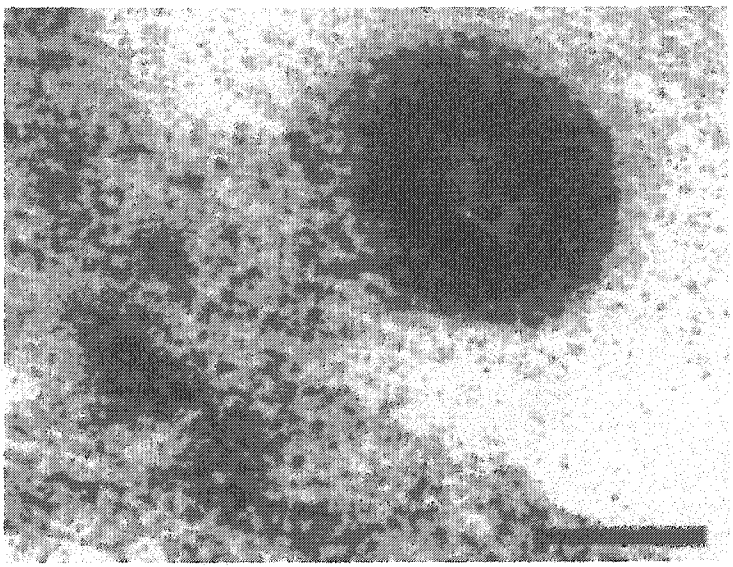


Figure 1-5: Schematic overview of the *EAV* life cycle. The *EAV* genome organization, including the replicase cleavage sites (arrowheads), is depicted at the top of the figure. Subgenomic (sg) mRNAs are illustrated at the right. Endoplasmic reticulum (ER); plasma membrane (PM); double membrane vesicle (DMV); nucleocapsid (NC). Reproduced from (36).

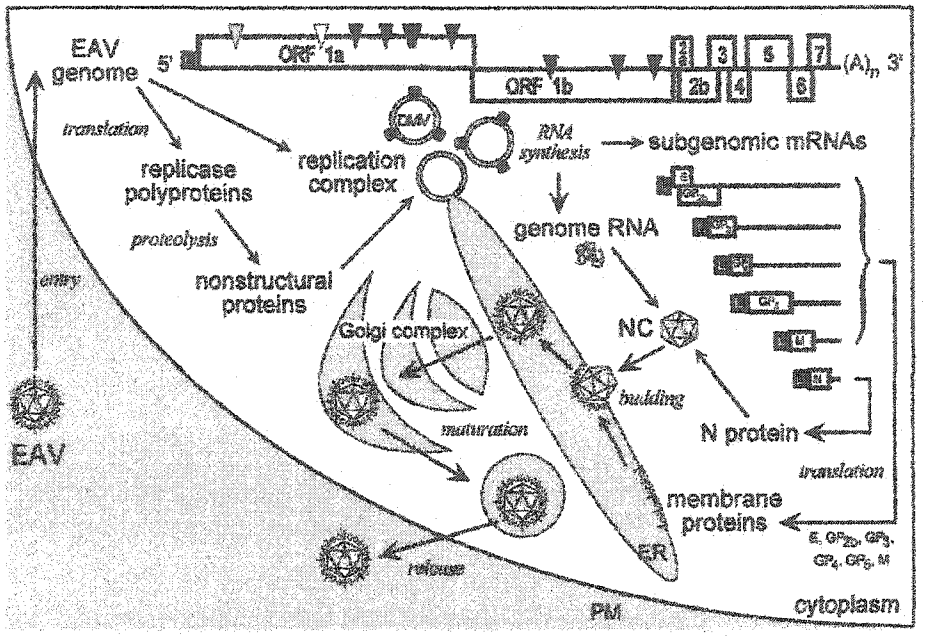


Figure 1-6: *EAV* proteolytic processing pathways **(A)** Overview of the proteolytic processing of the *EAV* replicase ORF1a and ORF1ab polyproteins. The three *EAV* proteinases (located in nsp1, nsp2, and nsp4), their cleavage sites, and the *EAV* nsp nomenclature are depicted. Abbreviations: PCP, papain-like cysteine proteinase; SP, serine proteinase; RdRP, RNA-dependent RNA polymerase; Z, zinc finger; Hel, helicase; N, Nidovirus-specific conserved domain. **(B)** Overview of the two alternative processing pathways which apply to the C-terminal half of the *EAV* ORF1a protein. The association of cleaved nsp2 with nsp3-8 (and probably also nsp3-12) was shown to direct the cleavage of the nsp4|5 site by the nsp4 proteinase (major pathway). Alternatively, in the absence of nsp2, the nsp5|6 and 6|7 sites are processed and the nsp4|5 junction remains uncleaved. The status of the small nsp6 subunit (fully cleaved or partially associated with nsp5 or nsp7) remains to be elucidated.

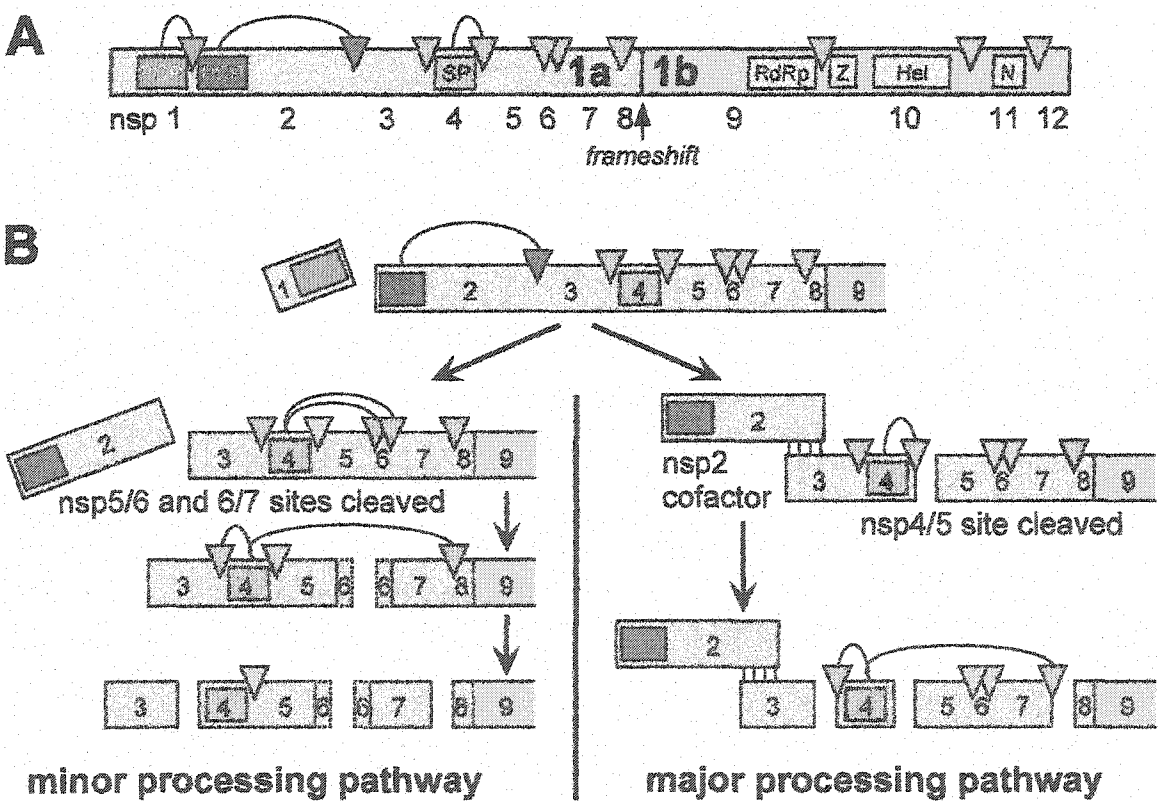


Figure 1-7: Multiple sequence alignment of representative members of the Potato II family of proteinase inhibitors. Secondary structural elements in TI-II are noted above the sequence, with boxes representing helices and arrows representing β-strands. Red and blue boxes indicate the boundaries of polypeptide segments forming TI-II Domains I and II respectively. Green boxes indicate the boundaries of reactive site loop residues P4 to P2'. Each sequence is preceded by its name in the Swiss-Prot database. LYCES indicates inhibitors from *Lycopersicon esculentum*, SOLTU *Solanum tuberosum*, TOBAC *Nicotiana tabacum* and CAPAN *Capsicum annuum.*

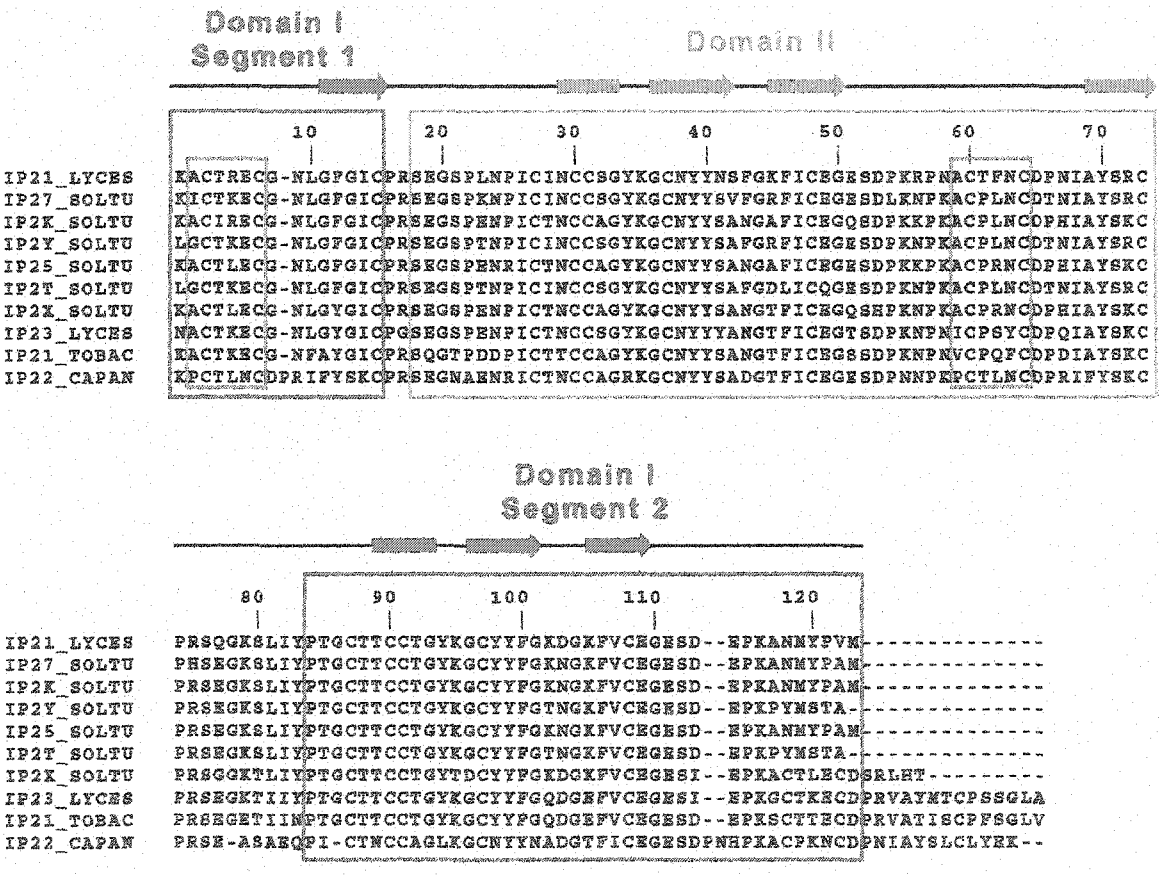


Figure 1-8: Structure of human pepsin bound to pepstatin (PDB code 1PSO) (98). Pepsin is drawn as blue ribbon and pepstatin as yellow ball-and-sticks. The active site Asp32 and Asp215 residues are drawn as pink ball-and-sticks.

Figure 1-9: Catalytic mechanism of aspartic proteinases. **(A)** Nucleophilic attack on carbonyl carbon by activated water molecule. **(B)** Noncovalently bound tetrahedral intermediate stabilized by hydrogen bonding with Asp32. **(C)** Breakdown of tetrahedral intermediate with Asp32 acting as general base and Asp215 acting as general acid. **(D)** Scissile peptide bond has been hydrolyzed and products are released. Adapted from (99).

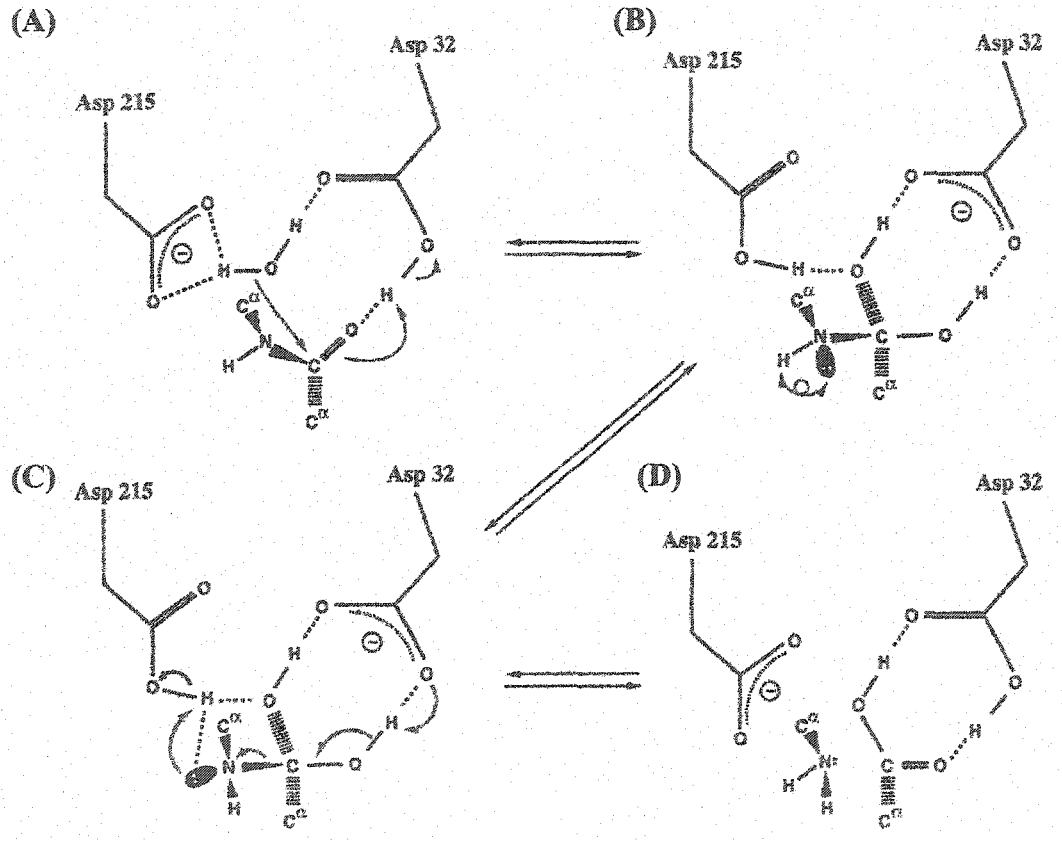
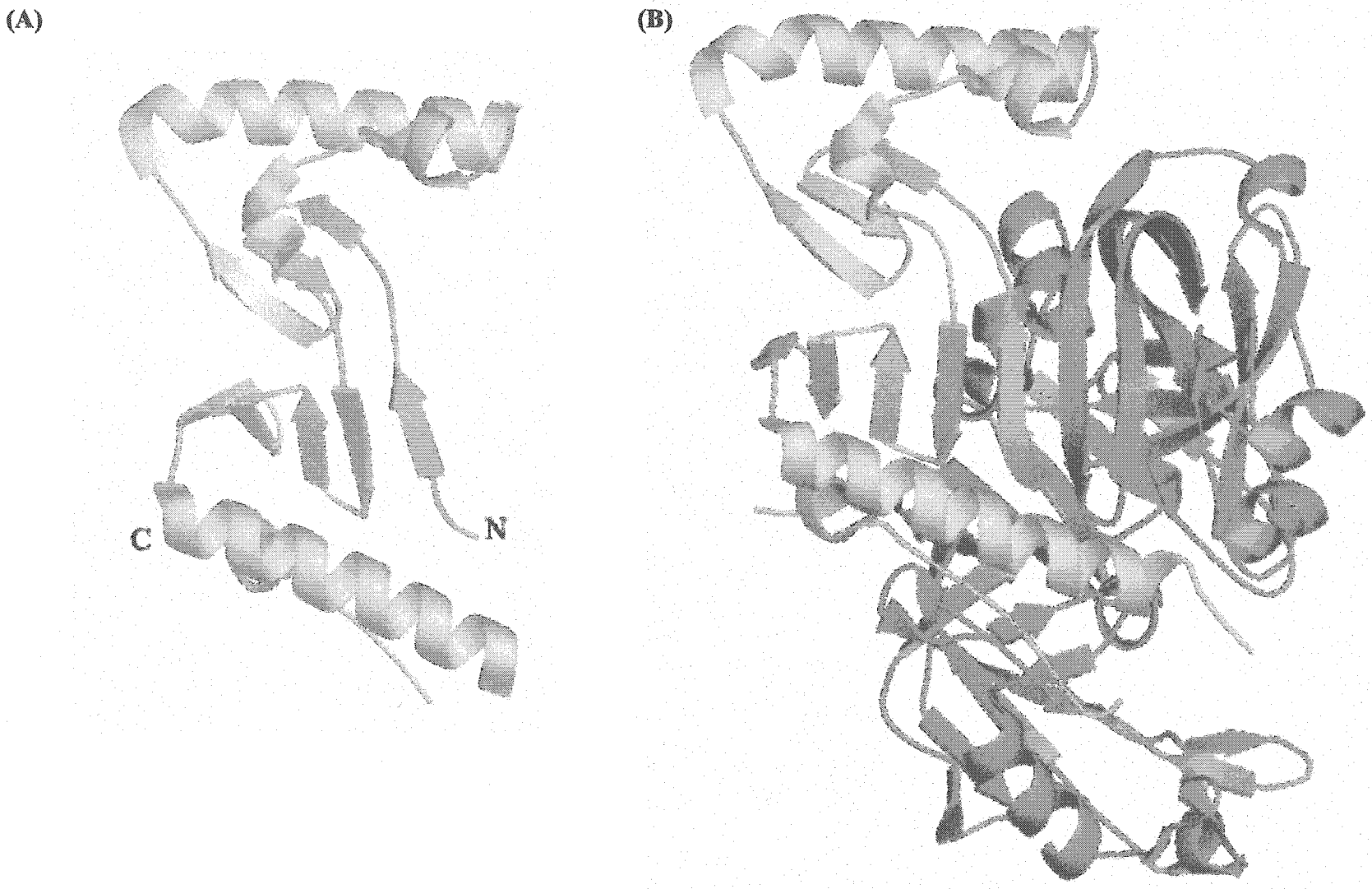


Figure 1-10: Structure of pepsin inhibitor-3 in **(A)** the unbound form (PDB code 1F32) and **(B)** the form bound to porcine pepsin (PDB code 1F34) (107).



1.4 References

1. Hedstrom, L. (2002) *Chem Rev* 102, 4429-4430
2. Rawlings, N. D., O'Brien, E. A., and Barrett, A. J. (2002) *Nucl. Acids Res.* 30, 343-346
3. Neurath, H. (1984) *Science* 224, 350-357
4. Otlewski, J., Krowarsch, D., and Apostoluk, W. (1999) *Acta Biochim. Pol.* 46, 531-565
5. Davie, E. W., Fujikawa, K., and Kisiel, W. (1991) *Biochemistry* 30, 10363-10370
6. Joseph, K., Ghebrehiwet, B., and Kaplan, A. P. (2001) *Biol. Chem.* 382, 71-75
7. Barros, C., Crosby, J. A., and Moreno, R. D. (1996) *Cell Biol Int* 20, 33-39
8. Edwards, D. R., and Murphy, G. (1998) *Nature* 394, 527-528
9. Watorek, W., Farley, D., Salvesen, G., and Travis, J. (1988) *Adv Exp Med Biol* 240, 23-31
10. Froelich, C. J., Zhang, X., Turbov, J., Hudig, D., Winkler, U., and Hanna, W. L. (1993) *J. Immunol.* 151, 7161-7171
11. Ryan, C. (1973) *Ann. Rev. Plant Physiol.* 24, 173-196
12. Laskowski, M. J., and Kato, I. (1980) *Ann. Rev. Biochem.* 49, 593-626
13. Ryan, C. A. (1990) *Annu. Rev. Phytopathol.* 28, 425-449
14. Abu-Erreisch, G. M., and Peanasky, R. J. (1974) *J Biol Chem* 249, 1558-1565
15. Dissanayake, S., Xu, M., Nkenfou, C., and Piessens, W. F. (1993) *Mol. Biochem. Parasitol.* 62, 143-146
16. Hong, X. Q., Mejia, J. S., Kumar, S., Perler, F. B., and Carlow, C. K. S. (1996) *Parasitology* 112, 331-338

17. Girdwood, K., Perler, F. B., and Berry, C. (1998) *Adv Exp Med Biol* 436, 419

18. Bode, W., and Huber, R. (2000) *Biochim Biophys Acta* 1477, 241-252

19. Hedstrom, L. (2002) *Chem Rev* 102, 4501-4523

20. Matthews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. (1967) *Nature* 214, 652-656

21. Wright, C. S., Alden, R. A., and Kraut, J. (1969) *Nature* 221

22. Liao, D., Breddam, K., Sweet, R. M., Bullock, T., and Remington, S. J. (1992) *Biochemistry* 31, 9796-9812

23. Lesk, A. M., and Fordham, W. D. (1996) *J. Mol. Biol.* 258, 501-537

24. Allaire, M., Chernaia, M. M., Malcolm, B. A., and James, M. N. (1994) *Nature* 369, 72-76

25. Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L., and Worland, S. (1994) *Cell* 77, 1-20

26. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* 27, 157-162

27. Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K., and Laskowski, M. J. (1997) *J. Mol. Biol.* 266, 441-461

28. Dixon, M. M., and Matthews, B. W. (1989) *Biochemistry* 28, 7033-7038

29. Katz, B. A., Finer-Moore, J., Mortezaei, R., Rich, D. H., and Stroud, R. M. (1995) *Biochemistry* 34, 8264-8680

30. Katona, G., Wilmouth, R. C., Wright, P. A., Berglund, G. I., Hajdu, J., Neutze, R., and C.J., S. (2002) *J. Biol. Chem.* 277, 21962-21970

31. Wilmouth, R. C., Edman, K., Neutze, R., Wright, P. A., Clifton, I. J., Schneider, T. R., Schofield, C. J., and Hajdu, J. (2001) *Nat Struct Biol* 8, 689-694

32. Wensvoort, G., Terpstra, C., Pol, J. M., ter Laak, E. A., Bloemraad, M., de Kluyver, E. P., Kragten, C., van Buiten, L., den Besten, A., and Wagenaar, F. (1991) *Vet. Q.* 13, 121-130

33. Doll, E. R., Bryans, J. T., McCollum, W. H. M., and Wallace, M. E. (1957) *Cornell Vet.* 47, 3-41

34. Timoney, P. J. (2002) *Equine Respiratory Diseases* (Lekeux, P., Ed.), International Veterinary Information Service (www.ivis.org), Ithaca, NY

35. Botner, A., Strandbygaard, B., Sorensen, K. J., Have, P., Madsen, K. G., Madsen, E. S., and Alexandersen, S. (1997) *Vet. Rec.* 141, 497-499

36. Snijder, E. J., and Meulenberg, J. J. (2001) in *Fields Virology* (Knipe, D. M., and Howley, P. M., eds) Vol. 1, 4th Ed. Ed., pp. 1205-1220, 2 vols., Lippincott Williams and Wilkins, Philadelphia, PA

37. Snijder, E. J., and Meulenberg, J. J. (1998) *J. Gen. Virol.* 79, 961-979

38. Ziebuhr, J., Snijder, E. J., and Gorbalenya, A. E. (2000) *J. Gen. Virol.* 81, 853-879

39. Pasternak, A. O., van den Born, E., Spaan, W. J., and Snijder, E. J. (2001) *EMBO J* 20, 7220-7228

40. Snijder, E. J., Wassenaar, A. L., van Dinten, L. C., Spaan, W. J., and Gorbalenya, A. E. (1996) *J. Biol. Chem.* 271, 4864-4871

41. den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C., and Spaan, W. J. (1991) *J. Virol.* 65, 2910-2920

42. van Dinten, L. C., Rensen, S., Gorbalenya, A. E., and Snijder, E. J. (1999) *J. Virol.* 73, 2027-2037

43. Wassenaar, A. L., Spaan, W. J., Gorbalenya, A. E., and Snijder, E. J. (1997) *J. Virol.* 71, 9313-9322

44. Snijder, E. J., Wassenaar, A. L., and Spaan, W. J. (1994) *J. Virol.* 68, 5755-5764

45. Pedersen, K. W., van der Meer, Y., Roos, N., and Snijder, E. J. (1999) *J. Virol.* 73, 2016-2026

46. Bode, W., and Huber, R. (1992) *Eur. J. Biochem.* 204, 433-451

47. Laskowski, M. J., and Qasim, M. A. (2000) *Biochim. Biophys. Acta* 1477, 324-337

48. Ozawa, K., and Laskowski, M. J. (1966) *J. Biol. Chem.* 241, 3955-3961

49. Bolognesi, M., Gatti, G., Menegatti, E., Guarni, M., Marquart, M., Papamokos, E., and Huber, R. (1982) *J. Mol. Biol.* 162, 839-868

50. Blow, D. M. (1974) in *Proteinase Inhibitors, Bayer Symposim V* (Fritz, H., Tschesche, H., Greene, L. J., and Truscheit, E., eds) Vol. 5, pp. 677-678, Springer-Verlag, New York

51. Read, R. J., Fujinaga, M., Sielecki, A. R., and James, M. N. G. (1983) *Biochemistry* 22, 4420-4433

52. Read, R. J., and James, M. N. G. (1986) in *Proteinase Inhibitors* (Barrett, A. J., and Salvesen, G., eds) Vol. 12, pp. 301-336, Elsevier, New York

53. Finkenstadt, W. R., Hamid, M. A., Mattis, J. A., Schrode, J., Sealock, R. W., Wang, D., and Laskowski, M. J. (1974) in *Bayer Symp. V* (Fritz, H., Tschesche, H., Greene, L. J., and Truscheit, E., eds), pp. 389-411, Springer-Verlag, Berlin

54. Read, J. W., and Haas, L. W. (1938) *Cereal Chem* 15, 59-68

55. Richardson, M. (1977) *Phytochemistry* 16, 159-164

56. Brown, W. E., and Ryan, C. A. (1984) *Biochemistry* 23, 3418-3422

57. Jongsma, M. A., and Bolter, C. (1997) *J. Insect. Physiol.* 43, 885-895

58. Janzen, D. H. (1979) in *Herbivores, their interactions with secondary plant metabolites* (Rosenthal, A., and Janzen, D. H., eds), pp. 331-350, Academic Press, New York

59. Bowles, D. J. (1990) *Annu Rev Biochem* 59, 873-907

60. Green, T. R., and Ryan, C. A. (1972) *Science* 175, 776-777

61. Johnson, R., Narvaez, J., An, G., and Ryan, C. (1989) *Proc Natl Acad Sci U S A* 86, 9871-9875

62. Thornburg, R. W., An, G., Cleveland, T. E., Johnson, R., and Ryan, C. A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 744-748

63. Boulter, D., Gatehouse, A. M. R., and Hilder, V. (1989) *Biotechnol. Adv.* 7, 489-497

64. Duan, X., Li, X., Xue, Q., Abo-El-Saad, M., Xu, D., and Wu, R. (1996) *Nat. Biotech.* 14, 494-498

65. Marchetti, S., Delledonne, M., Fogher, C., Chiaba, C., Chiesa, F., Savazzini, F., and Giordano, A. (2000) *Theor. Appl. Genet.* 101, 519-526

66. Plunkett, G., Senear, D. F., Zuroske, G., and Ryan, C. A. (1982) *Arch. Biochem. Biophys.* 213, 463-472

67. Pearce, G., Johnson, S., and Ryan, C. A. (1993) *Plant Physiol.* 102, 639-644

68. Pearce, G., Ryan, C. A., and Liljegren, D. (1988) *Planta* 175, 527-531

69. Bryant, J., Green, T. R., Gurusaddaiah, T., and Ryan, C. A. (1976) *Biochemistry* 15, 3418-3424

70. Stiekema, W. J., Heidekamp, F., Dirkse, W. G., van Beckum, J., de Haan, P., ten Bosch, C., and Louwerse, J. D. (1988) *Plant Mol. Biol.* 11, 255-269

71. Richardson, M. (1979) *FEBS Lett.* 104, 322-326

72. Antcheva, N., Patthy, A., Athanasiadis, A., Tchorbanov, B., Zakhariev, S., and Pongor, S. (1996) *Biochim Biophys Acta* 1298, 95-101

73. Atkinson, A. H., Heath, R. L., Simpson, R. J., Clarke, A. E., and Anderson, M. A. (1993) *Plant Cell* 5, 203-213

74. Pearce, G., Sy, L., Russell, C., Ryan, C. A., and Hass, G. M. (1982) *Arch. Biochem. Biophys.* 213, 456-462

75. Pearce, G., Strydom, D., Johnson, S., and Ryan, C. A. (1991) *Science* 253, 895-898

76. Scheer, J. M., and Ryan, C. A., Jr. (2002) *Proc Natl Acad Sci U S A* 99, 9585-9590

77. Li, L., Li, C., Lee, G. I., and Howe, G. A. (2002) *Proc Natl Acad Sci U S A* 99, 6416-6421

78. Ryan, C. A., and Moura, D. S. (2002) *Proc Natl Acad Sci U S A* 99, 6519-6520

79. Graham, J., Hall, G., Pearce, G., and Ryan, C. A. (1986) *Planta* 169, 399-405

80. Broadway, R. M., and Duffey, S. S. (1986) *J. Insect. Physiol.* 32, 673-680

81. Graham, J. S., Pearce, G., Merryweather, J., Titani, K., Ericsson, L. H., and Ryan, C. A. (1985) *J Biol Chem* 260, 6561-6564

82. Keil, M., Sanchez-Serrano, J., Schell, J., and Willmitzer, L. (1986) *Nucl. Acids Res.* 14, 5641-5650

83. Taylor, B. H., Young, R. J., and Scheuring, C. S. (1993) *Plant Mol. Biol.* 23, 1005-1014

84. Balandin, T., van der Does, C., Belles Albert, J. M., Bol, J. F., and Linthorst, H. J. M. (1995) *Plant Mol. Biol.* 27, 1197-1204

85. Heath, R. J., Barton, P. A., Simpson, R. L., Reid, G. E., Lim, G., and Anderson, M. A. (1995) *Eur. J. Biochem.* 230, 250-257

86. Lee, M. C. S., Scanlon, M. J., Craik, D. J., and Anderson, M. A. (1999) *Nat. Struct. Biol.* 6, 526-530

87. Greenblatt, H. M., Ryan, C. A., and James, M. N. (1989) *J Mol Biol* 205, 201-228

88. Nielsen, K. J., Heath, R. L., Anderson, M. A., and Craik, D. J. (1994) *J. Mol. Biol.* 242, 231-243

89. Nielsen, K. J., Heath, R. L., Anderson, M. A., and Craik, D. J. (1995) *Biochemistry* 34, 14304-14311

90. Lin, G., Bode, W., Huber, R., Chi, C., and Engh, R. A. (1993) *Eur J Biochem* 212, 549-555

91. Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. (2000) *J Mol Biol* 298, 477-491

92. Raj, S. S., Kibushi, E., Kurasawa, T., Suzuki, A., Yamane, T., Odani, S., Iwasaki, Y., and Ashida, T. (2002) *J Biochem (Tokyo)* 132, 927-933

93. Dunn, B. (2002) *Chem Rev* 102, 4431-4458

94. Cooper, J. B. (2002) *Current Drug Targets* 3, 155-173

95. Davies, D. R. (1990) *Annu. Rev. Biophys. Biophys. Chem.* 19, 189-215

96. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) *Nature* 337, 615-620

97. Fruton, J. S. (1976) *Adv. Enzymol.* 44, 1-36

98. Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C., and James, M. N. (1995) *Protein Sci* 4, 960-972

99. James, M. N., Sielecki, A. R., Hayakawa, K., and Gelb, M. H. (1992) *Biochemistry* 31, 3872-3886

100. Citron, M. (2002) *Neurobiology of Aging* 23, 1017-1022

101. Citron, M. (2002) *J. Neurosci. Res.* 70, 373-379

102. Umezawa, H., Aoyagi, T., Morishima, H., Matsuzaki, M., and Hamada, M. (1970) *J. Antibiot. Tokyo* 23, 259-262

103. Martzen, M. R., McMullen, B. A., Smith, N. E., Fujikawa, K., and Peanasky, R. J. (1990) *Biochemistry* 29, 7366-7372

104. Dreyer, T., Valler, M. J., Kay, J., Charlton, P., and Dunn, B. (1985) *Biochem J* 231, 777-779

105. Lenarcic, B., Ritonja, A., Strukelj, B., Turk, B., and Turk, V. (1997) *J Biol Chem* 272, 13899-13903

106. Christeller, J. T., Farley, P. C., Ramsay, R. J., Sullivan, P. A., and Laing, W. A. (1998) *Eur J Biochem* 254, 160-167

107. Ng, K. K.-S., Petersen, J. F., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., Dunn, B. M., Martzen, M. R., Peanasky, R. J., and James, M. N. (2000) *Nat Struct Biol* 7, 653-657

108. Li, M., Phylip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodawer, A., Kay, J., and Gustchina, A. (2000) *Nat Struct Biol* 7, 113-117

109. Wharton, C. W. (1998) in *Comprehensive Biological Catalysis: A Mechanistic Reference* (Sinnott, M., ed) Vol. 1, pp. 345-379, 4 vols., Academic Press Limited, San Diego, CA

CHAPTER 2:

Structure of Arterivirus nsp4: the smallest chymotrypsin-like proteinase with an α/β C-terminal extension and alternate conformations of the oxyanion hole[1]

2.0 Introduction

The arterivirus *Equine Arteritis Virus* (*EAV*) (1) is a major economic concern for the horse-breeding industry worldwide. This virus is widespread in the population, can establish persistent infections, and is easily transmitted via both respiratory and venereal routes.

