



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

**Engineering plant cysteine protease inhibitors for the transgenic
control of banana weevil, *Cosmopolites sordidus* (Germar)
(Coleoptera: Curculionidae) and other coleopteran insects in
transgenic plants**

By

ANDREW KIGGUNDU

Thesis submitted in partial fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

Department of Plant Sciences and Forestry and Agricultural Biotechnology
Institute (FABI)
in the

Faculty of Natural and Agricultural Sciences
University of Pretoria South Africa

Supervisor:
PROF. K.J. KUNERT

Co-supervisors:
PROF. A. VILJOEN
PROF. D. MICHAUD

May 2008



TABLE OF CONTENTS

ABSTRACT	x
THESIS COMPOSITION	xii
ACKNOWLEDGEMENTS	xiii
ABBREVIATIONS AND SYMBOLS	xv
CHAPTER 1 Introduction: The banana weevil and protease inhibitors.....	1
1.1 Plant improvement and Africa	2
1.2 The banana weevil.....	3
1.2.1 Weevil resistance	5
1.2.2 Weevil resistance screening	5
1.2.3 Resistance mechanisms	8
1.2.4 Resistance breeding.....	9
1.3 Protease/protease inhibitor system	12
1.3.1 Insect proteases	12
1.3.2 Plant protease inhibitors	14
1.3.3 Regulation of protease inhibitors	29
1.3.4 Structure of protease inhibitor genes	32
1.3.5 Protease inhibitors and insect control	34
1.3.6 Engineering of protease inhibitors.....	36
1.4 Study hypothesis aim and objectives	39
CHAPTER 2 Characterization of the digestive proteases in the banana weevil gut and the effects of recombinant phytocystatins on early larval growth and development.....	40
2.1 Abstract	41
2.2 Introduction	42
2.3 Material and methods	43



2.3.1 Reagents	43
2.3.2 Insect colony and maintenance	43
2.3.3 Gut extractions and protein concentration determination	44
2.3.4 Determination of pH optima	44
2.3.5 Fluorometric assay.....	45
2.3.6 Gelatin SDS-polyacrylamide gel electrophoresis	46
2.3.7 Cloning of OC-I and PC genes	47
2.3.8 Protein expression and purification.....	49
2.3.9 In-vitro assays with recombinant phytocystatins	50
2.3.10 Infiltration of banana stem with phytocystatin.....	51
2.4 Results	52
2.4.1 pH optima	52
2.4.2 Fluorometric assays	52
2.4.3 Gelatin SDS-polyacrylamide gel electrophoresis	55
2.5 Discussion	62
CHAPTER 3 Phylogenetic and structural comparisons of phytocystatins: A	
bioinformatics approach.....	64
3.1 Abstract	65
3.2 Introduction	66
3.3 Materials and methods	67
3.3.1 Sequence analysis	67
3.3.2 Protein structure modelling	68
3.3.3 Active site and docking.....	68
3.4 Results	68
3.5 Discussion	81



CHAPTER 4 Engineering of a papaya cystatin using site - directed mutagenesis to improve its activity against papain and weevil digestive cysteine proteases	85
4.1 Abstract	86
4.2 Introduction	87
4.3 Materials and methods	89
4.3.1 <i>Phylogenetic and structural model analysis</i>	89
4.3.2 <i>Detection of positive selection sites in PhyCys</i>	89
4.3.3 <i>Construction of over-expression vector for papaya cystatin</i>	90
4.3.5 <i>Mutagenesis primer design</i>	90
4.3.6 <i>Site-directed mutagenesis</i>	93
4.3.7 <i>Protein expression</i>	95
4.3.8 <i>Purification</i>	96
4.3.7 <i>Enzyme Kinetics of mutants</i>	97
4.4 Results	97
4.4.1 <i>Rational of mutations</i>	97
4.4.2 <i>Positive selection among plant cystatin genes</i>	100
4.4.3 <i>Mutation and expression of recombinant mutant papaya cystatins</i>	103
4.4.4. <i>Inhibition activity of papaya cystatin mutants</i>	104
4.5 Discussion	108
CHAPTER 5 General discussion and future outlook	110
5.1 Summary	111
5.2 Future outlook	114
REFERENCES	116



LIST OF FIGURES

Figure 1.1	The three broad methods of crop improvement compared..	3
Figure 1.2	The Adult Banana Weevil (<i>Cosmoplites sordidus</i>).....	4
Figure 1.3	Genetic engineering strategies currently in commercially produced crops.	12
Figure 1.4	Substrate-like mechanism of inhibition by two serine protease inhibitors.	21
Figure 1.5	General classification of the cystatin super-family phycostatins.....	22
Figure 1.6	Alignment of selected members of the 4 cystatin families illustrating the sequence conservation regions within the family members.....	23
Figure 1.7	The three dimensional structure of OC-I showing the characteristic 5 anti-parallel B strands, the single 5 turn a-helix, the N-terminal, the 1 st and 2 nd hairpin-like loops	25
Figure 1.8	Three-dimensional plot showing the complex between papain (blue and green) and Chicken egg white cystatin (CEW) colored light blue, red and yellow (PDB accession No. 1STF).	26
Figure 2.1	Schematic diagram of the construction of expression vectors pQOC-I and pQPC used in the study to express OC-I and PC in <i>E. coli</i> , respectively.....	48
Figure 2.2	Effect of pH on the hydrolysis of azocasein by banana weevil larval gut proteases.....	53
Figure 2.3	Cathepsin B, L and H like activities detected in banana weevil larval gut extracts. (B) Trypsin and chymotrypsin-like activities detected in the same extracts.	54
Figure 2.4	The effect of protease inhibitors on the proteolysis activity of banana weevil larval gut proteases revealed by separation in a mildly denaturing 15% SDS-PAGE co-polymerized with gelatin.	57
Figure 3.2	SDS-PAGE of (A) PC and (B) OC-I at different purification steps..	59



Figure 2.6 The effect of recombinant cysteine protease inhibitors rOC-I and papaya rPC on the cysteine protease activity of banana weevil mid-gut extracts. 60

Figure 2.7 (A) Illustration of the apparatus used to vacuum infiltrate banana flower stalk disks with cystatin solution. (B) Larvae on the left after developing on cystatin-free (control) disks for 10 days, while larvae on the right developed in cystatin treated disks over the same period. (C) The growth rate of larvae that were reared on banana stem disks vacuum-infiltrated with a solution of recombinant rOC-I and rPC. 61

Figure 2.1 (A) Amino acid sequence alignment of known phytocystatins showing residue conservation across the different cystatins studied. A consensus sequence was also generated. Identical amino acids are highlighted in black while similar ones are in grey. (B) Cartoon of the generalized secondary structural elements of phytocystatins.. 73

Figure 3.2 Phylogenetic tree for known phytocystatins based on the neighbour-joining method using PROTDIST and NEIGHBOR programs available in the PHYLIP (Phylogeny Inference Package) Version 3.57.. 75

Figure 3.3 Predicted three-dimensional structures of selected phytocystatins representing the major phytocystatin phylogenetic groups..... 79

Figure 3.4 Modelled complex between OC-I (top) and papain (bottom) in front and side views..... 80



Figure 4.1	Schematic representation of recombinant protein expression vector pQE31PC-I created to express papaya cystatin and in which site directed mutagenesis was performed.	90
Figure 4.2	Schematic representation of the site-directed mutagenesis protocol used (modified from QuickChange® Site-Directed Mutagenesis Kit)	93
Figure 4.3	Consensus sequence from a multiple alignment (see Chapter 3).To illustrate residues subjected to mutations.....	100
Figure 4.4	Location of positively selected codon sites (with Bayesian posterior probabilities greater than 60% under model M3) in Poaceae and Solanaceae cystatins.....	103
Figure 4.5	SDS-PAGE (12%) of the purified fractions of selected papaya cystatin mutants CYSI07D, CYSA53P CYSA32V and CYSW78P to establish purity of the purification.	104
Figure 4.6	Comparison of inhibition activity between wild-type papaya cystatin (red bar) and 18 mutants of the papaya cystatin gene.....	106
Figure 4.7	Inhibition activities between wild-type papaya cystatin (red bar) and 18 mutants of the papaya cystatin. Inhibitors were tested by monitoring change in reaction rates of banana weevil (A) and black maize beetle (B) gut extracts.	107



LIST OF TABLES

Table 1.1	Suggested sources of banana weevil resistance in <i>Musa</i>	7
Table 1.2	Mechanistic classes of proteases, amino acid residues constituting their active site, their optimum pH ranges and examples of the protease enzymes.....	16
Table 1.3	Transgenic crop plants reported to express serine protease inhibitor genes from plants with improved resistance to respective pests.....	19
Table 1.4	Insect pests with reported susceptibility to phytocystatins, either <i>in-vitro</i> , in artificial diet or in transgenic plants.....	27
Table 2.1	Inhibition of banana weevil gut proteases by cysteine (A) and serine (B) protease inhibitors.....	56
Table 3.1	Known phytocystatins obtained from sequence databases: EMBL= European Molecular Biology Laboratory, PIR=Protein information Resource, SP=SwissProt, GB=GeneBank and NCBI=National Centre for Biotechnology Information.....	70
Table 3.2	Percentage identity matrix of phytocystatins	76
Table 4.1	Sequence information of the mutagenic primer pairs used for the mutations. The mismatched bases are underlined.....	92
Table 4.2	Mutations performed on native papaya cystatin, the amino acid changes made and the respective rationale	98
Table 4.3	Evidence for positive selection events among codon sites of Poaceae and Solanaceae cystatins	102



Engineering plant cysteine protease inhibitors for the transgenic control of banana weevil, *Cosmopolites sordidus* (Germar) (Coleoptera: Curculionidae) and other coleopteran insects in transgenic plants

Andrew Kiggundu

National Banana Research Programme, Kawanda Agricultural Research Institute,
National Agricultural Research Organisation,
P. O. Box 7065, Kampala, Uganda

Department of Plant Science and the Forestry and Agricultural Biotechnology Institute
University of Pretoria, 74 Lunnon Road, Hillcrest,
Pretoria, 0002. South Africa

Supervisor:

Karl Kunert

Department of Plant Science and the Forestry and Agricultural Biotechnology Institute
University of Pretoria, 74 Lunnon Road,
Hillcrest, Pretoria, 0002. South Africa

Co-supervisors:

Altus Viljoen

Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology
Institute University of Pretoria, 74 Lunnon Road,
Hillcrest, Pretoria, 0002. South Africa

Currently at: Department of Plant Pathology, University of Stellenbosch, Private Bag X1,
Matieland 7602, South Africa.

Dominique Michaud

Département de Phytologie, Pavillon Paul-Comtois,
Université Laval, Sainte-Foy (Québec),
Canada G1K 7P4



ABSTRACT

Cysteine protease inhibitors (cystatins) are expressed in plants in response to wounding and insect herbivory and they form part of the native host-plant defence system. Cysteine proteases are enzymes important in the break down of dietary proteins mainly in the mid gut of coleopteran insects such as the banana weevil. The inhibition of these proteases has a direct effect on the digestive activity of the insect resulting in protein deficiency. This significantly affects insect development and survival. Based on these observations, strategies have been designed involving expression of cysteine protease inhibitors for the transgenic control of insect pests of several crop plants. For this study, it was hypothesized that the major proteases in banana weevil are cysteine proteases and can be effectively targeted by plant cystatins. It was further hypothesised that since plant cystatins are defense related, certain amino acid residues may have undergone positive selection. This provides an opportunity to increase their inhibitory potential to the weevil gut proteases via protein engineering. To prove the hypotheses, both *in-vitro* and *in-vivo* assays were set up thus allowing us to demonstrate the presence of cysteine type proteases banana weevil as well as the effect of cystatins on the weevil proteases and early development. Initial *in-vitro* experiments were able to characterize the proteolytic activity of the banana weevil gut proteases, which are mostly of the cysteine type, and in particular cathepsin B and L like. Two recombinant phytocystatins were further successfully produced using a 6xHis-tagged affinity chromatography system in *Escherichia coli* bacteria. The recombinant phytocystatins were used in a newly developed vacuum infiltration assay system using banana stems. Young weevil larvae were allowed to develop on phytocystatin-treated stems for up to 10 days. They had a 60% reduction in body weight and rate of growth compared to those that grew in untreated stems. By carrying out site-directed mutagenesis to improve the inhibition efficiency of a model papaya cystatin, more



than 8 amino acid residues were found to be subjected to positive selection. Mutation of amino acids yielded improved the inhibition potential of papaya cystatin against the model cysteine protease papain. Increased inhibition was greatest when amino acids were changed in the highly variable regions of the amino acid sequence very closely to the conserved regions.

This study has been able to show for the first time that banana weevils use cysteine protease as major protein hydrolysis enzymes and that these can be effectively targeted by plant cystatins. It has also created novel phytocystatins using engineering of single amino acid sites following an evolutionary approach to modulate them for improved activity and targeting specific proteases.



THESIS COMPOSITION

Chapter 1 introduces the banana weevil which is a coleopteran pest of banana that barrows through the underground stem of banana plants causing considerable damage. The chapter reviews conventional efforts towards screening the banana germplasm for resistance, resistance mechanisms, and cross breeding activities targeting the banana weevil as well as protease inhibitors as one group of genes that have potential for weevil control in a transgenic approach. **Chapter 2** reports on investigations into the nature of the banana weevil gut environment *vis a vis* protease activity reveals the protease profile of the gut and bioassays are developed and conducted to test the hypothesis that banana weevil use mostly cysteine protease in protein digestion and can be targeted by cysteine protease inhibitors from plants. **Chapter 3** relates to the phylogenic, structural and protein modelling analysis of plant cysteine protease inhibitors in an effort to understand evolutionary trends. This could assist a protein engineering strategy to improve the cystatin action against weevil and other coleopteran insects. **Chapter 4** combines evolutionary analysis to determine if positive selection has acted on the cysteine protease inhibitor amino acid residues to lead to the observed diversity. This was followed by protein engineering approaches using site-directed mutagenesis guided by evolutionary analysis to produce novel mutants of the papaya cystatin with increased inhibition capacity. Finally **Chapter 5** discusses the contributions of this thesis to our better understanding of these important plant proteins. It further discusses how best to make future use of them, not only in the improvement of resistance to banana weevil but also to other coleopteran crop pests.



ACKNOWLEDGEMENTS

The encouragement my late father gave me to pursue science as a career is highly appreciated. Even that he is no longer with us I am sure he is happy to know that this is part of his efforts many years back. To my mother I do not have the right words to say thank you for all the patience you had with me and the toughness that ensured I do not take a wrong detour.

I am very grateful to the National Agricultural Research Organisation, Uganda for allowing me to take time off work and for supporting my studies both financially and morally. Special thanks to Drs. Wilberforce Tushemereirwe and Eldad Karamura whose encouragement and moral support were very helpful not only towards this thesis but also in ensuring that what I have achieved finds usefulness in NARO and Uganda.

Very importantly, I am indebted to the Rockefeller Foundation for the scholarship support that allowed me to undertake this study. To Dr. Joe DeVries of the foundation for his strong belief in me and awarding me this scholarship, but also very insightful were the discussions and suggestions he gave that led to the work in this thesis.

My sincere thanks go to Prof. Karl Kunert who was more than willing to work in his lab under his supervision, way back in February 2002. I thank him for the very good mentoring discussions and useful suggestions that have led to the completion of this scientific accomplishment. I thank FABI, the Plant Science Department of the University of Pretoria for proving such a wonderful and high standard environment for me to achieve what I have. Doing this work at the University of Pretoria and in South Africa has really made me



different. Dr. Altus Viljoen for his encouragement, co-supervision and untiring support that made my stay at FABI as comfortable as possible.

My sincere gratitude also to colleagues and friends at Forestry and Agricultural Biotechnology Institute (FABI), the University of Pretoria, who supported me during my study and stay at FABI



ABBREVIATIONS AND SYMBOLS

BBTI	Bowman-Birk trypsin inhibitor
bp	Base pair
CaMV	Cauliflower Mozaic Virus
E-64	Trans-epoxysuccinyl-L-leucylamido (4-guanidino) butane
EDTA	Ethylenediaminetetraacetic acid
kDa	Killo Dalton
LB	Luria-Bertani
mL	Milliliter
nm	Nanometer
OC-I	oryzacystatin-I
PAGE	Polyacrylamide gel electrophoresis
PC	Papaya cystatin
PCR	Polymerase Chain Reaction
PI	Protease inhibitor
PMSF	Phenylmethysulphonyl fluoride
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SE	Standard error
U	Unit
Z-phe-arg-AMC	Benzyloxycarbonyl-phenylalanine-arginie aminomethylcoumarin
µg	Microgram
µl	Microlitre
µM	Micromolar
%	Percentage
°C	Degree Celsius
m	Metre



CHAPTER 1

Introduction: The banana weevil and protease inhibitors

Scientific Communication

Major sections of this introduction were published in:

Kiggundu, A., Pillay, M., Viljoen, A., Gold, C., Tushemereirwe, W., Kunert, K. 2003. **Enhancing banana weevil (*Cosmopolites sordidus*) by genetic modification. A perspective.** *African Journal of Biotechnology* 2, 563-566.



1.1 Plant improvement and Africa

The African continent, and specifically Sub-Saharan Africa, will be among the critical areas for future food production. During the 1990s, the pace of agricultural growth has already improved considerably in many African countries when compared to the 1970s and 1980s. Conventional plant breeding has, for example, helped in the development of new crop varieties with increased resistance to biotic stress, such as insect infestation, which is of importance for Africa (DeVries and Toenniessen, 2001). However, small-scale farmers in many African countries are not yet utilizing the advantages that modern biotechnologies offer. This includes tissue culture derived planting material that has been cleaned of disease and pest infestation as well as novel varieties of crops developed through genetic engineering (Figure 1.1). This is in contrast to farmers in industrialized countries, who are rapidly taking advantage of the modern technologies to overcome crop production constraints. For future agricultural development it is a vital necessity that African farmers also get access to recent developments in modern plant breeding where plants are improved through the enhancement of useful characteristics. Any conventional breeding approach is likely to deliver only part of the required yield increase needed for a growing population in Africa. In addition, crops derived from the application of plant biotechnology with superior characteristics, such as insect resistance, might further reduce the use of expensive and often toxic insecticides. Unfortunately, biotechnological breakthroughs are only very slowly evaluated and implemented on the African continent as a useful complement to conventional breeding. This is mostly due to the high cost, lack of existing skills in plant biotechnology, technology protection by developed countries, and to concerns about possible health or ecological risks from genetically modified (GM) plants (Dunwell, 1998).

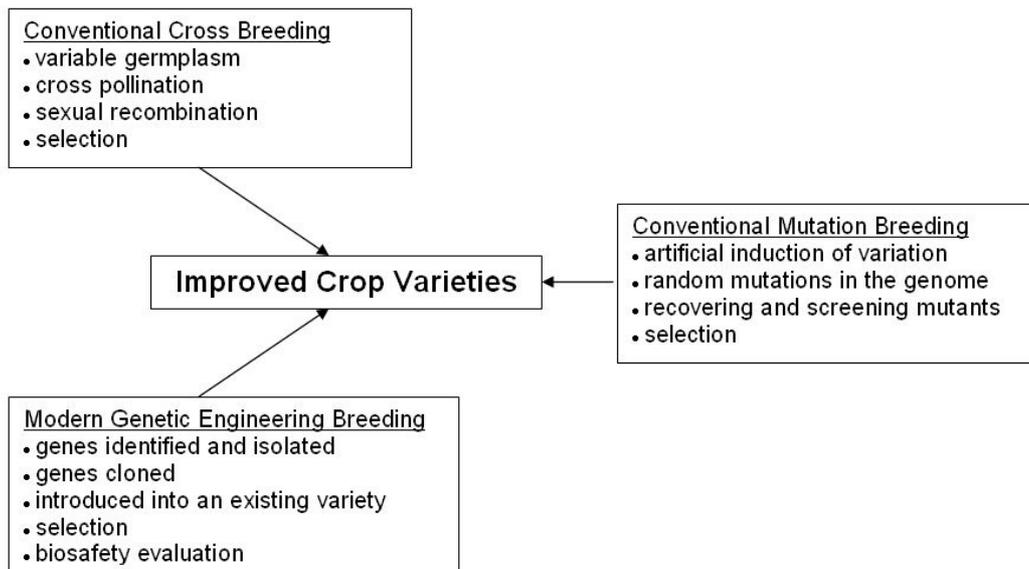


Figure 1.1 The three broad methods of crop improvement compared. Conventional cross breeding is limited by the availability of the required traits in the gene pool as well as sexual compatibility of the crop. Conventional mutation breeding relies on the use of artificial induction of variation by the use of radiation and chemical treatment. It requires laborious screening of a large number of mutants to find a desired trait. Modern genetic engineering offers the most significant advancement in crop improvement. Theoretically a characteristic from any organism of any species can be introduced into a plant to create new varieties with characteristic never though possible before.

1.2 The banana weevil

Among the targets for application of plant biotechnology is to increase resistance of banana to the banana weevil, *Cosmopolites sordidus* (Germar) (Coleoptera: Curculionidae). The weevil is a pest of considerable importance in Africa which significantly affects banana and plantain production (Ostmark, 1974; Gold, 1998; Gold and Messiaen, 2000; Swennen and Vulysteke, 2001; Fogein *et al.*, 2002). The weevil has been associated with rapid plantation decline in East Africa (Gold *et al.*, 1999b) and a phenomenon called “yield decline syndrome” in West Africa. The adult weevils are free living, have a nocturnal habit, and rarely fly. Their eggs are

deposited inside the plant tissue at the base of the pseudo-stem or on an exposed corm. On hatching, the larvae tunnel through the corm for feeding and development. Tunnelling reduces the water and mineral transport, thereby weakening the plant, reducing the bunch weight (yield) and causing plant toppling during windstorms. In severe weevil infestations, crop losses of up to 100% have been reported (Sengooba, 1989). The establishment of new plantings may fail (Price, 1994) and yield loss appears to increase gradually, reaching 44% in the fourth ratoon cycle (Rukazambuga *et al.*, 1998).

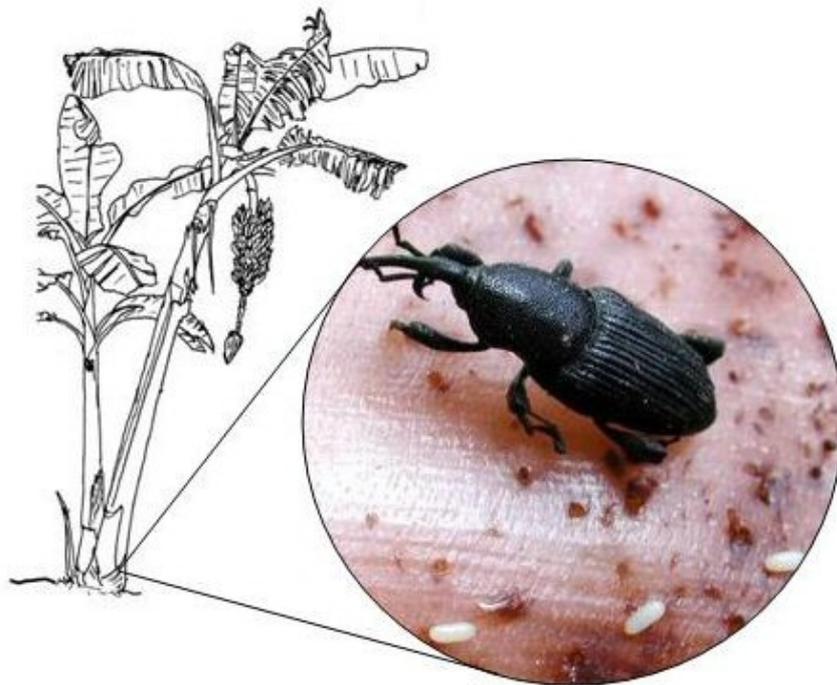


Figure 1.2 The adult banana weevil (*C. sordidus*) lays eggs on the banana plant just above the soil surface. When the eggs hatch the emerging larvae burrow through the underground stem leading to yield loss, structural weakness and toppling of the plant.

Weevil control is currently based on the application of cultural practices, such as the use of clean planting material, systematic trapping of adult weevils in an effort to control the weevil population, and field sanitation to remove residues that may form breeding grounds for the weevil (Gold, 2000; Gold and Messiaen, 2000). Although cultural control methods contribute to weevil management, both the high labour input and material requirements are often



limiting factors for adoption (Gold, 1998; Gold *et al.*, 2001). Application of effective pesticides is economically unfeasible for subsistence producers and, unfortunately, the banana weevil can developed resistance to a range of pesticides (Collins *et al.*, 1991; Gold *et al.*, 1999a). Consequently, development of resistant plants has been suggested as a potential long-term intervention for weevil control, especially on small-scale farms, as the inclusion of such plants might be part of an integrated pest management (IPM) framework (Seshu-Reddy and Lubega, 1993).

1.2.1 Weevil resistance

The development of weevil-resistant bananas and plantains is still in its infancy. Only a few breeding programs consider banana weevil resistance as a criterion for improvement. This is despite the fact that triploid plantains (AAB) and East African highland bananas (EAHB-AAA) are major sources of food in Africa, and that both are highly susceptible to weevil infestation (Fogain and Price, 1994; Gold *et al.*, 1994; Ortiz *et al.*, 1995; Musabyimana *et al.*, 2000; Kiggundu *et al.*, 2003a). Lack of considerable progress in the development of weevil-resistant banana has been, and still is, due to the cumbersome nature of techniques for resistance screening and the limited knowledge on resistance mechanisms.

1.2.2 Weevil resistance screening

Considerable work has been done on screening diverse *Musa* germplasm for weevil resistance in Africa (Pavis and Lemaire, 1997; Kiggundu *et al.*, 1999). Although plantains and EAHB were found to be the most susceptible, there are a few exceptions. For example, in India Padmanaban *et al.* (2001) found two plantain cultivars (Karumpoovan and Poozhachendu) resistant to the banana weevil, while Fogain and Price (1994) found that the cultivar Kedongkekang (plantain AAB) is also resistant. Kiggundu *et al.* (2003a) found some highland banana cultivars (cvs. Tereza, Nalukira and Nsowe) being intermediately resistant.



The large variability in weevil response observed in germplasm and hybrid testing indicates that useful sources of weevil resistance are indeed available in *Musa*. Possible candidates for use in conventional crosses have been therefore selected based on very low levels of weevil damage in the field, and on pollen fertility (Table 1.1). The AA genome progenitor *Musa accuminata* Colla is more susceptible to weevils than the BB progenitor *Musa balbisiana* Colla (Mesquita *et al.*, 1984), and it is expected that AA type sources of resistance might ultimately produce hybrids with better consumer acceptability.

Table 1.1 Suggested sources of banana weevil resistance in *Musa*.

Cultivar	Genome group	Reference
Yangambi km-5	AAA	Fogain and Price, 1994; Kiggundu <i>et al.</i> , 2003a
Sannachenkadali	AA	Padmanaban <i>et al.</i> , 2001
Sakkali	ABB	
Senkadali	AAA	
Elacazha	BB	
Njalipoovan	AB	
Pisang Awak	ABB	Kiggundu <i>et al.</i> , 2003a
FHIA03	AABB	
TMBx612-74	IITA hybrid	
TMB2x6142-1	IITA hybrid	
TMB2x8075-7	IITA hybrid	
TMB2x7197-2	IITA hybrid	
Long Tavoy	ABB	Ortiz <i>et al.</i> , 1995
Njeru	AA	Musabyimana <i>et al.</i> , 2000
Muraru	AA	
Calcutta-4	AA	Fogain and Price, 1994; Ortiz <i>et al.</i> , 1995; Kiggundu <i>et al.</i> , 2003a
Bluggoe	ABB	Fogain and Price, 1994; Kiggundu <i>et al.</i> , 2003a
<i>M. balbisiana</i>	BB	Fogain and Price, 1994



1.2.3 Resistance mechanisms

Classical resistance mechanisms (Painter, 1951) have been investigated in *Musa* germplasm. So far antibiosis (factors affecting larval performance), rather than antixenosis (attractivity), appears to be the most important weevil resistance mechanism (Ortiz *et al.*, 1995; Abera *et al.*, 1999). Although some differences in attracting adult weevils to different cultivars have been found, there were no direct correlations with plant damage (Budenburg *et al.*, 1993; Pavis and Minost, 1993; Musabymana, 1995; Abera *et al.*, 1999). Difference in attraction has been rather due to environmental factors, such as soil moisture, around a cultivar with high sucker number (Ityeipe, 1986).

Several phenological factors seem also to contribute to weevil resistance. Corm hardness was the first biophysical factor associated with resistance. Whereas Pavis and Minost (1993) found a small, negative correlation ($r = -0.47$) between corm hardness and weevil damage, Ortiz *et al.* (1995) found no relationship between the two factors in segregating plantain progenies. They rather suggested other weevil resistance factors such as chemical toxins or anti-feedants. Kiggundu *et al.* (2003a) found corm dry matter content, resin/sap production and suckering ability to negatively correlate with weevil damage.

The suggestion that biochemical compounds affected weevil performance further led to investigations of resistant selections by using high-performance liquid chromatography (HPLC). HPLC chromatograms from corm extracts of weevil-resistant AB and ABB cultivars (cvs. Kisubi and Kayinja) showed compound peaks that were absent not only in susceptible clones, but also in some resistant clones of the AA and AAA genomes (e.g. Calcutta-4 and Yangambi km-5). This result possibly indicates a type of antibiotic mechanism that may be based on toxic compounds. These compounds are seemingly present in weevil-resistant



cultivars with the B genome whereas a different form of resistance may be present in the genome of weevil-resistant AA cultivars.

In general, banana improvement for weevil resistance using existing resistance mechanisms appears complex and not well advanced due to a limited understanding of the genetics of resistance. Weevil resistance is probably controlled by a number of genes. These genes are different in the A and the B genome groups (Ortiz *et al.*, 1995; Ortiz, 2000). Resistance in the A genome might include corm hardness, which is less important for the B genome. Significant genetic correlations were observed between weevil damage, corm hardness, dry matter content, sap/resin production, and corm size, further indicating the complexity of weevil resistance in the diverse *Musa* germplasm. Conventional improvement for weevil resistance might ultimately also require multiple strategies in any conventional breeding program and therefore might render the overall process very slow and long-term.

1.2.4 Resistance breeding

1.2.4.1 Molecular markers

The application of DNA markers in banana has mostly been for germplasm characterisation (Crouch and Crouch, 1999; Visser, 2000; Pillay *et al.*, 2001). Molecular genetic techniques have recently been applied for improving the efficiency of *Musa* breeding. For example, markers for simple traits, such as parthenocarpy (Crouch *et al.*, 1998), earliness and regulated suckering (Vuylsteke *et al.*, 1997), and for a major quantitative trait like banana streak disease resistance (Carreel *et al.*, 1999; Lheureux *et al.*, 2003), have been developed for *Musa*. Despite these efforts, molecular biology-based breeding tools, such as Molecular Marker Assisted Selection (MAS), are still not highly developed for banana when compared to other major food crops in the world. MAS breeding has, however, the potential to markedly enhance the pace and efficiency of genetic improvement in *Musa* (Crouch *et al.*, 2000).



1.2.4.2 Genetic modification

Production of genetically modified (GM) banana has been attempted by several research groups. Although remarkable achievements have already been made in banana transformation, the identification and introduction of useful genes into banana to reduce losses caused by the banana weevil is still a major challenge. This is partially due to the lack of information on expression of endogenous banana genes after weevil infestation.

Several approaches can be followed. These include for example the production of transgenic banana expressing a plant lectin. Lectins confer a protective role against a range of organisms (Sharma *et al.*, 2000). They have been isolated from a wide range of plants including snowdrop, pea, wheat, rice and soybean and their carbohydrate-binding capability renders them toxic to insects. A lectin from snowdrop, *Galanthus nivalis* agglutinin (GNA), is toxic to several insect pests in the orders Homoptera, Coleoptera and Lepidoptera (Tinjuangjun, 2002). A study is currently being conducted to test the effect of GNA and the *Aegopodium podagraria* lectin (APA) among others on the mortality and reproduction of three nematode species pathogenic to banana (Carlens, 2002). Similar work could be extended to banana weevil using *in-vivo* assays. A major concern about the use of lectins, however, is that some of them, such as the wheat germ agglutinin (WGA), are toxic to mammals (Jouanin *et al.*, 1998). However, the snowdrop and garlic lectins are toxic only to insects (Boulter, 1993) and these deserve investigation for weevil control.

Expression and biological activity of the *Bacillus thuringiensis* (*Bt*) toxin has been extensively investigated in GM plants for insect control and represents a further approach for insect control in banana. *Bt* plants are currently the most widely used GM technology for Lepidopteran pest control in commercial crops (Krattiger, 1997). *Bt* genes products are a



group of more than fifty insecticidal crystal proteins. When ingested by an insect, they are solubilised in the alkaline environment characteristic of Lepidopteran insect midguts (e.g. Cry1 proteins). The proteins then become toxic by binding to apical border brush membranes of the columnar cells. This causes lysis of the cells and eventual death of the insect. On the other hand Coleopteran insects like the banana weevil do not have such high pH-induced solubilisation of Bt toxins (e.g. Cry3 proteins). The expression of a selected *Bt* gene for weevil resistance will therefore need a longer term strategy. *Bt* screening, however, is hampered by the lack of any artificial diet for the banana weevil, which is a pre-requisite for efficient screening under controlled conditions.

Alpha-amylase inhibitors (AI) and chitinase enzymes might also have a potential for weevil control. They are divided into two types, AI-1 and AI-2, isolated from common and wild beans (*Phaseolus vulgaris*), respectively. Alpha-amylase inhibitors operate by inhibiting the enzyme alpha-amylase which are responsible for the break down of starch to glucose in the insect gut (Le Berre-Anton *et al.*, 1997; Morton *et al.*, 2000). Ishimoto *et al.* (1996) produced transgenic adzuki beans with enhanced resistance to bean bruchids, which are Coleopteran insects. Since they are active against this type of insects, they might be of interest for banana weevil control in GM banana. Chitinase enzymes are produced as a result of invasion either by fungal pathogens or insects. Transgenic expression of chitinase has shown improved resistance to insect pests in tobacco against Lepidopteran insects (Ding *et al.*, 1998). Recently, a rice chitinase gene has been transformed into bananas directed towards the control of fungal pathogens in particular *Micosphaerella fijiensis* the causal agent of black sigatoka disease (Arinaitwe, 2002).

Among the proteins useful for a transgenic approach, protease inhibitors, such as cysteine and serine protease inhibitors, are possibly also useful candidates to protect plants against insect attack (Ryan, 1990; Pernas *et al.*, 2000; Ashouri *et al.*, 2001). They operate by disrupting protein digestion in the insect mid-gut via inhibition of proteases. These inhibitors have been investigated in this study in greater detail.

