PROTEASE INHIBTOR MEDIATED RESISTANCE TO INSECTS

NIKOLAY STOYANOV OUTCHKOUROV

Promotor:	Prof. dr. W.J. Stiekema, hoogleraar Bioinformatica, met speciale aandacht voor genomics, Wageningen Universiteit
Co-promotor:	Dr. Maarten Jongsma, clusterleider Biocides, BU Bioscience, Plant Research International, Wageningen UR
Promotiecommissie:	Prof. dr. M. Dicke, Wageningen Universiteit
	Prof. dr. R.J. Bino, Wageningen Universiteit
	Prof. dr. V. Turk, J. Stefan Institute, Ljubljana, Slovenia
	Prof. dr. R.G.T. Visser, Wageningen Universiteit

PROTEASE INHIBTOR MEDIATED RESISTANCE TO INSECTS

NIKOLAY STOYANOV OUTCHKOUROV

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. ir. L. Speelman,
in het openbaar te verdedigen
op woensdag 4 november 2003
des namiddags te 13.30 uur in de Aula.

Nikolay S. Outchkourov – Protease inhibitor mediated resistance to insects - 2003

Ph.D. Thesis Wageningen University, Wageningen, The Netherlands - 135 p.

ISBN: 90-5808-913-4

Table of contents

ABSTRACT OF THE THESIS IX

CHAPTER 1 1

GENERAL INTRODUCTION 1

ABSTRACT 1
CROP PROTECTION AGAINST INSECT PESTS 1
WESTERN FLOWER THRIPS 2
PROTEASE INHIBITORS AS PLANT PROTECTIVE MOLECULES 2
CLASSIFICATION AND MECHANISM OF ACTION OF PROTEASES AND INHIBITORS 4
THE ROLE OF MULTIDOMAIN PROTEASE INHIBITORS 6
SCOPE OF THE THESIS 6

CHAPTER 2 9

OPTIMIZATION OF THE EXPRESSION OF EQUISTATIN IN *PICHIA PASTORIS* . 9

ABSTRACT 10
INTRODUCTION 10
MATERIALS AND METHODS 11
Materials *11*
Oligonucleotides *11*
Cloning and expression of equistatin variants in P. pastoris *11*
Quantification and characterization of equistatin *12*
RNA analysis *12*
Large-scale equistatin purification *13*
RESULTS 13
Analysis of the effects of gene modifications on protein expression levels *13*
Initial protein secretion rate and RNA accumulation *15*
Bioreactor production and equistatin purification *16*
DISCUSSION 17

CHAPTER 3 21

THE PROMOTER TERMINATOR OF CHRYSANTHEMUM *RBCS1* DIRECTS VERY HIGH EXPRESSION LEVELS IN PLANTS 21

ABSTRACT 22
INTRODUCTION 22
MATERIALS AND METHODS 23
Cloning of the rbcS1 gene *23*
Sequence analysis of the chrysanthemum rbcS1 promoter *24*
Construction of the rbcS1 plant expression cassette *24*
Preparation of the plant transformation constructs *24*
Plant transformation and analysis *25*
RESULTS AND DISCUSSION 26
Isolation of the chrysanthemum rbcS1 gene and comparative sequence analysis *26*
Comparison of the rbcS1 promoter with the 35S promoter in tobacco and chrysanthemum *29*

Expression of equistatin and cystatin in potato and tomato plants ... *30*
Comparison with similar promoters ... *31*

CHAPTER 4 ... **35**

SPECIFIC CYSTEINE PROTEASE INHIBITORS ACT AS DETERRENTS OF WESTERN FLOWER THRIPS *FRANKLINIELLA OCCIDENTALIS* (PERGANDE) IN TRANSGENIC POTATO ... **35**

ABSTRACT ... 36
INTRODUCTION ... 36
EXPERIMENTAL PROCEDURES ... 37
Oligonucleotides ... *37*
Plant material and insects ... *37*
Inhibition of thrips proteolytic activity ... *38*
Preparation of constructs for plant transformation ... *38*
Transformation and regeneration ... *39*
Quantification and characterization of the expression in transgenic potato plants ... *39*
Thrips choice assays ... *39*
RESULTS ... 40
Inhibition of total thrips proteolytic activities ... *40*
Expression of the cysteine protease inhibitors in transgenic potato plants and WFT choice assays on detached leaves ... *41*
DISCUSSION ... 44

CHAPTER 5 ... **49**

EXPRESSION OF SEA ANEMONE EQUISTATIN IN POTATO: EFFECTS OF PLANT PROTEASES ON HETEROLOGOUS PROTEIN PRODUCTION ... **49**

ABSTRACT ... 50
INTRODUCTION ... 50
MATERIALS AND METODS ... 52
Plant material and insects ... *52*
Preparation of constructs for plant transformation ... *52*
Potato transformation ... *52*
Northern blot ... *52*
Dot blot imunological detection and quantification of equistatin in transgenic potato plants ... *53*
Western blot analysis ... *53*
Assay of inhibitory activity ... *53*
Purification of the equistatin degradation products from transgenic potato plants ... *54*
pH optimum of the proteolytic activities in potato leaf extracts ... *55*
Inhibitors of equistatin degradation ... *55*
Insect feeding trials ... *55*
RESULTS ... 56
Equistatin expression analysis ... *56*
Isolation and characterization of in-planta equistatin degradation products ... *60*
Proteolytic activities in potato plant extract ... *62*
Proteinaceous inhibitors of equistatin degradation ... *63*
Insect feeding trials ... *64*
DISCUSSION ... 65

Equistatin degradation in potato cv. Desirée ... 65
Identification of the plant proteases involved in equistatin proteolysis ... 65
Plant endosomal proteases as an impediment for protein production in plants ... 66
Strategies to reduce the degradation of heterologous proteins in plants ... 67
CHAPTER 6 ... **69**
ENGINEERING PROTEIN STABILITY OF EQUISTATIN IN POTATO PLANTS ... **69**
ABSTRACT ... 70
INTRODUCTION ... 70
MATERIALS AND METHODS ... 71
Site directed mutagenesis ... 71
Expression purification from Pichia pastoris and analysis of the inhibitory properties ... 71
Expression in plants ... 71
RESULTS AND DISCUSSION ... 72
Strategy for stabilization ... 72
Expression in Pichia pastoris and determination of the inhibitory properties ... 74
Expression in plants ... 76
Conclusion and recommendations ... 79
CHAPTER 7 ... **81**
ENGINEERED MULTI-DOMAIN CYSTEINE PROTEASE INHIBITORS YIELD RESISTANCE AGAINST WESTERN FLOWER THRIPS (*FRANKLINIELLA OCCIDENTALIS)* IN GREENHOUSE TRIALS ... **81**
ABSTRACT ... 82
INTRODUCTION ... 82
MATERIALS AND METHODS ... 83
Oligonucleotides ... 83
Preparation of multi-domain inhibitors EIMKACP and KACP ... 83
Transformation of potato ... 84
Expression in P. pastoris and subsequent protein purification and characterization ... 84
Artificial diet bioassay of WFT ... 85
Dot blot analysis ... 85
Western blot analysis ... 86
Greenhouse bioassay of WFT ... 86
RESULTS ... 86
Construction and characterization of the new multi-domain EIMKACP inhibitor ... 86
Effects of multidomain and single domain inhibitors on thrips fecundity ... 87
Expression of multi-domain protease inhibitors in plants. ... 89
Non-choice greenhouse trials ... 91
DISCUSSION ... 92
CHAPTER 8 ... **97**
CHARACTERISATION OF TWO *DROSOPHILA MELANOGASTER* GENES ENCODING MULTIPLE PUTATIVE PROTEASE INHIBITOR DOMAINS ... **97**
ABSTRACT ... 98
INTRODUCTION ... 98
MATERIALS AND METHODS ... 99

Oligonucleotides ... *99*
cDNA cloning of the two D. melanogaster genes ... *100*
Analysis of the CG2264 and CG5639 gene expression by quantitative real time polymerase chain reaction (RT-PCR) ... *100*
Preparation of Pichia pastoris expression constructs ... *101*
Expression the recombinant CG2264, and CG5639 proteins in Pichia pastoris ... *101*
Assays of the protease inhibitory activities ... *101*
RESULTS ... 102
cDNA cloning of CG2264 and CG5639 ... *102*
Analysis of the CG2264 and CG5639 gene expression ... *105*
Expression in Pichia pastoris and characterization ... *106*
Analysis of the inhibitory activities ... *107*
DISCUSSION ... 109
CHAPTER 9 ... 113
GENERAL DISCUSSION ... 113
ABSTRACT ... 113
Protease inhibitors as part of integrated pest management programs ... *113*
Important technological considerations for engineering applications of protease inhibitors in plants ... *114*
Heterologous protein production in Pichia pastoris ... *115*
An improvement of the expression of heterologous proteins in plants ... *115*
The role of endogenous proteases as an impediment for recombinant protein production in plants ... *116*
Strategies for stabilization of heterologous proteins in transgenic plants ... *116*
Issues with regard to genetically modified crop plants ... *118*
Future outlook ... *119*
ACKNOWLEDGEMENTS ... 120
CURRICULUM VITAE ... 122
LIST OF PUBLICATIONS ... 122
REFERENCES ... 123

ABSTRACT OF THE THESIS

Protease inhibitors are among the defensive molecules that plants produce in order to defend themselves against herbivores. A major aim of this thesis was to develop novel insect resistance traits using heterologous, non-plant protease inhibitors. To achieve this, it was necessary to express cysteine protease inhibitors in yeast in order to obtain sufficient protein for insect bioassays. To fully utilize the potential of the yeast *P. pastoris* for high protein production various gene variants encoding the sea anemone protease inhibitor equistatin were evaluated for the rate of protein production. A sequence element in the coding sequence of equistatin was identified which blocked the rate of expression. Removal of the element yielded up to 20-fold higher levels of protein production (1.7 g/L). Furthermore, high accumulation of selected inhibitors in transgenic plants was required. To this end, an expression cassette for expression in plants based on gene regulatory elements of a novel ribulose bisphosphate carboxylase gene (*rbcS1)* from chrysanthemum (*Chrysanthemum morifolium* Ramat.) was developed. In transgenic plants the new cassette yielded superior expression levels in the range of the endogenous rubisco small subunit protein (3-10% of TSP). High expression and insect resistance, however, were compromised by *in-planta* proteolytic degradation of a number of inhibitors in transgenic plants. The plant proteases involved in the degradation of recombinant equistatin were characterized, and two different strategies to prevent the proteolytic degradation of inhibitors in plants were designed. The first strategy, based on the rational choice of amino acid substitutions, only partly prevented the fast proteolytic degradation of equistatin in plants. In contrast, the second strategy, which employed inhibitors of the degrading proteases, resulted in stable, high level accumulation of novel recombinant multi-domain protease inhibitors. They were designed as synthetic custom made fusions of selected inhibitors that simultaneously bound different proteases. Thus, the new multidomain proteins were designed to prevent their own degradation *in-planta* and to target a wide range of cysteine proteases of insect pests such as western flower thrips (WFT) *Frankliniella occidentalis* (Pergande). As an extension of the multidomain concept natural multidomain proteins from *Drosophila melanogaster* were also cloned and expressed, but due to their limited inhibitory spectrum insect resistance applications failed to be developed. Tests of the single and multidomain custom-made cysteine protease inhibitors in transgenic plants for resistance against adult female WFT revealed that these insects can detect cysteine protease inhibitors in a matter of minutes after feeding on the plant and were strongly deterred by it. In a non-choice assay situation with purified inhibitors the WFT fecundity was up to 80% suppressed, and on plants with high levels of cysteine protease inhibitors the WFT population was up to 80% smaller after two weeks. Thus, it seems that female WFT select host plants with low levels of cysteine protease inhibitors where the development of the future population will be optimal. The dual insecticidal properties of the cysteine protease inhibitors (deterrence and suppression of population development) could be implemented in the resistance management strategies for protection against herbivorous insect pests.

CHAPTER 1

GENERAL INTRODUCTION

ABSTRACT

The development of novel insect resistance traits in field and greenhouse crops is an important challenge for plant biotechnology. Plant protease inhibitors are well known components of the plant defence strategy against insect herbivores. Insects, however, have developed various forms of resistance to inhibitors. Selection of heterologous inhibitors with better properties, and engineering their stable accumulation in plants are, therefore, critical in achieving transgenic plants with significantly improved defences against insects. In this thesis combinations of non-plant protease inhibitors are challenged against the adaptive potential of western flower thrips. Stable, high level expression of selected inhibitors in plants is approached by improving the promoter, the gene and the cellular environment. The results are combined into a sustainable solution for the control of Western flower thrips consisting of engineered multidomain inhibitors stably expressed in the endoplasmic reticulum of plant leaves using a strong promoter.

CROP PROTECTION AGAINST INSECT PESTS

Agricultural practices require crop plants to be grown in large genetically similar populations which are often replanted every year. This practice strongly favours the incidence of insect pests, and, if not controlled, considerable yield losses occur (Hilder and Boulter, 1999; Oerke et al., 1994; Smith, 1999). Traditional pest control involves the use of conventional pesticides, which in general are non-specific and wipe out the entire community of arthropods, pollutes the agro-ecosystem, and increases the cost of production. In contrast, genetically engineered insect-resistant crop varieties, which express insecticidal proteins derived from *Bacillus thuringiensis,* (Bt-toxins) have proven to provide an efficient way to control a number of major insect pests in crops like potato, cotton and maize (Cannon, 2000; Hilder and Boulter, 1999; Peferoen, 1997; Schuler et al., 1998, for reviews). The reduced use of conventional pesticides on these genetically modified crops has lead to an increased infestation by secondary pests such as thrips, aphids, and plant bugs (Cannon, 2000; Greene et al., 1999; Lynch et al., 1999; Turnipseed et al., 1995). Up to now there are no reports on Bt toxins or other principles which in transgenic plants are strongly active against these sucking pests, and in practice their control involves use of conventional chemicals (Cannon, 2000). In addition, the constant presence of Bt toxins in the crop plants and their acute toxicity to target insects creates a strong genetic selection for resistant phenotypes. Several instances of the development of resistance in target insects have been reported, although these cases all result from the use of conventional Bt sprays (Ferre and Van Rie, 2002, for review). In order to develop sustainable solutions for insect control more complex resistance management

strategies that involve several genes with different modes of action are needed. Precise characterisation of the effects of each individual gene is necessary. Principles already active in plants, such as protease inhibitors, may help to find such sustainable solutions.

WESTERN FLOWER THRIPS

Western flower thrips (WFT), *Frankliniella occidentalis* (Pergande) (Thysanoptera: Thripidae) is a highly polyphagous insect. WFT causes considerable economical losses on a number of vegetable, fruit, ornamental and plantation crops like onion, capsicum, cotton, tea, citrus, apple, strawberry and many more (Lewis, 1998). WFT is the most prevalent pest in the greenhouses throughout the world (Parrella, 1995; Shipp et al., 1998a; Shipp et al., 1998b; Shipp et al., 2000). WFT feeds by using their mouthparts to pierce through the plant epidermal cells and suck out the contents of underlying cells. The damaged plant parts collapse and result in formation of silvered patches or deformed plant growth of leaves, flowers, or fruits. In addition to the direct feeding damage, extensive economical losses occur when WFT transmit plant diseases (Tommasini and Maini, 1995; Wijkamp et al., 1995).

