IMPROVING THE INHIBITORY POTENCY OF PAPAYA CYSTATIN, USING SITE-DIRECTED MUTAGENESIS

by

STEFAN GEORGE VAN WYK

Dissertation submitted in partial fulfillment of the requirements for the degree

MAGISTER SCIENTIAE

In the Faculty of Natural and Agricultural Sciences

UNIVERSITY OF PRETORIA

Pretoria

SUPERVISORS: PROF. K.J. KUNERT

DR. B.J. VORSTER

DR. U. SCHLÜTER

April 2011

I, Stefan George van Wyk, declare that the dissertation, which I hereby submit for the degree Magister Scientiae at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: 

Date: 01 April 2011

TABLE OF CONTENTS **Page no**

ABSTRACT

Novel conserved amino acid variations of papaya cystatin (PC) were investigated by amino acid substitutions using oryzacystatin-I (OCI) as a model plant cystatin for comparison. These amino acid residues in the conserved motifs are involved in binding with cysteine proteases, these include the **GG** (Gly-Gly) in the N-terminal region for both OCI and PC, the (**Q**)**Q**VV**A**G (Gln-Val-Val-Ala-Gly) motif for OCI and (**Q**)**A**VV**E**G (Ala-Val-Val-Glu-Gly) motif for PC in the first inhibitory loop, and the P**W** (Pro-Trp) motif for OCI and L**W** (Leu-Trp) motif for PC in the second inhibitory loop. Recombinant OCI and PC mutant proteins were expressed in *Escherichia coli* and were tested for altered inhibitory activity against commercial cysteine proteases (papain and cathepsin L) and extracts from Colorado potato beetle (*Leptinotarsa decemlineata)* larvae, from banana weevil larvae (*Cosmopolites sordidus)* and tobacco leaf extracts (*Nicotiana benthamiana*). In all tests higher amounts of PC had to be used to obtain similar inhibition levels as OCI. Changing the amino acid Q at position 52 to E in OCI in the first inhibitory loop, had lowered the K_i value of the mutant against the commercial proteases. Concurrently the same amino acid string (EQ) in PC had resulted in a significantly decreased K_i value compared to PC wild-type and other mutants. All other OCI mutants were less efficient than the wild-type OCI, whereas all PC first inhibitory loop mutants had improved inhibitory activity against protease activity with the highest improvement against the protease extracts was found for the substitution of E with A at position 55. This study has shown the importance of the three conserved motifs and that it is possible to improve the binding capacity of a plant cystatins to cysteine protease activity by amino acid substitution using site-directed mutagenesis. By mutating individual amino acid residues in the first binding loop of the relatively “weak” papaya cystatin to amino acid residues found in OCI caused a significant improvement in inhibitory potency of PC.

COMPOSITION OF DISSERTATION

Chapter 1 of this dissertation outlines the cystatin superfamily, with specific focus on their characteristics and biological functions in plants. It also outlines the mechanism of interaction between cysteine proteases and cysteine protease inhibitors, as well as work done on altering this interaction mechanism. The aim and objectives of this study are then provided. At the end of this chapter previous work contributing to this study is described, which also includes the site-directed mutagenesis experiments to introduce the desired mutations. **Chapter 2** outlines the materials and methods used in this study. This includes the molecular cloning of the wild-type and mutant gene sequences, the expression thereof in a bacterial expression system and the purification and testing of these proteins against cysteine proteases. The various molecular biology techniques used are described, the cloning technique, the protein expression and purification techniques, the sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) technique and the fluorometric techniques. The simulation of the interaction between the models of the individual cystatins and a model of a cysteine protease are described. **Chapter 3** outlines the results obtained from the protein expression and purification experiments and further reports the observed changes in inhibitory potency. Finally, the simulated interaction data is reported. In **Chapter 4** the results obtained are discussed and results which have contributed to a better understanding of the cysteine protease cystatin interaction mechanism are highlighted. In **Chapter 5** the literature cited in this dissertation is listed.

LIST OF FIGURES **Page no**

LIST OF TABLES **Page no**

ABBREVEATIONS AND SYMBOLS

°C	Degree Celsius
%	Percentage
A	Alanine amino acid residue
N	Asparagine amino acid residue
D	Aspartate amino acid residue
BW	Banana weevil (*Cosmopolites sordidus)*
CPB	Colorado potato beetle (*Leptinotarsa decemlineata)*
C	Cysteine amino acid residue
dH_2O	Distilled water
DMSO	Dimethyl sulfoxide
E-64	*trans*-epoxysuccinyl-$_L$-leucylamido-(4-guanidino) butane
EDTA	Ethylenediaminetetraacetic acid
FU/min	Fluorescence units produced per minute
Q	Glutamine amino acid residue
E	Glutamic amino acid residue
G	Glycine amino acid residue
GSH	Glutathione

GST	Glutathione S-transferase
H	Histidine amino acid residue
I	Isoleucine amino acid residue
IPTG	Isopropyl-ß-D-thiogalactoside
kcal	Kilocalorie
kDa	Kilodalton
LB	Luria-Bertani
LBA	LB agar
L	Leucine amino acid residue
K	Lysine amino acid residue
K_i	Dissociation constant for inhibitor
$K_{i(app)}$	Apparent K_i
K_m	Michaelis constant
M	Molar
mg	Milligrams
µg	Micrograms
mL	Millilitres
µL	Microlitres
mol	Mole

mM	Millimolar
ng	Nanograms
nmol	Nanomole
OCI	Oryzacystatin-I
OD	Optical density
O/N	Overnight
F	Phenylalanine amino acid residue
P	Proline amino acid residue
PC	Papaya cystatin
PCD	Programmed cell death
Pfu	*Pyrococcus furiosus*
Rpm	Revolutions per minute
RT	Room temperature
S	Serine amino acid residue
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
TAE	Tris-acetate EDTA
Taq	*Thermus aquaticus*

T	Threonine
U	Enzyme units
W	Tryptophan amino acid residue
Y	Tyrosine amino acid residue
V	Valine amino acid residue
v/v	Volume per volume
w/v	Weight per volume
x	Any amino acid residue
x*g*	Times the force of gravity
Z-Phe-Arg-MCA	Z-phenylalanine-arginine-7-amido-4-methylcoumarin

1. INTRODUCTION

The importance of cystatins in plant metabolism has been reiterated by several studies confirming their involvement in controlling the activity of cysteine proteases in plants. Cystatin, initially considered to only control the turnover of storage proteins coinciding with seed germination and development, these proteins have been identified in a range of physiological processes were the activity of cysteine proteases needs to be regulated. These include regulating the plant's endogenous proteolytic activity during development, in the vegetative and reproductive organs, to inhibiting exogenous cysteine proteases used by insects, nematodes and microbial pathogens for digestive functions. Although major progress has been made to identify and characterize plant cystatins, knowledge on their *in vivo* regulation and regulatory functioning is still limited. Several studies have also looked at the potential cystatin in biotechnology applications, such as crop improvement and recombinant protein production. Understanding the role of individual amino acids involved in the binding with proteases will aid in the molecular engineering of improved cystatin variants, which are useful for various biotechnological applications (Benchabane *et al.*, 2010).

1.1 Cystatin super-family

The members of the cystatin superfamily are protein inhibitors which form tight, reversible bonds with papain-like cysteine proteases to prevent the enzyme's hydrolytic activity (Turk and Bode, 1991). The cystatin superfamily consists of four different subfamilies (Oliveira *et al.*, 2003), cystatins are subdivided into one of these based on characteristics specific to each family: Type 1 = the stefin family, which has

a molecular size of about 11 kDa, consists of a single domain, lacking sugar secondary groups and does not contain any disulphide bonds (Colella *et al.,* 1989; Turk and Bode, 1991); Type 2 = the cystatins family, which has a molecular size of about 13 kDa, consists of a single interacting domain with glycosylated groups and forms two disulphide bonds to stabilize the protein backbone (Colella *et al.,* 1989; Turk and Bode, 1991); Type 3 = the kininogen family, mainly consists of multiple stefin-like domains in tandem (similar to multi-domain plant cystatins) (Oliveira *et al.*, 2003). In comparison, plant cystatins or phytocystatins are independently grouped as they contain characteristics of the different animal cystatin families, but cannot be grouped into one specific group (Kondo *et al.,* 1991). For example, plant cystatins contain type 1 stefin characteristics, i.e. they do not contain any glycan moiety on the protein's structures and do not contain any disulfide bonds, but also have primary sequence similarities to type 2 cystatins (Kondo *et al*., 1991; Margis *et al.,* 1998). Similar to stefin-like cystatins, the most plant cystatins do not contain any cysteine amino acid residues, with the exception of the papaya cystatin (Song *et al.,* 1995). Phylogenetic analysis of animal and plant cystatins, groups plant cystatins into a separate evolutionary clade from animal cystatins, based on the leucine (L)-alanine (A)-arginine (R)-phenylalanine (F)-alanine (A)-valine (V)-aspartate (D)-glutamine (E)-histidine (H)-asparagine (N) motif (LARFAVDEHN), unique to the α-helix of plant cystatins (Margis *et al.*, 1998). Another differentiating characteristic of plant cystatins includes a unique and complex gene structure organization, e.g. defined size and positions of introns, and carboxy-terminal extension, which are not found for animal cystatin genes (Margis *et al.*, 1998). This possibly indicates an early divergence between plant and animal cystatins in their independent, parallel evolutionary course from a common eukaryotic cystatin ancestor (Kondo *et al.*, 1991;

Margis *et al.,* 1998). Additional evolutionary processes, such as gene duplication, alternative splicing and adaptive evolution, could also have caused several changes and further divergence between animal and plant cystatins (Benchabane *et al.*, 2010).

The gene sequences of more than 200 plant cystatins, and even more from the whole cystatin superfamily, are available on the NCBI database (Benchabane *et al.*, 2010). Structural comparisons have shown that plant cystatins range between 11 to 16 kDa in size (but can also be 23 kDa and even up to 87 kDa has been found for multi-domain plant cystatins) (Oliveira *et al*, 2003; Mosolov and Valueva, 2005). Structural similarities of plant cystatins to some animal cystatins (e.g. human stefin A, stefin B and chicken egg white cystatin) confirms an evolutionary relationship among the cystatin superfamily, these structures include the one or two glycine amino acid residues located on the N-terminal region, the Q (glutamine)-x-V (valine)-x-G (glycine) motif (QxVxG, where x represents any amino acid) in the first inhibitory loop and the P (proline)- W (tryptophan) (PW) amino acid residues in the second inhibitory loop (Kondo *et al.,* 1991; Margis *et al.,* 1998). As cystatins perform crucial roles in both plants and animals, regulating cysteine protease activity, a strong purifying selection pressure might have caused these domains to be conserved in both animal and plant cystatins (Kondo *et al.,* 1991; Margis *et al.,* 1998). The N-terminal region is suggested to be essential for animal cystatins, but dispensable in some plant cystatins (Abe *et al.,* 1988; Zhao *et al.,* 1996). Figure 1.1 shows the graphical representation of the conserved amino acid sequences typically found for each cystatin subfamily, the conserved motifs and intron positions are numbered and disulfide bonds are underlined. The position of the unique LARFAVDEHN motif in the α-helix of plant cystatins in also indicated.

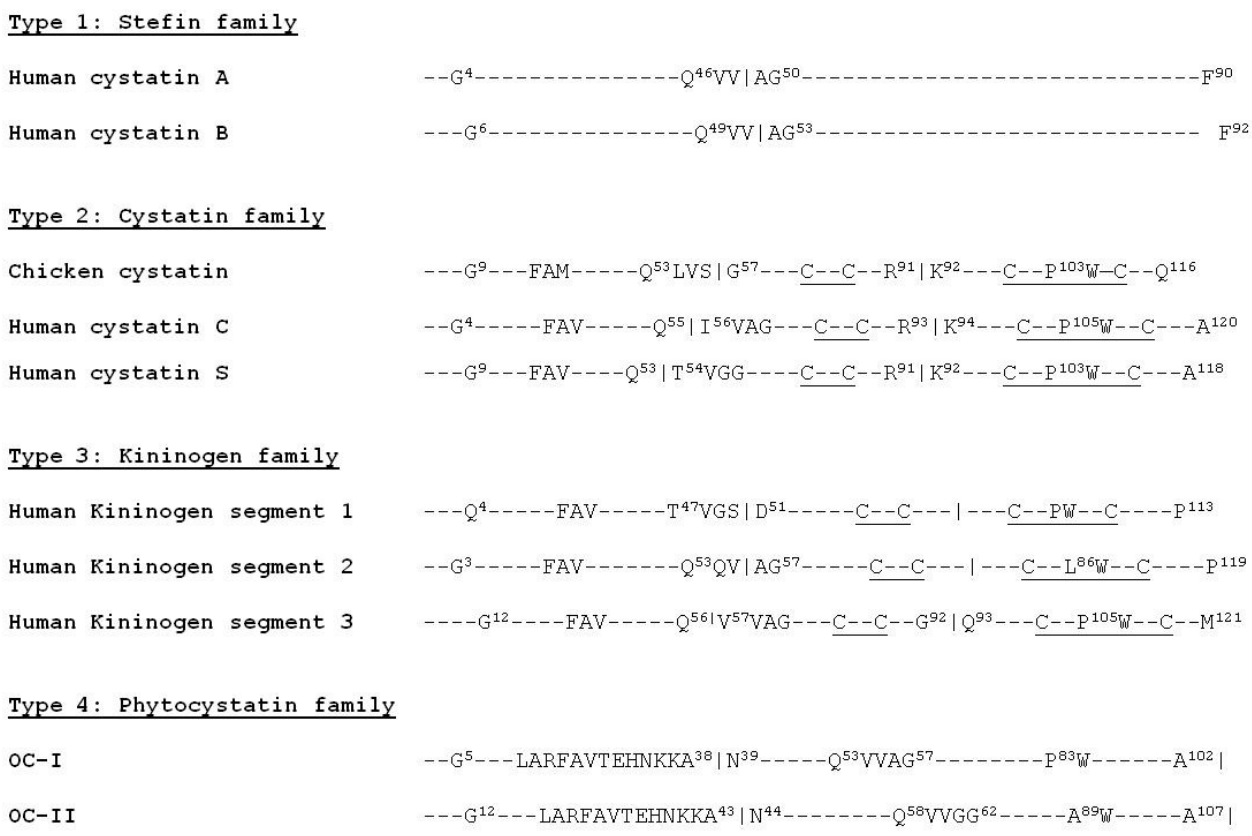

Figure 1.1: Amino acid sequence alignment of different members of the respective families in the cystatin superfamily. The amino acid residues conserved in the different motifs and junction sites are indicted and numbered, the | symbol represents intron positions and the underlined cysteine amino acid residues (C to C), represent the formation of disulfide bonds (figure adapted for illustrative purposes from Kondo *et al.,* 1991; Turk and Bode 1991; Križaj *et al.,* 1992; Margis *et al.,* 1998; Oliveira *et al*, 2003).

1.2 Cystatin structure

The gene organization of plant cystatins typically consists of 3 exons and 2 introns, and the TATA and CAAT sequence domains in the 5'-region of the transcription initiation point (Abe *et al*., 1996; Kondo *et al*., 1991). The first exon contains the sequence data for the first

43 amino acid residues of the cystatin (Kondo *et al*., 1991). The first intron (approximate length 408bp) separates exon 1 and 2, in animal cystatins this intron is mainly located in the coding region of the gene and disrupts the conserved QxVxG motif, while in plant cystatins the first intron is located upstream from the coding region for the QxVxG motif (Kondo *et al*., 1991). The second exon contains the sequence data for the rest of the cystatin protein, the stop codon (TAA) and part of the 3'- non-coding region (Kondo *et al*., 1991; Abe *et al*., 1996). The second intron (approximate length 132 bp), separates exons 2 and 3, the third exon only contains a 3'- non-coding region (the extended C-terminal structure of plant cystatins), the intron is located in the non-coding region of the gene, which differs from animal cystatins that also contain the second intron in the coding region (Kondo *et al*., 1991; Abe *et al*., 1996). Plant cystatins, such as OCI, soya cystatin or tomato cystatin (*Sl*CYS9), have a signal sequence at the N-terminal to direct the cystatin to the targeted area of effect (Doi-Kawano *et al*., 1998; Girard *et al*., 2006). Furthermore, the majority of plant cystatins are translated as either a prepro-protein or a pre-protein, which requires subsequent modifications to become active (Dubin, 2005; Fan and Wu 2005). Some unusual structures found in plant cystatins, include the occurrence of multi-domain cystatins, e.g. potato or tomato multi-cystatins which consist of eight, stefin-like cystatin domains that differ in sequence and can even be evolutionary distinct (Brown and Dziegielewska, 1997; Girard *et al*., 2006; Nissen *et al.,* 2009).

Plant cystatins typically consist of a central α-helix (containing the conserved LARFAVDEHN sequence, unique to plant cystatins), 5 beta-sheets in anti-parallel (Figure 1.2) and a conserved mechanism of inhibition found for most of cystatins, whereby the two inhibitory loops determine the active site inhibition and interference (Turk and Bode, 1991; Reis and Margis, 2001; Benchabane *et al.*, 2010).

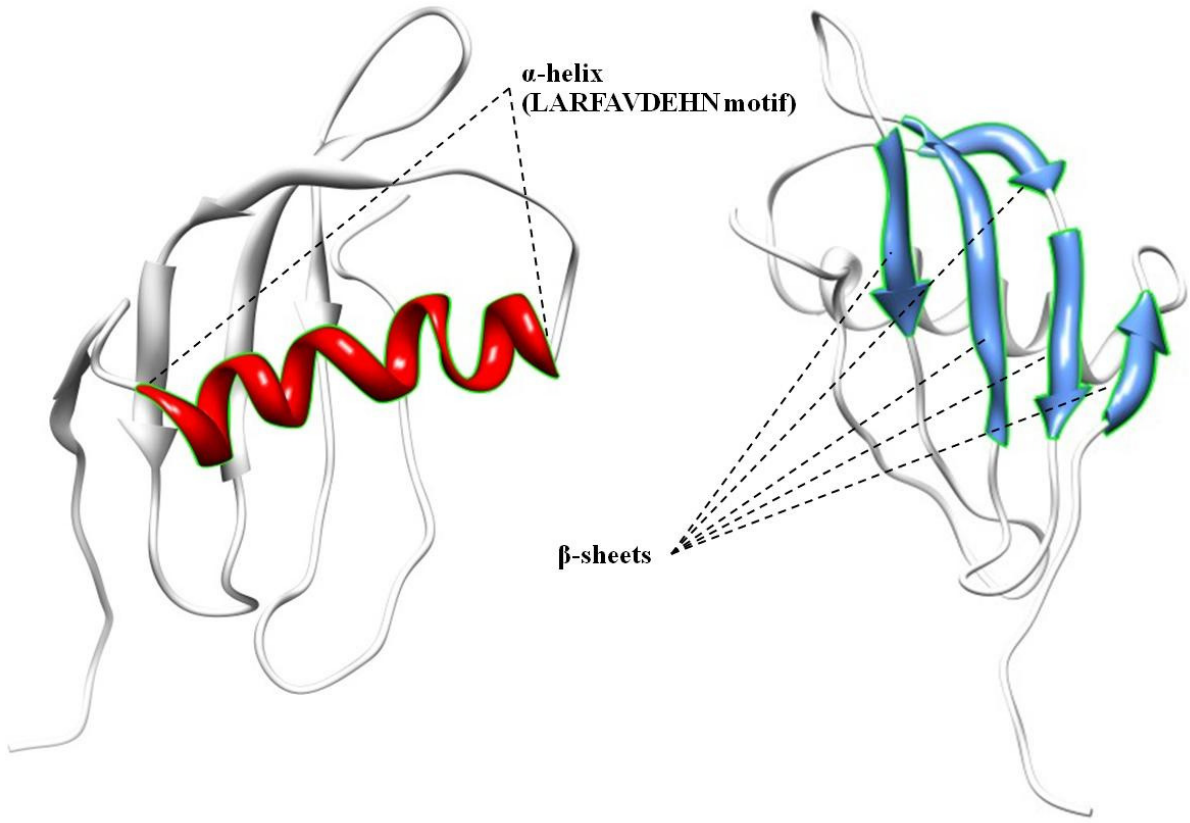

Figure 1.2: The molecular model of oryzacystatin-I (OCI) was used to illustrate the position of the α-helix and the 5 β-sheets, characteristic of plant cystatins (UCSF Chimera; Pettersen *et al.,* 2004).

During the formation of the binding interaction with a protease enzyme, the "tripartite wedge" is formed by the tertiary structures of the cystatin molecule (Figure 1.3) and has a high affinity fit with the groove of the target enzyme's active site (Stubbs *et al*., 1990; Benchabane *et al.*, 2010). This wedge is formed by the highly conserved QxVxG in the first inhibitory loop, and the P/AW motif of the second inhibitory loop (Arai *et al*., 1991).

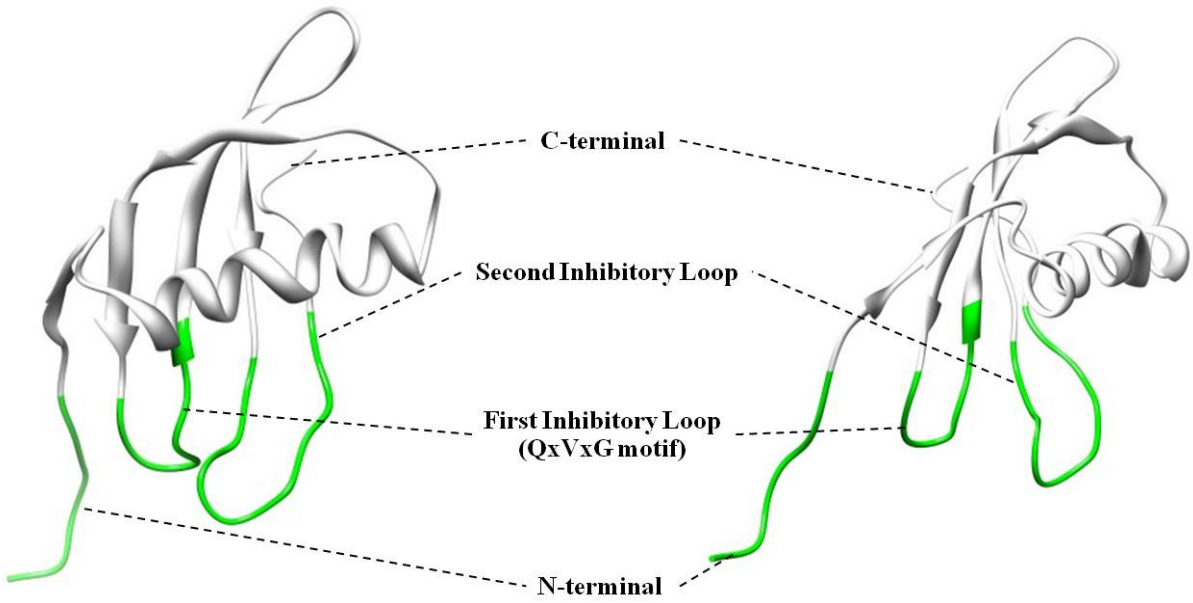

Figure 1.3: The side and front views of the OCI model showing the conserved motifs typically found in the structure of plant cystatins, the indicated regions contain amino acid residues which are involved in the interaction with the target enzyme: the N-terminal, containing the GG at amino acid position 10 and 11 (according to OCI numbering), the first inhibitory loop containing the QxVxG in amino acid position 53-57 and the second inhibitory loop containing the PW in amino acid position 84 and 85 (UCSF Chimera; Pettersen *et al.,* 2004).

The amide side chain of the first Q amino acid residue has a neutral side chain charge and is seemingly preferentially accommodated to the negatively charged side chain of E amino acid residue (Koiwa *et al*., 2001). Amino acids at this position is thought to form polar interactions with the protease molecule, the VVA amino acid residues are thought to be important for binding (Stubbs *et al*., 1990), with the G amino acid residue causing a local conformation around the first inhibitory loop forming the hairpin structure (Bode *et al*., 1988). The PW motif located in the second inhibitory loop is conserved in the third domain of kininogens and in the cystatin subfamily (type 2) (Turk and Bode, 1991). The P amino acid residue is found

in most plant cystatins, but can also be other amino acids (either an A/ Q/ L/ S, Table 1.2), and the W amino acid residue is found in almost all plant cystatins (Table 1.2). The N-terminal, which stabilizes the cysteine protease-cysteine protease inhibitor complex (Figure 1.4), forms hydrophobic interactions with the enzymes' protein surface, adjacent to the active site (Turk and Bode, 1991; Benchabane *et al.*, 2010). This forms a tight, reversible interaction complex of cysteine protease and cysteine protease inhibitor, and remains stable for a long period of time and prevents any further proteolysis of the protein substrate (Turk and Bode, 1991).