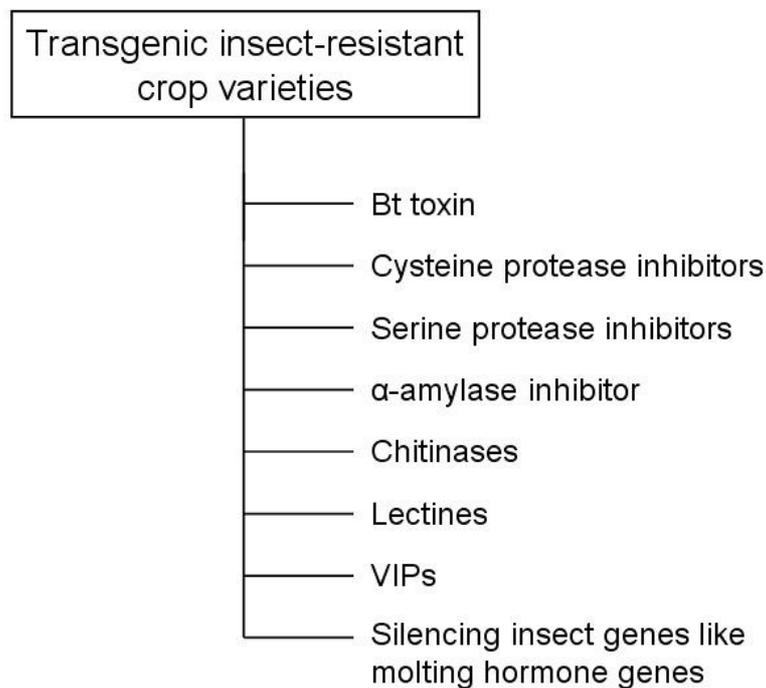


Figure 1.3 Genetic engineering strategies currently in commercially produced crops (only *Bt* toxin) and others being developed for increasing resistance to crop insect and nematode pests.

1.3 Protease/protease inhibitor system

1.3.1 Insect proteases

The term “protease” includes both “endopeptidases” and “exopeptidases” whereas; the term “proteinase” is used to describe only “endopeptidases” (Ryan, 1990). The digestive proteolytic enzymes in the different orders of commercially important insect pests belong to



one of the major classes of proteases. Serine proteases have been identified in extracts from the digestive tracts of insects from many families particularly those of Lepidoptera (Houseman *et al.*, 1989). Many of these enzymes are inhibited by protease inhibitors. The order Lepidoptera, which includes a number of crop pests, the pH of the gut environment is in the alkaline range of 9-11 (Applebaum, 1985) where serine proteases and metallo-exopeptidases are most active.

Coleopteran and Hemipteran species tend to utilize cysteine proteases (Murdock *et al.*, 1987) while Lepidopteran, Hymenopteran, Orthopteran and Dipteran species mainly use serine proteases (Ryan, 1990; Wolfson and Murdock, 1990). The effect of class specific inhibitors on the pest digestive enzymes is not always a simple inhibition of proteolytic activity. Recent studies have indicated that there are often two or more populations of digestive enzymes in target pests, some with susceptibility to inhibition and other insensitive to specific inhibitors (Michaud *et al.*, 1996; Bown *et al.*, 1997). Some insects respond to ingestion of plant PIs, such as soybean trypsin inhibitor (Broadway and Duffey, 1986) and oryzacystatin (Michaud *et al.*, 1996), by hyper-producing inhibitor-resistant enzymes.

Isolation and characterisation of midgut proteases from the larvae of Cowpea weevil, *Callosobruchus maculatus* (Fab.) (Col.: Bruchidae) (Kitch and Murdock, 1986; Campos *et al.*, 1989) and the Mexican bean weevil *Zabrotes subfasciatus* (Boheman) (Col.: Bruchidae) (Lemos *et al.*, 1987) confirmed the presence of a cysteine mechanistic class of protease in such insects. Similar proteases have been isolated from midguts of the Confused flour beetle *Tribolium castaneum*, Mexican bean beetle *Epilachna varivestis* (Mulsant) (Col.: Coccinellidae) (Murdock *et al.*, 1987) and the Common bean weevil *Acanthoscelides obtectus* (Say.) (Col.: Chrysomelidae) (Wieman and Nielsen, 1988).



In a study of the proteases from the midgut of several members of the order Coleoptera, 10 of 11 species representing 11 different families had gut proteases that were inhibited by p-chloromercuribenzenesulfonic acid (PCMBS), a potent sulphhydryl reagent (Murdock *et al.*, 1988) indicating that the proteases were of the cysteine mechanistic class. The optimum activity of cysteine proteases is usually in the pH range 5-7, which is the pH range of the guts of insects which use cysteine proteases (Murdock *et al.*, 1987).

1.3.2 Plant protease inhibitors

Protease inhibitors are widely produced in the plant kingdom, both in different plant species as well as tissue and organ/cell types. Currently, knowledge places them in different roles including functioning as resistance mechanisms to pest and pathogen attack as the most important. They operate as part of the host plant resistance arsenal to invading organisms ranging from insects, nematodes, fungi, bacteria and viruses. A further role is in programmed cell death in plants. Programmed cell death (PCD), also referred to as apoptosis, is a physiological process by which cells or organs that have reached a certain age are spontaneously killed to preserve the integrity of the whole organism. Cysteine proteases are involved in a key step in animal PCD and have recently also been found to be important in plants (Solomon *et al.*, 1999). Evidence of protease inhibitors being important in plant protection was first investigated by Mickel and Standish (1947). They observed that the larvae of certain insects were unable to develop normally on soybean products. Subsequently, trypsin inhibitors present in soybean were shown to be toxic to the larvae of the confused flour beetle *Tribolium confusum* (duVal) (Col.: Tenebrionidae) (Lipke *et al.*, 1954). Following these early studies, there have been many examples where protease inhibitors have been found to be active against certain insect species. These include both *in-vitro* assays



against insect gut proteases (Pannetier *et al.*, 1997; Koiwa *et al.*, 1997) and *in-vivo* artificial diet bioassays (Urwin *et al.*, 1997; Vain *et al.*, 1998).

The majority of protease inhibitors studied from the plant kingdom originate from three main families namely Leguminosae, Solanaceae and Gramineae (Richardson, 1991). Many of these protease inhibitors are rich in cysteine and lysine, contributing to better and enhanced nutritional quality (Ryan, 1998). Protease inhibitors also exhibit a very broad spectrum of activity including suppression of nematodes like the tobacco cyst nematode; *Globodera tabaccum* (Lownsbery & Lownsbery) Skarbilovich (Nematoda: Heteroderidae), potato cyst nematode; *Globodera pallida* (Stone) (Nematoda: Heteroderidae), and the root-knot nematode *Meloidogyne incognita* (Kofoid and White) Chitwood (Nematoda: Meloidogynidae) by CpTi (Williamson and Hussey, 1996), inhibition of spore germination and mycelium growth of *Alternaria alternata* by buckwheat trypsin/chymotrypsin (Dunaevskii *et al.*, 1997), and cysteine protease inhibitors from pearl millet inhibiting growth of many pathogenic fungi including *Trichoderma reesei* (Joshi *et al.*, 1998). These inhibitor families that have been found are specific for each of the four mechanistic classes of proteolytic enzymes. Based on the active amino acid in their “reaction center” (Koiwa *et al.*, 1997) they are classified as serine, cysteine, aspartic and metallo-proteases. There are four different classes of proteases and therefore protease inhibitors are classified and named based on the protease mechanistic class (Table 1.2) they inhibit. For example, cysteine proteases are inhibited by cysteine protease inhibitors or also called cystatins.



Table 1.2 Mechanistic classes of proteases, amino acid residues constituting their active site, their optimum pH ranges and examples of the protease enzymes (modified from Oliveira *et al.*, 2003)

Protease class	Active site amino acids	pH optima range	Example proteases
Serine protease	Serine, Histidine and cysteine	7-9	Trypsin, Chymotrypsin, Cathepsin-G
Cysteine protease	Cysteine	4-7	Papain, Ficin, Bromelain, Ananain, Cathepsins B, C, H, K, L, O, S, and W
Aspartic protease	Aspartin	> 5	Cathepsin D and E, Renin, Pepsin
Metallo-protease	Metal-ion	7-9	Caboxypeptidases A and B, Amino peptidases

1.3.2.1 Serine protease inhibitors

Serine protease inhibitors are highly varied and have been extensively studied in both animals and plants. They are reversible inhibitors of serine proteases mainly trypsin and chymotrypsin. Their functions seem to range from regulation of endogenous protease activity to storage proteins, as they tend to accumulate in large amounts in storage organs like tubers and seeds reaching concentrations of about 2% of total protein (Gatehouse *et al.*, 1983). Recently however, the body of evidence supporting serine protease inhibitors as defensive compounds in plants towards pests and diseases has accumulated. The fact that serine proteases accumulate in large amounts in plant tissue suggests that they have less of a regulatory role towards endogenous protease activity whose amounts in tissue are much lesser. Instead these serine protease inhibitors seem to be more important in the control of phytophagous animals, whose digestive proteases are of the serine class. Table 1.3 summarises reports in which serine protease inhibitors have been shown to increase resistance to various pests when over expressed in transgenic plants. The serine class of



proteases, such as trypsin, chymotrypsin and elastase belonging to the same protein super family, are responsible for the initial digestion of proteins in the gut of higher animals (Garcia-Olmedo *et al.*, 1987). *In vivo* they are used to cleave long intact polypeptide chains into short peptides, which are then acted upon by exopeptidases to generate amino acids, the end products of protein digestion. These three types of digestive serine proteases are distinguished based on their specificity. Trypsin is specifically cleaving the C-terminal into residues carrying a basic side chain (Lys, Arg), chymotrypsin showing a preference for cleaving C-terminal to residues carrying a large hydrophobic side chain (Phe, Tyr, Leu), and elastase showing a preference for cleaving C-terminal to residues carrying a small neutral side chain (Ala, Gly) (Ryan, 1990). All serine inhibitor families from plants are competitive inhibitors and all of them inhibit proteases with a similar standard mechanism (Laskowski and Kato, 1980).

1.3.2.1.1 Classification, nomenclature and structure

Serine protease inhibitors have been isolated and described in many plant species and found throughout the plant kingdom. Sixteen different classes of serine protease inhibitors have been described and about seven in plants all with a common mechanism of action. There are four groups of serine protease inhibitors that have been widely studied in plants. Two were isolated from soybean seeds and named after their discoverers the Kunitz and Bowman Birk families of protease inhibitors. Another two were isolated from potato; potato serine inhibitors I and II (Mello *et al.*, 2003). The first was discovered by Kunitz (1945), who found that an inhibitor in soybean seeds caused raw soybean meal to be more inferior in nutritional quality to steam-cooked soybean meal. Kunitz inhibitors are nonomeric with a length of approximately 190 amino acids and structurally reinforced by two intra chain disulfide bonds. Each molecule has a single binding site which is involved in strong protease interaction.

Kunitz inhibitors belong to the super family called STI-like (*Structural Classification of Proteins* SCOP database), which includes other proteins with whom they are structurally but not functionally related e.g. the tetanus neurotoxin from *Clostridium tetani* the bacteria that causes tetanus in humans.

Table 1.3 Transgenic crop plants reported expressing serine protease inhibitor transgenes and showing improved resistance to respective insect pests (Lawrence and Koundal 2002).

Inhibitor	Crop Plant	Crop Pest	Reference
Cowpea trypsin inhibitor (CpTi)	Tobacco	<i>Heliothis virescens</i> (Fabricius) (Lepidoptera: Noctuidae)	Hilder <i>et al.</i> , 1987
	Rice	<i>Chilo suppressalis</i> (Walker) (Lepidoptera: Pyralidae)	Xu <i>et al.</i> , 1996
	Potato	<i>Lacanobia oleracea</i> (Linnaeus) (Lepidoptera: Noctuidae)	Gatehouse <i>et al.</i> , 1997
	Strawberry	<i>Otiorhynchus sulcatus</i> (Fabricius) (Coleoptera: Curculionidae)	Graham <i>et al.</i> , 1997
	Tobacco	<i>Spodoptera litura</i> (Fabricius) (Lepidoptera: Noctuidae)	Sane <i>et al.</i> , 1997
	Cotton	<i>Helicoverpa armigera</i> (Hubner) (Lepidoptera: Noctuidae)	Li <i>et al.</i> , 1998
	Wheat	<i>Sitotroga cerealla</i> (Olivier) (Lepidoptera: Gelechiidae)	Alpteter <i>et al.</i> , 1999
	Pigeonpea	<i>H. armigera</i>	Lawrence <i>et al.</i> , 2001
CpTi + Snowdrop lectin	Sweet potato	<i>Cylas formicarius</i> (Fabricius) (Coleoptera: Curculionidae)	Newell <i>et al.</i> , 1995
Potato inhibitor II	Tobacco	<i>Manduca sexta</i> (Linnaeus) (Lepidoptera: Sphingidae)	Johnson <i>et al.</i> , 1989
	Bean/corn/eggplant	<i>Chrysodeixis eriosoma</i> (Doubleday) (Lepidoptera: Noctuidae)	McManus <i>et al.</i> , 1994
	Rice	<i>Sesamia inferens</i> (Walker) (Lepidoptera: Amphipyridae)	Duan <i>et al.</i> , 1996
Tomato inhibitor I and II	Tobacco	<i>M. sexta</i>	Johnson <i>et al.</i> , 1989
Sweet potato trypsin inhibitor (TI)	Tobacco	<i>M. sexta</i>	Yeh <i>et al.</i> , 1997
Soybean Kunitz TI	Rice	<i>Nilaparvata lugens</i> (Stal.) (Hemiptera: Delphacidae)	Lee <i>et al.</i> , 1999
Barley TI	Tobacco	<i>Agrotis ipsilon</i> (Hufnagel) (Lepidoptera: Noctuidae)	Carbonero <i>et al.</i> , 1993
<i>Nicotiana glauca</i> protease inhibitor (PI)	Tobacco	<i>Helicoverpa punctigera</i> (Wallengren) (Lepidoptera: Noctuidae)	Heath <i>et al.</i> , 1997
	Pea	<i>Plutella xylostella</i> (Linnaeus) (Lepidoptera: Plutellidae)	Charity <i>et al.</i> , 1999
Serpin type serine PI	Tobacco	<i>Bemisia tabaci</i> (Gennadius) (Hemiptera: Aleyrodidae)	Thomas <i>et al.</i> , 1995



The Bowman-Birk inhibitors were first isolated and characterised in soybean seeds (Bowman, 1946; Birk *et al.*, 1960) and are common in legume seeds. Their polypeptide chains range from 70 to 80 amino acids, which can form oligomers. The main polypeptide chain is rich in cysteine residues with which it forms several intra-chain disulphide bonds. The molecule has two binding loops (active sites) one on either side, making a single molecule bind to two protease molecules. Each of the binding sites may have different specificities (Chye *et al.*, 2006).

Recent X-ray crystallography structure of winged bean, *Psophocarpus tetragonolobus* Kunitz-type double headed alpha-chymotrypsin shows 12 anti-parallel beta strands joined in a form of beta trefoil with two reactive site regions (Asn 38-Leu 43 and Gln 63-Phe 68) at the external loops (Ravichandaran *et al.*, 1999; Mukhopadhyay, 2000). Structural analysis of the Indian finger millet (*Eleusine coracana*) bi-functional inhibitor of alpha-amylase/trypsin with 122 amino acids has shown five disulphide bridges and a trypsin-binding loop (Gourinath *et al.*, 2000). These structural analyses would greatly help in “enzyme engineering” of the native inhibitors to a potent form against the target pest species than the native protease inhibitors.

1.3.2.1.2 Mechanism of action

Basically a binding loop sticking out of the surface of the inhibitor contains an active site and a peptide bond. The inhibitor active site loop fits into protease active site and the inhibitor peptide bond may or may not be cleaved. However, the cleavage and hydrolysis of the inhibitor does not affect the interaction. The inhibitor therefore mimics a normal substrate but does not allow to be completely hydrolysed. Other residues in the vicinity of the interaction function in stabilising the complex, and are important in the strength and effectiveness of the inhibition. The Bowman-Birk inhibitors pose two binding loops and can thus inhibit two

molecules of protease per molecule of inhibitor and are therefore referred to in many publications as double headed (Figure 1.4). The loops (or reactive sites) are known to inhibit trypsin in monocots, while inhibiting trypsin, chymotrypsin and elastase in dicot plants (Mello *et al.*, 2003).

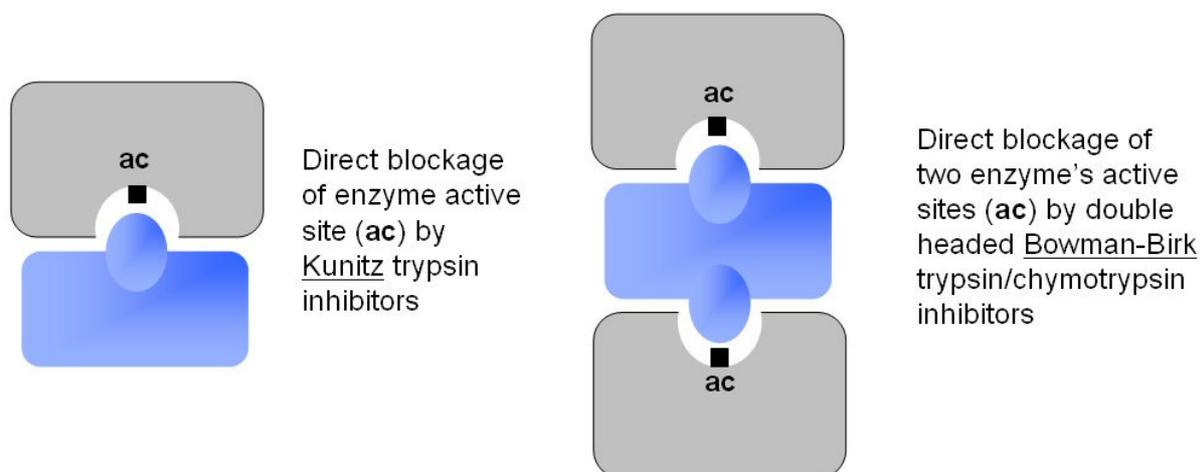


Figure 1.4 Substrate-like mechanism of inhibition by two serine protease inhibitor types, Kunitz and Bowman-Birk (Modified from Bode and Huber, 2000).

1.3.2.2. Cysteine protease inhibitors

Cysteine protease inhibitors, notably include cystatins and are reversible inhibitors of the cysteine class of proteases that include papain and its related proteases (Cathepsin B, H, L, ficin and bromelain). The first cystatin to be isolated was of animal origin and was isolated from chicken egg white (Colla *et al.*, 1989), while oryzacystatin (OC-I) was the first well characterised plant cystatin.

1.3.2.2.1 Classification

Cystatins are a group of related proteins both in structure and function and have been grouped into the cystatin super family. Before the discovery of phytocystatins (plant cystatins),

cystatin members were grouped into three families, the stefins, cystatin (same name as the super family) and the kininogens (Figure 1.5).

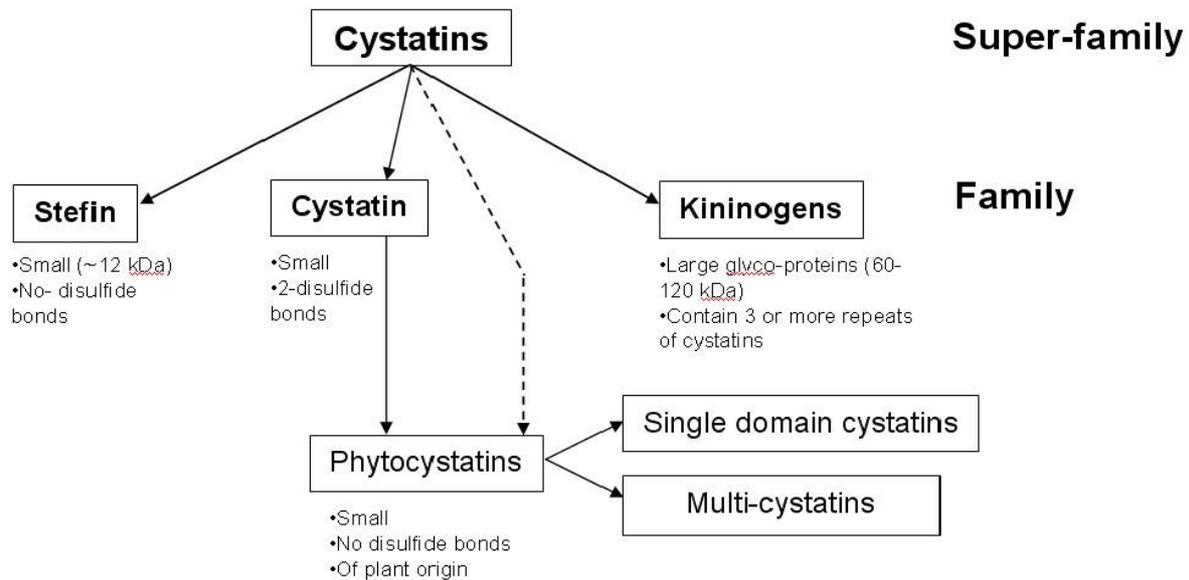


Figure 1.5 General classification of the cystatin super-family phytocystatins (plant cysteine protease inhibitors) are grouped as members of the cystatin family. Broken line indicates re-classification of phytocystatins as a separate family containing single and multidomain cystatins.

The classification in families is based on size, presence or absence of disulfide bonds and on primary amino acid sequence similarities (Figures 1.5 and 1.6). Members of the stefin family are small (approximately 12 kDa), lack both disulfide bonds and carbohydrate groups. The cystatin family contains members that have two disulfide bonds, are glycosylated and have molecular masses ranging from 13-24 kDa. Members of the third family, the kininogens, are large and complex, with sizes ranging from 60-120 kDa. They are known to have several domains in tandem that may have arisen due to two duplications of members of the cystatin family. When new members of the cystatin superfamily were discovered in dicot and monocot

plants, they were grouped into the cystatin family. However, due to their lack of disulfide bonds and also presence of several primary sequence differences, it has been proposed to re-classify plant cystatins into a separate family.

Stefin family

hca ---G-----QVVAG-----
hcb ---G-----QVVAG-----

Cystatin family

cc ---G-----FAM-----QLVSG-----C---C-----C---PW---C-----
hcc ---G-----FAV-----QIVAG-----C---C-----C---PW---C-----
bcc ---G-----FAV-----QVMSG-----C---C-----C---PW---C-----

Kininogen family

hk1 ---G-----FAV-----TVGSD-----C---C-----C---PW---C-----
hk2 ---G-----FAV-----QVVAG-----C---C-----C---PW---C-----

Phytocystatin family

OC-I ---G-----FAVTEHNKKAN-----QVVAG-----PW-----
OC-II ---G-----FAVTEHNKKAN-----QVVAG-----PW-----

Figure 1.6 Alignment of selected members of the four cystatin families illustrating the sequence conservation regions within the family members. The sequences are human cystatin-A (hca), human cystatin-B (hcb), chicken cystatin (cc), human cystatin-C (hcc), beef colostrums cystatin (bcc), human kininogen segment 1 (hk1) and segment 2 (hk2), oryzacystatin-I and oryzacystatin-II (OC-I and OC-II) (Modified from Oliveira *et al.*, 2003)



They can be divided into two major groups, one comprising members of a single domain, such as oryzacystatin-I and II from rice (Abe *et al.*, 1987; Kondo *et al.*, 1990), corn cystatin (Abe *et al.*, 1992), cowpea cystatin (Fernandes *et al.*, 1993), potato cystatin (Hildmann *et al.*, 1992) soybean cystatin I and II (Brzin *et al.*, 1990) and papaya (Song *et al.*, 1995) A second group of phytocystatins comprises members which are of multiple domains, such as sunflower multicystatin (Kouzuma *et al.*, 1996) and potato multicystatin (Waldron *et al.*, 1993; Walsh *et al.*, 1993) (See also Chapter 4 for a complete list and details of other phytocystatins). Purified phytocystatins have molecular masses ranging from 5 to 87 kDa with high stability at temperatures and pH extremes.

1.3.2.2.2 Structure

Oryzacystatin (OC-I), the first phytocystatin to be isolated (Abe *et al.*, 1987), has been well characterised and its crystal structure elucidated. Later, a similar cystatin to OC-I was isolated from rice seed leading to the renaming of OC to oryzacystatin-I (OC-I) and the new homolog oryzacystatin-II (OC-II). Based on the crystal structure of OC-I, phytocystatins are generally characterised by five stranded anti-parallel β -sheets which are a kind of wrap round one side of a central α -helix composed of about five turns (Rawlings and Barret, 1986; Turk and Bode, 1991) (Figure 1.7). Between the anti-parallel β -sheets are two hair-pin loops. The first one consists of the highly conserved QxVxG motif found in all members of the super family, while the sequence in the second loop with a PT motif is less conserved. The N-terminal region is a long arm extending outwards from the rest of the structure. It tends to acquire different conformations depending on the residues as exemplified by the solution structure of OC-I. However, a glycine residue in the N-terminal region is also highly conserved in all members of the super family.

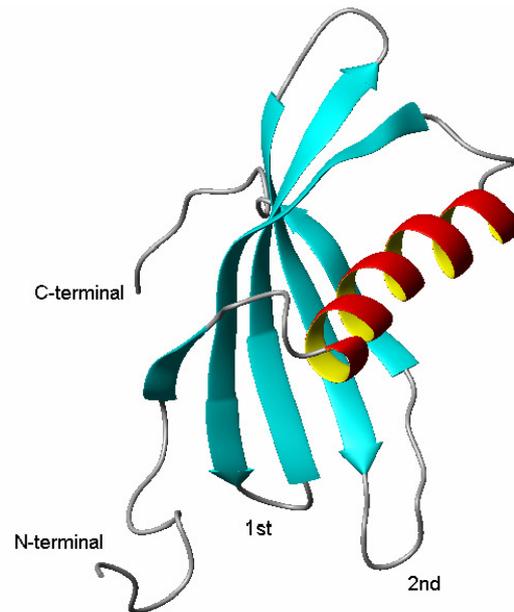


Figure 1.7 The three dimensional structure of OC-I showing the characteristic 5 anti-parallel B strands (blue), the single 5 turn a-helix (red), the N-terminal, the 1st and 2nd hairpin-like loops. Figure was drawn using MolMol version 2k.1.

1.3.2.2.3 Mechanism on interaction with cysteine proteases

Interaction models between cystatins and cysteine proteases have been proposed suggesting three regions of contact. The highly conserved N-terminal region and the two hairpin loops form a kind of wedge (Figure 1.8) which is also highly hydrophobic and complimentary to the active cleft of papain, a model cysteine protease from papaya (Bode *et al.*, 1988).

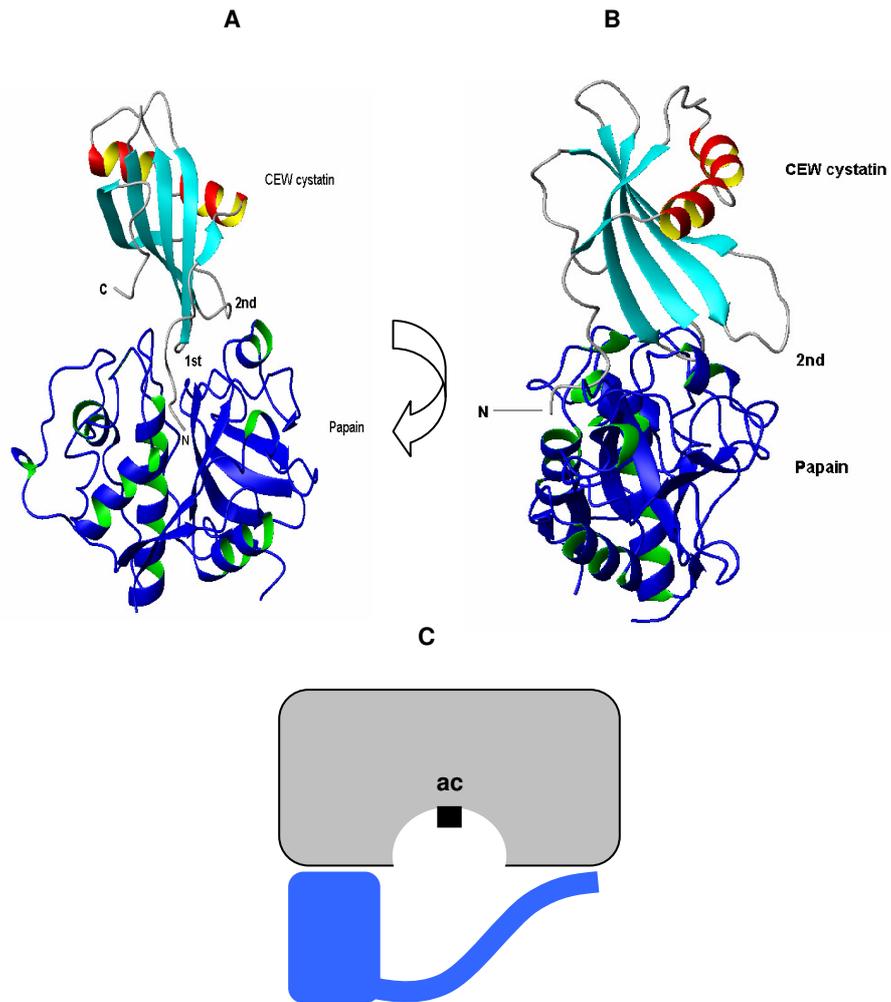


Figure 1.8 Three-dimensional plot showing the complex between papain (blue and green) and chicken egg white cystatin (CEW) colored light blue, red and yellow (PDB accession No. 1STF). (A) Complex is presented in front view to show the V-shaped active site of papain and how the N-terminal region of CEW fits into it. (B) Complex is rotated 90° on a vertical axis to show that the CEW N-terminal actually fits along the surface of the papain active site rather than inside. Note that the 1st cystatin loop fits deeper into the enzyme and therefore being more important the conserved 2nd loop. (C) Partially substrate-like mechanism of cystatin inhibition of cysteine proteases (Bode and Huber, 2000)

Table 1.4 Insect pests with reported susceptibility to phytocystatins, either *in-vitro*, in artificial diet or in transgenic plants

Insect pest	Order: family	Host plant	Phytocystatin	Nature of test	Reference
Alfalfa weevil (<i>Hypera postica</i>) (Gyllenhal)	Coleoptera: Curculionidae	Alfalfa	Oryzacystatin -I	<i>In-vitro</i> assays	Wilhite <i>et al.</i> , 2000
Bean beetle (<i>Callosobruchus chinensis</i>) (Linnaeus)	Coleoptera: Bruchidae	Common bean	Oryzacystatin -I & II	Artificial diet	Kuroda <i>et al.</i> , 2001
Bean bug (<i>Riptortus clavatus</i>) (Thunberg)	Heteroptera: Alydidae	Common bean	Oryzacystatin -I & II	Artificial diet	Kuroda <i>et al.</i> , 2001
Black vine weevil (<i>Otiorynchus sulcatus</i>) (Fabricius)	Coleoptera: Curculionidae	Forestry trees	Oryzacystatin -I	<i>In-vitro</i> assays	Michaud <i>et al.</i> , 1996
Colorado potato beetle (<i>Leptinotarsa decemlineata</i>) (Say)	Coleoptera: Chrysomelidae	Potato	Oryzacystatin -I	Transgenic potato	Lecardonnell <i>et al.</i> , (1999)
Maize grain weevil (<i>Sitophilus zeamais</i>) (Motschulsky)	Coleoptera: Curculionidae	Maize and rice	Corn cystatin (CC)	Transgenic rice	Irie <i>et al.</i> , 1996
Poplar leaf-beetle (<i>Chrysomela tremulae</i>) (Fabricius)	Coleoptera: Chrysomelidae	White poplar	Oryzacystatin -I	Transgenic poplar	Leple <i>et al.</i> , 1995
(<i>Chrysomela populi</i>) (Linnaeus)			Arabidopsis cystatin (<i>Atcys</i>)		Delledonne <i>et al.</i> , 2001
Southern corn rootworm (<i>Diabrotica undecimpunctata howardi</i>) (Barber)	Coleoptera: Chrysomelidae	Maize	Oryzacystatin -I	<i>In-vitro</i> assays	Edmonds <i>et al.</i> , 1996
			Potato cystatin (PCPI-10)	Artificial diet	Fabrick <i>et al.</i> , 2002
Western corn rootworm (<i>Diabrotica virgifera virgifera</i>) (LeConte)	Coleoptera: Chrysomelidae	Maize	Soyacystatin N (ScN)	Artificial diet	Zhao <i>et al.</i> , 1996 Koiwa <i>et al.</i> , 2000
Western flower thrip (<i>Frankliniella occidentalis</i>) (Pergande)	Thysanoptera: Thripidae	Capsicum, Cucumber Carnation, Chrysanthemum	Potato cystatin	<i>In-vitro</i> assays	Annadana <i>et al.</i> , 2002



1.3.2.3 Aspartic and metallo-protease inhibitors

There is far less knowledge on aspartic protease inhibitors and their inhibition in insect digestion. Aspartic proteases (cathepsin D-like proteases) together with cysteine proteases have been reported in species of six families of the order Hemiptera (Houseman and Downe, 1983). The low pH of midguts of many members of Coleoptera and Hemiptera provides more favourable environments for aspartic proteases (pH optima ~ 3-5) than the high pH of most insect guts (pH optima ~ 8-11) (Houseman *et al.*, 1987) where the aspartic and cysteine proteases would not be active. Therefore these inhibitors would be expected in Coleopteran insects. Wolfson and Murdock (1987) demonstrated that pepstatin, a powerful and specific inhibitor of aspartyl proteases, strongly inhibits proteolysis of the midgut enzymes of Colorado potato beetle, *Leptinotarsa decemlineata*. This indicates that an aspartic protease was present in the midgut extract. Aspartic PIs have been recently been isolated from sunflower (Park *et al.*, 2000), barley (Kervinen *et al.*, 1999) and cardoon (*Cyanara cardunculus*) flowers named as cardosin A (Frazao *et al.*, 1999).

At least two families of metallo-protease inhibitors, the metallo-carboxypeptidase inhibitor family in potato (Rancour and Ryan, 1968) and tomato plants (Graham and Ryan, 1981) and a cathepsin D inhibitor family in potatoes (Keilova and Tomasek, 1976), have been identified in plants. The cathepsin D inhibitor (27kDa) is unusual as it inhibits trypsin and chymotrypsin as well as cathepsin D, but does not inhibit aspartyl proteases such as pepsin, rennin or cathepsin E. The inhibitors of the metallo-carboxypeptidase from tissue of tomato and potato are polypeptides (4kDa). They strongly and competitively inhibit a broad spectrum of carboxypeptidases from both animals and microorganisms, but not the serine carboxypeptidases from yeast and



plants (Havkioja and Neuvonen, 1985). This type of inhibitor is found in tissues of potato tubers where it accumulates during tuber development along with the potato inhibitor I and II families belonging to the serine protease inhibitor type. The inhibitor is also induced and accumulates in potato leaf tissues in response to wounding (Graham and Ryan, 1981; Hollander-Czytko *et al.*, 1985). Thus, the inhibitor accumulated in the wounded leaf tissues of potato has the capacity to inhibit all the five major digestive enzymes i.e. trypsin, chymotrypsin, elastase, carboxypeptidase A and carboxypeptidase B of many insects (Hollander-Czytko *et al.*, 1985).