The family *Arteriviridae* (order *Nidovirales*) comprises enveloped, mammalian RNA viruses with a 12 to 16-kilobase positive-stranded genome (2-4). In the arterivirus prototype *EAV*, the replicase gene is translated into open reading frame 1a (ORF1a) and ORF1ab polyproteins of 1,727 and 3,175 amino acids, respectively, the latter product resulting from a ribosomal frameshift that can occur just prior to termination of ORF1a translation. As illustrated in Figure 1-6, the *EAV* ORF1a and ORFab proteins are cleaved by three different ORF1a-encoded proteinases, papain-like cysteine proteinases located in nonstructural protein 1 (nsp1) and nsp2, and the chymotrypsin-like serine proteinase nsp4 (5). Following the rapid autocatalytic release of nsp1 and nsp2, the remainder of the polyproteins (nsp3-8 and nsp3-12) is processed by nsp4, the main viral proteinase (6).

---

[1] *A version of this chapter has been published. Barrette-Ng, I.H., Ng, K.K.-S., Mark, B.L., van Aken, D., Cherney, M.M., Garen, C., Kolodenko, Y., Gorbalenya, A.E., Snijder, E. J., and James, M.N.G. (2002) J. Biol. Chem., 277(42), 39960-39966. Protein purification was performed by D. van Aken and E.J. Snijder and protein crystallization was done by M. Cherney.*

The nsp4 main proteinase processes the remaining eight cleavage sites, five in the ORF1a protein (7,8) and three in ORF1b-encoded part of the ORF1ab protein (9), which encodes (among other functions) the viral RNA-dependent RNA polymerase and helicase (5). Alternative processing pathways are used during cleavage of the nsp3-8 intermediate (8). Unfortunately, the molecular details of alternative proteolytic processing are at present poorly understood.

To provide a structural framework for understanding the complexities of proteolysis in arteriviruses, we have determined the crystallographic structure of *EAV* nsp4 at 2.0 Å resolution. The structure reveals the smallest chymotrypsin-like fold known as well as a novel C-terminal α/β extension domain.

2.1 Experimental procedures

2.1.1 Expression, purification and crystallization

Nsp4 (nts 3417 to 4025 of the *EAV* genome; EMBL accession number X53459) was cloned into pMAL-c2 (New England Biolabs, Inc.) and expressed as a MBP-nsp4 fusion in *E. coli* BL21(DE3) (Promega). Expression was induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside for 4 h at 37°C and cells were lysed using a French press. MBP-nsp4 was purified by amylose-affinity chromatography and cleaved with 1 U thrombin/mg protein in 20 mM Tris pH 8.1 and 150 mM NaCl. Following cleavage, nsp4 contained additional residues at its N and C terminus, Ser-Met and Leu-Ala-Ser, respectively. Further purification was achieved using Sepharose G-75 and Mono Q columns (Pharmacia Biotech). MBP and uncleaved fusion protein were removed using amylose resin. The integrity of nsp4 was confirmed by mass spectrometry. Crystals were grown at room temperature by the hanging drop vapour diffusion method by mixing 2 μL of protein (25 mg/mL) and 2 μL of 23 % (w/v) PEG MME 2000, 100 mM Tris, pH 8.5, 10 % (w/v) ethylene glycol, and 0.2 M ammonium acetate. Crystals grew within one to two weeks and had typical dimensions of 0.5 x 0.2 x 0.2 mm. The crystals belonged to space group P1 with cell dimensions $a = 56.6$, $b = 60.7$, $c = 62.8$ Å, $\alpha = 74.0°$, $\beta = 63.2°$ and $\gamma = 66.2°$. Four molecules were contained in the asymmetric unit.

2.1.2 Data collection, phasing and refinement

All data were collected on flash-frozen crystals at 100 K by transferring crystals from the mother liquor or heavy atom soaking solutions into a nitrogen gas stream. The structure of nsp4 was solved using the multiple isomorphous replacement technique with five heavy atom derivatives (Table 2-1). Data from the crystal soaked in mersalyl were

measured at beamline BM-14-C (BioCARS-CAT, Advanced Photon Source) using an ADSC Quantum-4 detector. All other data were measured using CuK_{α} radiation (Rigaku RU-300 generator, Osmic optics) and a R-AXIS IV++ image plate detector. Data were processed using the HKL suite (10). Experimental phases were calculated using SOLVE (11). RESOLVE (12) was used for density modification, improving the figure of merit from 0.57 to 0.69, and to generate a partial main-chain trace. The experimental electron density map was further improved by ARP/wARP (13). From this modified electron density map, a nearly complete trace of a single copy in the asymmetric unit was constructed using XFIT (14). A model for the complete asymmetric unit was generated by transposing the coordinates of the first copy onto partial models of the other copies, and making manual adjustments. Refinement against experimental phase information was performed using CNS (15). The model was further refined against the Native 1 data set using REFMAC v. 5.0.36 (16) with individual TLS parameters refined separately for the N-terminal barrel, C-terminal barrel and the C-terminal domain. The Native 1 data set was used for refinement instead of the higher resolution mersalyl data set because of systematic measurement errors in the latter. Non-crystallographic symmetry restraints were not used at any point during refinement, because of significant differences in conformation between different copies. Residues 1 to 6 are missing from all copies. In addition, the following residues are missing: 204 in copy A, 199 to 204 in copy B, 198 to 204 in copy C, and 8 and 204 in copy D. 88.6 % of residues lie in the most favoured regions of the Ramachandran plot and 10.7 % lie in the additional allowed regions (regions defined by PROCHECK) (17). Additional checks on geometry were performed using WHATCHECK (18). Data quality, phasing and refinement statistics are given in

Table 2-1. Figures were prepared with MOLSCRIPT (19), BOBSCRIPT (20) and RASTER3D (21).

2.2 Results and discussion

2.2.1 Structure of *EAV* nsp4

The structure of *EAV* nsp4 was determined using the multiple isomorphous replacement technique (Figure 2-1A; refer to Table 2-1 for crystallographic parameters). There are four copies of nsp4 (designated A, B, C and D) in the triclinic unit cell, each containing two β-barrels, as well as a unique C-terminal domain. The N-terminal barrel consists of six β-strands (A1 to F1), while the C-terminal barrel is composed of seven (A2 to G2) (Figure 2-1B). The G2 β-strand is also found in *Sindbis virus* core protein (SCP) (22) and *Semliki forest virus* core protein (SFCP) (23) but is not present in most other chymotrypsin-like proteinases (CLPs). The core of both β-barrels is comprised of conserved hydrophobic residues (Figure 2-1B). In addition, Trp114 is a conserved, fully solvent exposed hydrophobic residue that lies in a groove lined with other conserved residues, near the substrate binding pockets. This groove and Trp114 may mediate protein-protein interactions.

The most striking feature of the nsp4 structure is the presence of an additional C-terminal domain not found in most other CLPs. This third domain comprises residues 156 to 204 and consists of two short pairs of β-strands and two α-helices. The C-terminal domain interacts with the C-terminal barrel through an interface (buried surface area of 1638 $Å^2$) consisting of conserved hydrophobic residues: Leu105 and Leu112 from the C-terminal β-barrel and Val158, Leu163, Phe167, Ile182, Leu196 and Ile197 from the C-terminal domain. There is also an exposed patch of conserved solvent-

exposed hydrophobic residues (Figure 2-1C) that may form part of the interface with nsp5 in the nsp4-5 intermediate. This hydrophobic patch may also mediate interactions with nsp2, that associates with nsp3-8 to induce cleavage of the nsp4|5 site by nsp4 (8).

The overall conformation of the two β-barrels in all four copies of nsp4 is highly conserved, except for the loops between C1 and D1, and E2 and F2, both of which are flexible and have higher temperature factors (Figure 2-1D). The largest variation in conformation is seen in the position of the C-terminal domain relative to the β-barrels. The r.m.s.d.'s when superimposing the entire molecule range from 0.40 to 0.77 Å and are 11 to 91 % higher than when superimposing only the β-barrels. DYNDOM (24) reveals a rotation (residue 157 or 161 as the hinge) of the C-terminal domain relative to the proteinase domain of 11° when comparing copies A and C or C and D. The C-terminal domain can clearly adopt different orientations relative to the two β-barrels, which may facilitate substrate binding or autoproteolysis, as discussed below.

2.2.2 Comparison to other chymotrypsin-related proteinases

Despite having little sequence identity to other CLPs, nsp4 was proposed to have a CLP fold (4,7). The crystal structure confirms the existence of a CLP fold, which at 149 residues, is the smallest serine proteinase domain known. *Human Rhinovirus 2* (*HRV-2*) 3C is a viral CLP (25) with a similar P1 specificity of substrates to nsp4. A least squares superposition with nsp4 gives a r.m.s.d. of 0.96 Å (40 C_α pairs, cut-off 1.5 Å) (Figure 2-2A). Apart from the additional C-terminal domain in nsp4, a major difference between the two structures is the length of the loop separating strands B2 and C2. In *HRV-2* 3C and in most other CLPs, this β-hairpin is 10 to 20 residues longer. The shorter β-hairpin found in nsp4 provides enough space for the seventh β-strand of the C-

terminal barrel, G2. A short β-hairpin connecting B2 and C2 is also present in SCP and SFCP, both of which also have G2. The presence of this shorter β-hairpin in SCP may facilitate *cis* cleavage to produce the viral capsid protein (22). The structures of nsp4 and SCP align well, despite remote sequence similarity, with a r.m.s.d. of 0.95 Å (39 C_α pairs, cut-off 1.5 Å) (Figure 2-2B). The overall shape of nsp4 is much like SCP in being more compact than most other CLPs and results from shorter β-hairpins.

Another example of a CLP with a specificity for glutamic acid in the P1 position is *Streptomyces griseus* proteinase E (SGPE) (26) (Figure 2-2C). Nsp4 and SGPE align well with a r.m.s.d. of 0.86 Å (56 C_α pairs, cut-off 1.5 Å). The S1 specificity pocket in nsp4, SGPE and picornaviral 3C proteinases show remarkable structural similarities.

A recently reported structure of a coronavirus chymotrypsin-like cysteine main proteinase from the *Transmissible Gastroenteritis virus* in pigs reveals the only other structure of a CLP with an additional C-terminal domain (27). Coronaviruses comprise the distantly related second family in the order Nidovirales (4). Unlike nsp4, the structure of the C-terminal domain in the coronavirus proteinase is nearly twice as large at 110 residues and is comprised of only α-helices. In addition, the linker between the C-terminal domain and the C-terminal β-barrel differs markedly from that found in nsp4. Despite the lack of sequence and structural similarity, the C-terminal domains in both arteriviruses and coronaviruses may share a common functionality in facilitating substrate recognition (27).

2.2.3 Nsp4 active site

The crystal structure of nsp4 reveals that His39, Asp65 and Ser120 form a catalytic triad similar to that of other CLPs (Figure 2-1A), confirming the predictions

from sequence analysis (5) and site-directed mutagenesis (7). The active sites in all four copies of nsp4 adopt a very similar conformation (Figure 2-3A), except for the peptide bond between residues 117 and 118. The distances between Ser120 O^{γ} and His39 $N^{\varepsilon 2}$ range from 2.8 to 3.5 Å, and the distances between His39 $N^{\delta 1}$ and the two carboxylate oxygen atoms of Asp65 range from 2.8 to 3.0 Å. These distances are similar to values seen in other CLPs. The rotamer adopted by Ser120 is the same as in other CLPs ($\chi_1 \cong$ -80°).

In CLPs, the oxyanion hole (in nsp4, main chain amides from residues 118 and 120) stabilizes the negative charge on the P1 carbonyl oxygen atom in the tetrahedral intermediate. Remarkably, in three of the four copies, the amide nitrogen of Gly118 points away from Ser120, causing the collapse of the oxyanion hole (Figures 2-3A and 2-3B). In contrast, the oxyanion hole is properly formed in copy B (Figure 2-3C), where the only major difference between the other copies is the conformation of the peptide bond between residues 117 and 118. These two residues lie in a β-turn consisting of Thr116, Ser117, Gly118 and Asp119. The change in conformation of the peptide bond between residues 117 and 118 represents an interconversion of a type I (copies A, C and D) to a type II (copy B) β-turn. The psi angle of Ser117 in copy B is 136°, and for copies A, C and D it ranges between -41° and -51°. The phi angle of Gly118 in copy B is 116° whereas that for copies A, C and D ranges between -59° and -64°. This is a common type of peptide flip observed in pairs of homologous protein structures, especially with a glycine residue in the $i + 1$ position of the flipped peptide bond (28). Molecular orbital calculations indicate an energy barrier of ~3 kcal/mol for this flip (29). Nsp4 is the first example of a wild-type serine proteinase in which alternate conformations of the

oxyanion hole are observed. Previously, a collapsed oxyanion hole was observed in an active site cysteine to alanine mutant of HAV 3C proteinase (30). The collapsed oxyanion hole in nsp4 may be part of a novel mechanism regulating proteolytic activity.

An important structural feature stabilizing the collapsed conformation of the oxyanion hole is a water molecule that forms hydrogen bonds with Ser120 O^{γ} and His39 $N^{\varepsilon 2}$ (Figure 2-3B). A second water molecule hydrogen bonds with His134 $N^{\varepsilon 2}$ and the main-chain carbonyl oxygen atom of Thr116. These hydrogen bonds can only form if the oxyanion hole adopts a collapsed conformation and hence these water molecules are absent from copy B.

2.2.4 Substrate binding

Previous studies have shown that nsp4 has a specificity for Glu (and in one case Gln) at the P1 position (4,7,9). As in the picornaviral 3C proteinases and SGPE, which have specificity for either Gln or Glu in the P1 position, nsp4 also has a conserved histidine residue (His134 in nsp4, His161 in *HRV-2* 3C, His213 in SGPE) at the base of the S1 specificity pocket, and a conserved Ser/Thr residue lining one "wall" of the S1 pocket (Thr115 in nsp4, Thr142 in *HRV-2* 3C, Ser192 in SGPE) (Figure 2-4A). The histidine in the S1 pocket hydrogen bonds to the P1 side-chain carbonyl oxygen atom of an inhibitor bound to *HRV-2* 3C (25) and to one of the glutamate carboxylate oxygens of the P1 side chain of a peptide bound to SGPE (26). Thr142 in *HRV-2* 3C and Ser192 in SGPE also hydrogen bond to the carbonyl or carboxylate oxygen atoms of the P1 side chain. Because the structures of the active site and S1 pocket are conserved, His134 and Thr115 in nsp4 could also donate hydrogen bonds to the carboxylate oxygen atom of the P1 side chain in substrates. Site-directed mutagenesis studies confirm that His134 is

essential for the efficient processing of the *EAV* polyprotein cleavage sites (7). Thr115 can be replaced by serine, glycine and asparagine with variable effects on cleavage specificity and efficiency (7).

Both nsp4 and SGPE prefer glutamic acid over glutamine in the P1 position of substrates. In SGPE, Ser216 hydrogen bonds with the carboxylate oxygen atom (not interacting with His213 or Ser192) of the P1 side chain. Ser216 can not hydrogen bond with the amide nitrogen of a P1 glutamine side chain, because the lone pair already accepts a hydrogen bond from the main-chain amide of Gly219; this proposal is supported by site-directed mutagenesis studies (31). In nsp4, Ser137 occupies the same position as Ser216 in SGPE and should also recognize a negatively-charged P1 residue (Figure 2-4B). In picornaviral 3C proteinases, which prefer glutamine at the P1 position, a glycine is found in lieu of serine.

2.2.5 Nsp4 self-processing: a *cis* or *trans* event?

Nsp4 is known to cleave its N- and C-termini in the nsp3-8 intermediate (4,7-9). Although nsp4 can cleave several sites in *trans* (9), self-processing may also be a *cis* event. To act as both an enzyme and a substrate in *cis* reactions, and to act as an enzyme in *trans* reactions, nsp4 must adopt alternate conformations. The N- or C-termini must be either in or away from the substrate binding pocket for *cis* and *trans* cleavages, respectively. In the nsp4 crystal structure, the chain termini are remarkably close together and do not lie in the substrate binding pocket. The positions adopted by the chain termini likely differ from that adopted during *cis* cleavage to prevent self-inhibition of proteolytic activity, as in picornaviral proteinases (30,32). Each terminus can be brought into the appropriate position for *cis* cleavage by a conformational rearrangement

of the ten N-terminal residues or part of the C-terminal domain. In support of this possibility, the N-terminal six residues are disordered and residues 198 to 203 have higher than average temperature factors (in copies A and D) or are completely disordered (in copies B and C).

One of the few hydrogen-bonding interactions fixing the position of the C-terminus is the Asp119:Arg203 ion pair. Although Asp119 is conserved in arteriviral proteinases, Arg203 is not. Interestingly, Arg4, which is disordered in all four copies of nsp4, is conserved among arteriviral proteinases. Even though the N-terminal six residues are disordered, the position adopted by Lys7 indicates that Arg4 may be able to interact with Asp119 in an alternate conformation. When the C-terminus adopts the conformation required for *cis* cleavage, the Asp119:Arg203 ion-pair must be broken, thereby allowing Arg4 to interact with Asp119.

Coordinates. Coordinates have been deposited in the Protein Data Bank (accession code 1MBM).

Table 2-1 Data collection, phasing and refinement parameters

Data processing and heavy atom statistics

Dataset	Native 1	Native 2	$Pb(CH_3CO_2)_2$	$HgCl_2$	$UO_2(NO_3)_2$	$KPtCl_4$	Mersalyl
Resolution (Å)	16.2 – 2.00	40.0 – 2.10	40.0 – 2.06	40.0 – 2.50	40.0 – 1.90	40.0 – 1.90	40.0- 1.60
High-resolution range (Å)	2.07 – 2.00	2.16 – 2.10	2.13 – 2.06	2.57 – 2.50	1.97 – 1.90	1.97 – 1.90	1.64 – 1.60
Total reflections[1]	169,296 (15,362)	72,351 (5,208)	76,161 (2,292)	46,156 (3,558)	186,601 (9,932)	188,735 (10,212)	313,166 (11,610)
Unique reflections[1]	43,171 (4,151)	37,782 (2,942)	36,920 (1,188)	22,609 (1,834)	48,233 (2,916)	48,619 (2,934)	82,476 (3,261)
Completeness (%)[1]	95.0 (91.5)	94.5 (88.9)	81.6 (31.3)	96.6 (94.0)	90.3 (55.0)	90.3 (54.3)	91.7 (54.4)
I/σ [1]	22.4 (6.1)	11.2 (3.9)	13.0 (7.2)	12.6 (5.1)	18.6 (2.9)	18.5 (3.4)	19.1 (2.8)
R_{sym} (%)[1,2]	5.0 (17.9)	6.5 (16.2)	5.3 (10.2)	6.1 (15.5)	6.6 (35.4)	6.4 (28.5)	5.2 (64.4)
[Heavy atom] (mM)			20	1.0	2.5	2.5	2.0
Number of sites			4	4	3	5	4

Phasing statistics to 2.4 Å

R_{iso} (%)[3]	33.7 (33.9)	24.3 (25.6)	15.6 (18.9)	13.0 (12.9)	27.1 (30.7)
Phasing power[4]	0.61 (0.41)	0.86 (0.65)	0.37 (0.29)	0.74 (0.46)	0.59 (0.40)

Refinement statistics [numbers in parentheses refer to the outermost resolution bin (2.05-2.00 Å)]

Resolution range (Å)	16.2 – 2.00
R_{work}[5]	0.204 (0.222)
R_{free}[6]	0.261 (0.283)
Number of atoms or molecules	
Protein	5606
Water	400
R.m.s. deviation from ideal geometry	
Bond lengths (Å)	0.007
Bond angles (°)	1.271
Average B-factor (Å²)	15.6

[1]Values for the outermost resolution shell are given in parentheses.

[2]$R_{sym} = \Sigma|I - \langle I \rangle| / \Sigma I$, where I is the integrated intensity of a given reflection.

[3]$R_{iso} = \Sigma|F_{PH} - F_P| / \Sigma F_P$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively.

[4]Phasing power = <r.m.s. heavy atom structure factor> / <r.m.s. lack of closure>

[5]$R_{work} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$ for the 90% of the reflection data used in refinement.

[6]$R_{free} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$ for the remaining 10% of the reflection data excluded from refinement.

Figure 2-1A: Crystal structure of *EAV* nsp4. Ribbon diagram of the overall structure of *EAV* nsp4. β-strands of the N-terminal barrel (blue, A1 to F1) and the C-terminal barrel (green, A2 to G2) are indicated. The C-terminal extension domain is coloured in red. Members of the catalytic triad are shown in yellow. Residues Leu160, Leu171, Ile174, Ile178, Ile183, Ala184, and Val186 which comprise the solvent-exposed hydrophobic patch on the C-terminal extension domain are shown in black. Solvent-exposed Trp114 is also shown in black.

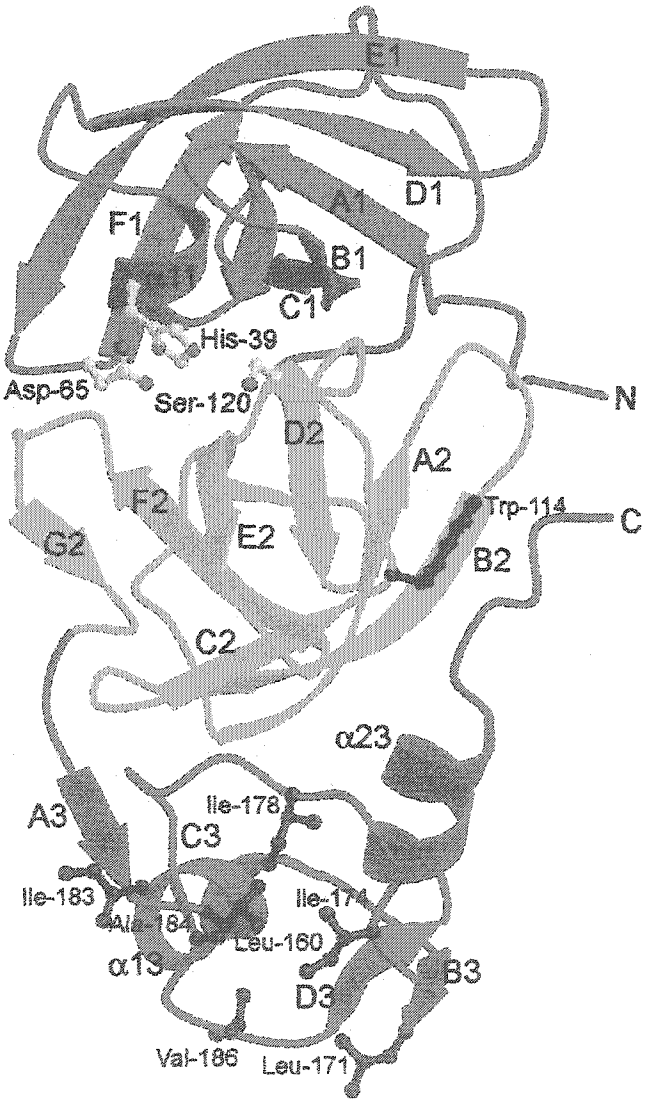


Figure 2-1B: Multiple sequence alignment of arteriviral 3C-like serine proteinases (for original references to sequences see (4)). Highly conserved residues are highlighted: hydrophobic residues at the interface between the N-terminal and the C-terminal β-barrels (mauve); hydrophobic core of the two β-barrels and the C-terminal extension domain (cyan); catalytic triad (red); hydrophobic residues at the interface between the C-terminal β-barrel and the C-terminal extension domain (dark blue); S1 pocket (yellow); Asp119, a buried and charged residue important for the conformation of the active site and the oxyanion hole (green). LDVC and LDVP, LDV neurovirulent type C and strain Plagemann, respectively; PRRSVLV and PRRSVVR, PRRSV strain Lelystad and strain ATCC VR-2332, respectively.

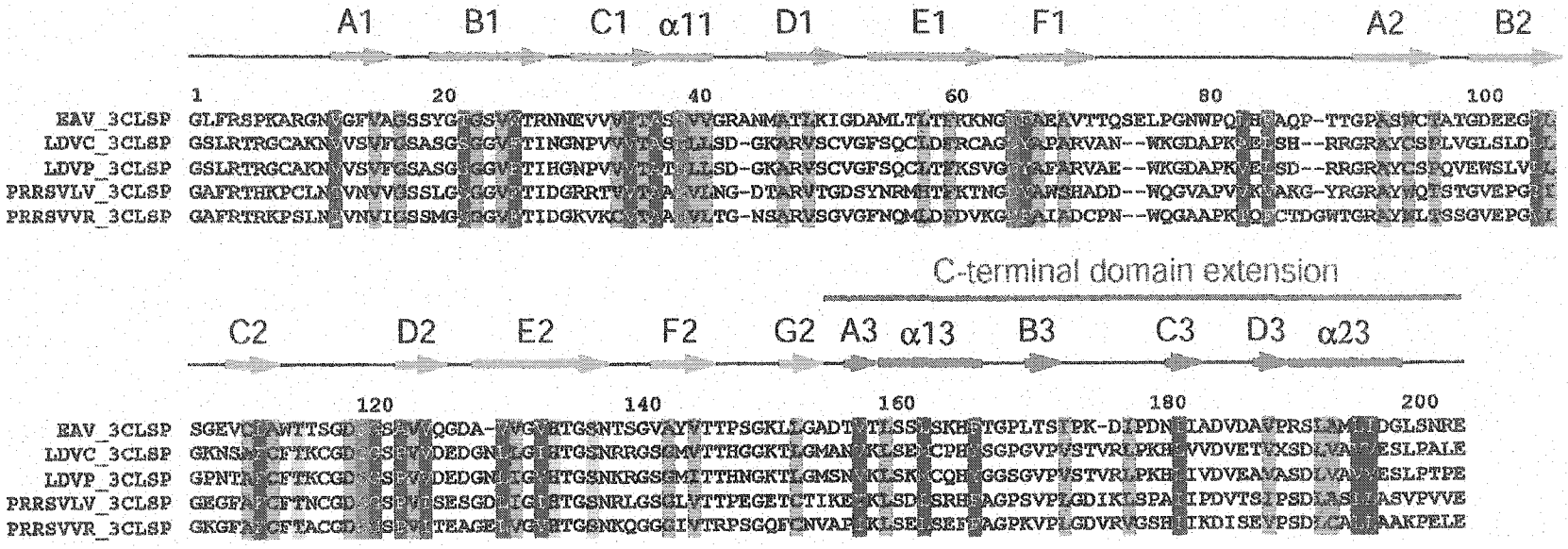


Figure 2-1C: Surface representation of nsp4. The highly conserved solvent-exposed hydrophobic residues in the C-terminal extension domain are shown in yellow whereas the catalytic triad is shown in red. Figure prepared using PYMOL (33).

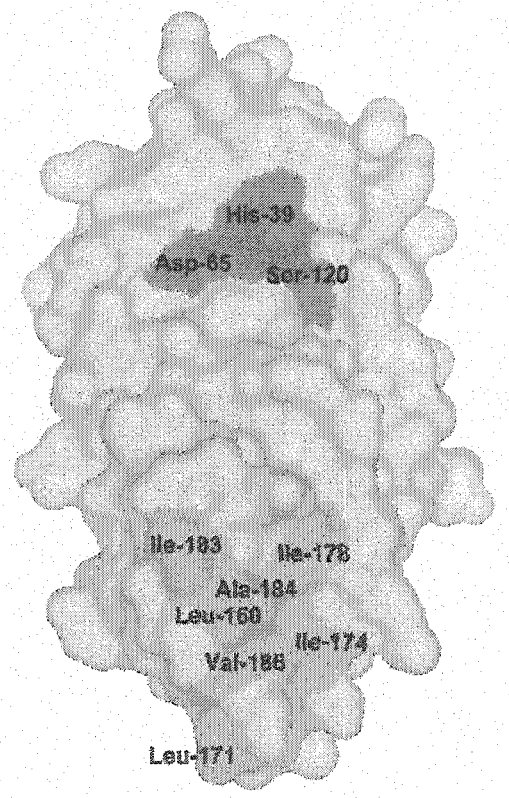


Figure 2-1D: Ribbon diagram of the superposition of the four copies of nsp4 in the asymmetric unit.

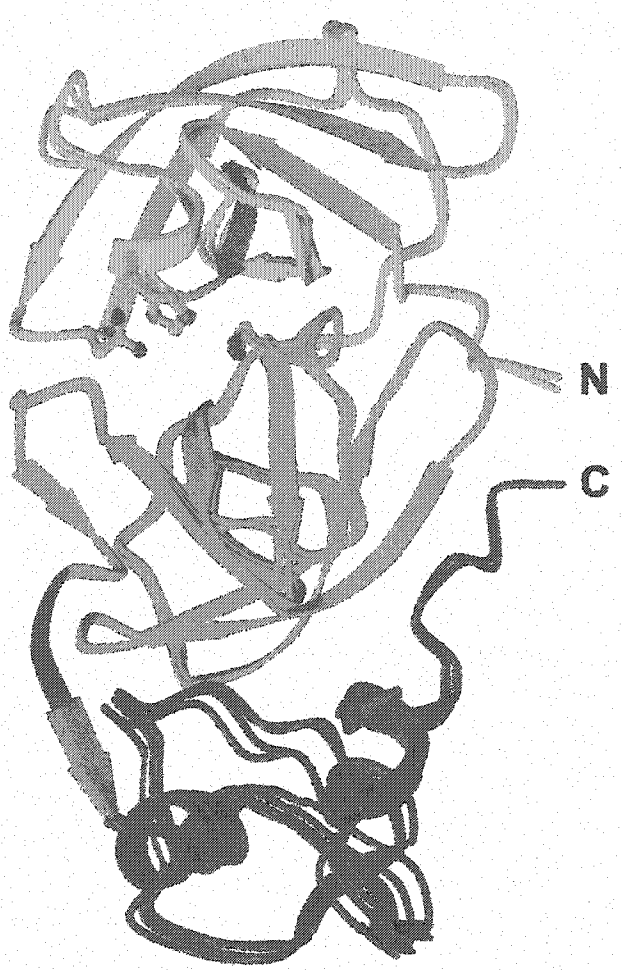


Figure 2-2A: Structural alignment between copy A of *EAV* nsp4 (shown in blue) and *Human Rhinovirus 2* 3C proteinase (PDB accession code 1CQQ) (shown in yellow).