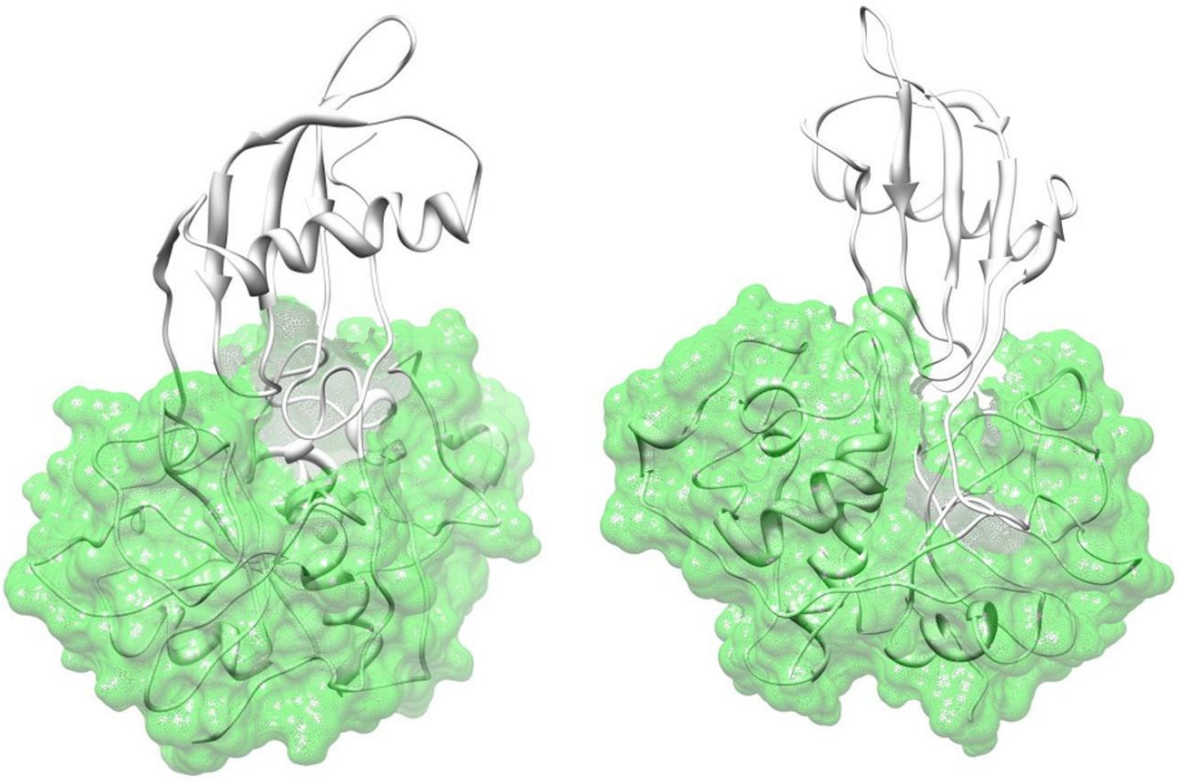

Figure 1.4: The side and rear views of the OCI model docking with the model cysteine protease, papain. The inhibitory loops can be seen entering the active site cleft of the enzyme and the N-terminal of the cystatin associates with the enzyme's protein surface adjacent to the cleft (UCSF Chimera; Pettersen *et al.,* 2004).

1.3 Plant cystatin function

The uncontrolled activity of proteases can have an extraordinarily destructive influence on normal biological functions and the coordinated activity of cysteine proteases is essential. Cystatins are natural inhibitors and are found in almost every form of life (Bode *et al.,* 1988; Turk and Bode, 1991; Oliveira *et al*., 2003). Four major classes of proteases exist, namely cysteine, serine, aspartyl and metallo-proteases, a corresponding class of protease inhibitors exists for each (Oliveira *et al*., 2003; Fan and Wan, 2005). Each protease inhibitor class has distinguishing characteristics, such as an optimum pH range and specific amino acid residues in their structures, which are required to efficiently bind to target proteases and prevent their activity (Fan and Wan, 2005). Plant cystatins are involved in various defence and developmental processes, the roles of cystatins have been inferred based on the expression profiles of cystatins during these different processes.

1.3.1 *Expression during developmental processes*

Cysteine proteases are one of the main enzyme classes active in plants and are responsible for the mobilization of storage proteins during germination (Grudkowska and Zagdańska, 2004). The processes of storage protein deposition, cystatin expression and protease down-regulation have been investigated in expression studies, e.g. during seed development, cystatin mRNA transcripts start to accumulate to eliminate protease activity and allow storage protein deposition and protecting these storage proteins until germination (Martinez *et al.* 2009; Benchabane *et al.*, 2010). Cysteine proteases have been found to be up-regulated in protein turnover during senescence, while cysteine and serine protease inhibitor genes are down-regulated (Martinez *et al.* 2009; Benchabane *et al.*, 2010). Cysteine proteases are also

involved in the different processes of programmed cell death (PCD) (e.g. process of tissue-differentiation, such as xylogenesis or different stages of senescence), the expression of cystatins was found to coincide with the activation of this signal transduction cascades, e.g. regulating the protease enzymes involved in PCD during the hypersensitive response (Solomon *et al*., 1999; Belenghi *et al*., 2003).

1.3.2 *Expression under abiotic and biotic stress*

Cysteine proteases have also been implicated in the response to abiotic (such as drought, temperature shock and salinity) and biotic stress (such as mechanical wounding, insect herbivory, fungal elicitors, abscisic acid and jasmonic acid) (Benchabane *et al.*, 2010). Cystatins used for housekeeping purposes and physiological regulation are thought to have broad range of expression patterns, whereas cystatins involved in developmental cues and in stress responses are thought to have a restricted expression pattern (Massonneau *et al.*, 2005). Five different cystatins from maize (*Zea mays*) were found to be down-regulated during drought conditions, while two maize cystatins were found to be induced by cold stress (Massonneau *et al.*, 2005).

The expression of the soybean cystatins N1 and R1 were found to be induced, both locally and systemically, by methyl-jasmonate or mechanical wounding, and it is suggested that these cystatins are involved in plant defence responses (e.g. hypersensitive response against pathogens, insects and nematodes) (Botella *et al*., 1996). Fungicidal characteristics were observed for barley cystatin proteins which were produced recombinantly, the proteins were capable of inhibiting the growth of phytopathogenic fungi *Botrytis cinerea* and *Fusarium oxysporum in vitro* (Martinez *et al.* 2005).

1.3.3 *Expression of exogenous cystatins in biotechnological applications*

Plant cystatins are being investigated for several biotechnological applications, specifically where the regulation of cysteine protease activity is desired. One such application is in food processing, e.g. proteolytic enzymes are used to modify the proteins to be used in a food product, for added value or improved properties. The activity of these protease enzymes is regulated by protease inhibitors, such as cystatins, to obtain the required degree of hydrolysis (hydrolyzed peptide bonds) (García-Carreño *et al*., 2000).

Plant cystatins have been successfully used to engineer crop plants for resistance against different types of pests, such as insects, nematodes and pathogens (Benchabane *et al.*, 2010; Schlüter *et al*., 2010). Recombinant cystatins have been tested in *in vitro* feeding trials to illustrate the deleterious effects of cystatins on insect larvae, specifically tested were the oryzacystatin-I (OCI) and the papaya cystatin (PC) (Kiggundu *et al*., 2010). The cowpea cystatin, found in cowpea (*Vigna unguiculata*) seeds, was found to inhibit the digestive proteases of targeted insect pests (the bean weevil, *Acanthoscelides obtectus,* and the Mexican bean weevil, *Zabrotes subfasciatus*) in *in vitro* assays, the protein was found to be effective evenat levels as low as 0.025% of total seed protein content. Molecular modelling studies of this cowpea cystatin revealed that five amino acid residues located on the N-terminal are involved in the interaction with the proteases, which form hydrophobic interactions to stabilize the enzyme-inhibitor complex (Aguiar *et al*., 2006). The functional modification of the eighth cystatin domain of the multi-cystatin from *Solanum lycopersicum* (*Sl*CYS8) had shown some variants to have enhanced ability to inhibit target digestive proteases from *Leptinotarsa decemlineata* (Colorado potato beetle) while at the same time

exhibiting lowered activity against non-targeted cysteine proteases (Goulet *et al*., 2008). The interaction between inhibitors and target proteases is dose-dependent, therefore enhancing the inhibitory potency of a cystatin would assist to decrease the dosage of cystatin required to successfully provide anti-nutritive/ pesticidal effects (Goulet *et al*., 2008).

The ectopic expression of OCI in transgenic tobacco resulted in enhanced fitness and tolerance to chilling stress along with an altered physiology, such as retarded stem elongation and leaf expansion under low light intensities (Van der Vyver *et al*., 2003; Demirevska *et al*., 2009). Expression of cystatins in the cytosol of a transgenic non-host plant have been found to alter the phenotype with potentially beneficial traits, such as improved tolerance against abiotic stress (Munger *et al*., 2009). The ectopic expression of maize cystatin-II in potato leaves led to the constitutive expression of PR-proteins and other protease inhibitor proteins, normally inducible during plant defence or stress response (Munger *et al*., 2009). Constitutive over expression of the *At*CYS1 gene in *Arabidopsis* led to the activation on signal transduction cascades. This was found to coincide with the regulating the protease enzymes involved in PCD during the hypersensitive response in the leaves and had a repressing effect of the overall degree of the response (Belenghi *et al*., 2003).

A further recent biotechnological application of cystatin expression is the protection of recombinant proteins expressed in transgenic plants. When a protease inhibitor, such as the tomato cathepsin D inhibitor or bovine aprotinin, was co-expressed along with the protein of interest, higher amounts of the protein could be harvested compared to the control plants (Rivard *et al*., 2006; Benchabane *et al*., 2009; Goulet *et al*., 2010).

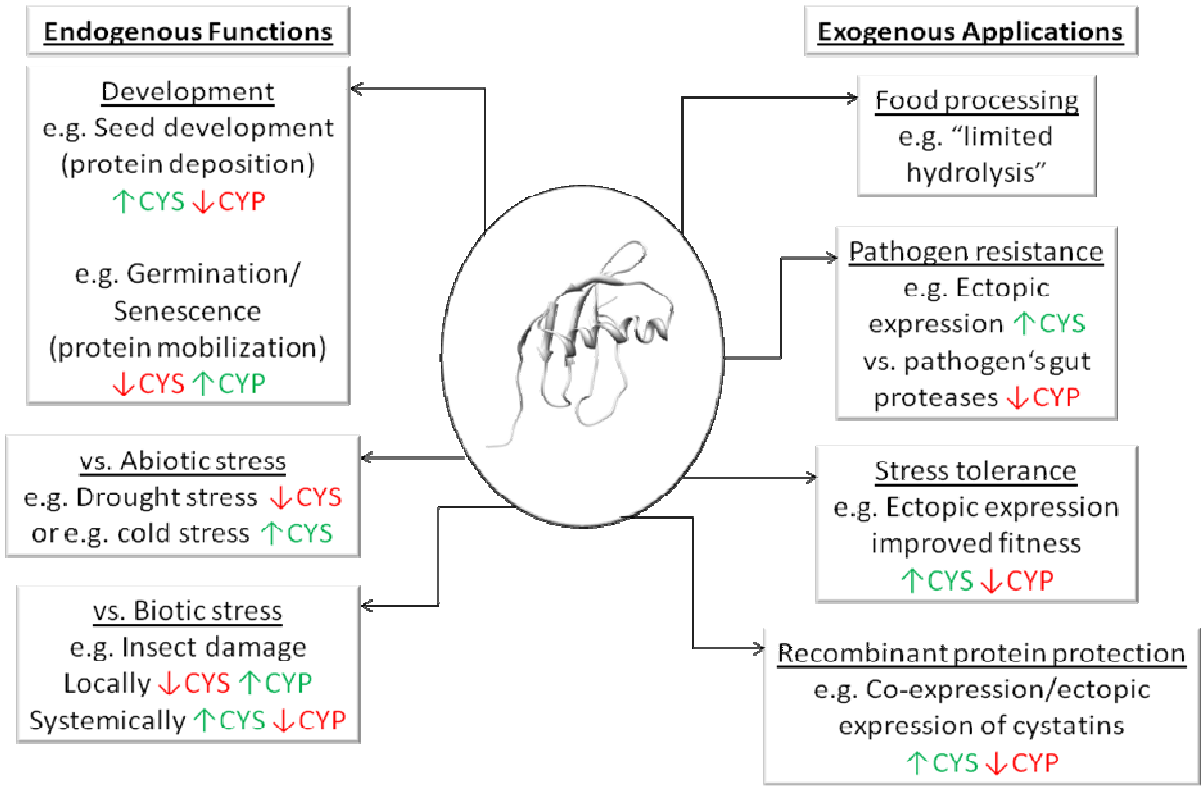

Figure 1.5: Various endogenous functions that cystatins are involved in and the exogenous applications that cystatins have been used for. ↑ indicates an increase in transcription and ↓ indicates a decrease in transcription, while CYS represents cystatins, CYP represents cysteine proteases.

1.4 Cystatin mutagenesis

Several research groups have attempted to improve the potency of cystatins against cysteine proteases by substituting amino acids in the cystatin sequence. However, there are contrasting ideas about the role of certain conserved amino acids in a specific position of a cystatin and their role in determining inhibitory potency. Structure/function models have been used to understand the role of specific amino acids or strings of amino acids in inhibitor potency against specific proteases. Many of these studies have been based on a structural model using the human stefin B-papain complex (Stubbs *et al.*, 1990), which has been instrumental in

identifying particular target sites for the improvement of cystatins against cysteine proteases. Molecular phage displays combined with random mutagenesis have also been used in protease inhibitor sequence regions to identify biologically active cystatin variants. Site-directed mutagenesis of selected amino acids has been applied as a technique to investigate the structure/function relationships in combination with computer algorithms to predict stronger binding interactions and has indeed resulted in more potent protease inhibitors to be engineered (Laboissiere *et al.*, 2002; Ceci *et al.,* 2003; Stoop and Craik, 2003). Several research groups have previously used these two techniques to investigate conserved amino acid residues located in the N-terminal, the first inhibitory loop (the QxVxG motif in the first inhibitory loop being the most conserved of these motives), as well as motives in the second inhibitory loop.

1.4.1 *Mutations in the N-terminal region*

Human cystatin C has been one of the first cystatins to be used as target for site-directed mutagenesis, to substitute particular amino acid residues in this cystatin. The conserved amino acid residue G (glycine) amino acid residue in position 11 was investigated by Hall *et al.* (1993) by replacing it with either a positively charged R (arginine), a negatively charged E (glutamic acid), a bulky hydrophobic tryptophan, or by an amino acid containing small side-chains, such as S (serine) or A (alanine). They found that G, which lacks a side-chain, at amino acid position 11 was important to allow the local confirmation in the N-terminal region, allowing for both high-affinity binding and efficient inhibition. Equilibrium constants for dissociation (K_i) of complexes with papain and human cathepsin B, had showed that human cystatin C variants, with serine (S) or alanine (A) in position 11, had K_i values about 20-fold higher than those of the wild-type cystatin C and variants with W, R or E in this

position had K_i values with a factor of about 2000-fold higher, implying less efficient inhibition of the target proteases (Hall *et al*., 1993). In another study, the importance of the N-terminal region in cystatin C was investigated by a method of random-centroid optimization. This method was used to introduce double-site mutations and had shown to improve the inhibitory activity of cystatin C, without any prior knowledge of the cystatin's molecular structure (Ogawa *et al*., 2002). The substitution mutations of G in position 12 to W, and of H (histidine) in position 86 to V (valine), had increased inhibitory activity 5-fold over wild-type cystatin C. Improved papain inhibition was also obtained by substitution of P (proline) in position 13 with F (phenylalanine), while mutations decreasing beta-sheet content resulted in reduced polymerization of cystatin C and had increased papain inhibition. Furthermore, when three N-terminal amino acid residues preceding the conserved G of cystatin A were replaced with a 10-amino acid residue long cystatin C segment, inhibitor affinity to cathepsin B had increased approximately 15-fold (Pavlova and Björk, 2003). This was predominantly due to a higher association rate constant. Furthermore, substituting G in position 75 with either W or H in the second inhibitory loop of cystatin A, a 10-fold higher papain affinity was observed due to both a higher association rate as well as a lower dissociation rate constant. By changing the individual residues in the N-terminal region of cystatin C, L in position 9 to W, or V in position 10 to either W, F or R as well as changing W in position 106 to G in the C-terminal region, the binding ability to cathepsins B and H was reduced (Mason *et al.*, 1998). This indicates that both cathepsins are repulsed by large aromatic residues in their S_2 and S_3 pockets, in contrast to cathepsin L, which preferentially accommodates larger aromatic residues in the S_2 pocket. This hypothetical binding pocket is formed by amino acids in the enzyme's active site, which interact with each other and contribute to form hydrophobic interactions, hydrogen- and van der Waals bonds with the inhibitor (or substrate). The inhibitor's overall shape and hydrophobicity that would be

tolerated in this binding pocket is dependent on these amino acids present in the enzyme's active site and their contribution towards binding. Mason *et al.* (1998) further found that introducing a charged residue into the S_2 pocket resulted in an inhibitor that showed selectivity towards cathepsin L and S compared with cathepsin B and H. Other mutants were shown to be capable of distinguishing cathepsin H from cathepsin B, or cathepsin S from cathepsin L, dependent on the amino acid residues at specific tertiary positions.

Several research groups have investigated the importance of the N-terminal region of plant cystatins, but despite several studies, the function of the N-terminal region is still rather unclear. In an early investigation, Abe *et al.* (1995) found distinct dissimilarity in the amino acid sequence of corn cystatin-I (CCI) and corn cystatin-II (CCII) around the N-terminal region, with CCII being a stronger inhibitor of cathepsin L compared to CCI. When OCI was studied, it was found that the first 21 residues of the N-terminal region are seemingly not essential for papain-inhibitory activity (Abe *et al.* 1988; Arai *et al.*, 1991; Urwin *et al.*, 1995a). In contrast, OCI truncated of 38 N-terminal residues was almost completely inactive (Abe *et al.*, 1988). Further studies have also shown the importance of individual amino acid residues for cystatin potency, when the conserved G in positions 5, 6, 10 or 11 in the N-terminal region was studied by deletion and mutation and observed inhibitory characteristics were compared to wild-type OCI. Only substitutions of the amino acid at position 10, changing G to either R, C (cysteine), E, Q, H, L (leucine), K (lysine), F, P, S, or Y (tyrosine) significantly changed the inhibitory potency of OCI, but not at any of the other amino acid sites (Urwin *et al.*, 1995a). This demonstrated the functional importance of the highly conserved G residue at position 10, for effective inhibition of papain. In another study, using N-terminal deletion mutants in the N-terminal of the sunflower cystatin, the amino acids isoleucine (I) and P at position 1 and 2 as well as the N-terminal G residues at position 3

and/or 4, were shown to play a functionally important role in papain inhibition (Doi-Kawano *et al.,* 1998). The N-terminal regions of plant cystatins have also been associated with anti-feedant and antifungal activity. Modification of either C, E/ D (aspartate), or R residues in the N-terminal sequence of a pearl millet cystatin resulted in the loss of the antifungal activity (Joshi *et al.,* 1999).

Proteins that are under direct selection pressure, such as defence proteins, have been shown to undergo adaptive evolution, whereby mutations that confer a selective advantage are maintained under selection. The non-synonymous substitutions contribute to the diversification and maintenance of cystatin variability at these interacting amino acids residues of cystatins that have been shown to undergo positive selection in response to the target proteases from the pathogen (Kiggundu *et al.* 2006). Single mutations at these positively selected amino acid sites have been extensively investigated in several studies to improve the anti-nutritional activity of plant cystatins, a possible link was established between hyper-variable amino acid sites in the plant cystatin amino acid sequence and inhibitory potency against cysteine proteases of the papain-like family (C1) (Kiggundu *et al.* 2006). In this study, a maximum likelihood approach was applied to assess plant cystatins for positive selection during evolution by comparing the rates of synonymous to non-synonymous mutations. This comparison had ultimately provided an indication to which amino acid sites have potential for improving the inhibitory activity of cystatins. Hyper-variable sites were located on each side of the conserved G residues in the N-terminal region, also within the first and second inhibitory loops and areas surrounding the conserved LARFAVDEHN motif in the α-helix of plant cystatins. Based on the study of Goulet *et al.* (2008), 29 mutants of the tomato multi-cystatin 9 subunit were produced with single substitutions at three positions shown to be under positive selection. Substituting the original

P at position 2, T (threonine) at position 6, or E at position 31, had resulted in mutants showing a range of increased or lowered inhibitory potency against cysteine proteases. In particular the P at position 2, the amino acid residue adjacent to the conserved GG motif in the N-terminal region of the cystatin, showed an improved inhibitory potency against the cysteine proteases of the Colorado potato beetle extract. Furthermore, the substitutions of P to F, I, L or Y resulted either in lowered or unchanged inhibitory potency against target proteases, potato leaf cysteine protease(s) or protease I, a cysteine protease utilized by the predatory insect *Perillus bioculatus*.

1.4.2 *Mutations in the first inhibitory loop*

The motif QxVxG is found in the first inhibitory loop, this motif is conserved among most members of the cystatin super-family and substitutions of amino acid residues in this motif can greatly change binding affinity to cysteine proteases. When the amino acid residues L and S in the QLVSG motif of chicken cystatin were changed to V and A, respectively, the binding affinity with the cysteine protease actinidin (papain-like enzyme) significantly increased more than 10-fold, the other cysteine proteases tested did not show the same degree of change (Auerswald *et al.*, 1996). Several researchers have investigated this conserved motif in plant cystatins, in the first study by Abe *et al.* (1988), it was found that an N-truncated OCI retaining the QVVAG motif inhibited papain as efficiently as a non-truncated OCI, demonstrating the importance of this motif for activity. Arai *et al.* (1991) provided more detailed information of the role of particular amino acid residues in the QVVAG motif of OCI (Figure 1.6). These results clearly indicate that the QVVAG sequence of the cystatin molecule is the primary region of interaction with the cysteine protease and is thus responsible for the inhibitory activity. In general, substituting Q at position 53 with L caused

significantly lowered inhibition of papain and a K_i value approximately 150-times higher than wild-type OCI. Furthermore, when substituting the same Q with P resulted in a completely inactive cystatin, this completely inactive mutant was also unable to bind to a papain column, indicating that the affinity site of OCI is also its reactive site. In a cystatin isolated from papaya, the papaya cystatin (PC), the Q of the QVVAG motif is substituted by A (Song *et al*., 1995, Figure 1.7). In this study, the function of this natural A substitution in PC was primarily investigated. Furthermore, substituting the central V residue only resulted in a moderate effect on activity, substituting V with G resulted in a protein which was as active as the wild-type OCI, while changing the V to D resulted in a 40-times higher *Ki* value than the wild-type OCI (Arai *et al.*, 1991, Figure 1.6). Nikawa *et al.,* (1989) claimed that the QVVAG motif was not essential for inhibitory activity, but the study had only looked at two mutants of human cystatin A. Changing the glutamine amino acid at position 46 to lysine (KVVAG) had lowered inhibitory activity against cathepsin B (0.6-times) and H (4.0-times), but improved against papain and cathepsin L (0.4-times) and changing the valine amino acid at position 48 to threonine (QVTAG) had performed close to that of the wild-type against the different enzymes assayed papain, cathepsin B and H, but had a 3.5-times reduction against cathepsin L. In another study, Koiwa *et al.,* (2001) confirmed the importance of the first inhibitory loop in the structure of soyacystatin for inhibition of the cysteine protease papain. By using a combination of phage display libraries and random mutations, targeting the QVVAG motif of the first hairpin loop, all functional soyacystatin variants that were studied had this motif conserved. This indicates the functional importance of this motif in the activity of soyacystatin.

Table 1.1: List of cystatins used in other studies, investigating individual amino acids and the observed effect of changing each of these.