The detailed structural analysis of prophytepsin, a zymogen of barley aspartic protease shows a pepsin-like bilobe and a plant specific domain. The N-terminal has 13 amino acids necessary for inactivation of the mature phytepsin (Kervinen *et al.*, 1999). The aspartic PI cardosin A from cardoon shows regions of glycolylations at Asn-67 and Asn 257. The Arg-Gly-Asp sequence recognises the cardosin receptor, which is found in a loop between two-beta strands on the molecular surface (Frazao *et al.*, 1999).

1.3.3 Regulation of protease inhibitors

Protease inhibitors are expressed in plants in response to wounding, insect herbivory and chemical signals such as jasmonic acid (JA) derivatives (Ryan, 1990; Koiwa *et al.*, 1997). Earlier research on tomato inhibitors has shown that the protease inhibitor initiation factor (PIIF), triggered by wounding or chemical elicitors, switches on the cascade of events leading to the synthesis of these inhibitor proteins (Melville and Ryan, 1973; Bryant *et al.*, 1976), and the newly synthesized PIs are primarily cytosolic (Hobday *et al.*, 1973).



Current evidence suggests that the production of the inhibitors occurs via the octadecanoid (OD) pathway. This pathway catalyzes the break down of linolenic acid and the formation of jasmonic acid (JA) to induce protease inhibitor gene expression (Koiwa *et al.*, 1997). There are four systemic signals responsible for the translocation of the wound response. This includes systemin, abscisic acid (ABA), hydraulic signals (variation potentials) and electrical signals (Malone and Alarcon, 1995). These signal molecules are translocated from the wound site through the xylem or phloem as a consequence of hydraulic dispersal. Systemin, an 18-mer peptide, has been intensely studied from wounded tomato leaves which strongly induced expression of protease inhibitor (PI) genes. Transgenic plants expressing prosystemin antisense cDNA exhibited a substantial reduction in systemic induction of PI synthesis, and reduced capacity to resist insect attack (McGurl *et al.*, 1994). Systemin regulates the activation of over 20 defensive genes in tomato plants in response to herbivorous and pathogenic attacks. The polypeptide activates a lipid-based signal transduction pathway in which linolenic acid is released from plant membranes and converted into an oxylipin signaling molecule, jasmonic acid (Ryan, 2000). A wound-inducible systemin cell surface receptor with an M(r) of 160,000 has also been identified and the receptor regulates an intracellular cascade including depolarization of the plasma membrane and the opening of ion channels thereby increasing the intracellular Ca(2+). This activates a MAP kinase activity and a phospholipase A(2). These rapid changes play a vital role leading to the intracellular release of linolenic acid from membranes and its subsequent conversion to JA, a potent activator of defence gene transcription (Ryan, 2000). The oligosaccharides, generated from the pathogen-derived pectin degrading enzymes i.e. polygalacturonase (Bergey *et al.*, 1999) and the application of systemin as well as wounding have been shown to increase the jasmonate levels in tomato



plants. Application of jasmonate or its methyl ester, methyl jasmonate, strongly induces local and systemic expression of PI genes in many plant species. This suggests that jasmonate has a ubiquitous role in the wound response (Wasternack and Parthier, 1997). Further, analysis of a potato PI-IIK promoter has revealed a G-box sequence (CACGTGG) as jasmonate-responsive element (Koiwa *et al.*, 1997). The model developed for the wound-induced activation of the protease inhibitor II (Pin2) gene in potato (*Solanum tuberosum*) and tomato (*Solanum lycopersicum*) establishes the involvement of the plant hormones, abscisic acid and jasmonic acid (JA) as the key components of wound signal transduction pathway (Titarenko *et al.*, 1997). Levels of ABA have been shown to increase in response to wounding, electrical signal, heat treatment or systemin application in parallel with PI induction (Koiwa *et al.*, 1997). Abscisic acid, originally thought to be involved in the signalling pathway, is now believed to weakly induce the mRNAs of wound response proteins. A concentration even as high as 100 mM induces only low levels of protease inhibitor as compared to systemin or jasmonic acid (Birkenmeiner and Ryan, 1998) suggesting the localized role of ABA.

There is evidence that wound induction, insect and pathogen defence pathways overlap considerably. Expression of wound and JA inducible genes can be positively and negatively regulated by ethylene or salicylic acid (SA), both of which are components of the pathogen-induced signalling pathway (Delaney *et al.*, 1994; Bent, 1996). The expression of thionins in *Arabidopsis* (Epple *et al.*, 1995) and lectin II in *Griffonia simplicifolia* (Zhu-Salzman *et al.*, 1998) was elicited by JA but suppressed by ethylene, showing their opposite effects on the wound signalling pathway.



1.3.4 Structure of protease inhibitor genes

Many protease inhibitors are products of multigene families (Ryan, 1990). The gene size and coding regions of serine inhibitors are generally small with no introns (Boulter, 1993). Bowman-Birk type double-headed protease inhibitors are assumed to have arisen by duplication of an ancestral single headed inhibitor gene and subsequently diverged into different classes i.e. trypsin/trypsin (T/T), trypsin/chymotrypsin (T/C) and trypsin/elastase (T/E) inhibitors (Odani *et al.*, 1983). The mature proteins comprise an easily identifiable ‘core’ region of about 62 amino acids. This covers the invariant cysteine residues and active centre serines, which are bound by highly variable amino and carboxy-terminal regions. The average number of amino acid replacements in this region from all pair-wise comparisons show that the differences between the different classes of inhibitor within a species (around 16.5/62 residues) are much greater than the differences within a class between different species (around 11/62 residues). Considering that 18 of the residues in this region are obligatorily invariant for proteins to be classified as Bowman-Birk type inhibitors, these are very high rates of amino acid substitutions. This highlights the problems likely to be encountered in attempting to draw conclusions about the evolutionary history of the rapidly diverging, multigenic protein families from sequences, which may be paralogous rather than orthologous. Corrected divergence between pair-wise combinations of sequences calculated according to the method of Perler *et al.* (1980) revealed that the average divergence between trypsin-specific and chymotrypsin-specific second domains (about 36%) is very similar to that between the first and second domains (about 40%). On an “evolutionary clock” model this would imply that the gene duplication leading to T/T and T/C families occurred very close to the duplication. This leads to the appearance of the double-headed inhibitors and that the



number of silent substitutions has reached saturation in all these genes (Hilder *et al.*, 1989).

Analysis of the winged bean Kunitz chymotrypsin inhibitor (WCI) protein shows that it is encoded by a multigene family that includes four putative inhibitor-coding genes and three pseudogenes. The structural analysis of the WCI genes indicates that an insertion at a 5' proximal site occurred after duplication of the ancestral WCI gene and that several gene conversion events subsequently contributed to the evolution of this gene family (Habu *et al.*, 1997). The 5' region of the pseudogene WCI-P1 contains frame-shift mutations, an indication that the 5' region of the WCI-P1 gene may have spontaneously acquired new regulatory sequences during evolution. Since gene conversion is a relatively frequent event and the homology between the WCI-P1 and the other inhibitor genes WCI-3a/b is disrupted at a 5' proximal site by remnants of an inserted sequence, the WCI-P1 gene appears to be a possible intermediate. This could be converted into a new functional gene with a distinct pattern of expression by a single gene-conversion event (Habu *et al.*, 1997). Molecular evolution of *wip-1* genes from four *Zea* species show significant heterogeneity in the evolutionary rates of the two inhibitory loops, in which one inhibitory loop is highly conserved, whereas the second is diverged rapidly. Because these two inhibitory loops are predicted to have very similar biochemical functions, the significantly different evolutionary histories suggest that these loops have different ecological functions (Tiffin and Gaut, 2001). Analysis of OC-I has further revealed the presence of two introns; the first a 1.4kbp region between Ala 38 and Asn 39 and a second region of 372bp in the 3' non coding region (Kishimoto *et al.*, 1994). OC-II, present on chromosome 5, also has introns in



the same positions (Kondo *et al.*, 1991). This suggests deviation from the earlier PIs that lacked introns.

1.3.5 Protease inhibitors and insect control

Protease inhibitor genes have advantages over genes encoding for complex pathways i.e. by transferring single defensive genes from one plant species to another and expressing them either from wound-inducible or constitutive promoters. It thereby imparts resistance against insect pests (Boulter, 1993) and may not interfere with other plant functions as pathway related proteins would. This was first demonstrated by Hilder *et al.* (1987) by transferring the trypsin inhibitor gene from *Vigna unguiculata* to tobacco. This conferred resistance to wide range of insect pests including Lepidopterans, such as *Heliothis* and *Spodoptera*, Coleopterans, such as *Diabrotica*, *Anthonomous*, and Orthoptera such as locust. Further, there is no evidence that it had toxic or deleterious effects on mammals.

These advantages make protease inhibitors an ideal choice to be used in developing transgenic crops resistant to insect pests. Further, transformation of plant genomes with protease inhibitor-encoding cDNA clones appears attractive not only for the control of plant pests and pathogens, but also as a means to produce protease inhibitors useful in alternative systems and the use of plants as factories for the production of heterologous proteins (Sardana *et al.*, 1998).

Additionally, serine protease inhibitors have anti-nutritional effects against several Lepidopteran insect species (Shulke and Murdock, 1983; Applebaum, 1985). Broadway and Duffey (1986) compared the effects of purified soybean trypsin inhibitor (SBTI) and potato inhibitor II (an inhibitor of both trypsin and



chymotrypsin) on the growth and digestive physiology of larvae of *Heliothis zea* (Boddie) (Lepidoptera: Noctuidae) and *Spodoptera exigua* (Hübner) (Lepidoptera: Noctuidae). They demonstrated that growth of larvae was inhibited at levels of 10% of the proteins in their diet. Trypsin inhibitors at 10% of the diet were toxic to larvae of the *Callosobruchus maculatus* (Fabricius) (Coleoptera: Bruchidae) (Gatehouse and Boulter, 1983) and *Manduca sexta* (Linnaeus) (Sphingidae: Sphinginae) (Shulke and Murdock, 1983). However, the mechanism of action of these protease inhibitors towards insect digestive enzymes seems rather complicated and has been a subject of investigation (Barrett, 1986; MacPhalen and James, 1987; Greenblatt *et al.*, 1989). Knowledge on mechanisms of protease action and their regulation *in vitro* and *in vivo* in animals, plants, microorganisms and more recently in viruses have contributed to many practical applications for inhibitor proteins in and agriculture.

The secretion of proteases in insect guts seems to depend upon midgut protein content rather than the food volume (Baker *et al.*, 1984). The secretion of proteases has been attributed to two mechanisms. This involves either a direct effect of food components (proteins) on the midgut epithelial cells, or a hormonal effect triggered by food consumption (Applebaum, 1985). Models for the synthesis and release of proteolytic enzymes in the midguts of insects proposed by Birk and Applebaum (1960) and Brovosky (1986) reveal that ingested food proteins trigger the synthesis and release of enzymes from the posterior midgut epithelial cells. The enzymes are then released from membrane-associated forms and stored in vesicles that are in turn associated with the cytoskeleton. The peptidases are secreted into the ectoperitrophic space between the epithelium. This is a particulate complex (Eguchi *et al.*, 1982) from where the proteases move transversely into the lumen of the gut where the food



proteins are degraded. Protease inhibitors then directly inhibit the protease activity of these enzymes and reduce the quantity of proteins that can be digested. These also cause hyper-production of the digestive enzymes which enhances the loss of sulfur amino acids (Shulke and Murdock, 1983). As a result, the insects become weak resulting in stunted growth and ultimate death.

Isolation of the midgut proteases from the larvae of cowpea weevil, *C. maculatus* (Kitch and Murdock, 1986; Campos *et al.*, 1989) and bruchid, *Z. subfasciatus* (Lemos *et al.*, 1987) confirmed the presence of cysteine mechanistic class of protease inhibitors. Similar proteases have been isolated from midguts of the flour beetle *T. castaneum*, Mexican beetle *E. varivestis* (Murdock *et al.*, 1987) and the bean weevil *A. obtectus* (Wieman and Nielsen, 1988). Cysteine proteases isolated from insect larvae are inhibited by both synthetic and naturally occurring cysteine protease inhibitors (Wolfson and Murdock, 1987). The optimum activity of cysteine proteases is usually in the pH range of 5-7, which is the pH range of the insect gut that uses cysteine proteases (Murdock *et al.*, 1987). Another puzzling aspect of studies with *C. maculatus* is the apparent effects of certain members of Bowman-Birk trypsin inhibitor family on the growth and development of these larvae. Although cysteine protease is primarily responsible for protein digestion in *C. maculatus*, it is not clear, how the cowpea and soybean Bowman-Birk inhibitors are exert their anti-nutritional effects on this organism.

1.3.6. Engineering of protease inhibitors

Despite several studies showing the promise of cystatin in pest control (Urwin *et al.*, 1995; Leplé *et al.*, 1995; Duan *et al.*, 1996; Atkinson *et al.*, 2004), successful use of



these proteins to protect plants remains somewhat limited. The presence of inhibitors in plant tissues, either naturally or engineered, has been shown to induce the synthesis of novel proteases in the midgut of several insects. This is one way of compensating for the loss of proteolytic functions (Jongsma and Bolter, 1997). These compensatory processes, together with the breakdown of inhibitors by alternative proteases in insect guts (Michaud 1997; Girard *et al.*, 1998; Zhu-Salzman *et al.*, 2003) and other variables, such as gut environment changes due to age of insect and diet variation (Mazumdar-Leighton and Broadway 2001), seem to help the target pests to overcome anti-digestive effects of protease inhibitors therefore limiting their effectiveness. The development of effective plant protection strategies based on protease inhibitors necessitates a strategy that takes these variables in consideration. Two strategies have been proposed to overcome this situation. Gene “pyramiding” would develop transgenic plants with more than one gene strategy with either different genes or variants of the same gene. The later strategy has been explored through protein engineering not only to improve activities of inhibitors but also changes their active site configuration. This renders them less recognizable by the insect gut proteases that would have degraded them.

Two principle methods are currently being used to modulate the binding properties of protease inhibitors. This includes random mutagenesis and selection of improved inhibitor variants by molecular phage display (Laboissière *et al.*, 2002; Stoop and Craik 2003) and rational site-directed mutagenesis of amino acids (Mason *et al.*, 1998; Ogawa *et al.*, 2002; Pavlova and Björk 2003). The availability of sequence data of many plant cystatins and structural data of animal cystatins (Bode *et al.*, 1988; Stubbs *et al.*, 1990) has been very instrumental in elucidating the mechanisms of



protease inhibition by cystatins and for guiding rational engineering of cystatin variants with altered specificities and improved inhibition. For example, mutations in the N-terminal trunk of chicken egg cystatin helped to prove the importance of the conserved glycine residue in this unique region of cystatins (Mason *et al.*, 1998; Pavlova and Björk, 2003). Animal cystatin structural models have also been used to understand interactions plant cystatins and their proteases with the aim of identifying potential target amino acids for mutagenesis. Urwin *et al.* (1995) successfully engineered a variant of OC-I by site-directed mutagenesis, in which the residue aspartate 86 was removed from the original sequence, which showed a 13-fold improvement in inhibition of papain.

The observation that most insect resistance in plants is polygenic may be too simplistic to expect that the over-expression of a single native plant gene will provide efficient and sustainable pest resistance. Recent evidence shows, however, that these sophisticated defence mechanisms have been lost during selection for domestication (Carlini and Grossi-de-Sa, 2001). Therefore, one approach would be to optimise a “resistance” gene by protein engineering, or a balanced interaction that involves the simultaneous expression of several protective proteins by using gene pyramiding or multiple resistances engineering (Winterer, 2002).



1.4 Study hypothesis, study aim and objectives

At the onset of this study it was hypothesized that rotease inhibitors and in particular phytocystatins can control the growth and development of banana weevil. It was further hypothesized that engineering of a native phytocystatin improves inhibition of a cysteine protease from the banana weevil. The overall aim of this study was therefore to investigate the suitability of phytocystatins to control growth and development of the banana weevil. To achieve the aim the following objectives were set up:

- (i) To identify the major class of proteolytic activity in the mid-gut of banana weevil larvae so that the usefulness of application of phytocystatins for preventing cysteine protease action in the weevil could be determined.
- (ii) To express and purify a recombinant native phytocystatin that could be incorporated into a feeding assay in order to assess the effect of phytocystatins on the early growth and development of banana weevil larvae.
- (iii) To carry out a phylogenetic, evolutionary, structural and modeling analysis on phytocystatins to predict which amino acid residues can be mutated to improve the inhibition capacity phytocystatins.
- (iv) To use site-directed mutagenesis to generate novel papaya cystatin mutated at various amino acid residues to evaluate novel phytocystatins for improved activity against papain and cysteine protease containing gut extracts of the banana weevil.



CHAPTER 2

Characterization of the digestive proteases in the banana weevil gut and the effects of recombinant phytocystatins on early larval growth and development

Scientific Communications

Conference presentation and proceedings:

Kiggundu A., Kunert K., Viljoen A., Pillay M. and Gold C. 2002. Designing protease inhibitors for banana weevil control, Proceedings of the 3rd International Symposium on Molecular and Cellular Biology of Bananas. INIBAP, Montpellier, France (<http://www.promusa.org/publications/leuven-abstracts.pdf>).

Publication in preperation:

Kiggundu, A., Van der Vyver C., Mukiibi J. M., Michaud D., Viljoen A., Schlüter U., Kunert K. Phytocystatins inhibit digestive cysteine protease activity of the banana weevil *Cosmopolites sordidus* G. (Coleoptera: Curculionidae) to Archives of Insect Biochemistry and Physiology



2.1 Abstract

It is well-documented that insects possess different protease forms used to digest dietary proteins. Therefore, studies to characterize the forms of protease are important to provide the basis for selecting appropriate protease inhibitors likely to be effective in a transgenic approach. In this study the protease activity in the gut of banana weevil was analysed in order to determine the potential of phycocystatins (OC-I and papaya cystatin) for the control of the banana weevil *Cosmopolites sordidus* G. (Coleoptera: Curculionidae). Extracts from complete weevil larvae guts were found to hydrolyse casein at an acidic pH optimum (pH 5.5). Lesser activity was also detected at alkaline pH conditions (pH 8.0). Cathepsin L and B like cysteine proteases were found in the larval gut as shown by hydrolysis of the specific substrates Z-Phe-Arg-MCA and Z-Arg-Arg-MCA, respectively. In addition, activity of trypsin and chymotrypsin-like serine proteases were also detected using the specific substrates Bz-Arg-MCA and N-Suc-Ala-Ala-Pro-Phe-MCA, respectively. OC-I and papaya cystatin produced as a His-tagged fusion protein in *Escherichia coli* and purified by affinity chromatography inhibited cysteine protease activity in the banana weevil gut homogenates by 66.2 and 81.6% and LD50's of 1×10^{-5} ng/ml and 2.1×10^{-5} ng/ml, respectively. A new bioassay was applied to evaluate the effect of OC-I on early growth and development of the larvae. After banana stem disks were vacuum infiltrated with purified OC-I, weight gain per day of larvae was inhibited by 77% at an inhibitor concentration of 0.6mg of cystatin/g fresh weight. This part of the study demonstrated that the banana weevil uses cysteine proteases similar to cathepsin L and B for protein digestion and metabolism in the gut while phycocystatins are potential control agents for banana weevil growth.



2.2 Introduction

Numerous protease inhibitors have been isolated from numerous plants species and there is evidence that they contribute to the natural defense against insect and pathogen attack (Green and Ryan, 1972; Jacinto *et al.*, 1998). Several studies have already demonstrated the effectiveness of protease inhibitors for the control of various pests when engineered into transgenic plants. Lecardonnel *et al.* (1999) found increased resistance to the Colorado potato beetle (*Leptinotarsa decemlineata*) by developing transgenic potatoes expressing OC-I. Furthermore, Newell *et al.* (1995) developed sweet potato plants expressing cowpea trypsin inhibitors and found resistance to the West Indian sweet potato weevil (*Euscepes postfasciatus*).

There are generally two major protease classes in the digestive systems of phytophagous insects, either the serine or the cysteine class. Serine protease activity is characteristic of Lepidoptera, Dictyoptera and Hymenoptera while the cysteine class is characteristic of Odoptera and Hemiptera. Initial investigations had concluded that Coleopteran insects mainly use cysteine proteases (Gatehouse *at al.*, 1985; Murdock *et al.*, 1988). However, from more recent work it appears that a combination of both serine and cysteine proteases is active in this more advanced order (Mochizuki, 1998) suggesting a higher diversity of proteases in these insects.

The objectives of this study were to identify the major classes of proteolytic activity in the gut of banana weevil larvae using *in-vitro* and *in-vivo* assays in order to determine the protease classes present in the weevil. A further objective was to evaluate the potential of OC-I and papaya cystatin to control growth of banana weevil larvae by targeting the cysteine proteases in the weevil gut.



2.3 Materials and methods

2.3.1 Reagents

Azocasein, N- Z-Arg-Arg-7-amido-4-methylcoumarin hydrochloride (Z-Arg-Arg-MCA), Z-Phe-Arg-7-amido-4-methylcoumarin hydrochloride (Z-Phe-Arg-MCA), Z-L-arginine-4-methyl-7-coumarinylamide hydrochloride (Z-Arg-MCA), N-Succinyl-Ala-Ala-Pro-Phe-7-amido-4-methylcoumarin hydrochloride (N-Suc-Ala-Ala-Pro-Phe-MCA), Benzoyl-L-arginine-7-amido-4-methylcoumarin hydrochloride (Bz-Arg-MCA), bovine serum albumin (BSA), trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane (E-64), gelatin (porcine type A), Triton X-100, phenylmethylsulfonyl fluoride, ethylenediamine tetra acetic acid (EDTA), Phenylmethanesulfonyl fluoride (PMSF), trypsin-chymotrypsin inhibitor from *Glycine max* (Soybean) (SBTi) and aprotinin were purchased from Sigma (Aston Manor, South Africa). Recombinant OC-I and OC-II, corn cystatin-II (CC-II), stefin-A from human (HSA) were a gift from Prof. D. Michaud, who expressed them using the S-transferase (GST) gene fusion system (Michaud *et al.*, 1994; Brunelle *et al.*, 1999).

2.3.2 Insect colony and maintenance

Adult banana weevils were collected from banana growers in Kwazulu Natal Province (South Africa) and maintained in the greenhouse at the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa. The weevils were kept in 10 liter plastic buckets and provided with fresh banana stem (pseudostem and corm) material to oviposit. After two days, weevils were moved to a different container to allow development of laid eggs. After one week, corms were dissected to collect 3th to 4th instar larvae. These were quickly stored at -20°C until required.



2.3.3 Gut extractions and protein concentration determination

Frozen larvae were thawed on ice and dissected in cold distilled water under a stereomicroscope to remove whole guts. The guts were then homogenized in liquid nitrogen followed by addition of 0.15M calcium chloride buffer containing 0.1% Triton X-100 at a tissue to buffer ratio of 0.2g/ml of buffer. The mixture was incubated on ice for 30min and then centrifuged at 15,000rpm for 10min. The clear supernatant was collected into fresh tubes and stored at -20°C. For extracts to be used in gelatin SDS-PAGE (see below), guts were homogenized directly in 100µl gelatin-PAGE sample loading buffer (62.5mM Tris-HCl pH 8.0, 2% sucrose, and 0.001% bromophenol blue). The protein concentration of both types of extracts was determined using the Bio-Rad protein assay kit (Bio-Rad, UK), which is based on the Bradford method with bovine serum albumin as the standard.

2.3.4 Determination of pH optima

To determine the pH optima of the crude larvae extracts, protease activity of the extracts was determined using azocasein as a protein substrate as described by Michaud *et al.* (1995). Basically, 50µl (50µg total soluble protein) of a gut extract were mixed with 450µl of assay buffer (0.1M citrate phosphate buffer for pH 4.0; pH 4.5; pH 5.0; pH 5.5; pH 6.0; pH 6.5 and pH 7.0; 0.1M Tris-HCl buffer for pH 7.5; pH 8.0; pH 8.5 and pH 9; 0.1M glycine buffer for pH 9.0, 9.5 and pH 10). All buffers were made to contain 5mM L-cysteine before use. After pre-activating proteases by incubating the mixture for 10min at 37°C, an equal volume of 2% azocasein (in the respective assay buffer) was added and the complete mixture incubated at 37°C for 3hrs. To stop the reaction, 100µl of 10% (w/v) trichloro-acetic acid was added to the mixture and the mixture incubated for 30min at 4°C. Residual azocasein was removed



by centrifugation at 12000rpm for 5min at 4°C. To 1.0ml of the supernatant, 1.0ml of 1N NaOH was added to precipitate the hydrolysed azocasein and finally the absorbance of this solution was determined at 440nm in a spectro-photometer. At this wavelength, one unit of protease activity is defined to be the amount of enzyme required to produce an absorbance change of 1.0 in a 1cm cuvette under the conditions of the assay (Sarath, 1989). Reactions were performed in triplicate on a micro-titre plate.

2.3.5 Fluorometric assay

Protease specific proteolytic activity and inhibition by specific inhibitors were investigated using the substrates Z-Arg-Arg-MCA (specific to cathepsin B), Z-Phe-Arg-MCA (specific to cathepsin L), Z-Arg-MCA (specific to cathepsin H), Bz-Arg-MCA (specific to trypsin) and N-Suc-Ala-Ala-Pro-Phe-MCA (specific to chymotrypsin). These are highly sensitive fluorometric substrates. When hydrolyzed by their specific proteases, bound α -amino 4-methylcoumarin (MCA) is released, which is highly florescent and MCA release is determined using fluorescence spectro-photometry.

Hydrolysis of the specific substrates by the gut extract was monitored using hydrolysis progress curves as described by Salvesen and Nagase (1989). For detection of cathepsin B, L and H like activity, reaction mixtures contained 10 μ l (10 μ g total soluble protein) of the gut extract, 1 μ l (1%) substrate solution in DMSO dissolved in 89 μ l reaction buffer; 0.1M citrate phosphate buffer pH 6.0 with 5mM L-cysteine freshly added for cysteine like activity or 0.1M Tris-HCL pH 8.0 for trypsin and chymotrypsion like activity. Hydrolysis was monitored at room temperature using a



spectro-fluorometer (BMG FluoStar Galaxy) with excitation and emission at 360nm and 450nm, respectively. Reaction rates represented by the slope of the curve were recorded as Fluorescence Units (FU) per unit time. All reactions were performed in triplicate.

Inhibitors for the different protease classes were used to evaluate inhibition of their activity. For that, 1µl of a 1% inhibitor solution (E-64, OCI, OCII, CCII, HSA, STBi, aprotinin and PMSF) prepared in the same reaction buffer was introduced into the protease reaction monitored in the spectro-fluorometer. The reactions were briefly mixed and detection of protease reaction continued until a steady rate was reached. Slope values were determined before addition and after addition of the inhibitor.

2.3.6 Gelatin SDS-polyacrylamide gel electrophoresis

Gelatin SDS-polyacrylamide gel electrophoresis, as described by Michaud (1998) was carried out to quantitatively identify protease activity in gut extracts by visually analyzing gel-separated proteases. Proteins in the gut extracts were separated on a 15% SDS-PAGE which had been co-polymerized with 0.1% gelatin as a protease substrate. After electrophoresis at 4°C and 100V, the gel was incubated in 2.5% Triton X-100 for 30min at room temperature to re-nature the proteases. The gel was then incubated in a proteolysis buffer (0.1M citrate phosphate buffer pH 6.0 and 10mM L-cysteine) at 37°C for 3hrs for protease action. The gel was subsequently transferred to a gel staining solution (25% isopropanol, 10% acetic acid and 0.1% coomassie blue). Protease activity was visualized as clear bands on a blue background. To test inhibition, 5µl of a 1% inhibitor solution of selected proteinacious inhibitors



(oryzacystatin, papaya cystatin, SBTi, aprotinin and EDTA) were pre-incubated with 5µl of extracts for 15min at 37°C before loading 10µl of the reaction mixture.

2.3.7 Cloning of OC-I and PC genes

The strategy followed for cloning the coding sequences of OC-I and papaya cystatin (PC) in frame for protein expression in *E. coli* is outlined in Figure 2.1. The vector system pQE30, 31 and 32 from the QIAexpressionist kit (Qiagen, Germany) was used. These vectors allow tagging the cloned coding sequence to a 6-histidine tag and purification by nickel chelation chromatography.

Coding sequences of OC-I and PC were excised from the cloning vectors pAOCI-3 and pBICYS1 using the restriction enzymes *EcoRI* and *PstI*. The *EcoRI/PstI* fragments were then first cloned into the *EcoRI/PstI* site of pBlueScript (Stratagene, USA) and then as a *BamHI/KpnI* fragment from pBlueScript into the vector pQE31 to achieve in-frame ligation. This sub-cloning procedure created the vectors pQOC-I and pQPC, which were transformed into *E. coli* cells (strain JM109) and stored in 10% glycerol stocks at -80°C. In a final step, vectors pQOC-I and pQPC were transferred into competent *E. coli* cells of strain M15 for expression according to the QIAexpressionist kit users manual (Qiagen, Germany).

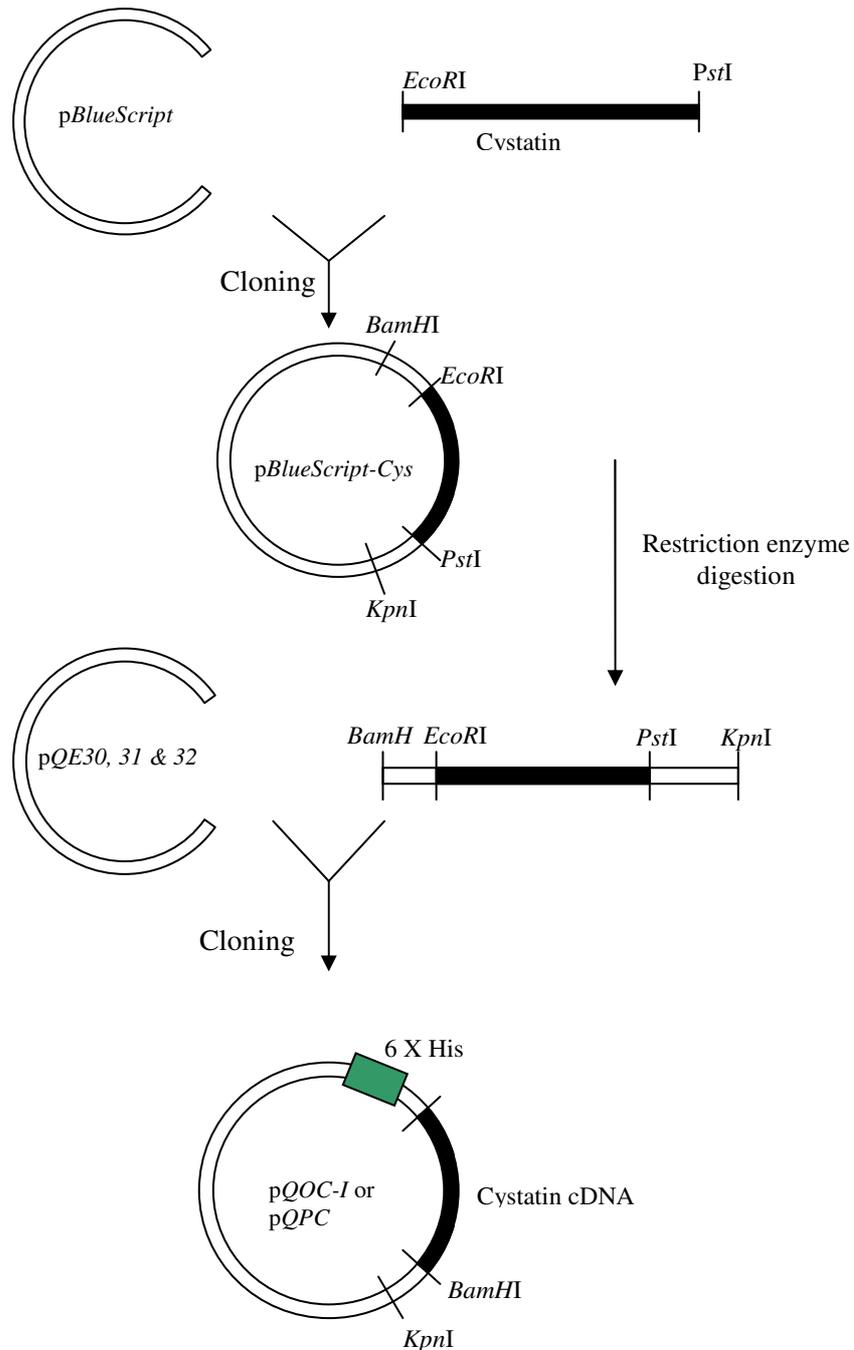


Figure 2.1 Schematic diagram of the construction of expression vectors *pQOC-I* and *pQPC* used in the study to express OC-I and PC in *E. coli*, respectively.



2.3.8 Protein expression and purification

Luria-Bertani (LB) medium (5ml) consisting antibiotics (100µg/ml kanamycin and 25µg/ml ampicillin) was inoculated with a single bacterial colony of M15 cells containing either *pQOC-I* or *pQPC-I* and grown overnight at 37°C under shaking at 210rpm. Pre-warmed LB medium (100ml) with antibiotics (100µg/ml kanamycin and 25µg/ml ampicillin) in a 250ml conical flask was inoculated with 5ml of the overnight culture and incubated at 37°C under shaking at 210rpm for 1hr. Isopropyl-β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 1mM to induce protein expression and bacterial cell growth in the presence of IPTG continued for another 4hr for PC expression or for 12hr for OC-I expression. Cells were finally harvested by centrifugation at 13000rpm in an Eppendorf 5414 Bench top Centrifuge at 4°C for 10min and then stored frozen at -20°C until purification.

The two fusion proteins were purified according to the standard protocol provided in the QIAexpressionist kit manual (Qiagen, Germany). For that, frozen cell pellets were thawed on ice for 30min, re-suspended in his-tag lysis buffer containing 50mM sodium di-hydrogen phosphate, pH 8.0; 300mM sodium chloride and 10mM imidazole at a rate of 2ml buffer per 1mg of cells and 1mg of lysozyme was added. This was mixed gently and incubated on ice for 1hr. The cell suspension was then sonicated using a Cell Disruptor B-30 sonicator (Branson Sonic Power Co./SmithKline Co.) fitted with a standard micro-tip and set to 20% duty cycle, 2 output control in pulse mode. The cells were sonicated using 10 bursts with 10sec cooling on ice between each burst, taking care not to create much frothing. Lysates obtained were centrifuged at 10,000rpm for 30min at 4°C in an Eppendorf centrifuge and the clear supernatant transferred into fresh Eppendorf tubes to which 800µl of



50% Ni-NTA slurry (Qiagen, Germany) was added. The tubes were shaken at 200rpm for 30min at 4°C after which the cell lysate mixture was poured into a short plastic column (setup with a 2ml syringe and a glass wool plug at the bottom) with a paper bottom cover in place. The cover was removed after the slurry settled and the flow-through was collected. Twice 1ml washing buffer (50mM sodium di-hydrogen phosphate, pH 8.0; 300mM sodium chloride; 50mM imidazole) was carefully poured over the column and the buffer was collected at the bottom. This was followed by pouring slowly four times 500µl elution buffer (50mM sodium di-hydrogen phosphate, pH 8.0; 300mM sodium chloride; 250mM imidazole) over the slurry. The elutions were collected separately in 500µl fractions. Five micro-liters of each fraction (flow-through, washes and elution fractions) were each added to 5.0µl SDS-PAGE sample buffer (6% β-mercaptoethanol, 6% SDS, 0.6% bromophenol blue, 20% glycerol) boiled for 10min and loaded onto a 15% polyacrylamide gel for evaluation of the purification process and detection of the recombinant proteins. The protein concentration of the elution fractions was finally determined using the Bio-Rad protein assay kit (Bio-Rad, U.K), and fractions were stored in aliquots at 4°C until required.