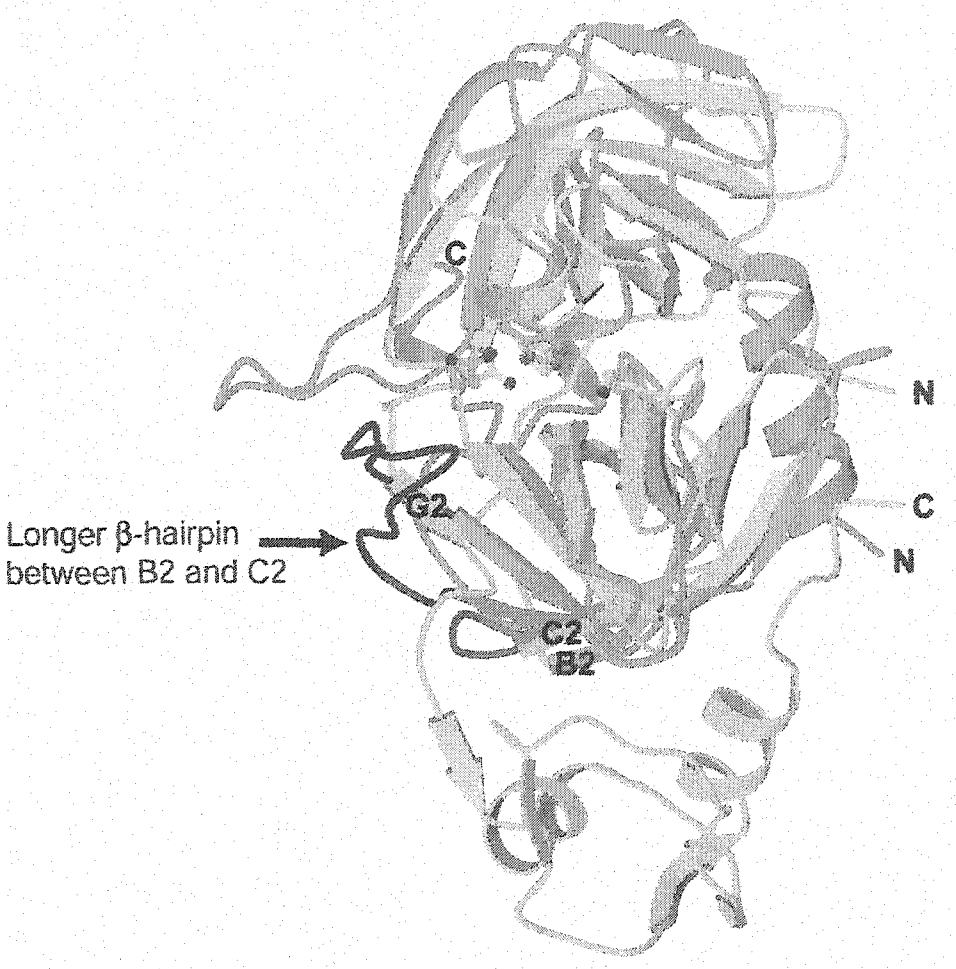


Figure 2-2B: Structural alignment between copy A of *EAV* nsp4 (shown in blue) and *Sindbis virus* core protein (PDB accession code 2SNV) (shown in yellow).

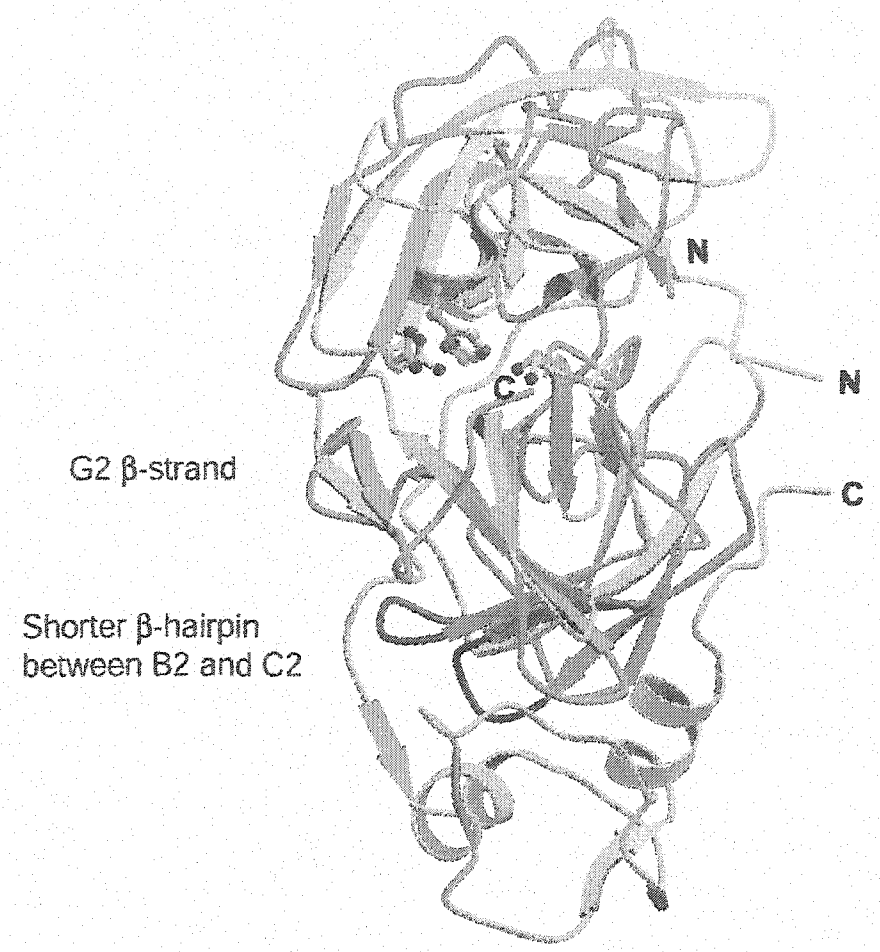


Figure 2-2C: Structural alignment between copy A of *EAV* nsp4 (shown in blue) and *Streptomyces griseus* proteinase E (PDB accession code 1HPG) (shown in yellow).

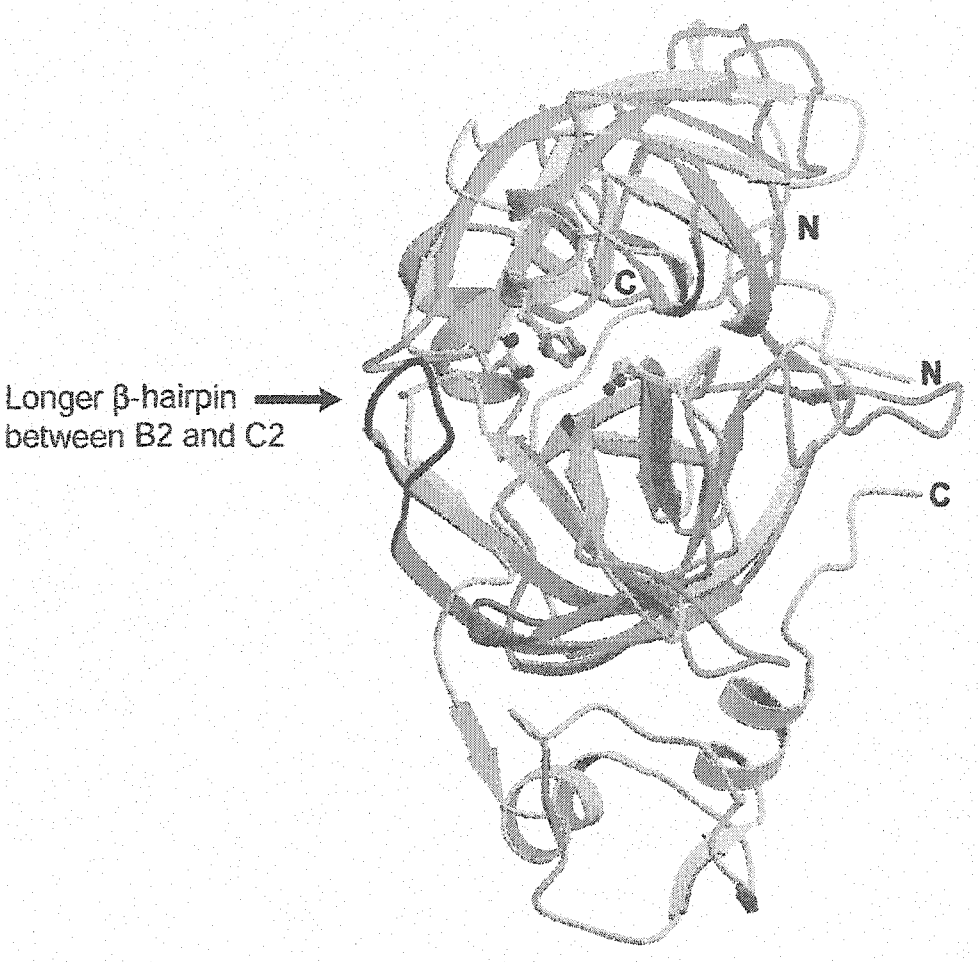


Figure 2-3A: The active site of *EAV* nsp4. Dashed black lines indicate hydrogen bonds. Superposition of the active site of all four copies in the asymmetric unit.

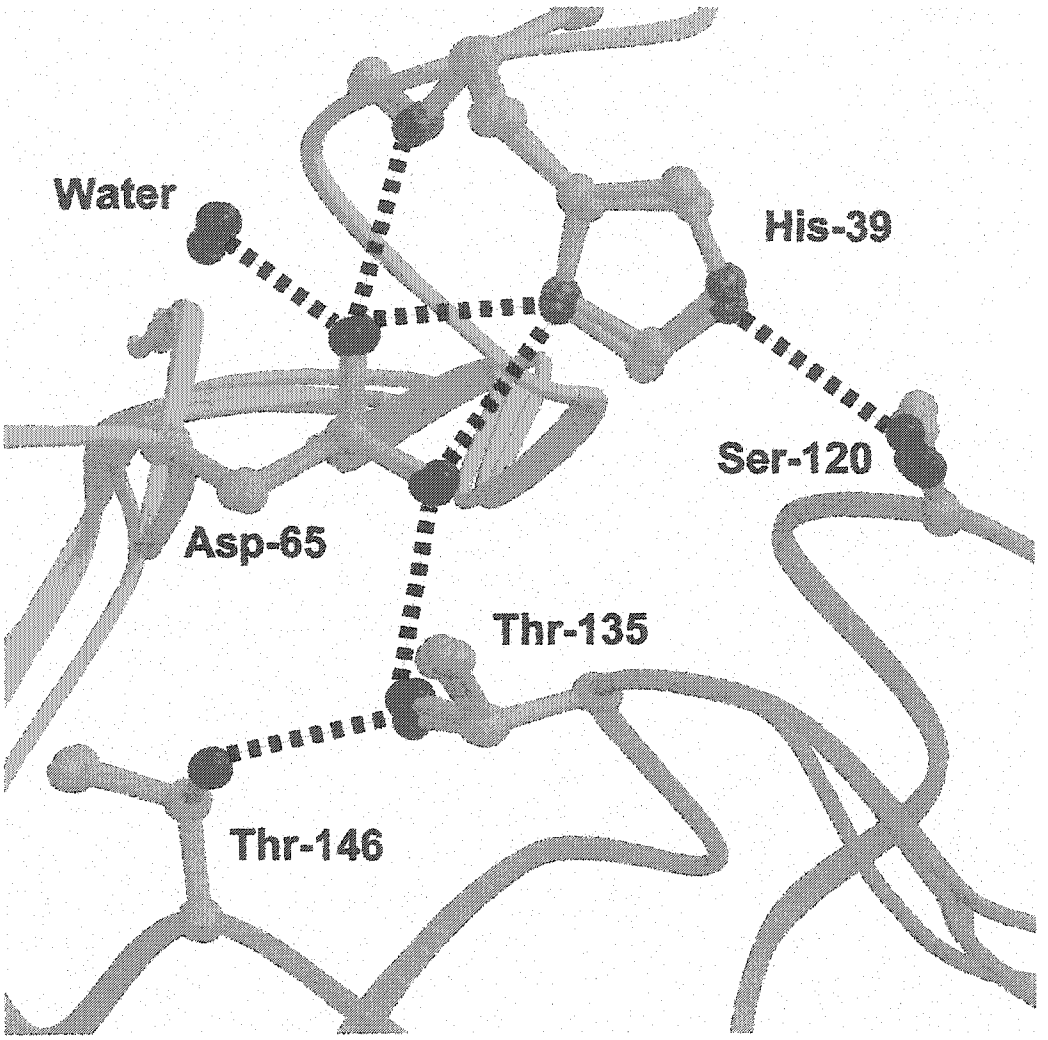
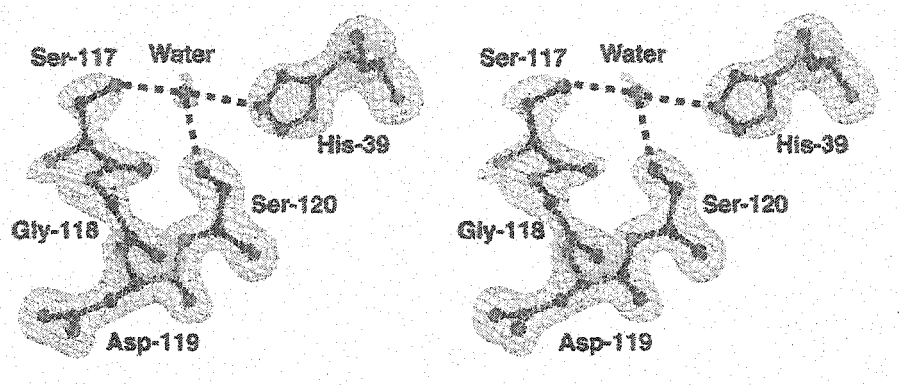


Figure 2-3: **(B)** Simulated annealing omit map for the collapsed oxyanion hole of copy A, calculated after omitting the labeled residues and refining the structure by the simulated annealing protocol, starting at 1000 K. Dashed black lines indicate hydrogen bonds.



(C) Simulated annealing omit map for copy B showing that the conformation of the peptide bond between residues 117 and 118 has flipped.

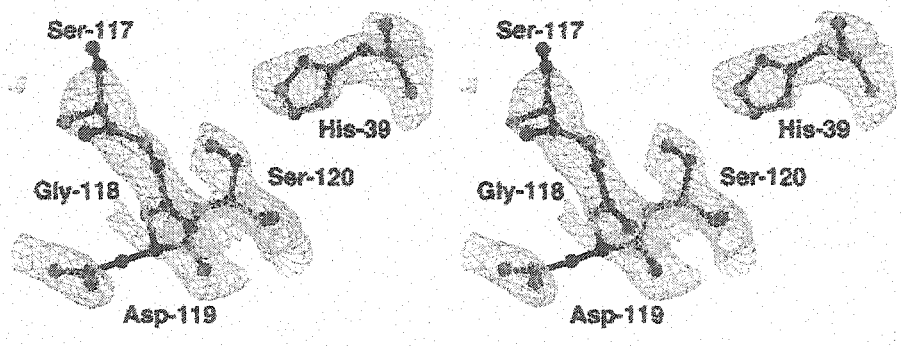
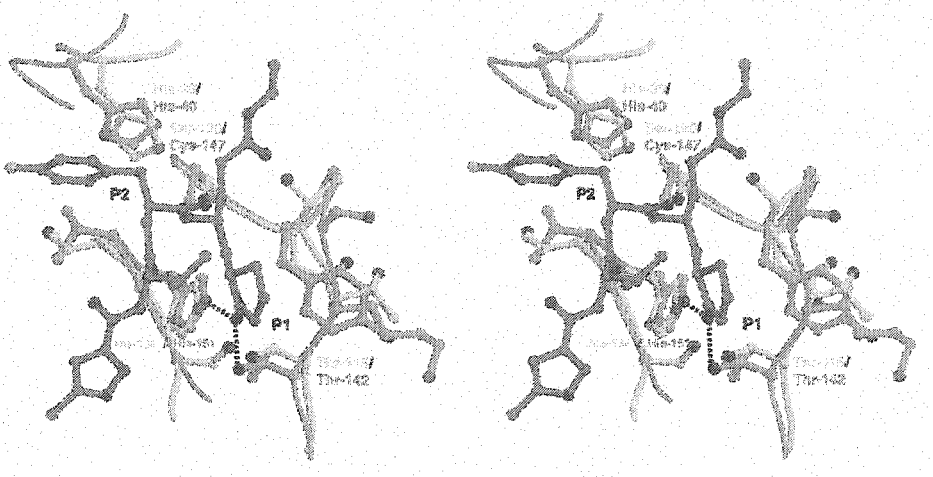
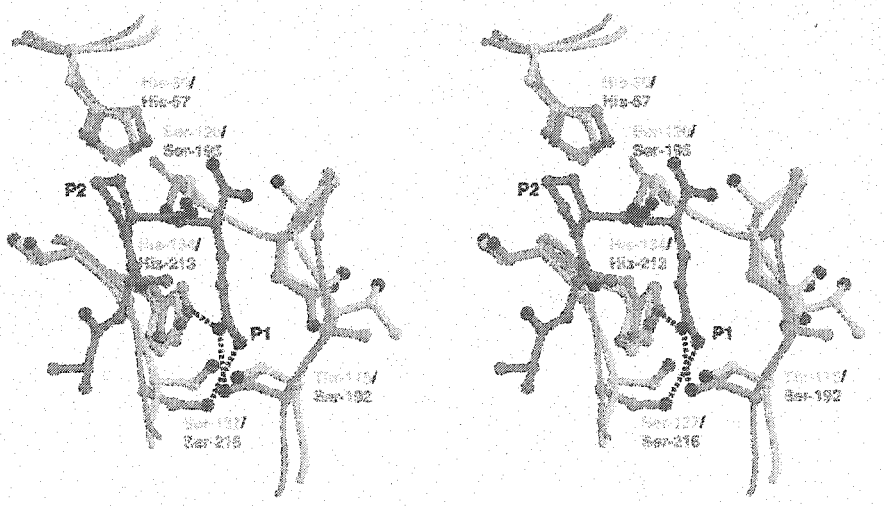


Figure 2-4: The S1 specificity pocket of *EAV* nsp4. Dashed black lines denote hydrogen bonds. Residues referred to in the text are labeled. Superposition of nsp4 (cyan) with **(A),** *HRV-2* 3C (yellow) bound to AG7088 inhibitor (magenta), and **(B),** *S. griseus* proteinase E (yellow) bound to tetrapeptide product (magenta).

(A)



(B)



2.3 References

1. Doll, E. R., Bryans, J. T., McCollum, W. H. M., and Wallace, M. E. (1957) *Cornell Vet.* 47, 3-41
2. Snijder, E. J., and Meulenberg, J. J. M. (2001) in *Fields Virology* (Knipe, D. M., and Howley, P. M., eds) Vol. 1, 4th Ed., pp. 1205-1220, 2 vols., Lippincott Williams and Wilkins, Philadelphia, PA
3. Snijder, E. J., and Meulenberg, J. J. (1998) *J Gen Virol* 79, 961-979.
4. Ziebuhr, J., Snijder, E. J., and Gorbalenya, A. E. (2000) *J Gen Virol* 81, 853-879.
5. den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C., and Spaan, W. J. (1991) *J. Virol.* 65, 2910-2920
6. Snijder, E. J., Wassenaar, A. L., and Spaan, W. J. (1994) *J. Virol.* 68, 5755-5764
7. Snijder, E. J., Wassenaar, A. L., van Dinten, L. C., Spaan, W. J., and Gorbalenya, A. E. (1996) *J Biol Chem* 271, 4864-4871.
8. Wassenaar, A. L., Spaan, W. J. M., Gorbalenya, A. E., and Snijder, E. J. (1997) *J. Virol.* 71, 9313-9322
9. van Dinten, L. C., Rensen, S., Gorbalenya, A. E., and Snijder, E. J. (1999) *J. Virol.* 73, 2027-2037
10. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307-326
11. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* 55, 849-861.
12. Terwilliger, T. C. (2001) *Acta Crystallogr D Biol Crystallogr* 57, 1755-1762.
13. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat Struct Biol* 6, 458-463.
14. McRee, D. E. (1999) *J. Structural Biology* 125, 156-165

15. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* 54, 905-921.

16. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr D Biol Crystallogr* 57, 122-133.

17. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) *Proteins* 12, 345-364.

18. Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* 381, 272.

19. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946-950

20. Esnouf, R. M. (1997) *J. Mol. Graph. Model.* 15, 132-134

21. Merritt, E. A., and Bacon, M. (1997) *Methods Enzymol.* 277, 505-524

22. Tong, L., Wengler, G., and Rossmann, M. G. (1993) *J. Mol. Biol.* 230, 228-247

23. Choi, H. K., Lu, G., Lee, S., Wengler, G., and Rossmann, M. G. (1997) *Proteins* 27, 345-359.

24. Hayward, S., and Berendsen, H. J. (1998) *Proteins* 30, 144-154.

25. Matthews, D. A., Dragovich, P. S., Webber, S. E., Fuhrman, S. A., Patick, A. K., Zalman, L. S., Hendrickson, T. F., Love, R. A., Prins, T. J., Marakovits, J. T., Zhou, R., Tikhe, J., Ford, C. E., Meador, J. W., Ferre, R. A., Brown, E. L., Binford, S. L., Brothers, M. A., DeLisle, D. M., and Worland, S. T. (1999) *Proc Natl Acad Sci U S A* 96, 11000-11007.

26. Nienaber, V. L., Breddam, K., and Birktoft, J. J. (1993) *Biochemistry* 32, 11469-11475.

27. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and Hilgenfeld, R. (2002) *Embo J* 21, 3213-3224

28. Hayward, S. (2001) *Protein Sci* 10, 2219-2227.

29. Gunasekaran, K., Gomathi, L., Ramakrishnan, C., Chandrasekhar, J., and Balaram, P. (1998) *J Mol Biol* 284, 1505-1516.

30. Allaire, M., Chernaia, M. M., Malcolm, B. A., and James, M. N. (1994) *Nature* 369, 72-76.

31. Stennicke, H. R., Birktoft, J. J., and Breddam, K. (1996) *Protein Sci* 5, 2266-2275.

32. Petersen, J. F., Cherney, M. M., Liebig, H. D., Skern, T., Kuechler, E., and James, M. N. (1999) *Embo J* 18, 5463-5475.

33. DeLano, W. L. (2002), DeLano Scientific, San Carlos, CA, USA

CHAPTER 3:

Structural Basis of Inhibition Revealed by a 1:2 Complex of the Two-headed Tomato Inhibitor-II and Subtilisin Carlsberg[1]

3.0 Introduction

PIs of the Potato II (Pot II) inhibitor family have been isolated from a wide range of commercially important crops including tomatoes, tobacco and potatoes (1-8). Pot II PIs can inhibit trypsin, chymotrypsin, subtilisin, oryzin and elastase (1,9), and accumulate systemically in plant tissue as a result of wounding or pest attack. Within 48 hours of insect attack or wounding, PIs can accumulate to levels of 2% or more of the total soluble protein in the leaves of tomato, potato and alfalfa plants (10,11) and are thought to have adverse effects on the digestive physiology of insects (12). The wide distribution and inducible expression of Pot II PIs in plants strongly suggest the fundamental importance of these proteins to the pest defence strategies of many commercially important crops.

Tomato inhibitor II (TI-II) contains two copies of a 54 amino acid repeat. High sequence identity exists among the Pot II PI sequences, although interesting variations (mostly in the reactive site loops) occur among different species and among different isoforms of the same species (Figure 1-7). These differences in sequence can give rise to a wide range of differences in cognate proteinase specificities. The expression of multiple isoforms of Pot II PIs with differing target proteinase specificities may be

---

[1] *A version of this chapter has been accepted for publication. Barrette-Ng, I.H., Ng, K.K.-S., Cherney, M.M., Pearce, G., Ryan, C.A., and James, M.N.G. (2003) J. Biol. Chem., electronically published on April 8, 2003 (in press). Protein purification was performed by Gregory Pearce and Dr. Clarence Ryan at Washington State University, and protein crystallization was performed by Maia Cherney at the University of Alberta.*

particularly important for protecting plants against predators with an arsenal of proteinases having a wide range of substrate specificity. Understanding the structural basis of PI inhibitory specificity is clearly crucial for understanding how variations in PI sequence give rise to different inhibitory specificities.

For the first time, the three-dimensional X-ray crystallographic structure of a ternary complex of a full-length two-headed inhibitor from the Pot II family, TI-II, bound to two molecules of subtilisin Carlsberg has been solved to 2.5 Å resolution. The novel three-dimensional structure of TI-II in a ternary complex with bacterial proteinases opens avenues of research previously difficult to undertake due to the lack of structural information on the mode of inhibition of two-domain Pot II inhibitors.

3.1 Experimental procedures

3.1.1 Purification and crystallization

Tomato inhibitor-II was prepared from transgenic tomato plants that overexpressed a prosystemin transgene, which resulted in the synthesis and accumulation of high levels of TI-II in the leaves (~1 mg/ml leaf juice). Leaves from 1000 young tomato plants were collected (~800 g leaf tissue) and blended with 1600 mL of a buffer containing 0.01 M sodium citrate, 0.5 M sodium chloride, and 0.7% sodium hydrosulfite, pH 4.3. The homogenate was expressed through 8 layers of cheesecloth by hand-squeezing, and the debris retained in the cloth was resuspended in 400 mL buffer and the mixture was expressed. The combined expressed liquid was clarified by centrifugation at 12,000 x g for 20 min. Proteins in the supernatant liquid were precipitated by adding solid ammonium sulfate to 80% saturation and stirred at 4°C for 2 hr. The precipitate was recovered by centrifugation as above. The pellet was solubilized in 600 mL water and the remaining debris removed by centrifugation. The resulting solution was placed in a flask and immersed in a boiling water bath with stirring until the temperature of the liquid was 70°C. The contents of the flask were then cooled rapidly in an ice bath to room temperature and the precipitated proteins repelleted by centrifugation. The solution was dialyzed overnight at 4° C against 0.01 M Tris, 0.10 M KCl buffer, pH 8.1. The retentate was passed through a 3 cm x 12 cm column containing chymotrypsin-Sepharose CL4B affinity resin (13) that had been equilibrated with the buffer. The column was then washed with 3 bed volumes of buffer and then with 8 M urea, pH 3, to elute the inhibitor proteins bound to chymotrypsin. The eluate was dialyzed against several changes of 50 mM ammonium carbonate and lyophilized and stored. The preparation yielded about 80

mg of dry material containing about 24 mg TI-II and 14 mg tomato inhibitor-I. TI-II was purified to homogeneity using reverse phase HPLC. A Vydac (Hesperia, CA) 218TP510 C18 semi-preparative column (10 x 250 mm, 5 micrometers, 300 Angstroms) was employed using a 90 min gradient of 0-60% acetonitrile in 0.1% TFA. TI-II eluted as a single peak at 53 min. The yields of inhibitor from this step were nearly quantitative. The TI-II protein obtained was homogeneous as judged by SDS-PAGE.

Crystals were grown at room temperature by the hanging drop vapour diffusion method by mixing 2 μL of a 1:1 mixture of TI-II and subtilisin Carlsberg (Sigma) (1 mM each) and 2 μL of 18 % (w/v) PEG 8000, 200 mM sodium malonate, 75 mM sodium citrate, pH 6.0, 10 % (w/v) ethylene glycol, and 5% (v/v) ethanol. Very small crystals grew within one to two weeks. Macroseeding was used to produce crystals with typical dimensions of 0.2 x 0.1 x ~0.01 mm. The crystals belonged to space group C2 with cell dimensions a = 155.0 Å, b = 55.1 Å, c = 91.3 Å, $\alpha = 90°$, $\beta = 119.5°$ and $\gamma = 90°$.

3.1.2 Data collection, structure solution and refinement

Data were collected on flash-frozen crystals at 100 K by transferring crystals briefly into a cryoprotectant solution identical to the crystallization solution, with the exception of an increased concentration of ethylene glycol (final concentration of 25% (w/v)). After a 3-5 second soaking time, the crystal was transferred directly into a nitrogen gas stream. Data were measured from a MAR 345 detector using 1.08 Å wavelength radiation at beamline 7-1 (Stanford Synchrotron Radiation Laboratory). Data were recorded as 220 1° oscillations at a crystal-to-detector distance of 220 mm and processed using the HKL suite of programs (14). Intensities were converted to amplitudes using TRUNCATE (15).

The solvent content of the proteinase-inhibitor complex crystals was calculated to be either 70% or 50% if either one or two copies of subtilisin were present in the asymmetric unit, along with one copy of TI-II. Molecular replacement calculations were carried out using AMoRe (16). The structure of subtilisin Carlsberg (PDB code 1SCN), with all the side chains and temperature factors retained, was used as the search model for molecular replacement calculations. Cross-rotation functions calculated using data to a maximum resolution of 3 Å gave two solutions (Patterson correlation coefficients of 0.32 and 0.24). Each rotation function solution was used to calculate a translation function using data to a maximum resolution of 3 Å, yielding two solutions (R-factors of 0.485 and 0.503 after rigid-body refinement). The relative positions of the two subtilisin molecules in the asymmetric unit were determined using a phased translation function, yielding a single solution (R-factor of 0.396). Rotation and translation searches using a search model for either a single domain or both domains of the inhibitor, as constructed from PCI-I (PDB code 4SGB) (17), did not reveal any clear solutions.

The electron density map calculated using the molecular replacement solutions of the two subtilisin molecules revealed a region of electron density adjacent to the proteinase active sites that had the features expected for TI-II. A search model for a single domain of TI-II was constructed using the coordinates of PCI-I (PDB code 4SGB) (17), and two single-domain search models could be placed by manual inspection using the molecular graphics program XFIT (18). Side chains that differed between the PCI-I search model and TI-II were modified, and the model was further adjusted by manual inspection of the electron density map. The model was subjected to iterative rounds of manual model building using XFIT and refinement against a maximum-likelihood target

using CNS (19). Analysis of interdomain and protein-inhibitor interfaces as well as buried surface area calculations were performed using CNS.

The final model contains two complete copies of subtilisin, and residues 1 to 73 and 86 to 116 of TI-II. Residues 74 to 85 and 117 to 121 of TI-II appear to be disordered and there was no electron density corresponding to these regions of the structure. 83.6 % of residues lie in the most favoured regions of the Ramachandran plot and no residues are in disallowed regions (regions defined by PROCHECK) (20). Additional checks on geometry were performed using WHATCHECK (21). Data quality and refinement statistics are given in Table 3-1. Figures 3-1, 3-2, 3-3a and 3-5 were prepared with MOLSCRIPT (22), BOBSCRIPT (23) and RASTER3D (24). Figures 3-3b and 3-3c were prepared with PyMOL (25).

3.2 Results and discussion

3.2.1 Structure of the TI-II:(Subtilisin)$_2$ complex

The structure of the TI-II:(Subtilisin)$_2$ complex was determined by the molecular replacement technique to 2.5 Å resolution, using the coordinates of subtilisin Carlsberg (26) (PDB code 1SCN) as the search model. There are two molecules of subtilisin bound to one molecule of the two-headed inhibitor TI-II (Figure 3-1; refer to Table 3-1 for crystallographic parameters). The two subtilisin molecules are bound at opposite ends of the elongated inhibitor molecule, forming a 673-residue ternary complex with dimensions of roughly 100 Å x 50 Å x 40 Å.