#	Cystatin	Region targeted	Amino acid change	Observed effect (vs. wild-type)	Degree of change	Cysteine protease used	Reference
1	Cystatin C	N-terminal	G11A/S	Reduced interaction stability	K_i > 20-fold higher	Papain and cathepsin B	Hall *et al., (*1993)
			G11E/R/W		K_i > 2000-fold higher		
2	Cystatin C	N-terminal and second inhibitory loop	G12W or H86V	Enhanced interaction stability	K_i < 5-fold higher	Papain	Ogawa *et al.*, (2002)
			P13F		≥ 1.56 activity vs. Wt		
3	Cystatin A	N-terminal	N(1-10)CC-cystatin A	Enhanced interaction stability	15-fold higher affinity	Cathepsin-B	Pavlova and Björk, (2003)
		Second inhibitory loop	G75W/H		10-fold higher affinity	Papain	
4	Cystatin C	N-terminal	L9W	Reduced interaction stability	- Cathepsin L < S	Cathepsins B, H, L and S	Mason *et al.*, (1998)
			V10W/F/R		> Cathepsin B > H		
		C-terminal	W106G		> Cathepsin L > S		
5	Oryzacystatin-I	N-terminal	Δ(1-21)	As effective as Wt	-	Papain	Abe et al. (1988)
			Δ(1-38)	Completely inactive	-		
6	Oryzacystatin-I	N-terminal	G5/6/10Δ	No significant alteration	-	Papain	Urwin et al., (1995a)
			G10R/C/Q/E/H/L/K /F/P/S/Y	Enhanced interaction stability	≥ 1.5 activity vs. Wt		
7	Sunflower cystatin	N-terminal	Δ(1-2)	Reduced inhibitory activity	n/a	Papain	Doi-Kawano *et al.,* (1998)
			Δ(1-4)/ Δ(1-13)	Loss of activity	-		

#	Cystatin	Region targeted	Amino acid change	Observed effect (vs. wild-type)	Degree of change	Cysteine protease used	Reference
8	Pearl millet cystatin	N-terminal	C/E/D/R	Loss of antifungal activity	-	Papain	Joshi *et al.,* (1999)
			R/H	Increased inhibitory activity	≤ 2-fold		
9	Tomato multi-cystatin	N-terminal	P2F/I/L/Y	Enhanced inhibitory activity	> 2-3-fold	Colorado potato beetle extract	Goulet *et al.* (2008)
			P2F/I/L/Y	Lowered/unchanged potency vs non target		Potato leaf cysteine protease and protease I from Perillus bioculatus	
			T6E/I/P/S	Reduced interaction stability	< 2-fold	Colorado potato beetle extract	
			E31A/F/G/ K/T	Enhanced interaction stability	≥ 1.5-fold	Colorado potato beetle extract	
10	Chicken cystatin	First inhibitory loop	L54V, S56A	Enhanced inhibitory activity	> 10-fold	Actinidin	Auerswald *et al.*, (1996)
11	OCI	First inhibitory loop	Q53L	Reduced inhibitory activity	K_i > 150-fold higher	Papain column	Arai *et al.*, (1991)
			Q53P	Completely inactive	-		
12	OCI	First inhibitory loop	V53G	As effective as Wt	-	Papain	Arai *et al.*, (1991)
			V53D	Reduced inhibitory activity	K_i > 40-fold higher		
13	OCI	C-terminal	Δ(91-102)	As effective as Wt	-	Papain	Arai et al., (1991)
14	OCI	C-terminal	Δ(67-102)	Reduced interaction stability	Immeasurable	Papain	Arai et al., (1991)
15	Soyacystatin	C-terminal	W78L	Enhanced inhibitory activity	K_i > 2-fold reduction	Papain	Koiwa *et al.,* (2001)

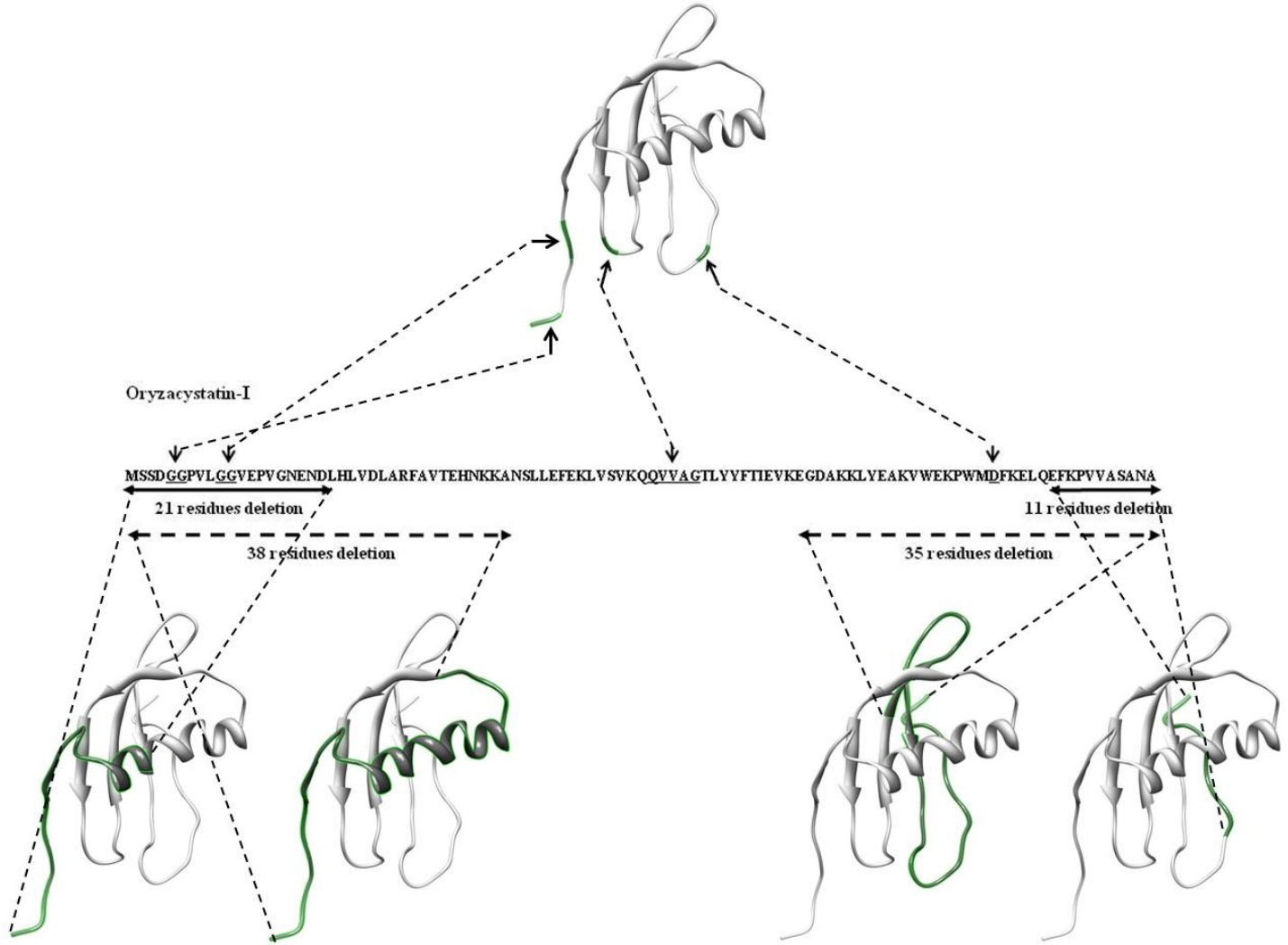

Figure 1.6: The amino acid sequence of OCI, used as model plant cystatin, is shown. The arrows indicate the position of the single, or strings of, amino acid residues which have been previously investigated in other studies using OCI (UCSF Chimera; Pettersen *et al.,* 2004).

1.4.3 *Mutations in the second inhibitory loop*

Research to date has also shown that the C-terminal region of cystatins plays an important role in inhibitory activity and this region can affect the target specificity of cystatins (Auerswald *et al.,* 1996). C-terminal truncated oryzacystatin-I, lacking the 11 C-terminal residues still exhibited potent papain-inhibitory activity (Abe *et al.*, 1988; Arai *et al.*, 1991; Figure 1.6). The possible role of the C-terminal region was also demonstrated by investigating biochemical interactions between digestive proteases of the Coleopteran pest, the black vine weevil (*Otiorynchus sulcatus*) and two plant cysteine protease inhibitors, oryzacystatin-I and oryzacystatin-II, when a C-terminal truncated form of oryzacystatin-II was detected (Michaud *et al.*, 1995). This truncated form was as active against papain and human cathepsin H as the non-truncated oryzacystatin-II. However, when the C-terminal of oryzacystatin-I was truncated by 35 residues, the truncated cystatin showed a considerably reduced inhibitory activity (Abe *et al.*, 1988; Arai *et al.*, 1991; Figure 1.6).

By investigating the second inhibitory loop in greater detail, a variant of oryzacystatin-I with improved activity against nematodes was produced by deleting the aspartate (D) amino acid residue at position 86 of the wild-type's sequence (Urwin *et al.*, 1995b; Figure 1.6). Koiwa *et al.,* (2001) had isolated several soyacystatin variants with improved inhibitory potency, from combining phage display libraries and after inducing random mutations, all the selected variants had tryptophan (W) in position79 in the second hairpin-loop motif. However, there was diversity in the original amino acid sequence in position 78, with preferentially accommodate hydrophobic and basic amino acids in this position. Changing the glutamic acid (E) in position 78 in the wild-type sequence to leucine (L) resulted in higher papain affinity than the wild-type cystatin and a two-fold lower K_i than wild-type soyacystatin.

Changes in this first residue may be more involved in determining target specificity during the association process, this was suggested by an increased association-rate constant of leucine (L)-tryptophan (W), without affecting basal level affinity. The LW motif in the second hairpin loop of soyacystatin, with greater affinity to papain, was also identified in the papaya cystatin. The occurrence of these amino acids in the papaya cystatin could be attributed to the co-evolution of the papaya cystatin and the cysteine protease papain in *Carica papaya* (Song *et al.*, 1995).

1.4.4 *Cysteine proteases*

The commercial enzymes selected to be used in this study included papain and cathepsin-L. Furthermore, gut extracts from insect larvae (Colorado potato beetle and banana weevil), as well as tobacco leaf extracts, were used since they are known to contain cysteine protease activity that can be inhibited by plant cystatins. Papain and cathepsin L belong to the same family (C1 proteases) and are characterized by two distinct domains, the N-terminal, mainly α-helixes and the C-terminal, mainly β-sheets. The active site is formed by a cysteine (C) amino acid residue at position 25, a histidine (H) amino acid residue at position 159 and a asparagine (N) amino acid at position 175 (according to papain numbering) (Oliveira *et al.,* 2003).

Papain exhibits broad specificity, cleaving peptide bonds of basic amino acids, e.g. leucine (L) or glycine (G), and also hydrolyzes esters and amides. Papain preferentially accommodates an amino acid bearing a large hydrophobic side chain at the P2 position, but does not accept valine (V) at the P1' position (IUBMB Enzyme Nomenclature). Papain preferentially cleaves at the site indicated by ↓. Hydrophobic amino acids include: alanine

(A), phenylalanine (F), isoleucine (I), leucine (L), valine (V), tryptophan (W), and tyrosine (Y).

P3	P2	P1	↓	P1’	P2’
Xaa	Hydrophobic	Arg	↓	not Val	Xaa
Xaa	Hydrophobic	Lys	↓	not Val	Xaa

Cathepsin L is a lysosomal enzyme and exhibits higher specific activity than cathepsin B or cathepsin H in the degradation of a variety of physiological protein substrates, both cellular and endocytosed macromolecules (Barrett and Kirschke, 1981; Mason *et al.,* 1985). Cathepsin L will preferentially cleave at the site indicated by ↓, with the hydrophobic amino acids listed as well as aromatic amino acids, phenylalanine (F), histidine (H), tryptophan (W) and tyrosine (Y), denoting cleaving specificity.

P4	P3	P2	P1	↓	P1’
Xaa	Hydrophobic	Phe	Arg	↓	Xaa
Xaa	Aromatic	Phe	Arg	↓	Xaa
Xaa	Hydrophobic	Arg	Arg	↓	Xaa
Xaa	Aromatic	Arg	Arg	↓	Xaa

1.5 Previous work on the project

1.5.1 *Plant cystatin sequence analysis*

At the commencement of this study, conserved amino acids in the N-terminal region, first and second inhibitory loops had been identified from 153 gene sequences of plant cystatins from 62 different plant species, including OCI and papaya cystatin, and several OCI and PC variants sequences were produced in an initial study by Dr Schlüter (U Schlüter, unpublished data; Table 1.2). Amino acid substitutions in the conserved amino acids of PC and OCI have been designed based on determined frequency of amino acid residues present in analyzed sequences with the specific objective to understand the particular function of residues in PC potency since this cystatin had some unique characteristics in comparison to OCI specifically in the QVVAG motif. Particular focus on these two plant cystatins was based on results of a previous study where Kiggundu *et al.* (2010) found better inhibition by OCI inhibition of cysteine protease in banana weevil gut extracts when compared to PC and the reason for the difference in inhibitory activity between the two cystatins is still unclear.

1.5.2 *OCI and PC mutagenesis*

The majority of analyzed cystatin sequences had been isolated by sequence homology, but only 27% of them had been analysed for their binding characteristics against cysteine proteases (mainly papain) and most of these sequences contain all the conserved motifs. The highest conservation was found for G in the GG motif (position 10 and 11 in OCI and 8 and 9 in PC) present in the N-terminal end of all analyzed plant cystatins. Some sequence

differences were identified with CLUSTAL W in the alignment of the GG motif, but shifts were limited to +/- one amino acid or two amino acid residues as found for PC when compared to OCI (Figure 1.7). In the plant cystatins analysed in the initial study, 100% of the sequences had at least one conserved G in the N-terminal region (Table 1.2). The importance of G at these positions is still unclear. N-terminal truncated OCI variants lacking both G have previously been found to be as active as non-truncated OCI (Arai *et al.*, 1991), whereas Urwin *et al.* (1995a) reported that substitutions of G in position 10 changed the inhibitory potency of OCI significantly. The GG motif was therefore mutated in the initial study to AA (GG to AA). These two residues, which are non-polar with a neutral side chain charge and without an extensive side chain, were similar to G and had been selected as substitution to provide information about any importance of the GG motive in the N-terminal end. Furthermore, A is naturally found in some cystatins in this position, having either AA or just one A in combination with a G in this position (Table 1.2).

The first inhibitory loop, containing the conserved motif, QVVA/SG, was found to be highly conserved. Q in position 53 in OCI occurs in 98% of sequences. V in position 54 in OCI and in position 52 in PC was found in 84%, with V in position 55 in OCI and 54 in PC found in 97%. At position 56 in OCI (position 55 in PC), A was found in 63% or S in 29%, and G in position 57 in OCI and 56 in PC was found in 98% of cystatins. The conserved residue Q was also present in OCI. The PC sequence contains two unique characteristics, an A in position 52 of PC, which is only found in PC sequence, and an E in position 55 of PC. This variation is only found in three other cystatin sequences. Most plant cystatins have an A in this position, although S is also very common (A and S together occur in 92% of all analysed sequences). This E amino acid residue is situated in the middle of the first inhibitory loop, which is proposed to be inserted into the active site cleft of the cysteine protease, and might

therefore have a significant effect on the binding capacity of the cystatin, as this amino acid is electrically charged compared to the hydrophobic amino acid A, normally found in this position. Since PC shows the greatest discrepancy to the majority of cystatin sequences in the conserved motif of the first inhibitory loop, mutations, shown in Figure 1.7, have been done in the initial study. This allowed the investigation of whether if substitution of these amino acid residues in the PC sequence will influence the activity or specificity of the mutant cystatins towards different cysteine proteases.

Furthermore, the conserved residue Q in OCI or A in PC is preceded in both sequences by Q (position 52 in OCI and 51 in PC). This amino acid residue is present in only 23% of plant cystatin sequences analyzed. Since this amino acid is directly preceding the first amino acid residue of the conserved motif in plant cystatins, which is still in the beta-sheet and not part of the first inhibitory loop, this might have a limited effect on cystatin binding. This might be especially relevant in PC, where the conserved Q is substituted by A. The preceding Q residue in both sequences was therefore substituted by E in the initial study by site-directed mutagenesis. E is found with the highest frequency (63 sequences) at this position in other plant cystatins (Table 1.2). Furthermore, double mutations were created in the study changing the amino acids (positions 52 and 53 in OCI and 51 and 52 in PC) simultaneously to obtain all four combinations of amino acids: QQVVAG as in wild-type OCI changed to Q**A**VVAG, **EA**VVAG and **E**QVVAG, and in wild-type PC the QAVVEG was changed to Q**Q**VVEG, **EQ**VVEG and **E**AVVEG (Figure 1.7).

In the second inhibitory loop, a conserved W was found in 88% of the sequences including OCI in position 84 and PC in position 83. So far only one plant cystatin (*Hv*CPI7) not containing this amino acid residue (F instead of W) has been experimentally tested for altered

binding capacity of the cystatin and no inhibitor activity was found against papain, cathepsin B and cathepsin H (Abraham *et al.*, 2006). AlthouYgh W is the most conserved amino acid in the second inhibitory loop of the plant cystatins, in the preliminary study it was found that variability in this position was higher than in the other conserved sites. A variety of other amino acids residues could be found at this position (Table 1.2). To determine the influence of this residue on OCI and PC activity, W was substituted with G in both sequences. G was found in 4 of the 146 sequences analysed and is found substituting for W more often than other amino acid residues.

Table 1.2: Comparison of phytocystatin amino acid sequences from different plant species. The number of cystatins with the corresponding amino acid in the sequence was scored. The first horizontal row indicates amino acid sequences of cystatins, whereas the first vertical row indicates all potential amino acids found at the particular position. The amino acid residue with the highest frequency at each position is highlighted (shadowed box).

	G	**G**	**G**	**L**	**A**	**R**	**F**	**A**	**V**	**D**	**E**	**H**	**N**	**E**	**Q**	**V**	**V**	**A**	**G**	**T**	**M**	**Y**	**Y**	**L**	**T**	**E**	**A**	**K**	**V**	**W**	**V**	**K**	**P**	**W**
A	11	6	2	6	**110**	1		**134**		18		4			1	2	1	**97**	1		1						**133**			1			11	
C								3												1		1	2										
D	4	2	1							**58**	17		2										4			17	1				1	2	
E	2	1				6				18	**103**	9	6	**63**			1	3	1				4		4			11			5	1	13	
F			1				**125**					5				1				1		2		9	2	**104**		2	1	4	53			1
G	**81**	**134**	**53**		41	5				2	1							1	**150**	2						2						1	4
H				2						1	2	**93**	5	2								45	45							1		1		
I	1		4	12					7	1						7	1			12	13		4	43	11			5	20		2		1	
K			4			13			1	3		2	3	17	1					11	15		3		11	12		**104**				**112**	4	
L	12			**124**		3			1	4			1			9				6	41		1	**99**		2		1	3	9	4	1	7	2
M	1		2											1						20	**48**			1	1						1			
N		4				7				1			**127**					1			28				2			1				1		1
O																							1											
P	1		4					1																								1	**89**	
Q	1		1			4				17	9	1	2	36	**150**			4					1			5		1				19	1	
R	3	1	1	1		**108**				2	1		1	6	1					2			20		6	2						10		1
S	1		4			4	2	9		15	2		1	8			1	45	1	2		1			3	1	7	1			1		8	1
T			1		1			1	1	5			1	18		6		2		**81**	1		3		**96**	2	9	1			2	1	3	
V	24		2	7		1		2	**144**	3				2		**128**	**149**			10	3	1		1	13	5	2	25	**128**	14	**80**	1	5	1
W							19															1								**107**				**135**
Y							7					31									1	**102**	**64**		1					14				
	142	148	80	152	152	152	153	150	154	148	135	145	149	153	153	153	153	153	153	148	151	153	152	153	150	152	152	152	152	150	148	149	145	146

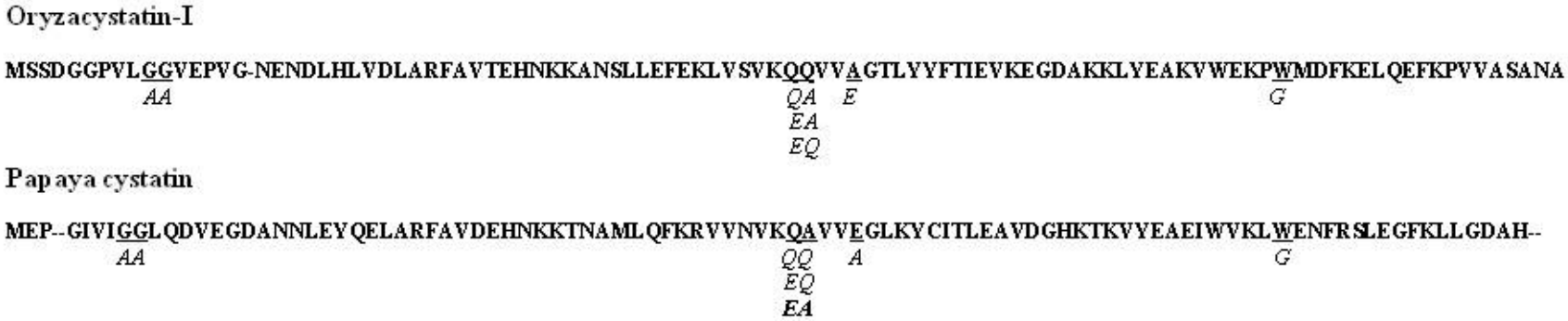

Figure 1.7: Amino acid sequence of oryzacystatin-I (OCI) and papaya cystatin (PC) which were mutated by site-directed mutagenesis in the initial study. Amino acids mutated are underlined and substitutions are in *italics*. **A** represents alanine, **Q** glutamine, **E** glutamic acid, **G** glycine and **W** tryptophan.

Table 1.3: Mutant OCI and PC generated from corresponding wild-type cystatin gene sequences in the site-directed mutagenesis experiments, the amino acid position, the non-synonymous amino acid substitutions and position in the tertiary position are shown. **A** represents alanine, **Q** glutamine **E** glutamic acid, **G** glycine and **W** tryptophan.

Name	Amino acid position	Amino acid substitution	Mutant code	Position in tertiary structure
OCI				
M1	53	**Q to A**	OCI-Q53A	Q**A**VVAG
M2	52	**Q to E**	OCI-Q52E	**E**QVVAG
M3	52, 53	**QQ to EA**	OCI-Q52EQ53A	**EA**VVAG
M7	56	**A to E**	OCI-A56E	QVV**E**G
M9	84	**W to G**	OCI-W84G	P**G**
M11	10, 11	**GG to AA**	OCI-G10AG11A	**AA**
PC				
M4	52	**A to Q**	PC-A52Q	Q**Q**VVEG
M5	51	**Q to E**	PC-Q51E	**E**AVVEG
M6	51, 52	**QA to EQ**	PC-Q51EA52Q	**EQ**VVEG
M8	55	**E to A**	PC-E55A	AVV**A**G
M10	83	**W to G**	PC-W83G	L**G**
M12	8, 9	**GG to AA**	PC-G8AG9A	**AA**

1.5.3 *Cystatin mutation and cloning*

The wild-type gene sequences of OCI and PC cloned into pGEM®-T Easy vector (Promega Corporation), the vector map is shown in Figure 1.8, served as the template for the site-directed mutagenesis PCR reactions. The gene sequences were amplified with oligonucleotides to allow integration of restriction enzyme recognition sites for *Bam*HI at the N-terminal and an *Eco*RI at the C-terminal of the gene sequences, which serves for in-frame cloning into the expression vector pGEX-3X® (vector map shown in Figure 1.9). Figure 1.10 shows a schematic representation of the gene constructs containing the gene sequences of oryzacystatin-I (OCI) and papaya cystatin (PC) in the pGEM®-T Easy vector. The constructs are collectively named pGEM-PC for PC and mutant gene sequences and pGEM-OCI for OCI and mutant gene sequences and the individual constructs were named as pGEMPWT, pGEMOWT, pGEM-M1, pGEM-M2, etc. Once the gene sequences were cloned into the pGEX-3X vector, the constructs were collectively named as pGEX-PC and pGEX-OCI and the individual constructs were named as pGEXPWT, pGEXOWT, pGEXOM1, pGEXOM2, etc.