2.3.9 In-vitro assays with recombinant phytocystatins

Assays were carried out using a modified method as described by Abrahamson (1994). For that gut extract samples containing 0, 10, 20, 50, 100 and 150µg/µl soluble protein were placed into micro-centrifuge tubes and diluted to 250µl with 0.1% Tween-20. Proteolysis buffer (125µl) was then added (340mM sodium acetate, 60mM acetic acid, 4mM di-sodium EDTA, pH 5.5) and then 8mM DTT was added shortly before use. The mixture was pre-incubated in the presence or absence of the



inhibitor for 1min in a water bath (37°C) before 125µl of substrate (20µM Z-Phe-Arg-AMC, prepared by diluting a 1mM stock in DMSO) was added. Incubation was continued for exactly 10min after which the reaction was stopped by the addition of 1 ml stopping reagent (10mM sodium monochloroacetate, 30mM sodium acetate, 70mM acetic acid, pH 4.3). The fluorescence of released AMC was determined with the use of a fluorescence spectrophotometer (Hitachi, Model F-2000) with excitation and emission set at 370nm and 460nm, respectively.

2.3.10 Infiltration of banana stem with phytocystatin

Banana inner stems, which form part of the fruit (bunch) stalk but running in the centre of the pseudo-stem from the bunch to underground stem (corm), were collected fresh from the field. In the laboratory, the stem was cut into 1cm disks and dipped into hot 60°C sorbic acid solution (1%) used as a preservative to prevent rapid oxidation and deterioration of the stem disks. Disks were then wrapped into polythene bags and stored at 4°C. Inhibitor solutions for infiltration were prepared by diluting purified his-tagged OC-I and PC to a 10µg/ml solution and 2ml of this solution was placed into a 5cm diameter petri-dish. As a negative control, elution buffer was used in the same way as the inhibitor solutions. Three 4cm long and 1mm diameter thick plastic rods were placed into the petri-dish. One banana stem disk was placed on the rods to prop the disk just above the bottom of the dish to provide the tissue uniform contact with the solution (Figure 2.7A). The complete set-up was then placed into a vacuum desiccator and the desiccator was attached to a vacuum pump (Savant SC100 SpeedVac equipped with a Savant RT100 refrigerated condensation trap). Vacuum was then applied until bubbling was observed on the surface of the tissue and on the solution. The vacuum was then rapidly removed by unplugging a conveniently placed



(between the desiccator and the pump) tap plunger. This caused the liquid to be drawn into the tissue rapidly. The tissue was removed and placed onto a paper filter in a clean petri-dish and a newly hatched banana weevil larvae were placed in a small hole made on the disk. The treated disks were stored in the dark at 25°C. After 10 days, when the disks were almost decayed, the larvae were dissected out. They were weighted and their head capsule lengths measured (dorsal inter-ocular plane) under a stereo microscope to determine their instar stage as described by Gold *et al.* (1999a).

2.4 Results

2.4.1 pH optima

The optimal hydrolysis of a general protein substrate, azocasein, by banana weevil larval gut homogenates was found to range from pH 5.5 to pH 7.0 with a peak at pH 6.5. There was also a smaller hydrolysis peak (pH 8.5) indicating the presence of both acidic and alkaline proteases in the weevil larval gut (Figure 2.2). Hydrolysis at pH 6.5 was at least 2.5-fold higher than that at pH 8.5. This suggests that acidic cysteine proteases were more predominant in the gut extracts.

2.4.2 Fluorometric assays

To further elucidate on the nature of cysteine and or serine proteases, activity assays were carried out using specific fluorescent substrates. Reaction rates were monitored by detection of the fluorescent product MCA. Two types of cysteine proteases, cathepsin-L and cathepsin-B, were the predominant cysteine protease types producing reaction rates of >1000FU/sec/μg total protein compared to <400FU/sec/μg total protein from cathepsin-H, trypsin and chymotrypsin (Figure 2.3). Of the serine

proteases, trypsin showed the lowest significant proteolysis at 177FU/sec/ μ g total protein (Figure 2.3).

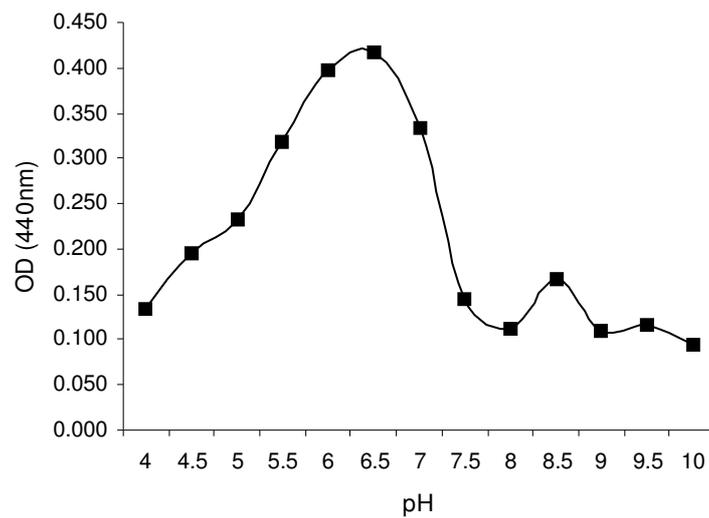


Figure 2.2 Effect of pH on the hydrolysis of azocasein by banana weevil larval gut proteases. Proteolysis was stopped by the addition of 1.0ml of 1N NaOH and the OD of the solution determined at 440nm. At this wavelength, one unit of protease activity is defined to be the amount of enzyme required to produce an absorbance change of 1.0. Reactions and measurements were performed at room temperature. Experiment was repeated twice and values shown are the mean of 3 individual experiments.

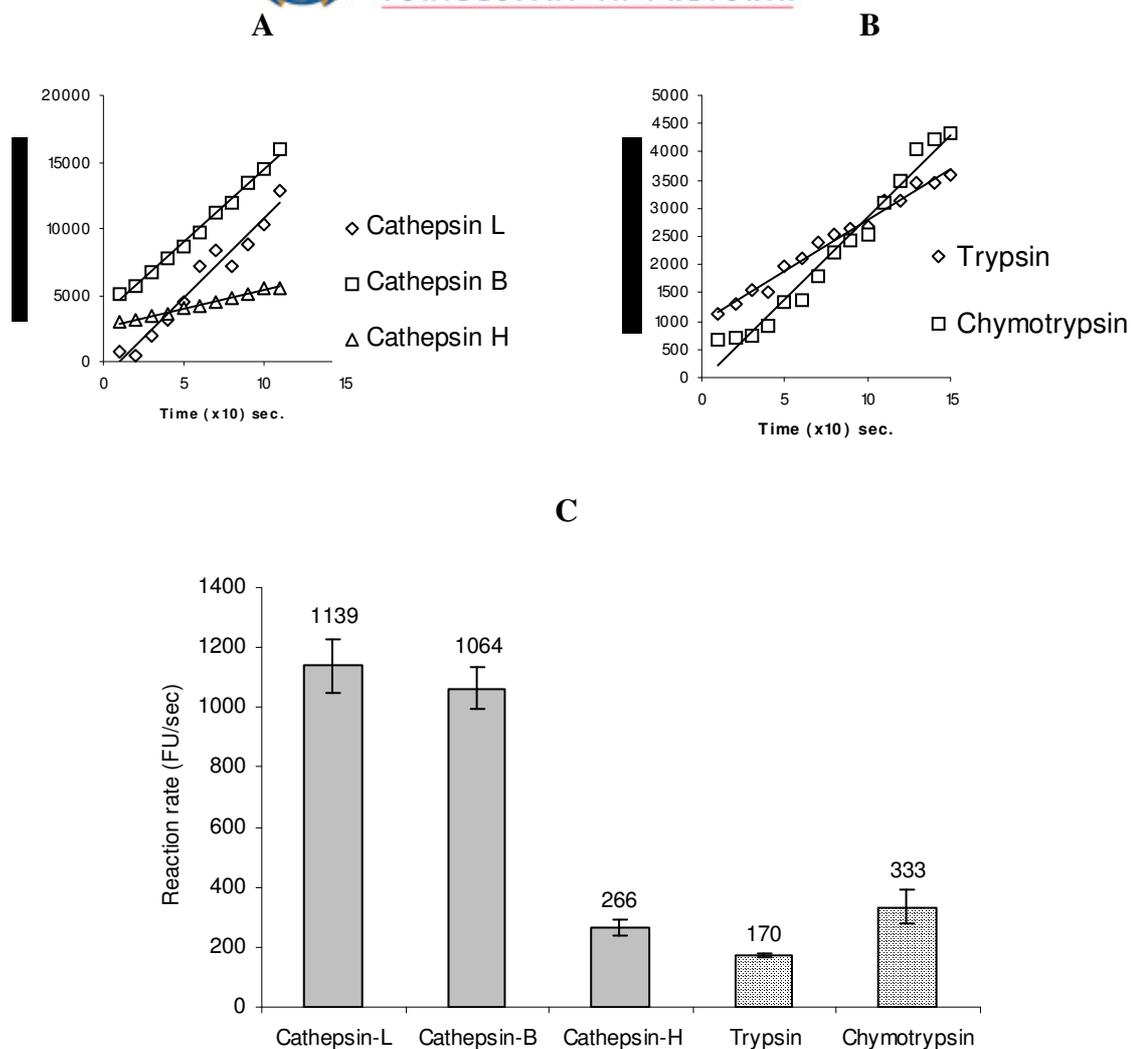


Figure 2.3 (A) Cathepsin B, L and H like activities detected in banana weevil larval gut extracts. Fluorometric assays were conducted using Z-Arg-Arg-MCA, Z-Phe-Arg-MCA and Z-Arg-MCA as substrates for cathepsin B, L and H like activities, respectively at pH 6.0. (B) Trypsin and chymotrypsin-like activities detected in the same extracts, using Bz-Arg-MCA and N-Suc-Ala-Ala-Pro-Phe-MCA as substrates respectively performed at pH 8.0. (C) Maximum activities of all proteases tested. Data points and graphs shown represent the means of three replications \pm SE.



The effects of selected inhibitors were accessed using a banana weevil gut extract and cysteine and serine protease specific substrates. E-64 was the most potent inhibitor of cathepsin L and B like activity with 96% and 85% inhibition of protease activity, respectively. OC-I was the most potent natural plant cysteine protease inhibitor of cathepsin L and B-like activity with 81% and 80% inhibition of protease activity, respectively (Table 2.1A). The soybean trypsin-chymotrypsin inhibitor was most the potent inhibitor against trypsin and chymotrypsin-like activity with 92% and 98% inhibition, respectively. Aprotinin, a serine protease inhibitor, showed lower inhibition of chymotrypsin-like activity when compared to inhibition of trypsin-like activity (Table 2.1B).

2.4.3 Gelatin SDS-polyacrylamide gel electrophoresis

The use of gelatin-containing PAGE gels offers a visual assessment of the protease profile in a crude extract by separating the proteases into their individual constituents. This provides a more detailed profile of protease activity. Extracts were therefore pre-incubated with selected inhibitors before separation on a 15% SDS-PAGE. Figure 2.3 shows that extracts contained at least five different proteases with different molecular sizes of 22, 25, 30 72 and 170kDa.



Table 2.1 Inhibition of banana weevil gut proteases by (A) cysteine and (B) serine protease inhibitors. Cysteine protease inhibitors tested were E-64, OC-I, OC-II, corn cystatin (CC-II) and human stefin A (HSA). Serine protease inhibitors tested were soybean trypsin and chymotrypsin inhibitor (STBi), aprotinin and phenylmethylsulphonylfluoride (PMSF). Reactions for serine protease inhibition were performed in 100mM Tris-HCl buffer (pH 8.0) at room temperature. Proteolytic activity was measured as a rate of reaction indicated by fluorescence units (FU) produced per second per μg of protein (FU/Sec/ μg). Control represents reaction in substrate without addition of an inhibitor. Data represent the mean of three replications $\pm\text{SE}$.

A

	Cathpesin L-like		Cathepsin B-like	
	FU/sec/ μg	Inhibition (%)	FU/sec/ μg	Inhibition (%)
Control	1278.2 \pm 131	-	1151.1 \pm 59	-
E-64	47.4 \pm 73	96	175.2 \pm 19	85
OCI	238.3 \pm 59	81	228.3 \pm 6	80
OCII	394.1 \pm 106	69	495.0 \pm 10	57
CCII	225.6 \pm 36	82	306.0 \pm 23	73
HSA	294.5 \pm 30	77	593.7 \pm 20	48

B

	Trypsin-like		Chymotrypsin-like	
	FU/Sec/ μg	Inhibition (%)	FU/Sec/ μg	Inhibition (%)
Control	185.0 \pm 15	-	297.8 \pm 32	-
SBTi	14.5 \pm 1	92	7.3 \pm 10	98
Aprotinin	18.3 \pm 11	90	168.9 \pm 15	43
PMSF	108.9 \pm 14	41	66.3 \pm 6	78

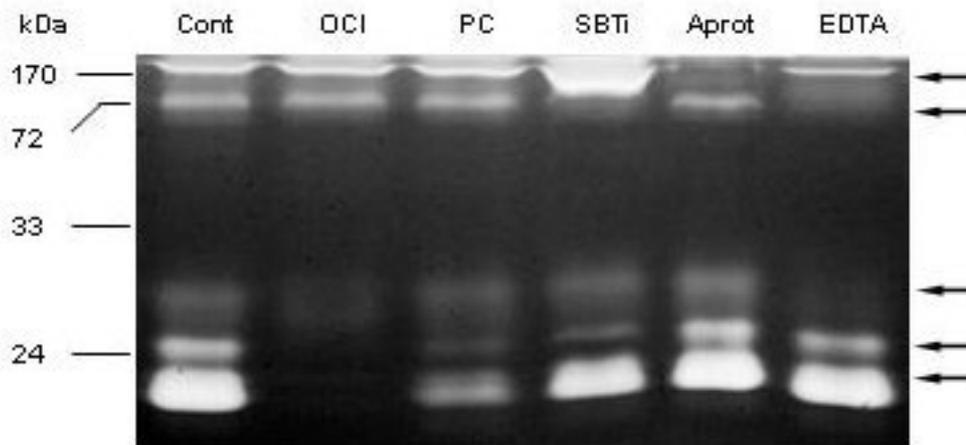
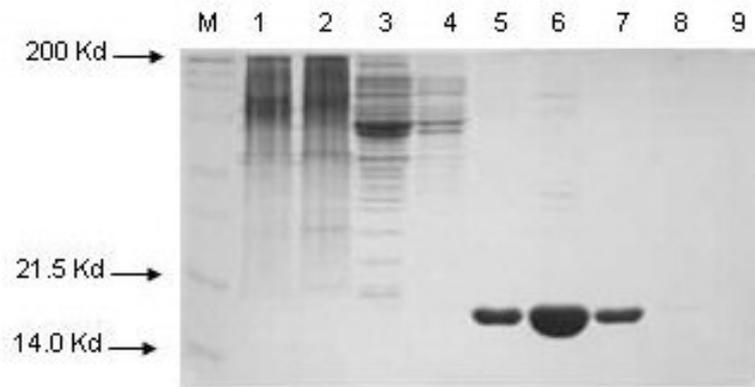


Figure 2.4 The effect of protease inhibitors on the proteolysis activity of banana weevil larval gut proteases revealed by separation in a mildly denaturing 15% SDS-PAGE co-polymerized with gelatin. Protease activity measurement was carried out in a buffer containing 100mM citrate phosphate and 10mM mercaptoethanol, pH 6.0. Visible clear bands indicate proteolysis of gelatin. (Cont) represents activity of crude gut extract (3 μ l from a 200 μ g/ μ l solution); (OC-I) pre incubation with 20 μ g of OC-I; (PC) pre-incubation with 20 μ g papaya cystatin; (SBTi) pre-incubation with 1% soybean trypsin-chymotrypsin inhibitor; (Aprot) pre-incubation with 1% aprotinin and (EDTA) pre-incubation with 100mM EDTA. Arrows indicate major protease activities.



The most potent inhibitor in this assay was OC-I followed by PC, which had not been used in the previous fluorometric assay. Both cystatins reduced the protease activity profile from 5 to 3 bands including a major activity band at 22kDa (Figure 2.4). The soybean trypsin-chymotrypsin inhibitor (SBTi) inhibited one band at 72kDa while aprotinin inhibited one band at 170kDa. EDTA showed inhibition of one band at 30kDa. This suggests the presence of some metallo-proteases in the gut extract. Figures 2.5A and B show the expression and purification of both PC and OC-I as his-tagged fusion proteins, respectively. Protein bands with the expected size of about 15.0Kd for PC and 18Kd for OC-I were found. However, expression levels of OC-I were much lower than of PC. This required the OC-I cultures to be incubated in the presence of IPTG for a longer time period (12hr) when compared to PC (4hr). Reduced growth might be due to OC-I toxicity in *E. coli*. Purified his-tagged PC and OC-I reduced cysteine protease (cathepsin L like) activity of weevil larval gut extracts by 66.2 and 81.6%, respectively (Figure 2.6). The calculated LD50 of inhibition by PC was 2.1×10^{-5} ng/ml and for OC-I 0.1×10^{-5} ng/ml.

A



B

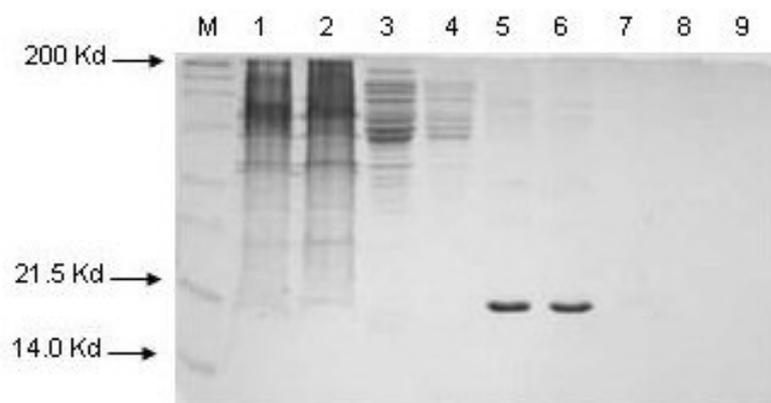


Figure 2.5 SDS-PAGE of (A) PC and (B) OC-I at different purification steps. Lane 1 represents a broad range protein marker (BioRad); lane 2 non-induced and lane 3 IPTG-induced proteins expressed in *E. coli* cells; lane 4 flow-through from Ni-NTA column; lane 5 wash-through from column and lanes 6-9 are four consecutive elutions from column used for purification. Slower migration speeds of recombinant cystatins may be due to presence of his-tags and high concentration of imidazol in the elution buffer.

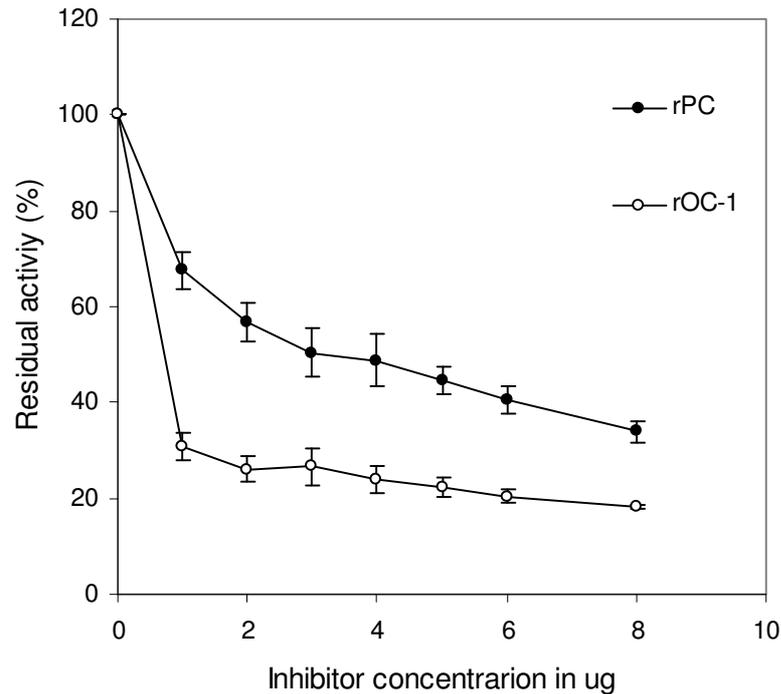
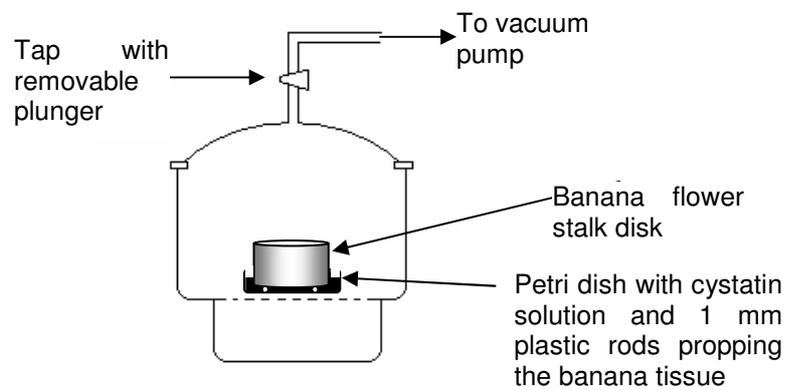


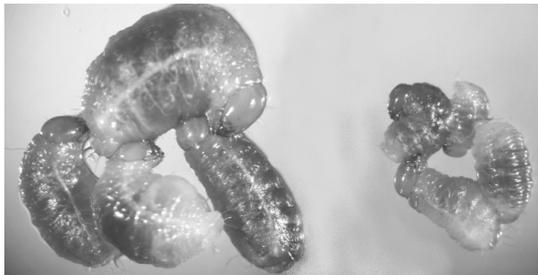
Figure 2.6 The effect of recombinant cysteine protease inhibitors rOC-I and rPC on the cysteine protease activity of banana weevil mid-gut extracts using Z-phe-arg-AMC as substrate.

When larvae fed on banana disks infiltrated with both purified phycocystatins, their development was significantly reduced. Early larval developmental rate was reduced for both phycocystatins. Body weight gain was 0.25mg/day for OC-I and 0.35mg/day PC compared to 1.1mg body weight gain per day in the control larvae (Figure 4.4 C). This represents a reduction in development of 77% for OC-I and 68% for PC at a concentration of 0.6mg/g fresh weight of infiltrated stem disk after re-extraction (Figures 2.7 B and C). However, there was no significant difference in weight gain / day between OC-I and PC-treated larvae ($p>0.05$).

A



B



C

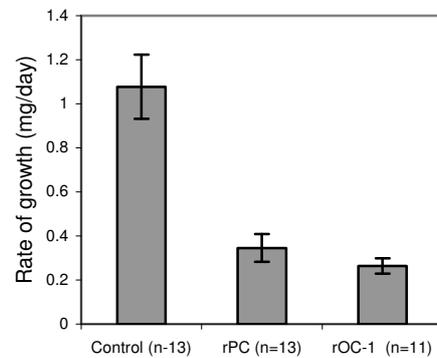


Figure 2.7 (A) Illustration of the apparatus used to vacuum infiltrate banana flower stalk disks with cystatin solution. (B) Larvae on the left after developing on cystatin-free (control) disks for 10 days, while larvae on the right developed in cystatin treated disks over the same period. (C) The growth rate of larvae that were reared on banana stem disks vacuum-infiltrated with a 100ug/ml (to give a final 0.6mg of recombinant protein per disk) solution of recombinant rOC-I and rPC. Values represent means of 39, 38 and 22 replicates for control, rPC and rOC-I, respectively



2.5 Discussion

Many efforts to develop insect resistance in a plant via the expression of protease inhibitors have resulted only in a few successes (Winterer, 2002). Several studies have shown that many insects have more than one protease forms and their activity in gut protein digestion and metabolism is influenced by several factors (Gatehouse *et al.*, 1993), such as gut pH (Michaud *et al.*, 1993), larval stage (Orr *et al.*, 1994) and the quantity and quality of the protein diet (Burgess *et al.*, 1991). This study has provided first evidence that the banana weevil larval, the most destructive stage of the pest's life cycle, expresses a variety of proteases, including cysteine proteases, in its gut. This protease can be blocked by phytocystatins. In contrast, this study showed that serine and metallo-proteases very likely play a less prominent role in protein digestion by larvae. Any strategy to use protease inhibitors to target the banana weevil needs therefore to consider that the weevil possesses more than one protease class. In this study strong evidence has been further provided that both OC-I and PC are able to control the development of the banana weevil by blocking gut cysteine proteases. Significant reduction in the body weight and thus rate of growth of the larvae due to inhibitor action contributed to the larvae underfeeding and interference with protein digestion and metabolism. However, it has always to be considered that even if serine types seemingly play a less significant role in the gut profile when compared to cysteine proteases, weevils might switch to serine type proteases to overcome the presence of cystatins in the diet.

Since there are no transformed banana plants available yet to express an endogenous phytocystatin, the developed vacuum infiltration assay was a very useful and simple tool to access the effects of phytocystatins on growth and development of the banana



weevil larvae. There is even further potential to scale up this infiltration assay so that the assay period is extended to the pupa and adult stage of the larvae. Although this experiment demonstrated the potential of phytocystatins to block weevil development, the infiltration experiments were carried out with a relatively high inhibitor concentration of 0.6mg/g of fresh weight after re-extracting the recombinant proteins from infiltrated stem disks. Such concentrations are difficult to achieve in transgenic approaches to effectively extenuate pest insect gut proteases. However, this study used a single dose of the phytocystatin whose effect may have deteriorated with time of culture. Further experiments have to demonstrate if a lower phytocystatin but continuously expressed in a transgenic plant might result in a similar growth inhibition.

Overall, this study confirmed that cysteine proteases are important protein digestive enzymes in the gut of the banana weevil. The two phytocystatins studied are able to significantly reduce developmental success of the banana weevil larvae. The newly developed bioassay system has been found to be a useful tool for testing bioactive compounds on banana weevil larvae growth and development. Finally, this study provided also first evidence that a transgenic strategy to use protease inhibitors expressed in banana in the control of the banana weevil is plausible. Due to the presence of both cysteine and serine protease in the gut, this study also suggests that simultaneous expression of cysteine and serine protease inhibitors might be a strategy to prevent larvae growth and development.



CHAPTER 3

Phylogenetic and structural comparisons of phytocystatins: A bioinformatics approach

Scientific Communications

Conference presentation and proceedings:

Kiggundu A., Kunert K. and Michaud D. 2005. The N-terminal trunk of plant cystatins determines their inhibitory specificity against cysteine proteases. Proceedings of Plant Canada 2005 Conference, June 15th -18th, Edmonton Canada.

Phylogenetic analysis and structural modelling from this study contributed to the publication:

Girard C., Rivard D., Kiggundu A., Kunert K., Gleddie S. C., Cloutier C., and Michaud D. 2007. A multicomponent, elicitor-inducible cystatin complex in tomato, *Solanum lycopersicum*. New Phytologist 173 (4), 841–851.

Manuscript in preparation:

Kiggundu A., Kunert K., Viljoen V., Van de Vyver C., and Michaud D. Phylogenetic and structural comparisons of phytocystatins: A bioinformatics approach.



3.1 Abstract

With the use of bioinformatics tools the phylogenetic relationships of phytocystatins based on amino acid sequence information was elucidated and their secondary and tertiary structures were investigated for structural comparisons. Sixty six distinct phytocystatins from 43 plant species and 5 different tissue types were investigated. Inhibition constants for inhibition of the model cysteine protease papain varied greatly from 0.00011nM for chelidocystatin to 19,000nM for a soybean cystatin. Phytocystatins could be divided into five distinct phylogenetic groups but their structural features were highly conserved. Amino acid sequence similarities ranged from 7 to 94%. A new highly conserved amino acid sequence motif, YEAKxKxWxKxF, in the C-terminal end being unique to phytocystatins was identified. The predicted 3D homology models showed a high conservation of the general central structure of the phytocystatins i.e. the 4-5 anti-parallel β -sheets, wrapping halfway round a single central α -helix, and particularly the three active site regions, the N-terminal, the 1st and 2nd hairpin loops. Any structural differences seem to be mainly in the length of the N and C terminal, the length of the 2nd hairpin loop and the 5th β -sheet. Via docking experiments, small heterogeneties were observed in the vicinity of the OC-I active sites that seemed to be influential in the binding process and stability of the resultant inhibitor-protease complex.



3.2 Introduction

Phytocystatins are proteinacious inhibitors of plant origin that inhibit specifically cysteine proteases by forming tight reversible bonds thus preventing the hydrolysis of proteins by proteases. The cystatin super family is subdivided into three families based mainly on the three criteria sequence homology, presence of disulfide bonds and on the molecular mass of the protein. These families are the stefins, cystatins and kininogens. Many different phytocystatins have been isolated from different plants and their gene sequences deposited on public databases.

Phylogenetic analysis provides an insight into the molecular evolution of proteins. Numerous bioinformatic and computational biology tools are now available online providing automated analysis of relationships of proteins at molecular and structural level. Public sequence databases have also provided a very useful and wide range of resources to perform such analyses. One of the key ideas in genomic bioinformatics is the concept of homology. This is used to predict the function of genes and proteins. This is followed by a next level where not only protein function can be predicted but also ere the primary, secondary and tertiary structures of a protein can be predicted. This is achieved through powerful computation methods referred to as *in-silico* analysis. Such analysis provides a better understanding of the microstructures on the protein surface that contribute or may even hinder its proper function.

Part of the aim of this study was therefore to analyse, based on available amino acid sequence information, the phylogenetic relationships of phytocystatins. This was carried out by a comparative study on the primary, predicted 2D and 3D structures of known phytocystatins. In particular the 3D positions of the amino acids involved in

binding, structures of active sites and the local structural variation among members of the proposed phytocystatin family were studied.

3.3 Materials and methods

3.3.1 Sequence analysis

Amino acid sequences of phytocystatins were obtained from various online databases (Table 5.1) using the sequence retrieval system (SRS) (<http://srs.embl-heidelberg.de:8000/srs5/>). The program BLAST (Altschul *et al.*, 1990) was used against the GenBank database to further obtain recent submissions that may not have reached the more advanced databases like European Molecular Biology Laboratory (EMBL) sequence database and the Protein Information Resource (PIR) database.

Multiple alignments were performed using the program CLUSTALX (Thomson *et al.*, 1997) with default settings and the alignment edited manually. Long sequences were truncated both at the N and C terminal to include only the domain region and the alignment was repeated. A consensus sequence and a PAM250 (Gonnet *et al.*, 1992) sequence similarity matrix were generated using BIOEDIT suite (Hall, 1999).

Phylogenetic inference was performed using the PHYLIP version 3.5 suite (Felsenstein, 1989). First a distance matrix was generated using the PRODIST program followed by the neighbour joining method using NEIGHBOR program. A consensus tree derived after 1000 bootstraps through the programs BOOTST and CONSES. An un-rooted phylogenetic tree was constructed using the TREEVIEW program (Page, 1996).



3.3.2 Protein structure modelling

The coordinate files (pdb) for OC-I and papain were obtained from the protein data bank (PDB) database. The OC-I pdb file was used to predict the 3D structures of selected representative from each of the phylogenetic groups. Structure modelling to predict the unknown structures was done using the program MODELLER (Sanchez and Sali, 2000) that determines structure using the satisfaction of spatial constraints. The input files consisted of the pdb file of OC-I and the amino acid sequence alignment between OC-I with the unknown sequence at greater than 30% sequence similarity with OC-I. Predicted models were evaluated for energy distribution. Stereochemical quality of the predicted structures was tested using the ENERGY command on MODELLER and PROCHECK (Laskowski, 1993) programs, respectively. Structures were visualised using both SWISS-PDB Viewer (Guex and Peitsch, 1997) and PYMOL (www.pymol.org).

3.3.3 Active site and docking

Based on the X-ray crystal structure of recombinant human stefin-B and papain (Studds *et al.*, 1990), the structure of OC-I bound to papain was extensively modelled manually followed by refinement using MULTIDOCK program in the 3D-DOCK suite (Jackson *et al.*, 1998). Since the binding structural motifs in stefin-B and animal cystatins are present in OC-I and in other phytocystatins, it was expected that OC-I would bind papain in the same manner (Nagata *et al.*, 2000).

3.4 Results

To date, phytocystatins have been isolated from at least 43 different plant species (Table 3.1) but the rate at which new members are identified and isolated is rapid. In

this study a total of 66 phytocystatins have been collected either deposited in sequence databases or reported in the literature.

For some of the known phytocystatins characterisation studies including inhibition kinetics have been carried out either on wild-type proteins extracted directly from the plant or recombinant proteins expressed and purified in the laboratory. The known inhibition constants (K_i) for the model cysteine protease papain are included in Table 3.1. Papain K_i values of known phytocystatins ranged from 0.00011nM for *Chelidonium majus* L. (Celandine-chelidocystatin) to 19,000nM for soybean domain L1 cystain, respectively. The celandine plant, from which the most potent phytocystatin known was found, is traditionally used in China and Europe as a herb to treat bacterial and viral infections (Rogel *et al.*, 1998) in humans.

Table 3.1 Known phytocystatins obtained from sequence databases: EMBL= European Molecular Biology Laboratory, PIR=Protein information Resource, SP=SwissProt, GB=GeneBank and NCBI=National Centre for Biotechnology Information.

