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, 563-56.

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).

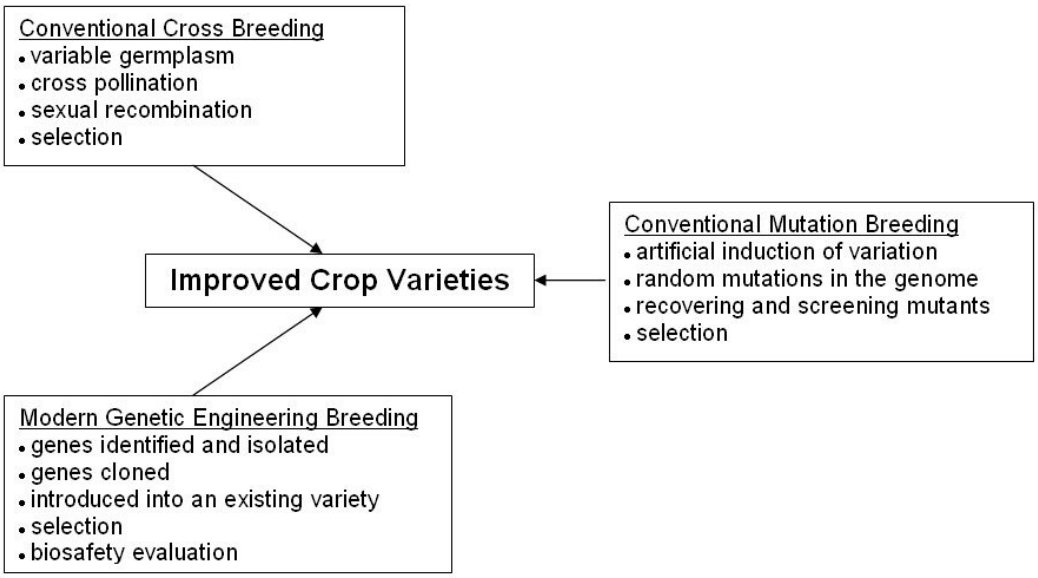

Figure 1.1 The three broad methods of crop improvement compared. Conventional cross breeding is limited by the availability of the required traits in the gene pool as well as sexual compatibility of the crop. Conventional mutation breeding relies on the use of artificial induction of variation by the 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).

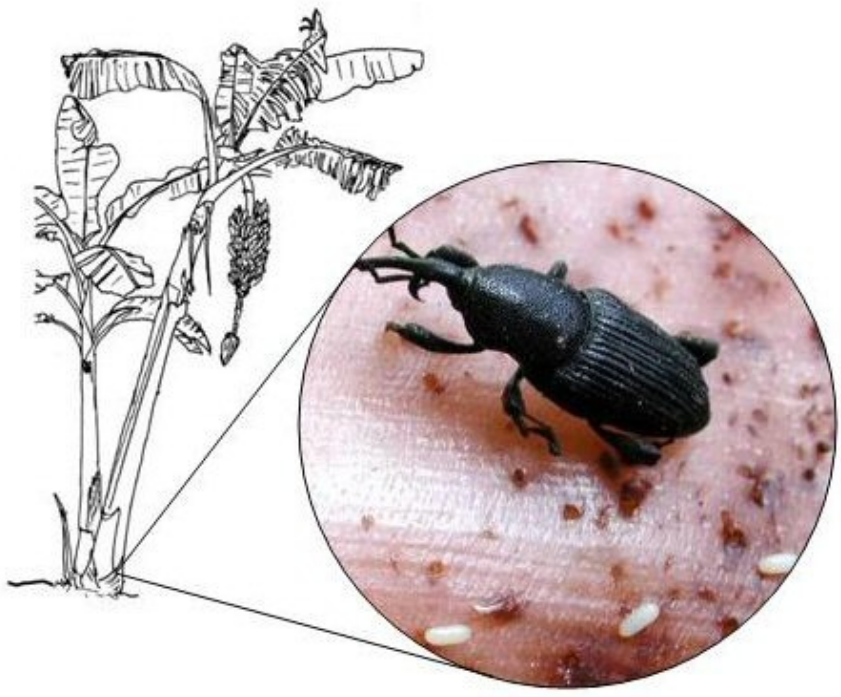

Figure 1.2 The adult banana weevil (*C. sordidus*) lays eggs on the banana plant just above the soil surface. When the eggs hatch the emerging larvae burrow through the underground stem leading to yield loss, structural weakness and toppling of the plant.