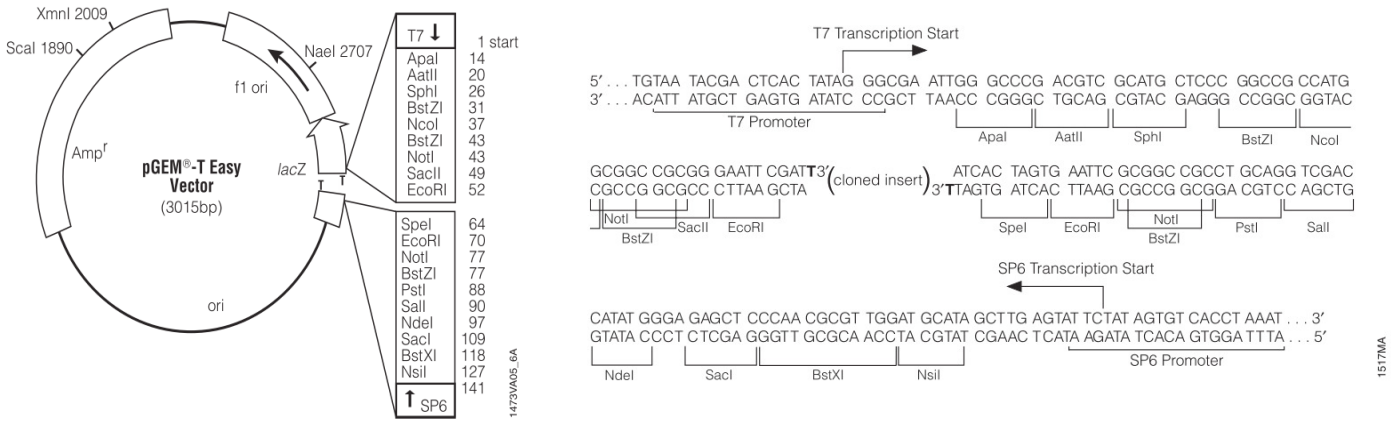

Figure 1.8: Vector maps of pGEM®-T Easy Vector (Promega Corporation, USA). The vector contains an ampicillin resistance gene for selection of transformants containing the plasmid. The 3'-T overhangs of the vector was used for the ligation of the PCR product (OCI/PC), the inserted gene sequences inactivates the α-peptide of the β-galactosidase enzyme, allowing blue-white selection of transformants with the gene insert.

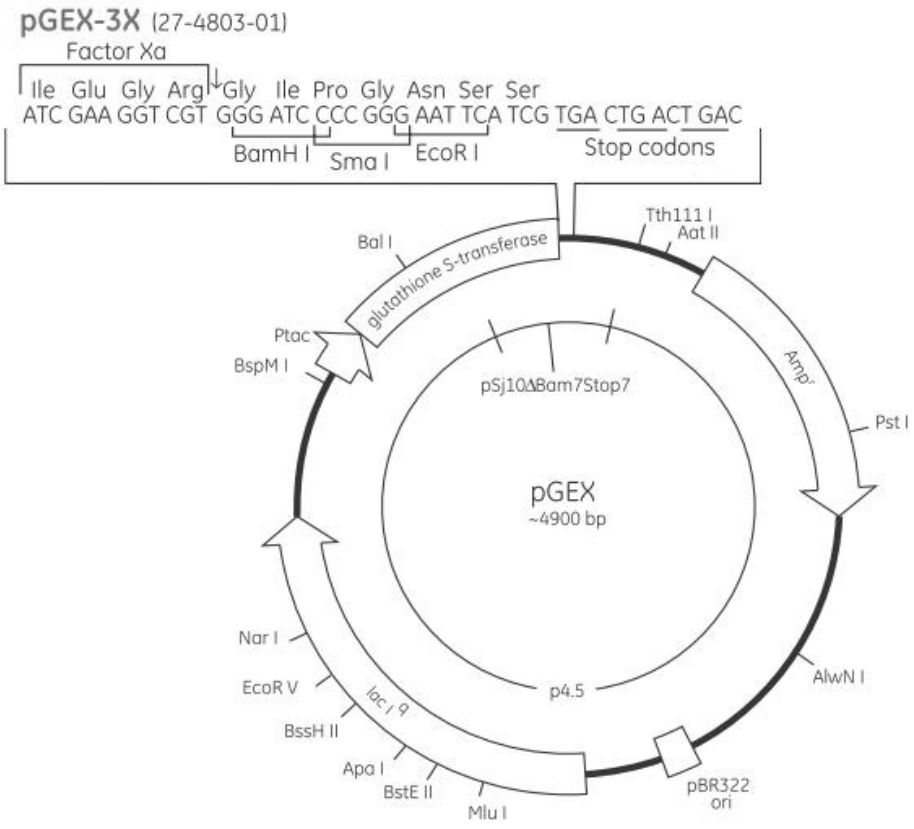

Figure 1.9: Vector map of pGEX-3X® *E. coli* expression vector (Amersham Pharmacia Biotech, UK), the vector has an ampicillin resistance gene for selection of transformants. The expression of the gene insert is controlled by the *lac*I^q gene, which is downstream from the pBR322 ori; the *lac*I^q gene produces a repressor protein regulating the operator of the *tac* promoter. The expression of the gene construct can be controlled by using IPTG, which inhibits the *lac*I^q gene expression and allows the *tac* operator to initiate expression of the gene construct. The gene construct of OCI and PC were cloned into the *Bam*HI-*Eco*RI sites of the vector for in-frame ligation and fusion to the glutathione-S-transferase (GST) protein-tag, the GST-tag and the Factor Xa site are crucial for the downstream purification of the recombinant cystatins by affinity chromatography.

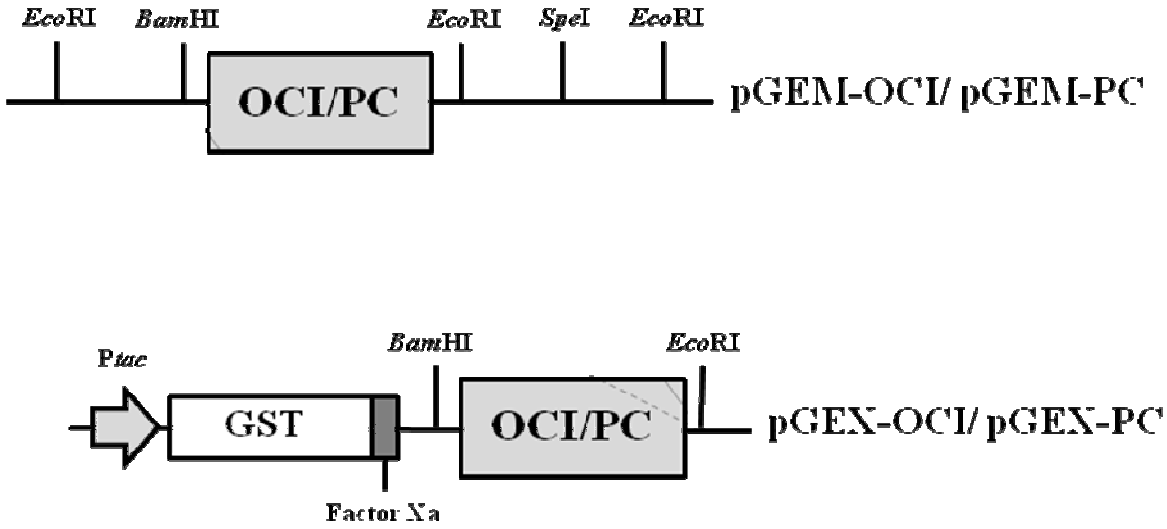

Figure 1.10: A schematic representation of the OCI and PC gene constructs, each flanked by a transcription initiation (ATG) and termination (TAA) codons. The *Bam*HI restriction enzyme site at the N-terminal and the *Eco*RI sites (adjacent to coding sequence and/or located on the vector backbone) are shown. The gene constructs were designed to be removed from the pGEM®-T Easy vector as a *Bam*HI-*Eco*RI fragment. The N-terminal *Bam*HI site is particularly important for the in-frame ligation into the pGEX®-3X expression vector.

The different primers used for the site-directed mutagenesis experiments are listed in Table 1.4. Briefly, the site-directed mutagenesis reaction was setup as follows: 3 µl Plasmid (PC/OCI in pGEM-T® Easy vector), 5 µL of a 10X *Pfu* buffer with $MgSO_4$ (Fermentas), 2.5 µL of a 10 µM primer stock solution of the forward primer or the reverse primer (Inqaba Biotec, South Africa), 1 µL of dNTPs (Fermentas) to have a final concentration of 10 µM in the final reaction volume, 1 µL *Pfu* DNA polymerase (Fermentas) and finally 36.5 µL dH_2O was added to have a final reaction volume of 50 µL per reaction. The temperature cycling conditions were as follows: 95°C for 3 min, followed by 4 cycles of 95°C for 1 min, 55°C for 1 min and 72°C for 6 min. The reaction products were mixed, using 25 µL of the total 50 µL from both the PCR reactions for the forward primer and the corresponding reverse primer,

and an additional 1 µL of *Pfu* enzyme was added. The following temperature cycling conditions were then used: 3 min at 95°C, followed by 20 cycles of 1 min at 95°C, 1 min at 55°C and 6 min at 72°C, the reaction was cooled down to 37°C and 1 µl of *Dpn*I (Fermentas) was added and the reaction was incubation for 90 min. From the reaction volume 10 µL was used to transformation 50 µL of *E. coli* JM109 competent cells, using the heat shock transformation method. One hundred and fifty microlitres from the transformation reactions were plated onto LB plates containing ampicillin at 100 mg/L. Plasmids containing mutated PC or OCI sequences were sequenced using a T7 or M13 primer for sequencing.

Table 1.4: Primer sequences used to generate the mutant cystatins by site directed mutagenesis. The regions in each primer sequence that are bolded indicate the position of the introduced nucleotides, corresponding to the desired amino acid change.

Name	Forward primer	Reverse primer
OCI		
M1 (Q**Q** to Q**A**)	5' GTGAGTGTGAAGCAG**GCA**GTTGTCGCTGGCAC 3'	5' GTGCCAGCGACAAC**TGC**CTGCTTCACACTCAC 3'
M2 (**Q**Q to **E**Q)	5' GAGAAGCTTGTGAGTGTGAAG**GAA**CAAGTTGTCGCTGGCACTTTG 3'	5'CAAAGTGCCAGCGACAACTTG**TTC**CTTCACACTCACAAGCTTCTC 3'
M3 (**QQ** to **EA**)	5' GAAGCTTGTGAGTGTGAAG**GAGGCA**GTTGTCGCTGGCACTTTG 3'	5' CAAAGTGCCAGCGACAAC**TGCCTC**CTTCACACTCACAAGCTTC 3'
M7 (V**A**G to V**E**G)	5' GAAGCAGCAAGTTGTC**GAA**GGCACTTTGTACTATTTC 3'	5' GAAATAGTACAAAGTGCC**TTC**GACAACTTGCTGCTTC 3'
M9 (P**W** to P**G**)	5' CTGGGAGAAACCA**GGG**ATGGACTTCAAG 3'	5' CTTGAAGTCCAT**CCC**TGGTTTCTCCCAG 3'
M11 (**GG** to **AA**)	5' GCCGGTGCTT**GCAGCC**GTCGAGCCGG 3'	5' CCGGCTCGAC**GGCTGC**AAGCACCGGC 3'
PC		
M4 (Q**A** to Q**Q**)	5' GGTTGTGAATGTAAAGCAG**CAG**GTGGTTGAAGGCTTAAAG 3'	5' CTTTAAGCCTTCAACCAC**CTG**CTGCTTTACATTCACAACC 3'
M5 (**Q**A to **E**A)	5' GAGGGTTGTGAATGTAAAG**GAA**GCAGTGGTTGAAGGC 3'	5' GCCTTCAACCACTGC**TTC**CTTTACATTCACAACCCTC 3'
M6 (**QA** to **EQ**)	5' GAGGGTTGTGAATGTAAAG**GAACAG**GTGGTTGAAGGCTTAAAGTAC 3'	5' GTACTTTAAGCCTTCAACCAC**CTGTTC**CTTTACATTCACAACCCTC 3'
M8 (V**E**G to V**A**G)	5' GCAGGCAGTGGTT**GCA**GGCTTAAAGTAC 3'	5' GTACTTTAAGCC**TGC**AACCACTGCCTGC 3'
M10 (L**W** to L**G**)	5' CTGGTTGAAGCTC**GGG**GAGAATTTCAGG 3'	5' CCTGAAATTCTC**CCC**GAGCTTCAACCAG 3'
M12 (**GG** to **AA**)	5' GAATTGTGATC**GCAGCT**TTGCAGGACG 3'	5' CGTCCTGCAA**AGCTGC**GATCACAATTC 3'

1.6 Research aim and objectives

Despite emerging evidence about certain amino acids in the cystatin sequence playing an important role in inhibitory activity, there is still a lack of detailed information about the possible function of individual amino acids, particularly in the conserved regions of the cystatin amino acid sequence. Therefore, in this study, mutant cystatins derived from rice (oryzacystatin-I) and papaya (papaya cystatin) were used to determine the importance of individual amino acids in the N-terminal and first and second inhibitory loops of these two cystatins for inhibitory activity against cysteine proteases and cysteine protease activity contained in insect and plant extracts. The two cystatins were selected due to the unique sequence characteristics of the papaya cystatin in the conserve motif and also due to their significantly different activity when a gut extract from banana weevils was used to test their inhibitory efficiency (Kiggundu *et al.*, 2010). In this study, 12 mutant cystatins were produced and purified and the results of their inhibitory activity against papain, cathepsin L and extracts of Colorado potato beetle larvae, gut extracts of banana weevil larvae and a *Nicotiana benthamiana* leaf extract are reported. To achieve the aim, the following three objectives were set:

1. Produce and purify wild-type and mutant cystatins by applying the GST-fusion protein technique and the GSH affinity chromatography technique to produce cystatins for *in vitro* testing against cysteine proteases and cysteine protease-containing insect and plant extracts.

2. Determine the inhibitory activity of wild-type and mutated cystatins using fluorometric assays with specific fluorescent substrates to evaluate the importance of individual amino acids in the conserved regions of cystatins.

3. Computer-simulate the interaction between mutant cystatins with the model cysteine protease papain to determine changed interactions of mutated amino acids in comparison to native amino acids.

2. MATERIALS AND METHODS

2.1 DNA work

2.1.1 *Preparation of* E. coli *competent cells*

Escherichia coli competent cells were prepared for strains DH5α and BL21 according to the publication of Inoue *et al.* (1990) and the transformation procedure that followed was the heat-shock transformation method also described in the article. The *E. coli* strain DH5α with the genotype: F-, φ80d*lac*ZΔM15 Δ(lacZYA-argF)U169, *deo*R, *rec*A1, *end*A1, *hsd*R17(r_k^- m_k^+ *pho*A, *sup*E44, λ^-, *thi*-1, *gyr*A96, *rel*A1) is deficient in DNases due to the *end*A1 mutation, allowing high plasmid yields to be obtained and has an enhanced insert stability due to the *rec*A1 mutation. The strain can be used for blue/white screening and can accept large plasmids due to the *deo*R mutation. The *E. coli* BL21 strain with the genotype: F-, *omp*T, *hsd*S(r^-_B,m^-_B), *gal*, *dcm*, *Ion*) was selected for recombinant protein expression, as the strain is compatible with the *tac* promoter on the pGEX-3X® expression vector, and lacks the *Ion* and *omp*T proteases, which could cause the degradation of heterologous proteins expressed in the strain (Sigma-Aldrich, 2006).

2.1.2 *Cloning into expression vector pGEX*

The pGEM-OCI variants (OCI-WT, M1, M2, M3, M7, M9 and M11) were maintained in *E. coli* JM109 cells which were stored at -80°C in a glycerol stock until used. The cells were streaked out onto LBA plates (10 g/L tryptone powder [Merck], 10 g/L NaCl [Merck], 5 g/L

yeast extract [Merck] and 15 g/L bacteriological agar [Merck] containing 100 µg/mL ampicillin [Sigma Aldrich]). The plates were incubated overnight (O/N) at 37°C, the following day a single colony was selected from the plates and used to inoculate 5ml of liquid LB (10 g/L tryptone powder, 10 g/L NaCl and 5 g/L yeast extract) containing 100 µg/mL ampicillin, these cultures were incubated O/N at 37°C with shaking at 200 rpm. The following day, the cells were harvested by centrifugation at 13200 rpm for 2 min at room temperature (RT). The cell pellet was used for a plasmid miniprep using the Fermentas – GeneJET™ Plasmid Miniprep Kit, according to the manufacturer's instructions with variations. The elution buffer (10 mM Tris-HCl, pH 8.5) supplied with the kit was not used, as the additional salt concentration might interfere with enzyme reactions still to follow. Distilled water was used instead and was pre-heated to ±60°C before used, while still following the recommended volumes and incubation times. The eluted plasmid was stored at -20°C until used.

The eluted plasmid for each of the pGEM-PC and pGEM-OCI variants were used for a restriction enzyme digest to remove the gene insert, which was cloned into the pGEX®-3X expression vector. All restriction enzymes and related buffers used were from Fermentas. Briefly, 20 µL of plasmid DNA was incubated in 4 µL of 10X Tango™ buffer, 0.5 µL of *Eco*RI (10 U/µL) and 1 µL of *Bam*HI (10 U/µL). The reaction compositions were followed as per manufacturer's instructions for a double digest in a single reaction containing 2X Tango™ buffer, *Eco*RI at 1 U/µL and *Bam*HI at 2 U/µL (2-fold excess of enzyme to compensate for lowered activity in the 2x Tango™ buffer). The reaction was incubated at 37°C for 1 h, before separation on a 1% agarose (Lonza, USA) gel at 100V for 30 min.

The agarose slice containing the DNA insert (±300 bp) was cut from the gel using a scalpel and the agarose was dissolved to release the *Bam*HI-*Eco*RI DNA fragment. The GFX Gel purification kit (GE Healthcare) was used according to the manufacturer's instructions to purify DNA from the agarose, again with the same variation from the manufacturer's instructions. The elution buffer (10 mM Tris-HCl, pH 8.5) supplied with the kit was not used, distilled water was used instead and pre-heated to ±60°C before used, the recommended volumes and incubation times were as per manufacturer's instructions. The eluted insert was stored at -20°C until used.

The purified insert for PC and OCI and each mutant were then used in a ligation reaction with a dephosphorylated pGEX-3X plasmid, pre-cut with *Bam*HI and *Eco*RI. The ligation reaction for each PC and OCI was prepared to contain a 3:1 ratio of insert to vector, 1 µL (5 U/ µL) T4 DNA Ligase (Fermentas), 2 µL of 10X ligation buffer (Fermentas), 7 µL of dephosphorylated pGEX-3X plasmid (*Bam*HI-*Eco*RI fragment removed) (±7.2 ng/µL) and 10 µL of insert (±15.0 ng/µL). The ligation reaction was incubated at RT for 1 h, prior to heat inactivation of the T4 DNA Ligase at 65°C for 10 min.

Five microlitres of the ligation mix was used to transform 50 µL of *E. coli* DH5α competent cells by heat shock transformation, the remaining ligation mix was stored at -20°C, until required for further use. The 100 µL of the transformation mixture was streaked out on LBA plates, containing 100 µg/mL ampicillin, these plates were incubated O/N at 37°C. A single colony from each was selected from the transformation plates, to inoculate 5 mL LB media containing 100 µg/mL ampicillin. The culture was incubated O/N at 37°C with 200 rpm of shaking. All the cells from the cultures were harvested by centrifugation at 13200 rpm for 2 min, the cell pellets were used for plasmid preparations, again using the Fermentas -

GeneJET™ Plasmid Miniprep Kit, performed as described above. The extracted plasmid for each of the PC and OCI variants was used in a *Bam*HI-*Eco*RI digestion reaction to confirm the insertion of the *Bam*HI-*Eco*RI gene fragment.

2.1.3 *DNA sequencing*

All obtained pGEX-PC and pGEX-OCI constructs were sequenced at FABI, using pGEX specific sequencing primers (Amersham Pharmacia Biotech) to confirm the correct mutations in PC and OCI coding sequence. The experimental procedure followed was according to the BigDye® Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems, US, and is available from www.appliedbiosystems.com and https://bi.fabi.up.ac.za/seqserve websites.

2.2 Protein work

2.2.1 *Protein expression*

Plasmid DNA (5 µL) for gene sequences cloned was used to transform 50 µL of competent DH5α cells to carry out the protein expression studies. From bacterial plates for each of the variants in *E. coli* DH5α cells (pGEX-OCI: OCI-WT, M1, M2, M3, M7, M9 and M11 and pGEX-PC: PC-WT, M4, M5, M6, M8, M10 and M12), a single colony was selected for each of the variants and used to inoculate a 5 ml LB culture, containing 100 µg/mL ampicillin. The culture was incubated O/N at 37°C with shaking at 200 rpm, from the culture 0.1 mL was used to inoculate 10 mL of LB media, containing 100 µg/mL ampicillin. The culture was incubated at 37°C with 200 rpm of shaking, until an OD_{600nm} of 0.6 was reached. The non-

induced sample was taken prior to adding IPTG to a concentration of 0.2 mM, the culture was incubated for a further 16 h under the same conditions. The induced sample was taken after IPTG treatment and was diluted to an OD_{600nm} of 2.0. The samples were diluted in a 5X SDS-PAGE sample buffer containing 62.5 mM Tris-HCl (Merck) at pH 6.8, 20% glycerol (Merck), 2% SDS (Merck), 5% β-mercaptoethanol (Merck) and 0.5% (w/v) bromophenol blue (Bio-Rad) in a 4:1 ratio of sample to buffer, the samples were heated to 95°C for 4 min before storing at -20°C until analysis on SDS-PAGE. Analysis of all samples was done on a 12% SDS-PAGE, which was setup according to the instructional leaflet "Reagent and Gel Preparation for Laemmli SDS Polyacrylamide Gel Electrophoresis" from Bio-Rad Laboratories. After confirming GST-tagged proteins based on the molecular marker, another 5 µL of pGEX-OCI/PC plasmid was used to transform 50 µL of competent *E. coli* BL21 cells, the expression procedure was followed as above. All of the cells from the 250 mL cultures were harvested by centrifugation at 8000 x*g* and 4°C for 5 min, the pellet for each was stored at -20°C until the purification of recombinant PC and OCI proteins.

2.2.2 *Protein purification*

The cells harvested from the 250 mL culture after IPTG treatment (or without IPTG treatment was used as a control) were freeze-thawed 3-times, before re-suspending the pellet in 3 mL of lysis buffer containing 50 mM Tris-HCl at pH 8.0, 5% (w/v) sucrose, 50 mM EDTA (Sigma-Aldrich) and 5% (v/v) Triton X-100 (Merck). The suspension was incubated on ice for 5 min, before the samples were centrifuged at 10000 x*g* at 4°C for 10 min. The supernatant was transferred to a 15 mL Falcon tube, the lysis supernatant sample was taken, and 250 µL of Glutathione Sepharose™ 4B (GE Healthcare) was added. The binding between the GST-tag and the GSH-Sepharose was allowed to take place for 30 min at RT under 35 rpm shaking,

followed by 90 min at 4°C with 35 rpm of shaking. The Sepharose was washed 3-times with 2.5 mL of 50 mM Tris at pH 8.0. The GSH-Sepharose was then incubated O/N at RT in 250 µL of 1x cleaving buffer and 3 µL of Factor Xa enzyme (Novagen, Merck), with slight agitation at 35 rpm on a shaker. After incubation, the reaction was centrifuged at 8000 x*g* for 10 min. The supernatant (±250 µL) was then transferred to a 1.5 mL Eppendorff tube. The Sepharose was washed again with 250 µL of 50 mM Tris at pH 8.0 to recover any remaining protein and the sample was centrifuged again and the two supernatant samples were pooled.

2.2.3 *Protein determination*

Protein concentration of all protein samples was determined using the commercially available Bradford protein determination assay reagent (Bio-Rad) and using a BSA standard (Bio-Rad). The OD_{595nm} measurements were done in a clear, flat-bottom, 96-well plate in a total volume of 200 µL. Triplicate measurements of all standards and samples were taken and mean values were used. A BSA stock solution of 10 µg/µL was used for preparation of the BSA standards, which were prepared as shown in Table 2.1.

Table 2.1: The BSA standards which were prepared to be used as comparison when determining the quantity of purified cystatin obtained from the purification procedure.

Standard concentration (µg/µL)	0	1	2	3	4	5	6	7	8	9
BSA stock (µL)	0	10	20	30	40	50	60	70	80	90
dH_2O (µL)	100	90	80	80	60	50	40	30	20	10

Only 2 µL of each of the standards were loaded in triplicate onto the plate and 5 µL of the cystatin samples were added in triplicates. The Bradford protein determination assay reagent was added to a final concentration of 20% (v/v) to the total of 200 µL. The OD_{595nm}

measurements for the standards were used to determine a linear regression curve, which serves as reference to determine the concentration of the cystatin samples. Figure 2.1 illustrates the result expected from the standard sample measurements.