The small size of WFT (1-2 mm), combined with its polyphagous nature (Loomans et al., 1995), resistance to many pesticides (Jensen, 2000) and preferred feeding site in the inner whorls of flowers and buds make this particular insect extremely difficult to control. Insecticides must be applied at least three times and to avoid development of resistance a rotation of insecticide classes every 4-6 weeks is suggested (Robb and Parrella, 1995). This results in a hazardous accumulation of synthetic pesticides leading to allergies, skin and lung problems. Biological control using WFT's natural enemies often fails and is economically unsafe since the presence of beneficial insects reduces the market value of the product. A genetically determined resistance to WFT has been observed in many crop plants. Resistant cultivars, however, often lack other beneficial crop traits and as a result their use is greatly reduced. Furthermore, the complex inheritance of these traits and costly assessment of the resistance has prevented the implementation of systematic breeding programs focussed on resistance. Therefore, it is desirable to identify novel genes active against WFT that could be used in transgenic plants.

PROTEASE INHIBITORS AS PLANT PROTECTIVE MOLECULES

Plants in contrast to animals cannot simply escape from being consumed by herbivores, and they protect themselves by accumulation of numerous defensive molecules in their tissues (Karban and Baldwin, 1997; Kessler and Baldwin, 2002). Protease inhibitors (PI) in plants are known to be part of the defences that plants utilize to counteract the adverse effects from herbivore and pathogen attack (Green and Ryan, 1972; Koiwa et al., 1997; Ryan, 1990; Atkinson et al., 1984; Atkinson et al., 1995; Lorito et al., 1994). Plant PIs function against herbivores by decreasing the ability of the herbivore to derive protein from the plant by complexing and inhibiting a wide range of proteases found in insect guts (De Leo et al., 2002). They accumulate in massive amounts (up to 50% (Pearce et al., 1988)) in organs like seeds, fruits and tubers, but are also induced to high levels upon insect feeding, pathogen infection, and mechanical wounding of the aerial part of the plants (Bolter and Jongsma, 1995; Green and Ryan, 1972; Jongsma et al., 1994; Jongsma and Bolter, 1997; Wolfson and

Murdock, 1990). The ecological cost of expressing PI was recently revealed in a study by Glawe et al., (2003). Genotypes from *Nicotiana attenuata* collected from Arizona (A) were found to lack the ability to produce trypsin PI at a transcriptional level. They also had decreased volatile production, but exhibited nicotine and growth responses that were not distinguishable from genotypes collected in Utah (U) which had normal PI responses. In field trials with naturally occurring herbivores and in lab experiments with *Manduca sexta* larvae, A genotypes were damaged more and sustained greater herbivore growth than the U genotypes. When A and U genotypes were grown in competition, A genotypes produced significantly more seed capsules than the U neighbour. Moreover, jasmonate elicitation, which dramatically increased trypsin PI production in only the U genotypes, reduced lifetime fitness measures of the U genotypes more than of the A genotypes, demonstrating that trypsin protease inhibitor production is correlated with a fitness cost.

Herbivores on the other hand co-evolved mechanisms of adaptation that allow them to circumvent or dissolve these plant defences. The induction of novel proteases insensitive to PIs, and secretion of proteases that digest PIs present in the diet have been recognised as common mechanisms by which insect herbivores adapt to the presence of protease inhibitors (Bown et al., 1997; Broadway, 1996; Giri et al., 1998; Jongsma, 1995; Jongsma and Bolter, 1997; Markwick et al., 1998; Michaud et al., 1995a; Michaud et al., 1996; Volpicella et al., 2003; Wu et al., 1997). The degradation of alpha-amylase inhibitors by proteases in the gut of herbivores, and induction of alpha-amylases resistant to the inhibitors have also been observed (Markwick et al., 1996; Ishimoto and Chrispeels, 1996). However, even “adapted” insect herbivores on host plants with impaired induction of protease inhibitors perform significantly better than on wild type host plants. (Orozco-Cardenas et al., 1993; Royo et al., 1999). It seems, therefore, that the adaptations to the inhibitors are not complete or include their own costs, so that insects are still suffering from the presence of inhibitors.

Over-expression of protease inhibitors from non-host origin that target “newly induced PI-insensitive proteases” and complement the natural defences was proposed as a novel and potentially efficient method for insect control (Jongsma and Bolter, 1997; Jongsma et al., 1996). The success of this approach will largely depend on the adaptive potential of a given insect, careful selection of appropriate protease inhibitors and the ability to functionally express them in transgenic plants at adequate levels. In principle, for every particular inhibitor the ratio to the target protease, and the equilibrium dissociation constant of the complex are the primary critical determinants for the reduction of the digestive process. Inhibitors with special properties against the diversity of gut proteases are needed: they must be highly effective (>95% inhibition of the proteolytic activities), target induced enzymes, resist degradation, be expressed to high levels, and not affect the plant phenotype.

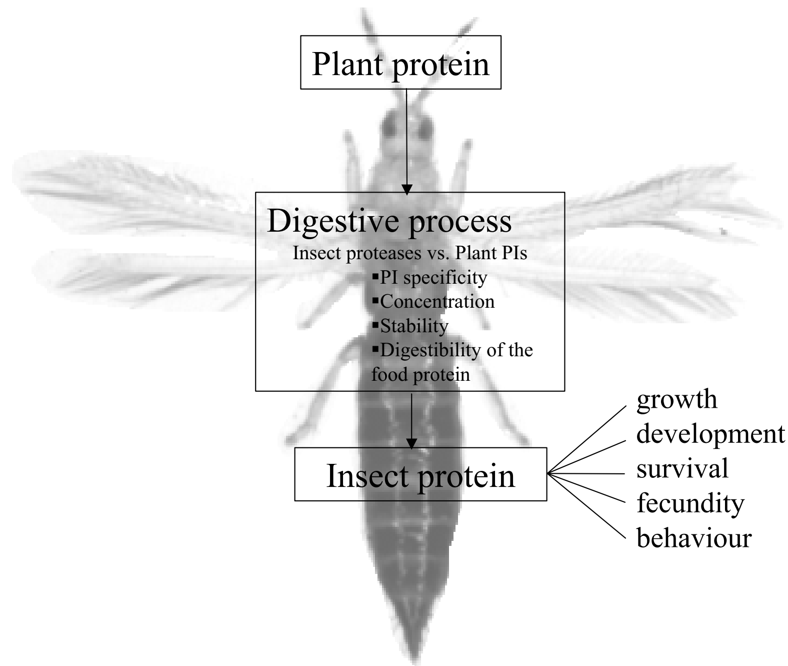


Figure 1. The efficacy of protease inhibitors for insect control in general depends on their ability to reduce the speed of the digestive process in the insect guts. Factors such as number of different proteases in the insect gut vs. inhibitors present in the plant, affinity of the PIs for proteases and concentration and stability of both are known to change the equilibrium of inhibition of the digestive process and thereby to change the effects on the target insects.

CLASSIFICATION AND MECHANISM OF ACTION OF PROTEASES AND INHIBITORS

To identify suitable inhibitors with the desired properties, it is important to understand the nature of the interaction between the proteases and their inhibitors and the complex task of not disturbing the plant enzymes, while fully blocking the herbivore ones.

Proteolytic enzymes, or proteases, catalyse the hydrolytic cleavage of peptide bonds in proteins. Based on the catalytic mechanism most eukaryotic proteases are classified into serine, cysteine, aspartic, and metallo-proteases (Barret et al., 1998) The MEROPS database (http://merops.iapc.bbsrc.ac.uk/) represents a comprehensive systematic classification of all known and putative types of proteases.

Proteinaceous protease inhibitors form stable complexes with proteases and this inactivates them. Protease inhibitors are generally classified on the basis of the proteases that they inhibit, and so there are inhibitors of serine, cysteine, aspartic, and metallo-proteases (Hartley, 1960; Laskowski and Kato, 1980). Further classifications are based on protein families of protease inhibitors are also in use.

Nature has invented different concepts for the interaction of proteases with inhibitors (Bode and Hubert, 2000, for review). Mostly protease inhibitors mimic the substrate of the protease, and directly contact, and thereby block the active site of the enzyme, i.e. “canonical” inhibitors. In other cases, the inhibitor does not bind directly to the substrate-binding site of the protease, but instead sterically prevents the uptake of the substrate. The above two interactions are very tight, but reversible. Serpins have evolved a different, third extraordinary “mousetrap” mechanism of inhibition in which by profound structural changes the serpin entraps the target protease in an irreversible complex (Huntington et al., 2000; Silverman et al., 2001). In all three cases, however, initial reversible protein-protein interaction in the active site of the protease or in adjacent position, between the protease and the inhibitor is a prerequisite for stable complex formation and inhibition (Bode and Hubert, 2000; Huntington et al., 2000; Silverman et al., 2001). Understanding the mechanism of interaction between the protease and the inhibitor is essential in understanding the mechanism of inhibition. For example the cysteine protease inhibitors, stefin B (cystatin family) and p41 (thyroglobulin type-1 or thyropin family), two proteins with totally different sequence and folding, demonstrate striking similarities in terms of interaction with papain-like cysteine proteases. Both inhibitors are wedge (V-) shaped structures, which have three interacting points with the active site cleft of the papain-like cysteine proteases (Guncar et al., 1999). This is interpreted by the authors as a case of convergent structural evolution of two unrelated groups of cysteine protease inhibitors.

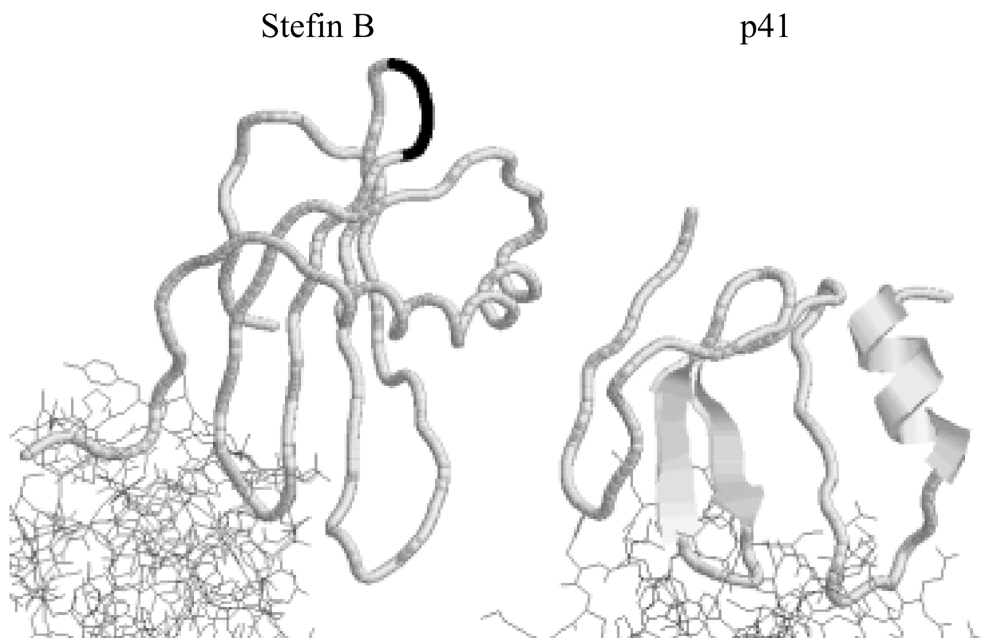


Figure 2. The structures of cystatin (Stefin B) in complex with papain (Stubbs et al., 1990)(left) and p41 invariant chain (thyropin) with cathepsin L (Guncar et al., 1999) (right). The inhibitors are shown as ribbons. The back-side loop involved in the inhibition of the second cysteine protease for some cystatins is mapped on Stefin B as a black ribbon fragment.

THE ROLE OF MULTIDOMAIN PROTEASE INHIBITORS

In addition to the various different types of single domain interactions with proteases, nature has also created many examples of so-called multi-headed or multidomain inhibitors, in which single proteins can bind multiple proteases. Multicystatins can bind up to eight cysteine proteases like papain (Kouzuma et al., 2000; Waldron et al., 1993a; Wu and Haard, 2000). Ovomucoid contains seven kazal inhibitory domains (Scott et al., 1987) and soybean bowman-birk, and potato PI2 each have two domains (Beekwilder et al., 2000; Song et al., 1999). PI2 serine protease inhibitors from *Nicotiana alata* are composed of four to six repeated domains (Atkinson et al., 1993; Miller et al., 2000; Schirra et al., 2001) Equistatin is an inhibitor consisting of three thyropin domains: domain I inhibits cysteine proteases, domain II inhibits aspartic proteases, and the function of domain III is not yet known (Lenarcic and Bevec, 1998; Lenarcic and Turk, 1999). Some members of the cystatin family, 10-15 kD in size, have evolved a loop on the "back" of the molecule that acts as a secondary reactive centre and enables some cystatins to bind simultaneously two proteases (Alvarez Fernandez et al., 1999) (Figure 1). Similar evolution has occurred in members of the serpin and plant Kunitz families where a single "globular domain" evolved the ability to inhibit both serine and cysteine proteases at two different reaction centres (Al-Khunaizi et al., 2002). This apparent multiple inhibitory function on a single reading frame seems to be advantageous. It may increase the inhibitory potential and spectrum of a given protein while saving resources for protein synthesis. By multivalent binding to multimeric proteases the primary stoichiometric interaction in each of the domains may turn into an irreversible complex that inactivates simultaneously multiple proteases, and possibly precipitates them out of a solution. Also the compact multidomain protein might be more resistant to proteolysis, or form crystalline inclusions that are dissolved only when the protein is needed by changing the pH, like in the case of the potato multicystatin (Walsh and Strickland, 1993).