TI-II has a novel two-domain structure; each domain adopts the fold previously described in the single-domain Pot II inhibitors (17,27-29). Each domain contains only a small amount of regular secondary structure in the form of a three-stranded antiparallel β-sheet, as well as a series of stretches of polypeptide chain interconnected by four disulphide bonds. As first predicted after the determination of the structure of the single-domain Pot II inhibitor PCI-I, the sequence repeats present in TI-II do not correspond to individual structural domains (Figure 1-7) (17). Instead, Domain I consists of the first fifteen residues and residues 86 to 116, whereas Domain II consists of residues 16-73. Even though Domain I is composed of two non-contiguous polypeptide sequences, it adopts the same fold as Domain II (r.m.s.d. = 0.86 Å, 41 C_α atom pairs; Figure 3-2a). The β-sheet scaffold of Domain I is similar to that seen in the single-domain C2 inhibitor from *N. alata* (r.m.s.d. = 0.99 Å, 19 C_α atom pairs; Figure 3-2b). Both Domain I and C2 consist of two noncontiguous segments of polypeptide connected through disulphide bonds, giving rise to a 'clasped bracelet' fold that is thought to be characteristic of

inhibitors of the Pot II family (29). Although Domain I shares the same folding topology and aligns very well with C2 in the β-sheet region, the rest of the domain, including the reactive site loops, differs from C2 substantially in conformation. These differences in conformation between Domain I and C2 reflect intrinsic structural differences between TI-II and C2, as well as the inability of NMR structure calculations to define precisely the conformation of residues A1 to A9 and B1 to B12 of C2 (Table 1 from ref. (29)).

The arrangement of the two domains in TI-II gives rise to an extended structure with overall dimensions of approximately 50 Å x 25 Å x 15 Å. The arrangement of the two domains differs substantially from an earlier prediction based on the structure of PCI-I, in which a pseudo 2-fold rotation axis relates the two domains across a continuous six-stranded antiparallel β-sheet (17). Instead, the three-stranded β-sheet of Domain I is packed against the C-terminus of the reactive site loop of Domain II. The arrangement of the two domains in TI-II also differs substantially from that proposed for the six-domain inhibitor precursor from *N. alata* based on the structure of the single-domain C2 inhibitor (29).

The interdomain interface in TI-II consists of a small cluster of highly conserved hydrophobic residues (Ile14, Pro16, Tyr98, Phe100 and Phe106 from Domain I; and Tyr34, Pro54 and Lys55 from Domain II). Although this interface is quite small (buried surface area of 487 $Å^2$), it appears to form a stable packing arrangement between the two domains. Preliminary data from a crystal form of the unbound inhibitor containing four copies of TI-II in the asymmetric unit indicate that the orientation of the two domains in TI-II observed in the TI-II:$(Subtilisin)_2$ complex is similar to that seen for the unbound inhibitor (see Chapter 4). The arrangement of domains observed in TI-II also suggests

how domains may be oriented relative to each other in other multidomain Pot II inhibitors, such as the six-domain inhibitor from *N. alata* (8).

3.2.2 Enzyme-inhibitor interactions

"Standard mechanism, canonical" proteinaceous inhibitors of serine proteinases from at least eighteen nonhomologous families (30) bind to the active sites of proteinases in a substrate-like manner (31) through the 'reactive site loops' (32). According to the nomenclature of Schechter and Berger (33), where P1 is the residue N-terminal to the scissile bond and P1' is the residue C-terminal to the scissile bond, residues P4 to P2' of each inhibitor almost always interact with binding pockets S4 to S2' on the proteinase. The reactive site loop adopts an approximate extended conformation and forms a distorted antiparallel β-strand pair with residues near the active site of the proteinase (34).

Each reactive site loop in TI-II interacts with a separate molecule of subtilisin in the standard, canonical manner as observed in other proteinase-inhibitor complexes. Main chain torsional angles in the reactive site loops are highly similar to those seen in other PIs (Table 3-2). The conformations adopted by the reactive site loops of each domain in TI-II are very similar with an r.m.s.d. of 0.49 Å for the 32 main-chain atoms of the residues at positions P5 to P3' (Figure 3-3a). As in other members of the Pot II family, disulfide bonds formed by cysteine residues at the P3 and P2' positions (Cys3 and Cys7 in Domain I and Cys60 and Cys64 in Domain II) help to hold the reactive site loop in a relatively rigid conformation that likely helps to prevent proteolytic cleavage of the inhibitor upon interaction with proteinases. Additional features of the structure of TI-II that contribute to the stability of both reactive site loops is the presence of a hydrogen bond between the side chains of the P2 and P1' residues (Thr4 and Glu6 in Domain I, and

Thr61 and Asn63 in Domain II), which is also seen in many other PIs (35). In addition, the $N^{\delta 2}$ atom of the side chain of Asn29 donates hydrogen bonds to the main chain carbonyl oxygen atoms of Thr61 and Asn63 at the P2 and P1' positions of the reactive site loop of Domain II, and the main chain carbonyl oxygen atom of Ile28 accepts a hydrogen bond from the main chain amide nitrogen atom of Asp65. The reactive site loop in Domain I is less well stabilized, with a van der Waals contact between the side chains of Thr89 and Glu6 replacing the hydrogen bonds donated by Asn29 in Domain II.

Residues in the P5 to P2' positions and the P6 to P2' positions of the reactive site loops from Domains I and II, respectively, interact with the substrate-binding clefts of separate subtilisin molecules (Figures 3-3b and 3-3c). Specific interactions between the reactive site loops of TI-II and subtilisin are illustrated schematically in Figure 3-4. A number of the inhibitor-enzyme contacts consist of van der Waals interactions, with the P1 residues contributing the largest number of contacts. In addition, the polypeptide backbone of the reactive site loops of TI-II forms a number of hydrogen bonds with the polypeptide backbone of subtilisin that are characteristic of standard, canonical inhibitor:proteinase complexes (Figures 3-4a and 3-4b).

Subtilisin has a broad substrate promiscuity compared to most other serine proteinases, mostly due to the relatively shallow S1 "specificity" pocket. The highly accommodating nature of the S1 pocket is dramatically illustrated by the binding of two large P1 residues, arginine and phenylalanine, in the TI-II complex. Although the binding mode of the reactive site loop in both domains is similar to that seen in other proteinase-inhibitor complexes (Table 3-2), there are some notable differences. The arrangement of the active site residues in the subtilisin molecule bound to Domain I is

very similar to that seen in subtilisin Carlsberg and subtilisin Novo BPN' bound to eglin-C (36-38), chymotrypsin inhibitor-2 (37) and *Streptomyces* subtilisin inhibitor (39,40). In the subtilisin molecule bound to Domain II, however, the active-site Ser-221 side chain lies closer to the carbonyl group of the P1 residue of the inhibitor than in other proteinase:inhibitor complexes. As a result, the active-site Ser-221 residue adopts a rotamer (χ_1 = -153°) in the molecule of subtilisin bound to Domain II that differs from the rotamer (χ_1 = -99°) seen in the molecule of subtilisin bound to Domain I, as well as in other subtilisin:inhbitor complexes (Figure 3-5).

The surface area of the inhibitor that is buried upon complex formation is 611 $Å^2$ for Domain I and 970 $Å^2$ for Domain II. Each domain of TI-II appears to bind the proteinase independently of the other domain, with the exception of a small number of additional contacts outside of the reactive site loops (Table 3-3). The reactive site loops of Domains I and II contribute 547 $Å^2$ and 611 $Å^2$ of buried surface area respectively, accounting for 90% and 62% of each domain's binding interface with subtilisin. Preliminary modelling studies in which other proteinases are docked onto the reactive site loops of TI-II suggest that TI-II may contact other proteinases at a wider range of sites outside of the reactive site loop region. For subtilisin and likely also for other proteinases, the two molecules of proteinase bound to the inhibitor are separated by a large distance and do not contact each other.

The structure of the TI-II:(Subtilisin)$_2$ ternary complex reveals for the first time the structural basis of inhibition by a multidomain Pot II family inhibitor, and shows a number of interesting similarities and differences with ternary complexes formed between the structurally unrelated Bowman-Birk family of proteinase inhibitors (BBIs)

and trypsin (41-43). Like TI-II, the structurally simpler BBIs also contain two domains, each of which presents a reactive-site loop at the extremities of the elongated inhibitor molecule. Also as seen in TI-II, a single BBI molecule can bind two proteinase molecules simultaneously. The fold of each domain and the arrangement of the two inhibitory domains of the BBIs differ strikingly from TI-II, however. The BBIs consist of two domains that are related by an approximate two-fold rotational symmetry axis, and each domain interacts in a very similar manner with target proteinases. Interactions between the BBIs and proteinases are almost completely restricted to the reactive-site loop, with buried surface areas of approximately 650-700 $Å^2$. These values are similar to that seen for Domain I in TI-II, but significantly smaller than that seen for Domain II (970 $Å^2$), where non-reactive site loop interactions account for roughly one-third of the inhibitor:proteinase interface.

Secondary contacts not involving the reactive-site loop in Domain II of two-domain Pot II family inhibitors may be of critical importance to determining proteinase inhibition specificity, as suggested by the results of site-directed mutagenesis studies on PI-II from potato (44). In PI-II, the capacity for trypsin inhibition in Domain II could not be transferred to Domain I by mutation of the P1 residue, the P2-P1-P1' sequence or even the entire stretch of sequence from P2 to P10'. These results underline the importance of regions of the inhibitor apart from the reactive site loop in determining inhibitor specificity. The determination of the structure of the TI-II:$(Subtilisin)_2$ complex provides an important step towards understanding the roles of secondary binding surfaces as well as reactive site loops in the mechanism of inhibition in the Pot II family of proteinase inhibitors.

Coordinates. Coordinates have been deposited in the Protein Data Bank (accession code 1OYV).

Table 3-1: Crystallographic Statistics

a) Data Processing

Resolution (Å)	40.0 - 2.5
Total reflections[1]	97405 (7394)
Unique reflections[1]	23099 (2173)
Redundancy[1]	4.2 (3.4)
Completeness (%)[1]	98.3 (94.1)
I/σ [1]	11.6 (2.4)
R_{sym} (%)[1,2]	12.3 (47.3)

b) Refinement

Resolution (Å)	40.0 - 2.5
R_{work}[3]	0.214 (0.321)
R_{free}[4]	0.270 (0.399)
Number of atoms	
Protein	4623
Water	128
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.007
Bond angles (°)	1.37
Avg. Temp. Factors (Å^2)	
TI-II	39.2
Subtilisin copy A	26.2
Subtilisin copy B	35.2
Water	28.0

[1]Values for the outermost resolution shell (2.59 - 2.50 Å) are given in parentheses.
[2]$R_{sym} = \Sigma_h\Sigma_i (|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|) / \Sigma_h\Sigma_i I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the i^{th} integrated intensity of a given reflection and $\langle I(\mathbf{h})\rangle$ is the weighted mean of all measurements of $I(\mathbf{h})$.
[3]$R_{work} = \Sigma_h ||F(\mathbf{h})_o| - |F(\mathbf{h})_c|| / \Sigma_h |F(\mathbf{h})_o|$ for the 90% of reflection data used in refinement.
[4]$R_{free} = \Sigma_h ||F(\mathbf{h})_o| - |F(\mathbf{h})_c|| / \Sigma_h |F(\mathbf{h})_o|$ for the 10% of reflection data excluded from refinement.

Table 3-2: Main chain torsion angles of reactive site loops

	P4		P3		P2		P1		P1'		P2'		P3'	
	φ	ψ	φ	ψ	φ	Ψ	φ	Ψ	φ	ψ	φ	ψ	φ	ψ
TI-II (D1)	-120	147	-126	158	-62	132	-85	36	-90	156	-92	125	-77	168
TI-II (D2)	-84	136	-134	152	-67	145	-97	49	-115	154	-81	143	-93	111
PCI-I	-82	132	-137	152	-65	151	-115	51	-102	165	-95	121	-71	132
OMTKY3[1]	-158	158	-126	147	-69	162	-119	45	-84	155	-100	115	-147	92
OMTKY3[2]	-129	136	-131	150	-68	160	-107	32	-74	159	-113	107	-141	76
CI-2	-93	140	-133	166	-64	147	-103	34	-91	146	-106	119	-118	113
Eglin-C	-76	138	-143	165	-65	151	-112	42	-96	176	-120	113	-119	115
BPTI	78	175	-77	-30	-70	156	-117	39	-88	164	-112	80	-106	122
sBBI[3]	-121	134	-139	145	-71	156	-95	45	-96	162	-108	97	-88	171

[1]Ovomucoid third domain bound to *Streptomyces griseus* proteinase B (45).
[2]Ovomucoid third domain bound to α-chymotrypsin (46).
[3]Soybean Bowman-Birk Inhibitor bound to bovine trypsin (42).

Angles for PCI-I (polypeptide chymotrypsin inhibitor-I from potatoes), OMTKY3, CI-2 (chymotrypsin inhibitor-2 from barley), eglin-C and BPTI (bovine pancreatic trypsin inhibitor) were obtained from ref. (17).

Table 3-3: Secondary contacts between TI-II and Subtilisin

Van der Waals contacts (less than 4.0 Å) are shown in regular type face and hydrogen-bonding contacts are shown in bold face.

TI-II	Subtilisin Carlsberg
F12	Y167, K170
C27	N155, F189
Y34	A129
C87	N155, F189
C91	S101
F100	S130
D103	K136, **Y171**

Figure 3-1A: Structure of the TI-II:(Subtilisin)$_2$ complex. Subtilisin molecules are drawn in yellow, and the two domains of TI-II are drawn in red and blue.

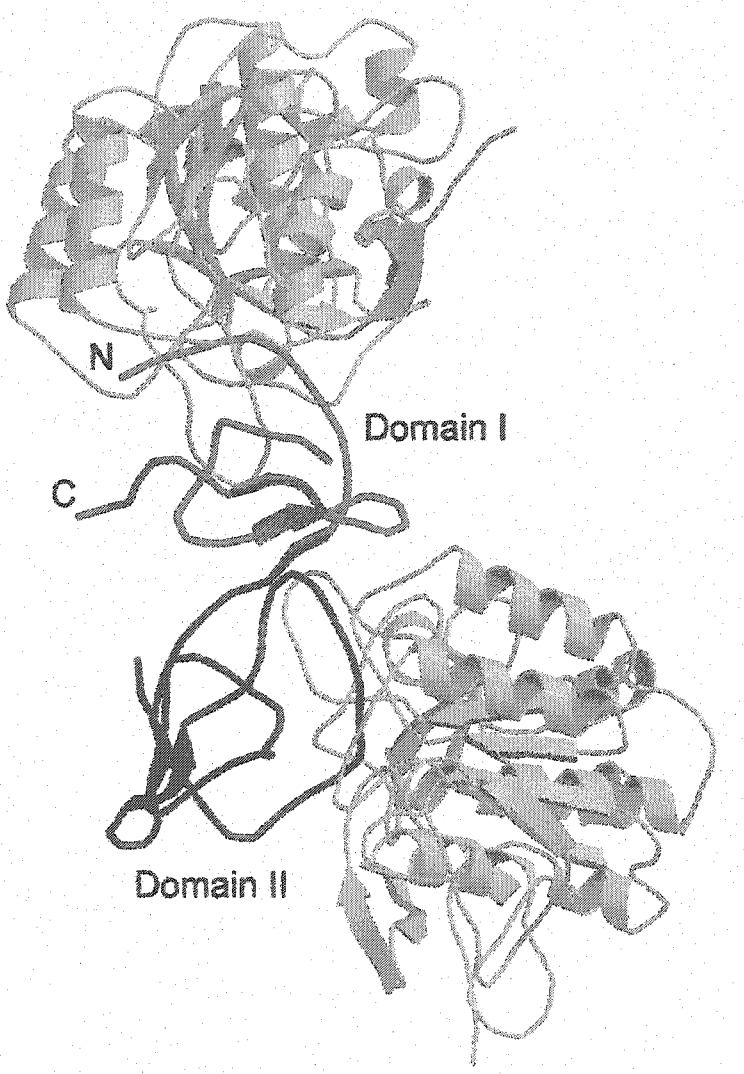


Figure 3-1B: Stereoscopic view of the structure of TI-II. C_α trace of TI-II, with cysteine residue side chains and disulfide bonds drawn in yellow and given residue numbers. Residues 74 to 85 and 117 to 123 are missing from the final model, because these portions of the structure were not defined in electron density maps. The approximate location of these residues is designated by dots drawn in magenta.

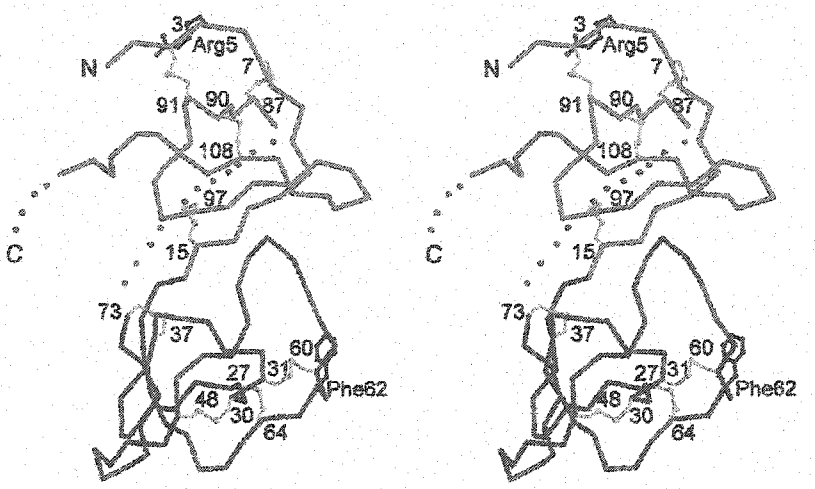
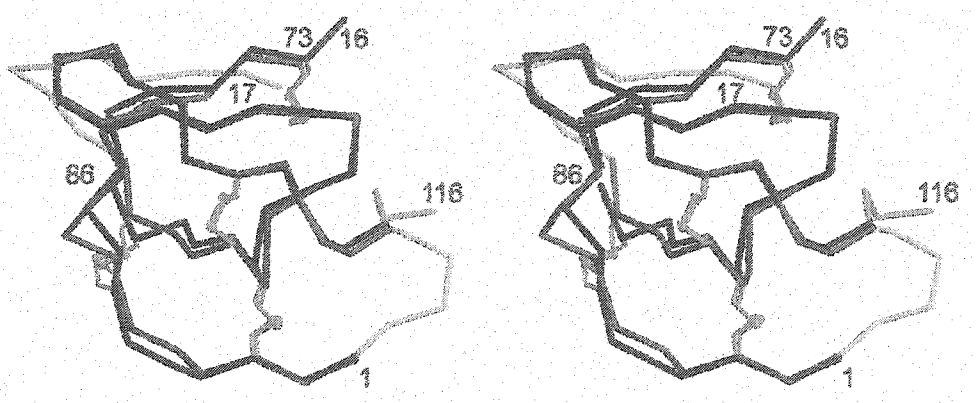


Figure 3-2: **(A)** Stereoscopic view of a least-squares superposition of TI-II Domains I and II. C_α atoms of Domains I and II that align within 3.5 Å of each other are drawn in red and blue respectively, whereas C_α atoms that do not align are drawn in magenta and cyan respectively. The N- and C-terminal residues of the polypeptide segments in each domain are labeled. **(B)** Stereoscopic view of a least-squares superposition of TI-II Domain I and the C2 inhibitor from *N. alata* (29). C_α atoms of Domain I and C2 that align within 3.5 Å of each other are drawn in red and blue respectively, whereas C_α atoms that do not align are drawn in magenta and cyan respectively. The N- and C-terminal residues of polypeptide segments in each domain are labeled.

(A)



(B)

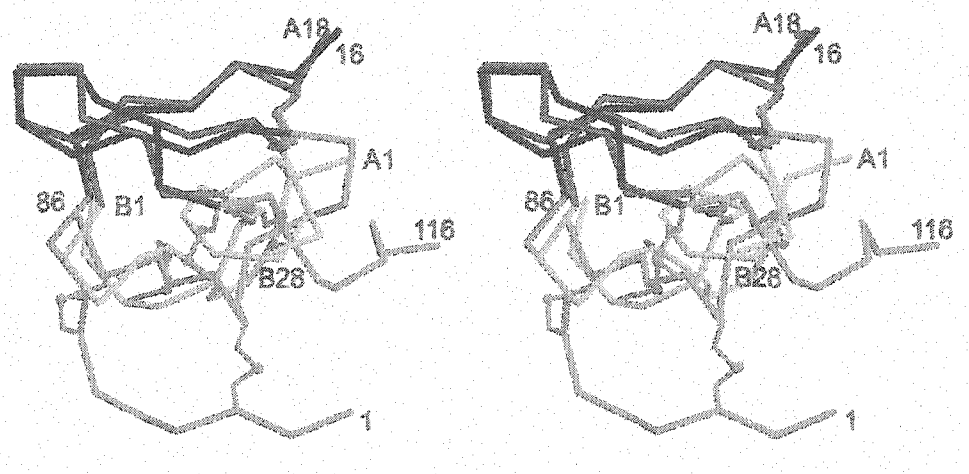
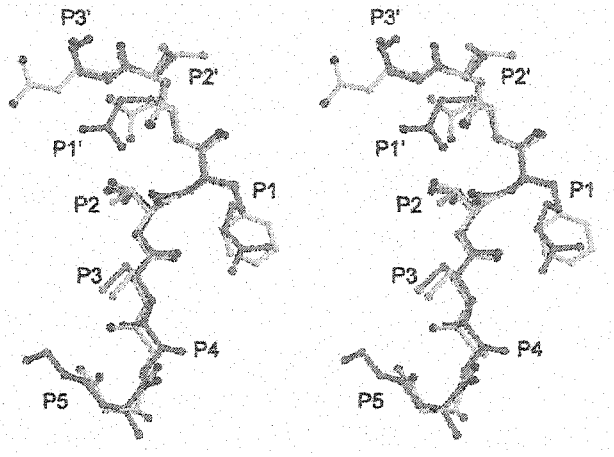
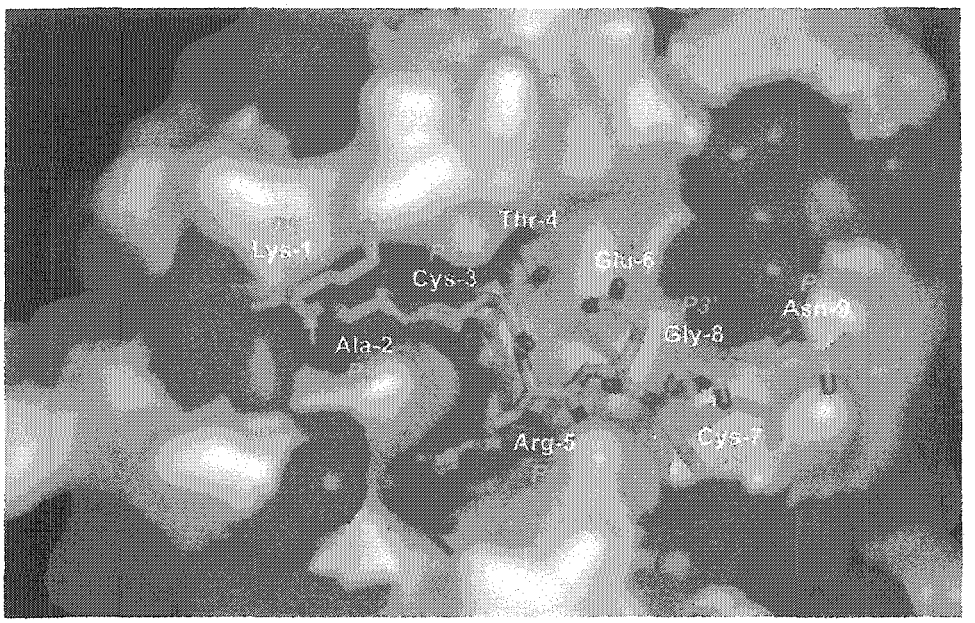


Figure 3-3A: Conformations of the reactive site loops. Stereoscopic view of a superposition of the reactive site loops of both domains.



Figures 3-3B and 3-3C: Conformation of the reactive site loops of **(B)**, Domain I and **(C)**, Domain II bound to subtilisin. The solvent-accessible surface of the subtilisin molecules are drawn, with negatively charged residues coloured red, positively charged residues coloured blue and hydrophobic residues coloured magenta.

(B)



(C)

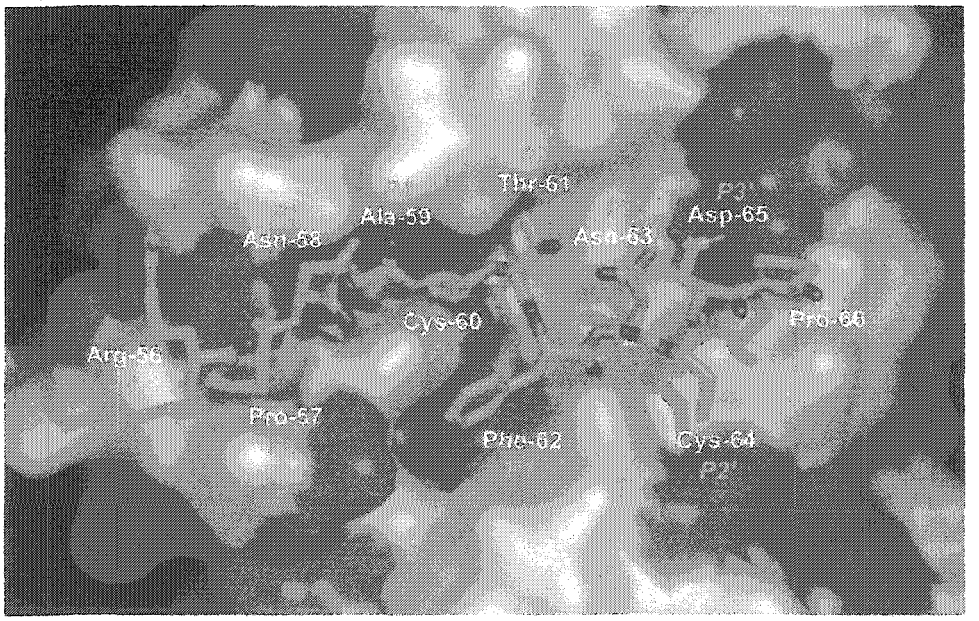


Figure 3-4A: Interactions between subtilisin and TI-II. Stylized representation of the interactions between subtilisin and the reactive site loops of Domain I. Residues of subtilisin making Van der Waals interactions with the reactive site loops of TI-II are shown as blue circles as well as in some instances as parallel lines.

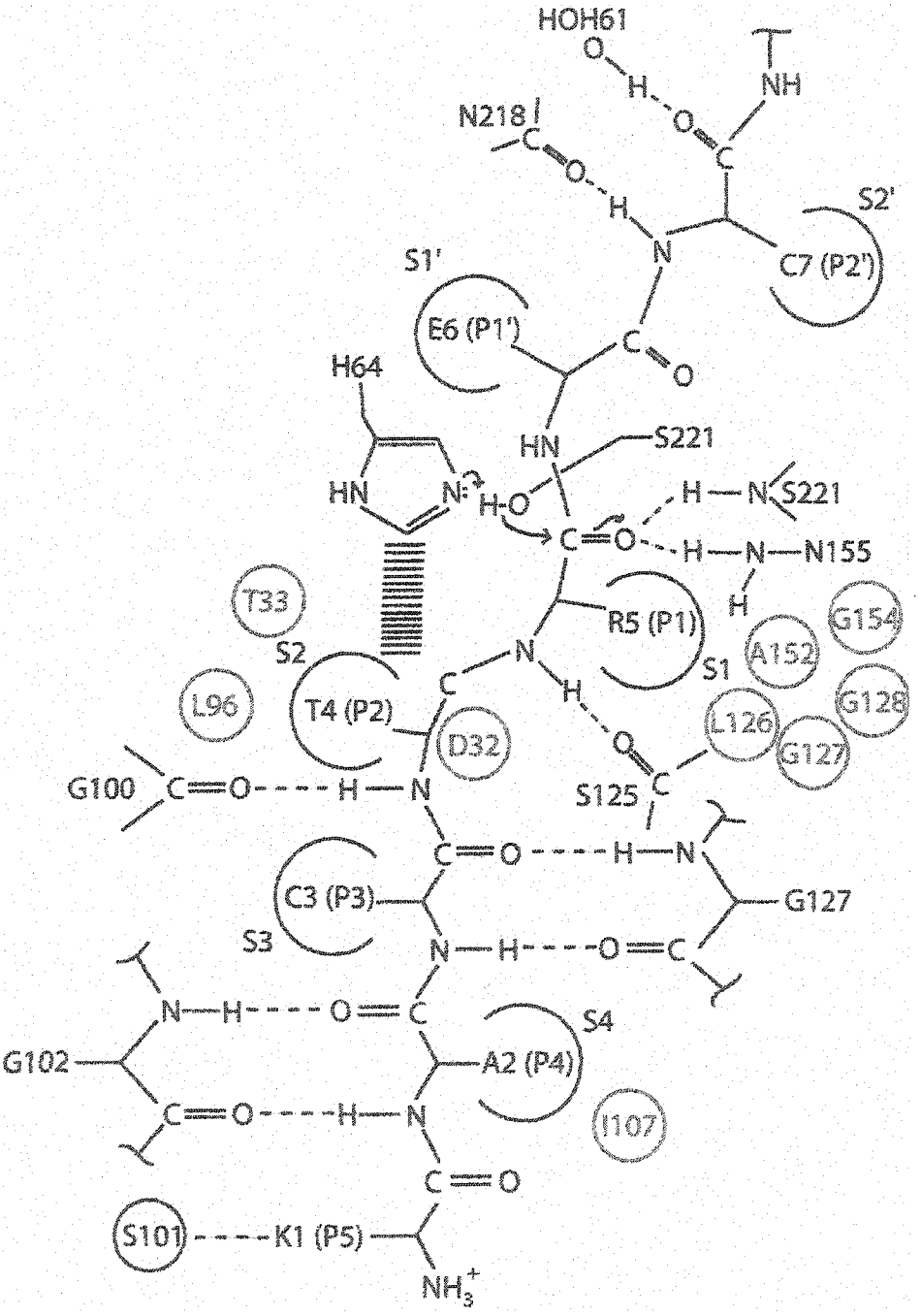


Figure 3-4B: Interactions between subtilisin and TI-II. Stylized representation of the interactions between subtilisin and the reactive site loops of Domain II. Residues of subtilisin making Van der Waals interactions with the reactive site loops of TI-II are shown as blue circles as well as in some instances as parallel lines.