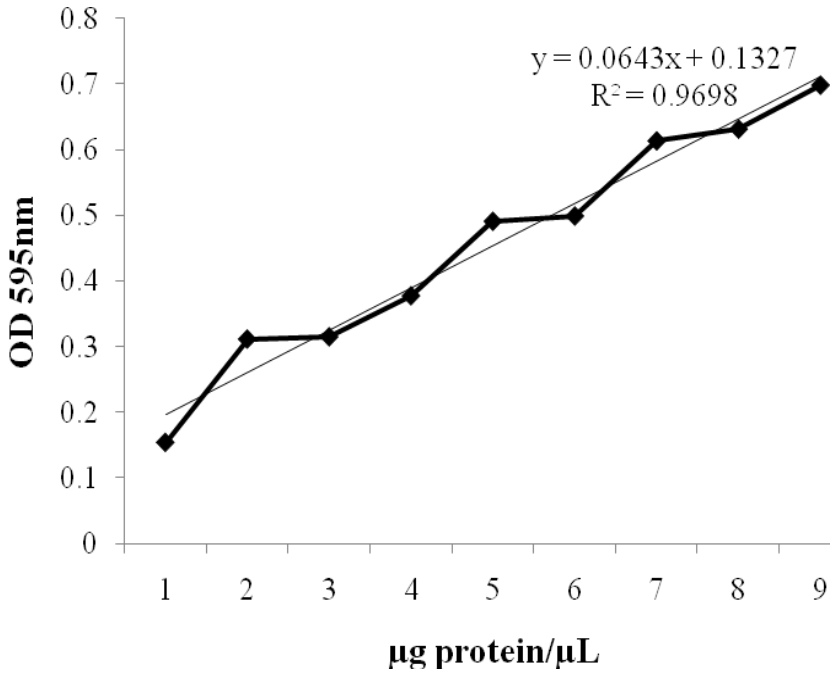

Figure 2.1: BSA standard curve to be generated from the OD_{595nm} measurements of the standard solutions. The linear regression equation was used to calculate the quantity of purified cystatin obtained.

2.2.4 *Determination of K_i values*

A papain (cysteine protease) stock was prepared from a commercially available latex preparation of C. *papaya* (Sigma-Aldrich) which contains a minimum of 3.2 U/mg of enzyme. Human cathepsin-L, another commercially available protease, was also purchased from Sigma-Aldrich. The cysteine protease activity was determined and the K_i values for the wild-type PC and OCI, and the mutant variants were determined. The measurements were performed using black, flat-bottom, polysorp, 96-well plates (Nunc). Each well contained a total reaction volume of 100 µL, the fluorogenic substrate Z-Phe-Arg-MCA (cathepsin L-like substrate from Sigma-Aldrich) was used at a final concentration of 8 µM from a 100 µM stock dissolved in dimethyl sulfoxide (Sigma-Aldrich).

Dissociation (inhibition) constants (K_i) for the interaction between the different PC and OCI variants, with the model cysteine proteases, papain and human cathepsin L, were determined according to the method outlined by Goulet *et al.* (2008). Substrate hydrolysis progress curves were monitored as described by Salvesen and Nagase (1989) and the linear equation was determined as described by Henderson (1972). Papain activity was measured in 100 mM sodium phosphate buffer, pH 7.0, and cathepsin L activity was measured in 100 mM citrate phosphate buffer, pH 5.5. The hydrolysis reactions were allowed to proceed at 25°C, under reducing conditions, by adding 10 mM L-cysteine to the reaction buffer. Cysteine protease activities were determined using a Fluostar Galaxy fluorimeter (BMG, Offenburg, Germany), using a 360 nm excitation filter and a 450 nm emission filter. K_i values were calculated, after firstly empirically estimating the $K_{i(app)}$ and K_m values, and then using the equation $K_i = K_{i(app)}/(1 + [S]/K_m)$. The K_m values which were used were 13.6 µM and 1.0 µM, for papain,

and cathepsin L, respectively, which were determined with their test substrates (Goulet *et al.*, 2008).

2.2.5 *Measurement of PC and OCI activity*

Total protein extracts from plant material were used as sources for cysteine proteases in assays to measure PCs and OCIs inhibitory activity. Extracts were prepared from leaf material from *Nicotiana benthamiana* by crushing in liquid nitrogen, weighed and a 100mM sodium phosphate buffer at pH 6.5, was added in a ratio of 1:2 (100 µg : 200 µL; sample : buffer). The solution was incubated for 30 min on ice before centrifuging at 15000 x*g* for 15 min at 4°C to remove debris. The supernatant was removed and total protein concentration determined. Total protein extracts from larvae of banana weevils (*Cosmopolites sordidus*) and Colorado potato beetles (*Leptinotarsa decemlineata*) were used as another source of cysteine proteases and were provided by Prof. K.J. Kunert. The total protein content of the extracts was determined using the same method for the quantification of the purified cystatins.

The concentration of inhibitor required to reduce the proteolytic activity by 40-60% was determined using OCI and PC for each of the different extracts, the determined concentration for the wild-type inhibitor was used for the respective mutants and allowed for comparative analysis of OCI against PC and against the mutant inhibitors. Cysteine protease inhibitor activities were assayed as before, by the monitoring the progression curves of substrate hydrolysis, using the synthetic fluorogenic substrate *Z*-Phe-Arg-MCA. Substrate hydrolysis by the plant and insect digestive proteases were allowed to proceed in a 100 mM phosphate buffer, pH 6.5, at 25°C. The inhibitory activities were monitored using a Fluostar Galaxy

fluorimeter (BMG), using the same excitation and emission filters as mentioned before. As a maximal inhibitory control, E-64 (from Sigma-Aldrich) was used as a broadspectrum inhibitor for cysteine proteases (Michaud *et al.*, 1993).

2.3 Bioinformatics work

Discovery Studio v2.5 (Accelrys Software Inc., USA) software package was used to produce the different cystatin models and to simulate the interaction energies between the cysteine protease, papain, and the PC and OCI variants generated in the project. The electrostatic interaction energies and the van der Waals interaction energies were calculated between the protease model and each of the cystatins. The binding interfaces of the models were refined and re-scored using the RDOCK protocol and the energy minimization was done using the CHARMm Polar H forcefield. The binding interface residues were grouped and the binding energies were calculated.

The UCSF Chimera software package (http://www.cgl.ucsf.edu/chimera) was used for manipulation of the individual models of OCI and PC for illustrative purposes in Figures 1.2-6 and 3.12 (Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco).

2.4 Statistical methods

The statistical significance of variations in mean values of both *in vitro* and *in silico* was determined using the Student's two-tailed *t* test at stringency level of $P \leq 0.05$.

3. RESULTS

3.1 DNA work

3.1.1 *Cloning into vector pGEX*

All OCI and PC wild type and mutant clones were sequenced to confirm the correctness of mutations introduced before a restriction enzyme digest was performed to obtain an OCI or PC fragment for further sub-cloning into the pGEX®-3X vector. Figure 3.1 shows the digestion of plasmid pGEMPWT containing the PC wild-type and the different pGEMPM vectors containing PC mutant coding sequences with restriction enzymes *Bam*HI and *Eco*RI. A DNA fragment with the predicted size of 300 bp for the coding sequence was detected after restriction enzyme digest, confirming correct cloning of coding sequences into vector pGEM®-T Easy. An identical result was obtained when either plasmid pGEMOWT containing the OCI wild-type or the different pGEMOM vectors containing OCI mutant coding sequences were digested with *Bam*.

After isolation of wild-type and mutant OCI and PC sequences from an agarose gel, purified *Bam*HI/*Eco*RI fragments were cloned into plasmid pGEX®-3X to create vectors pGEXOWT and pGEXOM vectors for OCI and pGEXPWT and different pGEXPM vectors for PC. Cloning into the plasmid pGEX®-3X to created an in-frame gene fusion with the GST coding sequence at the 5'-end of the OCI or PC coding sequence (Figure 2.1). To confirm that a correct *Bam*HI/*Eco*RI fragment was cloned into pGEX®-3X, all pGEX plasmids carrying

OCI or PC were also digested with restriction enzyme *Eco*RI or *Bam*HI alone to confirm the presence of unique *Eco*RI and *Bam*HI restriction sites in the plasmids. Figure 3.2 shows an example of the restriction enzyme analysis carried out with plasmid pGEXPWT, to confirm that the gene insert was ligated in-frame as a *Bam*HI/*Eco*RI fragment. A potential second *Eco*RI restriction might have been present upstream of the gene sequence, due to the *Eco*RI-*Eco*RI restriction enzyme sites flanking the TA overhangs on plasmid pGEM®-T Easy. An incomplete *Bam*HI digest could result in a similar sized *Eco*RI-*Eco*RI fragment being ligated and which would not be in-frame with the GST-tag (Figure 3.2).

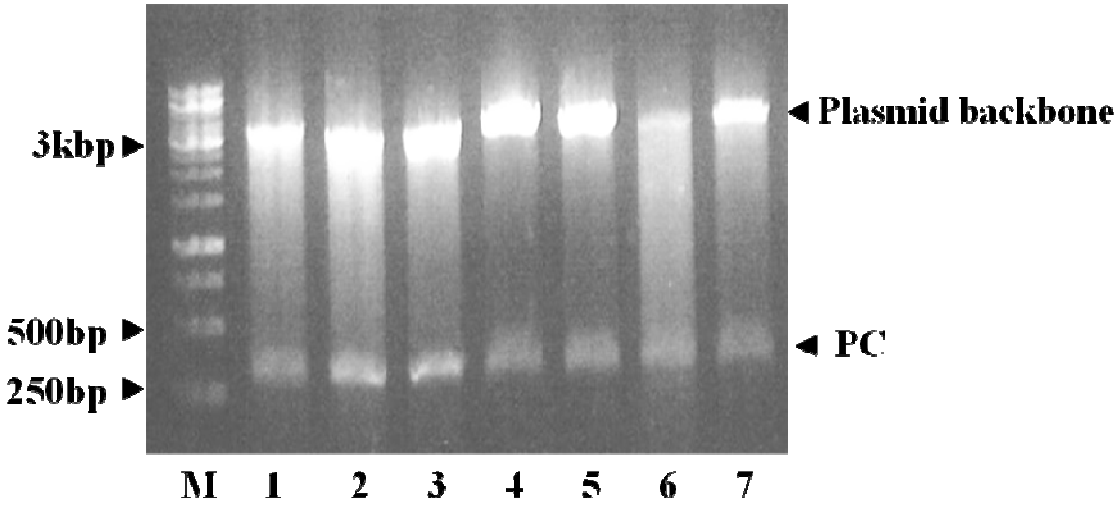

Figure 3.1: Digestion of plasmid pGEMPWT (lane 1) and pGEMPM4 (lane 2), M5 (lane 3), M6 (lane 4), M8 (lane 5), M10 (lane 6) and M12 (lane 7) with restriction enzymes *Bam*HI and *Eco*RI to release a 300 bp DNA PC fragment. M represents a DNA ladder and particular DNA fragment sizes and the remaining 3.1 kbp pGEM®-T Easy vector backbone are indicated.

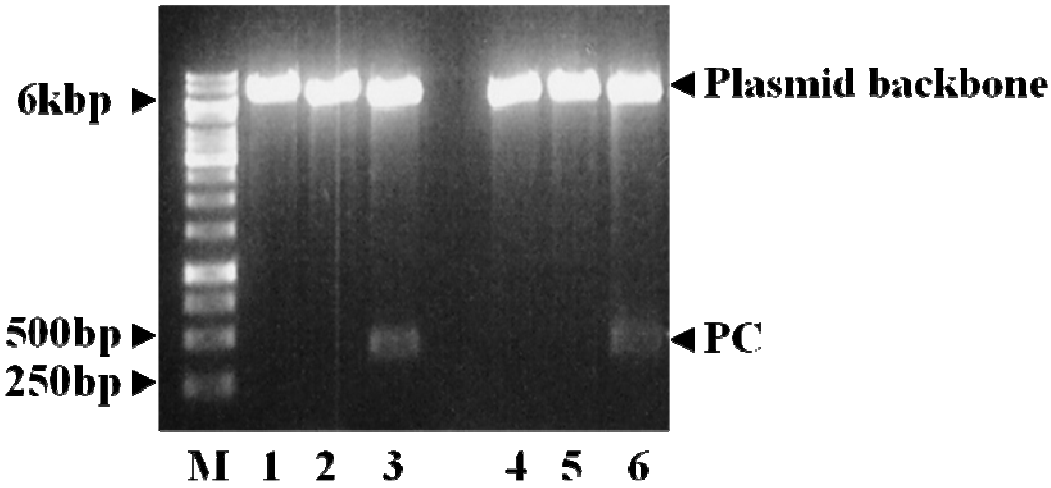

Figure 3.2: Digestion of plasmid pGEXPWT with *Bam*HI (lane 1), *Eco*RI (lane 2) and *Bam*HI/*Eco*RI combined (lane 3) to release a 300 bp PC DNA fragment. The digestion of plasmid pGEMPM4 with *Bam*HI (lane 4), *Eco*RI (lane 5) and *Bam*HI/*Eco*RI combined (lane 6) is also shown to release the expected 300 bp PC DNA fragment. M represents the DNA ladder and particular DNA fragment sizes and the remaining 4.9 kbp pGEX®-3X vector backbone are indicated.

All pGEX-OCI and pGEX-PC expression plasmids were re-sequenced to confirm the correctness of the junction site of the GST and either OCI or PC coding sequence for in frame expression of a fusion protein. All plasmids were finally maintained in *E. coli* DH5α, and stored at -80°C until protein expression studies.

3.2 Protein work

3.2.1 *Protein expression*

Expression of GST-PC fusion proteins in *E. coli* DH5α cells, which were used as a pre-screen to confirm correctness of plasmids allowing fusion protein production, is shown in figure 3.3. GST fusion proteins were produced after induction of expression by IPTG when the cell suspension had an OD_{600nm} of 0.6. Fusion protein samples were prepared from cells grown for 12 hrs in the presence of IPTG by first diluting the suspension to an OD_{600nm} of 1.0 with distilled water. Cells were then boiled in the presence of 1% SDS and proteins analyzed by SDS-PAGE on a 12% gel. The expected 37 kDa GST fusion protein (GSTF) was detected to be expressed in all induced cells (Figure 3.3) but no significant fusion protein production was found in non-IPTG-induced cells. An identical protein expression experiment was carried out for GST-OCI fusions and for all OCI proteins (wild-type and mutants) a GST fusion protein was detected (data not shown).

To optimize fusion protein expression, all plasmids were transferred into cells of *E. coil* strain BL21 by heat shock treatment. The *E. coli* strain BL21 is widely used for expression of recombinant proteins. Figure 3.4 shows the analysis of expressed proteins in non-induced and IPTG-induced *E. coli* BL21 cells (6 hrs protein expression induction) expressing the different pGEX-PC plasmids by a SDS-PAGE on a 12% gel. The expected 37 kDa GST fusion protein (GSTF) was detected to be expressed in all induced *E. coli* BL21 cells (Figure 3.4) but no significant fusion protein production was found in non-IPTG-induced cells. An identical protein expression experiment was carried out for GST-OCI fusions and for all OCI proteins (wild-type and mutants) a GST fusion protein was detected (data not shown).

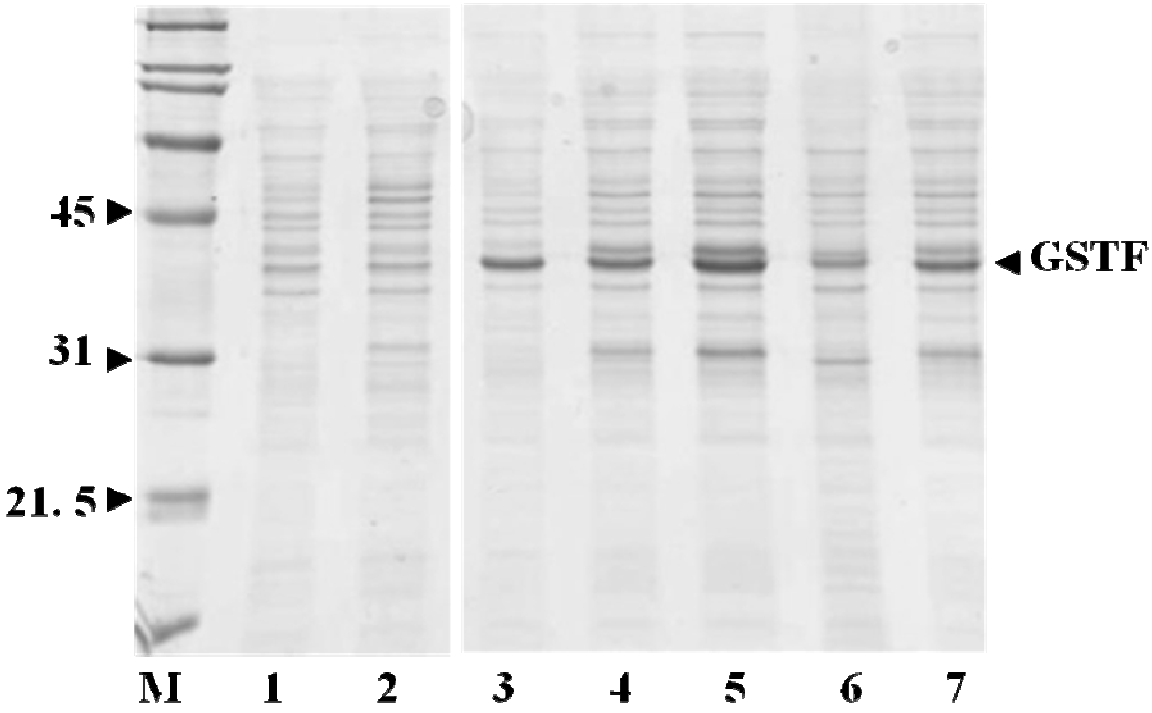

Figure 3.3: M represents the protein molecular weight marker and only particular sizes (in kDa) are indicated. Lane 1 represents the non-induced sample of PC; lane 2 represents the non-induced sample taken from M5. Lane 3 represents the induced expression of PC in *E. coli* DH5α. Lane 4 represents the induced expression samples for M4, lane 5 represents the induced expression sample of M5, lane 6 represents the induced expression sample of M6 and lane 7 represents the induced expression sample of M8.

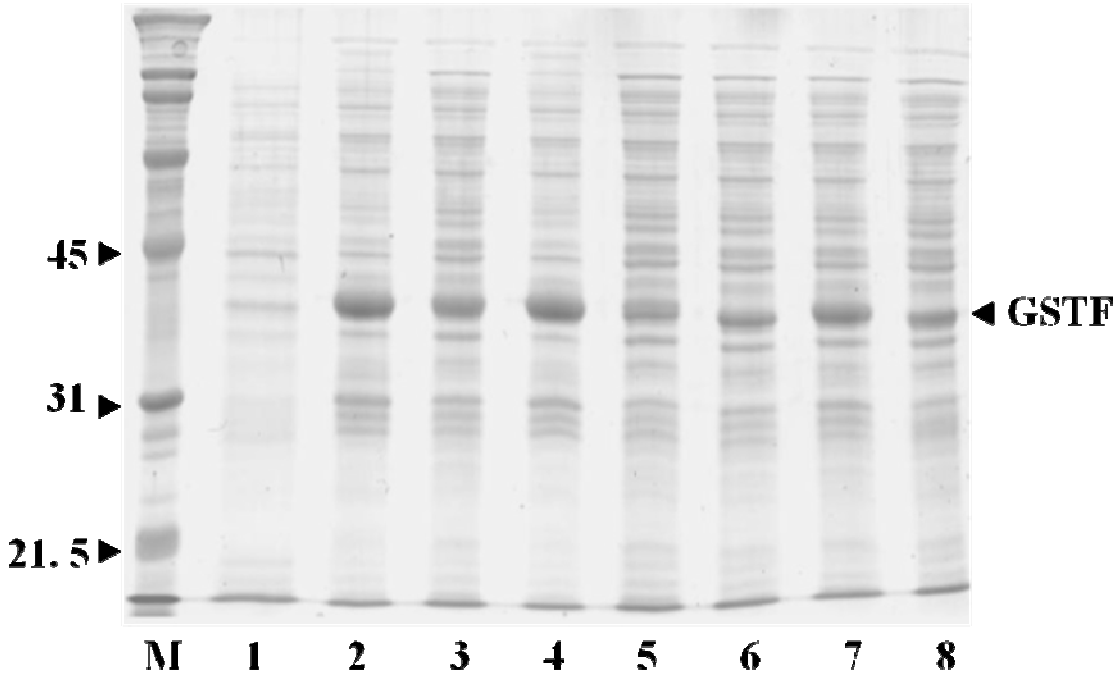

Figure 3.4: Expression of GST-PC fusions in *E. coli* BL21 cells and analysis of expressed proteins by SDS-PAGE on a 12% gel. Lane 1 represents protein expression in the non-IPTG induced *E. coli* BL21 cells; lane 2 expression in non-IPTG induced cells carrying plasmid pGEXPM5; lane 3-8 expression in cells carrying plasmids pGEXPM4, M5, M6, M8, M10 and M12. M represents protein molecular weight markers and particular sizes (in kDa) and position of the fusion protein (GSTF) on the gel are indicated.

3.2.2 *Fusion protein purification*

For fusion protein purification, expressed PC and OCI wild-type and mutant proteins fused to GST were released from IPTG-induced *E. coli* BL21 cells by applying a freeze-thawing procedure. After removing cell debris by centrifugation, the supernatant containing the soluble fusion protein was mixed with GSH-Sepharose 4B (Amersham Pharmacia Biotech) to bind the GST fusion protein. Figure 3.5 shows the different steps of the purification procedure including the step of binding the GST fusion protein to GSH-Sepharose 4B. All protein samples were analyzed by SDS-PAGE on a 12% gel and the expected 37 kDa GST fusion proteins bound to GSH-Sepharose 4B was detected (Figure 3.5). An identical procedure was carried out with OCI and mutant OCI and an identical result as found for PC was obtained (data not shown).

The GST fusion protein bound to the GSH-Sepharose was collected by centrifugation. Treatment of the different GST-fusions with protease Factor Xa released OCI and PC from the GST fusion protein and the released OCI and PC protein could be detected in the supernatant. Figure 3.6 shows the protein samples analyzed by SDS-PAGE on a 15% gel at the different purification steps for PC and OCI and the released purified PC (or OCI) was found with the predicted size of about 11kDa.

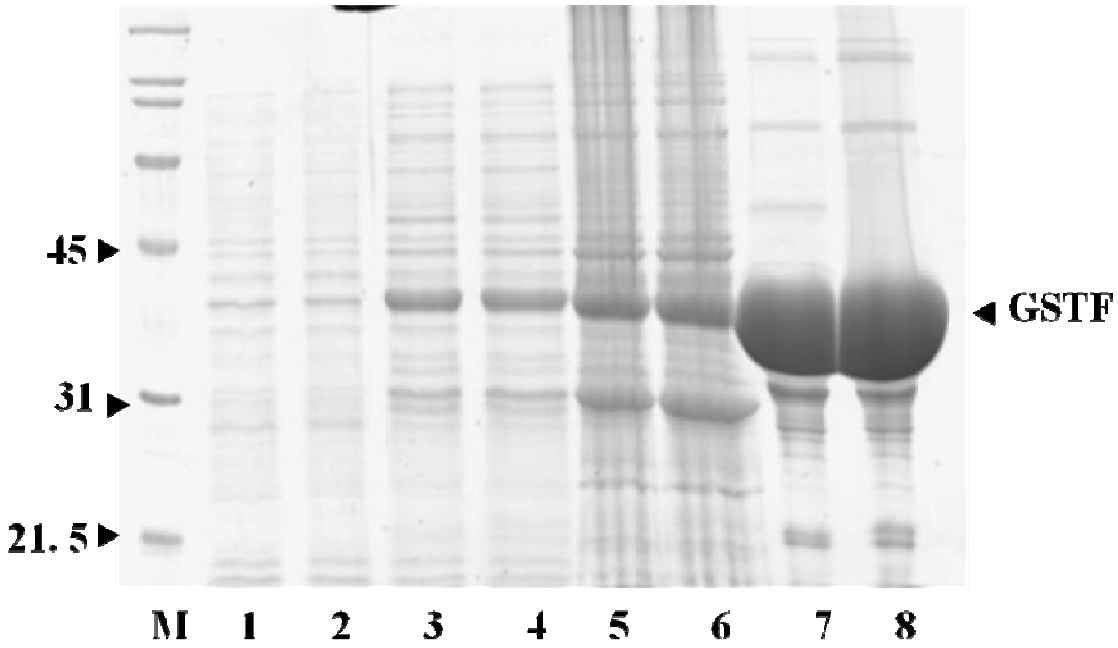

Figure 3.5: Expression of GST-PC fusions in *E. coli* BL21 cells and purification of expressed fusion proteins. Expressed proteins of different purification steps were analyzed by SDS-PAGE on a 12% gel. Lane 1 represents protein expression in the non-IPTG induced BL21 cells carrying pGEXPWT; lane 2 expression in non-IPTG induced cells carrying plasmid pGEXPM5; lanes 3 and 4 expression in IPTG induced cells carrying plasmids pGEXPWT (lane 3) and pGEXPM5 (lane 4); lanes 5 and 6 expressed proteins in supernatant after freeze-thawing (GST-PC wild-type in lane 5 and GST-PCM5 in lane 6), lanes 7 and 8 represent the GST fusion protein bound to GSH-Sepharose 4B (GST-PC wild-type in lane 7 and GST-PCM5 in lane 8. M represents protein molecular weight markers and particular sizes (kDa) and position of the fusion protein (GSTF) on the gel are indicated.