SCOPE OF THE THESIS

The aim of the work laid down in this thesis has been to obtain transgenic plants expressing different cysteine protease inhibitors at high levels and to assess the level of resistance to insects such as Colorado potato beetle and western flower thrips. Equistatin, a protein from sea anemone, was of major interest at the beginning of the research project, since it was proven to posses strong inhibitory and insecticidal properties (Gruden et al., 1998). Main focus was then to produce sufficient quantities of recombinant protein for insect bioassays in which the potential of the purified protein could be assessed. The production of equistatin in *P. pastoris* was optimised 20-fold, reaching a maximum of 1.66 g/L of expressed recombinant protein (Chapter 2). The purified inhibitor was then tested in a number of different bioassays by partners (Annadana et al., 2002b; Bown et al., in preparation; Deraison et al., in preparation) collaborating in the EU project "Novel PI crops" which co-funded a large part of the presented work. Furthermore, to engineer high expression levels in plants this thesis describes (Chapter 3) the cloning and use of novel gene regulatory elements of the small subunit of ribulose biphosphate carboxylase gene (*RbcS1)* from chrysanthemum (*Chrysanthemum morifolium* Ramat.). *RbcS1* yielded expression levels 8 fold better compared to the commonly used 35S promoter. Selected cysteine protease inhibitors (equistatin, kininogen domain 3, stefin A, cystatatin C and potato

cystatin) were then expressed under the control of the new *rbcS1* promoter, in transgenic potato plants, and the effects of accumulation of inhibitors on WFT food preferences were evaluated (Chapter 4). It was found that adult female WFT are deterred from host plants with high levels of functionally expressed inhibitors. The degradation of some inhibitors targeted to the secretory pathway in potato plants significantly hampered the insecticidal effect of most of the PIs. The proteases involved in the degradation of the heterologous Pi (equistatin) were then studied in detail (Chapter 5). Two different strategies to engineer the stability of heterologous inhibitors in plants: (i) exchanging amino acids sensitive for cleavage (P_1) by the endogenous protease (Chapter 6), and (ii) generation of a recombinant multidomain inhibitor (Chapter 7), were then evaluated. The latter method resulted in accumulation of undegraded multidomain inhibitors in potato plants and resulted in the development of WFT resistant plants. The cloning and characterization of two natural multidomain inhibitors found in the genomic DNA of the *Drosophila melanogaster* with potential for insect resistance applications is furthermore described in chapter 8. It is apparent that the artificial multidomain inhibitors posses the desired custom made properties in contrast to the ones that nature created in the fruit fly. In order to examine the potential for insect resistance applications of natural, non-plant multidomain inhibitors the cloning and characterization of two genes from fruit fly *D. melanogaster* is described in chapter 8. These inhibitors containing four different protein families of serine and cysteine inhibitors did not inhibit fruit fly proteases to a significant extend. The engineered multidomain approach described in this thesis, therefore, seems to be more successful.

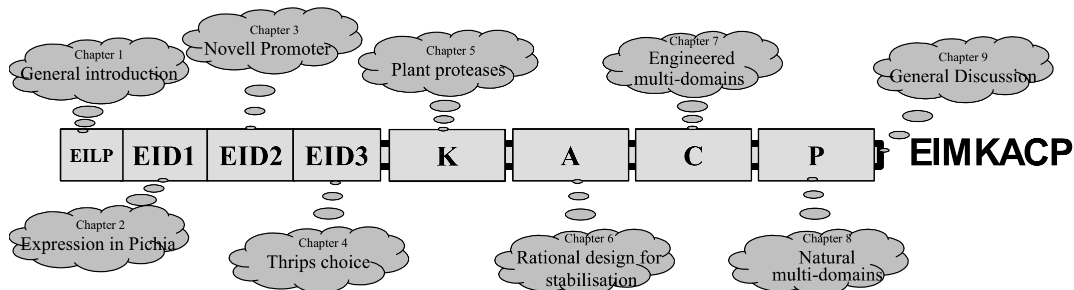


Figure 3. Metaphoric comparison of the chapters of the thesis with the recombinant seven-domain inhibitor EIMKACP. General introduction corresponds to the leader peptide of equistatin (EILP). The seven domains represent the research chapters, (EID1 to 3- the three domains of the sea anemone equistatin, K- human kininogen, A-human stefin A, C-cystatin C, P- potato cystatin) and the C-terminal KDEL retention signal corresponds to the general discussion.

CHAPTER 2

OPTIMIZATION OF THE EXPRESSION OF EQUISTATIN IN *PICHIA PASTORIS*

Nikolay S. Outchkourov[1], Willem J. Stiekema[2] and Maarten A. Jongsma[1]

Business Units Cell Cybernetics[1] and Genomics[2], Plant Research International, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands

Protein Expression and Purification (2002) 24:18-24

ABSTRACT

To improve the expression of equistatin, a proteinase inhibitor from the sea anemone *Actinia equina,* in the yeast *Pichia pastoris,* we prepared gene variants with yeast-preferred codon usage and lower repetitive AT and GC content. The full gene optimization approximately doubled the level of steady state mRNA and protein accumulated in the culture medium. The removal of a short stretch of 12 additional nucleotides from the multiple cloning site (MCS) sequence in the vector pPIC9 had an enhancement effect similar to full gene optimization (factor 1.5) at the mRNA level. However, at the protein level, this increase was 4- to 10-fold. The optimized gene without the MCS sequence yielded 1.66 g/L active protein in a bioreactor and was purified by a new two-step procedure with a recovery of activity that was >95%. This production level constitutes an overall improvement of about 20-fold relative to our previously published results. The characteristics of the MCS sequence element are discussed in the light of its apparent ability to act as negative expression regulator.

INTRODUCTION

Equistatin, a protein found in the sea anemone *Actinia equina,* is a potent inhibitor of both cysteine and aspartic acid proteases (Lenarcic and Turk, 1999). Its amino acid sequence reveals that it is structurally related to thyroglobulin type-1 domains (Lenarcic et al., 1997; Lenarcic and Turk, 1999). The thyroglobulin type-1 domains are found throughout the animal kingdom from nematodes to humans, generally as part of larger protein structures (Molina et al., 1996). Equistatin consists of three thyroglobulin type-1 domains. The first N-terminal domain acts as a cysteine protease inhibitor (K_i for papain of 0.57 ± 0.04 nM). The second as an aspartic acid protease inhibitor (K_i for cathepsin D of 0.3 ± 0.15 nM) (Lenarcic and Bevec, 1998; Strukelj et al., 2000).

Previously, the wild type equistatin gene was expressed in *E. coli* at 1.2 mg/L (Galesa et al., 2000) and in *Pichia pastoris* at up to 25 mg/L (Rogelj et al., 2000). The relatively low protein yield obtained in this yeast was in contrast with its reported high potential capacity for producing protein. The low yields suggested there might be problems at one or more points in the chain of gene expression events. To potentially solve the low expression problem, we optimized the gene by means of substituting rare codons with more frequently occurring ones and increasing the GC nucleotide content of the gene (Kotula and Curtis, 1991; Romanos et al., 1992; Sreekrishna et al., 1997).

The coding sequence of the wild type (wt) equistatin gene (EI) had previously been optimized for expression in the potato, *Solanum tuberosum,* by site-directed mutagenesis, resulting in modified equistatin (EIM) (Rogelj et al. unpublished; Outchkourov et al., 2003a). Codons with an abundance of <10% were replaced with more frequently occurring ones, and potential mRNA instability or polyadenylation motifs represented by five or more A/T or G/C repeats (Strizhov et al., 1996) were removed, whilst preserving the protein sequence. As the yeast *P. pastoris* has codon preferences similar to those of the potato, the optimization strategy for potato resulted

in a gene, which was also optimal for use in yeast. We expressed the modified equistatin (EIM) gene optimized for codon usage and repetitive AT or GC content in *P. pastoris* and expected to obtain much higher protein levels. In order to assess the actual improvement factor, we compared expression from the EIM gene to the wt EI gene inserted at both the same and a different restriction site in the multiple cloning site sequence of the commonly used *P. pastoris* expression vector pPIC9. Chimeras of the optimized and wt genes were also prepared to evaluate the effect of a partial gene optimization.

MATERIALS AND METHODS

Materials

The pPIC9 plasmid and *P. pastoris* strain GS115, which form part of the *P. pastoris* expression kit was purchased from Invitrogen, Carlsbad, CA, USA. Media components were obtained from Difco, Detroit, MI, USA. *Escherichia coli* strain XL1 Blue (Stratagene Inc., La Jolla, CA, USA) was used for the construct preparation. The chromatography materials used were exclusively from Amersham Pharmacia Biotech.

Oligonucleotides

The following primers (Eurogentec, Seraing, Belgium) were used.

pPIC9/EI-N 5'-CTCTCGAGAAAAGAGAGGCTGAAGCTAGTCTAACCAAATGCCAAC
pPIC9/EI-C 5'-GATGCGGCCGCTTAGCATGTGGGGCGTTTAAA
pPIC9/EIM-N 5'-CTCTCGAGAAAAGAGAGGCTGAAGCTAGTCTAACGAAATGCCAAC
pPIC9/EIM-C 5'-GATGCGGCCGCTTAGCATGTGGGACGTTTGAA

Cloning and expression of equistatin variants in *P. pastoris*

The coding regions of the wild type EI (Rogelj et al., 2000) (Accession No: AF184891) and optimized EIM genes (AY166597) were amplified by PCR using *Pwo* DNA polymerase (Roche). The following combinations of primers were used, pPIC9/EI-N plus pPIC9/EI-C for EI and pPIC9/EIM-N plus pPIC9/EIM-C for EIM, introducing *Xho*I and *Not*I restriction sites to the 5' and 3' ends. Use of the *Xho*I site, before the cleavage site of the α-mating factor signal peptide, necessitates the recreation of the sequence between the restriction site and signal peptide cleavages *KEX2* and *STE1* for removal of the signal peptide. These sequences were incorporated into the primers. Amplifications were carried out for 30 cycles comprising 30 sec at 94°C, 30 sec at 52°C and 1 min at 72°C. Fragments were digested with *Xho*I and *Not*I and cloned into the pPIC9 expression vector. A unique *Fsp*I site in the equistatin gene was used to prepare the fusion constructs EI/EIM and EIM/EI in a three-point ligation directly into pPIC9. All of the constructs were sequenced on an Applied Biosystems 370A sequencing machine. The constructed vectors were linearized with *Sal*I and transformed into *P. pastoris* strain GS115 (*his4*)

by electroporation using a Gene Pulser electroporator (Bio-Rad, USA). Transformants were selected on MD agar plates (1.34% yeast nitrogen base, 4 × 10^{-5}% biotin, 1% D-glucose, 1.5% agar). To study the influence of twelve additional nucleotides derived from the MCS of vector pPIC9 the two highest expressers selected from 48 *P. pastoris* transformants, cloned into the *Eco*RI-*Not*I site were used (EI+MCS 5 and EI+MCS 16)(Rogelj et al., 2000). Small-scale protein expression was performed in 24-well plates as described in (Volpicella et al., 2000). Shake flask expression studies were performed by pre-growing cultures in a shaking incubator at 250 rpm for 24 hours at 30°C in 1 liter baffled shake flasks containing 50 ml BMG medium (1.34% yeast nitrogen base, 4 × 10^{-5}% biotin, 1% glycerol, 0.1 M K-phosphate, pH 6.0). The cultures were centrifuged and the pellet containing the cells was resuspended in BMM (1.34% yeast nitrogen base, 4 × 10^{-5}% biotin, 0.5% methanol, 0.1 M K-phosphate, pH 6.0) and grown under the same conditions for 8 hours. Samples were taken at 4, 6 and 8 hours after the methanol induction to monitor the rate of protein secretion and RNA accumulation. Large-scale productions in the bioreactor were carried out as previously described (Rogelj et al., 2000).

Quantification and characterization of equistatin

The concentration of equistatin was determined by either immunological or activity measurements. For dot blot immunological measurements 20 µl culture supernatants were mixed in a ratio of 1:5 with 100 mM CAPS, 0.1% SDS, pH 11 and spotted on nitrocellulose membranes using the SRC 96 D Dot blot apparatus (Schleicher & Schuell, Germany). Membranes were blocked in TTBS (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween 20) containing 2% non-fat milk powder for 1 hour and then incubated with rabbit anti-EI antibodies (Eurogentec, Seraing, Belgium). The blots were subsequently washed and incubated with goat anti-rabbit IgG conjugated with horseradish peroxidase (Jackson Immuno Research, USA). The dots were visualized with Lumi-Light Western blotting substrate and scanned in the Lumi-Imager F1™ under the control of Lumi-Analist™ software (Roche). Inhibitory activity concentration was determined by titration against papain as described (Rogelj et al., 2000).

The protein was characterized by SDS-PAGE analysis on a 15% precast resolving gel (BioRad, USA). The gels were run according to the manufacturer's instructions and the bands were visualized with Coomassie Brilliant blue R250 staining. The amount of total protein in the samples was quantified according to Bradford's method (Bradford, 1976).

RNA analysis

For RNA analysis the cells from the shake flask cultures were grown for up to 8 hours after methanol induction as described above. The total RNA was extracted from the cells as described in (Schmitt et al., 1990) and the Northern blot procedure was performed in accordance with (Kevil et al., 1997). Filters were hybridized with a ^{32}P labeled probe of *Bgl*II-*Not*I 1792-bp fragment of EI/EIM construct containing the AOX1 promoter, alpha-mating factor signal peptide and a fusion of the wild and

optimized gene. The use of this probe ensured cross-hybridization to the 5' untranslated mRNA leader of the AOX1 endogenous transcript as well as to the equistatin. This made it possible to simultaneously quantify the steady state mRNA levels of the two genes under the control of the same promoter. For a positive control, RNA from the EIM gene synthesized with Riboprobe® in vitro Transcription Systems (Promega) was used. Band intensities were visualized by autoradiography and quantified using a BAS-2000 Phospho-imaging scanner (Fuji).

Large-scale equistatin purification

The cells were removed from the culture by two centrifugations at 28000 x g. Cell-free supernatant was filtered through a 0.45 μm filter, mixed with 20% column equilibration buffer 20 mM MES, 100 mM $(NH_4)_2SO_4$, pH 6.5 and applied to Phenyl Sepharose 200 ml column. The column was washed twice with two column volumes of 75 mM and 50 mM $(NH_4)_2SO_4$ respectively and bound equistatin was eluted with demineralized water. The entire procedure was monitored at 280 nm. A further purification step using anion exchange chromatography was performed on a HiTrap Q Sepharose 5 ml column. The pH of the solution was adjusted to 8.5 by adding 2 M Tris to the final concentration 16 mM and aliquots containing about 100 mg protein were applied. Stepwise elution was performed with 50 mM, 100 mM, 150 mM and 200 mM NaCl, respectively. After analysis on 15% SDS-PAGE, the fractions containing the undegraded molecular size band were pooled. The buffer was exchanged to 20 mM ammonium carbonate on Vivacell 70, 5.000 MWCO filters (Vivascience) and lyophilized.

RESULTS

Analysis of the effects of gene modifications on protein expression levels

The pPIC9 expression vector utilizes the *Saccharomyces cerevisiae* α-mating factor pre-propeptide to secrete the protein of interest into the culture medium. The protein to be expressed is fused in reading frame to this secretion signal. Proteolytic cleavage sites situated between *Xho*I and *Sna*BI restriction sites of the pPIC9 polylinker are utilized by KEX2 and STE13 proteases in the *P. pastoris* secretory pathway to remove the pre-propeptide. Therefore, subcloning in the *Sna*BI restriction site and other downstream sites (*Eco*RI, *Avr*II, *Not*I) result in additional amino acids at the N-terminus of the protein which are derived from the multiple cloning site and not from the inserted gene.

To compare expression levels, the modified equistatin EIM gene and the wt EI gene were cloned into the *Xho*I site of the pPIC9 vector in such a way that no additional amino acids were created between the pre-propeptide of the vector and the mature protein part of the equistatin gene. To assess the effects of gene regions on gene expression, chimaeric genes EI/EIM and EIM/EI containing modified and unmodified fragments of similar size were also constructed and cloned into the same *Xho*I site. A comparison with a previously published construct, which had the EI gene cloned in the *Eco*RI site, yielding a protein product with four additional amino

acids at the N-terminus, was also made (Rogelj et al., 2000)(Fig. 1). For every new construct, 20 His$^+$ clones were randomly chosen to run a statistical analysis of the expression levels. All of the clones were grown to the same density and methanol induced for 24 hours. The experiment was repeated twice. The results of the second experiment were representative for those of the first and are shown in Fig. 2**.** The accumulated equistatin in the supernatants was measured by a dot blot antibody analysis. The non-expressing clones (between 5 and 10% of the population) were excluded from further analysis. The levels of equistatin secreted in the supernatant are shown in the box plot graph. For each construct, this graph shows the percentile distribution in the population.