Code	Common name	Specie name	Database ¹ (Acc No.)	Papain K _i	Reference ²
Apple	Apple	<i>Malus domestica</i>	EMBL:AY173139	0.2-0.3nM	Ryan <i>et al.</i> , 1998
AraI	Arabidopsis	<i>Arabidopsis thaliana</i>	GB:AF315737	-	-
AraII	Arabidopsis	<i>Arabidopsis thaliana</i>	EMBL: BT002775	-	Yamada <i>et al.</i> , (unpub)
AraIII	Arabidopsis	<i>Arabidopsis thaliana</i>	EMBL: AAM64985	-	Haas <i>et al.</i> , (unpub)
Avo	Avocado	<i>Persea americana</i>	PIR: JH0269	-	Kimura <i>et al.</i> , 1995
Bar	Barley	<i>Hordeum vulgare</i>	EMBL:Y12068	0.02nM	Gaddour <i>et al.</i> , 2001
Bea	Bean	<i>Phaseolus vulgaris</i>	-	-	Santino <i>et al.</i> , 1998
Bit	Bitter dock	<i>Rumex obtusifolius</i>	EMBL:AJ428415	-	Tinney <i>et al.</i> (unpub.)
Broc	Broccoli	<i>Brassica oleracea</i>	EMBL:AY065838	-	Watson and Coupe 2001(unpub.)
CabI	Chinese cabbage	<i>Brassica rapa</i>	EMBL:L41355	-	Lim <i>et al.</i> , 1996
CabII	Chinese cabbage	<i>Brassica rapa</i>	EMBL:L42819	-	Kim and Chung 2000 (unpub.)
Car	Carnation (clove pink)	<i>Dianthus caryophyllus</i>	EMBL: AY028994	-	Sugawara <i>et al.</i> , (unpub.)
Carr	Carrot	<i>Daucus carrota</i>	PIR: T14323	-	Ojima <i>et al.</i> , 1997
Cass	Cassava	<i>Manihot esculenta</i>	EMBL:AF265551	-	Reilly <i>et al.</i> , (unpub.)
Cast	Castor	<i>Ricinus communis</i>	EMBL:Z49697	-	Szederkenyi and Schobert (unpub.)
Cau	Cauliflower	<i>Brassica oleracea</i>	TrEMBL:Q8VYX5	-	Watson and Coupe 2001(unpub.)
Chel	Celandine (Chelidocystatin)	<i>Chelidonium majus</i>	-	0.00011nM	Rogel <i>et al.</i> , 1998
ChesI	European chestnut (CsC)	<i>Castanea sativa</i>	EMBL: AJ224331	29nM	Pernas <i>et al.</i> , 1998
ChesII	American chestnut	<i>Castanea dentate</i>	EMBL:AF480168	-	Connors <i>et al.</i> , (unpub.)
Chrb	Christmas bells	<i>Sandersonia aurantiaca</i>	EMBL:AF469485	-	Eason 2002 (unpub.)
Cock	Cockscomb (Celosiacystatin)	<i>Celosia cristata</i>	EMBL:AJ535712	-	Gholizadeh <i>et al.</i> , 2005
CornI	Corn I (Maize)	<i>Zea mays</i>	EMBL:D10622	0.083nM	Abe <i>et al.</i> , 1992
CornII	Corn II (Maize)	<i>Zea mays</i>	EMBL:D38130	-	Abe <i>et al.</i> , 1995
Cow	Cowpea	<i>Vigna unguiculata</i>	EMBL:Z21954	-	Fernandes <i>et al.</i> , 1993
Cuc	Cucumber	<i>Cucumis sativus</i>	-	-	Yamakawa <i>et al.</i> , (unpub.)
Faba	Faba bean	<i>Vicia faba</i>	EMBL:AY237958	-	-
Job	Job's tears	<i>Coix lacryma-jobi</i>	-	190nM	Yoza <i>et al.</i> , 2002
Kid	Kidney bean	<i>Phaseolus vulgaris L.</i>	-	0.08nM	Brzin <i>et al.</i> , 1998
Kiwi-I	Kiwi fruit	<i>Actinidia deliciosa</i>	GB:AY390353	0.16nM	Rassam and Laing 2004
Kiwi-II	Kiwi fruit	<i>Actinidia deliciosa</i>	GB:AY390354	-	Rassam and Laing 2004
Mugb	Mugbean	<i>Vigna radiata</i>	-	-	Kang <i>et al.</i> , (unpub.)
Mugw	Mugwort	<i>Artemisia vulgaris</i>	EMBL:AF143677	-	Hubinger <i>et al.</i> , 1999
Mus	Mustard	<i>Brassica campestris</i>	PIR:S65071	-	Lim <i>et al.</i> , 1996
RiceI	Rice (Oryzacystatin I)	<i>Oryza sativa</i>	EMBL:J03469	30nM	Abe <i>et al.</i> , 1987, Kondo <i>et al.</i> , 1990

¹Entries without database accession number were obtained from the referred publication.

²Years on unpublished references indicate date sequences were deposited in the database.



Table 4.1 continued

Code	Common name	Specie name	Database ¹ (Acc No.)	Papain K _i	Reference ²
RiceII	Rice (Oryzacystatin II)	<i>Oryza sativa</i>	EMBL:J05595	8.3nM	Kondo <i>et al.</i> , 1990
Pap	Papaya	<i>Carica papaya</i>	EMBL:X71124	0.75nM	Song <i>et al.</i> , 1995
Pear	Pear	<i>Pyrus communis</i>	-	-	Gauillard <i>et al.</i> , (unpub.)
Pot	Potato	<i>Solanum tuberosum</i>	PIR:PQ0469	-	Hildmann <i>et al.</i> , 1992.
PMC1	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC2	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC3	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC4	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC5	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC6	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC7	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC8	Potato multicystatin	<i>Solanum tuberosum</i>	-	-	Michaud, D. (Per. Comm.)
PMC10-4	Potato multicystatin (10-4)	<i>Solanum tuberosum</i>	GB:AAB29661	0.5nM	Walsh <i>et al.</i> , 1993
PMC32	Potato multicystatin (32)	<i>Solanum tuberosum</i>	SPROT:P37842	0.7nM	Walsh <i>et al.</i> , 1993 Waldron <i>et al.</i> , 1993
Rag	Ragweed	<i>Ambrosia artemisiifolia</i>	PIR:JN0906	-	Rogers <i>et al.</i> , 1993
Sesa	Sesame	<i>Sesamum indicum</i>	-	-	Tai <i>et al.</i> , (unpub.)
Sorg	Sorghum	<i>Sorghum bicolor</i>	EMBL:X87168	-	Li <i>et al.</i> , 1996
SoyI	Soyabean	<i>Glycine max</i>	PIR:S10588	-	Brzin <i>et al.</i> , 1990
SoyII	Soyabean (N2)	<i>Glycine max</i>	EMBL:U51855	57nM	Zhao <i>et al.</i> , (unpub.); Botella <i>et al.</i> (unpu)
SoyII	Soyabean (L1)	<i>Glycine max</i>	-	19,000nM	Zhao <i>et al.</i> , 1996
SoyIV	Soyabean (R1)	<i>Glycine max</i>	-	21nM	Zhao <i>et al.</i> , 1996
Squ	Squash	<i>Cucurbita maxima</i>	-	-	Farley <i>et a.</i> , 1998
Sug1	Sugarcane	<i>Saccharum officinarum</i>	NCBI:AAM78598	-	Soares-Costa <i>et al.</i> , 2002
Sug2	Sugarcane	<i>Saccharum officinarum</i>	-	-	Reis and Margis 2001
Sug3	Sugarcane	<i>Saccharum officinarum</i>	-	-	Reis and Margis 2001
SMC-I	Sunflower (Sca)	<i>Helianthus annuus</i>	PIR:JC4791	0.005nM	Kouzuma <i>et al.</i> , 1996
SMC-II	Sunflower (Scb)	<i>Helianthus annuus</i>	PIR:JC4792	0.00017nM	Kouzuma <i>et al.</i> , 1996
SMC-III	Sunflower multicystatin	<i>Helianthus annuus</i>	PIR:JC7333	0.04nM	Kouzuma <i>et al.</i> , 2000
Swe	Sweet potato (Batate)	<i>Ipomoea batatas</i>	EMBL:AF117334	-	To <i>et al.</i> , 1999; Huang <i>et al.</i> , 2001
Taro	Taro (Cocoyam)	<i>Colocasia esculenta</i>	EMBL:AF525880	-	Yang <i>et al.</i> , (unpub.)
Tom1	Tomato	<i>Solanum lycopersicum</i>	PIR:A59155	4.7nM	Jacinto <i>et al.</i> , 1998
Tom2	Tomato	<i>Solanum lycopersicum</i>	EMBL AF198388	-	Girard and Michaud 1999 (unpub.)
Whe	Wheat	<i>Triticum aestivum</i>	EMBL:AB038393	-	Kuroda <i>et al.</i> , 2001
Wist	Wisteria	<i>Wisteria floribunda</i>	PIR:PX0039	-	Hirashiki <i>et al.</i> , 1990

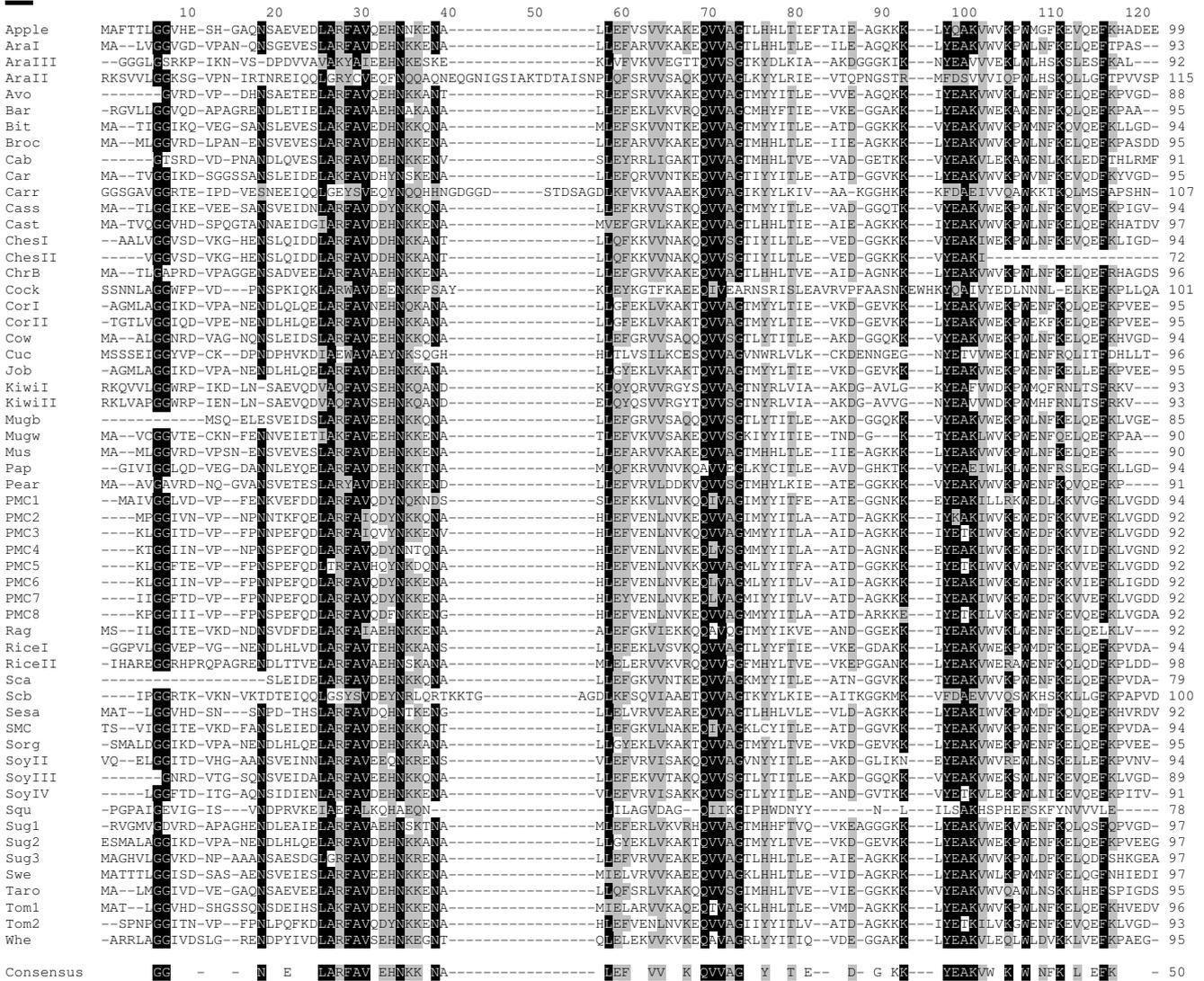
¹Entries without database accession number were obtained from the referred publication.

²Years on unpublished references indicate date sequences were deposited in the database

The multiple sequence analysis of the phytocystatins showed high levels of sequence homology and conservation especially around the regions involved in function and important structural features (Figure 3.1). The conserved glycine residue in the N-terminal region, known to be characteristic to this group of proteins and involved in N-terminal binding, was as expected present in all but three phytocystatins, mungbean (mugb), potato (pot) and sunflower multi-cystatin domain (sca). However this may have been due to incomplete sequences being deposited on the databases or intentional truncation of the gene by the research groups that provided the sequence. The QxVxG motif characteristic of all members of the cystatin super family and responsible for the second binding site (located in the 2nd hairpin loop) was clearly identified in the multiple alignments (Figure 3.1). Also found was the LARFAV motif in the N-terminal corresponds to the alpha-helix structure and is characteristic to phytocystatins only (Margis *et al.*, 1998). A new YEAKxKxWxKxF was identified in the C-terminal of the phytocystatins. This motif being unique to phytocystatins further adds to their qualification for a separate sub-family. This region is not as highly conserved as in animal cystatins. It has been reported to constitute the third binding region but with less binding capacity and probably more important in stabilising the complex with proteases. Since this region is characteristic only to phytocystatins, several workers have proposed that this group of proteins may constitute a separate sub-family within the cystatin family. From the multiple alignments, it is also clear that there is a very high correlation of conserved regions to important structural features used either for binding or structural conformity of the protein (Figure 3.1).



A



B

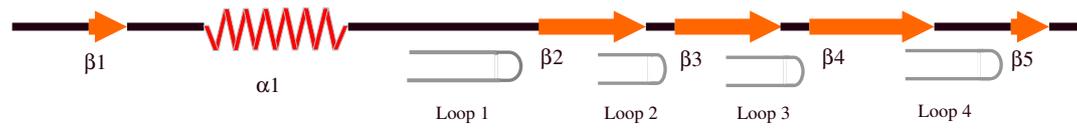


Figure 3.1 (A) Amino acid sequence alignment of known phycocystatins showing residue conservation across the different cystatins studied. A consensus sequence was also generated. Identical amino acids are highlighted in black while similar ones are in grey. (See Table 2.1 for a guide to abbreviated sequence titles). (B) Cartoon of the generalized secondary structural elements of phycocystatins. The orange arrows are the β -sheets (numbered from 1 to 5) and the red spiral representing the single α -helix. The positions where the loops occur are indicated with a gray paper clip mark and labelled 1 to 4.

Based on the neighbour joining phylogenetic tree that was generated, phytocystatins could be separated into five distinctive clades (Figure 2.2). Clades 5 and 6 seemed to be more primitive and may be progenitors of all the other groups of phytocystatins. The biggest clade, clade 1, could further be divided into two sub-clades, sub-clade 1 and 2 with the entire monocot cystatins grouping together in sub-clade 2. Sub-clade 1 included a rather more diverse group of phytocystatins. However, members of this group showed the lowest K_i values for papain (mean 0.37nM –data not shown) rendering them the most potent phytocystatins. They seem to have evolved from the monocot cystatins as deduced from a evolutionary distance tree (data not shown), which show the next lowest K_i values (mean 38.0nM - data not shown). Potency of phytocystatins seems to decrease down the tree with the exception of clade 5 (scb and SMC), which has a mean papain K_i of 0.2nM (data not shown).

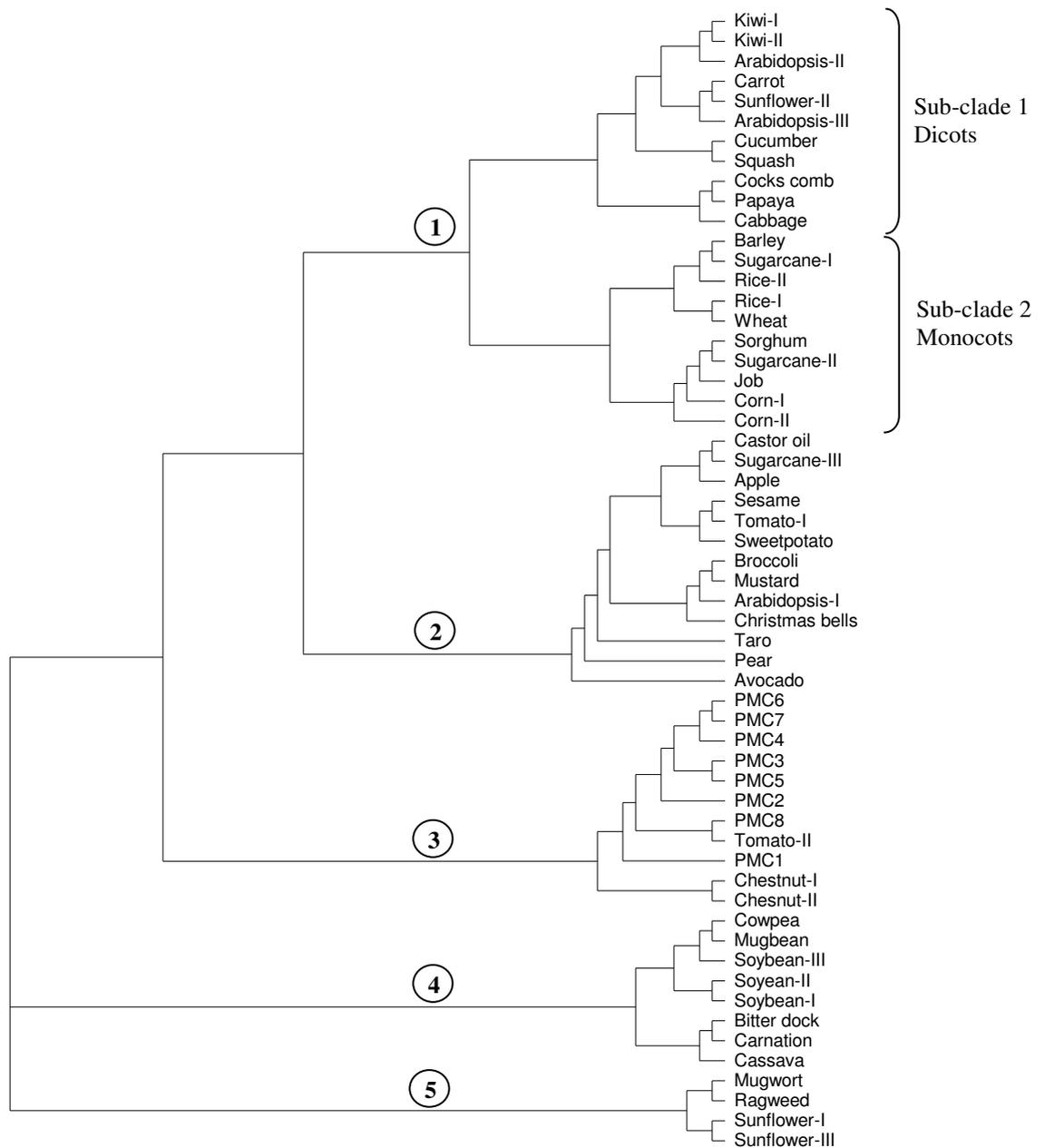


Figure 3.2 Phylogenetic tree for known phytocystatins based on the neighbour-joining method using PROTDIST and NEIGHBOR programs available in the PHYLIP (Phylogeny Inference Package) Version 3.57. Circled numbers indicate the 7 clades that were obtained.



Other clades also showed high plant taxa relationships, for example in clade 3 are members of the tomato and potato multi-cystatin, both plants belong to Solanaceae family and both phytocystatins are characterised by multiple domains. Clade 4 includes mostly members of the Fabaceae (Leguminosae) family for example the soybean multi-cystatin domains and cowpea cystatin. This clade seems to be evolutionary primitive. However, one of the domains clustered in clade 3, shows significant difference from its other domain cousins.

In a similarity matrix drawn to compare the sequence similarity of phytocystatins, percentage similarity ranged from 7% to 94% (Table 4.3). Phytocystatins with the least similarities included Arabidopsis-II and III, corks comb, cucumber and squash. High similarities were observed in avocado (Avo), barley (Bar), bitter dick (Bit) and broccoli (Broc). The highest similarity percentage was found between corn-I and job cystatins. This suggests that these are orthologs, genes that have maintained sequence and functional similarity even after speciation. A few more examples were identified in this similarity matrix; Arabidopsis-I (AraI) and broccoli (Broc) with 87% similarity, broccoli and Christmas bells, 79% similarity, corn-I and sorghum with 89% similarity (Table 3.2).

Modelled 3D structures of phytocystatins did show a few variations in the secondary structure elements and their arrangement. OC-I, which is the only phytocystatin whose crystal structure has been determined so far, was the only template structure used for the comparative modelling to determine the 3D models of unknown phytocystatins. Despite the diversity of origin (plant species and tissue types), phytocystatins structures have many structural features in common. The major



differences in the 3D structures were length of the N-terminal trunk, length of the 2nd hairpin loop, length of the 5th β -strand and length of the C-terminal (Figure 3.3). Further, as found from experimental structures of chicken egg white (Rawlings and Barret, 1990; Turk and Bode, 1991) and OC-I, other phytocystatins display the same general structural features i.e. five (in some cases four as in the corn cystatin and sunflower multi-cystatins sca and scb) anti-parallel β -sheets, wrapping halfway round a single central α -helix structure and three hairpin loops (Figure 3.3).

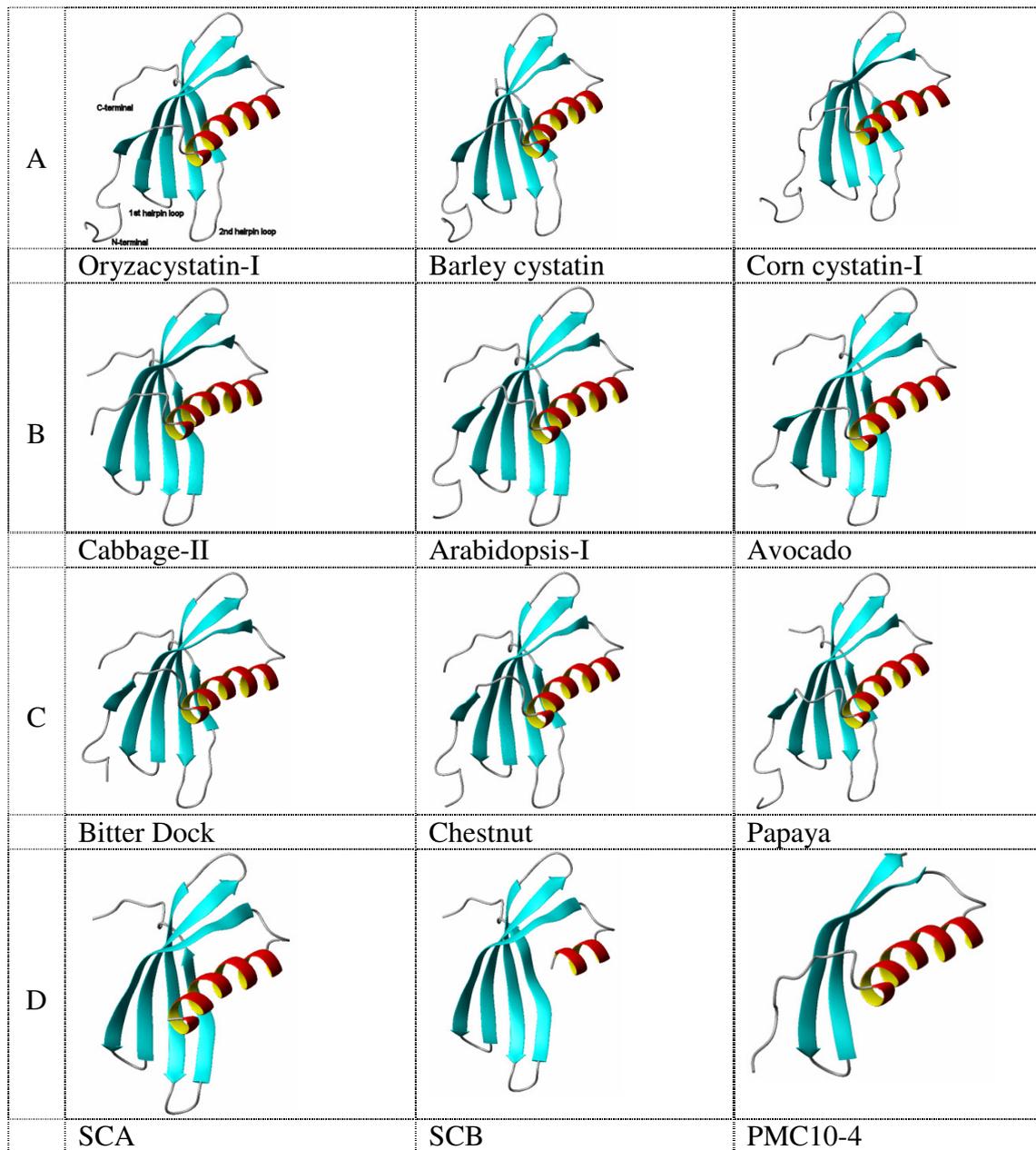


Figure 3.3 Predicted three-dimensional structures of selected phytocystatins representing the major phylogenetic groups (Figure 3.2; (A) group 1; (B) group 2; (C) group 3; (D) groups 4 and 5), and showing the secondary structural elements; five anti-parallel β -strands (blue), one α -helix (red), three hairpin loops, a long N-terminal trunk and a short C-terminal. The figures were made with MOLMOL program (Koradi *et al.*, 1996).

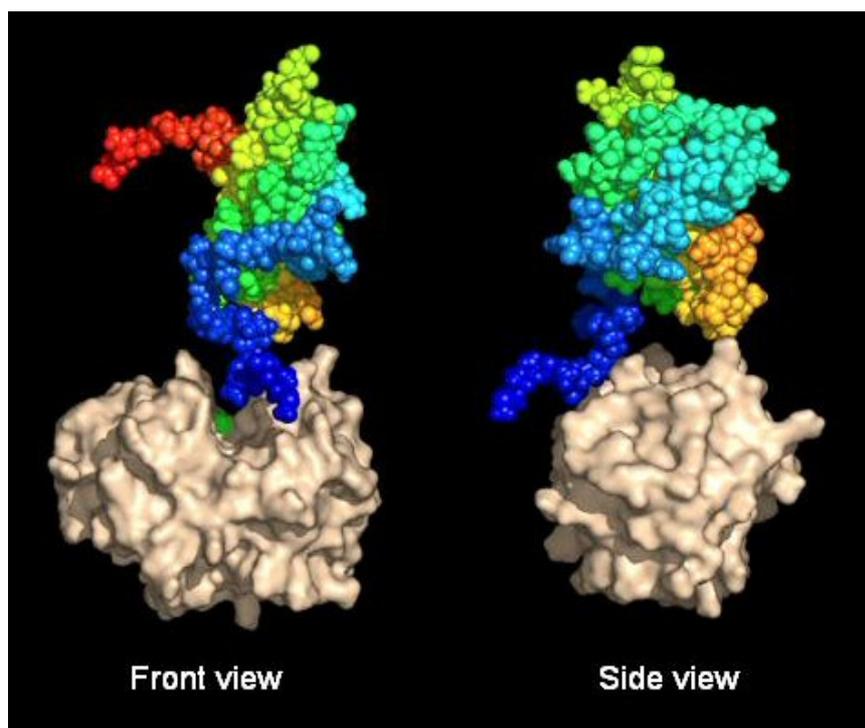


Figure 3.4 Modelled complex between OC-I (top) and papain (bottom) in front and side views. OC-I is shown in spheres and its structure coloured by rainbow. The N-terminal is blue towards the C-terminal red. The surface of papain is coloured light brown. The active site of papain appears as a trench extending from the front to the back of the molecule. The active cysteine residue of papain here coloured green (front view) occurs in the middle of this trench. The complex was initially modelled manually and the model refined using MULTIDOCK program based on minimisation algorithms. Visualisation and rendering graphics were done using PYMOL program.

Figure 3.4 shows the predicted binding of OC-I onto the papain active site cleft. *In-silico* docking experiments involving OC-I and papain revealed that OC-I attaches onto the active site cleft of papain and possibly in the same way with other cysteine



proteases (Figure 4.4). Due to electrostatic forces along these two molecules, an average distance of 1.8Å separates them from each other. During the docking process, one residue in the N-terminal of OC-I, aspartic acid (Asp4) prevented the inhibitor from docking closer into the papain active site.

3.5 Discussion

This study has provided new information about the structure of phytocystatins. Firstly, despite high structural similarity of phytocystatins, there is very wide variation in inhibition potential meaning that biochemical screening could yield selections of cystatins targeting a wide range of pests as well as other uses. It has been shown in this study that the experimentally determined structure of OC-I can effectively and successfully be used to predict structural conformations of unknown cystatins. This improves their analysis and evaluation as demonstrated by Girad *et al.*, (2007). Further, it has been elucidated through *in-silico* prediction of structure that the individual domains of multicystatins (e.g. tomato and potato), when separated, can fold into functional proteins individually. Docking and inhibitor- protease complex prediction was possible for the first time using phytocystatins. From the analyses of binding candidate residues to engineer for improved binding (and thus inhibition capacity) were inferred.

In some plant species, for example *Oryza sativa* (rice), *Zea mays* (corn) and *Solanum lycopersicum* (tomato), more than one highly homologous phytocystatin has been identified. In other plants, like *Glycine max* (soybean), *Helianthus annuus* (sunflower) and *Solanum tuberosum* (potato), multiple domain cystatins have been identified. It has been shown that when these domains, which occur in tandem, are cleaved by



enzyme digestion, the separated domains can fold into functional proteins retaining their inhibitory activity (Walsh *et al.*, 1993). This suggests that these forms of multi-domain cystatins may have arisen as a result of gene duplication events. The potato multicystatin for example has eight domains while the sunflower multi-cystatin has 4 active domains.

Evolutionary relationships among phytocystatins were inferred using an unrooted phylogenetic tree. As expected, most phytocystatins grouped together to reflect the plant taxonomic groups. However, some members of the multidomain cystatins tended to occur in distinctly different groupings. This suggests that plants, such as tomato, soybean, sunflower and sugarcane, contain complete cystatin coding genes that may have distinctly different evolutionary origins.

It is still structurally unclear how the phytocystatins with longer N-terminal trunks are more potent than the shorter forms. From nuclear magnetic resonance (NMR) data of OC-I it is known that the N-terminal trunk is highly flexible and does not form any ordered structure. *In-silico* observations in this study have shown the possible formation of a bond between residues GLY6 and VAL8 forming a loop structure in long enough N-terminals. This probably stabilises the trunk allowing a more precise binding more and rendering the complex more stable. Cystatins bind to proteases with a 1:1 stoichiometry and with varying affinities. However, it is not clear whether this is true for the multi-domain cystatins. The whole phytocystatin molecule is wedge shaped with the N-terminal and the two hairpin loops forming the sharp edge in some cases the N-terminal protrudes out into a long arm extending outwards from the rest of the structure (Figure 3.3) forming what has been referred to as a trunk. This sharp



edge is highly hydrophilic and complimentary to the active cleft of cysteine proteases (Bode *et al.*, 1988) forming the active site region. The active site itself is composed of a glycine residue in the N-terminal and this appears to be the most important binding site in many cystatins although its removal or absence does not seem to affect binding by other types of phytocystatins (Arai *et al.*, 1991). The sca and scb cystatin domains of the sunflower multi-cystatin do not have N-terminal trunks (Figure 3.3) despite retaining high affinity for cysteine proteases (Kouzuma *et al.*, 1996).

OC-I is still the only phytocystatin one whose tertiary structure has been experimentally determined. In this study, bioinformatics tools have been successfully used to predict the inhibitor-protease (OC-I and papain) complex. In general, the binding in the predicted complex was in agreement with that of the experimental complex structures previously reported between stefin-B and papain (Studds *et al.*, 1990) and stefin-A with cathepsin-H (Jenko *et al.*, 2003). In docking OC-I and papain, it was difficult to dock two residues of aspartic acid ASP4 in the N-terminal trunk and ASP86 in the 2nd binding hairpin loop of the C-terminal region. These two residues are close to the active sites and seemed to prevent closer binding of the active site to the target papain. Therefore, these sites appear to be potential targets for site-directed mutagenesis directed at improving binding and therefore potency of OC-I to papain and probably to other cysteine proteases. These might be replaced by either asparagine (ASN) or glutamic acid (GLU) based on the Doolittle amino acid substitution matrix (Mark *et al.*, 1993). This is one of the many database derived matrices showing evolutionally substitution of amino acids in many similar proteins. Through such a matrix amino acid substitutions can be done through site-directed



mutagenesis, to maintain the overall structure as much as possible but vary small parameters like bond distance and eventually levels of potency.

The wide variation in affinities found for phytocystatins in this study, and indeed also in other animal cystatins, is not explainable by a simple structural difference. For example, an inhibitor with highly similar structural features has a great difference in affinity. Nikawa *et al.*, (1989) reported that the 1st binding hairpin loop with the QVVAG highly conserved motif was not essential for cysteine protease inhibitory activity in cystatin-A. This study identified that PMC10-4, a potato multi-cystatin domain, retains high affinity despite having only the N-terminal and the 1st loop. Sca and scb retain high affinity (mean K_i is 0.003nM) despite not having an N-terminal. This means that the two rely on the 1st and 2nd hairpin loops for their activity. Therefore, it is possible that such functional differences may be explainable by small structural features at residue level that result in great differences in affinity of the inhibitor.



CHAPTER 4

Engineering of a papaya cystatin using site-directed mutagenesis to improve its activity against papain and weevil digestive cysteine proteases

Scientific Communication

Parts of this chapter led to the publication:

Kiggundu, A., Goulet MC., Goulet C., Dubuc JF., Rivard D., Benchabane M., Pépin G., Van der Vyver C., Kunert K. and Michaud D. (2006). Modulating the protease inhibitory profile of a plant cystatin by single mutations at positively selected amino acid sites. *The Plant Journal*, 48, 403–413.



4.1 Abstract

The usefulness of native phytocystatins for pest control is limited by the co-evolution between the pest and host-plant. This has allowed insects to develop ways of overcoming the presence of inhibitors in plant tissues. This includes the production of insensitive proteases in the variable gut environment helping insects to elude the anti-nutritive effects of cystatins. Protein engineering was employed in this part of the study to attempt to produce variants of a papaya cystatin with improved activity against a model protease papain and also against gut proteases of banana weevil and the black maize beetle *Heteronychus arator* Fabricius (Coleoptera: Scarabaeidae). Specific amino acids in the amino acid sequence of the papaya cystatin were changed using site-directed mutagenesis. An evolutionary and structural analysis strategy was applied to improve cystatin activity against cysteine proteases. The papaya cystatin was amendable to improvement and papaya cystatin mutants showed 1.5- to 6-fold improved inhibition of papain. Amino acid changes close to conserved regions of the protein provided the most improved inhibition against cysteine proteases. Improvement was not as high as for papain when banana weevil and black maize beetle gut extracts were tested. Improvements ranged from 1.5- to 2-fold in the mutant E52Q with a change from glutamic acid to glutamine. Novel cystatin mutants with increased inhibitory activity represent a first step in setting up a library of mutated phytocystatins with improved inhibition against both endogenous cysteine proteases and proteases derived from plant pests.