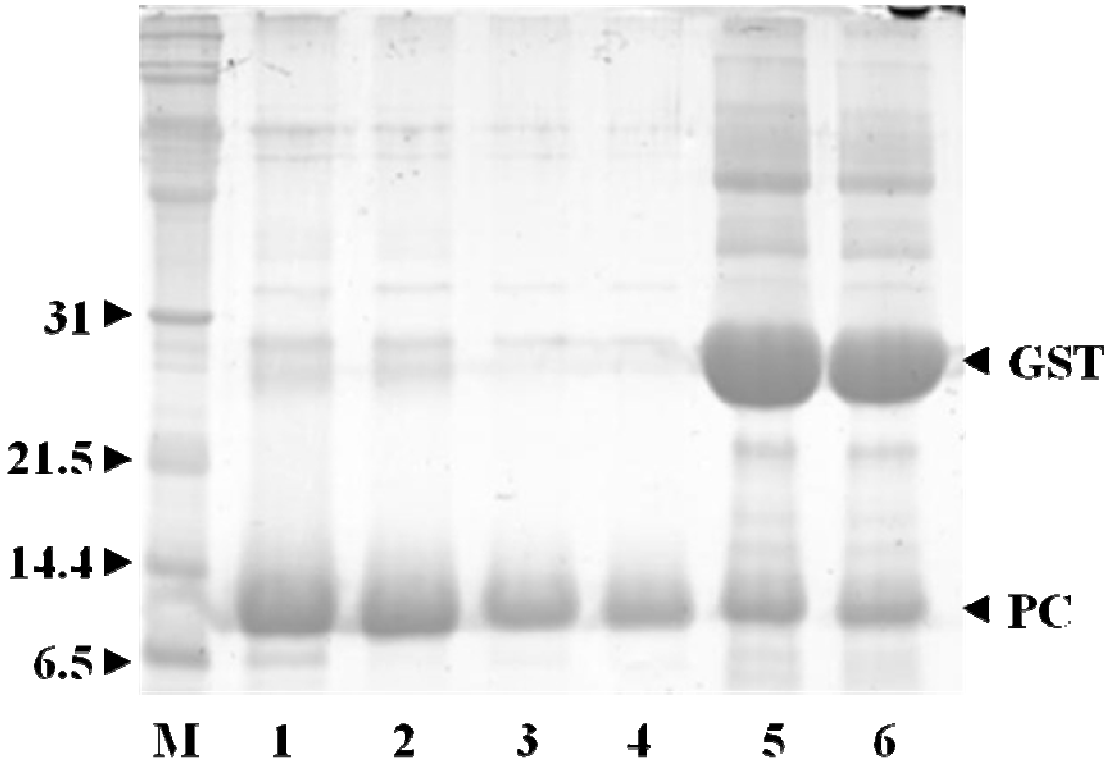

Figure 3.6: Purification of GST-PC by GSH-Sepharose 4B and treatment with Factor Xa. Proteins of different purification steps were analyzed by SDS-PAGE on a 15% gel. Lane 1 represents GST-PC wild-type in the supernatant after Factor Xa treatment; lane 2 GST-PCM5 mutant fusion also in supernatant after Factor Xa. Lane 3 and 4 represent PC (lane 3) and M5 (lane 4) after washing the Sepharose to recover residual cystatins. Lane 5 (PC) and 6 (M5) represent the GST protein still bound to the GSH-Sepharose after Factor Xa treatment. M represents protein molecular weight markers and particular sizes (in kDa) and position of the fusion protein (GSTF) and purified PC on the gel are indicated.

3.2.3.1 *Determination of relative inhibitory potency*

Inhibitory activity of OCI and PC mutants relative to wild-type OCI and PC were determined to show changes in the potency of individual mutated inhibitors to modulate activity against the two model cysteine proteases papain and cathepsin-L. Figure 3.7 shows the relative inhibitory activity of OCI wild-type and OCI mutants against (A) papain with 17.5 ng of papain per 100 µl of assay mixture and (B) with 2.81 ng cathepsin-L per 100 µl of assay mixture. Assays were carried out in 100 mM sodium phosphate buffer, pH 6.5. For papain all inhibitors tested were used at 0.625 ng per test and for cathepsin-L at 0.312 ng per test. Papain activity without inhibitor addition was 64.0± 4.3 FU/min (fluorescence units produced per minute) and cathepsin-L activity was 39.5± 0.7 FU/min. Inhibitory activity of each mutant OCI is shown relative inhibitory activity of wild-type OCI set at 100%. Relative activities greater than 100% correspond to improved and below 100% to decreased inhibitory activity of OCI mutants when compared to wild-type OCI. Data shown for mutants are the mean ± SD from 3 different individual measurements of protease activity in the presence of the inhibitor. Only OCIM1 tested against papain did not show statistically significantly ($P \leq 0.05$) altered the inhibitory potency compared to the wild-type, all other cystatins in all other tests had statistically significant change in inhibitory potency. The OCIM2 was the most active mutant for both papain and cathepsin-L inhibition improving potency by 80% and 50%, respectively, all other OCI mutants had lower inhibitory activity against papain and cathepsin-L when compared to wild-type OCI.

Figure 3.8 shows the relative inhibitory activity of PC wild-type and PC mutants against (A) papain with 17.5 ng of papain per 100 µl of assay mixture and (B) with 2.81 ng of cathepsin-L per 100 µl of assay mixture. Assay was carried out in 100 mM sodium phosphate buffer,

pH 6.5. All inhibitors tested for inhibition of papain activity were used at 1.25 ng per test and at 19 ng per test for cathepsin-L. Papain activity without addition of inhibitor was 64.0± 4.3 FU/min and cathepsin-L activity was 39.7 ± 0.7 FU/min. Inhibitory activity of each mutant PC is shown relative to the activity of wild-type PC, which was set at 100%. Relative inhibitory activities greater than 100% correspond to improved and below 100% to decreased inhibitory activity of mutants when compared to wild-type PC. Data shown for mutants are the mean ± SD from 3 different individual measurements of protease activity in the presence of the inhibitor. The inhibitory activity relative to wild-type PC increased in all PC mutants when papain was used as a cysteine protease activity source with the highest increase found for PCM6 (85%) followed by PCM5 (69%), PCM4 (33%) and PCM8 (23%). Both PCM10 (5%) and PCM12 (7%) showed only slightly improved inhibitory potency compared to the wild-type PC activity. When cathepsin-L was used as a cysteine protease source, PCM6 again, as already found for papain, was the most active mutant improving wild-type PC inhibition by 90% which was followed by PCM4 (83%), and PCM8 (63%). PCM5 only increased wild-type PC inhibitory potency by 14% and again PCM10 and PCM12 did not improve potency, which was also found when papain was used as a cysteine protease source. Only PCM10 and PCM12 cystatins tested against papain did not show statistically significant change in inhibitory potency ($P \leq 0.05$) compared to the wild-type, all other cystatins had statistically significant change in inhibitory potency.

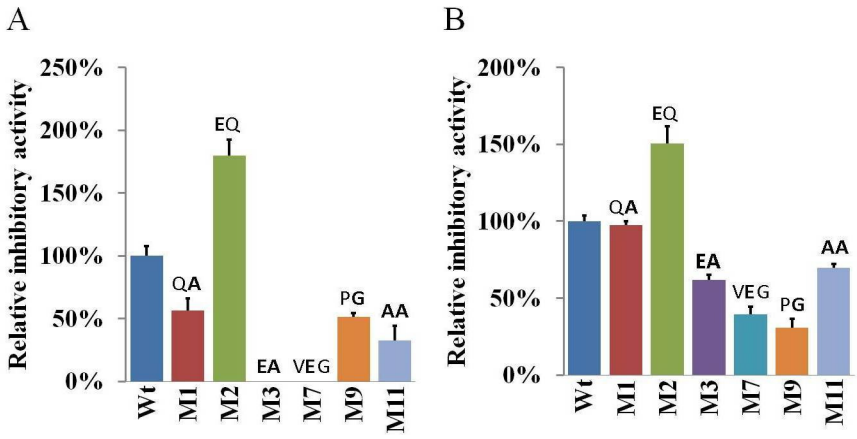

Figure 3.7: Relative inhibitory activity of OCI wild-type and OCI mutants against (A) papain and (B) cathepsin-L. Inhibitory activity of each mutant OCI is shown relative inhibitory activity of wild-type OCI set at 100%. Relative activities greater than 100% correspond to improved and below 100% to decreased inhibitory activity of OCI mutants when compared to wild-type OCI. Data shown for mutants are the mean ± SD from 3 different individual measurements of protease activity in the presence of the inhibitor.

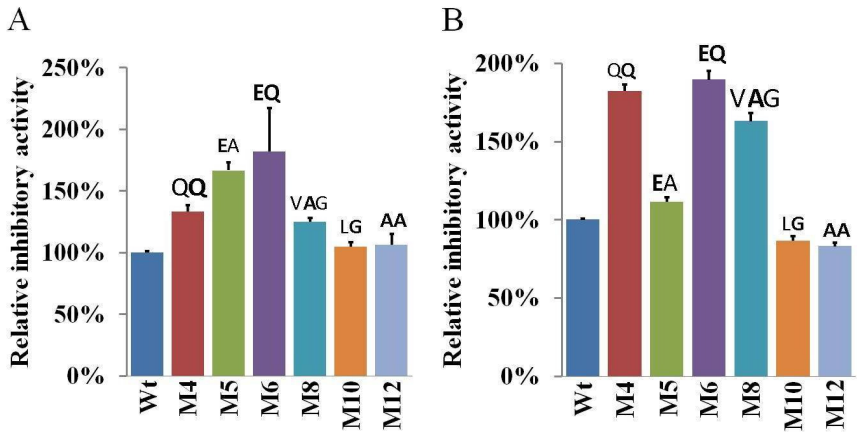

Figure 3.8: Relative inhibitory activity of PC wild-type and PC mutants against (A) papain and (B) cathepsin-L. Inhibitory activity of each mutant PC is shown relative to the activity of wild-type PC, which was set at 100%. Relative inhibitory activities greater than 100% correspond to improved and below 100% to decreased inhibitory activity of mutants when compared to wild-type PC. Data shown for mutants are the mean ± SD from 3 different individual measurements of protease activity in the presence of the inhibitor.

3.2.3.2 *Determination of K_i values*

Table 3.1 shows the *Ki* values for wild-type PC and OCI and PC and OCI mutants using papain and cathepsin-L for cysteine protease activity and Z-Phe-Arg-MCA as a cysteine protease substrate. When papain was used, only OCIM2 (K_i = 0.04 nM) had a greatly lowered K_i value compared to wild-type OCI (K_i = 0.30 nM) and all other OCI mutants did not lower the K_i value of wild-type OCI. All mutant PCs had significantly lowered K_i values compared to wild-type PC (K_i = 5.58 nM) with PCM5 (K_i = 0.47 nM) and PCM6 (K_i = 0.21 nM) the most active and PCM10 (K_i = 4.38 nM) and PCM12 (K_i = 4.04 nM) the least active PC mutant.

When cathepsin-L was used, only OCIM2 (K_i = 0.14 nM) had a significantly lower K_i value compared to wild-type OCI (K_i = 0.49 nM) (Table 3.1). All other OCI mutants had a higher K_i values compared to wild-type OCI with OCIM9 showing no inhibition of cathepsin–L activity. Except for PCM10 and PCM12, which both showed no inhibition of cathepsin-L all other PC mutants had showed lower K_i values compared to wild-type PC (K_i = 152 nM). However, PCM4 (K_i = 3.58 nM) and PCM6 (K_i = 1.95 nM) had the lowest K_i values and therefore the most improved potency against cathepsin-L.

Table 3.1: K_i values for inhibition of papain and cathepsin-L by wild-type and mutant cystatins. K_i values using Z-Phe-Arg-MCA as a cysteine protease substrate. Activity measurements were done in triplicates and mean K_i values are shown.

	Papain		Cathepsin-L	
Name	**K_i (nM)**	**Significance**	**K_i (nM)**	**Significance**
OCI				
Wt	0.30	-	0.49	-
M1 (Q**Q** to Q**A**)	0.54	-	0.53	-
M2 (**Q**Q to **E**Q)	**0.04**	**	**0.14**	*
M3 (**QQ** to **EA**)	1.37	-	1.74	-
M7 (V**A**G to V**E**G)	4.25	-	20.23	-
M9 (P**W** to P**G**)	0.58	-	no inhibition	-
M11 (**GG** to **AA**)	0.74	-	1.23	-
PC				
Wt	5.58	-	152	-
M4 (Q**A** to Q**Q**)	**1.51**	*	**3.58**	**
M5 (**Q**A to **E**A)	**0.47**	**	**70.8**	*
M6 (**QA** to **EQ**)	**0.21**	**	**1.95**	**
M8 (V**E**G to V**A**G)	**1.97**	*	**9.36**	**
M10 (L**W** to L**G**)	4.38	-	no inhibition	-
M12 (**GG** to **AA**)	4.04	-	no inhibition	-

- Reduced or no significant increase in inhibition was observed

* Significant increase in inhibition was observed (≥2-5xfold)

** Highly significant increase in inhibition was observed (>5xfold)

3.2.3.3 *Measurement of PC and OCI activity against plant and insect extracts*

The inhibitory potency of wild-type and mutant OCI and PC was determined against three natural sources of cysteine protease activity. These were a tobacco leaf extract *(N. benthamiana),* a gut extract of banana weevil (*C. sordidus)* larvae (BW), and a gut extract from Colorado potato beetle (*L. decemlineata*) larvae (CPB). Figure 3.9 shows the inhibition of proteolytic activity of 25ng CPB protein extract and inhibition by OCI. Reaction was carried out in 100mM sodium phosphate buffer, pH 6.5, using Z-Phe-Arg-MCA as a cysteine protease substrate. Measurements were done in triplicates and the mean values are shown.by different amounts of OCI at constant amount (CPB) cysteine protease activity.

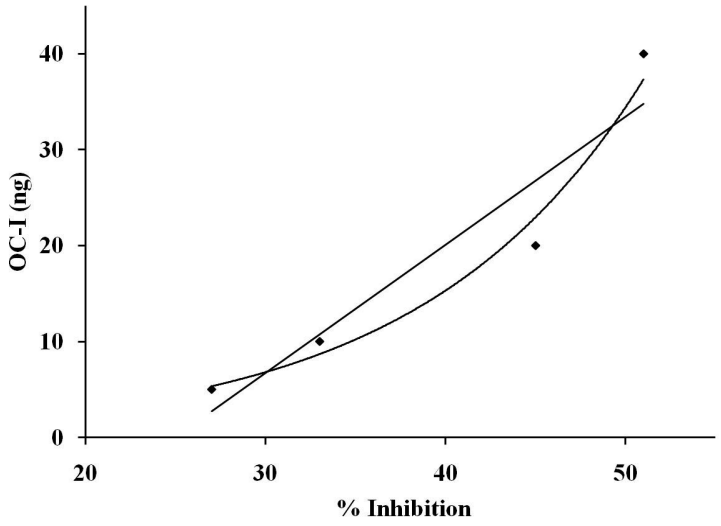

Figure 3.9: The concentration of inhibitor required to reduce the proteolytic activity by 40-60% was determined using either OCI or PC for each of the different extracts, the determined concentration for the wild-type inhibitor was used. Shown is the inhibition proteolytic activity of 25ng CPB protein extract by OCI. Measurements were done in triplicates and only the mean values are shown.

When the different inhibitory potency for wild-type and mutant OCI and PC against cysteine protease activity of different protein extracts (CPB, BW, tobacco leaf) was determined, higher amounts of wild-type or mutant PC had to be used than wild-type or mutant OCI to obtain comparable inhibition of cysteine protease activity (Figures 3.10 and 3.11). Relative inhibitory activity of the wild-type and mutants were tested against (A) 125 ng of CPB protein extract, (B) 5 µg of BW protein extract protein extract and (C) 4.6 µg of tobacco leaf protein extract. Assay mixture (100 µL) contained 100 mM sodium phosphate buffer, pH 6.5. For inhibition of cysteine protease activity of CPB protein extract, 10 ng of wild-type OCI or mutant OCI was used and 100 ng of wild-type PC or mutant PC. For BW protein extract, 400 ng of wild-type OCI or mutant OCI was used and 500 ng of wild-type PC or mutant PC. For tobacco leaf protein extract, 10 ng of wild-type OCI or mutant OCI was used and 50ng of wild-type PC or mutant PC. The cysteine protease activity of CPB protein extract activity without inhibitor addition was measured to be 50.1 ± 1.9 FU/min, for BW protein extract 21.1 ± 0.7 FU/min and for tobacco leaf protein extract 38.8 ± 1.9 FU/min. Inhibitory activity of each mutant OCI is shown relative to the inhibitory activity of wild-type OCI set at 100%. Relative activities greater than 100% correspond to improved and below 100% to decreased inhibitory activity of mutant OCI when compared to wild-type OCI. Data shown for mutants are the mean ± SD from 3 different individual measurements of protease activity in the presence of the inhibitor (Table 3.2; Figures 3.10 and 3.11). Only OCIM1, OCIM2, OCIM3 and OCIM11, tested against the tobacco leaf extract, did not show statistically significantly ($P \leq 0.05$) altered inhibitory potency compared to the wild-type, all other mutant cystatins in other tests had statistically significant change in inhibitory potency. All other OCI mutants did not increase potency of the inhibitor and were less effective than wild-type OCI (Table 3.2; Figure 3.10), only OCIM2 had a slight (6%) increase in inhibition when a CPB extract was used (Table 3.2; Figure 3.10A).

For wild-type PC, no inhibition of CPB cysteine protease activity was found even when 8 µg of PC was used in the assay (Table 3.2). In contrast, PC mutants showed improved the inhibitor potency compared to wild-type PC. CPB cysteine protease activity was inhibited by 40% when 100 ng of mutant PCM4 was used in the assay. Furthermore, using the same of amount of mutant inhibitor, a 61% inhibition was found for PCM8 and 22% inhibition for PCM6. However, no inhibition of CPB cysteine protease activity was found for mutant PCM10 and PCM12 when 100 ng of the inhibitor was used. When a BW extract was used as a cysteine protease activity source, none of the PC mutations greatly improved the potency of the wild-type PC against BW cysteine protease activity and a maximum improvement of 9% was found for PCM8 (Table 3.2 and Figure 3.11A). All other PC mutants performed close to that of the wild-type and PCM10 showing no inhibition at all. When using the tobacco leaf extract, 50 ng of inhibitor was used per test, PCM4 had shown an 8% improvement and PCM8 had shown a 16% improved inhibition against the measured cysteine protease activity in comparison to the wild-type PC (Table 3.2 and Figure 3.11C). Only PCM10 tested against the CPB extract, PCM6 and PCM12 tested against BW extract and PCM6 tested against the tobacco leaf extract did not show statistically significantly ($P \leq 0.05$) altered the inhibitory potency compared to the wild-type, all other cystatins in all other tests had statistically significant change in inhibitory potency.

Table 3.2: Inhibitory potencies of the wild-types OCI and PC and mutants, tested against the cysteine protease of CPB, BW or tobacco leaf. (- indicates that reduced or no significant increase in inhibition was observed ($P \leq 0.05$); * indicates that a significant increase in inhibition was observed [≥2-5xfold])

	Colorado potato beetle (CPB)		***Banana weevil (BW)***		***Tobacco***	
	Inhibition (%)	**Significance**	**Inhibition (%)**	**Significance**	**Inhibition (%)**	**Significance**
OCI						
Wt	38	-	37	-	80	-
M1 (Q**Q** to Q**A**)	15	**-**	11	**-**	76	**-**
M2 (**Q**Q to **E**Q)	44	**-**	27	**-**	81	**-**
M3 (**QQ** to **EA**)	4	**-**	13	**-**	77	**-**
M7 (V**A**G to V**E**G)	12	**-**	15	**-**	48	**-**
M9 (P**W** to P**G**)	14	**-**	20	**-**	54	**-**
M11 (**GG** to **AA**)	16	**-**	25	**-**	78	**-**
PC						
Wt	no inhibition	-	24	-	72	-
M4 (Q**A** to Q**Q**)	40	**-**	30	**-**	80	**-**
M5 (**Q**A to **E**A)	6	**-**	21	**-**	30	**-**
M6 (**QA** to **EQ**)	22	**-**	22	**-**	74	**-**
M8 (V**E**G to V**A**G)	**61**	*	33	**-**	**90**	*
M10 (L**W** to L**G**)	no inhibition	-	no inhibition	-	8	**-**
M12 (**GG** to **AA**)	no inhibition	**-**	21	**-**	23	**-**

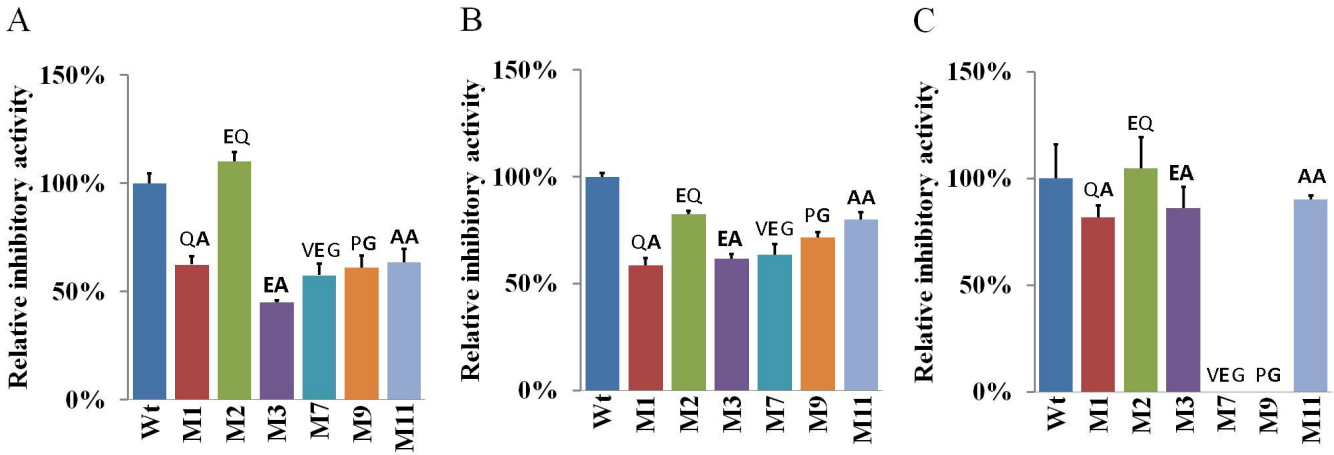

Figure 3.10: Relative inhibitory activity of OCI wild-type and OCI mutants were determined against (A) CPB protein extract, (B) BW protein extract protein and (C) tobacco leaf protein extract. Inhibitory activity of each mutant OCI is shown relative to the inhibitory activity of wild-type OCI set at 100%. Relative activities greater than 100% correspond to improved and below 100% to decreased inhibitory activity of mutant OCI when compared to wild-type OCI. Data shown for mutants are the mean ± SD from 3 different individual measurements of protease activity in the presence of the inhibitor.

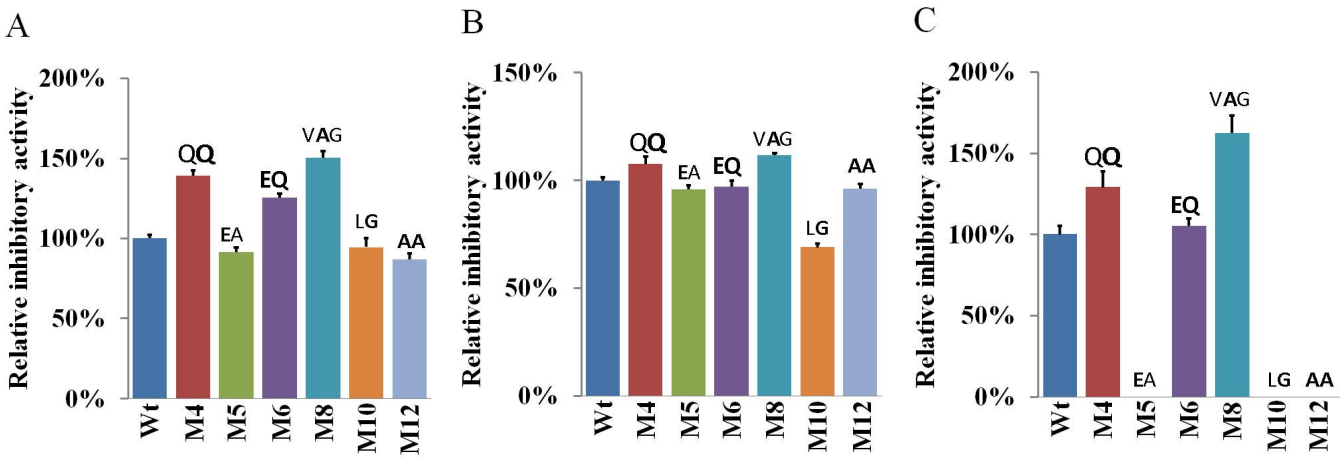

Figure 3.11: Relative inhibitory activity of PC wild-type and PC mutants were determined against (A) CPB protein extract, (B) BW protein extract protein and (C) tobacco leaf protein extract. Inhibitory activity of each mutant PC is shown relative to the inhibitory activity of wild-type PC set at 100%. Relative activities greater than 100% correspond to improved and below 100% to decreased inhibitory activity of mutant PC when compared to wild-type PC. Data shown for mutants are the mean ± SD from 3 different individual measurements of protease activity in the presence of the inhibitor.