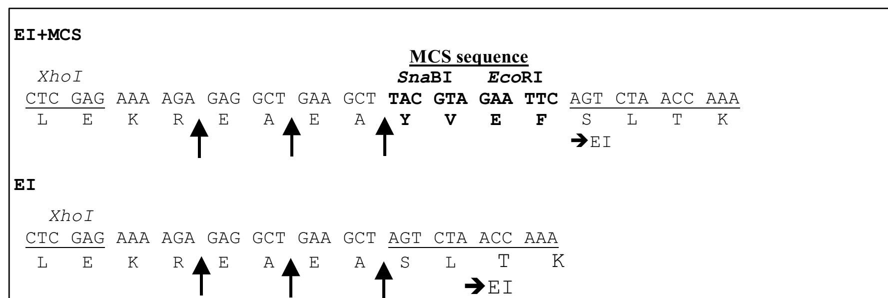


Figure 1. The fusion points of the mating factor alpha secretion signal sequence and equistatin mature protein in EI+MCS and EI constructs. The additional MCS sequence is indicated in bold. Arrows indicate the Kex2 and Ste13 cleavage positions of the signal peptide.

The largest variation in expression was observed in the EIM population. This was mainly caused by a relatively large number of high expressing clones (up to 26 mg/L). The EI, EI/EIM and the EIM/EI populations did not yield such high-expression clones (up to 14 mg/L), which resulted in a reduced variability within these populations. All of the constructs had a nearly equal percentile of low expressing clones (4 mg/L), hence the difference in median for the EIM population (11 mg/L), was moderate compared with other populations (7-9 mg/L).

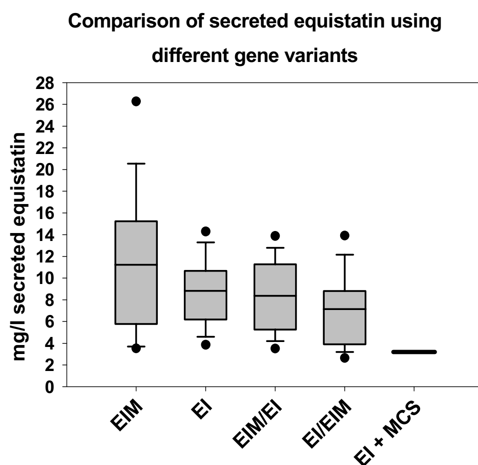


Figure 2. Effects of gene optimization on the equistatin expression levels after 24 hours growth in 500 µl cultures in 24-well plates. The box plot graphs represent statistical values indicating the level of expression in a group of 20 randomly chosen His+ transformants. The boundaries of the box are the 25th and 75th percentile. The line within the box marks the median. Whiskers above and below the box indicate the 90th and 10th percentiles. The 95th and 5th percentiles are indicated with circles. The horizontal line on the right indicates the average expression level of the two best expressing transformants with the EI gene cloned in the *Eco*RI site of pPIC9 and corresponds to the top circle on a box plot. Explanation of abbreviations: EIM, modified equistatin; EI, wild type (wt) equistatin; EIM/EI and EI/EIM, fusions of modified and wt equistatin; EI+MCS, EI plus 12 nucleotides derived from the multiple cloning site.

In contrast, the two highest expressing clones selected from the 48 transformants cloned in the *Eco*RI site (EI+MCS 5 and EI+MCS 16) as previously published (Rogelj et al., 2000) were expressing in the range of the lowest transformants cloned in the *Xho*I site (3-4 mg/L).

Initial protein secretion rate and RNA accumulation

For a precise evaluation of the differences in protein secretion rate and steady state mRNA level between the constructs, glycerol cultures of the two highest expressing clones with an identical initial OD_{600}=1 were induced with methanol and grown in shake flasks. At 4, 6 and 8 hours after the induction samples were taken for analysis. At the protein level, a significant improvement (10-fold to 20-fold) was observed for both EI and EIM cloned into the *Xho*I site compared with the previous *EcoR*I cloned EI+MCS construct (Fig. 3B). The gene optimization itself did not seem to have a significant effect on the initial rate of foreign protein synthesis.

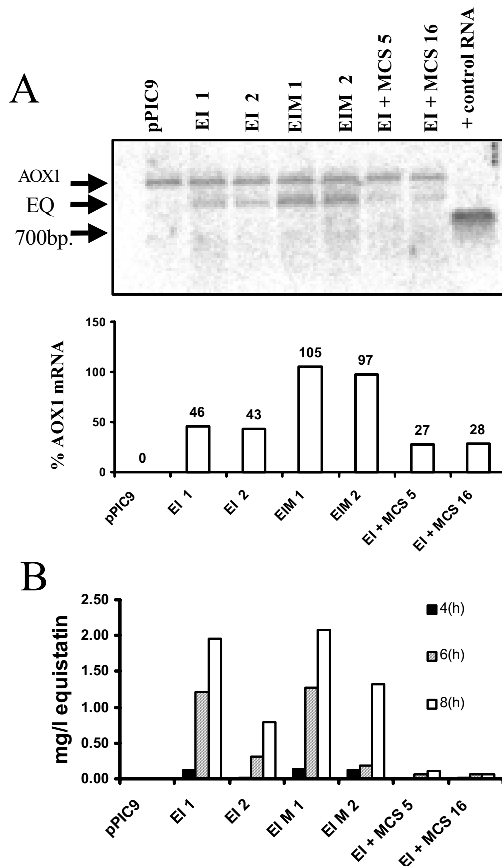


Figure 3. Initial levels of equistatin steady state mRNA and protein secretion of the two highest expressing clones per construct. **A.** Northern blot of the steady state mRNA levels with quantification in the graph below. A *Bgl*II-*Not*I 1792-bp ^{32}P labeled fragment of the EI/EIM construct containing the AOX1 promoter, alpha-mating factor signal peptide MF, and the fused wt and optimized gene was used as a probe. The AOX1 promoter sequence cross-hybridized with the 5' mRNA leader of the endogenous transcript and resulted in an additional band. The band was used as an internal control to quantify the differences. **B.** Protein secretion levels 4, 6 and 8 hours after methanol induction were measured by means of a dot blot immunological detection.

The effects on the steady state mRNA levels, for the same cultures, 8 hours after methanol induction (Fig. 3A) were quite different from the effects on the protein levels. For these northern blots, a probe was used which was chimaeric for the *AOX1*

untranslated 5' mRNA, the *S. cerevisiae* mating factor alpha signal sequence and the equistatin gene, which in turn was chimaeric for the modified and wt equistatin gene (EI/EIM). The presence of the *AOX1* sequence resulted in an additional hybridizing band which was larger than the equistatin mRNA and was also present in the control strain. We used this band as an internal control to normalize the differences in mRNA levels of the equistatin gene relative to the *AOX1* transcript. A further advantage of this probe was that both genes were under the control of the same promoter and differences in the relative amount of mRNA within each sample were a useful indication for gene optimization. The fully modified gene (EIM) had steady state mRNA levels similar to the *AOX1* gene while reductions of ≈56% and ≈73% were observed for the EI and EI+MCS clones. The 10-fold to 20-fold reduction in protein yield observed in the EI+MCS relative to the EI clones was, therefore, not supported by a similar reduction in mRNA (1.5-fold), suggesting a post-transcriptional cause for the expression difference. For all of the constructs only full-length equistatin transcripts were observed without significant amounts of degradation or premature polyadenylation.

Bioreactor production and equistatin purification

The optimized gene (EIM 1) was expressed in the bioreactor, with densities of up to 450 OD_{600}. After 85 hours of methanol induction this yielded 1.66 g/L of active equistatin (Fig. 4). We observed a direct correlation, between the duration of fermentation and equistatin degradation during the subsequent storage and purification.

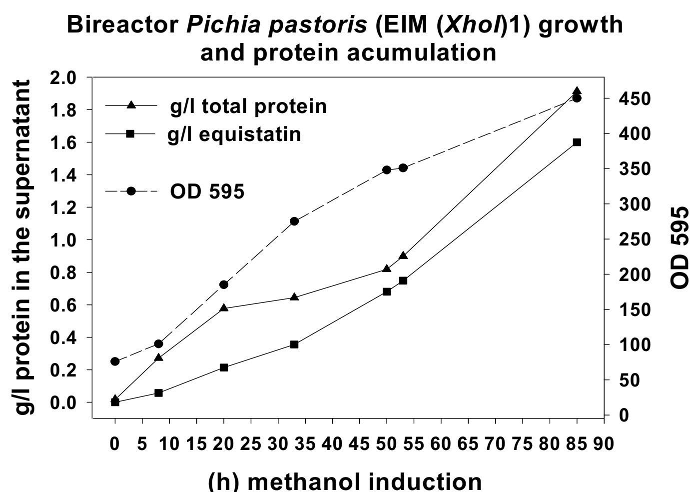


Figure 4. Time-course of equistatin (EIM1) fermentation in a bioreactor. The cell density, level of secreted protein, and the corresponding level of papain inhibitory activity in the supernatant are related to the time after methanol induction.

The highest yield of fully intact, undegraded protein was obtained by running the fermentation for no longer than 30-35 hours, which yielded 350 mg/L of active equistatin. A new two-step protein purification protocol was developed with a complete recovery of activity. After the first step, phenyl sepharose chromatography, equistatin was eluted in water as the only protein band visible on Coomassie stained gel. The second purification step by means of anion exchange chromatography was

necessary to remove brownish colorants from the preparations. In this purification step, equistatin was eluted at 150 mM NaCl as a clear solution with >95% of the initial activity recovered. On SDS-PAGE, the mobility of the purified protein was found to correspond to a native molecular weight of 22 kDa. The absence of a reducing agent increased the mobility of this highly disulfide-bonded protein and did not reveal the extra bands indicative of partially degraded equistatin (Fig. 5). Apparently, the disulfide bonds keep the protein intact despite the cleavage of a small portion of the peptide bonds.

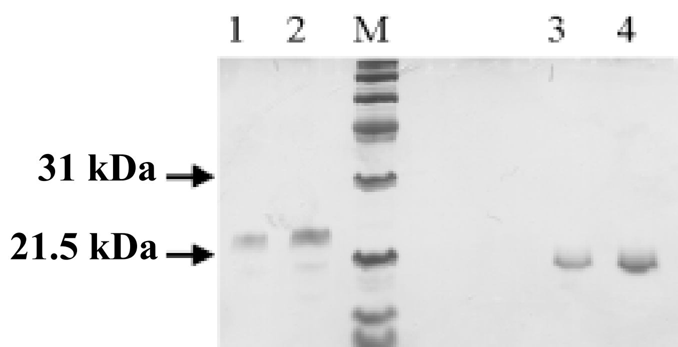


Figure 5. SDS-PAGE analysis of purified equistatin protein. *Lanes 1 and 2*, equistatin after DTT treatment; *lanes 3 and 4,* non-reduced equistatin, 300 and 600 ng respectively. *M,* molecular weight standards.

DISCUSSION

Protein production in *P. pastoris* has been the subject of much research in recent years and many proteins have been produced with varying degrees of success. The reasons why some proteins have low or intermediate expression levels have largely remained unclear. We have evaluated the effect of gene optimization by means of improved codon usage and lower repetitive AT and GC content, as well as the effect of utilizing a different cloning insertion site. The combined effect of these changes was an estimated 20-fold increase in protein production compared with our EI+MCS construct as previously published (Rogelj et al., 2000).

Gene optimization generated a gene, which was improved for *P. pastoris* in terms of codon usage and lower repetitive AT and GC content. It resulted in a subset of *P. pastoris* transformants with a higher relative expression level than the wt control group. The effect was at most twofold and was consistently observed in the population of transformants after 24 hours of induction. For the clones with the highest expression level, the initial rates of protein synthesis did not consistently differ between the optimized EIM gene and the wild-type EI gene. However, the experimental error in these immunological measurements is significant, especially at low signal levels. For the same cultures, we observed a consistent difference (~twofold) in the amount of equistatin mRNA which had accumulated after 8 hours. We believe this measurement was more accurate since we had a good internal control and we, therefore, assume that the gene optimization resulted in a net effect of an approximately twofold increase in the protein synthesis level supported by an approximately twofold increase in the steady state mRNA level. It was not possible to evaluate which modification was the most significant: improved codon usage or lower

repetitive AT and GC content. Furthermore, the chimaeric EI/EIM and EIM/EI genes did not appear to exhibit significantly improved expression levels relative to EI alone. Therefore, it would seem that the fully modified gene is required for this effect to be significant and measurable.

Our most interesting finding was the significant effect of the vector-cloning site on the expression levels. We previously described expression of equistatin with the EI+MCS construct in *P. pastoris* where a maximum expression level of 25 mg/L was obtained after 48 hours fermentation in a bioreactor (Rogelj et al., 2000). With the optimized gene construct described here, we obtained 1660 mg/L after 85 hours and 350 mg/L after 35 hours. This difference could not be explained by our gene optimization efforts, which appeared to have only a twofold effect at most. Eight hours after induction with methanol, the mRNA level was 1.5 times higher when the gene was cloned in the *Xho*I site (EI) as opposed to the *Eco*RI site (EI+MCS). However, this was not sufficient to explain the 10 to 20-fold increase in the protein expression levels observed with these clones.

To gain insight into a potential cause of the protein level improvement, we examined the nature of the additional stretch of 4 amino acids/12 nucleotides (TACGTAGAATTC), which contain the *Sna*BI and *Eco*RI sites that were present in our earlier construct. In the EI+MCS mRNA we observed the presence of a potential hairpin structure, not present in EI, consisting of 5 paired nucleotides and a loop of 12 nucleotides containing 9 nucleotides of the MCS. This structure was much smaller, with much less base pairing relative to structures observed in the rest of the gene. Thus, we concluded that a hairpin formed in part by these 12 additional bases was not likely to be responsible for the reduction in the production of equistatin protein we observed.

A search of the yeast literature for homology between these added bases and known elements which affect transcription/translation, revealed that the part of the sequence represented by the *Sna*BI site (TACGTA) can act as an efficiency element in 3'-RNA formation in the yeast *Saccharomyces cerevisiae* (Irniger and Braus, 1994; van Helden et al., 2000). The yeast efficiency element is part of three sequence components that constitute the yeast polyadenylation signal. The other two are a positioning element and the actual polyadenylation site (Guo and Sherman, 1995; Guo and Sherman, 1996; Russo et al., 1993). Using mutational analysis, the efficiency element was initially revealed to be an element for determining the efficiency of polyadenylation in *in vivo* experiments (Russo et al., 1993). Recently, a deeper understanding of its role has been gained by the demonstration that heterogeneous nuclear ribonucleoprotein (Hrp1), an essential component for mRNA 3'-end formation, can be cross-linked to the efficiency element of mRNA (Minvielle Sebastia et al., 1998). The ability of Hrp1 to travel from the nucleus to the cytoplasm suggested a role in mRNA transport (Chen and Hyman, 1998; Kessler et al., 1997). Usually, efficiency elements are highly over-represented downstream of the stop codon in *S. cerevisiae* mRNAs (van Helden et al., 2000). We suggest that the presence of this type of efficiency element in the coding part of our EI gene may have inhibited protein translation. The fact that these efficiency elements bind proteins, which are also present in the cytoplasm (Kessler et al., 1997), could be sufficient to greatly reduce the efficiency of ribosomal translation and thereby the levels of EI protein synthesized.