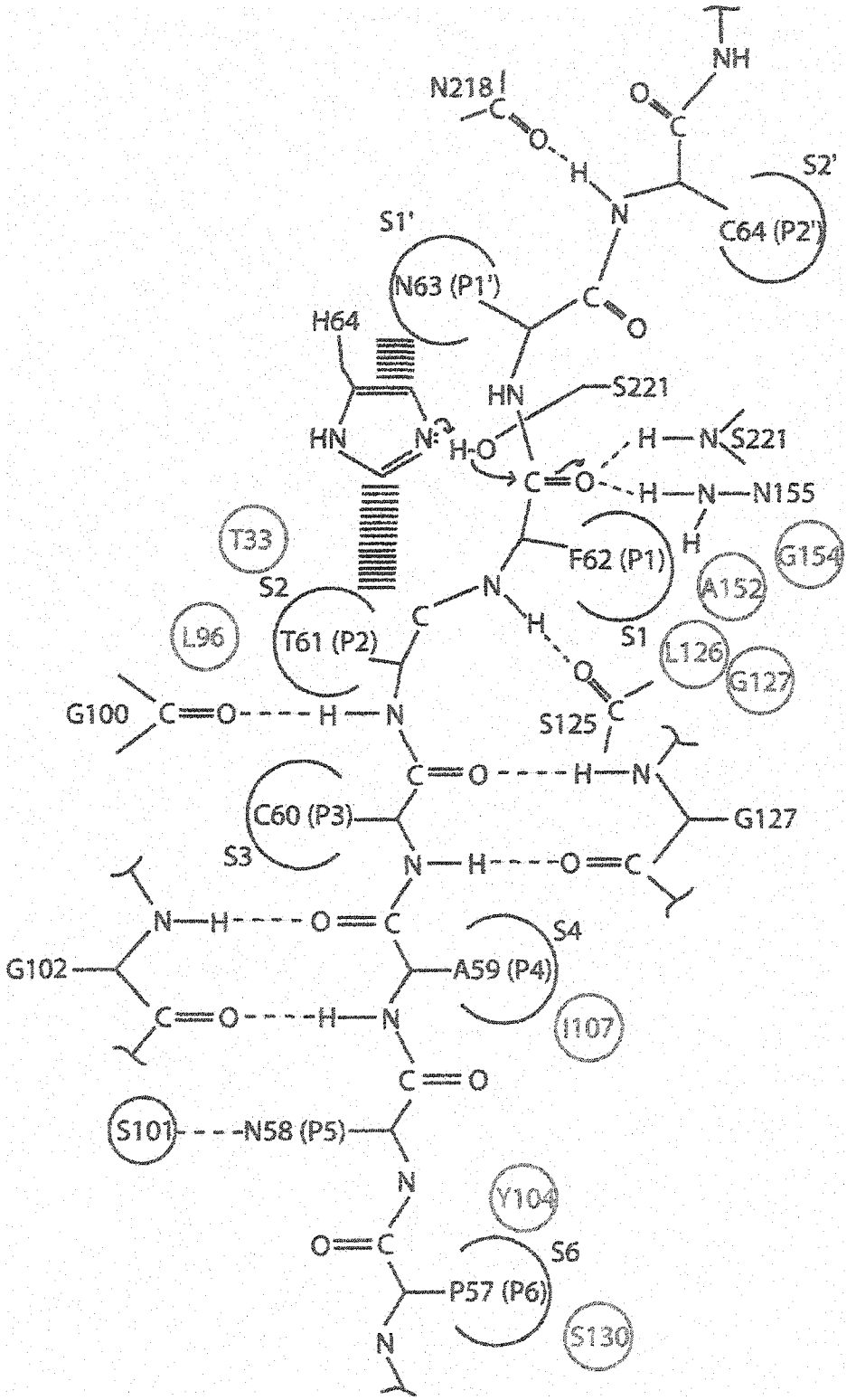
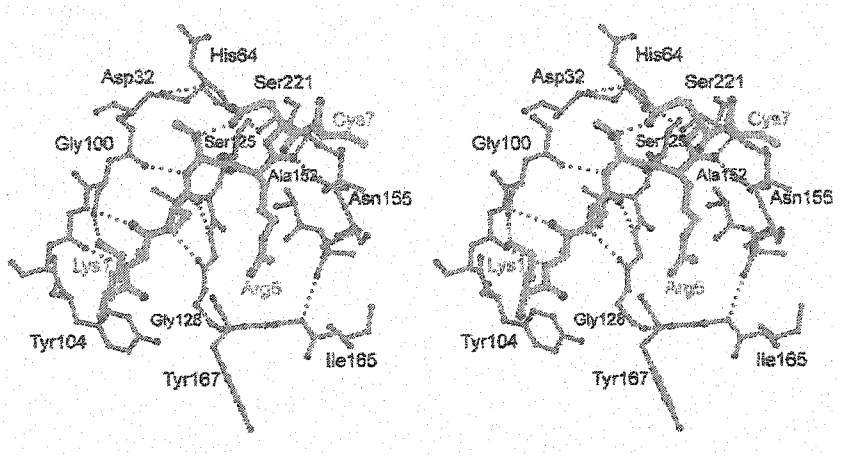


Figure 3-4C and 3-4D: Stereoscopic view of the interactions between subtilisin and **(C)**, the reactive site loops of Domain I and **(D)**, Domain II. Inhibitor drawn in green and subtilisin drawn in black.

(C)



(D)

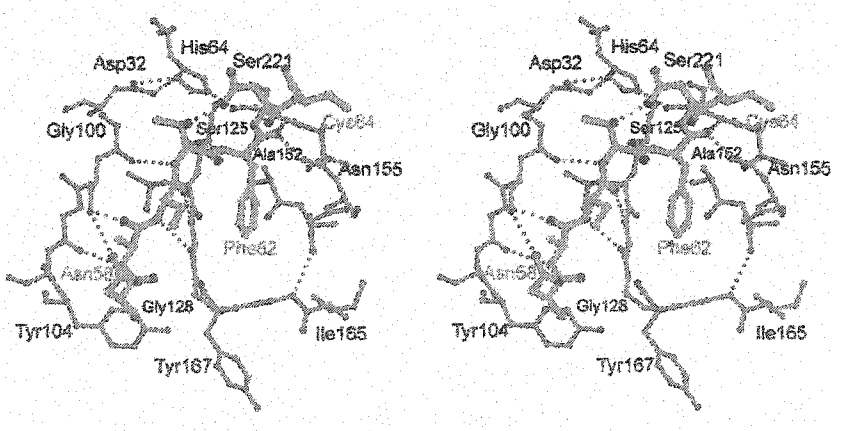
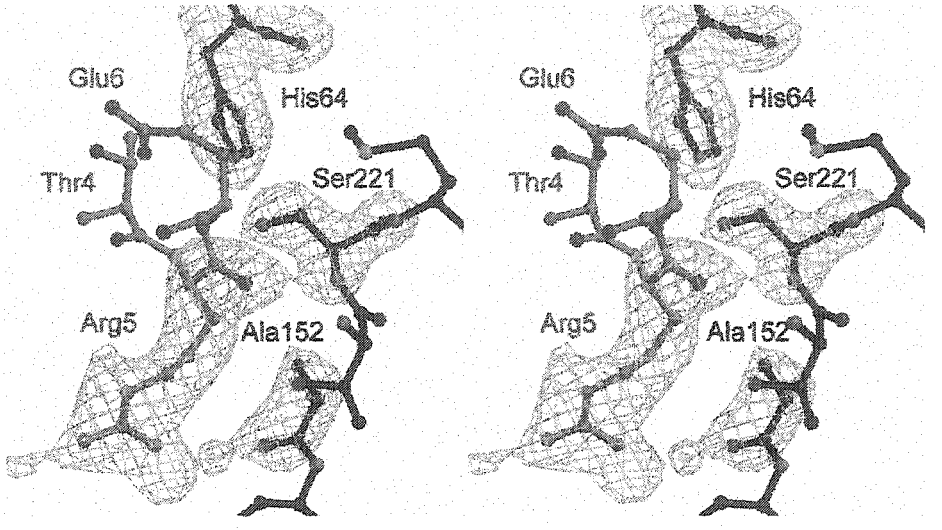
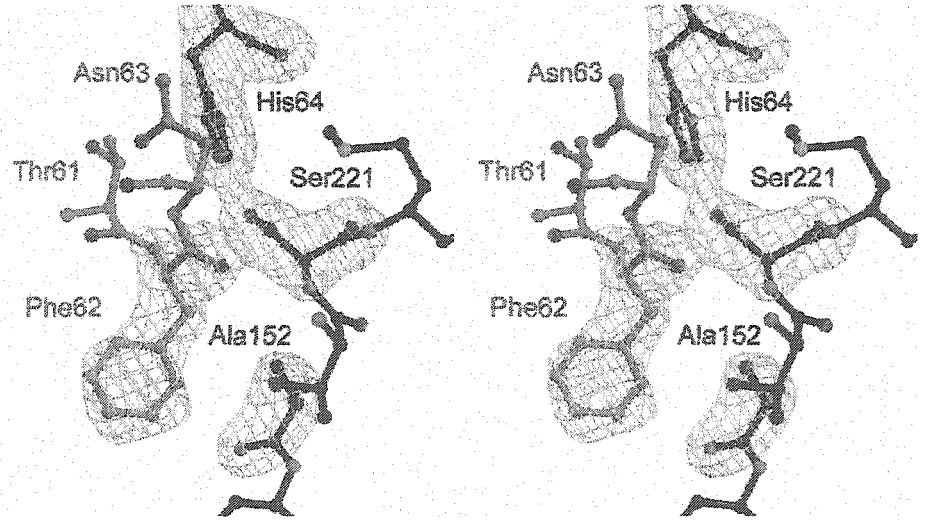


Figure 3-5: Stereoscopic view of simulated annealing omit maps for the active site regions of subtilisin bound to **(A)**, Domain I and **(B)**, Domain II. Electron density maps were calculated after omitting subtilisin residues 64, 152 and 221, as well as the P1 residue from the inhibitor, and refining the structure by the simulated annealing protocol, starting at 1000 K (19).

(A)



(B)



3.3 References

1. Plunkett, G., Senear, D. F., Zuroske, G., and Ryan, C. A. (1982) *Arch. Biochem. Biophys.* 213, 463-472
2. Pearce, G., Johnson, S., and Ryan, C. A. (1993) *Plant Physiol.* 102, 639-644
3. Pearce, G., Ryan, C. A., and Liljegren, D. (1988) *Planta* 175, 527-531
4. Bryant, J., Green, T. R., Gurusaddaiah, T., and Ryan, C. A. (1976) *Biochemistry* 15, 3418-3424
5. Stiekema, W. J., Heidekamp, F., Dirkse, W. G., van Beckum, J., de Haan, P., ten Bosch, C., and Louwerse, J. D. (1988) *Plant Mol. Biol.* 11, 255-269
6. Richardson, M. (1979) *FEBS Lett.* 104, 322-326
7. Antcheva, N., Patthy, A., Athanasiadis, A., Tchorbanov, B., Zakhariev, S., and Pongor, S. (1996) *Biochim Biophys Acta* 1298, 95-101
8. Atkinson, A. H., Heath, R. L., Simpson, R. J., Clarke, A. E., and Anderson, M. A. (1993) *Plant Cell* 5, 203-213
9. Pearce, G., Sy, L., Russell, C., Ryan, C. A., and Hass, G. M. (1982) *Arch. Biochem. Biophys.* 213, 456-462
10. Brown, W. E., and Ryan, C. A. (1984) *Biochemistry* 23, 3418-3422
11. Graham, J., Hall, G., Pearce, G., and Ryan, C. A. (1986) *Planta* 169, 399-405
12. Broadway, R. M., and Duffey, S. S. (1986) *J. Insect. Physiol.* 32, 673-680
13. Cuatrecasas, P., and Anfinsen, C. B. (1971) *Methods Enzymol.* 22, 345-378
14. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307-326
15. Collaborative Computational Project, N. (1994) *Acta Crystallogr. D* 50, 760-763
16. Navaza, J. (2001) *Acta Crystallogr D Biol Crystallogr* 57, 1367-1372

17. Greenblatt, H. M., Ryan, C. A., and James, M. N. (1989) *J. Mol. Biol.* 205, 201-228

18. McRee, D. E. (1999) *J. Structural Biology* 125, 156-165

19. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* 54, 905-921.

20. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) *Proteins* 12, 345-364

21. Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* 381, 272.

22. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946-950

23. Esnouf, R. M. (1997) *J. Mol. Graph. Model.* 15, 132-134

24. Merritt, E. A., and Bacon, M. (1997) *Methods Enzymol.* 277, 505-524

25. DeLano, W. L. (2002), DeLano Scientific, San Carlos, CA, USA

26. Steinmetz, A. C., Demuth, H. U., and Ringe, D. (1994) *Biochemistry* 33, 10535-10544

27. Nielsen, K. J., Heath, R. L., Anderson, M. A., and Craik, D. J. (1994) *J. Mol. Biol.* 242, 231-243

28. Nielsen, K. J., Heath, R. L., Anderson, M. A., and Craik, D. J. (1995) *Biochemistry* 34, 14304-14311

29. Lee, M. C. S., Scanlon, M. J., Craik, D. J., and Anderson, M. A. (1999) *Nat. Struct. Biol.* 6, 526-530

30. Laskowski, M. J., and Qasim, M. A. (2000) *Biochim. Biophys. Acta* 1477, 324-337

31. Laskowski, M. J., and Kato, I. (1980) *Ann. Rev. Biochem.* 49, 593-626

32. Ozawa, K., and Laskowski, M. J. (1966) *J. Biol. Chem.* 241, 3955-3961

33. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* 27, 157-162

34. Bolognesi, M., Gatti, G., Menegatti, E., Guarni, M., Marquart, M., Papamokos, E., and Huber, R. (1982) *J. Mol. Biol.* 162, 839-868

35. Apostoluk, W., and Otlewski, J. (1998) *Proteins* 32, 459-474

36. Bode, W., Papamokos, E., and Musil, D. (1987) *Eur J Biochem* 166, 673-692

37. McPhalen, C. A., and James, M. N. (1988) *Biochemistry* 27, 6582-6598

38. Heinz, D. W., Priestle, J. P., Rahuel, J., Wilson, K. S., and Grutter, M. G. (1991) *J Mol Biol* 217, 353-371

39. Takeuchi, Y., Satow, Y., Nakamura, K. T., and Mitsui, Y. (1991) *J Mol Biol* 221, 309-325

40. Takeuchi, Y., Noguchi, S., Satow, Y., Kojima, S., Kumagai, I., Miura, K., Nakamura, K. T., and Mitsui, Y. (1991) *Protein Eng* 4, 501-508

41. Lin, G., Bode, W., Huber, R., Chi, C., and Engh, R. A. (1993) *Eur. J. Biochem.* 212, 549-555

42. Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. (2000) *J. Mol. Biol.* 298, 477-491

43. Raj, S. S., Kibushi, E., Kurasawa, T., Suzuki, A., Yamane, T., Odani, S., Iwasaki, Y., and Ashida, T. (2002) *J. Biochem. (Tokyo)* 132, 927-933

44. Beekwilder, J., Schipper, B., Bakker, P., Bosch, D., and Jongsma, M. (2000) *Eur J Biochem* 267, 1975-1984

45. Fujinaga, M., Read, R. J., Sielecki, A. R., Ardelt, W., Laskowski, M. J., and James, M. N. G. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4868-4872

46. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M. J., and James, M. N. (1987) *J. Mol. Biol.* 195, 397-418

CHAPTER 4:

Unbound Form of Tomato Inhibitor-II Reveals Interdomain Flexibility and Conformational Variability in the Reactive Site Loops[1]

4.0 Introduction

The crystal structure of the TI-II:(subtilisin)$_2$ complex revealed that TI-II consists of two structurally related domains, each containing a reactive-site loop capable of independently binding a proteinase molecule (1). The combination of multiple inhibitory domains within a single PI likely plays an important function, but interactions between the multiple inhibitory domains within a single PI remain poorly characterized. Crystallographic studies of multidomain Pot II inhibitors, starting with the structure of the TI-II:(subtilisin)$_2$ complex, are just beginning to suggest the importance of multiple inhibitory domains within a single PI.

A central paradox in understanding the mechanism of proteinase inhibitors is how these proteins can bind tightly to proteinases in a substrate-like manner without being rapidly hydrolyzed like a substrate. Many structural and biochemical studies (2,3) on a wide range of structurally diverse PIs and their cognate proteinase have revealed a variety of mechanisms allowing PIs to achieve their biologically important functions. These mechanisms include various strategies for stabilizing the reactive site loop using disulfide bridges, extensive hydrogen bonding networks, salt bridges and van der Waals contacts to

---

[1] *A version of this chapter has been accepted for publication. Barrette-Ng, I.H., Ng, K.K.-S., Cherney, M.M., Pearce, G., Ghani, U., Ryan, C.A., and James, M.N.G. (2003) J. Biol. Chem., electronically published on June 4, 2003 (in press). Protein purification was performed by Gregory Pearce and Dr. Clarence Ryan at Washington State University, and protein crystallization was performed by Maia Cherney at the University of Alberta.*

prevent hydrolysis and to promote tighter binding by reducing the loss in entropy upon binding. In addition to the need for resistance to hydrolysis and tight binding, however, reactive site loops from broad-spectrum PIs must also retain sufficient conformational flexibility to allow the recognition by a variety of proteinase active site clefts (2-7).

To further extend the earlier X-ray crystallographic structure determination of a ternary complex consisting of TI-II bound to two molecules of subtilisin Carlsberg (1), the structure of unbound TI-II has now been determined to 2.15 Å resolution. Four copies of the unbound inhibitor are contained within the asymmetric unit of the crystalline unit cell, thereby providing a unique opportunity to examine the significant range of conformational flexibility present in the global structure of the inhibitor, as well as in the local structure of the reactive site loops. An analysis of the range of conformations sampled by the four independent molecules in the asymmetric unit provides substantial insight into the importance of conformational flexibility between domains as well as within reactive site loops.

4.1 Experimental procedures

4.1.1 Purification and crystallization

TI-II was prepared from transgenic tomato plants that overexpressed a prosystemin transgene, resulting in the synthesis and accumulation of high levels of TI-II in the leaves (~1 mg/ml leaf juice). Leaves from 1000 young tomato plants were collected (~800 g leaf tissue) and blended with 1600 mL of a buffer containing 0.01 M sodium citrate, 0.5 M sodium chloride, and 0.7% sodium hydrosulfite, pH 4.3. The homogenate was expressed through 8 layers of cheesecloth by hand-squeezing, and the debris retained in the cloth was resuspended in 400 mL buffer and the mixture was further expressed. The combined, expressed liquid was clarified by centrifugation at 12,000 x g for 20 min. Proteins in the supernatant liquid were precipitated by adding solid ammonium sulfate to 80% saturation and stirred at 4° C for 2 hr. The precipitate was recovered by centrifugation as above. The pellet was solubilized in 600 mL water and the remaining debris removed by centrifugation. The resulting solution was placed in a flask and immersed in a boiling water bath with stirring until the temperature of the liquid was 70° C. The contents of the flask were then cooled rapidly in an ice bath to room temperature and the precipitated proteins repelleted by centrifugation. The solution was dialyzed overnight at 4°C against 0.01 M Tris, 0.10 M KCl buffer, pH 8.1. The retentate was passed through a 3 cm x 12 cm column containing chymotrypsin-Sepharose CL4B affinity resin (8) that had been equilibrated with the buffer. The column was then washed with 3 bed volumes of buffer and then with 8 M urea, pH 3, to elute the inhibitor proteins that were bound to chymotrypsin. The eluate was dialyzed against several changes of 50 mM ammonium bicarbonate and lyophilized and stored. The preparation

yielded about 80 mg of dry material containing about 24 mg TI-II and 14 mg inhibitor I. TI-II protein was purified to homogeneity using reverse phase HPLC. A Vydac (Hesperia, CA) 218TP510 C18 semi-preparative column (10 x 250 mm, 5 micrometers, 300 Å) was employed using a 90 min gradient of 0-60% acetonitrile in 0.1% TFA. TI-II eluted as a single peak at 53 min. The yields of inhibitor from this step were nearly quantitative. TI-II protein obtained was homogeneous as judged by SDS-PAGE (data not shown).

Crystals were grown at room temperature by the hanging-drop, vapour-diffusion method by mixing 2 μL of a 1 mM solution of TI-II and 2 μL of 7 % (w/v) PEG 10000 and 50 mM ammonium sulphate. Small crystals grew within one to two weeks. Macroseeding was used to produce crystals with typical dimensions of 0.2 x 0.1 x ~0.05 mm. The crystals belonged to space group C2 with cell dimensions a = 58.31 Å, b = 105.94 Å, c = 81.35 Å, $\alpha = 90°$, $\beta = 104.14°$ and $\gamma = 90°$. The V_m for four TI-II molecules (13,500 Da) per asymmetric unit is 2.26 $Å^3$/Da.

4.1.2 Data collection, structure solution and refinement

Crystals used for data collection were prepared by harvesting into a solution identical to the crystallization solution with the addition of 5% (w/v) PEG 400 and equilibrating overnight against a well solution containing the crystallization solution with the addition of 20% (w/v) PEG 400. Immediately prior to flash freezing in a nitrogen gas stream at 100 K, the crystals were briefly transferred into a cryoprotectant solution identical to the crystallization solution, with the addition of 25% (w/v) PEG 400. Data were measured from a ADSC Quantum-315 CCD detector using 0.97 Å wavelength radiation at beamline 9-2 (Stanford Synchrotron Radiation Laboratory). Data were

recorded as 180 1° oscillations at a crystal-to-detector distance of 250 mm and processed using the HKL suite of programs (9). Intensities were converted to amplitudes using TRUNCATE (10).

The solvent content of the unbound inhibitor crystals was calculated to be 59%, 45% or 32% if three, four or five copies of TI-II were present in the asymmetric unit, respectively. Molecular replacement searches were performed using BEAST (11) and CNS (12). The structure of TI-II from the TI-II:(subtilisin)$_2$ complex (PDB code 1OYV), with all the side chains and temperature factors retained, was used as the search model for molecular replacement calculations. Cross-rotation functions calculated using data to a maximum resolution of 3.5 Å in BEAST (11) gave two solutions (log-likelihood-gain scores of 23.0 and 21.1). Each rotation function solution was used to calculate a translation function using data to a maximum resolution of 3.5 Å, each yielding one clear solution (log-likelihood-gain scores of 22.5 and 14.9). Using CNS (12), the two solutions were subjected to Patterson correlation refinement and the relative orientation of the two solutions was determined using a translation search, followed by rigid body refinement (R-factor of 0.51). A third solution was identified using CNS (12) by carrying out Patterson correlation refinement and translation searches on other high-scoring solutions from the cross-rotation search. The three solutions were subjected to rigid body refinement, yielding a R-factor of 0.46. The fourth solution could not be identified using Patterson correlation refinement and translation searches. Instead, a non-crystallographic two-fold rotational symmetry element relating two of the three positioned molecules was identified and was applied to the third molecule to generate the fourth molecule in the

asymmetric unit. Following rigid body refinement of the four molecules, the R-factor was 0.42.

The electron density map calculated using the molecular replacement solution revealed a good fit of the model to the core regions of the inhibitor. In addition, there was clear electron density for the interdomain residues 74 to 85 in all four copies of TI-II in the asymmetric unit. Because these residues were not modeled in the structure of TI-II in complex with subtilisin, they were placed by manual inspection using the molecular graphics program XFIT (13). The N- and C-termini, and the loop encompassing residues 101 to 105 showed the greatest variation amongst the four copies. The model was subjected to iterative rounds of manual model building using XFIT (13) and refinement against a maximum-likelihood target using CNS (12). Further refinement of the structure was performed using REFMAC v. 5.0.36 (14) with individual TLS parameters refined separately for Domain 1 and Domain 2. Non-crystallographic symmetry restraints were not used at any point during refinement, because of significant differences in conformation among different copies. Residues 1 and 117 to 123 are missing from all four copies of TI-II. In addition, the following residues are missing: 2 in copy B, and 116 in copies B, C and D. 90.4 % of residues lie in the most favoured regions of the Ramachandran plot and 9.6 % lie in the additional allowed regions (regions defined by PROCHECK (15)). Additional checks on geometry were performed using WHATCHECK (16). Data quality and refinement statistics are given in Table 4-1. Figures were prepared with MOLSCRIPT (17), BOBSCRIPT (18) and RASTER3D (19). Analysis of interdomain interfaces as well as buried surface area calculations were performed using CNS (12).

4.2 Results and discussion

4.2.1 Global structure of the unbound form of TI-II

The structure of unbound TI-II (Figure 4-1) was determined by the molecular replacement technique to 2.15 Å resolution, using the coordinates of TI-II in complex with two molecules of subtilisin Carlsberg (1) (PDB code 1OYV) as the search model (refer to Table 4-1 for crystallographic parameters). There are four copies of TI-II (designated as A, B, C and D) in the asymmetric unit and the overall conformation of each copy is similar in most regions to that of TI-II from the complex (1).

TI-II adopts an elongated shape with approximate overall dimensions of 50 Å x 25 Å x 15 Å; it consists of two structurally-similar inhibitory domains (Domains I and II). Each inhibitory domain adopts the fold previously determined for the single-domain Pot II inhibitors (20-23) and consisting of only a small amount of regular secondary structure in the form of a small antiparallel β-sheet, as well as a series of polypeptide chain segments interconnected by four disulphide bonds. Both domains also contain one turn of a 3_{10} helix, comprising residues 30 to 32 in Domain II and residues 90 to 92 in Domain I. An inhibitory reactive site loop is found in each domain and these loops are positioned at opposite ends of the elongated molecule (Figure 4-1), thereby allowing a single inhibitor to bind to two proteinase molecules simultaneously.

As first deduced from the structure of PCI-I in complex with SGPB and proven by the structure of TI-II in complex with subtilisin Carlsberg, the sequence repeats present in TI-II do not correspond to individual structural domains (1,20). Instead, Domain I consists of the first fifteen residues and of residues 80-116, whereas Domain II consists of residues 18-73. Although both domains adopt the same fold, the structures of each

domain differ in detail. The two most important structural differences between Domain I and Domain II are: (1) the N-terminus of the reactive site loop in Domain I lies at the N-terminus of the protein but in Domain II, this segment is connected by a loop to the last β-strand of the small β-sheet (corresponding to the C-terminus of Domain I), and (2) the loop composed of residues 20 to 26 in Domain II is shortened by four residues in Domain I and includes the only *cis*-peptide bond in the structure (Tyr83 to Pro84).

Domains I and II are linked together by two polypeptide segments, consisting of residues 16 to 17 and 74 to 79. The linker consisting of residues 74 to 79, as well as several adjacent residues (residues 80 to 85), appeared to be disordered in the structure of TI-II bound to subtilisin, because electron density corresponding to this region could not be detected. In contrast, in all four copies of unbound TI-II, the entire region was well defined by unbiased electron density maps calculated prior to modeling the structure of this region. The linker region consisting of residues 74 to 79 is highly solvent exposed and appears to be stabilized primarily by three hydrogen bonds. It is not clear why there is such a dramatic change in the conformational flexibility of residues 74 to 85 between the unbound and bound forms of TI-II. In the TI-II:$(subtilisin)_2$ complex, the loop of subtilisin comprising residues from 158 to 164 is located adjacent to where residues 81 to 84 of the inhibitor would be expected. It is puzzling why this region of the inhibitor appears to be disordered only in the presence of bound proteinase, because it might be expected that the proteinase would interact with this region and possibly stabilize it.

A possible explanation for the conformational flexibility seen in the linker region may lie in the different orientations of Domain I relative to Domain II that are observed in the bound and unbound forms of TI-II. DYNDOM (24) reveals a rotation (residues 17

and 75 as the hinge residues) of Domain II relative to Domain I of 7.4° and 8.8° when comparing copies A to C, and A to D, respectively. The two inhibitory domains can clearly adopt different orientations relative to each other, which may allow the multidomain inhibitor molecule to accommodate the binding of multiple proteinase molecules simultaneously. In the complex of TI-II with subtilisin Carlsberg, the two proteinase molecules are well separated from each other, but complexes of TI-II with other proteinases may require slight changes in the orientation of the inhibitory domains to avoid steric clashes between bound proteinase molecules.

A feature of the structure of TI-II that allows for interdomain flexibility is the relatively small interdomain interface. An analysis of this interface reveals that the interdomain buried surface area found in the unbound form of TI-II ranges from 559 to 627 $Å^2$, which is 15 to 34% larger than that found for the bound form (487 $Å^2$). Similar to the bound form, the interdomain interface in unbound TI-II consists of a small cluster of highly conserved hydrophobic residues (Ile14, Pro16, Tyr98, Phe100 and Phe106 from Domain I; and Tyr34, Pro54 and Pro57 from Domain II), as well as Arg17 and Lys55. When comparing the different copies of bound and unbound TI-II, there are a number of changes in the positions and conformations of residues at the interdomain interface, most notably Ile14, Arg17, Pro54, Pro57 and Phe100. Perhaps the most striking difference in conformation amongst the interface residues arises in Arg17. In copies A and D, the side chain points towards Domain I, forming a salt bridge with Glu114, whereas in copies B and C, as well as in the TI-II:$(subtilisin)_2$ complex the side chain extends out into solution (Figure 4-2). Because these two conformations place the side chains on opposite sides of the linker segment consisting of residues 77 to 78, it appears that a concerted

conformational change involving the linker and Arg17 would be required to interconvert between the two side chain conformations.

In solution, it is likely that both bound and unbound forms of TI-II exhibit conformational variability in the orientation of one inhibitory domain relative to the other and that this conformational flexibility is mediated by slight changes in packing at the interdomain interface. The range of conformations as seen in different copies of both the bound and unbound forms of TI-II likely arises from the trapping of different solution conformations of the inhibitor by the crystal lattice. Table 4-2 reveals the overall range of conformational variability seen among the four copies of unbound TI-II in the asymmetric unit and in comparison with the bound form. Globally, copies A and B (r.m.s.d. of 0.62 Å) are similar, as are copies C and D (r.m.s.d. of 0.44 Å). This is also reflected in comparisons of Domain II by itself with r.m.s.d.'s of 0.19 Å and 0.24 Å when comparing copies A and B, and copies C and D, respectively. In contrast, Domain I differs strikingly between copies A and B (r.m.s.d. of 0.75 Å), although it is quite similar when comparing copies C and D (r.m.s.d. of 0.38 Å).

4.2.2 Comparison of unbound TI-II with that of bound TI-II

The overall conformation of the core of the unbound TI-II reveals no major structural rearrangements between the bound and unbound forms (Figure 4-3) but there are clear differences in the orientation of the two domains relative to one another as well as localized variations in the structures of several loops. Superposition of all main-chain atoms reveals large r.m.s.d.'s among all four copies of the unbound inhibitor in comparison with the bound form of the inhibitor (refer to Table 4-2). A hinge motion consisting of a rotation of 7.5° between Domains I and II is revealed when comparing

copies C and D with the bound form of TI-II. A similar hinge motion of 5° can be detected when comparing copies A and B with the bound form of TI-II. The structure of the proteinase:inhibitor ternary complex suggests that proteinase binding restricts the orientation of the two inhibitory domains, because the proteinase molecule bound to Domain II can form interactions with residues in Domain I (1). In the only other structure of a proteinase inhibitor bound to multiple proteinase molecules, individual trypsin molecules bind to single domains of two-headed Bowman-Birk inhibitors and do not appear to affect the orientation of the two inhibitory domains (25-27). In contrast, the flexible orientation of the two domains of TI-II may allow the inhibitor to fine-tune its interactions with different proteinase molecules bound to Domain II.

In addition to differences in orientations of domains, the bound and unbound forms of TI-II reveal significant localized changes in the conformations of several loops. In particular, the loop consisting of residues 100 to 106 in Domain I, as well as the reactive site loops and adjacent regions in both domains show large conformational differences. The loop consisting of residues 100 to 106 is situated at the interdomain interface and is near to residues 130 to 132 of the subtilisin molecule bound to Domain II in the TI-II:$(subtilisin)_2$ complex. The different conformations of this loop appear to reflect differences in the orientation of the two inhibitory domains as well as the presence of a proteinase molecule bound to Domain II.

Although there have been many previous studies (4-7,28-30) examining the conformations of bound and unbound forms of reactive site loops in canonical, standard mechanism proteinase inhibitors, the structures of four independent copies of unbound TI-II in the crystalline asymmetric unit provide a unique opportunity to examine the

range of conformations adopted by reactive site loops of two different sequences in both unbound and bound forms. As in other members of the Pot II family, the reactive site loops are flanked by disulfide bonds formed by cysteine residues at the P3 and P2' positions (Cys3 to Cys91 and Cys7 to Cys87 in Domain I and Cys60 to Cys31 and Cys64 to Cys27 in Domain II). These disulfide bridges are thought to hold the reactive site loop in a relatively rigid conformation that may help to prevent proteolytic cleavage of the inhibitor upon interaction with proteinases. Reduced conformational flexibility in the loop is also thought to enhance proteinase binding by reducing entropic loss in order to achieve tighter binding (4,31).

4.2.3 Conformational flexibility

As previously analyzed (1), main chain torsional angles of the reactive site loops in the TI-II:(subtilisin)$_2$ complex are quite similar to those seen in other PIs (Table 4-3). Although the overall conformation of reactive site loops from different inhibitors bound to a variety of serine proteinases is generally conserved, each proteinase substrate binding cleft is unique. Broad-specificity inhibitors that have evolved to inhibit a wide range of proteinases must balance the need for flexibility in recognizing proteinases with different substrate recognition sites against the need to maintain conformational rigidity (6).

The range of conformational flexibility in the reactive site loops of TI-II is revealed by variations in main chain torsional angles, side chain rotamers and hydrogen bonding networks in the four copies of unbound TI-II. Some of these features underlying the conformational flexibility of reactive site loops have been analyzed previously (5-7,30). Significant changes in main chain torsional angles are found in the reactive site loops from Domains I and II (refer to Table 4-3 and Figure 4-4). Most of these changes

reflect the higher flexibility of the unbound reactive site loop. The overall path of the polypeptide chain is quite similar for residues P3 to P5', but numerous changes in main chain torsional angles cause carbonyl oxygen and amide nitrogen atoms to occupy positions differing significantly from those seen in the bound form of TI-II.