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 Kedong-kekang (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 *et al.,* 2003a
Sannachenkadali	AA	Padmanaban *et al.,* 2001
Sakkali	ABB	
Senkadali	AAA	
Elacazha	BB	
Njalipoovan	AB	
Pisang Awak	ABB	Kiggundu *et al.,* 2003a
FHIA03	AABB	
TMBx612-74	IITA hybrid	
TMB2x6142-1	IITA hybrid	
TMB2x8075-7	IITA hybrid	
TMB2x7197-2	IITA hybrid	
Long Tavoy	ABB	Ortiz *et al.,* 1995
Njeru	AA	Musabyimana *et al.,* 2000
Muraru	AA	
Calcutta-4	AA	Fogain and Price, 1994; Ortiz *et al.,* 1995; Kiggundu *et al.,* 2003a
Bluggoe	ABB	Fogain and Price, 1994; Kiggundu *et al.,* 2003a
M. balbisiana	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 pathenocarpy (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 Homotera, 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. B*t* plants are currently the most widely used GM technololgy 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 midgets (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.

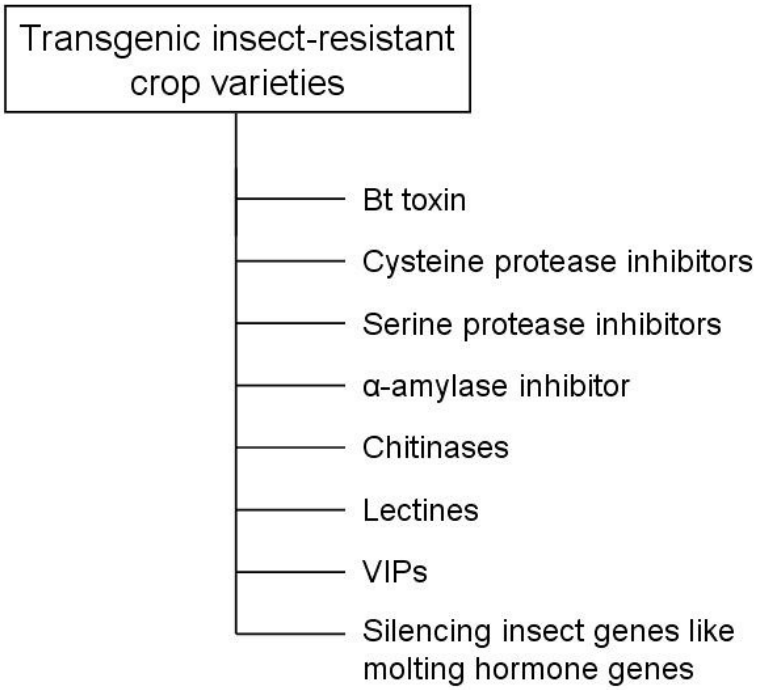

Figure 1.3 Genetic engineering strategies currently in commercially produced crops (only *Bt* toxin) and others being developed for increasing resistance to crop insect and nematode pests.