CHAPTER 3

THE PROMOTER TERMINATOR OF CHRYSANTHEMUM *RBCS1* DIRECTS VERY HIGH EXPRESSION LEVELS IN PLANTS

Outchkourov NS, Peters J, de Jong J, Rademakers W, Jongsma MA

Plant Research International, PO Box 16, NL-6700 AA Wageningen, The Netherlands

Planta (2003) 216: 1003-1012

ABSTRACT

Transgenic plants are increasingly used as production platforms for various proteins, yet protein expression levels in the range of the most abundant plant protein, ribulose-1,5-bisphosphate carboxylase have not yet been achieved by nuclear transformation. Suitable gene regulatory 5' and 3' elements are crucial to obtain adequate expression. In this study an abundantly transcribed member (*rbcS1*) of the ribulose-1,5-bisphosphate carboxylase small-subunit gene family of chrysanthemum (*Chrysanthemum morifolium* Ramat.) was cloned. The promoter of *rbcS1* was found to be homologous to promoters of highly expressed *rbcS* gene members of the plant families Asteraceae, Fabaceae and Solanaceae. The regulatory 5 and 3 non-translated regions of *rbcS1* were engineered to drive heterologous expression of various genes. In chrysanthemum, the homologous *rbcS1* cassette resulted in a glucuronidase (gusA) accumulation of, at maximum, 0.88% of total soluble protein (population mean 0.17%). In tobacco (*Nicotiana tabacum* L.), the gusA expression reached 10% of total soluble protein. The population mean of 2.7% was found to be 7- to 8-fold higher than for the commonly used cauliflower mosaic virus (CaMV) 35S promoter (population mean 0.34%). *RbcS1*-driven expression of sea anemone equistatin in potato (*Solanum tuberosum* L.), and potato cystatin in tomato (*Lycopersicon esculentum* Mill.) yielded maximum levels of 3-7% of total soluble protein. The results demonstrate, that the compact 2-kb *rbcS1* expression cassette provides a novel nuclear transformation vector that generates plants with expression levels of up to 10% of total protein.

INTRODUCTION

In genetically modified plants, high-level, tissue-specific transgene expression is required for a variety of traits, and numerous promoters to ensure this have been described in the literature. Known strong constitutive promoters in plants include the cauliflower mosaic virus (CaMV) 35S promoter (Ow et al., 1987), opine synthase promoters (Harpster et al., 1988), actin promoters (McElroy et al., 1991) and ubiquitin promoters (Kawalleck et al., 1993). Other non-constitutive strong promoters include the light-regulated chlorophyll a/b-binding protein promoter (Nap et al., 1993) or the promoters of the small subunit of ribulose bisphosphate carboxylase (rbcS) (Gittins et al., 2000; Khoudi et al., 1997; Nomura et al., 2000).

Ribulose bisphosphate carboxylase (RBC) is the primary enzyme of the carbon fixation process (EC 4.1.1.39). It constitutes approximately 50-60% of total leaf protein (Dean and Leech, 1982) and is arguably the most abundant protein on earth. RBC is composed of 8 small subunits of 14 kDa (rbcS) encoded by a gene family of 2-12 genes on the nuclear genome and 8 large subunits of 55 kDa (rbcL) encoded by a single gene on the chloroplast genome. The expression of the *rbcS* genes is light-dependent and co-ordinated with chloroplast development (Dean and Leech, 1982). A great number of light responsive cis-acting regulatory units have been identified in the *rbcS* promoters. However, the factors and sequence elements involved in chloroplast-dependent regulation still remain elusive. The different

members of the *rbcS* gene family are not expressed at similar levels. Usually one or two gene members are very strongly transcribed, representing up to 50% of total rbcS transcripts, while the rest are less active. Diversity in the promoter, coding part, and terminator sequences of *rbcS* genes contribute to the quantitative differences in their gene expression (Dean et al., 1989c). It can be calculated that the most abundantly expressed rbcS gene contributes 5-6% of total soluble protein, so that similar expression levels would be expected from constructs using such a promoter. Various investigators have engineered the regulatory elements of abundantly expressed *rbcS* genes to express heterologous proteins in plants. They were found to be broadly active across different plant species, and their expression was generally localized in the green plant parts (Gittins et al., 2000), yet the expression levels in general did not exceed 1% of total soluble protein.

With the aim of obtaining regulatory sequences with strong transcriptional activity the gene encoding the most abundant *rbcS* transcript of cultivated chrysanthemum was isolated. Cultivated chrysanthemum (*Chrysanthemum morifolium* Ramat.) is a hexaploid member of the relatively recently evolved Asteracea family (Meagher et al., 1989). A sequence relationship of the new *rbcS1* gene with known members of the *rbcS* gene family was investigated by homology comparison. One-kilobase promoter and terminator elements were subcloned to create a cassette for cloning and expression of genes of interest. The expression strength of the new expression cassette fused to the gusA reporter gene was analyzed in chrysanthemum and tobacco, and compared with that of the commonly used CaMV 35S promoter. Two variants of the equistatin gene: EI-wild type (AF184891) and an optimized version for expression in the potato EIM gene (AY166597) were fused to *rbcS1* and expressed in potato. In addition the expression level of the single domain of potato cystatin driven by *rbcS1* was analyzed in transgenic tomato plants. The results demonstrate a successful strategy to engineer nuclear expression levels of up to 10% of total soluble protein in transgenic plants.

MATERIALS AND METHODS

Cloning of the rbcS1 gene

RNA was isolated from green leaves of chrysanthemum (*Chrysanthemum morifolium* Ramat. cv. 1581; (Annadana et al., 2002a)) according to the hot-phenol method (Verwoerd et al., 1989). mRNA was subsequently purified from total RNA using the Quickmicroprep-mRNA purification kit (Amersham Pharmacia Biotech, Uppsala, Sweden) and cDNA was synthesized using the Great Lengths cDNA synthesis kit (Clontech, Palo Alto, Calif., USA). The cDNAs were cloned into SmaI-digested pBluescript SK+ (Stratagene, La Jolla, Calif., USA) vector, resulting in a library of 1,200 bacterial colonies. The colonies were transferred to ten 96-well plates containing 200 l of LB medium per well and grown overnight. The library (960 clones) was subsequently transferred to nitrocellulose filters using the SRC 96 D Dot blot apparatus (Schleicher & Schuell, Germany) and screened using a ^{32}P-labeled rbcS cDNA probe from potato (Nap et al., 1993). Three clones reacted positively, two of them (7C10 and 7C11) were identical and the third one (7A11) differed in a stretch of 100 bp of the 3 untranslated region, starting four nucleotides behind the stop codon.

The same genotype chrysanthemum (1581) was used to construct a genomic library in bacteriophage lambda FIX II (Stratagene) yielding a titer of 3×10^6 primary phage-forming units (pfu). One million clones were plated and screened using a ^{32}P-labeled probe of 7C10 *rbcS* cDNA. The DNA of five positive phages was then purified and digested with several different restriction enzymes. A digestion using HindIII restriction enzyme yielded, in one of the clones, a unique hybridising band of 2.6-kb. After subcloning into pBluescript vector the above clone (2H2) was found to contain the *rbcS1* gene fully identical to the 7C10 cDNA, together with a 1-kb promoter and terminator region (accession number AY163904).

Sequence analysis of the chrysanthemum rbcS1 promoter

The location and distribution of cis-regulatory sequence elements in the chrysanthemum *rbcS1* promoter region (1,004 bp) were analysed by a signal scan search in the PLACE database (http://www.dna.affrc.go.jp/htdocs/PLACE/signalscan.html). The closest homologues to the chrysanthemum *rbcS1* promoter were identified by a homology-based search in the PLACE database. The identified *rbcS* homologous fragments were aligned to the chrysanthemum gene using the software program MegAlign 4.05 (DNASTAR Inc.) (Wilbur and Lipman, 1983), and subsequently manually improved.

Construction of the rbcS1 plant expression cassette

The chrysanthemum *rbcS1* gene was used to construct an expression cassette for heterologous gene expression. A promoter and terminator fragment of 1 kb were present on the 2.6-kb HindIII fragment and these were amplified with primer pairs P1 (AATACGACTCACTATAG) and P2 (CCCCCCATGGTTTCTAGATACTTAGGAGGAATGGAC), and with P3 (GGGGCCATGGGAGCTCAGATCTCATAAGCCCGATGGCTACTA) and P4 (CCCCAAGCTTGATGAATTCCA AGAAGAAGGGGAAAGAGGC). The fragments were digested with HindIII-NcoI and NcoI-EcoRI, respectively, and cloned with a three-point ligation into the shuttle vector pUCAP (van Engelen et al., 1995) to generate pUCRBC. Thus, a multiple cloning site (XbaI, NcoI, SacI and BglII) was created between the promoter and terminator for versatile cloning of genes of interest (Fig. 1).

```
                    Start ---//-----RubiscoSS          Stop
rbcS gene      AGTATCTAAA----ATGGCC...........CATAAGCCCGATGGCTACTAAGT…
Expr.cassette  AGTATCTAGAAACCATGGGAGCTCAGATCTCATAAGCCCGATGGCTACTAAGT…
                   XbaI     NcoI  SacI  BglII
```

Figure 1. The original sequence at the translational start and stop of the *rbcS1* compared to the multiple cloning site of the newly prepared *rbcS1* expression cassette. The *rbcS1* start and stop codons are indicated in bold. The shuttle vector pUCAP was chosen for easy subcloning of the expression cassette into the binary pBINPLUS vector (van Engelen et al., 1995).

Preparation of the plant transformation constructs

The gusA reporter gene from pMOGEN410 (Jong et al., 1994) was excised and inserted into XbaI-SacI digested pUC-RBC to generate pUC-RBC-GUS. The expression cassette PrbcS-gusA-TrbcS was subcloned into EcoRI-HindIII digested pBINPLUS to generate pRBC-GUS. An existing construct 35S-GUS pLM5 (Mlynarova et al., 1994) in a binary vector was used as a control. Potato cystatin (PC) was subcloned by PCR amplification of the gene described by Annadana et al. (2003) using the primers GGGGGGATCCATGGCAATCGTAGGGGGCATTATCAATGTTCC and GGGGGGGATCCCTACTTTGTAGCATCACCAACAAGTTTAAATTCTTG followed by digestion with NcoI and BamHI and cloning into the NcoI/BglII sites of pUCRBC. The resulting vector pUCRBC-PC was digested with EcoRI/HindIII and the expression cassette was subcloned into pBINPLUS yielding pRBC-PC. The coding regions of the EI (AF184891) and EIM genes (AY166597) (Outchkourov et al., 2002) were amplified with primer pairs EI1 (AAAAACCATGGCTCTTAGCCAAAACC) and EI2 (AAAAAAGATCTTTAGCATGTGGGGCGTTTAAATC) for EI, and EIM1 (AAAAACCATGGCTCTTAGCCAGA ACC) and EIM2 (AAAAAAGATCTTTAGCATGTGGGACGTTTGAATC) for EIM. The PCR products were digested using NcoI and BglII restriction enzymes and cloned into a pUCRBC vector digested with the same enzymes. The expression cassettes containing the *rbcS1* promoter, EI or EIM coding part and the *rbcS1* terminator were subcloned into a HindIII site of pBluescript SK+ vector, and then excised from the vector using SalI and SmaI. The pBin19-based binary vector pLM9 (Mlynarova et al., 1994) was digested with BamHI, and treated with Klenow polymerase to create blunt ends. After subsequent digestion with SalI the old expression cassette was removed. The PrbcS-EI-TrbcS, PrbcS-EIM-TrbcS cassettes were subsequently inserted in the above prepared pLM9 vector. All constructs were electroporated into Agrobacterium tumefaciens (AGLO) (Lazo et al., 1991) using a Gene Pulser electroporator (Bio-Rad).

Plant transformation and analysis

Chrysanthemum (*C. morifolium* Ramat.) cv. 1581 (Annadana et al., 2002a) was transformed as described (Seiichi et al., 1995). The tobacco (*Nicotiana tabacum* L.) SR1 transformation was conducted as described in (Jongsma et al., 1995). Tomato (*Lycopersicon esculentum* Mill.) T201 was transformed according to (Pozueta Romero et al., 2001). Potato (*Solanum tuberosum* L.) cv. Line V plants were transformed according to (Visser et al., 1989). Protein concentrations were determined according to the method of (Bradford, 1976) using bovine serum albumin as the reference protein. The gusA enzymatic activity measurements were conducted as previously described (Mlynarova et al., 1994). In order to correlate the gusA enzyme activity with the amount of enzyme, a purified -glucuronidase (GUS; Sigma) was used as quantitative reference. The amount of enzyme activity from the purified GUS was measured in the presence of the corresponding plant extract (tobacco and chrysanthemum) derived from a non-transformed plant. There was no significant effect of the plant extract on the enzyme activity. The experimentally determined equation $Y=0.0066.X^{1.111}$ was found to correlate the amount of gusA protein with the gusA enzyme activity. Y is the percentage of total soluble protein and X is the gusA activity {nmol 4-methylumbelliferone (mg protein^{-1}) min^{-1}}.

The expression levels of the tomato plants expressing potato cystatin were determined by immunological measurements. Purified potato cystatin as a quantitative protein reference was obtained from the study of (Annadana et al., 2003). Fully expanded young leaves were collected and ground in liquid nitrogen to a fine powder. The powder was resuspended in 300 µl extraction buffer {100 mM TrisHCl (pH 7.6), 25 mM Na-diethyldithiocarbamate, 50 mM EDTA, and 10% polyvinylpolypyrrolidone (PVPP)} and this crude extract was twice centrifuged for 5 min at 14,000 rpm (Eppendorf 5417R centrifuge) and 4 °C. Each time the supernatant was transferred into a new tube. Equal amounts of proteins were loaded on 8-16% gradient SDS-PAGE pre-cast resolving gels (BioRad). The gels were run according to the manufacturer's instructions and stained with Coomassie Brilliant Blue R250 or blotted onto a nitrocellulose membrane using the mini-Protean II electrotransfer system (BioRad). Subsequently, the membranes were blocked in TTBS {10 mM TrisHCl (pH 7.5), 150 mM NaCl, 0.1% Tween 20} containing 2% non-fat milk powder for 1 h and then incubated with rabbit anti-cystatin antibodies (Eurogentec, Seraing, Belgium). The blots were washed and incubated with goat anti-rabbit IgG conjugated with horseradish peroxidase (Jackson ImmunoResearch, West Grove, Pa., USA). The membranes were visualized with Lumi-Light Western blotting substrate and scanned in a Lumi-Imager F1 under the control of Lumi-Analist software (Roche). Two protein extraction protocols were used for quantification of the level of equistatin in the rbcS1-EI and rbcS1-EIM potato plants. One protocol was the same as that described for the tomato plants using liquid nitrogen for grinding, and the other used a freezethaw procedure. Briefly, leaf discs were punched into separate wells of a 96-well microtiter plate and to each well 200 µl of protein extraction buffer {80 mM TrisHCl (pH 7.6), containing 25 mM diethyldithiocarbamate, and 50 mM Na_2EDTA} was added. The samples were then frozen three times to 20 °C and thawed to room temperature. The extract was pipetted to a new microtiter plate, separating it sufficiently from the insoluble leaf material. Both methods were found to extract equal amounts of total soluble protein per leaf mass. For dot-blot analysis 2.5 µg of protein extract obtained by the freezethaw extraction procedure was transferred to Trans Blot (BioRad) nitrocellulose membranes using the SRC 96 D Dot blot apparatus (Schleicher & Schuell). Membranes were blocked and proteins visualized as described for Western blotting of the tomato plants, except that a primary rabbit anti-EI antibody was used (Eurogentec). EIM protein, produced as in Outchkourov et al. (2002), was used to construct a reference curve.