Particularly dramatic changes are seen in the ψ angle of the P1 residue and the φ angle of the P1' residue in the reactive site loop of Domain II (refer to Table 4-3 and Figure 4-5). The reactive site loop in Domain II appears to adopt two distinct conformations in the unbound state with copies A and B adopting one conformation and copies C and D adopting the other. Both of these conformations appear to differ significantly from the bound conformation. The change in the φ and ψ angles flanking the scissile bond causes the carbonyl carbon to carbonyl oxygen vector of the P1 residue to point either 45° towards (copies A and B) or 70° away from (copies C and D) Asn29. The P1 side chain in copies A and B (χ_1 = -78° and χ_2 = -21°) adopt a similar conformation to that seen in the bound form of the inhibitor (χ_1 = -88° and χ_2 = -4°) but in copies C and D, the P1 side chain adopts a distinctly different rotamer (χ_1 = -173° and χ_2 = 76°).

4.2.4 Conformations and interactions of reactive site residues

Side chains of other residues in the reactive site loop also adopt distinct rotamers in the unbound form of the inhibitor. In the reactive site loop of Domain I, the most dramatic changes can be seen in the side chains of the P2, P1 and P1' residues. The P1 arginine residue adopts a somewhat unusual rotamer to fit into the S1 pocket when bound to subtilisin Carlsberg (1). In the absence of constraints imposed by a bound proteinase molecule, this highly solvent exposed side chain adopts an extended conformation. The

P2 and P1' side chains in both domains also change significantly in conformation in the unbound form of TI-II, resulting in the loss of a commonly found hydrogen bond formed between the side chains of these two residues; this hydrogen bond is thought to confer stability in the reactive site loop (7). In Domain I of the unbound form of TI-II, the conformation of Thr4 at the P2 position is quite similar to that seen in the bound form but the conformation of Glu6 at the P1' position is distinctly different. In copies A and B, the rotamer adopted by Glu6 places the carboxylate oxygen atoms too far from Thr4 to form a direct hydrogen bond whereas in copies C and D, a hydrogen bond can still be formed between the P2 and P1' residues. In Domain II, the rotamers adopted by Thr61 (P2) and Asn63 (P1') all differ markedly from the rotamers adopted in the bound form of TI-II which disrupts the hydrogen bond between the P2 and P1' residues in all four copies of the unbound inhibitor.

4.2.5 Interactions of reactive site (P2, P1 and P1') with core of inhibitor

A second dramatic alteration in the hydrogen bonding network stabilizing the reactive site loop in Domain II is also seen in the unbound form of the inhibitor. In the bound form of the inhibitor, a key set of interactions stabilizing the reactive site loop in Domain II are contributed by the hydrogen bonds between the $N^{\delta 2}$ atom of the side chain of Asn29 and the main chain carbonyl oxygen atoms of Thr61 and Asn63 at the P2 and P1' positions, as well as the hydrogen bond between the main chain carbonyl oxygen atom of Ile28 and the main chain amide nitrogen atom of Asp65. The same set of interactions are observed in copies A and B of the unbound inhibitor, but in copies C and D, the hydrogen bond between Asn29 and Thr61 is missing because the reactive site loop appears to have moved slightly farther away from the core of Domain II. The

corresponding interaction in Domain I consists of a van der Waals contact between the side chains of Thr89 and Glu6 replacing the hydrogen bonds donated by Asn29 in Domain II. This van der Waals contact is missing in all four copies of the unbound inhibitor, because the side chain of Glu6 adopts different rotamers that place it farther from Thr89.

In addition to the many differences seen in the conformations of the reactive site loops of the unbound inhibitor, conformational differences are also seen in a region immediately N-terminal to the reactive site loop in Domain II consisting of residues 53 to 59. This loop adopts a significantly different conformation in all four copies of unbound TI-II compared to the conformation of the bound form. This region is close to the interdomain interface and small changes in conformation in this loop may stabilize different orientations of the two inhibitory domains.

4.2.6 Comparison with inhibitors from the Potato II family

Apart from the structure of TI-II in complex with subtilisin Carlsberg (1), the only other crystallographic structure of a Pot II family inhibitor is the structure of PCI-I in complex with SGPB (20). The structure of PCI-I in complex with SGPB is very similar to that of Domain II in TI-II, as the two domains share 86% sequence identity (refer to Table 4-3). Most of the conformational changes observed between the structure of Domain II of TI-II bound to subtilisin Carlsberg and the unbound inhibitor structures can also be seen when comparing the bound form of PCI-I and the unbound forms of TI-II. One interesting difference between PCI-I and TI-II is the identity of the P2 residue in the reactive site loop. In PCI-I and most other Pot II family PIs, proline occupies this position, whereas in TI-II it is occupied by threonine in both reactive site loops. In the

bound form of TI-II, this residue forms a hydrogen bond with P1', but this hydrogen bond is not present in six of the eight independent reactive site loops of the unbound form. It has been proposed that proline at the P2 position helps to rigidify the reactive site loop and that threonine in the P2 position can also help to rigidify the reactive site loop by forming a hydrogen bond with the P1' residue (7). These two different mechanisms for stabilizing reactive site loop conformation may be of functional significance because the hydrogen bond between the threonine at P2 and the P1' residue can be broken in the unbound form of the inhibitor to provide additional flexibility. On the other hand, proline at P2 presumably rigidifies both the bound and unbound forms of the reactive site loops.

The structures of several single domain Pot II family inhibitors from *Nicotiana alata* have been determined by NMR (21-23). These studies reveal that the solution structures of Pot II family inhibitors are similar to the crystallographically determined structures. The regions of φ-ψ space occupied by residues in the reactive site loop appear to be similar to those occupied in both the bound and unbound structures of TI-II and PCI-I. The NMR structures reveal that the atomic r.m.s.d.'s over the reactive site loop are significantly larger than the other regions of the inhibitor, which is consistent with the greater conformational variability seen in the reactive site loops of the unbound form of TI-II.

A highly conserved feature of the bound form of reactive site loops in many Pot II family inhibitors is a hydrogen bond between the side chain of the P1' residue and its amide nitrogen. It has been proposed that this hydrogen bond stabilizes the reactive site loop from proteolytic cleavage at the scissile bond immediately adjacent to the amide

nitrogen (20,32,33). Although this hydrogen bond is seen in the bound forms of PCI-I and TI-II, it is missing in six of the eight reactive site loops of unbound TI-II. The loss of this hydrogen bond is consistent with NMR studies of unbound T1 to T4 (22) where this hydrogen bond also appears to be absent. The absence of this hydrogen bond in unbound forms of proteinase inhibitors may provide additional flexibility to the unbound form of the reactive site loop that may be of importance in proteinase recognition and binding. However once bound, the P1' NH to P1' side chain oxygen (Glu, Asp or Asn) is important to the inhibitory nature of the domain.

4.2.7 Conformational changes in reactive site loops upon complexation: comparison with inhibitors from other families

The increased conformational variability of the reactive site loops in unbound TI-II is similar to that seen in other families of proteinase inhibitors (6,7). Main chain torsional angles particularly at the P2 to P1' positions show greatly increased variability in the unbound form of a wide range of PIs and this is also seen in TI-II (refer to Table 4-3 and Figures 4-4 and 4-5). A similar increase in variability in side chain torsional angles is also seen particularly at positions P1, P1' and P2', which is also the case in TI-II. Surprisingly, variation at the P2 position in other inhibitors appears to be quite limited but in TI-II, Thr61 which occupies the P2 position in the reactive site loop in Domain II can adopt three different rotamers ranging over 250° in χ_1 torsion angle space in the unbound form. Variation in side chain conformation at this position as well as at the P1' position in unbound inhibitors appears to be coupled to the loss of the stabilizing hydrogen bond between the P2 and P1' residues as discussed above. Dramatic changes in the conformations of the P2 threonine residue and the P1' glutamate residue in the

unbound form of CI-2 (30) also accompany the loss of a stabilizing hydrogen bond between the P2 and P1' residues in the bound form of the inhibitor. In contrast, smaller changes in conformation between bound and unbound forms are seen in other PIs, such as the Kazal family of PIs, where the network of hydrogen bonds and van der Waals contacts stabilizing the reactive site loop are preserved in the unbound form (5). Specifically, the hydrogen bonding network between the side chain of Asn33 and the main-chain carbonyl groups of Thr17 and Glu19 (P2 and P1' residues of the reactive site loop) are seen in both the bound and unbound forms of the ovomucoid third domain inhibitors.

Coordinates. Coordinates have been deposited in the Protein Data Bank (accession code 1PJU).

Table 4-1: Crystallographic Statistics

a) Data Processing

Resolution (Å)	50.0 - 2.15
Space group	C2
Unit cell dimensions	a = 58.31 Å, b = 105.94 Å, c = 81.35 Å
	$\alpha = 90°$, $\beta = 104.14°$, $\gamma = 90°$
Total reflections[1]	96074 (7412)
Unique reflections[1]	26506 (2619)
Redundancy[1]	3.6 (2.8)
Completeness (%)[1]	99.8 (99.7)
I/σ [1]	16.2 (4.3)
R_{sym} (%)[1,2]	7.4 (30.4)

[1]Values for the outermost resolution shell (2.23 - 2.15 Å) are given in parentheses.
[2]$R_{sym} = \Sigma_h\Sigma_i (|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|) / \Sigma_h\Sigma_i I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the i^{th} integrated intensity of a given reflection and $\langle I(\mathbf{h})\rangle$ is the weighted mean of all measurements of $I(\mathbf{h})$.

b) Refinement

Resolution (Å)	50.0 - 2.15
R_{work}[3]	0.206 (0.321)
R_{free}[4]	0.250 (0.399)
Number of atoms	
Protein	3436
Water	175
sulphate ions	8
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.008
Bond angles (°)	1.12
Avg. Temp. Factors (Å^2)	
TI-II copy A	22.3
TI-II copy B	22.3
TI-II copy C	22.0
TI-II copy D	21.9
Sulphate ions	69.5
Water	44.8

[3] $R_{work} = \Sigma_h ||F(\mathbf{h})_o| - |F(\mathbf{h})_c|| / \Sigma_h |F(\mathbf{h})_o|$ for the 90% of reflection data used in refinement.
[4] $R_{free} = \Sigma_h ||F(\mathbf{h})_o| - |F(\mathbf{h})_c|| / \Sigma_h |F(\mathbf{h})_o|$ for the 10% of reflection data excluded from refinement.

Table 4-2: RMSDs of superimposed coordinates for bound and unbound forms of TI-II[a]

A. Global Superpositions[b]

		Unbound TI-II				Bound TI-II
	Copy No.	A	B	C	D	
Unbound TI-II	**A**	-	0.62	0.77	0.79	0.88
	B	-	-	0.88	0.87	0.86
	C	-	-	-	0.44	1.01
	D	-	-	-	-	0.92

B. Superpositions of Single Domains[c]

Unbound TI-II	**A**	-	0.75	0.33	0.42	0.71
	B	0.19	-	0.74	0.54	0.71
	C	0.61	0.61	-	0.38	0.76
	D	0.57	0.57	0.24	-	0.69
Bound TI-II		0.77	0.79	0.81	0.76	-
PCI-I		0.70	0.72	0.67	0.64	0.59

C. Superpositions of Reactive Site Loops[d]

Unbound TI-II	**A**	-	0.29	0.20	0.18	0.40
	B	0.12	-	0.41	0.30	0.55
	C	1.01	1.03	-	0.26	0.36
	D	0.98	1.00	0.10	-	0.49
Bound TI-II		0.61	0.62	0.78	0.75	-
PCI-I		0.68	0.71	0.70	0.67	0.37

[a] Only main chain atoms are included in the superpositions. All superpositions were performed using the program LSQMAN (34). The "Copy No." entry refers to the individual copies in the asymmetric unit of the crystal.
[b] 452 atom pairs
[c] Above diagonal: Domain I of unbound TI-II (176 atom pairs). Below diagonal: Domain II of unbound TI-II (236 atom pairs).
[d] Above diagonal: Reactive site loop of Domain I of unbound TI-II (28 atom pairs). Below diagonal: Reactive site loop of Domain II of unbound TI-II (28 atom pairs).

Table 4-3: Main chain torsion angles of reactive site loops

	P4		P3		P2		P1		P1'		P2'		P3'	
	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ
ubTI-II (1A)			-117	155	-72	167	-79	-2	-58	136	-85	115	-75	164
ubTI-II (1B)					-103	171	-87	6	-69	131	-84	116	-71	156
ubTI-II (1C)			-62	139	-68	161	-76	-1	-72	152	-92	111	-68	167
ubTI-II (1D)			-94	147	-78	166	-78	-1	-45	141	-102	117	-80	162
ubTI-II (2A)	-75	113	-136	162	-78	166	-80	-3	-58	125	-62	135	-84	113
ubTI-II (2B)	-76	115	-139	157	-75	163	-83	0	-70	132	-63	141	-88	112
ubTI-II (2C)	-90	127	-139	135	-60	127	-64	129	-164	165	-106	120	-75	119
ubTI-II (2D)	-88	126	-133	143	-68	125	-61	123	-158	166	-112	115	-67	117
bTI-II (D1)	-120	147	-126	158	-62	132	-85	36	-90	156	-92	125	-77	168
bTI-II (D2)	-84	136	-134	152	-67	145	-97	49	-115	154	-81	143	-93	111
PCI-I	-82	132	-137	152	-65	151	-115	51	-102	165	-95	121	-71	132
OMTKY3[1]	-158	158	-126	147	-69	162	-119	45	-84	155	-100	115	-147	92
OMTKY3[2]	-129	136	-131	150	-68	160	-107	32	-74	159	-113	107	-141	76
CI-2	-93	140	-133	166	-64	147	-103	34	-91	146	-106	119	-118	113
Eglin-C	-76	138	-143	165	-65	151	-112	42	-96	176	-120	113	-119	115
BPTI	78	175	-77	-30	-70	156	-117	39	-88	164	-112	80	-106	122
sBBI[3]	-121	134	-139	145	-71	156	-95	45	-96	162	-108	97	-88	171

[1]Ovomucoid third domain bound to *Streptomyces griseus* proteinase B (33).
[2]Ovomucoid third domain bound to α-chymotrypsin (5).
[3]Soybean Bowman-Birk Inhibitor bound to bovine trypsin (26).

Angles for PCI-I (polypeptide chymotrypsin inhibitor-I from potatoes), OMTKY3, CI-2 (chymotrypsin inhibitor-2 from barley), eglin-C and BPTI (bovine pancreatic trypsin inhibitor) were obtained from ref. (20).

Figure 4-1A: Crystallographic structure of unbound TI-II. Ribbon diagram of unbound TI-II. Domain I is coloured in red while Domain II is coloured in blue. Reactive site loops 1 and 2 are shown in black and green, respectively. The side chain of the P1 residue in each reactive site loop is drawn in black. Disulfide bonds are drawn in yellow and are assigned residue numbers in black.

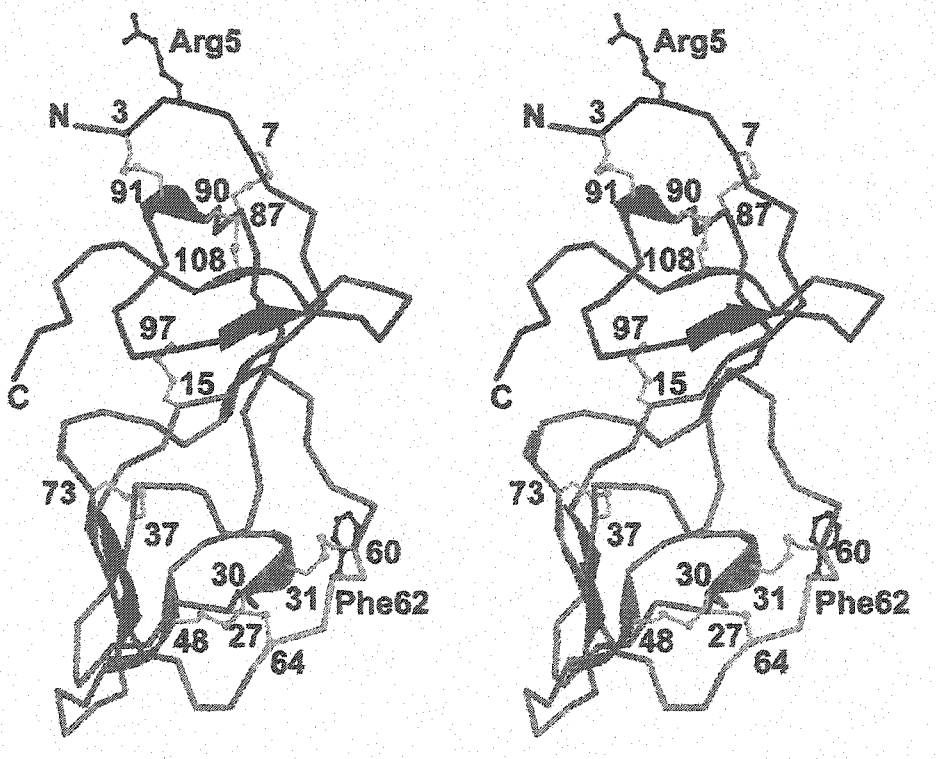


Figure 4-1B: Stereoscopic view of the four copies in the asymmetric unit of unbound TI-II. Domain I is coloured in red while Domain II is coloured in blue. Domain II of each copy has been superimposed using main chain atoms.

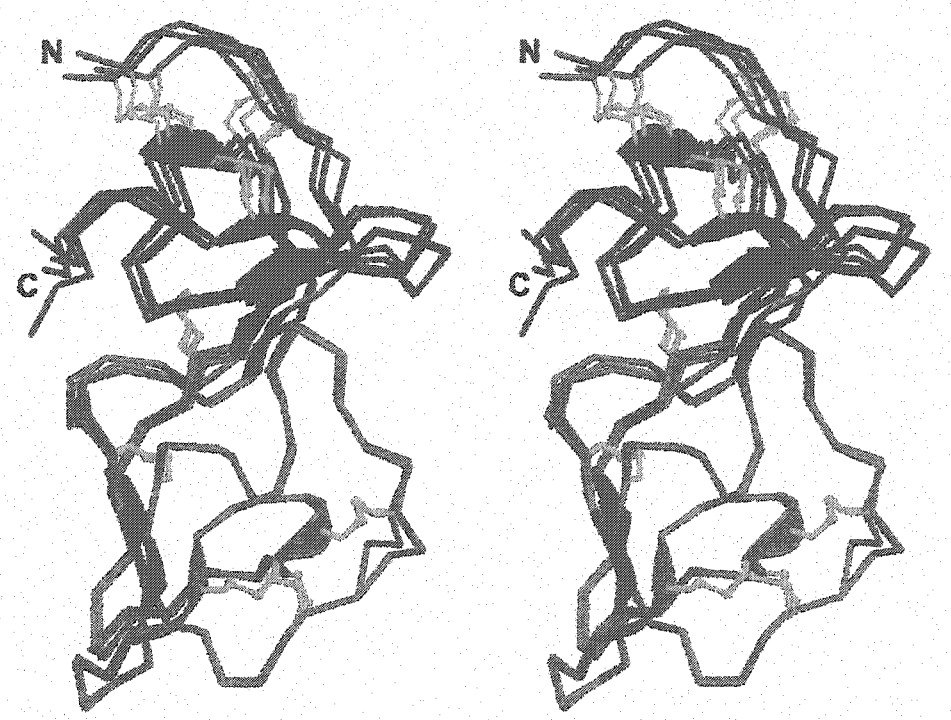
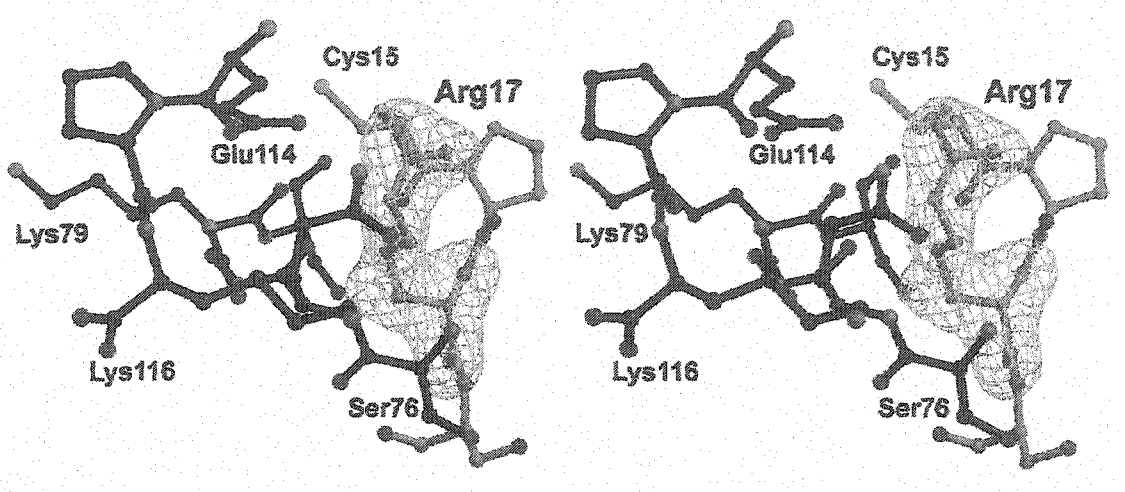


Figure 4-2: Conformational change in linker region of unbound TI-II. Stereoscopic view of linker residues 15 to 18 (coloured in magenta) and residues 76 to 80 (coloured in black) of **(A)** copy A and **(B)** copy B. Omit map electron density corresponding to Arg17 has been contoured at 3.4σ.

(A)



(B)

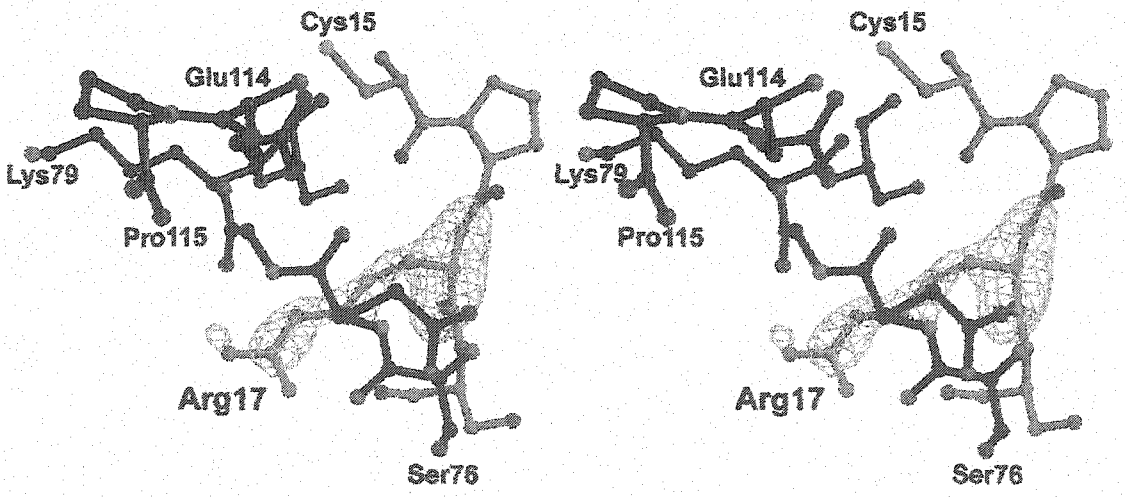
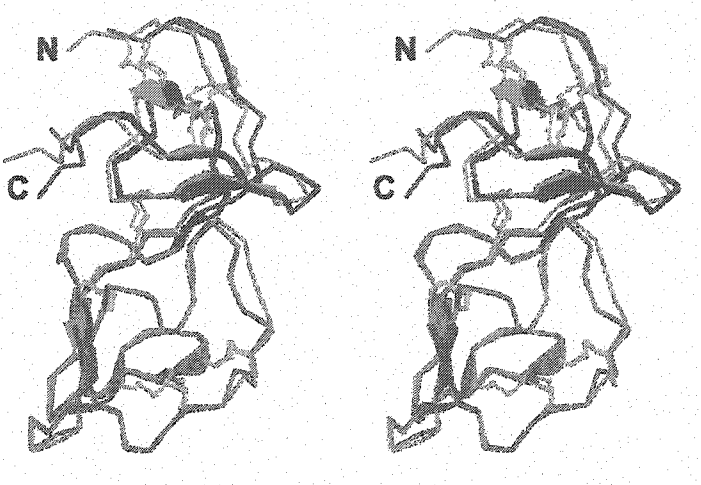


Figure 4-3: Least-squares superpositions of unbound TI-II onto bound TI-II. Stereoscopic view of a least-squares superposition of **(A)** copies A and B and **(B)** copies C and D onto the structure of TI-II from the TI-II:(subtilisin)$_2$ complex. The bound form of TI-II is drawn in magenta, whereas the unbound forms are coloured red in Domain I and blue in Domain II.

(A)



(B)

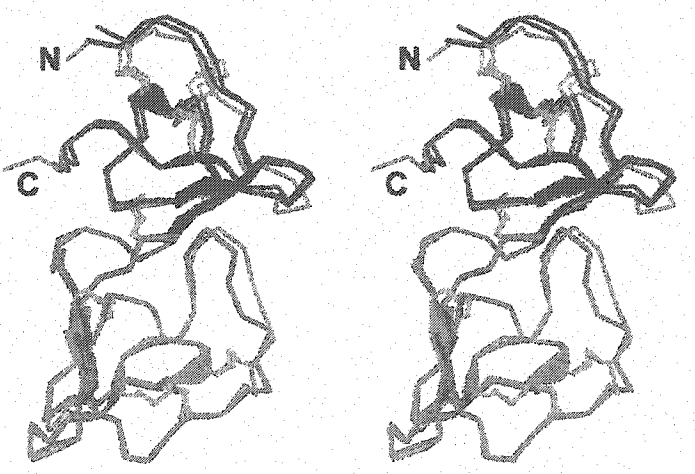


Figure 4-4: Conformations of the reactive site loops. The reactive site loops from Domain I of unbound TI-II copies A, B, C, and D are shown superimposed onto that of bound TI-II in panels **(A)**, **(B)**, **(C)**, and **(D)**, respectively. The reactive site loops from Domain II of unbound TI-II copies A, B, C, and D are shown superimposed onto that of bound TI-II in panels **(E)**, **(F)**, **(G)**, and **(H)**, respectively. The bound form of TI-II is coloured magenta and the unbound form is coloured cyan.

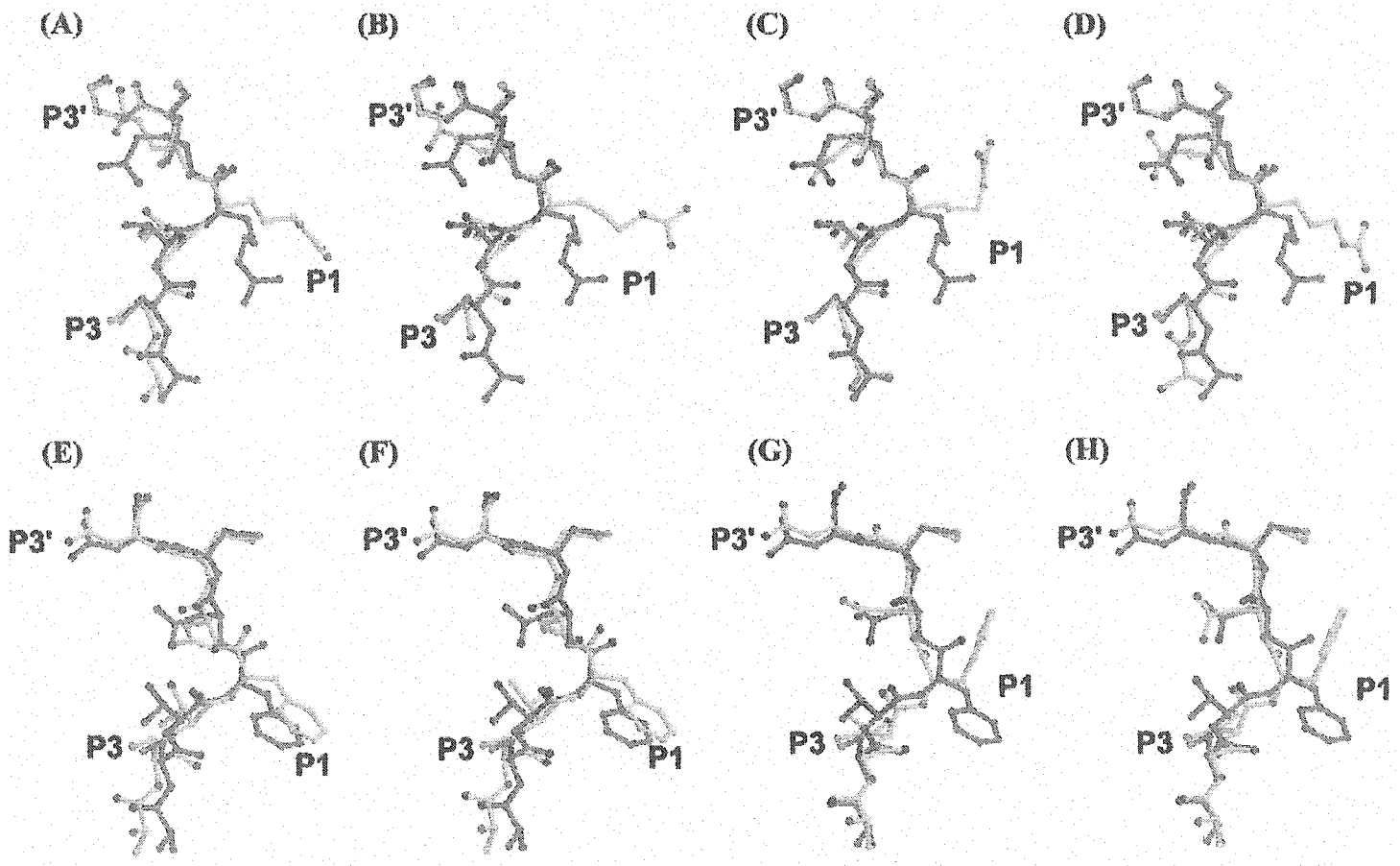


Figure 4-5A: Variations in main chain torsional angles between the bound and unbound forms of TI-II in reactive site loop from Domain I. Angles from copies A and B of unbound TI-II are shown in blue filled circles. Angles from copies C and D of unbound TI-II are shown in red filled circles. Angles from bound TI-II are shown in open black circles.