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-chloromercuribenzene sulfonic acid (PCMBS), a potent sulphydryl 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 insect 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 were 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
Aspertic protease	Aspertin	> 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. There functions seem to range from regulation of endogenous protease activity to storage proteins, as they tend to accumulate in large amount 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 deceases 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 increases 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	*Heliothis virescens* (Fabricius) (Lepidoptera: Noctuidae)	Hilder *et al*., 1987
	Rice	*Chilo suppressalis* (Walker) (Lepidoptera: Pyralidae)	Xu et al., 1996
	Potato	*Lacanobia oleracea* (Linnaeus) (Lepidoptera: Noctuidae)	Gatehouse *et al*., 1997
	Strawberry	*Otiorhynchus sulcatus* (Fabricius) (Coleoptera: Curculionidae)	Graham *et al*., 1997
	Tobacco	*Spodoptera litura* (Fabricius) (Lepidoptera: Noctuidae)	Sane *et al*., 1997
	Cotton	*Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae)	Li *et al*., 1998
	Wheat	*Sitotroga cerealla* (Olivier) (Lepidoptera: Gelechiidae)	Alpteter *et al*., 1999
	Pigeonpea	*H. armigera*	Lawrence *et al*., 2001
CpTi + Snowdrop lectin	Sweet potato	*Cylas formicarius* (Fabricius) (Coleoptera: Curculionidae)	Newell *et al*., 1995
Potato inhibitor II	Tobacco	*Manduca sexta* (Linnaeus) (Lepidoptera: Sphingidae)	Johnson *et al*., 1989
	Bean/corn/eggplant	*Chrysodeixis eriosoma* (Doubleday) (Lepidoptera: Noctuidae)	McManus *et al*.,1994
	Rice	*Sesamia inferens* (Walker) (Lepidoptera: Amphipyrinae)	Duan *et al*.,1996
Tomato inhibitor I and II	Tobacco	*M. sexta*	Johnson *et al*.,1989
Sweet potato trypsin inhibitor (TI)	Tobacco	*M. sexta*	Yeh *et al*.,1997
Soybean Kunitz TI	Rice	*Nilaparvata lugens* (Stal.) (Hemiptera: Delphacidae)	Lee *et al*., 1999
Barley TI	Tobacco	*Agrotis ipsilon* (Hufnagel) (Lepidoptera: Noctuidae)	Carbonero *et al*.,1993
Nicotiana alta protease inhibitor (PI)	Tobacco	*Helicoverpa punctigera* (Wallengren) (Lepidoptera: Noctuidae)	Heath *et al*., 1997
	Pea	*Plutella xylostella* (Linnaeus) (Lepidoptera: Plutellidae)	Charity *et al*., 1999
Serpin type serine PI	Tobacco	*Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae)	Thomas *et al*., 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 ologomers. 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).

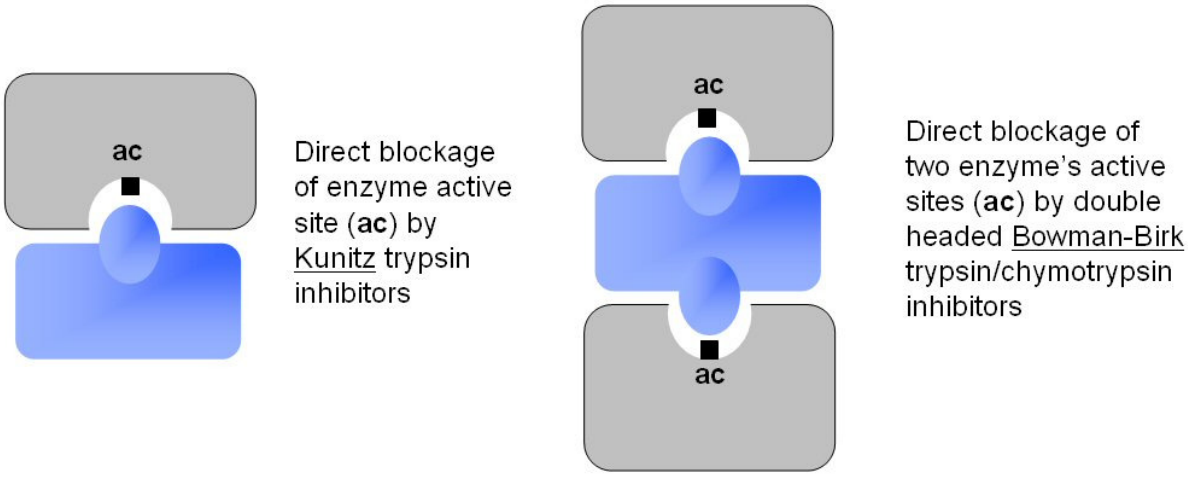

Figure 1.4 Substrate-like mechanism of inhibition by two serine protease inhibitor types, Kunitz and Bowman-Birk (Modified from Bode and Huber, 2000).

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).