3.3 Bioinformatics work

The Discovery Studio v2.5 (Accelrys Software Inc., USA) software package was used to simulate the interaction energies between the model cysteine protease model papain and both wild-type and mutant OCI and PC generated in the study. Figure 3.12 shows the models generated and the positions of the individual mutations are shown in the structural model of both OCI and PC. The amino acid changes in the OCI model include Q53A, Q52E, A56E, W84G, G10A and G11A. The amino acid changes indicated on the PC model include A52Q, Q51E, A55E, W83G, G8A and G9A. The interaction energies were first simulated based on the interaction between the entire models, which calculates the contribution all amino acids in binding and taking into account the spacial arrangement of the protein's structure in secondary and tertiary folding, giving an overall interaction energy value (Table 3.3). The interaction energies were then simulated based on the individual contributions that the mutated amino acids contribute towards the interaction with the papain model and the individual cystatin models, giving the exact value that each amino acid contributes to the binding energy (Table 3.4). The simulations were done to determine similarities to the biochemical data obtained in the study. The interaction energy between the enzyme-inhibitor complex is expected to decrease when a more stable interaction complex (i.e. more favourable) is formed and is expected to increase when a less stable interaction complex (i.e. less favourable) is formed.

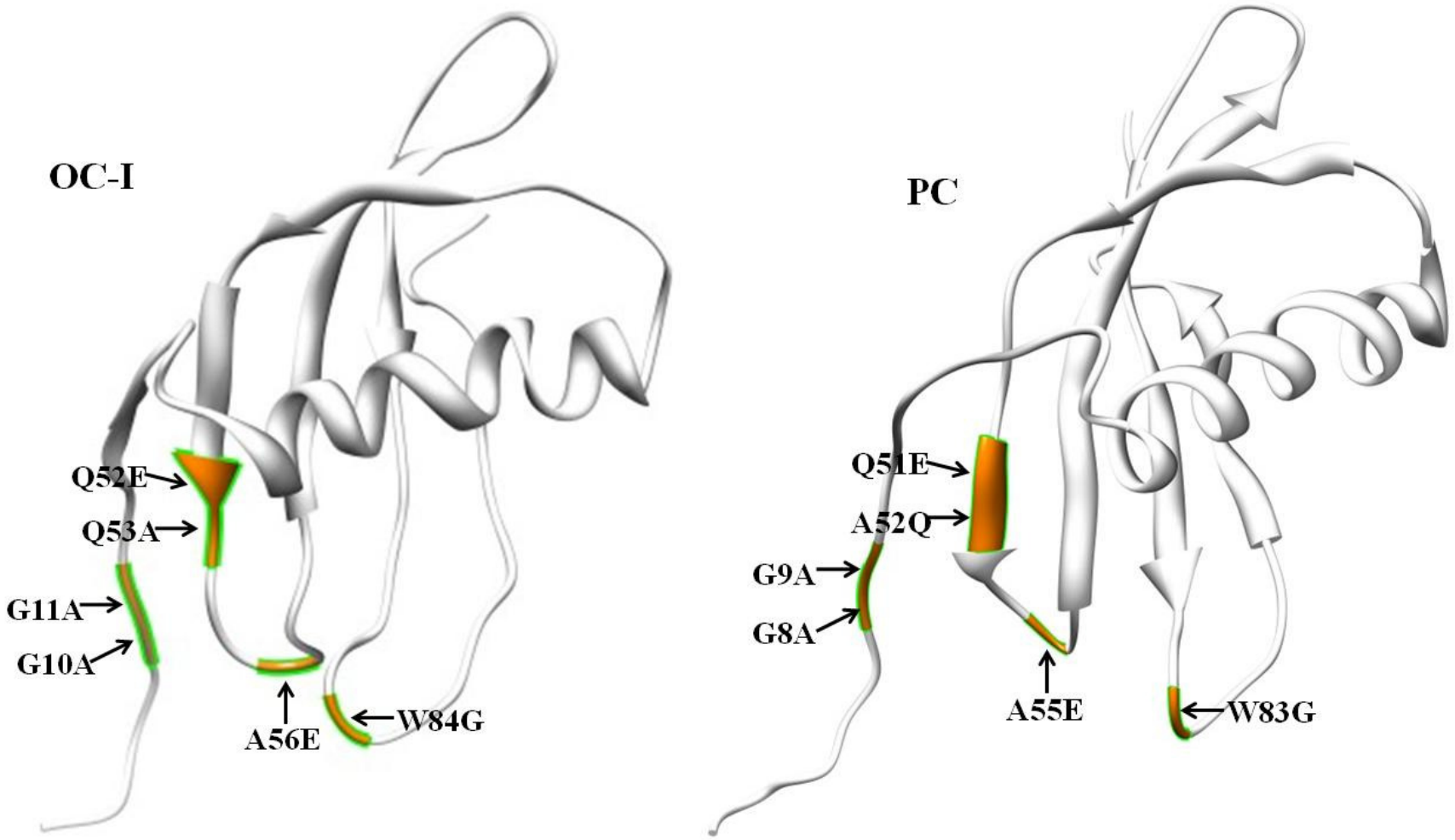

Figure 3.12: The wild-type cystatin models used to simulate the interactions energies are shown, the amino acid positions which were targeted are indicated and the corresponding amino acid changes are shown (UCSF Chimera; Pettersen *et al.,* 2004).

The predicted interaction energies (kcal/mol) calculated between the model cysteine protease papain and the entire OCI or PC inhibitor is shown in Table 3.3. All OCI and PC mutants had shown a decreased predicted total interaction energy and therefore predicts a less stable association complexes being formed between the papain and the cystatins (4.5-11.8%) when compared to wild-type OCI or wild-type PC. The highest decrease in interaction energy, which relates to a reduced complex stability, was found for mutant OCIM7 (11.8%) and PCM8 (11.3%), which contain the mutation A56E for OCI and E55A for PC within the same tertiary position, indicating the importance of this site for binding. However, this reduced complex stability observed for OCIM7 and PCM8, was not directly comparable to the biochemical data obtained. The OCIM2 and PCM6 mutations were predicted to have a destabilizing effect, but from the biochemical data was observed to have the most improved inhibitory activity against papain.

Table 3.4 shows the predicted interaction energy (kcal/mol) based on the interaction between papain and the individually mutated amino acid residues in the OCI and PC. For OCI, increased interaction energies were only predicted for changing the Q (glutamine) at position 52 in the first inhibitory loop to E (glutamic acid) (OCIM2) and changing the conserved Q at position 53 to A (OCIM1). The interaction energy at the site of mutation was predicted to decrease from -15.74 kcal/mol for OCI wild-type to -23.8 kcal/mol for OCIM2, with a further decrease to -27.77 kcal/mol in the double mutant Q52EQ53A (OCIM3)**.** The lowered interaction energy values indicate that the individual amino acids are cooperating to form a more stable interaction complex. The overall change in interaction energy for OCIM3 was determined by adding the individual contributions of the amino acid residues in the wild-type (-44.13 kcal/mol) and compared to that of the mutant (-50.0 kcal/mol). This was directly related to an increased binding ability of OCIM2 to papain, increasing the potency of the

mutant, having a much lower K_i value than wild-type OCI and the lowest K_i value of all OCI mutants (Tables 3.1 and 3.4). All other OCI mutations had a decrease in predicted interaction energy, having a destabilizing effect, also resulted in higher K_i values and lower binding affinity to papain than wild-type OCI. The greatest change of all the OCI mutants, was observed for OCIM7 with an A56E substitution, which had negative interaction energy (-41.93 kcal/mol) and was changed to positive interaction energy with an overall change of (+55.4 kcal/mol), this change coincides with the observed change in K_i. This indicates that the introduced amino acid E is possibly competing during the interaction with an amino acid(s) in the papain molecule, destabilizing the interaction complex.

Similar to OCI, when changing the Q at position 52 in the first inhibitory loop of OCI (OCIM2) to an E and having a conserved Q at position 53 adjacent, had the same amino acid sequence EQ in the motif as PCM6. These mutants had a predicted decrease in interaction energy at the site of change, from -12.24 kcal/mol for PC wild-type to -61.87 kcal/mol for PCM5 and a further increase for the double mutant Q51EA52Q (PCM6) to -65.03 kcal/mol**.** Again, the overall change for PCM6 was determined by adding the individual contributions of the amino acid residues in the wild-type (-39.44 kacl/mol) and comparing to that of the mutant (-90.37 kcal/mol). These two mutants PCM5 and PCM6 had had shown largest decreases in interaction energy at the site of the change of all PC mutants, indicating a stabilizing effect on the interaction complex. The PCM6 and PCM5 had a lowered K_i value of 27-times and 12-times, respectively, compared to the wild-type PC. The predicted interaction energy for the other mutants decreased in comparison to the wild-type PC, but these were observed to have lower K_i values than wild-type PC against papain and cathepsin L.

Table 3.3: Predicted interactions energies (kcal/mol) for each of the mutants. The interaction energies were calculated based on the interaction energies of all the amino acids in the cystatin molecules. TIE: total interaction energy, TVDWE: total van der Waals energy, TEE: total electrostatic energy.

Name	TIE	TVDWE	TEE	% Change from WT
OCI				
Wt	-1170.82	-79.80	-1091.02	-
M1 (Q**Q** to Q**A**)	-1078.85	-80.24	-998.61	+7.9%
M2 (**Q**Q to **E**Q)	-1094.67	-80.35	-1014.32	+6.5%
M3 (**QQ** to **EA**)	-1094.28	-80.29	-1013.99	+6.5%
M7 (V**A**G to V**E**G)	-1032.33	-56.08	-976.24	+11.8%
M9 (P**W** to P**G**)	-1060.23	-73.99	-986.24	+9.4%
M11 (**GG** to **AA**)	-1073.52	-79.01	-994.51	+8.3%
PC				
Wt	-1281.68	-91.49	-1190.19	-
M4 (Q**A** to Q**Q**)	-1172.58	-88.03	-1084.55	+8.5%
M5 (**Q**A to **E**A)	-1224.61	-87.71	-1136.90	+4.5%
M6 (**QA** to **EQ**)	-1223.70	-88.16	-1135.54	+4.5%
M8 (V**E**G to V**A**G)	-1137.49	-85.58	-1051.91	+11.3%
M10 (L**W** to L**G**)	-1155.27	-80.49	-1074.79	+9.9%
M12 (**GG** to **AA**)	-1161.69	-87.92	-1073.77	+9.4%

Table 3.4: The table shows the predicted interactions energy (kcal/mol) of the individual amino acids that were targeted and changed, indicated by the coloured blocks in the respective mutants. The other amino acid residues surrounding the coloured blocks had not shown considerable change and were therefore not shown.

OCI	Mutation	Gly (10)	Gly (11)	Gln (52)	Gln (53)	Ala (56)	Trp (84)
Wt		-27.05	-27.71	-15.74	-28.39	-41.93	-56.15
M1	Q53A				**-22.20**		
M2	Q52E			**-23.80**			
M3	Q52EQ53A			**-27.77**	**-22.23**		
M7	A56E					**+13.47**	
M9	W84G						**-29.01**
M11	G10AG11A	**-14.76**	**-24.27**				

PC	Mutation	Gly (8)	Gly (9)	Gln (51)	Ala (52)	Glu (55)	Trp (83)
Wt		-30.90	-15.17	-12.24	-27.20	-93.02	-59.90
M4	A52Q				**-26.29**		
M5	Q51E			**-61.87**			
M6	Q51EA52Q			**-65.03**	**-25.34**		
M8	E55A					**-43.85**	
M10	W83G						**-35.15**
M12	G8AG9A	**-29.70**	**-0.71**				

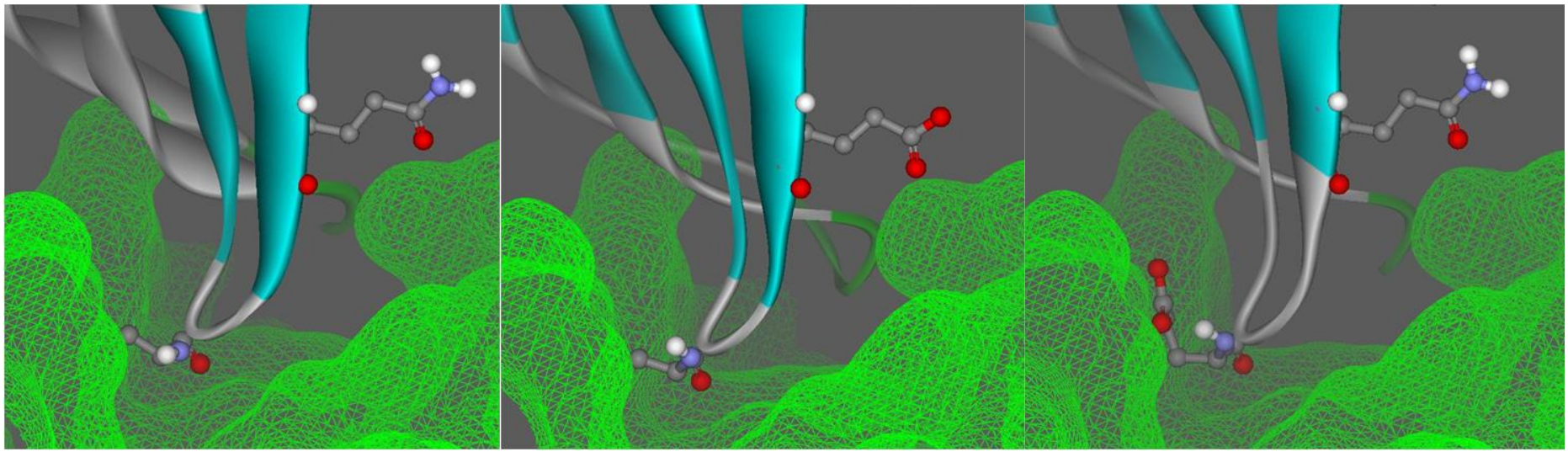

Figure 3.13: The wild-type OCI model was used indicate the potential changes in local conformation of the cystatin when interacting with the papain model. The wild-type model is shown in the picture on the left, with the amino acid residues targeted in the study shown. The change from QQVVAG to **E**QVVAG (OCIM2) is shown in the middle and the change from QQVVAG to QQVV**E**G (OCIM7) is shown right. The altered amino acid side chains can be seen at the site of mutation. The second inhibitory loop can be seen in the background (Discovery studio).

4. DISCUSSION

During the interaction of a cysteine protease enzyme and cystatin, hydrophobic contacts are established between the amino acid residues of the cystatin involved in binding and the corresponding amino acid residues of the enzyme which form the binding pockets. Knowledge of a given cysteine protease enzyme's preferential selection toward amino acids at given tertiary positions in a cystatin, such as cathepsins B/H which are repulsed by large aromatic residues in their S_2 and S_3 pockets, in contrast to cathepsin L which preferentially accommodates larger aromatic residues in the S_2 pocket (Mason *et al.*, 1998), would assist in designing cystatins with enhanced inhibitory potency. Despite the similar mechanism of action and structural homology among cystatins, they naturally exhibit distinct differences in enzyme affinities. Looking at the frequency and occurrence of evolutionary conserved amino acids at a specified tertiary position, such as the first inhibitory binding loop of cystatins, can give an indication of what amino acids are preferentially accommodated and what might be tolerated in the enzymes' binding pockets. Using site-directed mutagenesis allows the relative importance of amino acids at specific tertiary positions to be investigated, without detrimentally altering the structural backbone of the cystatin. Although changing a single amino acid residue alters the complementary shape, interacting charge and polarity of the binding face, it also influences the neighboring amino acids which are dependent on the local changes in amino acid side chains and the resulting functional charges. Previous studies have followed a similar approach to investigate changes in conserved amino acid residues within the N-terminal region and the first (QxVxG) and second (PW) inhibitory loops. However, there is still no consensus based on the different observations found in previous studies on which amino acids contribute to the inhibitory potency of cystatins. The importance of the N-terminal region of cystatins is known to facilitate binding and stabilize the interaction

complex, but contrasting evidence from other studies obscures the importance and involvement of the conserved glycine (G) amino acid residues in this region. The QxVxG motif in the first binding loop of cystatins is conserved among most cystatins with a degree of variance occurring at these variable amino acid sites. In the second inhibitory loop, the W amino acid is known to be highly conserved and important in this position, yet no study has looked at altering this amino acid residue using a mutagenesis approach. Based on the hypervariable sites identified by Kiggundu *et al*. (2006), which could potentially improve the inhibitory activity of cystatins and following a directed mutagenesis approach, the relative importance of specific amino acid residues of cystatins may be elucidated and inhibitors with improved inhibitory potency can be engineered.

This study was carried out to investigate the contribution of individual amino acid residues in the N-terminal and the first and second inhibitory loops of the two plant cystatins, OCI and PC, by using mutant OCI and PC inhibitors. These two cystatins were selected for this study due to their recently identified differences in inhibitory activity against a gut extract from banana weevils containing cysteine protease activity (Kiggundu *et al.*, 2010) and also due to unique sequence characteristics of PC in the conserved motif.

In a preceding study carried out by Dr Schlüter (unpublished data) the amino acid sequence plant cystatins were compared, from this *in silico* analysis, PC was identified to have unique amino acid residue characteristics in the conserved domain of the first inhibitory loop, a unique A (alanine) in position 52, not found in any other cystatins, as well as an E (glutamic acid) in position 55 (Q**A**VV**E**G). The particular sequence variation (E in position 55) has only been identified in three other characterized plant cystatin sequences: *Celosia cristata* cystatin, a multi-cystatin unit of cowpea VuCYS2.3 and barley HvCYS7 cystatin. In this

study it was found that changing the unique A (alanine) in position 52 of PC to Q (glutamine) (mutant PCM4), to obtain a QQ string in the conserved motif (QAVEG to Q**Q**VEG), which is found naturally in many other plant cystatins, resulted in increased inhibitory activity of PC against both papain and cathepsin-L without significantly changing the predicted interaction energies. The change from a non-polar, hydrophobic amino acid residue A to a polar, uncharged residue Q had possibly affected the shape of the binding loop and side-chain polarity resulting in improved PC inhibitory potency. Removing the hydrophobic side chain of A and replacing with the much larger, yet polar and uncharged side chain of Q, had favourably altered the hydrophobicity of the binding face, however, why PC contains a unique A residue and what role this residue has in PC, is still unclear. There is currently no information available if this residue possibly plays a role in PC specificity against a particular cysteine protease of a papaya pest or against a specific endogenous papaya cysteine protease. In general, wild-type PC, isolated from mature papaya leaves, is active against chymopapain and papaya protease IV, which is found in the latex of *Carica papaya* (Taylor *et al.,* 1999; Song *et al.*, 1995). Papaya protease IV however, lacks activity against synthetic protease substrates normally hydrolyzed by other papaya proteases (papain-like) and this protease is also not inhibited by chicken cystatin (Buttle *et al.*, 1989). Since PC potency was found in this study to be much higher against papain (a papaya cysteine protease) than against human cathepsin-L, PC might have a specific role in regulating endogenous papaya cysteine protease activity, rather than a defensive role such as preventing cathepsin-L type cysteine protease activity in the digestive system of herbivorous insects. Such interaction with an endogenous papaya cysteine protease might require a unique amino acid residue(s) in the first binding loop for optimal binding which might be investigated in more detail in a future study.

The importance of Q (a polar, uncharged amino acid) in the binding loop of the tertiary structure for plant cystatin activity was confirmed in this study when Q53 in OCI was mutated to a non-polar, hydrophobic A (mutant OCIM1 - QQVVAG to Q**A**VVAG), likened to the unique A (alanine) found in the first inhibitory loop of PC. This change had significantly decreased OCI activity, as reflected by an increase in the K_i value against papain and cathepsin-L. This was further associated with an increase of +6 kcal/mol in interaction energy with papain from -28 to -22 kcal/mol as predicted *in silico*. This increase in energy will result in a less stable complex. The importance of the Q residue in OCI activity has also been reported by Arai *et al.* (1991) where substitution of Q in OCI with P (proline) (from QQVVAG to Q**P**VVAG) resulted in a completely inactive OCI. In addition, substituting Q with L (leucine) (from QQVVAG to Q**L**VVAG) resulted in a 150-fold increase in K_i, value when compared to wild-type OCI causing a significant reduction in papain inhibition (Arai *et al.,* 1991).

A further new aspect of this study was the finding that substituting E with A in position 55 of PC (PCM8) (from AVVEG to AVV**A**G) located in the first binding loop significantly increased PC potency against both papain and cathepsin-L. Most other plant cystatins have either this non-polar, neutral A or a polar, neutral S (serine) in this position (Table 1.2) whereas E (glutamic acid) of PC in position 55 of PC is only found in three other plant cystatins. However, the exact role of a polar, negatively charged E in this position is still unknown. Since this amino acid residue is situated in the middle of the first inhibitory binding loop inserting into the cleft of the target protease, mutations in this position might significantly affect the binding capacity of PC. The alanine amino acid in this position is considered to be a relatively small amino acid in comparison to the bulky E with a longer side-chain. Also, changing the charge of the amino acid residue from a polar, negatively

charged E to a non-polar, neutral A will greatly influence the inhibitor's binding ability. The introduced amino acid change had affected the shape, size and charge of the resulting binding loop and interaction with the given enzyme, which is dependent on the space available and charge of the accepting binding cavity. The results observed within this study clearly demonstrated that a change from E to A (PCM8) significantly increases PC potency against the two model cysteine proteases papain and cathepsin-L by replacing the bulky E with the smaller A. However, interaction energies were predicted to increase *in silico* by +49 kcal/mol against the papain model (from -93 to -43 kcal/mol) resulting in a less stable interaction complex being formed. However, this was done only using a papain model and will probably have different observations when using a cathepsin-L model. This was further correlated to an increase in predicted interaction energies of all amino acids in the inhibitor molecule. The importance of A in this position in plant cystatin inhibitory activity against papain and cathepsin-L protease activity was further confirmed by the dramatic decrease in OCI (OCM7) activity against the two tested cysteine proteases (papain and cathepsin-L) and a large increase in binding energy of +55 kcal/mol with papain from -42 to +13 kcal/mol as predicted *in silico*. The introduction of the larger, bulky E in the place of the smaller A at this position of OCI had interfered with the shape and charge of the binding loop, negatively affecting the interaction with the target enzyme's binding cavity resulting in a less stable interaction complex being formed. The addition of an electron to the binding interface of OCI, where there was none, may now be competing with an electron in the binding pocket of papain, causing the electrons to repel each other. This excess energy destabilizes the interaction complex.

A very interesting new result of this study was the superior functionality of the amino acid residue string EQ in both PC positions 51 and 52 (QAVVEG to **EQ**VVEG) and OCI

positions 52 and 53 (QQVVAG to **E**QVVAG), when compared to other mutated strings tested at these positions. Mutating to the EQ string caused the greatest change in inhibitory activity of the two inhibitors against both papain and cathepsin-L. It also greatly decreased predicted interaction energies, a stronger overall change in PC of -51 kcal/mol, from -39 to -90 kcal/mol, and smaller overall change in OCI with -6 kcal/mol from -44 to -50 kcal/mol. The reduced predicted interaction energies indicate that a more stable interaction complex is formed (i.e. more favourable). The Q (PC position 51; OCI position 52) is located on the beta-sheet of the protein and is not part of the binding loop, but based on the modelling with papain, bearing in mind that this could be different against a cathepsin-L model, this amino acid is identified as it is not directly involved in the interaction, so it is possible that substituting this amino acid had changed the local conformation of neighbouring amino acids which are indeed involved in the binding interaction. Furthermore, this substitution with a polar negatively charged E increases the polarity of the surrounding amino acids and also increases the electrostatic strength of interaction with the target enzyme. Interestingly, the Q residue located on the beta-sheet of the protein has also been previously identified as a positively selected codon site (Kiggundu *et al.,* 2006). Such sites possibly evolve under selective pressure in response to challenges such as insect attack, implicating the diversification of digestive proteases and providing a selective advantage to the host (Tiffin and Gaut, 2001; Barbour *et al.,* 2002; Lopes *et al.,* 2004; Goulet *et al.*, 2008). However, the superiority of this EQ string in both plant cystatins could not be confirmed when either beetle extracts or a tobacco plant extract was used as a source of cysteine proteases. This might be because of the presence of various different forms of cysteine proteases in extracts, some of which the inhibitory potency of the mutant inhibitor(s) might be increased coupled with decreased activity against some of the other protease forms present.