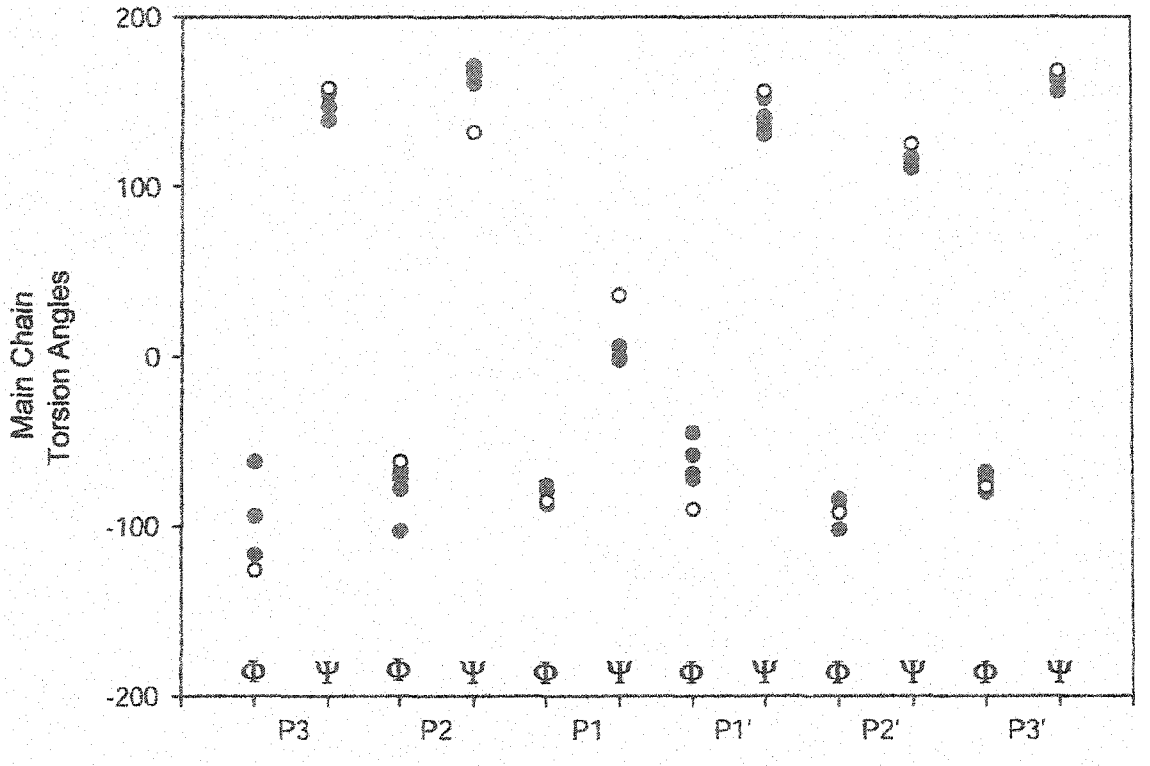
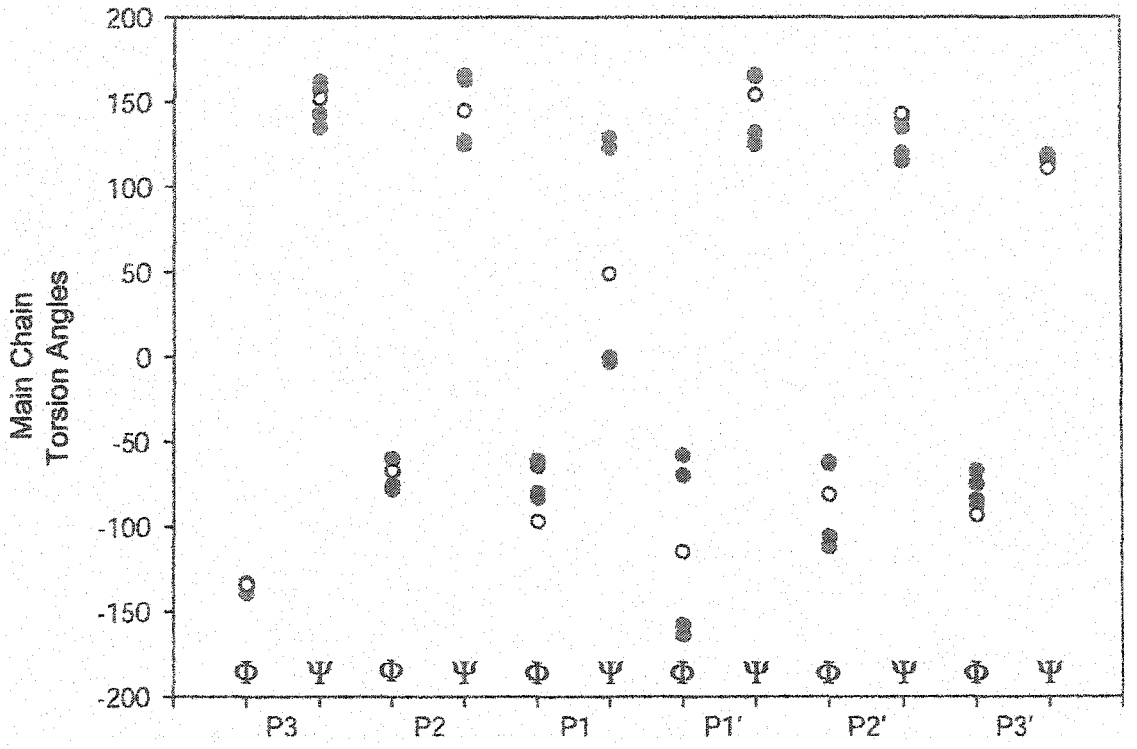


Figure 4-5B: Variations in main chain torsional angles between the bound and unbound forms of TI-II in reactive site loop from Domain II. Angles from copies A and B of unbound TI-II are shown in blue filled circles. Angles from copies C and D of unbound TI-II are shown in red filled circles. Angles from bound TI-II are shown in open black circles.



4.3 References

1. Barrette-Ng, I. H., Ng, K. K.-S., Cherney, M. M., Pearce, G., Ryan, C. A., and James, M. N. (2003) *J. Biol. Chem.*, (*in press*)
2. Laskowski, M. J., and Qasim, M. A. (2000) *Biochim. Biophys. Acta* 1477, 324-337
3. Bode, W., and Huber, R. (1992) *Eur. J. Biochem.* 204, 433-451
4. Read, R. J., and James, M. N. G. (1986) in *Proteinase Inhibitors* (Barrett, A. J., and Salvesen, G., eds) Vol. 12, pp. 301-336, Elsevier, New York
5. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M. J., and James, M. N. (1987) *J. Mol. Biol.* 195, 397-418
6. Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991) *J. Mol. Biol.* 220, 507-530
7. Apostoluk, W., and Otlewski, J. (1998) *Proteins* 32, 459-474
8. Cuatrecasas, P., and Anfinsen, C. B. (1971) *Methods Enzymol.* 22, 345-378
9. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307-326
10. Collaborative Computational Project, N. (1994) *Acta Crystallogr. D* 50, 760-763
11. Read, R. J. (2001) *Acta Cryst. D* 57, 1373-1382
12. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* 54, 905-921.
13. McRee, D. E. (1999) *J. Structural Biology* 125, 156-165

14. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr. D Biol. Crystallogr.* 57, 122-133.

15. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) *Proteins* 12, 345-364

16. Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* 381, 272.

17. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946-950

18. Esnouf, R. M. (1997) *J. Mol. Graph. Model.* 15, 132-134

19. Merritt, E. A., and Bacon, M. (1997) *Methods Enzymol.* 277, 505-524

20. Greenblatt, H. M., Ryan, C. A., and James, M. N. (1989) *J. Mol. Biol.* 205, 201-228

21. Nielsen, K. J., Heath, R. L., Anderson, M. A., and Craik, D. J. (1994) *J. Mol. Biol.* 242, 231-243

22. Nielsen, K. J., Heath, R. L., Anderson, M. A., and Craik, D. J. (1995) *Biochemistry* 34, 14304-14311

23. Lee, M. C. S., Scanlon, M. J., Craik, D. J., and Anderson, M. A. (1999) *Nat. Struct. Biol.* 6, 526-530

24. Hayward, S., and Berendsen, H. J. C. (1998) *Proteins* 30, 144-154

25. Lin, G., Bode, W., Huber, R., Chi, C., and Engh, R. A. (1993) *Eur. J. Biochem.* 212, 549-555

26. Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. (2000) *J. Mol. Biol.* 298, 477-491

27. Raj, S. S., Kibushi, E., Kurasawa, T., Suzuki, A., Yamane, T., Odani, S., Iwasaki, Y., and Ashida, T. (2002) *J. Biochem. (Tokyo)* 132, 927-933

28. Song, H. K., and Suh, S. W. (1998) *J. Mol. Biol.* 275, 347-363

29. Suzuki, A., Yamane, T., and Ashida, T. (1993) *J. Mol. Biol.* 234, 722-734

30. McPhalen, C. A., and James, M. N. (1987) *Biochemistry* 26, 261-269

31. Bode, W., Epp, O., Huber, R., Laskowski, M. J., and Ardelt, W. (1985) *Eur. J. Biochem.* 147, 387-395

32. Huber, R., and Bode, W. (1978) *Acc. Chem. Res.* 11, 114-122

33. Fujinaga, M., Read, R. J., Sielecki, A. R., Ardelt, W., Laskowski, M. J., and James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U S A* 79, 4868-4872

34. Kleywegt, G. J., and Jones, T. A. (1996) *Acta Cryst. D* 52, 842-857

CHAPTER 5:

DESIGN, CONSTRUCTION AND EVALUATION OF A SINGLE DOMAIN DELETION MUTANT OF *ASCARIS* PEPSIN INHIBITOR-3

5.0 Introduction

The nematode *Ascaris suum* is an extremely common intestinal parasite in pigs and its close relative, *Ascaris lumbricoides*, infects over 1 billion people (1). The *A. suum* life cycle consists of a number of stages (Figure 5-1) (2). Adult *A. suum* worms reproduce sexually and the eggs, which can remain viable for years in soil, are released by the female in the host's feces. Hosts are infected through the ingestion of *A. suum* eggs. After passing through the stomach and into the small intestine of the host, the larvae shed the egg casings before burrowing into the walls of the small and large intestines. These second stage larvae enter the host's bloodstream, go through the liver and enter the lungs where molting into third stage larvae occurs. Once in the lungs, the larvae produce irritants which make the host cough. Coughing allows the larvae to enter the throat where they are then swallowed. After passing through the digestive system of the host for a second time, the larvae reach maturity and remain in the small intestine.

Because *Ascaris* lives in the digestive tract of its host, it produces a large number of proteinase inhibitors to protect itself from digestion. *A. suum* produces three protein inhibitors of pepsin, the most abundant being pepsin inhibitor-3 (PI-3). PI-3 is a potent inhibitor of the gastric enzymes pepsin and gastricsin, as well as the non-gastric aspartic proteinase cathepsin E (K_i values ~ 1 to 75 nM) (3-5).

PI-3 consists of 149 residues and contains three disulfide bonds (4,6). The crystal structure of PI-3 reveals an extended structure consisting of two domains, each containing an antiparallel β-sheet flanked by a single α-helix (7). In the crystal structure of the complex formed between PI-3 and porcine pepsin, there are three main contact points between enzyme and inhibitor: (1) the N-terminal β-strand of PI-3 (residues 4 to 8) forms antiparallel β-sheet hydrogen bonding interactions with residues 70 to 74 of pepsin, (2) the C-terminal polyproline helix (residues 138 to 142) forms van der Waals contacts with residues 289 to 294 of pepsin, and (3) the N-terminal three residues of the inhibitor occupy the S1' to S3' binding pockets of the enzyme (7). The primary mechanism of inhibition appears to be the steric blockage of substrate binding by the N-terminus of the inhibitor.

Remarkably, all of the contacts between inhibitor and enzyme are restricted to the C-terminal domain. In addition, there are few contacts between the two domains in the extended structure of PI-3. Thus, the structure suggests that if the N-terminal domain (residues 11 to 69) were deleted and a new β-turn were allowed to form between the N-terminal β-strand and the second β-strand in the C-terminal domain, the resulting single-domain deletion mutant would still retain inhibitory activity against pepsin.

A single-domain deletion mutant of PI-3 that retains inhibitory activity against aspartic proteinases would be an attractive target for the further engineering of inhibitory specificity. Native PI-3 is difficult to express in recombinant expression systems, probably at least in part because of the presence of three disulfide bonds. The need to refold the protein *in vitro* is a serious deterrent to efforts at re-engineering the binding specificity of PI-3 to produce potent inhibitors of medically important aspartic

proteinases like HIV protease, β-secretase, renin and various fungal aspartic proteinases. Because two disulfide bonds are formed in the N-terminal domain, deletion of this domain would be expected to improve the yield of properly folded protein in recombinant expression systems.

As a test of the hypothesis that a single-domain deletion mutant of PI-3 would retain inhibitory activity against pepsin, a synthetic gene was designed and assembled, the single-domain deletion mutant was expressed and re-folded, and the secondary structure and inhibitory activity of the deletion mutant was evaluated using circular dichroism spectroscopy and enzyme inhibition assays.

5.1 Materials and methods

5.1.1 Construction of mPI3Δ1 expression vector

Four overlapping oligonucleotides were synthesized by Sigma Genosys (St. Louis, MO) to construct the synthetic gene for mPI3Δ1 (Figure 5-2). Codons were chosen for optimal expression of mPI3Δ1 in *E. coli*. Oligonucleotides were annealed and extended with Klenow fragment of DNA polymerase I according to the scheme presented in Figure 5-3. The product from each primer extension reaction was purified from an 8% non-denaturing polyacrylamide gel, extracted using the "crush-and-soak" technique and amplified with Vent polymerase. The amplified products were gel purified from 1% agarose using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and combined to form the template for the final PCR amplification of full-length mPI3Δ1. A C-terminal histidine tag was then added to the C-terminal end of mPI3Δ1 using the following oligonucleotide: 5'-CGGGATCCTTAGTGATGATGATGGTGGTGGCCCTGCACGG TGCAAAAGCTCGG-3'. The histidine tag was inserted at the C-terminal end of mPI3Δ1 so as not to interfere with its interaction with pepsin, which according to the crystallographic structure of PI-3 (7), occurs through the N-terminus. Full-length mPI3Δ1 with a C-terminal histidine tag was digested with restriction enzymes, and inserted into the *Bam*H1-*Nde*1 site of the expression vector pET3a (Novagen, Madison, WI). All molecular biology procedures were performed as previously described (8) to construct the mPI3Δ1 expression vector. All enzymes utilized in constructing the expression vector were obtained from New England BioLabs (Beverly, MA). The identity of mPI3Δ1 was confirmed by DNA sequencing (DNA Core Facility, Department of Biochemistry, University of Alberta).

5.1.2 Expression and purification of mPI3Δ1

Freshly transformed overnight cultures of *E. coli* BL21(DE3)pLysS cells containing the plasmid for expression of mPI3Δ1 were added at 37°C to 1 L of Luria broth (LB) medium that contained ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL). At OD_{600} = 0.5, expression was induced with isopropyl β-D-galactoside (IPTG) to a final concentration of 1 mM. The cells were further grown at 30°C, harvested 8 h after induction (final OD_{600} of 0.9), centrifuged, resuspended in 50 mL of buffer A (6 M guanidinium hydrochloride, 10 mM Tris-HCl, 100 mM NaH_2PO_4, and 10 mM β-mercaptoethanol, pH 8.0) and centrifuged at 12,000 rpm for 30 minutes. The high-affinity column refolding technique outlined by Zahn *et al.* (9) was used with modifications to purify and refold mPI3Δ1 from inclusion bodies. The soluble fraction of the resuspended cells was incubated with gentle rocking at 4°C for 60 minutes with 15 mL of nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen) previously washed in buffer A. The resulting slurry was then poured into a column and left to settle by gravity at 4°C for 12 hours. The unbound fraction was collected for analysis and the bound resin was washed with 60 mL of buffer A. The wash fraction was also collected for analysis. A 200 mL gradient of buffer A to buffer B (10 mM Tris-HCl, 100 mM NaH_2PO_4, pH 8.0) was then applied. Protein impurities devoid of histidine tails were removed from the Ni-NTA resin with 40 mL of buffer C (10 mM Tris-HCl, 100 mM NaH_2PO_4 and 50 mM imidazole, pH 8.0). The His-tagged mPI3Δ1 was eluted with 50 mL of buffer D (10 mM Tris-HCl, 100 mM NaH_2PO_4 and 500 mM imidazole, pH 5.8). Purified protein was analyzed by SDS-polyacrylamide gel electrophoresis, N-terminal sequencing, and MALDI-TOF mass spectrometry.

5.1.3 Circular dichroism spectroscopy

Far-UV (192-250 nm) CD spectra were obtained using a Jasco J-720 spectrapolarimeter that was interfaced with Jasco software and equipped with a 0.02-cm quartz cuvette cell. For circular dichroism measurements, mPI3Δ1 was used at a concentration of 0.4 mg/mL in 50 mM Tris-Cl, pH 7.8, 1 mM EDTA. The percentage of secondary structure in the sample was estimated using the methods of Provencher and Glockner (10) and Venyaminov (11,12).

5.1.4 Protein homology modeling

Residues 11 to 69 inclusively were deleted from the crystal structure of PI-3 (PDB code 1F32) (7). The Monte Carlo Simulated Annealing algorithm implemented in the program LPSA (13) was used to generate and minimize models of the loop connecting residues 10 and 70 in the deletion mutant mPI3Δ1.

5.1.5 Enzyme inhibition kinetics

Proteinase inhibitory activity was assessed by measuring the inhibition of the initial rate of cleavage of a chromogenic peptide substrate (Pro-Thr-Glu-Phe-(NO_2-Phe)-Arg-Leu, Bachem) by porcine pepsin (14). The inhibition of pepsin was measured after preincubating the enzyme with mPI3Δ1 for 20 minutes at 30°C in 0.85 to 0.99 mL of assay buffer (0.1 M sodium acetate, pH 5.0). Incubation time, temperature and pH were chosen based on a survey of reaction conditions to optimize interactions between the proteinase and the inhibitor, while retaining high intrinsic activity and stability in pepsin (14). 0.01 to 0.15 mL of substrate (1 mM) was added to the pepsin-inhibitor mixture, bringing the final volume to 1 mL. The change in absorbance at 300 nm was measured for two minutes using a spectrophotometer (Ultrospec 2000, BioChrom Ltd.) equipped

with a Peltier-heated cell holder set to 30°C. The initial rate kinetics were fit to a Michaelis-Menten model using nonlinear regression methods implemented in Sigmaplot using the following equation:

$$V_o = V_{max}[S]/(K_m + [S])$$

K_i was determined by fitting data to a competitive inhibition model:

$$V_i = V_{max}[S]/(K_m(1 + [I]/K_i) + [S])$$

5.2 Results and discussion

5.2.1 Design and construction of a synthetic gene for mPI3Δ1

The structure of PI-3 suggests that residues 11-69 can be deleted to create a single-domain deletion mutant (mPI3Δ1) consisting of residues 1-10 and 70-149. Although this single-domain mutant would be expected to retain all of the binding interfaces present in the complex of full-length PI-3 and pepsin, there are at least two additional problems that may impair the inhibitory activity of the deletion mutant. First, a new β-turn must form to connect residues 10 and 70. Second, hydrophobic residues in the deletion mutant may be exposed to the solvent if these residues lie at the interface between the two domains of PI-3 and pack against residues in the N-terminal domain.

To address the first concern, models of the deletion mutant were constructed by deleting residues 11-69 from the structure of PI-3, and new β-turns were constructed by the program LPSA (13). The resulting models were inspected for distorted geometry and unsatisfied hydrogen bonding as possible indicators of introducing an unfavourable β-turn as a result of deleting residues 11-69. Several models produced by LPSA appeared to give reasonable turn geometry and one is shown in Figure 5-5. There are no disallowed main chain torsion angles in the modeled loop, and a hydrogen bond is formed between the amide nitrogen of Gly9 and the carbonyl oxygen of Pro10 (Figure 5-6). Hydrogen bonds that stabilize the conformation of the loop are also formed between the side chain hydroxyl group of Thr8 and the amino group of Lys13, as well as between the carboxylate oxygen of Asp11 and the main chain amide nitrogen of Met12.

To address the second concern, a survey of exposed hydrophobic residues was conducted using the model of mPI3Δ1 constructed using LPSA. Only two hydrophobic

residues, Leu23 and Phe28 in mPI3Δ1 (Leu82 and Phe87 in PI-3), appeared to be more exposed to solvent as a result of deleting residues 11-69. Several different side chains were modeled as possible replacements for these hydrophobic residues. It appeared that replacing these residues with tyrosine would provide a good balance of hydrophobic packing and hydrophilicity to replace the roles of these hydrophobic residues without introducing a new and potentially destabilizing hydrophobic surface patch in mPI3Δ1.

Once these design considerations were evaluated, synthetic oligonucleotides encoding mPI3Δ1 were obtained (Figure 5-2). The sequence of the oligonucleotides was designed to maximize the use of the most commonly found codons in *E. coli* and to introduce restriction enzyme cleavage sites to facilitate cassette mutagenesis studies in the future. The overlapping oligonucleotides were annealed and the complete coding sequence of mPI3Δ1 was assembled using several steps of PCR and agarose gel purification of intermediate products (Figure 5-3).

5.2.2 Re-folding and assessment of the folded structure of mPI3Δ1

Although it was hoped that the reduced number of disulfide bonds in mPI3Δ1 compared with PI-3 would facilitate the correct folding of mPI3Δ1 in recombinant expression systems, high-level expression of mPI3Δ1 in *E. coli* produced mostly insoluble mPI3Δ1. This insoluble protein was assumed to be mis-folded and was solubilized in guanidinium hydrochloride and β-mercaptoethanol. The denatured protein was bound to Ni-NTA agarose and re-folded by passing a gradient of decreasing concentrations of denaturant and β-mercaptoethanol through the column. Approximately 4 milligrams of soluble protein were recovered from 2 litres of bacterial culture (Figure 5-4).

The re-folded protein could be concentrated to at least 2 mg/mL. A sample was analyzed by MALDI-TOF mass spectroscopy to determine the molecular mass (10019.6 Da, predicted mass of 10027.38 Da) and a sample was also subjected to Edman degradation to confirm the integrity of the N-terminus (TQFLF). To evaluate the secondary structural content of the re-folded protein, the sample was analyzed by far-UV circular dichroism spectroscopy. Comparisons of the observed spectrum with spectra obtained from a library of proteins with known amounts of secondary structure (10-12) indicated the presence of approximately 18% α-helix, 37% β-sheet and 20% β-turn, which compares quite reasonably with values calculated from the model of mPI3Δ1 (20% α-helix, 24% β-sheet and 17% β-turn) (Figure 5-7).

5.2.3 Enzyme inhibition kinetics

The ability of mPI3Δ1 to inhibit the hydrolysis of a chromogenic peptide substrate by porcine pepsin was evaluated using spectrophotometry. Initial rate kinetics both in the presence and in the absence of mPI3Δ1 followed the Michaelis-Menten model and clearly revealed the inhibition of pepsin catalyzed peptide hydrolysis at concentrations of inhibitor ranging from 0.28 to 0.73 μM (Figure 5-8). The pattern of inhibition was consistent with competitive inhibition, but only a narrow range of inhibitor concentrations could be evaluated due to limitations in the sensitivity of the assay. Nonlinear regression was used to fit kinetic parameters, yielding a K_i value of 0.5 μM. This value is dramatically higher than the K_i value of 2 nM reported for PI-3 inhibition of porcine pepsin (3-5).

The significantly weaker inhibition of pepsin by mPI3Δ1 may reflect the destabilization of the structure of the inhibitor caused by the deletion of the N-terminal

domain. Alternatively, it is possible that the sample of re-folded mPI3Δ1 may not be conformationally homogeneous due to the incomplete re-folding of denatured mPI3Δ1. Preliminary gel filtration experiments suggest that a fraction of re-folded mPI3Δ1 is digested by pepsin, and this fraction may represent mis-folded mPI3Δ1. If only a fraction of the mPI3Δ1 sample is properly folded, then the actual K_i value would be lower than the value obtained when assuming that the entire sample is properly folded and active.

Figure 5-1: *Ascaris suum* life cycle. In *A*, Females lay eggs that are excreted in the host's feces. In *B* and *C*, the egg develops into a second stage larvae. In *D*, insects such as earthworms sometimes ingest the eggs and carry the parasite. In *E*, pigs are infected through the ingestion of second stage larvae or insects. In *F* through *I*, the larvae travel through the digestive system, then to the liver and the lungs where molting occurs. Coughing leads to swallowing of the parasite which then passes through the digestive system again to remain in the small intestine. Reproduced from (2).

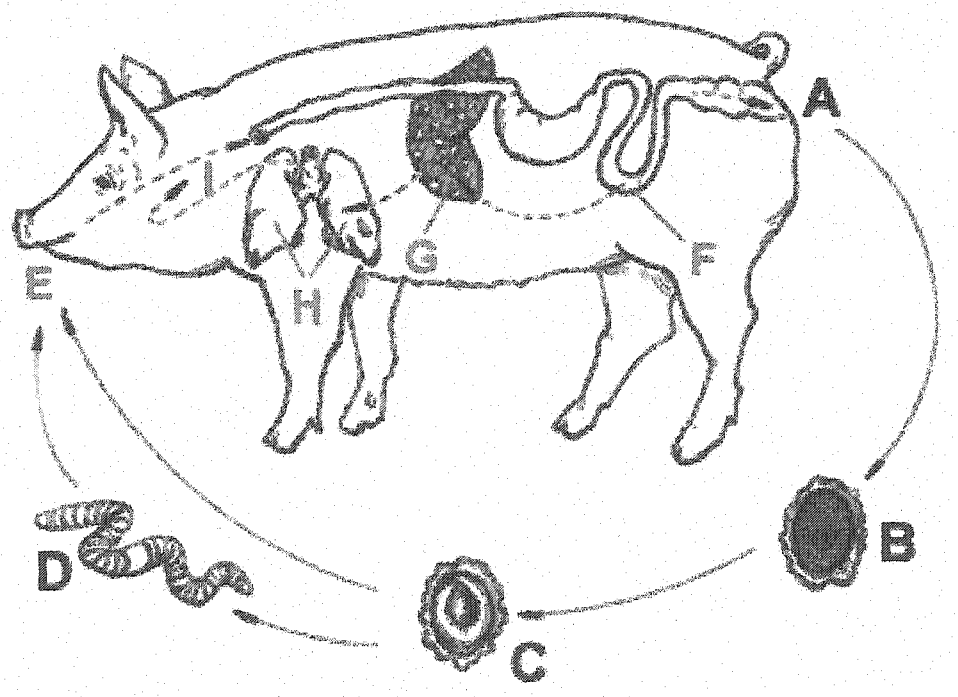


Figure 5-2: Synthetic gene of *A. suum* mPI3Δ1. Codons were chosen for optimal expression of mPI3Δ1 in *E. coli*. The four synthetic oligonucleotides are denoted by grey shading; the oligonucleotides of the upper DNA sequence are read 5' to 3' (left to right) and the oligonucleotides of the lower DNA sequence are read 3' to 5' (left to right).

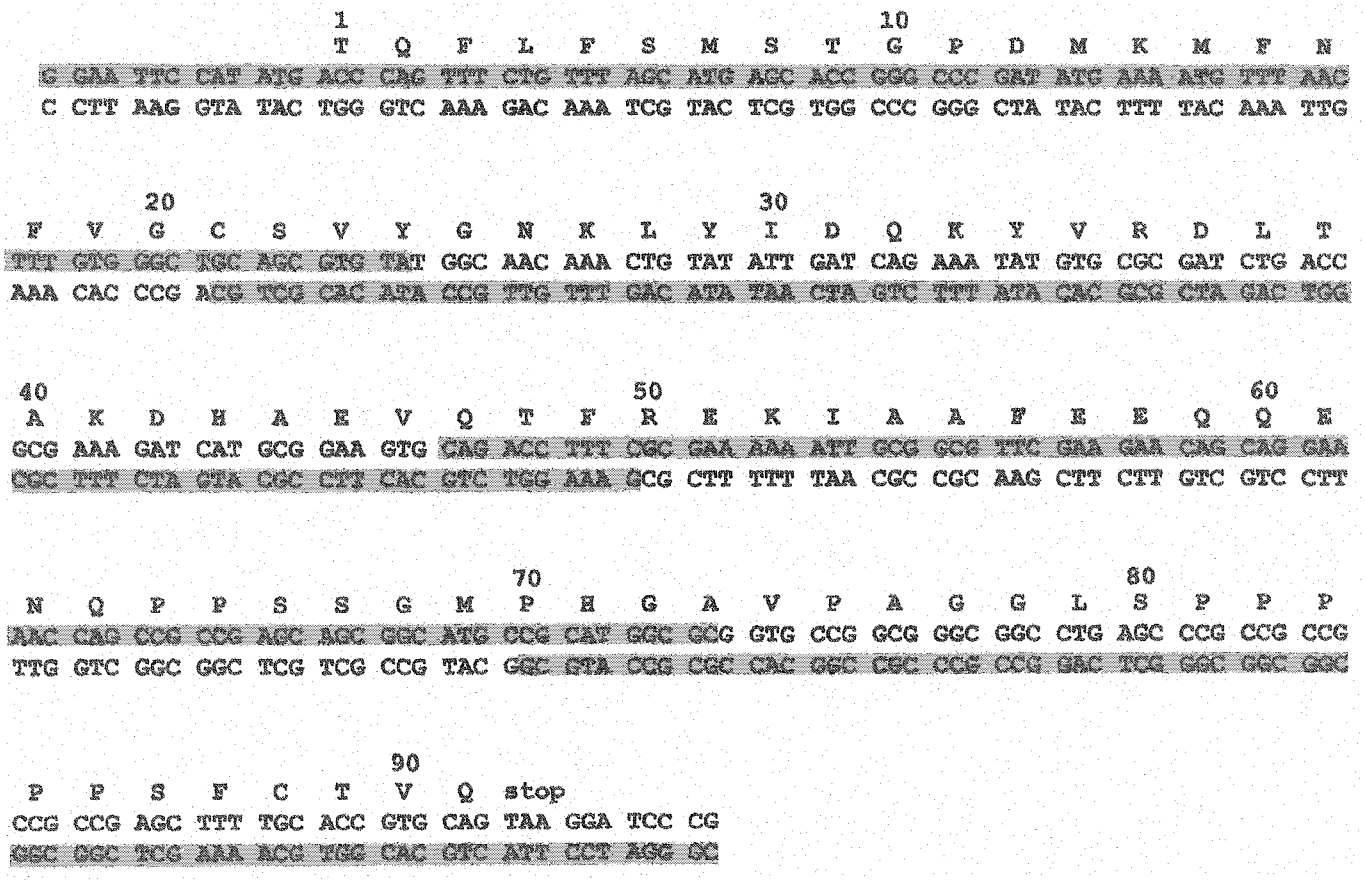


Figure 5-3: Schematic diagram of *A. suum* mPI3Δ1 gene synthesis.

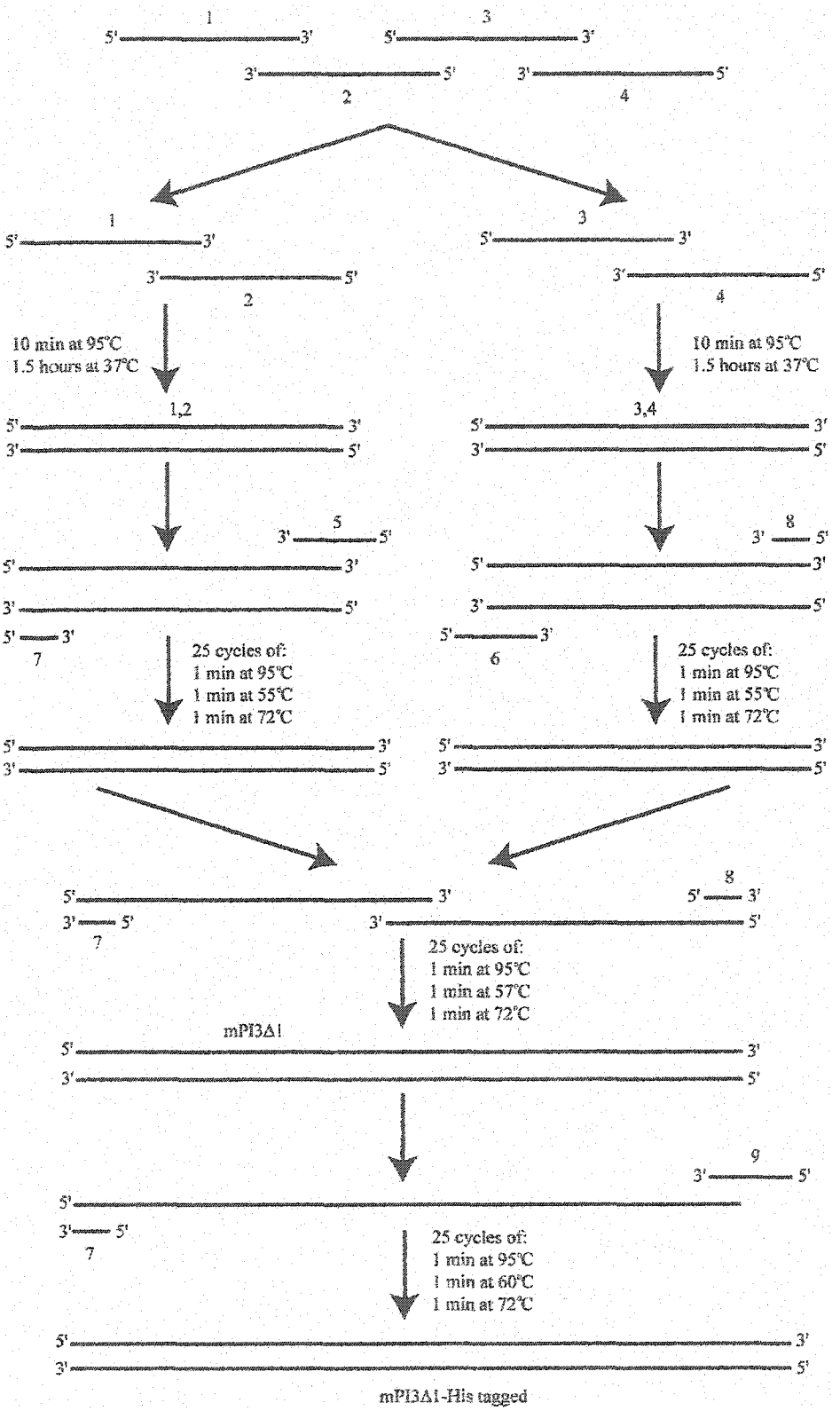


Figure 5-4: SDS-PAGE of *A. suum* mPI3Δ1 purification. Lane 1, native PI3 with expected molecular weight of 16.5 kDa; lane 2, molecular weight markers; lane 3, mPI3Δ1 after elution from Ni-NTA resin (expected molecular weight of 10 kDa).