4.2 Introduction

The usefulness of native protease inhibitors for pest control is limited due to the fact that over evolutionary time insects have developed ways of overcoming the presence of inhibitors in plant tissues. This is mainly due to the production of insensitive proteases in insects and the variable gut environment which helps insects to elude the anti-nutritive effects of phytocystatins. Engineering of phytocystatins by changing amino acids in the amino acid sequence for better binding to proteases is one strategy to possibly improve the efficiency of cystatins. In a first approach, using cystatin engineering, better protection against a plant pest has been found with transgenic plants expressing an engineered OC-I (Urwin *et al.*, 1997, Irie *et al.*, 1996). Site-directed mutagenesis is applied as a tool to alter the amino acid sequence through the replacement of single or several nucleotide bases to alter amino acid sequence of the respective protein.

There are mainly two general strategies for protein engineering (i) rationale design and (ii) directed evolution. In rationale design, detailed knowledge of the structure and function of the target protein is used to make changes in the sequences that through site-directed mutagenesis leads to the desired modulation of function and properties (Carter, 1986; Young and Dong, 2003). The second method known as directed evolution mimics natural evolution. This method is performed by application of random mutagenesis to a protein followed by a high throughput selection to identify variants that have the desired qualities. This method has been shown to successfully produce improved proteins. However, it requires large amounts of recombinant DNA which has to be mutated. Also, the products screened often require expensive robotic equipment for automated selection assays.



However, for the successful application of the rationale design approach, additional evolutionary analysis has to be carried out. This includes positive site selection at the amino acid level as a guide to mutagenesis. A number of studies have shown that proteins involved in host defence responses are subject to adaptive evolution. This results from direct selection pressure on amino acid residues that directly interacts with target molecules of invading or predatory organisms (Barbour *et al.*, 2002; Bishop *et al.*, 2000; Sawyer *et al.*, 2005). Most genetic variation detected at the molecular level is assumed to result from randomly generated mutations so that mutations that confer a selective advantage to the host are maintained in evolutionary time (Yang and Bielawski, 2000). At the gene level, this process of positive selection (maintenance of mutations that confer advantage) can be detected by comparing the rate of non-synonymous codon substitutions (dN), where the original amino acid is substituted for an alternative residue, and the rate of synonymous substitutions (dS), where the original amino acids are preserved. In practice, the ratio of dN to dS, referred to as ω , is considered to be a reliable measure of the directional selection exerted on the protein (Yang, 2005). For amino acid sites with little or no impact on the activity of the protein, the ω ratio will be close to 1 as nonsynonymous mutations will be fixed at the same rate as synonymous mutations by neutral selection. Conserved amino acid sites, where any amino acid substitution would strongly compromise biological activity, will typically show a ω ratio close to 0 as a result of negative (or purifying) selection. In contrast, amino acid substitutions giving the organism a selective advantage will tend to be readily fixed in the population, resulting in calculated ω values greater than 1 for the corresponding amino acid site. Statistical methods based on maximum-likelihood models have been developed to detect positive selection by



the estimation of ω values (Yang and Bielawski, 2000). These methods allow the identification of specific codon and amino acid sites subject to positive selection (Bielawski *et al.*, 2004; Ivarsson *et al.*, 2003; Sawyer *et al.*, 2005; Yang *et al.*, 2000).

The objective of this part of the study was therefore to use an evolutionary guided rationale for engineering of a papaya cystatin for improved inhibition of a cysteine protease. Maximum-likelihood models were used to detect amino acid sites in Poaceae (monocots; seven species) and Solanaceae (dicots; potato and tomato) that have, over evolutionary time, been subjected to positive selection. Possible sites for mutations were selected that can improve the activity or inhibitory profile of phytocystatins to assess if actually positive selection has occurred in phytocystatins.

4.3 Materials and methods

4.3.1 Phylogenetic and structural model analysis

Phylogenetic analysis as well as the protein structural modelling analysis that was used to predict potential mutable sites has been outlined in Chapter 3 of this thesis.

4.3.2 Detection of positive selection sites in PhyCys

Positive selection sites for phytocystatin genes were detected using maximum likelihood models M0, M1, M2, M3, M7, M8, R1 and R2, which are present in the software package Phylogenetic Analysis Maximum Likelihood (PAML) version 3.14 (<http://abacus.gene.ucl.ac.uk/software/paml.html>) (Yang, 1997). PAML includes a suite of codon-based models that can be used to estimate ω , the ratio of the rate of non-synonymous substitutions per non-synonymous codon site (dN) to the rate of

synonymous substitutions per synonymous site (dS) as well as calculation of posterior Bayesian probabilities needed to identify positively selected sites in genes.

4.3.3 Construction of over-expression vector for papaya cystatin (PC)

The PC coding sequence was excised from the cloning vector *pBLCYS1* using the restriction enzymes *EcoRI* and *PstI*. The *EcoRI/PstI* fragment was then first cloned into the *EcoRI/PstI* site of *pBlueScript* (Stratagene, USA) and then as a *BamHI/KpnI* fragment from *pBlueScript* into the vector *pQE31* to achieve in-frame expression of a 6Xhis-tagged protein. This sub-cloning procedure created the plasmid *pQE31PC-I* (Figure 4.1). This plasmid was transformed into *E. coli* cells (strain JM109) for storage and into *E. coli* strain M15 for expression according to the QIAexpressionist kit user's manual (Qiagen, Germany). Site-directed mutagenesis to engineer PC was done directly in the expression vector *pQE30XaCYS*.

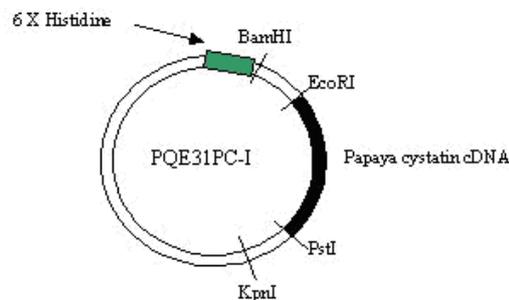


Figure 4.1 Schematic representation of recombinant protein expression vector *pQE31PC-I* created to express PC. In this vector site-directed mutagenesis was also performed.

4.3.5 Mutagenesis primer design

For each site to be mutated, two mutagenic oligonucleotide primers were designed. These primers contained the desired mutation with at least 10 to 13 bases flanking the

mutation were exactly complimentary to the template DNA. This was achieved by using PrimerX (<http://bioinformatics.org/primerx/>), a web-based program developed to automate the design of mutagenic PCR primers for application in site-directed mutagenesis. Based on input (DNA or amino acid sequence), the program compares a template sequence that already incorporates the desired mutation. It then generates several forward and reverse primer sequences by encoding the mutation and finally computes for other necessary primer information like melting temperature and GC content for each primer pair. The primers, which were used in this study, are outlined in Table 4.1.



Table 4.1 Sequence information of the mutagenic primer pairs used for the mutations. Mismatched bases are underlined. *Mutation 16 was an N-terminal truncation to remove seven amino acids. The primer pairs were created to cut out 21 bases and include part of the vector backbone.

Mutation number	Mutant Code	Primer sequence Forward and Reverse
1	CYSP03F	5' GAGGGAAGGATGGAG <u>T</u> TCGGAATTGTGATC 3' 5' CCGATCACAATTCCG <u>A</u> ACTCCATCCTTCCC 3'
2	CYSP03S	5' GGGGAAGGATGGAG <u>T</u> CCGGAATTGTGATC 3' 5' GATCACAATTCCGG <u>A</u> CTCCATCCTTCCC 3'
3	CYSV06R	5' GAGCCCGGAATT <u>C</u> GGATCGGTGGTTTG 3' 5' CAAACCACCGAT <u>C</u> CGAATTCCGGGCTC 3'
4	CYSI07L	5' CCCGGAATTGTG <u>C</u> TCGGTGGTTTGC 3' 5' GCAAACCACCGA <u>G</u> CACAATTCCGGG 3'
5	CYSI07A	5' CCGGAATTGTG <u>G</u> CAGGTGGTTTGCAG 3' 5' CTGCAAACCAC <u>T</u> GCCACAATTCCGG 3'
6	CYS07V	5' CCCGGAATTGTG <u>G</u> TCGGTGGTTTGC 3' 5' GCAAACCACCGA <u>C</u> CACAATTCCGGG 3'
7	CYSI07D	5' CCCGGAATTGTG <u>G</u> ACGGTGGTTTGC 3' 5' GCAAACCACCG <u>T</u> CCACAATTCCGGG 3'
8	CYSA32V	5' CGCCGTCGATG <u>T</u> GCCACAACAAAG 3' 5' CTTTGTGTGGC <u>A</u> CATCGACGGCG 3'
9	CYSA52P	5' GTGAATGTAAAGCAG <u>C</u> CAGTGGTTGAAGGC 3' 5' GCCTTCAACCACTG <u>G</u> CTGCTTTACATTAC 3'
10	CYSA52Q	5' GAATGTAAAGCAG <u>C</u> CAGTGGTTGAAGGC 3' 5' GCCTTCAACCACTG <u>G</u> CTGCTTTACATTAC 3'
11	CYSE55A	5' CAGGCAGTGGTTG <u>C</u> AGGCTTAAAGTAC 3' 5' GTA <u>C</u> TTTAAGCCTGCAACCACTGCCTG 3'
12	CYSC60T	5' GTTGAAGGCTTAAAGTAC <u>A</u> CCATCACTTTGGAGGCTG 3' 5' CAGCCTCCAAAGTGATG <u>G</u> TACTTTAAGCCTTCAAC 3'
13	CYSI78V	5' GTATATGAGGCCGAG <u>G</u> TCTGGGTGAAGCTC 3' 5' GAGCTTACCCAG <u>A</u> CTCGGCCTCATATAC 3'
14	CYSW79P	5' GAGGCCGAGAT <u>C</u> CGGTGAAGCTCTGG 3' 5' CCAGAGCTTCA <u>C</u> CGGATCTCGGCCTC 3'
15	CYSE84A	5' GTGAAGCTCTGGG <u>C</u> GAAATTCAGGAGC 3' 5' GCTCCTGAAATTC <u>G</u> CCCAGAGCTTAC 3'
16	CYSN85X	5' GAAGCTCTGGGAGTTCAGGAGCTTG 3' 5' CAAGCTCCTGAACTCCAGAGCTTCC 3'
15	CYSR87C	5' CTGGGAGAAATTT <u>C</u> TGCAGCTTGGAGGGATT 3' 5' GAATCCCTCCAAGCT <u>G</u> CAGAAATTCCTCCA 3'
16*	CYS _t NT	5' CTGGTATCGAGGGAAGGATGGGTTTGCAGG 3' 5' CCCTCGACGTCTGCAAACCAATCCTTCCC 3'

4.3.6 Site-directed mutagenesis

Site-directed mutagenesis was done following a modified *Quick Change* mutagenesis method (Stratagene, USA), which was an effective and simple method with which mutations can be carried out inside expression vectors (Fisher and Pei, 1997).

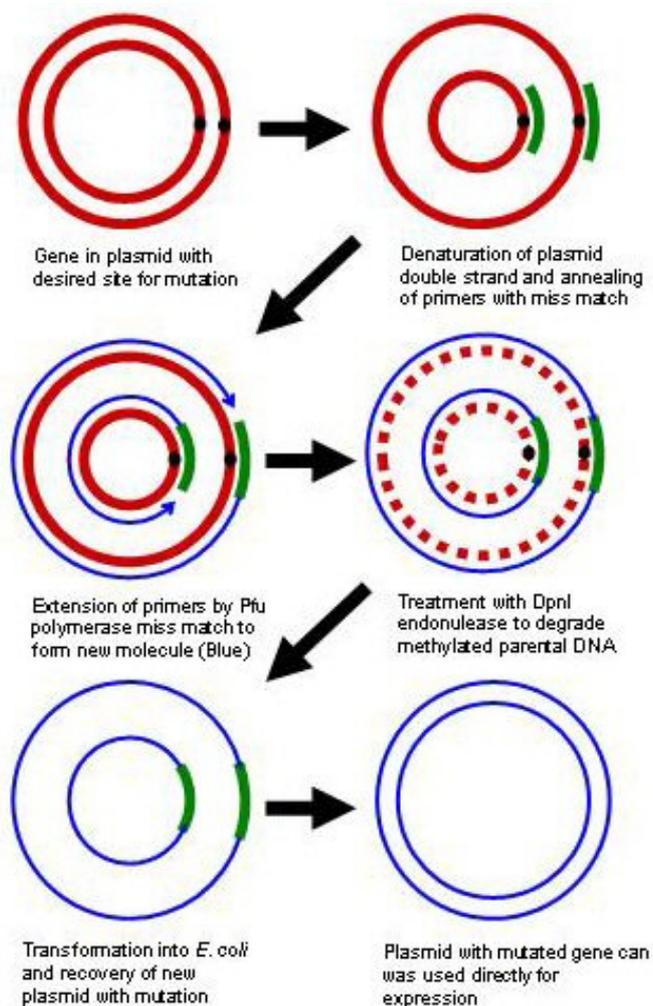


Figure 4.2 Schematic representation of the site-directed mutagenesis protocol used (modified from QuickChange® Site-Directed Mutagenesis Kit Manual #200518, Stratagene, USA)

The whole method is divided into three stages; amplification of mutant DNA, degradation of parental DNA (methylated) and transformation into *E. coli* cells. Briefly, primers containing miss-matched nucleotides at site of intended mutation



resulting in the desired modification anneal to complementary opposite strands of the template plasmid DNA. These are then extended in a PCR using *Pfu* DNA polymerase. The PCR reaction is then treated with *DpnI* endonuclease enzyme, which specifically targets methylated DNA, in this case the parental plasmid DNA. The new double-stranded DNA containing the desired mutations is then transformed into *E. coli* competent cells and stored until needed for further use.

In this particular study, the amplification step was modified into a two-stage PCR as described by Wang and Malcom (1999). Two separate primer extension reactions one for each primer were set up as follows; 10-15ng template plasmid (5 μ l), 10X *Pfu* buffer (5 μ l), 25 μ M primer-1 (1 μ l), 10mM dNTPs (1 μ l), 2.5 units *Pfu* DNA polymerase (Fermentas #EP0571) (1 μ l) and sterile distilled water up to 50 μ l total reaction volume. Exactly the same reaction was set up for the second primer. The PCR conditions setup on an automated cycler (Palm Cycler, Corbett Life Science, Australia) were denaturation at 94°C for 30sec, 4 cycles of; 95°C 30sec, 55°C 1min, 68°C 7min (2 minutes/kb of plasmid length and pQE31PC-1 is 3500bp). The reaction was held at 4°C on completion of cycling.

In the second PCR stage, 25 μ l from each of the separate reactions above were combined in a new tube, 1 μ l *Pfu* polymerase added, mixed and incubated as above except that the cycles were increased to 18. To degrade parental DNA, 10 units of *DpnI* enzyme were added to the cooled reaction, mixed well and incubated at 37°C for 1hr.



Finally, the reaction containing mutant DNA was used for transformation in competent *E. coli* cells. This was done by placing 200µl competent JM109 cells in a Falcon tube on ice plus 2µl of the digested PCR reaction and incubated on ice for 20min. Heat shock was performed by placing the cell/DNA mix at 42°C for 60sec and then returned on ice for 2min. LB (500µl) broth was added and then the cells were incubated at 37°C with shaking at 200rpm for 1hr after which 100µl of this culture was plated on solid LB containing 100mg/l ampicillin and incubated overnight 37°C. Three individual colonies were picked and inoculated into LB broth (50ml) containing 100mg/l ampicillin and again incubated at 37°C overnight with shaking at 180rpm. Minipreps were made and pure plasmid DNA sent for sequencing.

4.3.7 Protein expression

All the mutants were expressed directly in the pQE31P-1 vector in which the mutations were done using the QIAexpressionist kit (Qiagen, Germany) as described in the manufactures manual and also described in Chapter 2 (Section 2.3.8) of this thesis. Briefly, LB medium (5ml) with antibiotics (50mg/l kanamycin and 100mg/l ampicillin) was inoculated with a single bacterial colony of *E. coli* (strain M15) cells containing pQE31PC-1 mutants and grown overnight at 37°C with shaking at 200rpm. Pre-warmed LB medium (100ml) with antibiotics (as above) in a 250ml conical flask was inoculated with 5ml of the overnight culture and incubated at 37°C with shaking as above until the optical density at 600nm (OD600) reached 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 1mM to induce expression and incubation continued for another 4hrs. Bacterial cells were harvested by centrifugation (13000rpm at 4°C) for 10min and stored frozen at -20°C until purification.



4.3.8 Protein purification

Purification was performed under native conditions to preserve the conformational integrity of the protein. Frozen cell pellets were thawed on ice for 30min, re-suspended in his-tag lysis buffer (50mM sodium di-hydrogen phosphate, pH 8.0; 300mM sodium chloride; 10mM imidazole) at a rate of 2ml per 1mg of cells and 1mg lysozyme was added. This was mixed gently and incubated on ice for 1hr. The cell suspension was then sonicated using a sonicator (Cell Disruptor B-30, Branson Sonic Power Co./SmithKline Co.) fitted with a standard micro-tip and set to 20% duty cycle, 2 output control and in pulse mode. The cells were sonicated using 10 bursts with 10sec cooling on ice between each burst, taking care not to create much frothing. The lysates thus obtained were centrifuged at 10,000rpm for 30min at 4°C in a centrifuge and the clear supernatant transferred into fresh Eppendorf tubes to which 800µl of 50% Ni-NTA slurry (Qiagen, Germany) was added. The tubes were shaken at 200rpm for 30min at 4°C after which the cell lysate mixture was poured into a short plastic column (made with a 2.5ml syringe and a glass wool plug at the bottom) with the bottom cover in place. The cover was removed after the slurry settled and the flow-through collected. Two-times 1ml wash buffer (50mM sodium di-hydrogen phosphate, pH 8.0; 300mM sodium chloride; 50mM imidazole) was carefully poured over the column and collected at the bottom. This was followed by pouring slowly 4-times 500µl elution buffer (50mM sodium di-hydrogen phosphate, pH 8.0; 300mM sodium chloride; 250mM imidazole) over the slurry. The elutions were collected separately in 500µl fractions. Five micro-liters of each fraction (flow-through, washes and elution fractions) were each added to 5.0µl SDS-PAGE sample buffer (6% β-mercaptoethanal, 6% SDS, 0.6% bromophenol blue, 20% glycerol) heated to 37 °C for 10min and loaded onto a 15% polyacrylamide gel for evaluation of the purification



process and detection of the recombinant proteins. The purity of the inhibitors was assessed using 15% (w/v) SDS-PAGE analysis as described in (Sambrook *et al.*, 1989). The protein concentration of the elution fractions was finally determined using the Bio-Rad protein assay kit (Bio-Rad, South Africa), and fractions were stored in aliquots at 4°C until required.

4.3.7 Enzyme kinetics of mutants

Dissociation constants ($K_{i(\text{app})}$) for the interaction and inhibition of a model cysteine protease, papain, by the papaya cystatin variants obtained were determined by the monitoring of substrate hydrolysis progress curves as described by Salvesen and Nagase (1989). Papain activity was measured in 50mM Tris-HCl, pH 6.0 containing 5mM L-cysteine as reducing agent using the synthetic substrate *N*-CBZ-Phe-Arg-7-amido-4-methylcoumarin. Hydrolysis was allowed to proceed at room temperature while monitoring progress on the spectro-fluorometer with excitation and emission filters at 360nm and 450nm, respectively. When the reaction reached a steady state, the inhibitors were added and monitoring continued until a new steady state was reached. The difference in the initial vs final reaction rates was used to compute the apparent K_i value of the inhibitor.

4.4 Results

4.4.1 Rationale of mutations

Table 4.2 below outlines the particular amino acid change. Some of these changes were prompted by literature reports. In particular the truncation of the N-terminal has been reported not being important in some cystatins. However the modelling study showed that it may be important in stabilising the protein at the active site. Figure 4.3



below illustrates that mutations at sites 52 and 55 flanking the major functional motif ‘VV’ gave the highest improvement in inhibition. An indication that activity differences in phytocystatins could be explained by the sequence variability close to the active sites.

Table 4.2 Mutations performed on native PC, the amino acid changes made and the respective rationale. The mutant code refers to the amino acid changes made, for example CYSC60T refers to a mutation where cysteine at position 60 was replaced with threonine.

<i>Mutation number</i>	<i>Mutant code</i>	<i>Amino acid change</i>	<i>Rationale</i>
1	CYSP03F	Proline (position 3) to phenylalanine	Mutation in positive selection site in the N-terminal
2	CYSP03S	Proline (position 3) to serine	Mutation in positive selection site in the N-terminal
3	CYSV06R	Valine (position 6) to arginine	Random mutation in a less conserved region close to the N-terminal active site.
4	CYSI07L	Isoleucine (position 7) to leucine	Random mutation in a less conserved region close to the N-terminal active site.
5	CYSI07A	Isoleucine (position 7) to alanine	Random mutation in a less conserved region close to the N-terminal active site.
6	CYSI07V	Isoleucine (position 7) to valine	Random mutation in a less conserved region close to the N-terminal active site. Both isoleucine and valine are aliphatic and hydrophobic.
7	CYSI07D	Isoleucine (position 7) to aspartic acid	Random mutation in a less conserved region close to the N-terminal active site. Isoleucine and aspartic acid have very different chemical properties, however aspartic acid has the smallest side chain, than would less interfere with binding of the N-terminal.
8	CYSA32V	Alanine (position 32) to valine	Random mutation in a less conserved region, both are small and hydrophobic.
9	CYSA52P	Alanine (position 52) to proline	Random mutation in a positively selected site close to the 2 nd loop active site. Proline substitution showed increased bond number in the 2 nd loop and may improve structural strength.
10	CYSA52Q	Alanine (position 52) to Glutamine	Random mutation in a positively selected site close to the 2 nd loop active site. Proline substitution showed increased bond number in the 2 nd loop and may improve structural strength.
11	CYSE55A	Glutamic acid (position 55) to alanine	Random mutation in a positively selected site close to 2 nd binding site.



<i>Mutation number</i>	<i>Mutant code</i>	<i>Amino acid change</i>	<i>Rationale</i>
12	CYSC60T	Cysteine (position 60) to threonine	Cysteine was found to be a very rare amino acid in phytocystatins so it was changed to threonine also a small and hydrophobic amino acid.
13	CYSI78V	Isoleucine in (position 78) to valine	Random mutation in a positively selected site close to the C-terminal active site.
14	CYSW79P	Tryptophan (position 79) to proline	Random mutation in a positively selected site close to the C-terminal active site.
15	CYSE84A	Glutamic acid (position 84) to alanine	Random mutation in a positively selected site close to the C-terminal active site.
16	CYSN85X	Deletion of asparagine in position 85	Random mutation in a less conserved region close to the C-terminal active site. Asparagine deleted from the sequence.
17	CYSR87C	Arginine (position 87) to cysteine	Random mutation in a less conserved region close to the C-terminal active site. Arginine's long side chain seemed to interfere with C-terminal binding.
18	CYS _t NT	Truncation of the first 7 amino acids of the N-terminal	Truncation of N-terminal to reduce interference in binding
19	CYSA52QE55A	Combined 9 and 10	Combining two improved mutations



4.4.2 Positive selection among plant cystatin genes

Maximum-likelihood tests were carried out to predict positive selection among codon sites in a combined dataset of Poaceae and Solanaceae cystatin coding sequences. This was based on the phylogenetic analysis outlined in Chapter 3 of this thesis according to the methods described by Yang *et al.* (2000). Based on codon substitution models M0, M1, M2, M3, M7 and M8 (Yang *et al.*, 2000), 18 codon sites showing Bayesian posterior probabilities greater than 60% were considered to have been subjected to positive selection during the evolutionary advancement of these genes (Table 4.3; Figure 4.3). Three models M2, M3 and M8 allowing for positive selection fitted the data significantly better than M0, M1 and/or M7, with $p < 0.01$ for all likelihood ratio tests (Table 4.3). Models M3 and M8, which included 5 and 4 parameters respectively, gave a ω value greater than 1 (1.27) for the codons 1, 2, 6, 10, 15, 16, 17, 25, 29, 31, 45, 47, 51, 57, 58, 60, 76, 84 (Table 4.3). When the ratio of the rate of non-synonymous codon substitutions to rate of synonymous substitutions is greater than 1, the substitution at this site has given the organism a selective advantage and is largely fixed in the population. As expected and previously reported with other data sets (Yang *et al.*, 2000), positive selection could not be detected under M2 ($\omega < 1$), whereas M3 and M8, which are more powerful as they allow for heterogeneous distributions of ω ratios among codon sites (Yang and Bielawski, 2000), gave ω (ω for M8) values greater than 1.

Posterior Bayesian probabilities were calculated to estimate the probability of each individual codon belonging to an alternate codon assuming that the codon is being subject to positive selection. Eight sites, showing posterior probabilities greater than 95%, were thus identified to be highly positively selected. This included codons 1, 2,



6, 10, 45, 47, 76 and 84 (Figure 4.4). Site 2, subjected to several mutations to investigate mutants at this positively selected site, would improve activity and inhibition of proteases *in-vitro*.

Table 4.3 Evidence for positive selection events among codon sites of Poaceae and Solanaceae cystatins (n=21)

<i>Model</i>	p^a	Ω	l	<i>Positively selected sites^b</i>
M2	3	0.23	-2690.5	
M3	5	1.27	-2688.4	1, 2, 6, 10, 15, 16, 17, 25, 29, 31, 45, 47, 51, 57, 58, 60, 76, 84
M8	4	1.27	-2688.4	1, 2, 6, 10, 15, 16, 17, 25, 29, 31, 45, 47, 51, 57, 58, 60, 76, 84
R1		0.42	-2762.7	
R2		0.34, 0.65^c	-2757.4	

^a p , number of parameters in the model.

^bCodon numbering was based on the 2nd codon before the GG conserved motif in the N-terminal as number 1 ring to the last codon in the C-terminal

^cFor Poaceae and Solanaceae respectively

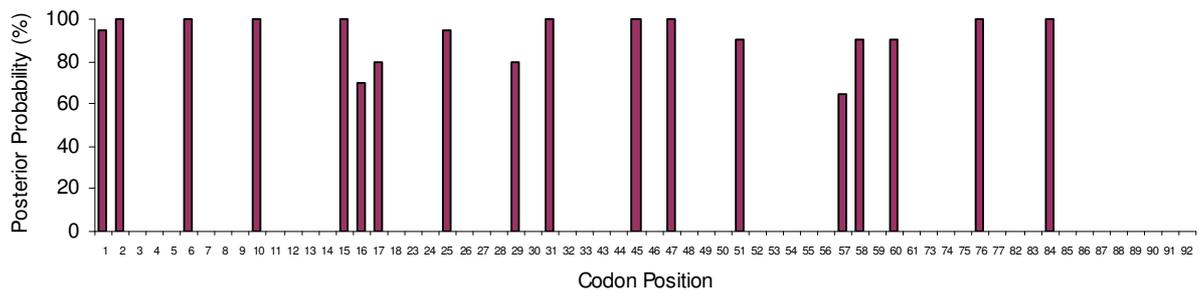


Figure 4.4 Location of positively selected codon sites (with Bayesian posterior probabilities greater than 60% under model M3) in Poaceae and Solanaceae cystatins. Codon 1 corresponds to the second codon before the conserved ‘GG’ motif. The dashed line indicates a posterior probability of 95%.

4.4.3 Mutation and expression of recombinant mutant papaya cystatins

Mutation success using the modified quick change protocol resulted in about 100 colonies on each plate after transformation of *E. coli*. Usually 2 out of 3 colonies picked for sequencing were positively mutated. Expression and purification under native conditions followed the recommended protocols in the QIAexpressionist kit manual. Figure 4.5 shows a 12% SDS-PAGE with successful expression and purification of mutants CYSI07D, CYSA53P, CYSA32V and CYSW78P. Mutant proteins were highly pure resulting in a single band on the SDS-PAGE gel, but tended to precipitate after buffer exchange. This problem was solved by adding Triton X100 up to the exchange buffer.

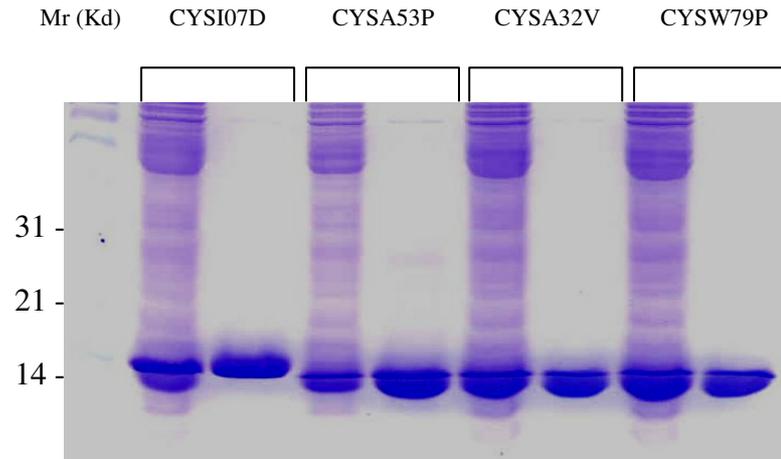


Figure 4.5 SDS-PAGE (12%) of the purified fractions of selected papaya cystatin mutants CYSI07D, CYSA53P CYSA32V and CYSW78P to establish purity during the purification process; lane 1: Molecular weight markers, lanes 2, 4, 6, and 8 crude extracts from an *E. coli* culture; lanes 3, 5, 7 and 9 the respective 1st elution from the purification column.

4.4.4 Inhibition activity of papaya cystatin mutants

To determine if the various mutations performed on the papaya cystatin resulted in any improvement inhibition, enzymatic assays were performed with all mutants and compared to the original native PC using papain, banana weevil and also black maize beetle gut extracts. As shown in Figure 4.5, 10 out of the 18 mutants showed a significant increase in inhibition of papain *in-vitro*. Mutant CYSA52P and CYSE55A further had the highest increase (6-fold) compared to the native PC. Mutant CYSE84A had a 5-fold, while CYStNT (truncation of the N-terminal trunk), CYSW79P, CYSI07D and CYSI07L all had a 2-fold increase. All increases greater than 5-fold were under positive selection pressure. Five mutants did not show any



improvement from the native PC, while 2 mutants, CYSC60T and CYSN85X, had a significant reduction in inhibitory activity compared to the native PC.

When the mutants were tested against banana weevil and black maize beetle gut enzymes, the increases in activity were less than for papain (Figure 4.6). Ten mutants showed significant increases ($p < 0.05$) against banana weevil gut proteases with CYSE84A (2-fold) while CYSA52P and CYSI07L with a 1.5-fold increase. For the black maize beetle, only CYSA52P showed a 2.5-fold increase. Mutants CYSE84A, CYSN97P, CYSE52Q, CYSI07L and CYSP03F, had non-significant increases of less than 1.5-fold ($p < 0.05$) in their inhibition capacity.

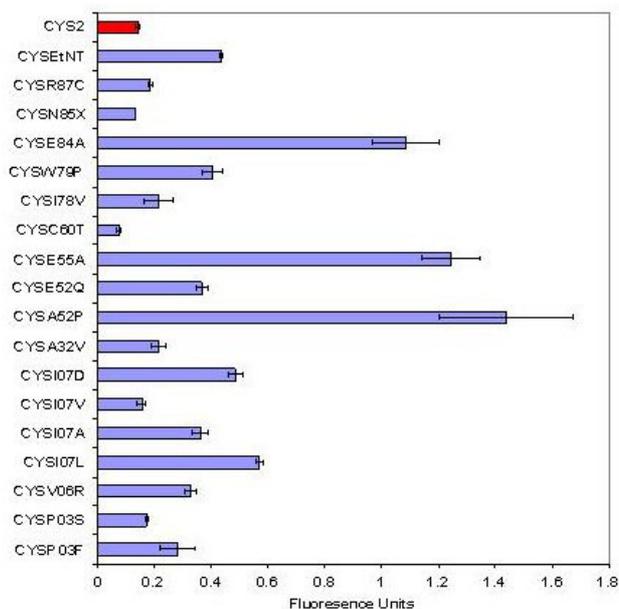
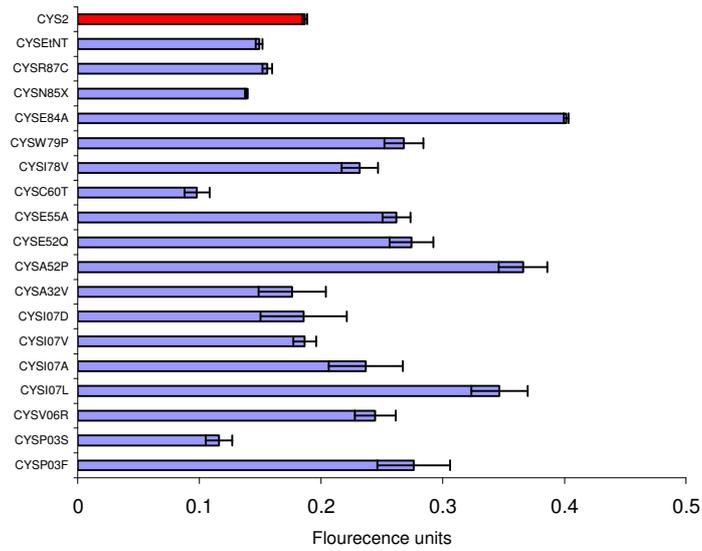


Figure 4.6 Comparison of inhibition activity between native PC (red bar) and 18 PC mutants. The inhibitors were tested by monitoring change in reaction rates of papain hydrolyzing Z-Phe-Arg-AMC a cathepsin-L specific substrate after addition of the inhibitor. Bars represent the mean \pm SE of 3 replications of difference in reaction rate before and after addition of the inhibitor in fluorescence units.

A



B

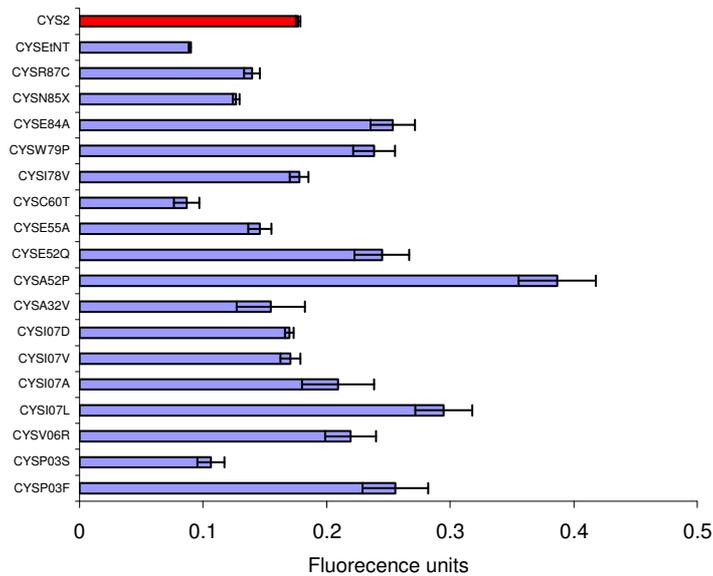


Figure 4.7 Inhibition activities between native PC (red bar) and 18 mutants of the papaya cystatin. Inhibitors were tested by monitoring change in reaction rates of banana weevil (A) and black maize beetle (B) gut extracts hydrolyzing Z-Phe-Arg-AMC substrate after addition of inhibitor. Bars represent mean \pm SE of 3 replications of difference in reaction rate before and after addition of inhibitor in florescent units.