RESULTS AND DISCUSSION

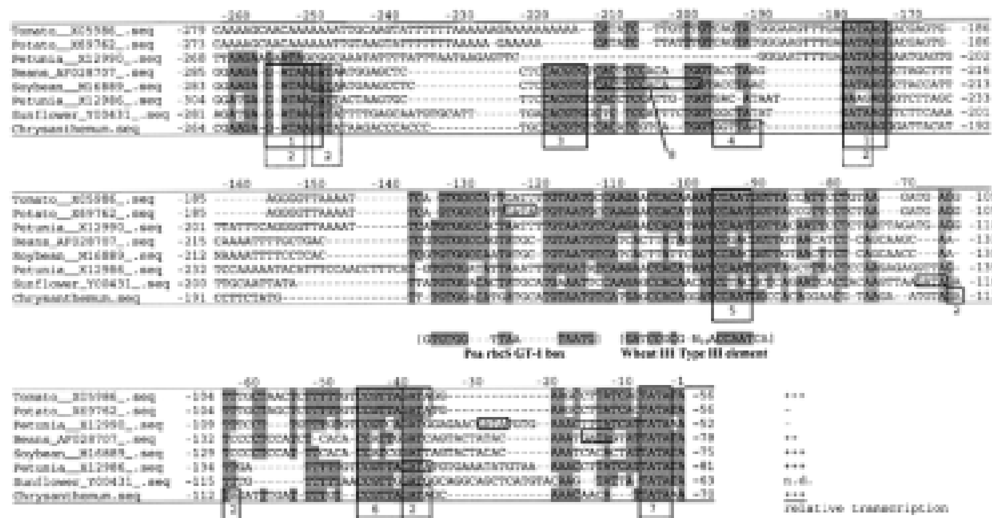
Isolation of the chrysanthemum rbcS1 gene and comparative sequence analysis

The aim of this study was to obtain regulatory gene elements that provide strong expression in the green parts of plant species like chrysanthemum. For this purpose, the cDNA clone 7C10, encoding an abundant *rbcS* transcript was first isolated from a small leaf-specific cDNA library from chrysanthemum using an rbcS probe from potato (Nap et al., 1993). The 7C10 cDNA clone was subsequently used as a probe to select positive plaques from a chrysanthemum genomic phage lambda library. A phage containing the entire *rbcS1* gene with 1-kb up- and downstream regions (2.6 kb

in total) on a HindIII restriction fragment was identified, and the fragment was subcloned and sequenced (AY163904). The gene contained two introns in the same positions as most dicot *rbcS* genes. Sequence analysis of the highly conserved coding region revealed that the chrysanthemum *rbcS1* gene was most similar to *rbcS* genes from the Asteraceae, with alignment bit scores of 326-315, compared with 294-265 for the Fabaceae, 292-288 for the Rosaceae, and 286-265 for the Solanaceae. This was in accordance with a previous analysis indicating that the *rbcS* members from the Asteracea are closely related to members from the Solanacea and Fabaceae (Meagher et al., 1989).

The high degree of divergence of promoter and terminator regions among *rbcS* genes makes a homology alignment of the full promoter and terminator sequences not very informative. However, nucleotide BLAST analysis of the *rbcS1* promoter region -264 to -70 resulted in several significant hits to related *rbcS* promoters of three plant families: Asteraceae (sunflower: Y00431; (Waksman et al., 1987)), Solanaceae {tomato: X05986 (rbc3c) (Sugita et al., 1987); petunia: X12990 (SSU112) and X12986 (SSU301) (Dean et al., 1987); potato: X69762 (rbcS2c) (Fritz et al., 1993)} and Fabaceae {beans: AF028707 (rbcS2) (Knight and Jenkins, 1992); soybean: M16889 (SRS4) (Grandbastien et al., 1986)}. Figure 2 provides the promoter alignment of these different plant species. Previous sequence comparisons of rbcS promoters have only used the rbcS promoters from a single or closely related species (Dean et al., 1985; Fluhr et al., 1986; Fritz et al., 1993; Grandbastien et al., 1986). The prepared alignment in this study reveals structural elements of the *rbcS* promoter that have been well preserved in three different plant families during evolution. Interestingly, five out of seven homologous promoters for which there is available data are abundantly expressed *rbcS* members in their endogenous host (tomato: X05986, petunia: X12986; beans: AF028707; soybean: M16889 and chrysanthemum) (Fig. 2). It is not clear why in the provided alignment the potato (GenBank Acc. No. X69762) and petunia (GenBank Acc. No. X12990) genes are not abundantly expressed, when for example the gene from potato is almost identical to that from tomato. However, the promoter itself is only one of several factors involved in the expression differences: other parts of the rbcS sequences, especially the terminator and upstream region, may also play a role (Dean et al., 1989a; Dean et al., 1989b). Tentatively, we conclude, therefore, that the five genes at the bottom of the alignment with the strongest cross-homology and conservation of sequence motifs are correlated with the highest abundance of rbcS transcripts for several different plant species. This provides a possible link from structural to functional organization of the above promoters. Potentially, the regulatory elements of these genes could be suitable to engineer high levels of expression in the green leaves of plants. A number of conserved promoter elements within these genes were found to represent plant cis-acting regulatory elements. The functions of these elements are listed in the table of Fig. 2. Starting from the distal region of about 250, an I-box element overlapping with a GATA-box was found. Common to both these elements is their role in promoting strong transcriptional activity. Next, a G-box, known to participate in several regulatory processes, including light responsiveness, was identified (Menkens et al., 1995). Downstream, a second I-box was found, this time without the overlapping GATA-motif (overall, the 1-kb promoter contained 3 I-boxes and 12 GATA-motifs). Further downstream, in the 140- to 80-bp promoter region, the highest degree of sequence homology was observed between the different promoters. No complete

identity to known cis-acting regulatory elements was found except for the CCAAT-motif, for which, in contrast to animals, no clear biological function has been defined in plants (Lekstrom-Himes and Xanthopoulos, 1998). Based on a lower score similarity the conserved sequence may represent two different motifs. The first 140- to 110-bp region is similar to box II from the pea rbcS promoter (Kuhlemeier et al., 1988)(Fig. 2). The matching part of the pea rbcS box-II sequence {5-TGTGG$(N)_{12}$TAATG} is conserved in all genes of the alignment. This element is also similar to an SV40 core enhancer TGTGGAA, and in both plants and mammals the two GG are critical for binding of transcriptional factors and subsequent expression (Green et al., 1987; Herr and Clarke, 1986). The second 110- to 85-bp region is similar to the so-called Type-III element, which is specific for histone promoters and responsible for S-phase-specific transcriptional activation (Taoka et al., 1999). Both of the above elements were shown to be important for high-level expression (Kuhlemeier et al., 1988; Taoka et al., 1999). Further downstream, a conservative binding site for a Myb transcriptional factor is present with possible involvement in dehydration stress and flavonoid biosynthesis. On the whole 1-kb rbcS1 promoter sequence, two more potential Myb-sites were found, indicating the possible involvement of Myb-related transcriptional factors in the regulations of the rbcS1 gene.



N	Cis-acting element	Sequence	Function	Reference
1	I-box	GATAAG	Important for high promoter activity. Binding site for LeMYB1,a myb like protein.	(Gilmartin et al. 1990)
2	GATA-box	GATA	Required for high levels, light regulated and tissue specific activity	(Giuliano et al. 1988; Donald and Cashmore, 1990; Rose et al. 1999)
3	G-box	CACGTG	Role in the high promoter activity and responsiveness to light, anaerobiosis, p-coumaric acid, abscisic acid, ethylene and methyl jasmonate.	(Menkens et al. 1995)
4	GT-1 core	GGTTAA	Important for the high level of expression of the pea rbcS-3A gene. Binding site for GT-1 protein.	(Green et al. 1987)
5	CCAAT motive	CCAAT	Common sequence found in the 5'-non coding part of eucaryotic genes.	
6	MYB	CNGTTA	Binding site for plant MYB transcriptional factors involved in dehydratation stress and flavonoid biosynthesis.	(Urao et al. 1993; Abe et al. 1997)
7	TATA-box	TATA	Binding site for mRNA polymerase 2	
8	SURE	CACTTCCACATG	Sucrose responsive element	([illegible] and [illegible], 1997)

Figure 2. A. DNA sequence relationship of the new *rbcS1* promoter from Chrysanthemum to other homologous promoters in the region –300 to 0. Shaded, bold lettering indicate at least five identical sequences. The sequences that represent cis-acting regulatory elements are shown in boxes and are numbered. The relative abundance of the rbcS transcript within the genome of origin is indicated: +++ most abundant, ++ second, - low, n. d. no data. **B.** The table summarises the number and function of cis-acting elements found in the *rbcS1* promoter

Comparison of the rbcS1 promoter with the 35S promoter in tobacco and chrysanthemum

The promoter constructs (35S and rbcS1) fused to the gusA reporter gene were introduced into chrysanthemum (*C. morifolium* cv. 1581) and tobacco (*N. tabacum* cv. SR1). The relative expression strength of both promoters was quantified by measuring the gusA enzyme activity of leaf extracts. Young leaves of greenhouse-grown plants were collected during different periods of the year (July, September and October) and the average enzyme activities of the three gusA measurements are shown in Fig. 3. In tobacco, comparison of the mean gusA activity of the *rbcS1* plant population (21 independent transgenic plants) with the 35S population (18 independent transgenic plants) shows that the *rbcS1* construct results in 8-fold higher levels than the 35S promoter (2.7% vs. 0.34% of total soluble protein). In chrysanthemum, the 35S promoter did not yield detectable activities and the *rbcS1* promoter showed 14-fold lower mean gusA activities than in tobacco (0.34%). However, these values are still similar (2 times less) to those obtained with the 35S promoter in tobacco, so that the promoter can still be considered to be strong also in chrysanthemum.

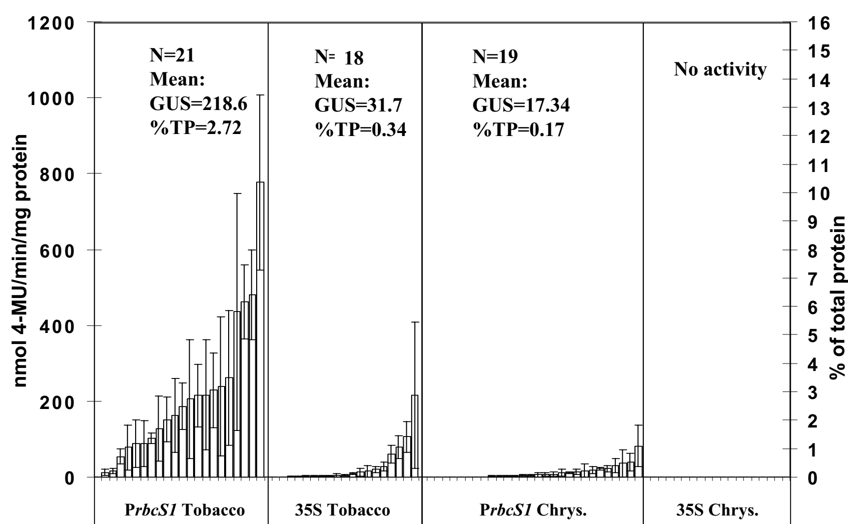


Figure 3. Average gusA activities of three separe measurements (July, September and October) for tobacco and chrysanthemum plants using 35S and *rbcS1* promoters. "N" indicates the number of independently generated plant lines. Bars indicate the standard error. Mean *gusA* activities for all plant populations are numbered on the top of the plant groups. The gusA activities were calculated to percentage of total soluble protein using the experimentally determined equation using purified β–glucoronidase, $Y=0.0066.X^{1.111}$, where "Y" is the percentage of total soluble protein and "X" is the gusA activity (nmol 4-MU/min/mg protein).

The fact that in chrysanthemum the 35S promoter yielded no gusA activity was previously described by (Annadana et al., 2002a). The gusA quantification assay there was adapted in order to detect the low levels of activity. The possible reasons for the low expression of the 35S promoter, as discussed in Annadana et al. (2002), may be

attributable to genotype-linked factors. There, the LhcaSt.1 promoter from potato also expressed GUS much less in chrysanthemum than in tobacco, suggesting that the lower expression is not specific for the homologous *rbcS1* promoter, but rather a consequence of the genetic background of chrysanthemum.

Expression of equistatin and cystatin in potato and tomato plants

Transgenic potato plants (*S. tuberosum* cv. line V) expressing equistatin constructs EI (wild-type gene, AF184891) and EIM (optimized gene, AY166597) under the control of the *rbcS1* promoter were analyzed for expression by dot-blot analysis. The optimization changed 4.2% of all nucleotides. Codons with an abundance of <10% were replaced with more abundant ones, and a potential mRNA instability motif and all polyadenylation motifs represented by five or more A/T or G/C repeats were removed. Equistatin accumulated up to maximum levels of about 7 % of total soluble protein, and about 40% of the analyzed plants accumulated equistatin at levels of 1% or more (Fig. 4).

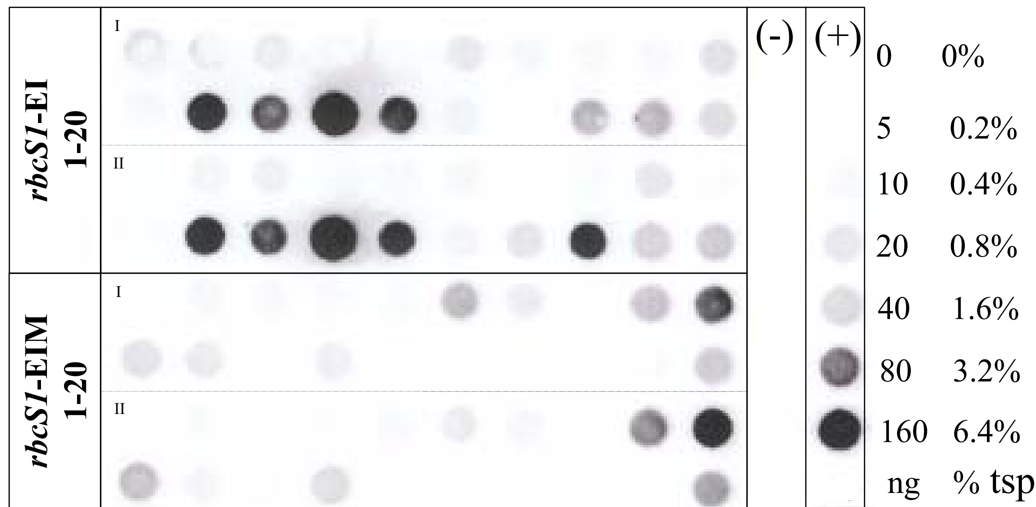


Figure 4. Dot blot immunological analysis of the potato transgenics transformed with *rbcS1*-EI and *rbcS1*-EIM constructs. Twenty plants per construct were analyzed. Two samples per plant, of 2.5 μg of total soluble protein each, were loaded on the membrane. On the left the abbreviated names are given. On the right the numbers indicate the amount of EI protein used as a reference in nanograms (ng) and the corresponding expression levels in % of the total soluble protein (% tsp). (+) positive control of dilution reference protein; (-) Line V potato plants transformed with the pBINPLUS empty vector.