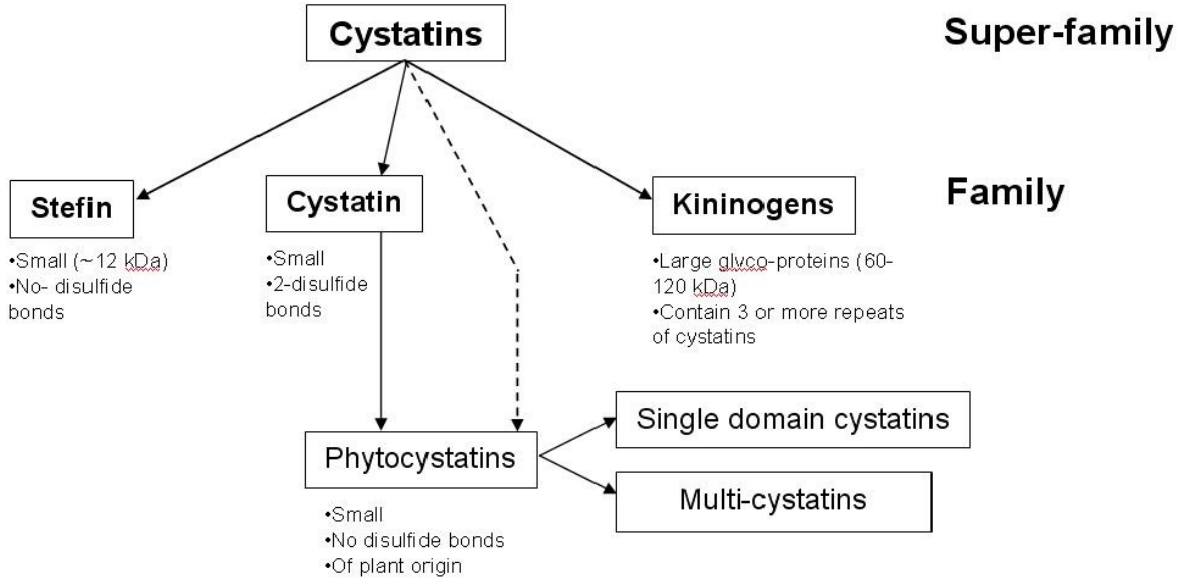

Figure 1.5 General classification of the cystatin super-family phytocystatins (plant cysteine protease inhibitors) are grouped as members of the cystatin family. Broken line indicates re-classification of phytocystatins as a separate family containing single and multidomain cystatins.

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.