4.5 Discussion

This part of the study showed that while there is high sequence conservation particularly in areas that are important to protein structure and function, there were also areas with high variability in amino acid sites in close proximity to the active site e.g the first binding loop. Such variability close to inhibitory sites has been documented for serine-type inhibitors of animal origin. This has been further shown to generate inhibitor variants with significantly different affinities for serine proteases (Creighton and Darby, 1989).

The occurrence of hypervariable amino acid sites among plant protease inhibitors as well as the high variability in affinity supports the idea that these inhibitors have been under selective pressure to evolve in response to herbivorous insect pests and protease diversification in the insects (Lopes *et al.*, 2004). The use and diversity of digestive proteases in coleopteran insects has been well documented (Murdock *et al.*, 1987). Walter *et al.* (1998) showed that the more advanced insects had the highest diversity of cysteine proteases in their gut. This indicates that they have evolved to overcome inhibitors and are perhaps have a more polyphagous feeding habit. Other studies have found that herbivorous insects are able eluding the inhibitory effects of phytocystatins by the use of 'cystatin-insensitive' digestive cysteine proteases (Cloutier *et al.*, 2000; Girard *et al.*, 1998; Michaud *et al.*, 1996) or by breakdown of cystatins using non-target proteases (Girard *et al.*, 1998; Michaud, 1997; Zhu-Salzman *et al.*, 2003).

Some proof has been provided in this study that at molecular level positive selection is active in phytocystatins. This is most likely within the variable amino acid residues in the active site cleft. Differential sensitivity to cystatins was previously identified in



the coleopteran insect *Callosobruchus maculatus*, challenged with soyacystatin, a wound-inducible cystatin from soybean (Moon *et al.*, 2004; Zhu-Salzman *et al.*, 2003). The accumulation of cystatin-insensitive proteases following cystatin ingestion, also observed for the potato herbivore *Leptinotarsa decemlineata* (Cloutier *et al.*, 2000; Gruden *et al.*, 2004), clearly supports the hypothesis of a co-evolution process. This is possibly driven by positive selection explaining the long-term interactions of cystatins with digestive cysteine proteases in plant–insect systems. In this study, the PC mutants CYSA52P, CYSE55A, CYSE84A and CYSI07L showed the highest and consistent improvement in inhibition of papain and protease activity of both banana weevil and black maize beetle. These mutations were all either at positive selection sites or in variable regions close to the active site of the phytocystatin.

In practice, the search for positive selection events among insect digestive cysteine proteases and phytocystatins could be useful in forthcoming years in interpreting the complex structural interactions taking place naturally between these presumably co-evolving proteins. It could further help in developing rationale strategies for the molecular improvement of phytocystatin variants with potential in plant protection. As a first step, the novel phytocystatin variants from this study should therefore also be tested in transgenic plants to prove their improved activity *in planta*. The identification of positively selected sites in phytocystatins could further be of general interest in biotechnology. Accumulated data on the functional characteristics of phytocystatins over the last 10 years have made these proteins not only attractive genes for pest control in plants but also for the control of cysteine proteases in various industrial and medical systems (Arai *et al.*, 2002).



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

CHAPTER 5

General discussion and future outlook



5.1 Summary

Crop improvement for pest resistance has continued to be a challenging task for many plant breeders worldwide. This is partly due to the fact that pest resistance is largely controlled by multiple genes and introgressing them into elite cultivar presents numerous challenges. Banana weevil is not an exception. Breeding of banana is difficult due to sterility, polyploidy and long generation time, and screening large populations of breeding material (hybrids) for weevil resistance is difficult to achieve with conventional breeding techniques. Breeding for such vegetatively propagated long generation crops has been rather by selection of naturally occurring resistance than conventional breeding. Yet, in many cases these selected lines may not have the productivity level as elite varieties. Regardless of the difficulties of generating pest resistance in crops, insect pests continue to destroy crop not only affecting yield in the field but damaging food already in storage.

The most important question facing the future of agriculture is therefore: How can the increased demand for food and other related products be met in the ever increasing world population? In advanced agricultural systems, increased use of fertilizer and pesticides may provide limited benefits, as they are already reaching optimum levels and are damaging to the environment. Similarly, future productivity cannot rely on solely increased irrigation and simply opening up of new lands. These options are also not available to the many resource poor largely subsistence farmers of Africa, Asia and South America. Thus, agricultural productivity and enhanced end-use quality in order to continue supporting humanity will need to exploit the new technologies of modern genetics coupled with environmentally sound cropping systems. Therefore, there has been a recent shift from conventional breeding to biotechnology involving



either molecular markers to pinpoint resistance traits in QTLs or direct engineering of genes from diverse species to crops to enhance resistance to pests and other diseases. One of the most successful control strategies for crop pests has been in recent year the development of *Bt* crops (mainly maize and cotton) that have revolutionalised these agricultural systems. However *Bt* technology is specific to Lepidopteran insects and most Coleopterans insects cannot be controlled in the same way.

At the onset of this PhD study, it was hypothesized that protease inhibitors, in particular cysteine protease inhibitors from plants, are potential candidates for the development of banana weevil resistance in banana. To prove this hypothesis, a vacuum infiltration assay was developed in which banana stems were infiltrated with recombinant phytocystatins and then fed to first instar weevil larvae. A first step to prove the correctness of the hypothesis was the finding that the banana weevil mainly employs cysteine proteases in particular cathepsin B and L for protein digestion. A second step of proof was that for the first time a modified *in-vivo* assay could used in which banana weevil larvae were fed stems infiltrated with phytocystatins. It was shown that early developmental rates were significantly reduced by more than 70% compared to the control. However, the presence of multiple forms will present a challenge to the strategy of using a single phytocystatin to target the weevil. Clearly the strategy should consider the use of multiple protease inhibitor forms including both serine type and cysteine types. In this regard, Ortega *et al* (1998) reported higher levels of mortality of larvae of the weevil *Aubeonymus mariaefranciscae* Roudier (Col.: Curculionidae) when fed to diets containing a combination of more than one inhibitor suggesting synergistic toxic effects. Serine proteases are, however, present in mammal digestive systems and would raise considerable food safety concerns when



used in a transgenic crop. Alternatives have to be employed by using either tissue specific or wound inducible promoters instead of constitutive promoter so that the transgene is targeted to a more specific site and time of expression.

A third step of proof was that action of phytocystatins can be improved by site-directed mutagenesis. All protein engineering strategies start with the hypothesis that the target protein has not yet achieved its maximum potential. In this study it has been shown that site-directed mutation was applied to papaya cystatin, 10 mutants showed improvement against papain, 10 against banana weevil gut extracts and 8 against the black maize beetle. This further illustrates the diversity of function and some specificity as some mutants had increased activity in only one of the insects.

By searching for sites for cystatin engineering the evolutionary dynamics of inhibitor-protease interactions and natural selection were also investigated in greater detail. This allowed understanding of the evolutionary relationships as well as the diversity in structure and function of phytocystatins. In general, there is a high diversity among phytocystatins. Diversity contains cystatins with functional multiple domains and some cystatins showed up in different taxonomic groups when a phylogenetic analysis was carried out. The diversity of evolutionary mechanisms in a single protein family is likely to be due to repeated interactions between these proteins and the continuous pressure to create variation.

By investigating evolutionary relationships the process of positive selection has been proved to occur in phytocystatins. Phytocystatins have amino acid sites that have during evolutionary time undergone positive selection. This study also provided



some evidence that mutations at these sites adding advantage to the host plant. Such positive selection sites offer the opportunity to modulate phytocystatins for improved activity or specificity.

Overall, this study has ultimately contributed to the advancement of science by providing new findings on the diversity of phytocystatins. It has also provided evolutionary evidence of positive selection in phytocystatins and has shown the importance of functional diversity both in plant defence proteins *vis a vis* pest protease. It has finally shown that cystatins can be improved by changing particular amino acid sites and several novel engineered cystatins have been created that can contribute to developing resistance to the banana weevil and also other Coleopteran insect pests.

5.2 Future outlook

The technology for developing insect-resistant transgenic plants is expanding very rapidly. Such plants have the potential to become in the future a part of the integrated pest management systems both for large commercial plantations but also helping resource poor farmers in Africa. With the development of several transgenic plants expressing *Bt* toxin, which are already on the market, clearly illustrates that gene technologies are a good strategy for developing insect pest resistance that is safe. In this regard phytocystatins have the advantage of very likely less regulatory concern since mammals and humans do not use cysteine proteases in their digestive systems.

More novel engineered cystatin mutants could therefore be created and tested in transgenic plants. In this regard, transgenic tobacco and banana are currently

produced expressing native and engineered phytocystatins to test for their efficiency to control insects. Transgenic approaches also include expression of combinations of different protease inhibitors (serine + cysteine protease inhibitor) or expression of cystatins with multiple mutations to possibly delay resistance in pests. However, the challenge with these proteins will be to develop cystatin variants tailored for the inhibition of specific proteases or sets of proteases. Activity of these target proteases should be efficiently controlled by phytocystatins without interfering with activity of endogenous plant proteases or being degraded by non-target cysteine proteases. Looking at the broader ecosystem level, it may be interesting to use poorly specific inhibitors to increase the number of target proteases, but maybe not an ideal choice as these may inhibit non target proteases unless if their activity can be modulated to make them specific to the targets and less affiants against non-target proteases.



REFERENCES

- Abe K. and Arai S. (1985) Purification of a cysteine protease inhibitor from rice, *Oryza sativa* L. japonica. *Agriculture Biology and Chemistry*, 49, 3349-3350.
- Abe K., Emori Y., Kondo H., Suzuki K. and Arai S. (1987) Molecular cloning of a cysteine protease inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds. *Journal of Biological Chemistry*, 262, 16793-16797.
- Abe M., Abe K., Domoto C. and Arai S. (1995) Two distinct species of corn cystatin in corn kernels. *Bioscience, Biotechnology and Biochemistry*, 59, 756-758.
- Abe M., Abe K., Kuroda M. and Arai S. (1992) Corn kernel cysteine protease inhibitor as a novel cystatin super family member of plant origin. Molecular cloning and expression studies. *European Journal of Biochemistry*, 209, 933-937.
- Abera A. M. K., Gold C.S. and Kyamanywa S. (1999) Timing and distribution of attack by the banana weevil (Coleoptera: Curculionidae) in East Africa highland banana (*Musa* spp.). *Florida Entomologist*, 82, 631-641.
- Abrahamson, M. (1994) Cystatins. In: *Methods in Enzymology* (Barrett, A. J. ed.). Academic Press, San Diego, pp. 685-700.
- Alpteter F., Diaz I., Mcauslane H., Gaddour K., Carbonero P. and Vasil I.K. (1999) Increased insect resistance in transgenic wheat stably expressing trypsin inhibitor. *Molecular Breeding*, 5, 53-63.
- Altschul S.F., Gish W., Miller W., Myers E.G. and Lipman D.J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215, 403-410.



- Annadana S., Peters J., Gruden K., Schipper A., Outchkourov N.S., Beekwilder M.J., Udayakumar M. and Jongsma M.A. (2002) Effects of cysteine protease inhibitors on oviposition rate of the western flower thrips, *Frankliniella occidentalis*. *Journal of Insect Physiology*, 48, 701–706.
- Applebaum S.W. (1985) Biochemistry of digestion. In: *Comprehensive Insect Physiology, Biochemistry and Pharmacology* (Kerkot, G.A. and Gilbert, L.I. (eds.)), New York, Pergamon Press 4, 279-311.
- Arai S., Watanabe H., Kondo H., Emori Y. and Abe K. (1991) Papain inhibitory activity of oryzacystatin, a rice seed cysteine protease inhibitor, depends on the central Gln-Val-Val-Ala-Gly region conserved among cystatin superfamily members. *Journal of Biochemistry*, 109, 294-298.
- Arai S. and Abe, K. (2000) Cystatin-based control of insects, with special reference to oryzacystatin. In: *Recombinant Protease Inhibitors in Plants* (Michaud, D., ed.). Georgetown, TX: Landes Bioscience, pp. 27–42.
- Arai S., Matsumoto, I., Emori, Y. and Abe, K. (2002) Plant seed cystatins and their target enzymes of endogenous and exogenous origin. *Journal of Agriculture and Food Chemistry*, 50, 6612–6617.
- Arinaitwe G. (2002) Genetic improvement of East African Highland Banana (*Musa* spp) via engineering for fungal disease resistance. In: *Novel Approaches to the Improvement of Banana Production in Eastern Africa- The Application of Biotechnological Methodologies Project Report (Annexes) INIBAP*, Montpellier France pp 45-53.
- Ashouri A., Michaud D. and Cloutier C. (2001) Recombinant and classically selected factors of potato plant resistance to the Colorado potato beetle, *Leptinotarsa*



decemlineata, variously affect the potato aphid parasitoid *Aphidius nigripes*.

Biocontrol 46: 401-18.

Atkinson H.J., Grimwood S. Johnston K. and Green J. (2004) Prototype demonstration of transgenic resistance to the nematode *Radopholus similis* conferred on banana by a cystatin. Transgenic Research, 13, 135–142.

Baker J.E., Woo S.M. and Mullen M.A. (1984) Distribution of proteinases and carbohydrates in the midgut of the larvae of the sweet potato weevil *Cyclas formicarius* and response of protease to inhibitors from sweet potato. Entomologia Experimentalis et Applicata, 36, 97-105.

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

Barrett A.J. (1986) The classes of proteolytic enzymes. In: Plant proteolytic enzymes (Dalling, M.J. ed.), Florida, CRC Press Inc. vol. 1, p. 1-16.

Belenghi B., Acconcia F., Trovato M., Perazzolli M., Bocedi, A., 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.

Bent A.F. (1996) Plant disease resistance genes: Function meets structure. Plant Cell, 8, 1757-1771.

Bergey D.R., Orozco-Cardenas M., de Moura D. S. and Ryan C. A. (1999) A wound- and systemin-inducible polygalacturonase in tomato leaves. Proceedings of the National Academy of Sciences of the United States of America, 96, 1756-1760.



- Bielawski J.P. and Yang Z. (2003) Maximum likelihood methods for detecting adaptive evolution after gene duplication. *Journal of Structural and Functional Genomics*, 3, 201–212.
- Birk Y., and Applebaum Y. (1960) Effect of soybean trypsin inhibitors on the development and midgut proteolytic activity of *Tribolium castaneum* larvae. *Enzymologia Acta Biocatalytica*, 22, 318-326.
- Birkenmeier G.F. and Ryan, C.A. (1998) Wound signaling in tomato plants: evidence that ABA is not a primary signal for defense gene activation. *Plant Physiology*, 17, 687-693.
- Bishop J.G., Dean A.M. and Mitchell-Olds T. (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant–pathogen coevolution. *Proceedings of the National academy of Sciences of the United States of America*, 97, 5322–5327.
- Bode W. and Huber, R. (2000) Structural basis of the endoproteinase–protein inhibitor interaction. *Biochimica et Biophysica Acta*, 1477, 241–252.
- Bode W., Engh R., Musil D., Thiele U., Huber R., Karshikov A., Brzin J., Kos J. and Turk, V. (1988) The 2.0 Å crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO Journal*, 7, 2593–2599.
- Boulter D. (1993) Insect pest control by copying nature using genetically engineered crops. *Phytochemistry*, 34, 1453-1466.
- Bown D.P., Wilkinson H.S. and Gatehouse, J.A. (1997) Differentially regulated inhibitor sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. *Insect Biochemistry and Molecular Biology*, 27, 625-638.



- Broadway RM. (2000) The adaptation of insects to protease inhibitors. In: Recombinant Protease Inhibitors in Plants (Michaud, D. ed.), Landes Bioscience/Eurekah.com, Georgetown, TX pp. 80–88.
- Broadway R.M. and Duffey, S.S. (1986) Plant protease inhibitors: mechanism of action and effect on the growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exigua*. *Journal of Insect Physiology*, 32, 827-833.
- Brovosky D. (1986) Proteolytic enzymes and blood digestion in the mosquito *Culex nigripalpus*. *Archives of Insect Biochemistry and Physiology*, 3, 47-160.
- Brunelle F., Nguyen-Quoc B., Cloutier C. and Michaud D. (1999) Protein hydrolysis by Colorado potato beetle, *Leptinotarsa decemlineata*, digestive proteases: the catalytic role of cathepsin D. *Archives of Insect Biochemistry and Physiology*, 42, 88-98.
- Bryant J., Green T.R. Gurusaddaiah T. and Ryan, C.A. (1976) Protease inhibitor II from potatoes: isolation and characterization of its promoter components. *Biochemistry*, 5, 3418-3424.
- Brzin J., Popovic T., Ritonja A., Puizdar V. and Kidric M. (1998) Related cystatin inhibitors from leaf and from seed of *Phaseolus vulgaris* L. *Plant Science*, 138, 17-26.
- Brzin J., Ritonja A., Popovic T. and Turk V. (1990) Low molecular mass protein inhibitor of cysteine proteinases from soybean. *Biological Chemistry Hoppe-Seyler*, 371, 167-170.
- Buddenburg W.J., Ndiege I.O., Karago F.W. and Hanson B.S.S. (1993) Behavioural and electrophysiological responses of the banana weevil *Cosmopolites sordidus* to host plant volatiles. *Journal of Chemical Ecology*, 19, 267-277.



- Burgess E.P.J. Steven P.S., Keen G.K., Laing W.A. and Christeller J.T. (1991) Effects of protease inhibitors and dietary protein level on the black field cricket *Teleogryllus commodus*. *Entomologia Experimentalis et Applicata*, 61, 123-130.
- Campos F.A.P., Xavier-Filho J., Silva C.P. and Ary M.B. (1989) Resolution and partial characterization of proteinases and α -amylases from midguts of larvae of the bruchid beetle *Callosobruchus maculatus* (F.). *Comparative Biochemistry and Physiology-B*, 92, 51-57.
- Carbonero P., Royo J., Díaz I., García-Maroto F., Gonzalez-Hidalgo E., Gutierrez C. and Casanera, P. (1993) Cereal inhibitors of insect hydrolases (α -amylases and trypsin): genetic control, transgenic expression and insect pests. In: *Workshop on engineering plants against pests and pathogens* (Bruening, G.J., García-Olmedo, F. and Ponz, F.J. eds.). Instituto Juan March de Estudios Investigaciones, Spain.
- Carlens K. (2002) Nematode control in East African regions by the transgenic banana (*Musa* spp.) expressing different types of lectins. In: *Novel Approaches to the improvement of banana production in Eastern Africa-the application of biotechnological methodologies Project Report (Annexes)* INIBAP, Montpellier France.
- Carlini C.R., Grossi-de-Sa, M.F., (2001) Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. *Toxicon*, 40, 1515-1539.
- Carreel F., Abadies C., Carlier K., Tomekpe K., Lagoda P.J.L. and Bakry F. (1999) Genome mapping and genetic analysis of the black leaf streak resistance in



banana. In: Abstracts of the International Symposium on the Molecular and Cellular Biology of banana, Ithaca, New York INIBAP.

Carter P. (1986) Site-directed mutagenesis. *Biochemistry Journal*, 237, 1-7

Charity J.A., Anderson Marilyn A., Bittisnich DJ., Whitecross M. and Higgins, T.J.V. Transgenic tobacco and peas expressing a protease inhibitor from *Nicotiana glauca* have increased insect resistance. *Molecular Breeding*, 5, 357-365.

Chen, M.S., Johnson B., Wen L., Muthukrishnan S., Kramer K.J., Morgan T.D., and Reeck G.R. (1992) Rice cystatin: bacterial expression, purification, cysteine protease inhibiting activity and insect growth suppressing activity of a truncated form of the protein. *Protein Expression and Purification*, 3, 41-49.

Chye M. L., Sin S.F., Xu Z.F. and Yeung E.C. (2006) Serine proteinase inhibitor proteins: exogenous and endogenous functions, *In Vitro Cellular and Development Biology – Plant*, 42, 100–108.

Cloutier C., Jean C., Fournier M., Yelle S., Michaud D. (2000) Adult colorado potato beetles, *Leptinotarsa decemlineata* compensate for nutritional stress on oryzacystatin I-transgenic potato plants by hypertrophic behaviour and over-production of insensitive proteinases. *Archives of Insect Biochemistry and Physiology*, 44, 69-81.

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

Collins P.J., Treverrow N.L. and Lambkin T.M. (1991) Organophosphorous insecticide resistance and its management in the banana weevil borer, *Cosmopolites sordidus* Germar (Coleoptera: Curculionidae) in Australia. *Crop Protection*, 10, 215-221.



- Creighton T.E. and Darby N.J. (1989) Functional evolutionary divergence of proteolytic enzymes and their inhibitors. *Trends in Biochemical Sciences*, 14, 319–324.
- Crouch J. H., Vuylsteke D., and Ortiz R. (1998). Perspectives on the application of biotechnology to assist the genetic enhancement of plantain and banana (*Musa spp*). *Electronic Journal of Biotechnology*, 1, <http://www.ejb.org>.
- Crouch J.H. and Crouch H.K. (1999) Application of DNA markers in plantain and banana. In: *DNA Marker-assisted Improvement of the Staple Crops of Sub-Saharan Africa* (Crouch J.H. and Tenkouano A. eds.), IITA Ibadan, Nigeria, pp. 163-167.
- Crouch J.H., Ortiz R., Crouch H.K., Jarret R.L., Ford-Lloyd B.V., Howell E.C. and Newbury H.J. (2000) Utilization of molecular genetic techniques in support of plantain and banana improvement. *Acta Horticulturae*, 540, 185-191.
- Delaney T.P., Uknes S., Vernooij B., Friedrich L., Weymann K., Negrotto D., Gaffney T., Gut-Rella M., Kessmann H., Ward E. and Ryals J. (1994) A central role of salicylic acid in plant disease resistance. *Science*, 266, 1247-1250.
- Delledonne M., Allegro G., Belenghi B., Balestrazzi A., Picco F., Levine A., Zelasco S., Calligari P. and Confalonieri M. (2001) Transformation of white poplar (*Populus alba* L.) with a novel *Arabidopsis thaliana* cysteine proteinase inhibitor and analysis of insect pest resistance. *Molecular Breeding*, 7, 35-42.
- DeVries J. and Toenniessen G. (2001) *Securing the harvest: biotechnology, breeding and seed systems for African crops*. CABI International, New York.



- Ding X., Gopalakrishnan B., Johnson L.B., White F.F., Wang X., Morgan T.D., Kramer K.J. and Muthukrishnan S. (1998). Insect resistance of transgenic tobacco expressing an insect chitinase gene. *Transgenic Research*, 7, 77-84.
- Duan X., Li X., Xue Q., Abo-el-Saad M., Xu D. and Wu R. (1996) Transgenic rice plants harbouring an introduced potato protease inhibitor II gene are insect resistant. *Nature Biotechnology*, 14, 494-498.
- Dunaevskii Y.E. Gladysheva I.P., Pavlukova E.B., Beliakova G.A., Gladyshev D.P., Papisova A.I., Larionova N.I. and Belozersky M.A. (1997) The anionic protease inhibitor BBWI-1 from buckwheat seeds: kinetic properties and possible biological role. *Physiologia Plantarum*, 100, 483-488.
- Dunwell J.M. (1998) Novel food products from genetically modified crop plants: methods and future prospects, *International Journal of Food Science and Technology*, 33, 205-213.
- Edmonds H.S., Gatehouse L.N., Hilder V.A. and Gatehouse J.A. (1996) The inhibitory effects of the cysteine protease inhibitor, oryzacystatin, on digestive proteases and on larval survival and development of the southern corn rootworm (*Diabrotica undecimpunctata howardi*). *Entomologia Experimentalis et Applicata*, 78, 83-94.
- Eguchi M., Iwamoto A. and Yamaguchi K. (1982) Interaction of proteases from the midgut lumen, epithelial and peritrophic membrane of the silkworm *Bombyx mori* L. *Comparative Biochemistry and Physiology-A*, 72, 359-363.
- Epple P., April K. and Bohlmann H. (1995) An *Arabidopsis thaliana* thionin gene is inducible via a signal transduction pathway different from that for pathogenesis related protein. *Plant Physiology*, 109, 813-820.



- Fabrick J., Behnke C., Czapla T., Bala K., Rao A. G., Kramer K.J. and Reeck G.R. (2002) Effects of a potato cysteine proteinase inhibitor on midgut proteolytic enzyme activity and growth of the southern corn rootworm, *Diabrotica undecimpunctata* Howardi (Coleoptera: Chrysomelidae), *Insect Biochemistry and Molecular Biology*, 32, 405-415.
- Farley P.C., Ramsay R.J., Sullivan P.A. and Laing W.A. (1998). Purification, characterization and cloning of an aspartic protease inhibitor from squash phloem exudate. *European Journal of Biochemistry*, 254, 160-167.
- Felsenstein J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164-166.
- Fernandes K.V., Sabelli P.A., Barrat D.H., Richardson M., Xavier-Filho J. and Shewry P.R. (1993) The resistance of cowpea seeds to bruchid beetles is not related to levels of cysteine protease inhibitors. *Plant Molecular Biology*, 23, 215-219.
- Fisher C. L. and Pei G. K. (1997) Modification of a PCR-based site-directed mutagenesis method. *Biotechniques*, 23, 570-574.
- Fogain R., Price, N.S., (1994) Varietal screening of some *Musa* cultivars for susceptibility to the banana weevil, *Cosmopolites sordidus* (Coleoptera: Curculionidae). *Fruits*, 49, 247-251.
- Forgain R., Messiaen S. and Foure E. (2002) Studies on the banana borer weevil in Cameroon. *InfoMusa* 11, 1- 4.
- Frazao C., Isabel B., Julia C., Claudio M.; Verissimo P., Faro C., Pires E. Cooper J. and Carrondo M.A. (1999) Crystal structure of cardosin A, a glycosylated and Arg-Gly-Asp- containing aspartic protease from the flowers of *Cyanara cardunculus* L. *Journal of Biological Chemistry*, 274, 27694-27701.



- Gaddour K., Vicente-Carbajosa J., Lara P., Isabel-Lamoneda I., Diaz I., Carbonero P. (2001) A constitutive cystatin-encoding gene from barley (Icy) responds differentially to abiotic stimuli. *Plant Molecular Biology*, 45, 599-608.
- Garcia-Olmedo S. F., Sanchez M. G., Gomez R.L., Royo J. and Carbonero, P. (1987) Plant proteinaceous inhibitors of proteinases and α -amylases. *Oxford Survey Plant Molecular and Cell Biology*, 4, 275-334.
- Gatehouse A.M.R., Shi Y., Powell K.S., Brough C., Hilder V. A., Hamilton W.D.O., Newell C.A., Merryweather A., Boulter D. and Gatehouse J.A. (1993) Approaches to insect resistance using transgenic plants. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences*, 342, 279-286.
- Gatehouse A.M.R. and Boulter D. (1993) Assessment of the antimetabolic effects of trypsin inhibitors from cowpea (*Vigna unguiculata*) and other legumes on development of the bruchid beetle *Callosobruchus maculatus*. *Journal of the Science of Food and Agriculture*, 34, 345-350.
- Gatehouse A.M.R., Butler K.J., Fenton K.A. and Gatehouse, J.A., (1985) Presence and partial characterisation of a major proteolytic enzyme in the larval gut of *Callosobruchus maculatus*. *Entomologia Experimentalis et Applicata*, 39, 279-286.
- Gatehouse A.M.R., Davidson G.M., Newell C.A., Merryweather A., Hamilton, W.D.O., Burgess E.P.J., Gilbert R.J.C. and Gatehouse J.A. (1997) Transgenic potato plants with enhanced resistance to the tomato moth, *Lacanobia oleracea*: growth room trials. *Molecular Breeding*, 3, 49-63.
- Gerald R.R., Kramer K.J., Baker J.E., Kanost J.F. and Behke C.A. (1997) Protease inhibitors and resistance of transgenic plants to insects. In: *Advances in*



- Insect Control the Role of Transgenic Plants (Carozi, N. and Koziel, M., eds.), Taylor and Francis, London, pp. 157-183.
- Gholizadeh A., Santha I., Lodha M. and Kapoor H. (2005) Cloning and expression of a small cDNA fragment encoding strong antiviral peptide from *Celosia cristata* in *Escherichia coli*. *Biochemistry*, 70, 1005-1010.
- Girard C., Rivard D., Kiggundu A., Kunert K., Gleddie S.C., Cloutier C. and Michaud D. (2007) A multicomponent, elicitor-inducible cystatin complex in tomato, *Solanum lycopersicum*. *New Phytologist*, 173, 841–851.
- Girard C., Le Métayer M., Bonadé-Bottino M., Pham-Delègue M.H. and Jouanin L. (1998) High level of resistance to protease inhibitors may be conferred by proteolytic cleavage in beetle larvae. *Insect Biochemistry and Molecular Biology*, 28, 229-237.
- Gold C.S., Bagabe M.I. and Sendege R., (1999a). Banana weevil, *Cosmopolites sordidus* (Germar) (Coleoptera: Curculionidae): tests for suspected resistance to carbofuran and dieldrin in the Masaka District, Uganda. *African Entomology*, 7, 189-196.
- Gold C. S., Karamura E. B., Kiggundu A., Bagamba F. and Abera A. M. K. (1999b). Geographic shifts in highland cooking banana (*Musa* spp. group AAA-EA) production in Uganda. *International Journal of Sustainable Development and World Ecology*, 6, 45-59.
- Gold C. S., Speijer P.R., Karamura E.B. and Rukazambuga D.N. (1994) Assessment of banana weevils in East African highland banana systems and strategies for control. In: *Banana Nematodes and Weevil Borers in Asia and the Pacific* (Valmayor, R.V., Davide, R.G., Stanton, J.M., Treverrow N.L. and Roa V.N.eds.), INIBAP, Los Banoss, Philippines. pp 170-190.



- Gold C.S. (1998) Banana weevil: ecology pest status and prospects for integrated control with emphasis on East Africa. In: Proceedings of a Symposium on Biological Control in Tropical Habitats: Third International Conference on Tropical Entomology (Saini, S.K. ed.), ICIPE, Nairobi, Kenya, pp. 49-74.
- Gold C.S. (2000) Biology and integrated pest management of banana weevil, *Cosmopolites sordidus* (Germar). In: Advancing banana and plantain R&D in Asia and the Pacific (Molina, A.B., Roa, V.N. and Maghuyop, M.A.G. eds.), INIBAP-ASPNET, pp. 28-33.
- Gold C.S. and Messiaen S. (2000) The banana weevil *Cosmopolites sordidus*. *MusaPest* INIBAP Fact Sheet No 4.
- Gold C.S., Peña J.E. and Karamura E.B. (2001) Biology and integrated pest management of the banana weevil, *Cosmopolites sordidus* (Germar) (Coleoptera: Curculionidae). *Integrated Pest Management Reviews*, 6, 79-155.
- Gonnet G.H., Cohen M.A. and Benner S.A. (1992) Exhaustive matching of the entire protein sequence database. *Science*, 256, 1443-5.
- Gourinath S., Alam N., Srinivasan A., Betzel C. and Singh T.P. (2000) Structure of the bifunctional inhibitor of trypsin and alpha-amylase from ragi seeds at 2.2Å resolution. *Acta Crystallography D Biology Crystallography*, 56, 287-293.
- Graham J., Gordon S.C. and Mcnicol R.J. (1997) The effect of the CpTI gene in strawberry against attack by vine weevil (*Otiorhynchus sulcatus* F. (Coleoptera: Curculionidae). *Annals of Applied Biology*, 131, 133-139.



- Graham J.S. and Ryan, C.A. Accumulation of metallocarboxy-peptidase inhibitor in leaves of wounded potato plants. *Biochemical and Biophysics Research Communication*, 101, 1164-1170.
- Green T.R. and Ryan C.A. (1972) Wound-induced protease inhibitors in plant leaves: a possible defence mechanism against insects. *Science*, 175, 776-777.
- Greenblatt H.M., Ryan C.A. and James M.N.G. (1989) Structure of the complex *Streptomyces griseus* protease B and polypeptide chymotrypsin inhibitor-I at 2.1 Å resolution. *Journal of Molecular Biology*, 205, 201-228.
- Gruden K., Kuipers A.G.J., Guncar G., Slapar N., Strukelj B. and Jongsma M.A. (2004) Molecular basis of Colorado potato beetle adaptation to potato plant defence at the level of digestive cysteine proteinases. *Insect Biochemistry and Molecular Biology*, 34, 365–375.
- Guex N. and Peitsch M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis*, 18, 2714-2723.
- Habu Y., Peyachoknagul S., Sakata Y., Fukasawa K. and Ohno T. (1997) Evolution of a multigene family that encodes the Kunitz chymotrypsin inhibitor in winged bean: a possible intermediate in the generation of a new gene with a distinct pattern of expression. *Molecular General Genetics*, 254, 73-80.
- Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98.
- Havkioja E. and Neuvonen L. (1985) Induced long-term resistance to birch foliage against defoliators: defence or incidental. *Ecology*, 66, 1303-1308.



- Heath R.G., McDonald J.T., Christeller M., Lee K., Bateman J., West R., Van Heeswijck and Anderson M.A. (1997) Protease inhibitors from *Nicotiana glauca* enhance plant resistance to insect pests. *Journal of Insect Physiology*, 9, 833-842.
- Hilder V.A., and Boulter D. (1999) Genetic engineering of crop plants for insect resistance: a critical review. *Crop Protection*, 18, 177-191.
- Hilder V.A., Gatehouse A.M.R., Sheerman S.E., Barker R.F. and Boulter D. (1987) A novel mechanism of insects resistance engineered into tobacco. *Nature*, 330, 160-163.
- Hilder V.A., Barker R.F., Samour R.A., Gatehouse A.M.R., Gatehouse J.A. and Boulter D. (1989) Protein and cDNA sequences of Bowman-Birk protease inhibitors from the cowpea (*Vigna unguiculata* Walp.). *Plant Molecular Biology*, 13, 701-710.
- Hildmann T., Ebneith M., Pena-Cortes H., Sanchez-Serrano J.J., Willmitzer L. and Prat S. (1992) General roles of abscisic and jasmonic acids in gene activation as a result of mechanical wounding. *Plant Cell*, 4, 1157-1170.
- Hirashiki I., Ogata F., Yoshida N., Makisumi S., Ito A. (1990) Purification and complex formation analysis of a cysteine protease inhibitor (cystatin) from seeds of *Wisteria floribunda*. *Journal of Biochemistry*, 108, 604-608.
- Hobday S.M., Thurmaen D.A. and Barber D.J. (1973) Proteolytic and trypsin inhibitory activities in extracts of germinating *Pisum sativum* seeds. *Phytochemistry*, 12, 1041-1046.
- Hollander-Czytko H., Andersen J.L. and Ryan C.A. (1985) Vacuolar localization of wound-induced carboxy peptidase inhibitor in potato leaves. *Plant Physiology*, 78, 76-79.