In general, mutating PC in the first binding loop to obtain an amino acid sequence similar to OCI had significantly improved the activity of PC when compared to wild-type PC. This was particularly evident against cysteine protease activity from a Colorado potato beetle extract and to a certain degree against cysteine protease activity from tobacco (*N. benthamiana)* leaf extracts. Amongst all PC mutants tested, PCM8 with the E to A substitution at position 55 had the highest activity against the potato beetle cysteine protease activity. In contrast, most OCI mutations had caused a decrease in inhibitory potency against protease activity in the various extracts tested. Overall, this result has provided further evidence that cysteine protease targets are seemingly different for OCI and PC. OCI very likely inhibits papain and cathepsin-L type cysteine protease activities found in insect pests, such as the potato beetle, as shown by the low amount of OCI needed, whereas to achieve a comparable inhibitory activity similar to OCI, PC had to be used at a 10 times higher concentration. In all test systems used, OCI was the better inhibitor compared to PC, however, by changing amino acid residues in first inhibitory loop of the relatively "weak" PC to amino acid residues found in OCI caused a significant improvement in inhibitory potency of PC against cysteine protease activity in the potato beetle extract.

The observed *in vitro* and predicted *in silico* results when using the entire cystatin molecule to calculate the interaction energy (Table 3.3) did not correlate. Although only a papain model was used in the analysis, different predicted interaction energies would be expected when used in conjunction with e.g. a cathepsin-L model as well. The interaction energies for all mutant cystatins were predicted to increase, which would lead to a less stable enzyme-inhibitor complex being formed (i.e. less favourable), yet several cystatins were shown to have enhanced binding capabilities. The observed change was correctly predicted when only looking at the contribution of the individual amino acids involved in the binding interaction

(Table 3.4). It is possible that additional algorithmic calculations to determine the spatial interaction of all the amino acids in the whole cystatin molecule caused obscured interaction energy to be predicted for all mutants.

Several amino acid substitutions were able to alter the binding capacities of the mutants by just one amino acid substitution, e.g. OCIM2, PCM4, PCM5 and PCM8. In previous studies and in this one, wild-type OCI is shown to have naturally high inhibitory activity against both plant and insect cysteine proteases activity. Activity against endogenous proteases is considered a "house-keeping" function and is considered to be conserved, as appose to activity against exogenous proteases which is considered to be more divergent. A functional interplay between one or several cystatins, with varying degree of activity against different proteases (endogenous or exogenous) are responsible for the regulation of everyday plant metabolism and are involved in elicited responses against attack. The OCI mutant M2 improved inhibitory potency, the introduced amino acid variation is frequently found in other cystatins (42%) and is thought to allow better interaction with certain sub-set of cysteine proteases (papain and cathepsin-L), the same observations were not seen against the protein extracts. For the mutants of wild-type PC, changing its unique and unusual amino acid variations within the first binding loop, improved the inhibitory potency dramatically. It is speculated that these amino acid variations in PC are conserved due to their functional significance in regulating an endogenous protease in papaya, or possibly an unknown exogenous target and changing these to the more frequently occurring amino acid variations, such as in OCI, changes the cystatin's target specificity likened to OCI.

Finally, this study has also confirmed the importance of the amino acid residue W (tryptophan), which is the most conserved amino acid in the second inhibitory loop and

present in the majority (93%) of plant cystatins (Table 1.2; Koiwa *et al.,* 2001) and the G (glycine) residue(s) in the N-terminal region of plant cystatins. Mutating the non-polar, neutral amino acid residue W, containing a hydrophobic side chain to a non-polar, neutral G had drastically decreased inhibitory activity against both papain and cathepsin-L, as well as the cysteine protease-containing extracts. This substitution strongly influences the side-chain polarity and hydrophobicity of a cystatin affecting both the overall secondary and tertiary structure at this site. This influences the binding ability of the molecule which is then directly related to decreases in both the *in silico* predicted interaction energies and inhibitory activity against papain and cathepsin-L or even results in no inhibition at all. The N-terminal is known to be critical in facilitating the binding interaction stabilizing the plant cystatin interaction with the active site by associating with the surface of the protease (Bode *et al.,* 1988; Turk and Bode, 1991). All known plant cystatins contain at least one G in the N-terminal region, but a previous study had demonstrated that substitutions of the conserved G at position 10 could enhance the inhibitory activity of OCI and since G in position 11 lacks a side-chain, it forms a local confirmation in the N-terminal region allowing for both high-affinity binding and efficient inhibition (Urwin *et al.*, 1995a).

In summary this study has provided new insight regarding the importance of single amino acid residues in the different conserved regions of plant cystatins through amino acid mutation. These changes did not greatly affect the overall structure of plant cystatins. This study has further shown that it is possible to improve the binding capacity of plant cystatins to papain and cathepsin-L, as well as to a potato beetle extract, by mutating individual amino acid residues in the first binding loop of the relatively “weak” papaya cystatin. In particular, mutation to the amino acid string EQ has been found to be superior for inhibitory potency in both plant cystatins, with E located on the beta-sheet of the cystatin and not forming part of

the first binding loop, when papain and cathepsin-L were used as a source of cysteine protease activity. In future studies, this potent loop mutation may be used in conjunction with the approach followed by Goulet *et al.* (2008) to identify positively selected amino acid sites in the N-terminal region of the two plant cystatins, to further improve plant cystatin inhibitory potency. Also, future strategies may include completely changing the first binding loop of PC to completely resemble the OCI binding loop and *vice versa* in order to determine how much the first binding loop contributes to the overall plant cystatin inhibitory potency.

Future research may include determination of predicted binding energies for all natural amino acid variations with papain to identify natural substitutions that significantly increase predicted binding energies. In particular for PC, the function of the unique A in the first binding loop should be investigated in more detail by comparing the inhibitory potency of PC mutations against cysteine protease activities from various papaya extracts, such as latex or mature leaves, in order to determine any superiority of PC compared to OCI. In addition, more detailed knowledge about the type and composition of cysteine proteases in beetle and plant extracts will be required to design more potent inhibitors tailored to improve inhibitor potency against complex cysteine protease systems. Knowing the recombinant cystatin's inhibitory range, which proteases are inhibited and knowing what changes in amino acid sequence would allow targeted specificity to certain proteases, optimizing and directing the inhibitory potency of the cystatin, would not only allow the enhanced control of selective proteases, but from an ecological point of view, would allow specific pests to be targeted while minimizing the effects on complex multitrophic interactions within the ecosystem. From a physiological point of view, improving the inhibitory potency of a cystatin and directing its activity, would aid to minimize the interference of the cystatin in other metabolic processes (e.g. defence protein induction, programmed cell death, response to abiotic/biotic

stress, etc.) when used in recombinant applications. Furthermore, the interaction between the protease and inhibitor is dose-dependent, improved cystatin variants will aid to achieve the desired protective effects in whatever biotechnological application, while minimizing the input requirements. Finally, “synthetic” model plant cystatins could be engineered with greatly improved inhibitory potency against a specific cysteine protease or cysteine protease class. The same approach would be possible for other protease families.

5. LITERATURE CITED

Abe, M., Domoto, C., Watanabe, H., Abe, K., and Arai, S. (1996) Structural organization of the gene encoding corn cystatin. Bioscience Biotechnology and Biochemistry. 60: 1173-1175.

Abe M., Abe K., Domoto C., Arai S. (1995) Two distinct species of corn cystatin in corn kernels. Bioscience Biotechnology and Biochemistry. 59: 756-758

Abe, K., Emori, Y., Kondo, H., Arai, S., and Suzuki K. (1988). The NH_2-terminal 21 amino acid residues are not essential for the papain-inhibitory activity of oryzacystatin, a member of the cystatin superfamily. Journal of Biological Chemistry. 263: 7655-7659.

Abraham, Z., Martinez, M., Carbonero, P., Diaz, I. (2006). Structural and functional diversity within the cystatin gene family of *Hordeum vulgare*. Journal of Experimental Botany. 15: 4245-4255.

Aguiar J.M., Franco O.L., Rigden D.J., Bloch Jr C., Monteiro A.C.S., Flores V.M.Q., Jacinto T., Xavier-Filho J., Oliveira A.E.A., Grossi-de-Sa M.F., Fernandes K.V.S. (2006) Molecular modelling and inhibitory activity of cowpea cystatin against bean bruchid pests. Proteins: Structure, Function, and Bioinformatics. 63: 662–670.

Arai S., Watanabe, H., Kondo, H., Emori, Y., and Abe, K. (1991). Papain-inhibitory activity of oryzacystatin, a rice seed proteinases inhibitor, depends on the central Gln-Val-Val-Ala-

Gly region conserved among cystatin superfamily members. Journal of Biochemistry. 109: 294-298.

Auerswald, E.A., Nägler, D.K., Gross, S., Assfai.G-Machleidt, I., Stubbs, M.T., Eckerskorn, C., Machleidt, W., and Fritz, H. (1996). Hybrids of chicken cystatin with human kininogen domain 2 sequences exhibit novel inhibition of calpain, improved inhibition of actinidin and impaired inhibition of papain, cathepsin L and cathepsin B. European Journal of Biochemistry. 23: 534-542.

Barbour, K.W., Goodwin, R.L., Guillonneau, F., Wang, Y., Baumann, H., and Berger, F.G. (2002). Functional diversification during evolution of the murine a1-proteinase inhibitor family: role of the hypervariable reactive center loop. Molecular Biology and Evolution. 19: 718–727.

Barrett, A.J., and Kirschke, H. (1981). Cathepsin B, Cathepsin H and Cathepsin L. Methods in Enzymology. 80: 535-561.

Belenghi, B., Acconcia, F., Trovato, M., Perazzolli, M., Bocedi, S., Polticelli, F., Ascenzi, P., and Delledonne, M. (2003). AtCYS1, a cystatin from *Arabidopsis thaliana*, suppresses hypersensitive cell death. European Journal of Biochemistry. 270: 2593-2604.

Benchabane, M., Schlüter, U., Vorster J., Goulet, M.C., and Michaud D. (2010). Plant cystatins. Biochimie. 92: 1657-1666.

Benchabane, M., Rivard, D., Girard, C., Michaud, D. (2009). Companion protease inhibitors to protect recombinant proteins in transgenic plant extracts. Methods in Molecular Biology. 483: 265-273.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988). The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. The EMBO Journal. 7: 2593-2599.

Botella, M.A., Xu, Y., Prabha, T.N., Zhao, Y., Narasimhan, M.L., Wilson, K.A., Nielsen, S.S, Bressan, R.A., and Hasegawa, P.M. (1996). Differential expression of soybean cysteine proteinase inhibitor genes during development and in response to wounding and methyl jasmonate. Plant Physiology. 112: 1201-1210.

Brown, W.M., and Dziegielewska K.M. (1997). Friends and relations of the cystatin superfamily - new members and their evolution. Protein Science. 6: 5-12.

Buttle, D.J., Kembhavi, A.A., Sharp, S.L., Shute, R.E., Rich, D.H., and Barrett, A.J. (1989). Affinity purification of the novel cysteine proteinase papaya proteinase IV, and papain from papaya latex. Journal of Biochemistry. 261: 469–476.

Ceci, L.R., Volpicella, M., Rahbé, Y., Gallerani, R., Beekwilder, J., and Jongsma, M.A. (2003). Selection by phage display of a variant mustard trypsin inhibitor toxic against aphids. Plant Journal. 33: 557-566.

Colella, R., Sakaguchi, Y., Nagase, H., and Bird, J.W.C. (1989). Chicken egg white cystatin; Molecular cloning, nucleotide sequence and tissue distribution. The Journal of Biological Chemistry. 264: 17164-17169.

Demirevska, K., Simova-Stoilova, L., Fedina, I., Georgieva, K., and Kunert., K. (2009). Response of Transformed with Oryzacystatin I Tobacco Plants to Drought, Heat and Light Stress. Journal of Agronomy and Crop Science. 196: 90-99.

Doi-Kawano, K., Kouzuma, Y., Yamasaki, N., and Kimura, M. (1998). Molecular cloning, functional expression, and mutagenesis of cDNA encoding a cysteine proteinase inhibitor from sunflower seeds. Journal of Biochemistry. 124: 911-916.

Dubin, G. (2005). Proteinaceous cysteine protease inhibitors. Cellular and Molecular Life Sciences. 62: 653-669.

Fan, S.G., and Wu, G.J. (2005). Characteristics of plant proteinases inhibitors and their applications in combating phytophagous insects. Botanical Bulletin of Academia Sinica. 46: 273-292.

Fermentas Life Sciences, Canada. The Fermentas DoubleDigest™ Engine (www.fermentas.com/doubledigest)

García-Carreño, F.L., An, H., Haard, N.F. (2000). Protease inhibitors in food processing. pg. 215-223 in: Michaud D. (Ed.), Recombinant Protease Inhibitors in Plants. Landes Bioscience/Eurekah.com, Georgetown, TX, USA.

Girard, C., Rivard, D., Kiggundu, A., Kunert, K., Gleddie, S.C., Cloutier, C., and Michaud, D. (2006). A multicomponent, elicitor-inducible cystatin complex in tomato, *Solanum lycopersicum*. New Phytologist. 173: 841-851.

Goulet, C., Benchabane, M., Anguenot, R., Brunelle, F., Khalf, M., Michaud, D. (2010). A companion protease inhibitor for the protection of cytosol-targeted recombinant proteins in plants. Journal of Plant Biotechnology. 8: 142-154.

Goulet, M.C., Dallaire, C., Vaillancourt, L.P., Khalf, M., Badri, A.M., Preradov, A., Duceppe, M.O., Goulet, C., Cloutier, C., and Michaud, D. (2008). Tailoring the specificity of a plant cystatin toward herbivorous insect digestive cysteine proteases by single mutations at positively selected amino acid sites. Plant Physiology. 146: 1010–1019.

Grudkowska, M., and Zagdańska, B. (2004). Multifunctional role of plant cysteine proteinases. Acta Biochimica Polonica. 51: 609-624.

Hall, A., Dalbøge, H., Grubb, A. and Abrahamson, M. (1993). Importance of the evolutionarily conserved glycine residue in the N-terminal region of human cystatin C (Gly-11) for cysteine endopeptidase inhibition. Journal of Biochemistry. 291: 123-129.

Henderson, P.J.F. (1972). A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors. Journal of Biochemistry. 127: 321–333.

Inoue, H., Nojima, H., and Okayama H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene. 96, 23-28.

International Union of Biochemistry and Molecular Biology (IUBMB), Enzyme Nomenclature (2010): www.chem.qmul.ac.uk/iubmb/enzyme

Joshi, B.N., Sainani, N., Bastawade, K.B., Deshpande, V.V., Gupta, V.S., and Ranjekar, P.K. (1999). Pearl millet cysteine proteinase inhibitor. Evidence for the presence of two distinct sites responsible for antifungal and anti-feedent activities. European Journal of Biochemistry. 265: 556-563.

Kiggundu, A., Muchwezi, J., Van der Vyver, C., Viljoen, A., Vorster, J., Schlüter, U., Kunert, K., and Michaud, D. (2010). Deleterious effects of plant cystatins against the banana weevil *Cosmopolites sordidus*. Archives of Insect Biochemistry and Physiology. 73: 87–105.

Kiggundu, A., Goulet, M.C., Goulet, C., Dubé, J.F., Rivard, D., Benchabane, M., Pépin, G., Van der Vyver, C., Kunert, K., Michaud, D. (2006). Modulating the proteinase inhibitory profile of a plant cystatin by single mutations at positively selected amino acid sites. The Plant Journal. 48: 403-413.

Koiwa, H., D'Urzo, M.P., Assfalg-Machleidt, I., Zhu-Salzman, K., Shade, R.E., An, H.J., Murdock, L.L., Machleidt, W., Bressan, R.A., and Hasegawa, PM. (2001). Phage display selection of hairpin loop soyacystatin variants that mediate high affinity inhibition of a cysteine proteinase. The Plant Journal. 27: 383-391.

Kondo, H., Abe, K., Emori, Y., and Arai, S. (1991). Gene organization of oryzacystatin-II, a new cystatin superfamily member of plant origin, is closely related to that of oryzacystatin-I but different from those of animal cystatins. FEBS Letters. 278: 87-90.

Križaj, I., Turk, B., and Turk, V. (1992). The complete primary structure of bovine stefin B. FEBS Letters. 2: 237 – 239.

Laboissiere, M.C., Young, M.M., Pinho, R.G., Todd, S., Fletterick, R.J., Kuntz, I., Craik, C.S. (2002). Computer-assisted mutagenesis of ecotin to engineer its secondary binding site for urokinase inhibition. Journal of Biological Chemistry. 277: 26623-26631.

Lopes, A.R., Juliano, M.A., Juliano, L., and Terra, W.R. (2004). Coevolution of insect trypsins and inhibitors. Archives of Insect Biochemistry and Physiology: 55: 140–152.

Margis, R., Reis, E.M., Villeret, V. (1998). Structural and phylogenetic relationships among plant and animal cystatins. Archives of Biochemistry and Biophysics. 359: 24-30.

Martinez, M., Cambra, I., Carrillo, L., Diaz-Mendoza, M., and Diaz, I. (2009). Characterization of the entire cystatin gene family in barley and their target cathepsin L-like cysteine-proteases, partners in the hordein mobilization during seed germination. Plant Physiology. 151: 1531-1545.

Martinez, M., Abraham, Z., Gambardella, M., Echaide, M., Carbonero, P., Diaz, I. (2005). The strawberry gene *Cyf1* encodes a phytocystatin with antifungal properties. Journal of Experimental Botany. 56: 1821-1829.

Mason, R.W., Sol-Church, K., and Abrahamson, M. (1998). Amino acid substitutions in the N-terminal segment of cystatin C create selective protein inhibitors of lysosomal cysteine proteinases. Journal of Biochemistry. 330: 833-838.

Mason, R.W., Green, G.D.J., and Barrett, A.J. (1985). Human liver cathepsin-L. Journal of Biochemistry. 226: 233-241.

Massonneau, A., Condamine, P., Wisniewski, J.P., Zivy, M., and Rogowsky, P.M. (2005). Maize cystatins respond to developmental cues, cold stress and drought. Biochimica et Biophysica Acta. 1729:186-199.

Michaud, D., Cantin, L., and Vrain, T.C. (1995). Carboxy-terminal truncation of oryzacysatin-II by oryzacytatin-insensitive insect digestive proteinases. Archives of Biochemistry & Biophysics. 322: 469-474.

Michaud, D., Nguyen-Quoc, B., and Yelle, S. (1993). Selective inhibition of Colorado potato beetle cathepsin H by oryzacystatins I and II. FEBS Letters. 331: 173-176.

Mosolov, V.V., and Valueva, T.A. (2005). Proteinase inhibitors and their function in plants: a review. Applied Biochemistry and Microbiology. 41: 227-246.

Munger, A., Goulet, C., Vaillancourt, L.P., Schlüter, U., Kiggundu, A., Kunert, K., Goulet, M.C., Michaud, D. (2011). Constitutive expression of endogenous defense proteins in transgenic potato lines expressing the Cys protease inhibitor corn cystatin II. Aspects of Applied Biology. 96 (in press).

National Centre for Biotechnology Information (NCBI). (2010): www.ncbi.nlm.nih.gov

Nikawa, T., Towatari, T., Ike, Y., and Karunuma N. (1989). Studies on the reactive site of the cystatin superfamily using recombinant cystatin A mutants: Evidence that the QVVAG region is not essential for cysteine proteinases inhibitory activities. FEBS Letters. 255: 309-314.

Nissen, M.S., Kumar, M.G.N, Youn, B., Knowles, B.D., Lam, K.S. Ballinger, W.J., Knowles, N.R., and Kang, C.H. (2009). Characterization of *Solanum tuberosum* multicystatin and its structural comparison with other cystatins. The Plant Cell. 21: 861-875.

Ogawa, M., Nakamura, S., Scaman, C.H., Jing, H., Kitts, D.D., Dou, J., Nakai, S. (2002). Enhancement of proteinase inhibitory activity of recombinant human cystatin C using random-centroid optimization. Biochimica et Biophysica Acta. 1599: 115-124.

Oliveira, A.S., Xavier-Filho, J., and Sales, M.P. (2003). Cysteine proteinases and cystatins. Brazilian Archives of Biology and Technology. 46: 91-104.

Pavlova, A., and Björk, I. (2003). Grafting of features of cystatins C or B into the N-terminal region or second binding loop of cystatin A (stefin A) substantially enhances inhibition of cysteine proteinases. Biochemistry. 42: 11326-11333.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry. 25: 1605-1612.

Promega Corporation, USA. (2007). Technical Manual: pGEM®-T and pGEM®-T Easy Vector Systems. Fig. 1 and Fig. 3, pg. 2, 4.

Reis, M.E., and Margis, R. (2001). Sugarcane phytocystatins: Identification, classification and expression pattern analysis. Genetics and Molecular Biology. 24: 291-296.

Rivard, D., Anguenot, R., Brunelle, F., Le, V.Q., Vézina, L.P., Trépanier, S., and Michaud, D. (2006). An in-built proteinase inhibitor system for the protection of recombinant proteins recovered from transgenic plants. Journal of Plant Biotechnology. 4: 359-368.

Salvesen, G., and Nagase, H. (1989). Inhibition of proteolytic enzymes. In Bond, J.S., and Beynon, R., eds., Proteolytic Enzymes. A Practical Approach. IRL Press, New York, pp 83–104.

Schlüter, U., Benchabane, M., Munger, A., Kiggundu, A., Vorster, J., Goulet, M.C., Cloutier, C., and Michaud, D. (2010). Recombinant protease inhibitors for herbivore pest control: a multitrophic perspective. Journal of Experimental Botany. 61: 4169-4183.

Sigma Aldrich, USA. (2006). Competent Cell Compendium: Tools and Tips for Successful Transformations: www.sigma-aldrich.com

Solomon, M., Belenghi, B., Delledonne, M., Menachem, E., and Levine, A. (1999). The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. Plant Cell. 11: 431-444.

Song, I., Taylor, M., Baker, K., Bateman Jr., R.C. (1995). Inhibition of cysteine proteinases by *Carica papaya* cystatin produced in *Escherichia coli*. Gene. 162: 221-224.

Stoop, A.A. and Craik, C.S. (2003). Engineering of a macromolecular scaffold to develop specific protease inhibitors. Nature Biotechnology. 21: 1063-1068.

Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990). The refined 2.4A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinases inhibitor interaction. The EMBO Journal. 9: 1939-1947.

Taylor, M.A.J., Al-sheikh, M., Revell, D.F., Sumner, I.G., and Connerton, I.F. (1999). cDNA cloning and expression of *Carica papaya* prochymopapain isoforms in *Escherichia coli.* Plant Science: 145: 41-47.

Tiffin, P., and Gaut, B.S. (2001). Molecular evolution of the wound induced serine protease inhibitor *wip1* in *Zea* and related genera. Molecular Biology and Evolution. 18: 2092–2101.

Turk, V., and Bode, W. (1991). The cystatins: protein inhibitors of cysteine proteinases. FEBS Letters. 285: 213-219.

Urwin, P.E., Atkinson, H.J., and McPherson, M.J. (1995a). Involvement of the NH_2-terminal region of orzyacystatin-I in cysteine proteinase inhibition. Protein Engineering. 8: 1303-1307.

Urwin, P.E., Atkinson, H.J., Waller, D.A., and McPherson, M.J. (1995b). Engineered oryzacystatin-expressed in transgenic hairy roots confers resistance to *Globodera pallida*. The Plant Journal. 8: 121-131.

UCSF Chimera package (v1.4) from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco: http://www.cgl.ucsf.edu/chimera

Van der Vyver, C., Schneidereit, J., Driscoll, S., Turner, J., Kunert, K., Foyer, C.H. (2003). Oryzacystatin I expression in transformed tobacco produces a conditional growth phenotype and enhances chilling tolerance. Journal of Plant Biotechnology. 1: 101-112.

Zhao, Y., Botella, M.A., Subramanian, L., Niu, X., Nielsen, S.S., Bressan, R.A., and Hasegawa, P.M. (1996). Two wound-inducible soybean cysteine proteinases inhibitors have greater insect digestive proteinases inhibitory activities than a constitutive homolog. Plant Physiology. 111: 1299-1306.