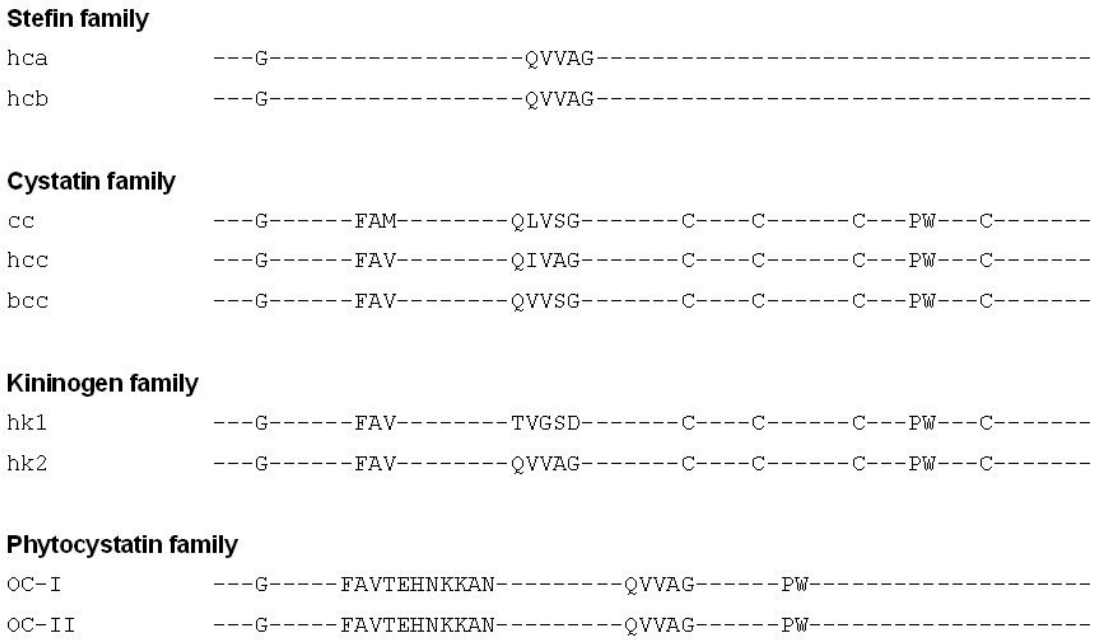

Figure 1.6 Alignment of selected members of the four cystatin families illustrating the sequence conservation regions within the family members. The sequences are human cystatin-A (hca), human cystatin-B (hcb), chicken cystatin (cc), human cystatin-C (hcc), beef colostrums cystatin (bcc), human kininogen 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.

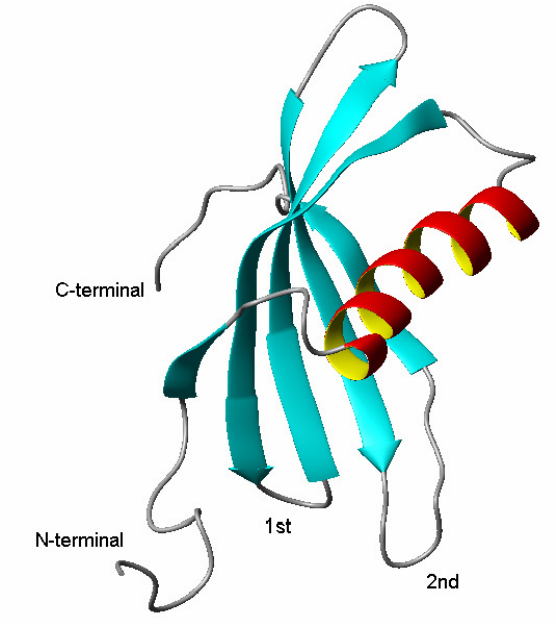

Figure 1.7 The three dimensional structure of OC-I showing the characteristic 5 anti-parallel B strands (blue), the single 5 turn a-helix (red), the N-terminal, the 1st and 2nd hairpin-like loops. Figure was drawn using MolMol version 2k.1.

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).