- Houseman J.G. and Downe A.E.R. (1983) Cathepsin D-like activity in the posterior midgut of Hemipteran insects. *Comparative Biochemistry and Physiology-B*, 75, 509-512.
- Houseman J.G., Campbell F.C. and Morrison P.E. (1987) A preliminary characterization of digestive proteases in the posterior midgut of the stable fly *Stomoxys calcitrans* (L.) (Diptera: Muscidae). *Insect Biochemistry*, 17, 213-218.
- Houseman J.G., Downe A.E.R. and Philogene B.J.R. (1987) Partial characterization of protease activity in the larval midgut of the European corn borer *Ostrinia nubilalis* Hubner (Lepidoptera: Pyralidae). *Canadian Journal of Zoology*, 67, 864-868.
- Huang Y.J., To K.Y., Yap M.N., Chiang W.J., Suen D.F. and Chen S.C.G. (2001) Cloning and characterization of leaf senescence up-regulated genes in sweet potato. *Physiologia Plantarum*, 113, 384-391.
- Hubinger G., Gruber P., Vollmann U., Kraft D., Ferreira F. and Himly M. (1999) Molecular cloning and nucleotide sequence analysis of a cDNA (Accession No. AF143677) Encoding the cystatin homolog from Mugwort (*Artemisia vulgaris*) pollen (PGR 99-078). *Plant Physiology*, 120, 634-634.
- Irie K., Hosoyama H., Takeuchi T., Iwabuchi K., Watanabe H., Abe M., Abe K. and Arai S. (1996) Transgenic rice established to express corn cystatin exhibits strong inhibitory activity against insect gut proteases. *Plant Molecular Biology*, 30, 149-157.
- Ishimoto M., Sato T., Chrispeels M.J. and Kitamura K. (1996) Bruchid resistance of transgenic azuki bean expressing seed α -amylase inhibitor of common bean. *Entomologia experimentalis et applicata*, 79, 309-315.



- Ittyeipe K. (1986) Studies on host preference of banana weevil borer, *Cosmopolites sordidus* GERM. (Curculionidae-Coleoptera). *Fruits*, 41, 375-79.
- Jacinto T., Fernandes K.V.S., Machando O.L.T. and Siqyueira-Junior C.L. (1998) Leaves of transgenic tomato plants over expressing prosystemin accumulate high levels of cystatin. *Plant Science*, 138, 35-42.
- Jackson R.M., Gabb H.A. and Sternberg M.J.E. (1998) Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *Journal of Molecular Biology*, 276, 265–285.
- Jenko S., Dolenc I., Guncar G., Dobersek A., Podobnik M. and Turk D. (2003) Crystal structure of stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo- and exopeptidases. *Journal of Molecular Biology*, 326, 875-885.
- Johnson R., Narraez J., An G. and Ryan C.A. (1989) Expression of protease inhibitors I and II in transgenic tobacco plants: effects on natural defense against *Manduca sexta* larvae. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 9871-9875.
- Jongsma M.A. and Bolter C. (1997) The adaptation of insects to plant protease inhibitors. *Journal of Insect Physiology*, 43, 885-895.
- Joshi B., Sainani M., Bastawade K., Gupta V.S. and Ranjekar P.K. (1998) Cysteine protease inhibitor from pearl millet: a new class of antifungal protein. *Biochemical and Biophysical Research Communications*, 246, 382-387.
- Jouanin L., Bonade-Bottino M., Girard C., Morrot G. and Giband M. (1998) Transgenic plants for insect resistance. *Plant Science*, 131, 1-11.



- Keilova H. and Tomasek V. (1976) Isolation and properties of cathepsin D inhibitor from potatoes. *Collection of Czechoslovak Chemical Communication*, 41, 489-497.
- Kervinen J., Tobin G.J., Costa J., Waugh D.S., Wlodawer A. and Zdanov A. (1999) Crystal structure of plant aspartic protease prophytepsin: inactivation and vacuolar targeting. *Journal of European Molecular Biology Organization*, 18, 3947-3955.
- Kiggundu A., Gold C.S, Labuschagne M., Vuylsteke H.K. and Louw S.V.D.M. (2003a) Levels of host plant resistance to banana weevil *Cosmopolites sordidus* (Germar) (Coleoptera: Curculionidae) in African *Musa* germplasm. *Euphytica*, 133, 267-277.
- Kiggundu A., Pillay M., Viljoen A., Gold C., Tushemereirwe W. and Kunert K. (2003b) Enhancing banana weevil (*Cosmopolites sordidus*) by genetic modification. A perspective. *African Journal of Biotechnology*, 2, 563-569.
- Kiggundu A., Vuylsteke D. and Gold C. (1999) Recent advances in host plant resistance to banana weevil, *Cosmopolites sordidus* (Germar) In: *Mobilizing IPM for sustainable banana production in Africa Proceedings of a workshop on banana IPM* (Frison, E. A., Gold, C.S., Karamura, E. B. and Sikora, R. A., eds), INIBAP, Montpellier, France, pp. 87-96.
- Kimura M., Ikeda T., Fukumoto D., Yamasaki N. and Yonekura M. (1995) Primary structure of a cysteine protease inhibitor from the fruit of avocado (*Persea americana* Mill) *Bioscience Biotechnology and Biochemistry*, 59, 2328-2329.
- Kishimoto N., Higo H., Abe K., Arai S., Saito A. and Higo K. (1994) Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of



cDNA markers of known functions. *Theoretical and Applied Genetics*, 88, 722-726.

Kitch L.W. and Murdock L.L. (1986) Partial characterization of a major gut thiol protease from larvae of *Callosobruchus maculatus* (F.). *Archives of Insect Biochemistry and Physiology*, 3, 561-575.

Koiwa H., Bressan R.A. and Hasegawa P.M. (1997) Regulation of protease inhibitors and plant defense. *Trends in Plant Science*, 2, 379-384.

Koiwa H., Shade R.E, Zhu-Salzman K., D'Urzo M.P., Murdock L.L., Bressan R.A. and Hasegawa P.M. (2000) A plant defensive cystatin (soyacystatin) targets cathepsin L-like digestive cysteine proteinases (DvCALs) in the larval midgut of western corn rootworm (*Diabrotica virgifera virgifera*). *FEBS Letters*, 471, 67-70.

Koiwa K., Shade R.E., Zhu-Salzman K., Subramanian L., Murdock L.L., Nielsen S.S., Bressan R.A. and Hasegawa P.M. (1998) Phage display selection can differentiate insecticidal activity of soybean cystatins. *Plant Journal*, 14, 371-379.

Kondo H., Abe K., Nishimura I., Watanabe H., Emori Y. and Arai S. (1990) Two distinct cystatin species in rice seeds with different specificities against cysteine proteinases. Molecular cloning, expression, and biochemical studies on oryzacystatin-II. *Journal of Biological Chemistry*, 265, 15832-15837.

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.



- Koradi R., Billeter M., Wüthrich K. 1996. MOLMOL: A program for display and analysis of macromolecular structures, *Journal of Molecular Graphics*, 14, 51-55.
- Kouzuma Y., Kawano K., Kimura M., Yamasaki N., Kadowaki T. and Yamamoto K. (1996) Purification, characterization, and sequencing of two cysteine protease inhibitors, Sca and Scb, from sunflower (*Helianthus annuus*) seeds *Journal of Biochemistry*, 119, 1106-1113.
- Kouzuma Y., Inanaga H., Doi-Kawano K., Yamasaki N. and Kimura M. (2000) Molecular cloning and functional expression of a cDNA encoding the cysteine protease inhibitor with three cystatin domains from sunflower seeds *Journal of Biochemistry*, 128, 161-166.
- Krattiger A.F. (1997) Insect resistance in crops: a case study of *Bacillus thuringiensis* (Bt) and its transfer to developing countries. International Service for the Acquisition of Agri-biotech Applications (ISAAA) Briefs, Ithaca, USA 2: p.42.
- Kunitz M. (1994) Crystallization of a trypsin inhibitor from soybean. *Science*, 101, 668-669.
- Kuroda M., Kiyosaki T., Matsumoto I., Misaka T., Arai S. and Abe K (2001) Molecular cloning, characterization, and expression of wheat cystatins. *Bioscience Biotechnology and Biochemistry*, 65, 22-28.
- Laboissiere M.C.A., Young M.M., Pinho R.G., Todd S., Fletterick R.J., Kuntz I. and 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.



- Laskowski R.A., MacArthur M.W., Moss D.S. and Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26, 283-291.
- Laskowski M. Jr. and Kato, I. (1980) Protein inhibitors of proteinases. *Annual Review of Biochemistry*, 49, 685-693.
- Lawrence P.K. and Koundal K.R. (2002) Plant protease inhibitors in control of phytophagous insects. *Electronic Journal of Biotechnology*, 5, 93-109.
- Lawrence P. K., Nirmala J. and Koundal K.R. (2001) Nucleotide sequence of a genomic clone encoding a cowpea (*Vigna unguiculata* L.) trypsin inhibitor. *Electronic Journal of Biotechnology*, 4, 1 <http://www.ejbiotechnology.info/content/vol4/issue1/full/4/index.html>. ISSN 0717-3458.
- LeBerre-Anton, V., Bompard-Gilles C., Payan F., Rouge P. (1997) Characterization and functional properties of the alpha-amylase inhibitor (alpha A-1) from kidney bean (*Phaseolus vulgaris*) seeds. *Biochemica et Biophysica Acta*, 1343, 31-40.
- Lecardonnell A., Chauvin L., Jouanin L.O., Beaujean A., Prevost G. and Sangwan-Novrreel B. (1999) Effect of rice cystatin I expression in transgenic potato on Colorado potato beetle larvae. *Plant Science*, 140, 71-79.
- Lemos F.J.A., Xavier-Filho J. and Campos F.A.P. (1987) Proteinases of the midgut of *Savrotes subfasciatus* larvae. *Agricultural Biology and Technology*, 80, 46-49.
- Leple J.C., Bonade M., Bontt S., Augustin G., Pilate V., Dumanoise L. R, Delplanque A., Cornu D. and Jouanin L. (1995) Toxicity to *Chrysomela tremulae* (Coleoptera: chrysomelidae) of transgenic polars expressing a cysteine inhibitor. *Molecular Breeding*, 1: 319-328.



- Lheureux F., Carreel F., Jenny C. Lockhart B.E.L. and Iskra-Caruana M.L. (2003) Identification of genetic markers linked to banana streak disease expression in inter-specific *Musa* hybrids. *Theoretical and Applied Genetics*, 106, 594-598.
- Li Z., Sommer A., Dingermann T. and Noe C.R. (1996) Molecular cloning and sequence analysis of a cDNA encoding a cysteine protease inhibitor of *Sorghum bicolor* seedlings. *Molecular and General Genetics*, 251, 499-502.
- Li, Y.E., Zhu Z., Chen Z.X., Wu X., Wang W. and Li S.J. (1998) Obtaining transgenic cotton plants with cowpea trypsin inhibitor. *Acta Gossypii Sinica*, 998, 237-243.
- Liang C., Brookhart G., Feng G.H., Reeck G.R. and Kramer K.J. (1991) Inhibition of digestive proteases of stored grain Coleoptera by oryzacystatin, a cysteine protease inhibitor from rice seed. *FEBS Letters*, 278, 139-142.
- Lim C.O., Lee S.I., Chung W.S., Park S.H., Hwang I., Cho M.J. (1996) Characterization of a cDNA encoding cysteine protease inhibitor from Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*) flower buds. *Plant Molecular Biology*, 30, 373-379.
- Lipke H., Fraenkel G.S. and Liener I.E. (1954) Effects of soybean inhibitors on growth of *Tribolium confusum*. *Journal of the Science of Food and Agriculture*, 2, 410-415.
- 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.
- MacPhalen C.N. and James M.N.G. (1987) Crystal and molecular structure of the serine protease inhibitor CI-2 from barley seeds. *Biochemistry*, 26, 261-269.



- Malone M. and Alarcon J.J. (1995) Only xylem-borne factors can account for systemic wound signalling in the tomato plant. *Planta*, 196, 740-746.
- Margis R., Reis E.M. and Villeret V. (1998) Structural and phylogenetic relationships among plant and animal cystatins. *Archives of Biochemistry and Biophysics*, 359, 24-30.
- Mark S.J. and Overington J.P. (1993) A Structural Basis for Sequence Comparisons: An Evaluation of Scoring Methodologies *Journal of Molecular Biology*, 233, 716-738.
- 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. *Biochemistry Journal*, 330, 833–838.
- Mazumdar-Leighton S. and Broadway R.M. (2001) Identification of six chymotrypsin cDNAs from larval midguts of *Helicoverpa zea* and *Agrotis ipsilon* feeding on the soybean (Kunitz) trypsin inhibitor. *Insect Biochemistry and Molecular Biology*, 31, 633–644.
- McGurl B., Orozco-Cardenas M., Pearce G. and Ryan C.A. (1994) Over expression of the prosystemin gene in transgenic tomato plants generates a systemic signal that constitutively induces protease inhibitor synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 9799-9802.
- McManus M.T., White W.R. and McGregor P.G. (1994) Accumulation of a chymotrypsin inhibitor in transgenic tobacco can affect the growth of insect pests. *Transgenic Research*, 3, 50-58.



- Mello M.O., Tanaka A.S. and Silva-Filho M.C. (2003) Molecular evolution of Bowman–Birk type proteinase inhibitors in flowering plants. *Molecular Phylogenetics and Evolution*, 27, 103–112.
- Melville J.C. and Ryan C.A. (1973) Chymotrypsin inhibitor I from potato: large scale preparation and the characterization of its subunit components. *Journal of Biological Chemistry*, 247, 3415-3453.
- Mesquita A.L.M., Alvers E.J. and Caldas R.C. (1984) Resistance of banana cultivars to *Cosmopolites sordidus* (GERMAR, 1824). *Fruits*, 39, 254-257.
- Michaud, D. and Vrain T.C. (1998) Expression of recombinant protease inhibitors in plants. In: *Recombinant proteins from plants: Production and isolation of clinically useful compounds* (Cunningham, C. and Porter, A.J.R. (eds.), Humana Press Inc., Totowa, NJ, pp. 49-64.
- Michaud D. (1997) Avoiding protease mediated resistance in herbivorous pests. *Trends in Biotechnology*, 15, 4-6.
- Michaud D., Nguyen-Quoc B. and Yelle S. (1993) Selective inactivation of Colorado potato beetle cathepsin H by oryzacystatin I and II. *FEBS Letters*, 331, 173-176.
- Michaud D., Bernier-Vandnais N., Overney S. and Yelle S. (1995) Constitutive expression of digestive cysteine protease forms during development of the Colorado potato beetle, *Leptinotarsa decemlineata* Say (Coleoptera: Chrysomelidae). *Insect Biochemistry and Molecular Biology*, 25, 1041-1048.
- Michaud D., Nguyen-Quoc B. and Yelle S. (1993) Selective inactivation of Colorado potato beetle cathepsin H by oryzacystatin I and II. *FEBS Letters*, 331, 173-176.



- Michaud D., Nguyen-Quoc B. and Yelle S. (1994) Production of oryzacystatins I and II in *Escherichia coli* using the glutathione *S*-transferase gene fusion system. *Biotechnology Progress*, 10, 155-159.
- Michaud D. (1998) Gel electrophoresis of proteolytic enzymes. *Analytica Chimica Acta*, 372, 173-185.
- Michaud D., Cantin L. and Vrain T.C. (1995) Carboxy terminal truncation of oryzacystatin II by oryzacystatin insensitive insect digestive proteinases. *Archives of Biochemistry and Biophysics*, 322, 469-474.
- Michaud D., Nguyen-Quoc B., Vrain T.C., Fong D. and Yelle S. (1996) Response of digestive cysteine proteinases from the Colorado potato beetle (*Leptinotarsa decemlineata*) and the black vine weevil (*Otiorynchus sulcatus*) to a recombinant form of human stefin A. *Archives of Insect Biochemistry and Physiology*, 31, 451-64.
- Mickel C.E. and Standish J. (1947) Susceptibility of processed soy flour and soy grits in storage to attack by *Tribolium castaneum*. University of Minnesota Agricultural Experimental Station Technical Bulletin, 178, 1-20.
- Mochizuki A. (1998) Characteristics of digestive proteases in the gut of some insect orders. *Applied Entomology and Zoology*, 33, 401-407.
- Moon, J., Salzman R.A., Ahn J. E., Koiwa H. and Zhu-Salzman K. (2004) Transcriptional regulation in cowpea bruchid guts during adaptation to a plant defence protease inhibitor. *Insect Molecular Biology*, 13, 283–291.
- Morton R.L., Schoroeder H.E., Bateman K.S., Chrispeels M.J., Armstrong E. and Higgins T.J.V. (2000) Bean α -amylase inhibitor-I in transgenic peas (*Pisium sativum*) provided complete protection from pea weevil (*Bruchus pisorum*)



under field conditions. Proceedings of the National Academy of Science of the United States of America, 97, 3820-3825.

- Mukhopadhyay D. (2000) The molecular evolutionary history of an winged bean alpha-chymotrypsin inhibitor and modeling of its mutations through structural analysis. *Journal of Molecular Evolution*, 50, 214-223.
- Murdock L.L., Brookhart G., Dunn P.E., Foard D.E., Kelley S., Kitch L., Shade R.E., Shuckle R.H. and Wolfson J.L. (1987). Cysteine digestive proteinases in Choleptera. *Comparative Biochemistry and Physiology*, 87, 783-787.
- Murdock L.L., Shade R.E., Pomeroy M.A. (1988) Effects of E-64, a cysteine protease inhibitor on cowpea weevil growth, development, and fecundity. *Environmental Entomology*, 17, 467-469.
- Musabyimana T., Seshu Reddy K.V. and Ngode L. (2000) Evaluation of banana cultivars for resistance to the banana weevil *Cosmopolites Sordidus* and nematode complex in western Kenya. *Acta Horticulturae*, 540, 233-238.
- Nagata K., Kudo N., Abe K., Arai S. and Tanokura M. (2000) Three-dimensional solution structure of Oryzacystatin-I, a cysteine protease inhibitor of the Rice, *Oryza sativa* L. japonica. *Biochemistry*, 39, 14753-14760.
- Newell C., Lowe J. and Merryweather A. (1995) Transformation of sweetpotato (*Ipomoea batatas* (L.) Lam.) with *Agrobacterium tumefaciens* and regeneration of plants expressing cowpea trypsin inhibitor and snowdrop lectin. *Plant Science*, 107, 215-227.
- Nikawa T., Towatari T., Ike Y. and Katunuma 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 proteinase inhibitory activities. *FEBS Letters*, 255, 309-314.



- Odani S., Koide T. and Ono T. (1983) The complete amino acid sequence of barley trypsin inhibitor. *Journal of Biological Chemistry*, 258, 7998-8003.
- Ogawa M., Nakamura S., Scaman C.H., Jing H., Kitts D.D., Dou J. and Nakai S. (2002) Enhancement of proteinase inhibitory activity of recombinant human cystatin C using random-centroid optimization. *Biochimica et Biophysica Acta* 1599, 115–124.
- Ojima A., Shiota H., Higashi K., Kamada H., Shimma Y., Wada M. and Satoh S. (1997) An extracellular insoluble inhibitor of cysteine proteinases in cell cultures and seeds of carrot. *Plant Molecular Biology*, 34, 99-109.
- Oliveira A.S., Xavier-Filho J. and Sales M.P. (2003) Cysteine Proteinases and cystatins. *Brazilian Archives of Biology and Technology*, 46, 91-104.
- Orr G.L., Strickland J.A. and Walsh T.A. (1994) Inhibition of *Diabrotica* larval growth by a multicystatin from potato tubers. *Journal of Insect Physiology*, 40, 893-900.
- Ortiz R., Vuylsteke D., Dumpe B. and Ferris R.S.B. (1995) Banana weevil resistance and corm hardness in *Musa* germplasm. *Euphytica*, 86, 95-102.
- Ortiz R. (2000) Understanding the *Musa* genome: an update. *Acta Horticulturae*, 540, 157-168.
- Ostmark H.E. (1974) Economic insect pests of bananas *Annual Reviews of Entomology*, 19, 161-176.
- Padmanaban B., Sundararaju P., Velayudhan K.C. and Sathiamoorthy S. (2001) Evaluation of *Musa* germplasm against banana weevil borers *Infomusa* 10, 26-28.
- Page R.D.M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, 12, 357-358.



- Painter R.H. (1951) Insect resistance in crop plants. The Macmillan Co., New York. 520.
- Pannetier C., Giband M., Couzi P., Letan V., Mazier M., Tourneur J. and Hau B. (1997) Introduction of new traits into cotton through genetic engineering: insect resistance as an example. *Euphytica*, 96, 163-166.
- Park H., Yamanaka N., Mikkonen A., Kusakabe I. and Kobayashi H. (2000) Purification and characterization of aspartic protease from sunflower seeds. *Bioscience Biotechnology and Biochemistry*, 64, 931-939.
- Pavis C. and Lemaire L. (1997) Resistance of *Musa* germplasm to the banana borer weevil, *Cosmopolites sordidus* Germar (Coleoptera: Curculionidae) *InfoMusa*, 5, 3-9.
- Pavis C. and Minost C. (1993) Banana resistance to the banana weevil borer *Cosmopolites sordidus* Germar (Coleoptera: Curculionidae); role of pseudostem attractivity and physical properties of the rhizome. In: Proceedings of international symposium on genetic improvement of banana for resistance to diseases and pests, CIRAD-FLHOR, Montpellier France, pp. 129-142.
- Pavlova A. and Bjork 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.
- Perler F., Efstratiadis A., Lomedico P., Gilbert N., Kolodner R. and Oodgson J. (1980) The evolution of genes: the chicken preproinsulin gene. *Cell*, 20, 555-556.



- Pernas M., Sanchez-Monge R., Gomez L. and Salcedo G. (1998) A chestnut seed cystatin differentially effective against cysteine proteinases from closely related pests. *Plant Molecular Biology*, 38, 1235-1242.
- Pernas M., Sanchez-Mong R. and Salcedo G. (2000) Biotic and abiotic stress can induce cystatin expression in chestnut. *FEBS Letters*, 467, 206-210.
- Pillay M., Ogundiwin E., Nwakanma D.C., Ude G. and Tenkouano A. (2001) Analysis of genetic diversity and relationships in East African banana germplasm. *Theoretical and Applied Genetics*, 102, 965-970.
- Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S. and Rafalski A. (1996) The comparison of RFLP, RAPD, AFLP, and SSR microsatellite markers for germplasm analysis. *Molecular Breeding*, 2, 225-238.
- Price N.S. (1994) Alternate cropping in the management of *Rodopholus similes* and *Cosmopolites sordidus*, two important pests of banana and plantain. *International Journal for pest Management*, 40, 237-244.
- Rancour J.M. and Ryan C. A. (1968) Isolation of a carboxypeptidase B inhibitor from potatoes. *Archives of Biochemistry and Biophysics*, 125, 380-382.
- Rassam M. and Laing W.A. (2004) Purification and characterization of phytocystatins from kiwifruit cortex and seeds. *Phytochemistry*, 65, 19-30.
- Ravichandaran S., Sen U., Chakrabarti C. and Dattagupta J.K. (1999) Cryocrystallography of a Kunitz type serine protease inhibitor: 90K structure of winged bean chymotrypsin (WCI) at 2.13Å resolution. *Acta Crystallography D Biology Crystallography*, 55, 1814-1821.
- Rawlings N.D. and Barrett A.J. (1995) Evolutionary families of metalloproteinases. In: *Methods in Enzymology* (Barrett A.J. (ed.). New York, Academic Press, 248, 183-228.



- Reis E.M. and Margis R. (2001) Sugarcane phytocystatins: Identification, classification and expression pattern analysis. *Genetics and Molecular Biology*, 24, 291-296.
- Richardson M.J. (1991) Seed storage proteins: The enzyme inhibitors. In: *Methods in Plant Biochemistry* (Richardson, M.J ed.), New York, Academic Press, 259-305.
- Rogel B., Tatjana P., Ritonja A., Strukelj J. and Brzin J. (1998) Chelidocystatin, A novel phytocystatin from *Chelidonium majus*. *Phytochemistry*, 49, 1645-1649.
- Rogers B.L., Pollock J., Klapper D.G. and Griffith I.J. (1993) Sequence of the protease-inhibitor cystatin homologue from the pollen of *Ambrosia artemisiifolia* (short ragweed). *Gene*, 133, 219-221.
- Rukazambuga N.D.T.M., Gold C.S. and Gowen S. (1998) Yield loss in East African highland banana (*Musa* spp., AAA-EA group) caused by the banana weevil, *Cosmopolites sordidus* Germar. *Crop Protection*, 17, 581-589.
- Ryan C.A. (1990) Protease inhibitors in plants: genes for improving resistance against insects and pathogens. *Annual Review of Phytopathology*, 28, 425-449.
- Ryan S.N., Laing W. A. and McManus M. (1998) A cysteine protease inhibitor from apple fruit. *Phytochemistry*, 49, 957-963.
- Ryan C.A. (1989) Insect-induced chemical signals regulating natural plant protection responses. In: *Variable plants and herbivores in natural and managed systems*. Denno, R.F. and McClure M.S. (eds.). New York, Academic Press, p. 43-60.
- Ryan C.A. (2000) The systemin signalling pathway: differential activation of plant defensive genes. *Biochimica et Biophysica Acta*, 1477, 112-121.



- Salvesen G. and Nagase H. (1989) Inhibition of proteolytic enzymes. In: Proteolytic enzymes. A practical approach, IRL Press, New York, pp 83-104.
- Sambrook J., Fritsh E.F. and Maniatis T. (1989) Molecular cloning: A laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Springs Houbour, NY.
- Sánchez R. and Sali A. (2000) Comparative protein structure modelling: Introduction and practical examples with MODELLER. In: Protein Structure Prediction: Methods and Protocols (Webster, D.M. ed.). Humana Press. pp. 97-129.
- Sane V.A., Nath P., Aminuddin L. and Sane P.V. (1997) Development of insect-resistant transgenic plants using plant genes: expression of cowpea trypsin inhibitor in transgenic tobacco plants. *Current Science*, 72, 741-747.
- Santino A., Gallo A., Leone A., Poltronieri P. and Zacheo G. (1998) Purification and characterization of a novel papain inhibitor from common bean (*Phaseolus vulgaris*) seed. *Physiologia Plantarum*, 104, 293-298.
- Sarath G. (1989) Protease assay methods. In: Proteolytic enzymes: a practical approach (Beynon, R. J. and Bond J. S. (rend eds.). Oxford: IRL Press.
- Sardana R.K., Ganz P.R., Dudani A.K., Tackaberry E.S., Cheng X. and Altosaar I. (1998) Synthesis of recombinant human cytokine GMCSF in the seeds of transgenic tobacco plants. In: Recombinant proteins from plants. Production and isolation of clinically useful compounds (Cunningham, C. and Porter A.J.R., eds). Totowa NJ, Humana Press, p. 77-87.
- Sawyer S.L., Wu L.I., Emerman M. and Malik H.S. (2005) Positive selection of primate TRIM5a identifies a critical species-specific retroviral restriction domain. *Proceedings of the National academy of Sciences of the United States of America*, 102, 2832–2837.



- Scribner K.T. and Pearce J.M. (2000) Microsatellites: evolutionary and methodological background and empirical applications at individual, population, and phylogenetic levels. In: Molecular Methods in Ecology (Baker, ed). Blackwell Science Limited, London, pp. 235-271.
- Sengooba T. (1986) Survey of banana pest problem complex in Rakai and Masaka Districts in Uganda. August 1986: Preliminary trip report, Namulonge Research Station, Namulonge, Uganda.
- Seshu-Reddy K.V. and Lubega M.C. (1993) Evaluation of banana cultivars for resistance to tolerance of the weevil *Cosmopolites sordidus* Germar, In: Breeding Banana and Plantain for Resistance to Diseases and Pests (Ganry, J. ed). CIRAD/INIBAP, Montpellier, France, pp. 143-148.
- Shulke R.H. and Murdock L.L. (1983) Lipoxigenase trypsin inhibitor and lectin from soybeans: effects on larval growth of *Manduca sexta* (Lepidoptera: Sphingidae). Environmental Entomology, 12, 787-791.
- Soares-Costa A., Beltramini L.M., Thiemann O.H. and Henrique-Silva F. (2002) A sugarcane cystatin: recombinant expression, purification, and antifungal activity. Biochemical and Biophysical. Research Communications, 296, 1194-1199.
- Solomon M., Belenghi B., Delledonne M., Menachem E. and Levine M. (1999) The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. The Plant Cell, 11, 431-443.
- Song I., Taylor M., Baker K. and Bateman 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 2.4Å x-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO Journal*, 9, 1939–1947.
- Swennen R. and Vuylsteke D. (2001) Banana *Musa L.* In: *Crop Production in Tropical Africa* (Raemaekers, R.H. ed.). DGIC Brussels, Belgium, pp.530-552.
- Thomas J.C., Adams D.G., Keppenne V.D., Wasmann C.C., Brown J.K., Kanosh M.R. and Bohnert H.J. (1995) Proteinase inhibitors of *Manduca sexta* expressed in transgenic cotton. *Plant Cell Reports*, 14, 758-762.
- Tiffin P. and Brandon G.S. (2001) Molecular evolution of the wound- induced serine protease inhibitor wip1 In *Zea* and related genera. *Molecular Biology and Evolution*, 18, 2092-20101.
- Tinjuangjun P. (2002) Snowdrop lectin gene in transgenic plants: its potential for Asian agriculture. www.agbiotechnet.com (ABN091).
- Titarenko E., Rojo E., Leon J. and Sanchez-Serrano J.J. (1997) Jasmonic acid-dependent and -independent signalling pathways control wound-induced gene activation in *Arabidopsis thaliana*. *Plant Physiology*, 115, 817-826.
- To K.Y., Suen D.F. and Chen S.C.G. (1999) A sweet potato leaf cDNA (Accession No. AF117334) encoding cysteine proteinase inhibitor. (PGR99-056). *Plant Physiology*, 119, 1568-1568.



- Tripathi L. (2003) Genetic engineering for improvement of *Musa* production in Africa. *African Journal of Biotechnology*, 2, 503-508.
- Turk V. and Bode W. (1991) The cystatins: protein inhibitors of cysteine proteinases. *FEBS Letters*, 285, 213-219.
- Urwin P.E., Atkinson H.J., Waller D.A. and Mcpherson M.J. (1995) Engineered oryzacystatin-I expressed in hairy roots confers resistance to *Globodera pallida*. *Plant Journal*, 8, 121-131.
- Urwin P.E., Lilley C.J., Mcpherson M.J. and Atkinson H.J. (1997) Resistance to both cyst and rootknot nematodes conferred by transgenic *Arabidopsis* expressing a modified plant cystatin. *Plant Journal*, 12, 455-461.
- Ussuf K.K., Laxmi N.H. and Mitra R. (2001) Proteinase inhibitors: Plant-derived genes of insecticidal protein for developing insect-resistant transgenic plants. *Current Science*, 80, 847- 552.
- Vain P., Worland B., Clarke M.C., Richard G., Beavis M., Liu H., Kohli A., Leech M., Snape J. and Christou P. (1998) Expression of an engineered cysteine proteinase inhibitor (Oryzacystatin-I delta 86) for nematode resistance in transgenic rice plants. *Theoretical and Applied Genetics*, 96, 266-271.
- Van der Vyver C., Schneidereit J., Driscoll S., Turner J., Kunert K.J. and Foyer C.H. (2003) Oryzacystatin-I expression in transformed tobacco produces a conditional growth phenotype and enhances cold tolerance. *Plant Biotechnology Journal*, 1, 101-112.
- Viljoen A., Kunert K., Kiggundu A. and Escalant J.V. (2004) Biotechnology for sustainable banana and plantain production in Africa: the South African Contribution. *South African Journal of Botany*, 70, 67-74.



- Visser A. A. (2000) Characterization of banana and plantain using random amplified polymorphic DNA markers. *Acta Horticulturae*, 540, 113-123.
- Vuylsteke D., Ortiz R., Ferris S. and Crouch J. (1997) Plantain improvement. *Plant Breeding Reviews*, 14, 267-320.
- Waldron C., Wegrich L.M., Merlo P.A. and Walsh T.A. (1993) Characterization of a genomic sequence coding for potato multicystatin, an eight-domain cysteine proteinase inhibitor *Plant Molecular Biology*, 23, 801-812.
- Walsh T.A. and Strickland J.A. (1993) Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. *Plant Physiology*, 103, 1227-1234.
- Wang W. and Malcolm B.A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuickChange Site-Directed Mutagenesis. *Biotechniques*, 26, 680-682.
- Wasternack C. and Parthier B. (1997) Jasmonate-signalled plant gene expression. *Trends in Plant Science*, 2, 302-307.
- Wieman K.F. and Nielsen S.S. (1988) Isolation and partial characterization of a major gut proteinase from larval *Acanthoscelides obtectus* Say (Coleoptera: bruchidae). *Comparative Biochemistry and Physiology-B*, 89, 419-426.
- Wilhite S.E., Elden T.C., Brzin J. and Smigocki A.C. (2000) Inhibition of cysteine and aspartyl proteinases in the alfalfa weevil midgut with biochemical and plant derived proteinase inhibitors. *Insect Biochemistry and Molecular Biology*, 30, 1181-1188.
- Williamson V.M. and Hussey R.S. (1996) Nematode pathogenesis and resistance in plants. *Plant Cell*, 8, 1735-1745.



- Winterer J. (2002) The mixed success of proteinase inhibitors to combat insect pests in transgenic crops. *www.AgBiotech.Net*, 4, 1-7.
- Wolfson J.L. and Murdock L.L. (1990) Diversity in digestive proteinase activity among insects. *Journal of Chemical Ecology*, 16, 1089-1102.
- Wolfson J.L. and Murdock L.L. (1987) Suppression of larval Colorado beetle growth and development by digestive proteinase inhibitors. *Entomologia Experimentalis et Applicata*, 44, 235-240.
- Xu D.P., Xue, Q.Z., Mcelroy D., Mawal Y., Hilder V.A. and Wu R. (1996) Constitutive expression of a cowpea trypsin inhibitor gene, CpTi, in transgenic rice plants confers resistance to two major rice insect pests. *Molecular Breeding*, 2, 167-173.
- Yang Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences*, 13, 555–556.
- Yang Z. and Bielawski J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15, 496–503.
- Yang Z., Nielsen R., Goldman N. and Pedersen A.M.K. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, 155, 431–449.
- Yeh K.W., Lin M.L., Tuan S.J., Chen Y.M., Lin C.Y. and Kao S.S. (1997) Sweet potato (*Ipomea batatas*) trypsin inhibitors expressed in transgenic tobacco plants confer resistance against *Spodoptera litura*. *Plant Cell Reporter*, 16, 696-699.
- Young, L. and Dong Q. (2003) TAMS technology for simple and efficient in vitro site-directed mutagenesis and mutant screening. *Nucleic Acids Research*, 31, 11-13.



- 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 proteinase inhibitors have greater insect digestive proteinase inhibitory activities than a constitutive homologue. *Plant Physiology*, 111, 1299-1306.
- Zhu-Salzman, K., Salzman R.A., Koiwa H., Murdock L.L., Bressan R.A. and Hasegawa P.M. (1998) Ethylene negatively regulates local expression of plant defense lectin genes. *Physiologia Plantarum*, 104, 365-372.