The gene optimization carried out on the equistatin gene appeared not to be relevant as similarly high levels were obtained in both plant populations. Transgenic tomato plants (*L. esculentum*, cv. T201) expressing a single domain of potato multicystatin were analyzed for expression by Western blot analysis. About half of the obtained

transgenic plants accumulated the single domain of potato multicystatin at levels above 1% of total soluble protein. A maximum expression level of 3.3% of total soluble protein was recorded in line C6 as shown in Fig. 5.

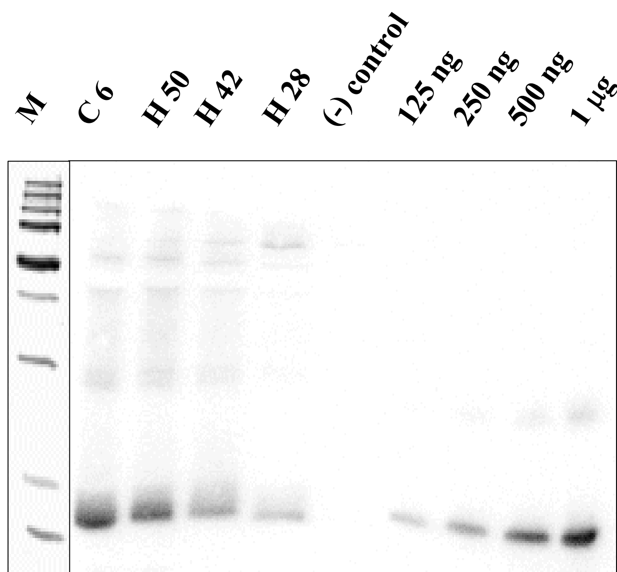


Figure 5. Western blot analysis of the transgenic tomato plants expressing the single domain of potato multicystatin, 30 µg of total leaf soluble protein are loaded per lane.

Comparison with similar promoters

The expression levels of gusA, equistatin and cystatin in tomato, tobacco and potato are the highest to date for a Rubisco small-subunit promoter (Table 1). An expression level of around 5-6% of total soluble protein is the calculated amount of Rubisco small-subunit protein generated by the strongest expressing endogenous Rubisco small-subunit gene. All *rbc*S genes together are responsible for 10-12% of total soluble protein. Thus, the foreign gene expression levels of 3-10% are in line with what can be expected from an rbcS promoter. Surprisingly, a survey of the available literature containing quantitative data reveals that protein expression levels reach at maximum 0.6-1.35% of total soluble protein (Christiansen et al., 2000; Dai et al., 2000; Edwards et al., 1991; Wong et al., 1992), about 5-fold lower than expected. Several factors could account for this, such as the absence of introns in some of the foreign genes, the possible negative effect of targeting to a cellular compartment other than the chloroplast, suboptimal codon usage and the relative instability of the proteins expressed. The promoter fragments used in those reports are similar in size or larger and derived from similarly abundantly transcribed *rbcS* genes, so that an explanation in that direction is less likely (Table 1). One difference from the construct described in this report is that all published uses of the *rbcS* regulatory sequences until now have applied only the 5-promoter region and fused it to 3-terminator elements derived from different rbcS genes or genes such as nopaline synthase or CaMV 35S to

terminate transcription (Table 1). The relevance of the terminator for efficient expression of rbcS genes was suggested by the study of (Dean et al., 1989b). It was found in petunia that sequences both 5 and 3 to the rbcS-coding region are important in explaining the differences in transcription among the native genes (Dean et al., 1989a; Dean et al., 1989b). The terminator from the poorly expressed petunia SSU911 gene was shown to decrease the expression of the SSU301 promoter 3-fold, while the terminator of the highly expressed SSU301 could improve expression of SSU911 at least 10-fold.

Table 1: Quantitative expression levels achieved with different *rbcS* and 35S promoter constructs in various genetic backgrounds

Host plant	**Gene**	**Promoter**	**Terminator**	**Max**	**Mean**	**Ref.**
		Gus activity (nmol/min/mg protein)*				
Tobacco	Gus	• RbcS1, Chr. • dCaMV35S	• RbcS1,Chr. • Nos	776 (10.7%) 216 (2.6%)	218 (2.72%) 31 (0.34%)	This study
Chrysanthemum	Gus	• RbcS1, Chr. • dCaMV35S	• RbcS1,Chr. • Nos	82 (0.88%) 0	17 (0.17%) 0	This study
Apple	Gus	• RbcS3C, Tom. • RbcSRS, Soy. • dCaMV35S	• CaMV35S • CaMV35S • CaMV35S	35 20 n. d.	11 10 18	(Gittins et al., 2000)
Alfalfa	Gus	• RbcSK-1A, Alf.	• CaMV35S	150	92	(Khoudi et al., 1999)
Rice	Gus	• RbcS type I, Rice • RbcS-m3, maize	• Nos • Nos	19 4	6 4	(Nomura et al., 2000)
		Other proteins (% of total soluble protein)				
Tomato	Cystatin	• RbcS1, Chr.	• RbcS1,Chr.	3.0		This study
Potato	Equistatin	• RbcS1, Chr.	• RbcS1,Chr.	7.0		This study
Potato	Pea lectin	• RbcSSU, Tob, • CaMV35S	• Nos • Nos	0.7 1.2		(Edwards et al., 1991)
Tobacco	CryIA(c)	• RbcAts1A, Ara. • RbcAts1A+TP24, Ara.	• n. d.	0.3 0.8		(Wong et al., 1992)
Tobacco	Endoglucanase	• RbcS-3C Tom.	• T5 and T7 octopine	1.35		(Dai et al., 2000)
Trifolium repens	Seed albumin	• RbcS Ara • RbcS Luc • CaMV35S	• RbcS, tob • RbcS, tob • CaMV35S	0.1 0.01 0.04		(Christiansen et al., 2000)

* GUS enzyme activities for this study were converted into percentage of total soluble protein using purified GUS enzyme spiked into the assay as a reference and are shown in brackets.

Table abbreviations:
Chr, chrysanthemum; Tom., tomato; Alf, alfalfa; Soy, soybean; Ara, Arabidopsis; Luc, Lucerne; Tob., tobacco; n.d., no data.

Although for pea rbcS genes similar switching of 3 terminator sequences did not seem relevant at the level of transcription (Kuhlemeier et al., 1988), the importance at the level of translation was shown in amaranth (Berry et al., 1986). A complementary role for 5- and 3-untranslated mRNA in nuclear transport of mRNA and translation efficiency has also been found (Pesole et al., 2002; Pesole et al., 2001). Hence, the

combination of an abundantly expressed *rbcS* promoter with its own 3-untranslated sequences, as reported here, may explain the improved protein expression levels compared with others.

There are few similar reports on high nuclear gene expression. In the given case, the expression cassette was constructed by only removing the rbsS-coding region from the original gene environment, replacing it with the open reading frames (ORFs) GUS, equistatin and cystatin. The expression levels in this study are similar to the high levels reported for monocot species by (Ku et al., 1999), where a fully intact phosphoenolpyruvate carboxylase gene with a 1.5-kb promoter and a 2.5-kb terminator from maize expressed in transgenic rice plants resulted in levels of 12% of total protein. The levels reported by Ku et al. (1999) are similar to the normal levels found in maize, suggesting that leaving the complete gene intact was sufficient to generate similarly high levels of expression. We cannot exclude alternative explanations for the uncommonly high expression levels, such as simply an unusually strong promoter. Studies in that direction might benefit from the alignment of the chrysanthemum rbcS1 promoter with abundantly expressed rbcS genes from other plant species. The identified conserved promoter elements could serve to help in the design of more powerful promoters. It is expected that the new *rbcS1* expression cassette may prove useful in applications requiring high expression levels of recombinant proteins in plants, such as biopharmaceuticals, or in engineering other traits.

CHAPTER 4

SPECIFIC CYSTEINE PROTEASE INHIBITORS ACT AS DETERRENTS OF WESTERN FLOWER THRIPS *FRANKLINIELLA OCCIDENTALIS* (PERGANDE) IN TRANSGENIC POTATO

Nikolay S. Outchkourov[1], Willem Jan de Kogel[1], Antje Schuurman-de Bruin[1], Magnus Abrahamson[2], Maarten A. Jongsma[1]

[1] Plant Research International, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands

[2] Department of Clinical Chemistry, Institute of Laboratory Medicine, Lund University Hospital, S-221 85 Lund, Sweden

Submited for publication

ABSTRACT

In this study the effects of accumulation of cysteine protease inhibitors on food preferences of adult female Western flower thrips, *Frankliniella occidentalis* (Pergande) were investigated. Representative members of the cystatin and thyropin gene families (stefin A, cystatin C, kininogen domain 3, and equistatin) were expressed in potato *Solanum tuberosum* cv. Impala, Kondor and Line V plants. In choice assays a strong, time- and concentration-dependent deterrence from plants expressing stefin A and equistatin was observed. Cystatin C and kininogen domain 3 were found to be not active. All tested inhibitors were equally or more active than stefin A at inhibiting thrips proteolytic activity, but in contrast to stefin A they were all expressed in potato as partially degraded proteins. Resistance of cysteine protease inhibitors against degradation *in-planta* by endogenous plant proteases is, therefore, relevant in explaining the observed differences in deterrence of thrips. The results demonstrate, that when given a choice Western flower thrips will select plants with low levels of certain cysteine protease inhibitors. The novel ecological implications of the defensive role of plant cysteine protease inhibitors as both deterrents and anti-metabolic proteins are discussed.

INTRODUCTION

Western flower thrips (WFT), *Frankliniella occidentalis* (Pergande) (Thysanoptera: Thripidae) is a highly polyphagous insect (Jensen, 2000) that causes considerable economic losses on many vegetable, fruit, ornamental and plantation crops (Lewis, 1998). Moreover, WFT is the most prevalent pest in greenhouses throughout the world (Parrella, 1995; Shipp et al., 1991). The small size of WFT (1-2 mm), combined with its polyphagous feeding habit (Loomans et al., 1995), resistance against many pesticides (Jensen, 2000) and concealed feeding behaviour in the inner whorls of flowers and buds, make this particular insect extremely difficult to control. It was previously established, that the proteolytic activity of extracts of adult WFT has an optimum at pH 3.5, and is nearly completely inhibited by protease inhibitors specific for cysteine proteinases. The hypothesis that cysteine proteases are predominant in WFT digestive tracts was subsequently supported by the gradual reduction in WFT oviposition rate, when purified potato cystatin and equistatin were fed to adult females in combination with a protein rich pollen diet (Annadana et al., 2002b). This demonstrated the potential protective role of cysteine protease inhibitors in plants, and the heterologous expression of cysteine protease inhibitors was proposed as a novel means to control WFT.

Protease inhibitors (PIs) are proteins that form stoichiometric high affinity complexes with proteases and inhibit their hydrolytic activity. Ryan *et al.* (1978) first proposed that in plants PIs play an important biological role by inhibiting the digestive proteases of potential insect herbivores. Some insect herbivores adapt to protease inhibitors of host plants by the induction of proteases that are insensitive to inhibition or by inactivation of the inhibitor by proteolysis (Jongsma and Bolter, 1997, for review). The effects of protease inhibitors on insect development, survival and

fecundity have been reported in numerous studies. Only a limited number of studies have reported effects of protease inhibitors on insect behaviour, but these studies used only artificial diets (Girard et al., 1998; Pham et al., 2000; Rahbe et al., 2003). Some studies have shown that jasmonate-induced responses in plants, which include the accumulation of protease inhibitors, are associated with changes in behavioural food preferences. However, these changes in food preference have so far not been associated with specific protease inhibitors (Thaler et al., 2001; van Dam et al., 2000).

The aim of this research was to investigate the effect of heterologous cysteine protease inhibitors expressed in susceptible model plants on thrips food preferences. For this purpose we transformed thrips-susceptible potato varieties with five different cysteine proteinase inhibitors that are representative of cystatins and thyropins (kininogen domain3, stefin A, cystatin C, potato cystatin, equistatin). We chose to localise all different cysteine protease inhibitors to the endoplasmic reticulum (ER), because tomato plants that expressed high levels of potato cystatin in the cytosol under the control of the *rbcS1* promoter showed an undesirable early senescent phenotype (Outchkourov et al., 2003a). In contrast, the abundant accumulation of equistatin in the endoplasmatic reticulum (ER) of potato leaf cells did not result in a deviating phenotype (Outchkourov et al., 2003a; Outchkourov et al., 2003b). Here, the food preferences of WFT were assessed in choice assays. The results demonstrate that two of the five cysteine protease inhibitors expressed in transgenic potato plants are highly deterrent to thrips.

EXPERIMENTAL PROCEDURES

Oligonucleotides

EIM-SP1 5'-AAAAAAACATGTCTCTTAGCCAGAACCAGG
EIM-SP2 5'-GTTAGACTTGGATCCATGGGACTAGCTTCAGTTGAAGTGATAG
EIM-MP1 5'-TCCCATGGATCCAAGTCTAACGAAATGCCAACAG
EIM-MP2 5'-AGATCTGAGCTCTTAAAGTTCGTCCTTTTGATCACCTCCACCTCCGCATGTGGGACGTTTGAATC
PC1 5'-CCCCCCCCATGGATCCAGCAATCGTAGGAGGCCTTG
PC2 5'-GGCCCCGGAGCTCCTATTGATCACCTCCACCTCCTGTACTATCATCACCAAC
709 5'-TAAATGGCGCCCATGGATCCCATGATACCTGGAGGCTTATCT
710 5'-TAAATGAGCTCCTATTGATCACCTCCACCTCCAAAGCCCGTCAGCTCGTCA
711 5'-TAAATGGCGCCCATGGATCCCTCTTCTCCGGGTAAACCGC
712 5'-TAAATGAGCTCCTATTGATCACCTCCACCTCCGGCGTCCTGACAGGTGGA
713 5'-TAAATGGCGCCCATGGATCCCGGGAAGGATTTTGTACAACCA
714 5'-TAAATGAGCTCCTATTGATCACCTCCACCTCCTGAGATCATTCCCAGTGGTTG

Plant material and insects

Potato *Solanum tuberosum* cultivars Impala and Kondor were used for plant transformation with kininogen domain 3, stefin A, cystatin C, and potato cystatin containing constructs. In addition, potato *Solanum tuberosum* cv. Line V previously transformed with the equistatin genes *EI* (AF184891) and *EIM* (AY166597), both encoding an identical equistatin protein (Outchkourov et al., 2003a) was used in the WFT choice bioassays. A population of western flower thrips (WFT) *Frankliniella occidentalis* was reared on flowering chrysanthemum (*Chrysanthemum morifolium* Ramat.) cv. Sunny Casa in a greenhouse at 25 °C.