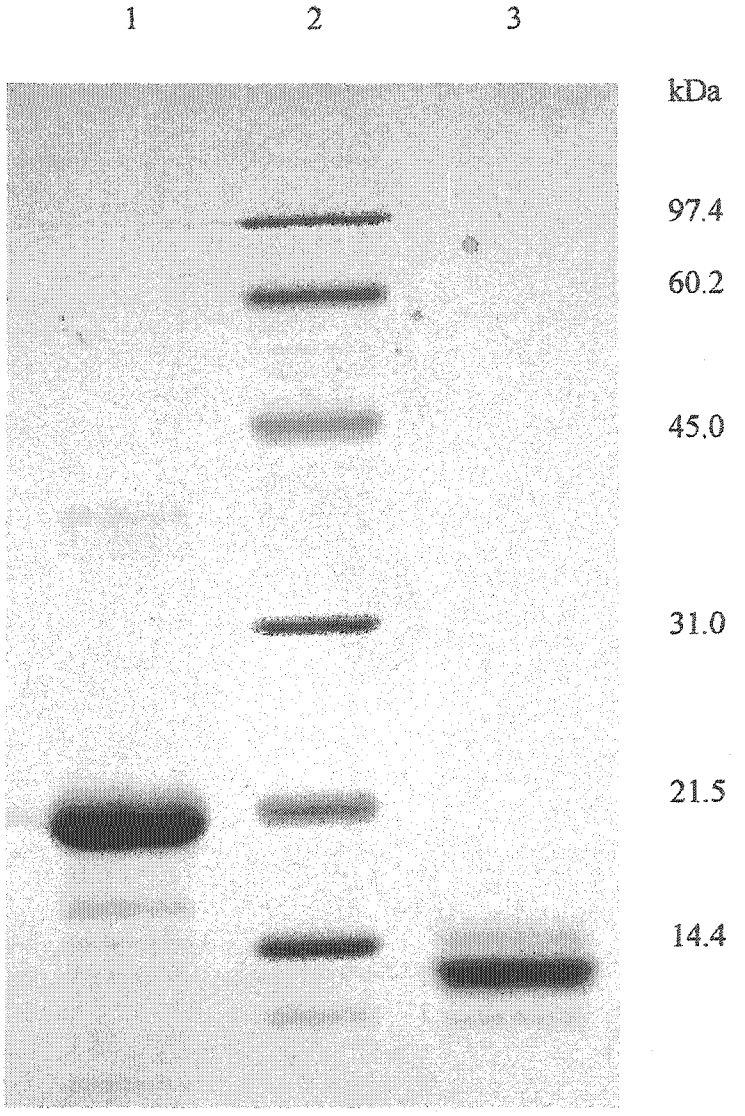


Figure 5-5: Model of the structure of mPI3Δ1. β-strands are coloured in cyan and the α-helix is coloured in red. The loop replacing residues 11 to 69 is marked on the figure as residues 9, 10 and 11.

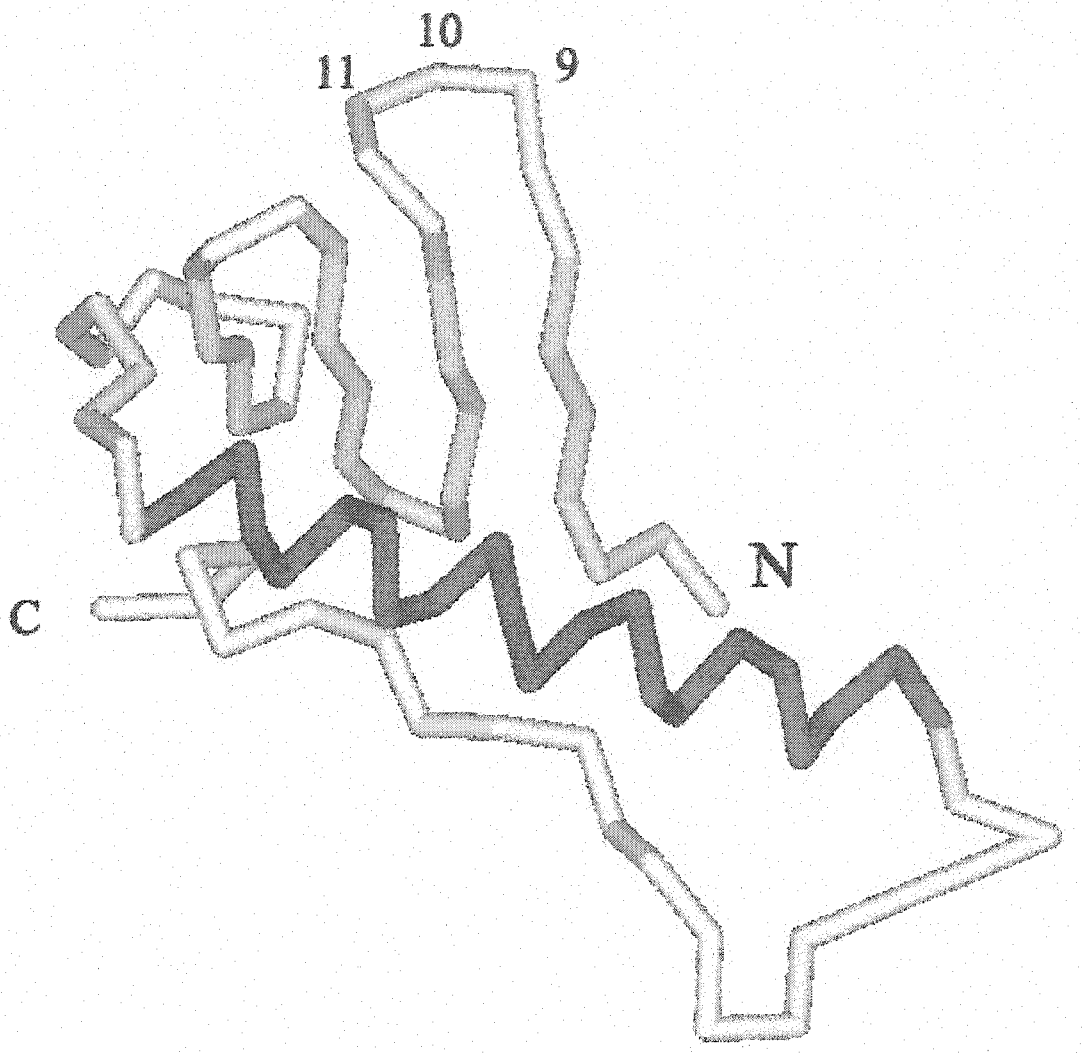


Figure 5-6: Modeled structure of the loop replacing residues 11 to 69. Hydrogen bonds are shown as green lines.

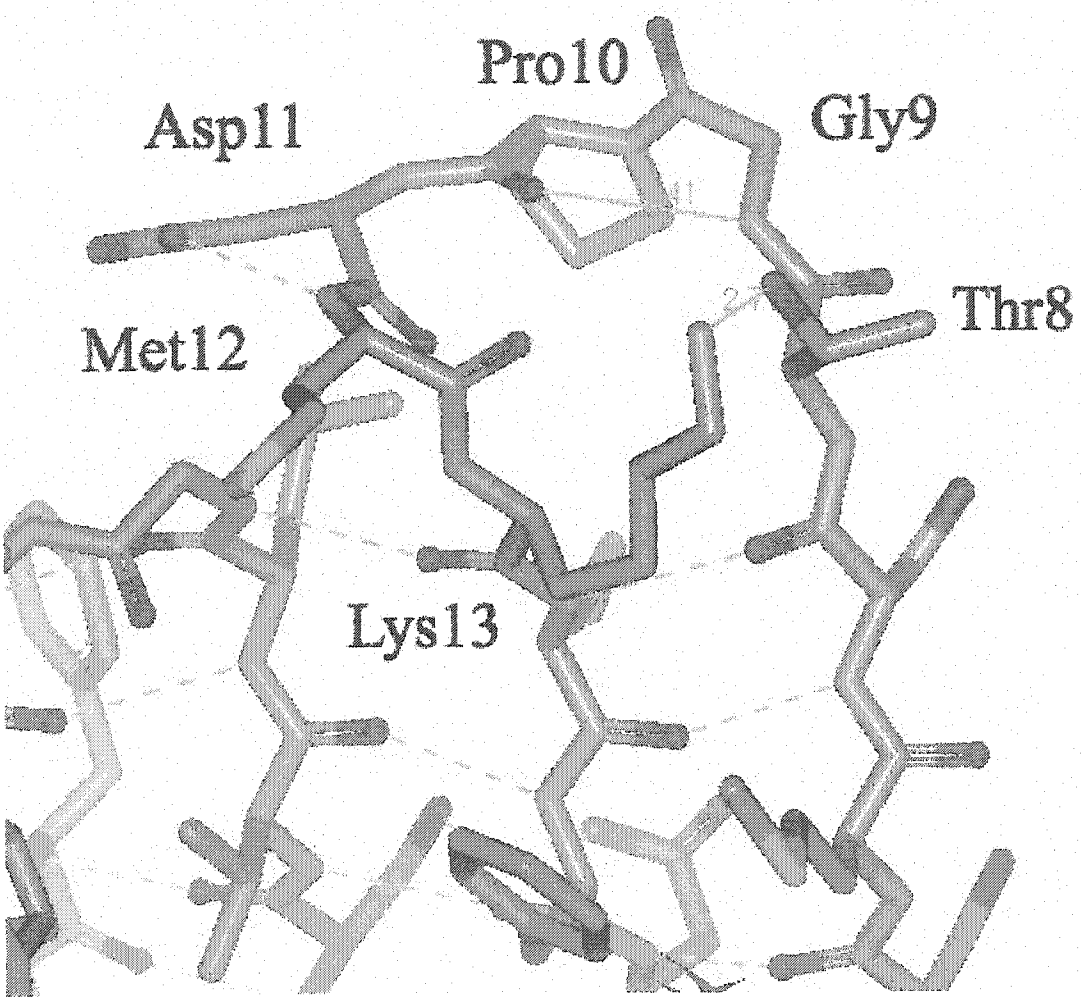


Figure 5-7: Far UV CD spectra obtained for mPI3Δ1 at a concentration of 0.4 mg/mL in 50 mM Tris-Cl, pH 7.8, 1 mM EDTA. The spectra was obtained using a Jasco J-720 spectrapolarimeter that was interfaced with Jasco software and equipped with a 0.02-cm quartz cuvette cell.

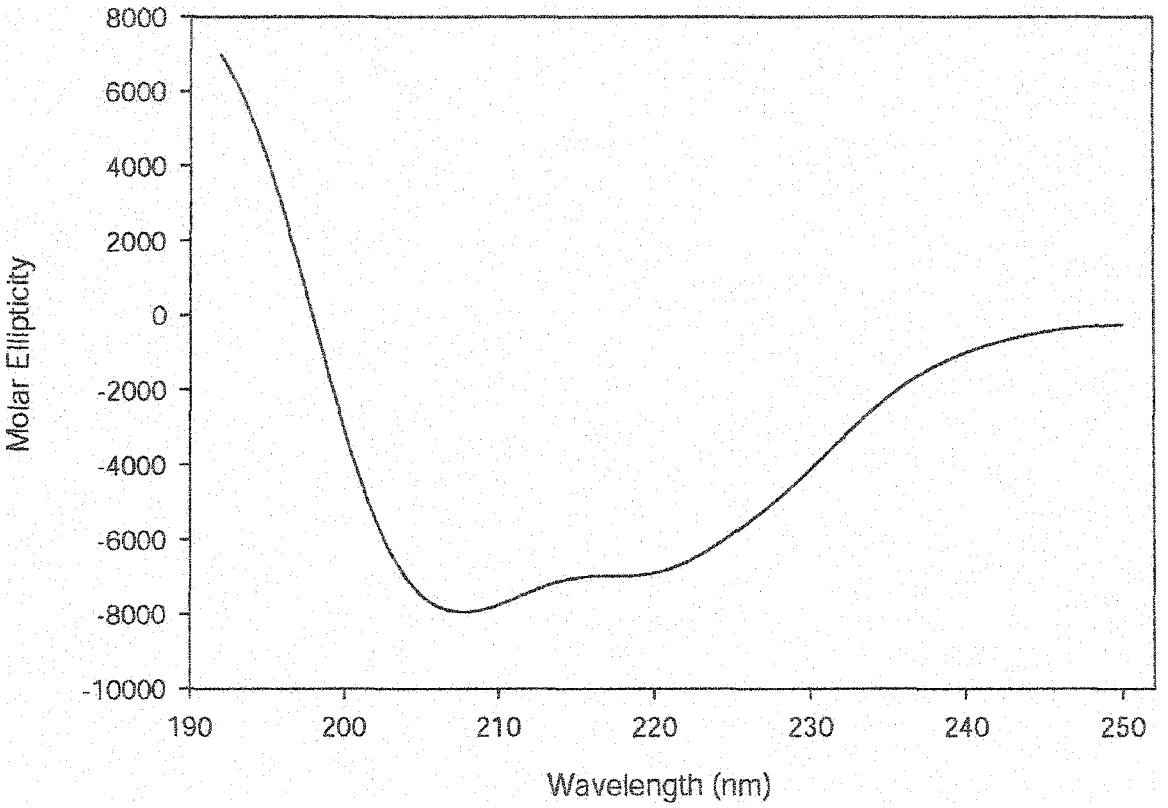
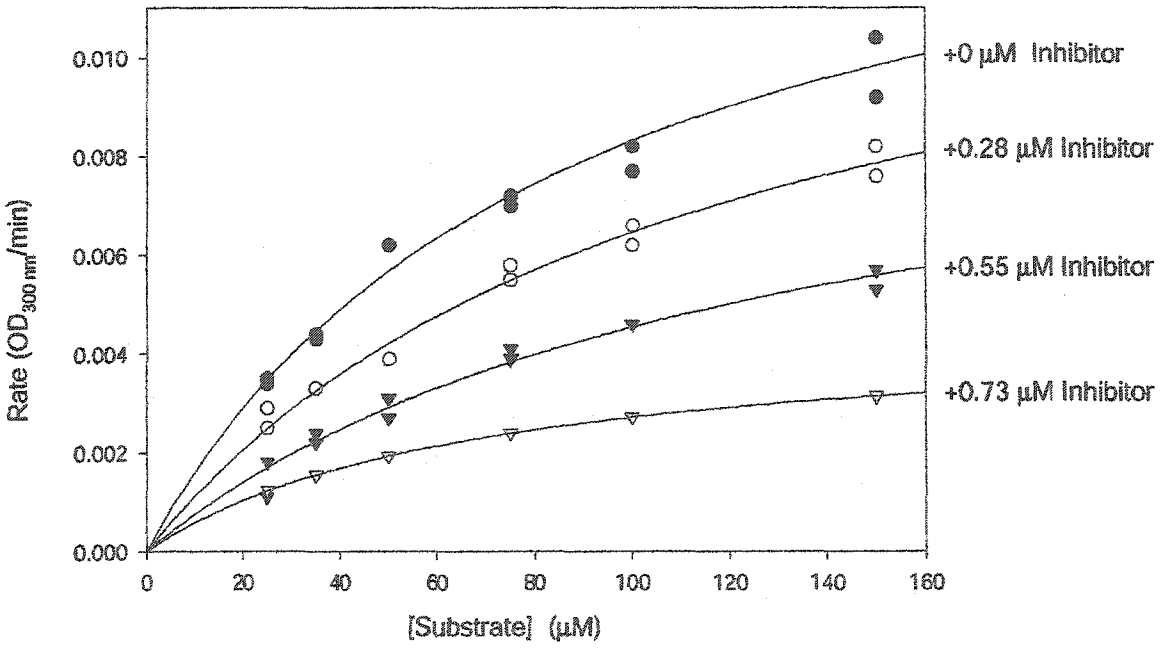


Figure 5-8: Initial rate of hydrolysis of the substrate Pro-Thr-Glu-Phe-(NO_2-Phe)-Arg-Leu by porcine pepsin in the presence of mPI3Δ1 concentrations ranging from 0 to 0.73 μM. Data are fit to a Michaelis Menten model using non-linear regression.



5.3 References

1. Levine, N. D. (1980) *Nematode parasites of domestic animals and of man*, Burgess Publishing Co., Minneapolis, Minnesota
2. Johnstone, C. (2000), http://cal.nbc.upenn.edu/merial/Default.htm
3. Kageyama, T. (1998) *Eur. J. Biochem.* 253, 804-809
4. Abu-Erreish, G. M., and Peanasky, R. J. (1974) *J. Biol. Chem.* 249, 1566-1571
5. Valler, M. J., Kay, J., Aoyagi, T., and Dunn, B. M. (1985) *J. Enzyme Inhibition* 1, 77-82
6. Martzen, M. R., McMullen, B. A., Smith, N. E., Fujikawa, K., and Peanasky, R. J. (1990) *Biochemistry* 29, 7366-7372
7. Ng, K. K., Petersen, J. F., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., Dunn, B. M., Martzen, M. R., Peanasky, R. J., and James, M. N. (2000) *Nat Struct Biol* 7, 653-657
8. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning, A Laboratory Manual*, 3rd edition Ed. 3 vols., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
9. Zahn, R., von Schroetter, C., and Wuthrich, K. (1997) *FEBS Let.* 417, 400-404
10. Provencher, S. W., and Glockner, J. (1981) *Biochemistry* 20, 33-37
11. Venyaminov, S., Baikalov, I. A., Wu, C. S., and Yang, J. T. (1991) *Anal Biochem* 198, 250-255
12. Venyaminov, S., Baikalov, I. A., Shen, Z. M., Wu, C. S., and Yang, J. T. (1993) *Anal Biochem* 214, 17-24

13. Zhang, H., Lai, L., Wang, L., Han, Y., and Tang, Y. (1997) *Biopolymers* 41, 61-72

14. Dunn, B. M., Kammermann, B., and McCurry, K. R. (1984) *Anal Biochem* 138, 68-73

CHAPTER 6:

CONCLUSIONS AND FUTURE DIRECTIONS

6.0 Looking beyond the structures

The studies undertaken in this thesis have attempted to obtain basic physical information about the structure and design of proteinases and their inhibitors. In the first part of this thesis, the serine proteinases are studied. In particular, the viral chymotrypsin-like serine proteinase nonstructural protein 4 (nsp4) of *Equine Arteritis Virus* (*EAV*) is examined. To elucidate the structural basis of proteolytic processing in arteriviruses, the structure of *EAV* nsp4 has been solved to 2.0 Å resolution using the multiple isomorphous replacement technique (see Chapter 2).

This crystallographic study reveals that nsp4 adopts the smallest known chymotrypsin-like fold with a canonical catalytic triad of Ser-120, His-39 and Asp-65, as well as a novel α/β C-terminal extension domain (see Figure 2-1A). Remarkably, nsp4 adopts alternate conformations of the oxyanion hole, which may be a novel means of regulating proteolytic activity. Also, nsp4 contains two β-barrel domains with the N-terminal barrel containing six strands (A1 to F1) and the C-terminal containing seven (A2 to G2). The seventh β-strand in the C-terminal β-barrel is not common in chymotrypsin-like proteinases (CLPs), and has, to date, only been reported in the structures of *Sindbis virus* core protein (1) and *Semliki forest virus* core protein (2).

The most striking feature of the nsp4 structure is the presence of an additional C-terminal domain not found in most other CLPs. This third domain comprises residues 156 to 204 and consists of two short pairs of β-strands and two α-helices. The only other

three-dimensional structure of a CLP with an additional C-terminal domain is the recently reported structure of a coronavirus chymotrypsin-like cysteine main proteinase from the *Transmissible Gastroenteritis virus* in pigs (3). Coronaviruses comprise the distantly related second family in the order Nidovirales (4). However, unlike nsp4, the structure of the C-terminal domain in the coronavirus proteinase is nearly twice as large at 110 residues and is comprised of only α-helices. In addition, the linker between the C-terminal domain and the C-terminal β-barrel differs markedly from that found in nsp4 due mostly to the presence of the seventh β-strand in the C-terminal β-barrel in nsp4. Despite the lack of sequence and structural similarity, the C-terminal domains in both arteriviruses and coronaviruses may share a common functionality in facilitating substrate recognition.

Deletion mutants in the coronavirus proteinase indicate that residues 1 to 5, as well the α-helical extension domain are involved in proteolytic activity by maintaining a presumed substrate-binding loop consisting of residues 184 to 199 in a certain position (3). Similar studies on the α/β-extension domain are currently being pursued by Dr. Eric Snijder's group. Comparisons of the results from these studies with those obtained for coronavirus main proteinase will be informative for understanding the function of the extra domain even though major differences exists between these two enzymes. This C-terminal extension domain may mediate the formation of multi-protein complexes to control proteolytic processing pathways critical to the viral life cycle. The structure determination of other members of the order Nidovirales should help to reveal the function of this interesting extension domain. Further studies should also concentrate on obtaining the structure of nsp4 in complex with substrate or an inhibitor which would

greatly advance our understanding of proteolytic processing in arteriviruses. Understanding the structural details of proteolytic processing in arteriviruses and other nidoviruses will allow for the design of novel antiviral agents.

The second part of this thesis explores multidomain proteinaceous inhibitors of serine proteinases. In Chapter 3, the crystallographic study of the ternary complex between tomato inhibitor-II (TI-II) and subtilisin Carlsberg is presented. The structure of this ternary complex was solved to 2.5 Å using molecular replacement. The 1:2 complex of TI-II bound to subtilisin Carlsberg reveals how multidomain Pot II inhibitors can simultaneously inhibit several proteinases within a single complex. Each domain of TI-II interacts with a single proteinase molecule primarily through the reactive site loop but also through secondary contacting regions mostly within the same domain (see Figure 3-1A). The extended conformation adopted by TI-II enables this inhibitor to form 1:2 complexes without encountering steric clashes between the proteinase molecules. The ternary complex of TI-II with subtilisin Carlsberg represents the first structure of a full-length Pot II inhibitor. However, no electron density could be observed for the linker joining the two domains of TI-II. Recently, a new crystal form diffracting to higher resolution has been obtained for the ternary complex. Crystallographic analysis on this crystal is currently underway and it will be interesting to see whether electron density will be present for this missing region.

Chapter 4 describes the crystallographic structure determination of unbound TI-II which was solved to 2.15 Å using molecular replacement. The structures of four independent copies of the unbound form of the two-headed tomato inhibitor-II (TI-II) reveal significant conformational flexibility that may be of functional importance in the

inhibition of a wide range of proteinases (see Figure 4-1A). TI-II is an integral part of the constitutive and inducible defensive mechanisms that protect plants from attacking pests (bacteria, fungi and insects). To this effect, the structure must have sufficient flexibility to be able to accommodate to the substrate binding clefts of a wide range of proteinases. High-resolution structures of TI-II in bound and unbound forms reveal that flexibility in the orientation of the two inhibitory domains as well as in the reactive site loops allow the inhibitor to achieve a balance between tight binding and broad specificity.

The structure determination of TI-II in the bound and unbound forms represents an initial step towards understanding how a large number of multidomain Pot II inhibitors utilize multiple inhibitory domains to defend plants against predation by a wide variety of pests. These structures provide a structural framework for the design of multidomain inhibitors carrying different specificities within a single polypeptide. When questions concerning the exact mechanism of inhibition are better answered, engineering better inhibitors for the purpose of controlling pests in economically important crops becomes more feasible.

During each growing season, it is estimated that 14% of the world's agricultural production (52% in wheat, 83% in rice, 59% in maize, 74% in potato, 58% in soybean and 84% in cotton) is lost due to plant pathogens (5). Although chemical insecticides are important for crop protection, concerns about toxicity and development of resistance in pests have prompted a demand for alternative, integrated pest management strategies. The feasibility of using PIs of the Pot II family in transgenic plants to increase their resistance to pathogens is alluring and has already been examined in tobacco (6) and rice (7).

Serine proteinases are the main digestive enzymes in many insect families (8-10). These insect families are consequently great targets for the engineering of high potency Pot II inhibitors. Other plant pathogens such as bacteria (e.g., *Pseudomonas syringae*, which causes bacterial speck disease in tomato (11)) and fungi (e.g., *Phytophthora infestans* which causes potato blight (12)) have also been shown to elicit the accumulation of TI-II in tomato plants. Proteinase inhibitors isolated from tomato were also reported to inhibit proteinases from *Fusarium solani* (13) as well as inhibiting the growth of microorganisms present in soil or rotting tissues (14). Clearly, naturally-occurring TI-II provides an effective defence against a wide range of pathogens, and variants may be engineered to have higher specificity and potency against a broader range of pathogens.

Using the structure of the bound form of TI-II, initial modeling studies were performed where the structure of TI-II was docked onto bovine chymotrypsin and trypsin. Table 6-1 summarizes the contacts observed between TI-II and the proteinases. Similar analyses could be performed using pest chymotrypsin or trypsin sequences. Several structures and sequences are now becoming available for pest proteinases (15-19). Future studies should focus on obtaining structures of complexes of TI-II with pest proteinases in order to gain a better understanding of the structural and specificity requirements for the engineering of better inhibitors.

The final part of this thesis concerns the design of proteinaceous inhibitors of aspartic proteinases (see Chapter 5). A single-domain inhibitor was designed based on the structure of the two-domain pepsin inhibitor-3 from *Ascaris suum* (20). A model for the single-domain inhibitor indicated that the sequence could be expected to form a stable

structure in the absence of residues 11-69 of PI-3 (see Figure 5-5). A synthetic gene was constructed using synthetic oligonucleotides and several rounds of PCR. The single-domain inhibitor was expressed at high levels in an insoluble form in *E. coli*. The insoluble inhibitor could be renatured, yielding an inhibitor of porcine pepsin with K_i of 0.5 μM. Inhibitory activity is retained in the absence of the N-terminal domain but the potency of the inhibitory appears to be dramatically reduced. Future work to improve the potency of the single-domain inhibitor could include modifying the connection between the N-terminal β-strand and the remainder of the β-sheet in the C-terminal domain, optimizing the protein re-folding protocol and further purification of inhibitor to remove misfolded conformers. Another avenue of research would be to explore designing novel inhibitory specificity in mPI3Δ1 to allow inhibition of more medically important proteinases such as β-secretase which is involved in Alzheimer's disease (21) and cathepsin D which is involved in metastasis and breast cancer (22).

Table 6-1: Contacts between TI-II and Proteinases. van der Waals contacts are coloured black, side-chain hydrogen bonding contacts are coloured red and main-chain hydrogen bonding contacts are coloured blue.

(A) Reactive-site loop of chymotrypsin-binding domain

	TI-II	Subtilisin	Chymotrypsin
P6	P57	Y104, S130	S217, S221, T224
P5	N58	S101, G102	S218
P4	A59	L96, I107, G127	W215, V227
P3	C60	G127	S217, S218, G216
P2	T61	D32, T33, H64, L96	
P1	F62	A152, N155, T22, S221, S125, N155	S189, S190, C191, M192, V213, G216, S217, C220, S214, G193,
P1'	N63	H64, S221	S195
P2'	C64	F189, N218	F39, H40, F41, T151, G193, F41

(B) Reactive-site loop of trypsin-binding domain

	TI-II	Subtilisin	Trypsin
P4	A2	I107, G102	W215, G216, S217
P3	C3	G127	G216
P2	T4	D32, T33, H64, L96, G100	H57, L99, S195, S214, W215
P1	R5	L126, G127, G128, A152, G154, N155, G166, A153, S125, T220	S190, G219, C220, S214, C191, Q192, V213, W215
P1'	E6	N155, N218, S221	C42, H57, S195
P2'	C7	N155, N218	F41

(C) Secondary binding interfaces

TI-II	Subtilisin	Chymotrypsin	Trypsin
F12	Y167, K170	V17, N18, G187, V188, S221, T222, S223	
I14		V17, N145, Y146, T219, C220	
C15		R145, Y146	
C27	N155, F189	T151, M192, G193	
I28		L143, Y146, M192	
S32		S217, S218, M192, T219	
G33		Y146, S218	
Y34	A129	Y146, T219, Y146	
Y39		Y146	
S52		T219, S218	
C87	N155, F189		Y151, Q192, G193
C91	S101		T149, Y151
T92			Q192
Y94			S146, S147
F100	S130	S221, T222, S223	
D103	K136, Y171	Y171, A185	
G104		A185, T222	
K105		S223	
F106		S223	
P115			S146, S147, Q221

6.1 References

1. Tong, L., Wengler, G., and Rossmann, M. G. (1993) *J. Mol. Biol.* 230, 228-247
2. Choi, H. K., Lu, G., Lee, S., Wengler, G., and Rossmann, M. G. (1997) *Proteins* 27, 345-359
3. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and Hilgenfeld, R. (2002) *EMBO J* 21, 3213-3224
4. Ziebuhr, J., Snijder, E. J., and Gorbalenya, A. E. (2000) *J. Gen. Virol.* 81, 853-879
5. Oerke, E. C., Dehne, H. W., Schonbeck, F., and Weber, A. (1994) *Crop production and crop protection: Estimated losses in major food and cash crops*, Elsevier, Amsterdam
6. Johnson, R., Narvaez, J., An, G., and Ryan, C. (1989) *Proc Natl Acad Sci U S A* 86, 9871-9875
7. Duan, X., Li, X., Xue, Q., Abo-El-Saad, M., Xu, D., and Wu, R. (1996) *Nat. Biotech.* 14, 494-498
8. Broadway, R. M., and Duffey, S. S. (1986) *J. Insect. Physiol.* 32, 673-680
9. Houseman, J. G., Downe, A. E. R., and Philogene, B. J. R. (1989) *Can. J. Zool.* 67, 864-868
10. Ferreira, C., Capella, A. N., Sitnik, R., and Terra, W. R. (1994) *Comp. Biochem. Physiol.* 107A, 631-640
11. Pautot, V., Holzer, F. M., and Walling, L. L. (1991) *Molec. Plant-Microbe Inter.* 4, 284-292
12. Peng, J. H., and Black, L. L. (1976) *Phytopathology* 66, 958-963

13. Mosolov, V. V., Loginova, M. D., Fedurkina, N. V., and Benken, I. I. (1976) *Plant Sci. Lett.* 7, 77-80

14. Brown, A. E., and Adikaram, N. D. B. (1983) *Phytopathol. Z.* 106, 239-251

15. Botos, I., Meyer, E., Nguyen, M., Swanson, S. M., Koomen, J. M., Russell, D. H., and Meyer, E. F. (2000) *J. Mol. Biol.* 298, 895-901

16. Zhu, Y.-C., and Baker, J. E. (2000) *Arch. Insect Biochem. Physiol.* 43, 173-184

17. Rypniewski, W. R., Ostergaard, P. R., Norregaard-Madsen, M., Dauter, M., and Wilson, K. S. (2001) *Acta Cryst. D* 57, 8-19

18. Murphy, J. M., and Walton, J. D. (1996) *Mol. Plant Microbe Interact.* 9, 290-297

19. Carlile, A. J., Bindschedler, L. V., Bailey, A. M., Bowyer, P., Clarkson, J. M., and Cooper, R. M. (2000) *Mol. Plant Microbe Interact.* 13, 538-550

20. Ng, K. K.-S., Petersen, J. F., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., Dunn, B. M., Martzen, M. R., Peanasky, R. J., and James, M. N. (2000) *Nat Struct Biol* 7, 653-657

21. Citron, M. (2002) *Neurobiology of Aging* 23, 1017-1022

22. Cooper, J. B. (2002) *Current Drug Targets* 3, 155-173