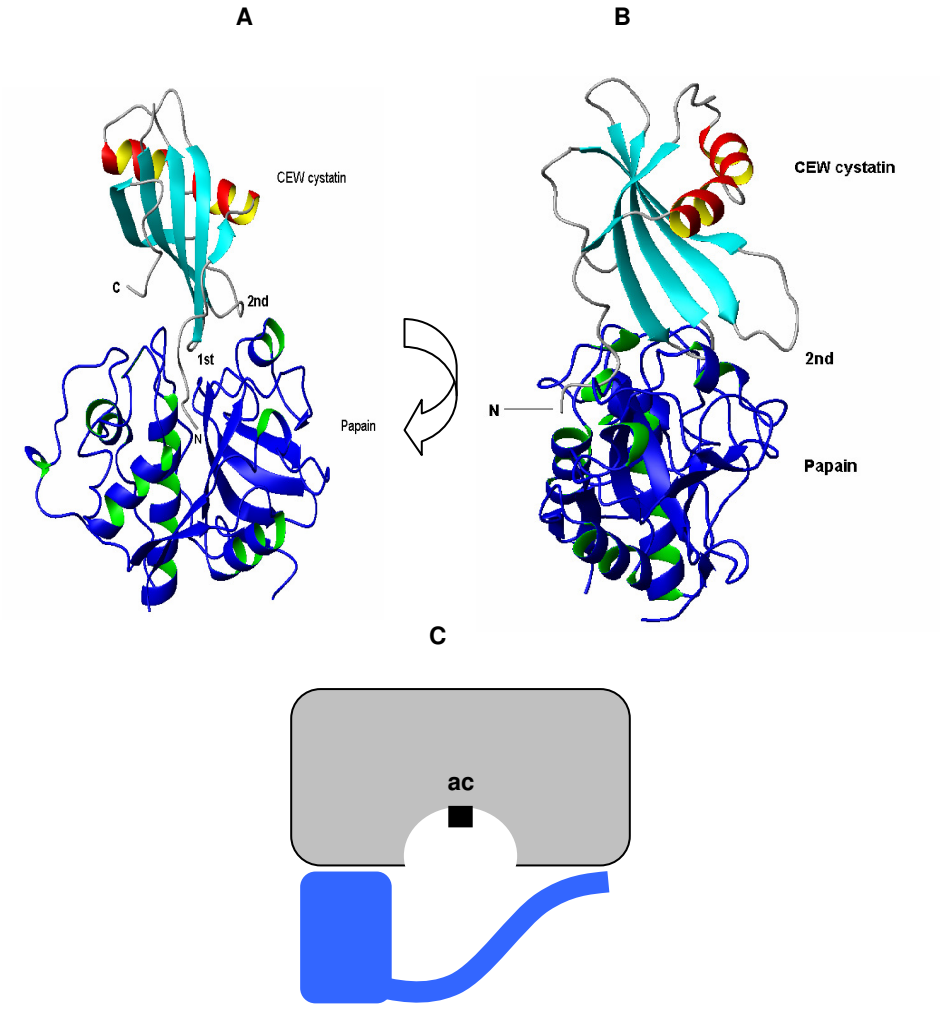

Figure 1.8 Three-dimensional plot showing the complex between papain (blue and green) and chicken egg white cystatin (CEW) colored light blue, red and yellow (PDB accession No. 1STF). (A) Complex is presented in front view to show the V-shaped active site of papain and how the N-terminal region of CEW fits into it. (B) Complex is rotated 90^{o} 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 (*Hypera postica*) (Gyllenhal)	Coleoptera: Curculionidae	Alfalfa	Oryzacystatin -I	*In-vitro assays*	Wilhite *et al*., 2000
Bean beetle *(Callosobruchus chinensis)* (Linnaeus)	Coleoptera: Bruchidae	Common bean	Oryzacystatin -I & II	Artificial diet	Kuroda *et al*., 2001
Bean bug *(Riptortus clavatus)* (Thunberg)	Heteroptera: Alydidae	Common bean	Oryzacystatin -I & II	Artificial diet	Kuroda *et al*., 2001
Black vine weevil *(Otiorhynchus sulcatus*) (Fabricius)	Coleoptera: Curculionidae	Forestry trees	Oryzacystatin -I	*In-vitro* assays	Michaud *et al*., 1996
Colorado potato beetle (*Leptinotarsa decemlineata)* (Say)	Coleoptera: Chrysomelidae	Potato	Oryzacystatin -I	Transgenic potato	Lecardonnel *et al*., (1999)
Maize grain weevil *(Sitophilus zeamais)* (Motschulsky)	Coleoptera: Curculionidae	Maize and rice	Corn cystatin (CC)	Transgenic rice	Irie *et al*., 1996
Poplar leaf-beetle *(Chrysomela tremulae)* (Fabricius)	Coleoptera: Chrysomelidae	White poplar	Oryzacystatin -I	Transgenic poplar	Leple *et al*., 1995
(Chrysomela populi) (Linnaeus)			Arabidopsis cystatin (*Atcys*)		Delledonne *et al*., 2001
Southern corn rootworm (*Diabrotica undecimpunctata howardi)* (Barber)	Coleoptera: Chrysomelidae	Maize	Oryzacystatin -I	*In-vitro* assays	Edmonds *et al*., 1996
			Potato cystatin (PCPI-10)	Artificial diet	Fabrick *et al*., 2002
Western corn rootworm (*Diabrotica virgifera virgifera*) (LeConte)	Coleoptera: Chrysomelidae	Maize	Soyacystatin N (ScN)	Artificial diet	Zhao *et al*., 1996 Koiwa *et al*., 2000
Western flower thrip *(Frankliniella occidentalis)* (Pergande)	Thysanoptera: Thripidae	Capsicum, Cucumber Carnation, Chrysanthemum	Potato cystatin	*In-vitro* assays	Annadana *et al*., 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 to 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 hypothesiszed 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 suitablity 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 access 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.