Inhibition of thrips proteolytic activity

Recombinant cystatin C and kininogen domain 3 were expressed in *E. coli* and purified as described earlier (Abrahamson et al., 1988; Auerswald et al., 1993). The production of stefin A will be described in detail elsewhere (Romero, Jongsma and Abrahamson, manuscript in preparation). In brief, a cDNA segment encoding human stefin A was cloned by RT-PCR from human liver RNA and was cloned in expression vector pHD389, following the same procedures as used for cystatin D expression (Freije et al., 1993). Equistatin was produced and purified as described by (Outchkourov et al., 2002) and potato cystatin by (Annadana et al., 2002a) The concentration of the inhibitors was quantified by titration against papain using Z-Phe-Arg-pNA (Bachem, Switzerland) as a substrate as described by Rogelj *et al.* (2000). The concentration of papain was determined by titration with E-64 (Sigma). Inhibition of the WFT proteolytic activity in extracts of adult insects was conducted as described by (Annadana et al., 2002b) using FITC-labeled albumin (Sigma).

Preparation of constructs for plant transformation

The pUCRBC-EIM construct (Outchkourov et al., 2003a) was modified to allow subcloning of the four other protease inhibitor genes into an expression cassette with an identical N-terminal signal peptide derived from equistatin and the same C-terminal KDEL sequence for retention in the ER. For this purpose cloning sites were created behind the signal sequence of equistatin and before the KDEL signal (see Figure 1 for a schematic overview).

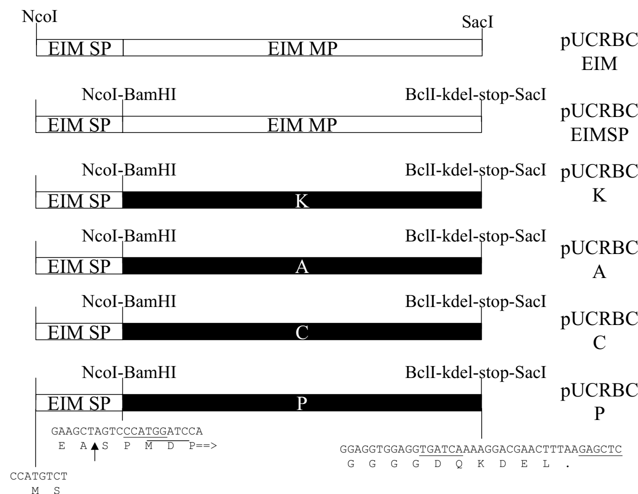


Figure 1. Schematic representation of the constructs used in this study. EIM SP- equistatin signal pep tide. EIM MP- equistatin mature protein. K- kininogen domain 3; A- stefin A; C – cystatin C; P- potato cystatin.

Briefly, a PCR reaction was performed using pUCRBC-EIM as template and primer pairs EIM-SP1/2 and EIM MP1/2, to amplify the DNA sequence encoding the equistatin signal peptide (*SP*) and equistatin mature protein (*MP*) separately. The primers were designed to create or remove restriction sites in the coding part of equistatin. The obtained fragments were digested with *Afl*III-*Bam*HI for *SP*, and *Bam*HI-*Sac*I for *MP*, and cloned with a three point ligation into pUCRBC

(Outchkourov et al., 2003a), that was digested with *Nco*I-*Sac*I. The ligation of the *Afl*III sticky ends into the *Nco*I site of pUCRBC destroyed the existing *Nco*I and AflIII sites. This together with the new restriction sites from the PCR reaction resulted in plasmid pUCRBC-EIM-SP (Fig. 1). The coding part of kininogen domain 3 was amplified from pDR1311 (Auerswald et al., 1993) using primer pair 713/714. Stefin A was amplified from pHD389/CysAmod (Romero, Jongsma and Abrahamson, manuscript in preparation) using primer pair 709/710. Cystatin C was amplified from pCysCmut283-Nco (Mason et al., 1998) using primer pair 711/712. Potato cystatin was amplified from pUCRBC-PC (Outchkourov et al., 2003a) using primers PC1 and PC2. The obtained fragments (kininogen domain 3, stefin A, cystatin C and potato cystatin) were digested with *Nco*I-*Bcl*I and cloned into the pUCRBC-EIM-SP vector, after removing the coding sequence of equistatin mature protein using the same restriction sites, to create pUCRBC-K, pUCRBC-A, pUCRBC-C, pUCRBC-P respectively. These four new expression cassettes containing the *rbcS1* promoter fused to four cysteine protease inhibitors were excised from the pUC based vectors by digestion with *Asc*I-*Eco*RI and inserted into the multiple cloning site of the pBINPLUS vector (van Engelen et al., 1995) to create RBC-K, pRBC-A, pRBC-C, and pRBC-P.

Transformation and regeneration

All pBINPLUS constructs were electroporated into electrocompetent *Agrobacterium tumefaciens* AGL0 (Lazo et al., 1991). Potato plants (*Solanum tuberosum cv.* Impala and Kondor) were transformed according to (Hoekema et al., 1989) using tuber discs as a source of explants.

Quantification and characterization of the expression in transgenic potato plants

Dot blot and western blot immunological detection procedures were carried out as described in (Outchkourov et al., 2002) using rabbit: anti-equistatin, anti-potato cystatin (Eurogentec, Seraing, Belgium), anti-kininogen, anti-stefin A, and anti-cystatin C (Abrahamson et al., 1986) as primary antisera, and anti-rabbit IgG conjugated with horseradish peroxidase (Jackson Immuno Research, USA) as secondary antiserum.

Thrips choice assays

Adult female thrips of mixed ages were collected with an aspirator from the rearing greenhouse. They were starved overnight, only provided with water, in perspex ring cages (Murai, 1990) before experiments. The WFT females were briefly anaesthetised with CO_2 before they were transferred to petri dishes using a fine brush. Dual choice assays were used to test the preference of thrips for leaf discs from potato plants with low versus high PI expression. For this purpose two leaf discs of 21 mm in diameter, punched from mature leaves of similar position, were placed with the abaxial side up on a thin layer of 1% wateragar in a petri dish. In between the leaf discs was a small piece of filter paper (5 x 5 mm) on which 10 starved female thrips were released at 10.00 a.m.. The petri dishes were closed and sealed with parafilm to prevent the thrips

from escaping and placed in a climate chamber (T=22 °C, L16:D8, R.H.=70%). Petri dishes (six per treatment) were randomised to eliminate the effect of light or temperature gradients that could affect choice. The number of thrips on each leaf disc was recorded at different time points after the start of the experiment.

As the repeated measurements were taken from the same petri dishes the data were not independent. To circumvent this dependency, data from the multiple time points measurements were averaged into an early period (0 to 6 hours after the start of the experiment), and late period (ranging from 21 to 26 hours after the start of the experiment). For each of the two periods the null hypothesis H_o: no preference of WFT for leaf discs against alternative hypothesis H_a : preference for leaf with low PI level, was tested using a one-sided Wilcoxon-signed rank test on the absolute differences of average numbers of WFT on leaf discs for the low and high expressing groups (Hollander and Wolfe, 1973).

RESULTS

Inhibition of total thrips proteolytic activities

The inhibitory properties of the purified inhibitors: kininogen domain 3, stefin A, cystatin C, potato cystatin and equistatin were compared to those of the synthetic cysteine protease inhibitor E-64. For this comparison, we used *in-vitro* assays with total WFT homogenate as an enzyme source and FITC-labelled albumin as a substrate. Two concentrations of each inhibitor were tested (10 and 100 nM). At the high concentration (100 nM) most inhibitors efficiently inhibited 90-95% of proteolytic activity in total extracts of WFT (Figure 2).

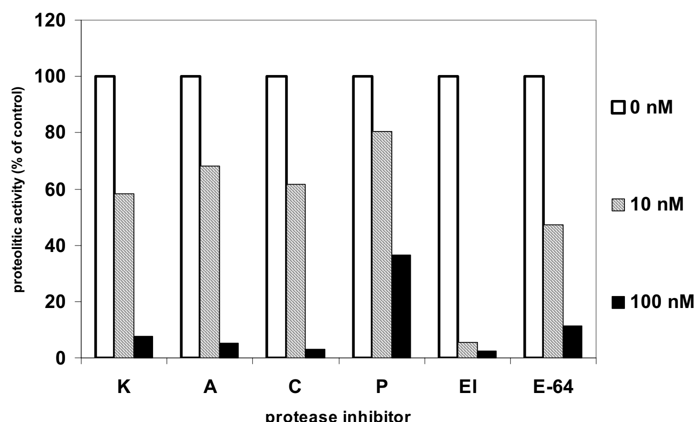


Figure 2. Effects of protease inhibitors on hydrolysis of FITC-labelled albumin by proteases present in whole insect homogenates of *F. occidentalis* adults. The legend provides the concentrations of tested inhibitors. Abbreviations: K- kininogen domain 3; A- stefin A; C – cystatin C; P- potato cystatin; EI-equistatin.

Potato cystatin inhibited only 60% of the activity at this concentration, and, therefore, classifies as a weak inhibitor. In contrast, equistatin was a strong inhibitor as > 90% of total thrips proteolytic activity was inhibited at the lower 10 nM concentration. The other three inhibitors kininogen domain 3, stefin A, and cystatin C were intermediate with around 40% inhibition at 10 nM inhibitor concentration. E-64 inhibited up to

90% of the total proteolytic activity but was less efficient than equistatin. The strong activity of equistatin compared to the other inhibitors may be partly due to the fact that this inhibitor is active against both cysteine and aspartic proteases. However, based on the maximum inhibition values obtained for the cysteine protease inhibitors the contribution of aspartic protease activity is not more than 10%.

Expression of the cysteine protease inhibitors in transgenic potato plants and WFT choice assays on detached leaves

The cysteine protease inhibitors kininogen domain 3, stefin A, cystatin C, and potato cystatin were expressed in potato cv. Kondor and Impala. Previously obtained transgenic plants of potato cv. Line V expressing equistatin were also included in the bioassays (Outchkourov et al., 2003a). All inhibitors were expressed under the control of the rubisco promoter/terminator *RbcS1* from chysanthemum (Outchkourov et al., 2003a), and targeted to the endoplasmatic reticulum. This was achieved by fusing the cDNA fragments encoding the mature proteins at the N-terminus to the equistatin signal sequence, and at the C-terminus to the KDEL signal. The amounts of the expressed proteins were measured by dot blot immunological detection.

Table 1: Quantitative expression levels achieved with *rbcS1* promoter constructs from Chrysanthemum in different potato cultivars.

Gene	Potato cultivar	Expression (% of total soluble protein)	
		Mean	Maximum
Kininogen domain 3	• Impala	0.81 (n=9)*	1.7
	• Kondor	0.31 (n=8)*	0.63
Stefin A	• Impala	0.18 (n=9)*	0.32
	• Kondor	0.28 (n=5)*	0.50
Cystatin C	• Impala	0.40 (n=8)*	1.5
	• Kondor	0.91 (n=9)*	2.7
Potato cystatin	• Impala	n.d.	n.d.
	• Kondor	n.d.	n.d.
Equistatin	• Line V	2.3 (n=22[#])*	7.5

* Between brackets plant numbers analysed.
[#] Transgenic plants obtained from the study of (Outchkourov et al., 2002). The levels of expression were quantified again in this study and differed slightly from those published in (Outchkourov et al., 2002).
n.d.- not determined due to background signal.

It was not possible to quantify the amount of potato cystatin by dot blot analysis, because of the presence of endogenous potato multi-cystatin in potato leaves that interfered with the analysis (see Fig.3 Western blot). The expression levels were strongly dependent on the proteins involved as shown in Table 1. The high expression levels of equistatin in the potato Line V could not be equalled in the Impala and

Kondor cultivars by the other inhibitors kininogen domain 3, stefin A, cystatin C, and potato cystatin.

The stability of the expressed proteins was assessed by Western blot analysis (Fig.3). Specific antibodies were used for each of the proteins. For potato cystatin varying (sample dependent) amounts of endogenous multicystatin inhibitor of 80 kD were detected and distinguishable from the transgenic inhibitor of 10 kD. For kininogen domain 3 and cystatin C a degradation pattern was observed which was identical in both potato cultivars Impala and Kondor. In contrast, stefin A was found as a single band, in both Impala and Kondor. The expression of stefin A, however, was weaker compared to the other inhibitor proteins. Multiple degradation bands appeared for kininogen domain 3 and equistatin, indicating multiple cleavage sites in the polypeptide chains of these inhibitors. Cystatin C appeared as a double band and was probably cleaved *in-planta* most efficiently at only one position.

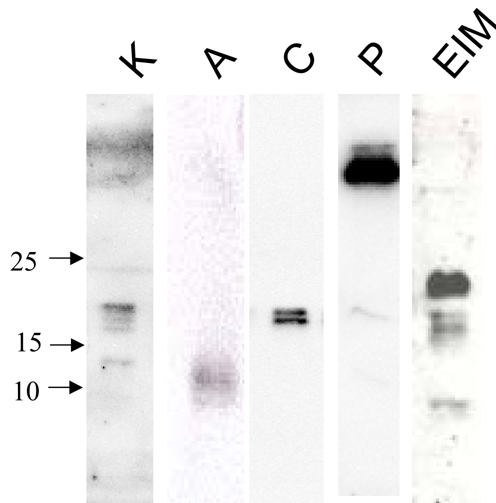


Figure 3. Combined results of different western blot experiments on the transgenic potato plants that express: K- kininogen domain 3; A- stefin A; C - cystatin C; P- potato cystatin; EIM-equistatin in potato cv. Impala. Identical degradation patterns were obtained in potato cv. Kondor (not shown). EIM was expressed in potato cv. Line V. Arrows the approximate position of molecular weight standarts in kilodaltons. The additional lower bands are degradation products.

The preferences of WFT when given a choice between plants expressing cysteine protease inhibitors at low and high levels were monitored *in vitro* using detached leaves. During the time course of 24 hours the ingestion of high concentrations of stefin A (Fig. 4) or equistatin (Fig. 5) led to increasing deterrence of WFT. The preference for the lower expresser tended to be most significant at the end of the 24-26 hour experimental period. The effects on WFT behaviour correlated very well to the expression level differences. When "high" expression level differences of stefin A were tested (Figure 4E: 0.34% TSP is the difference between inhibitor expressed at 0.50% and 0.16% of total soluble protein (TSP)) females responded more rapidly and significantly more strongly to the accumulated inhibitor for all the time points. Choice assays with smaller expression differences of 0.16% to 0.21% TSP stefin A (Figure 4B-D) resulted in significant thrips deterrence only after a prolonged period of feeding.

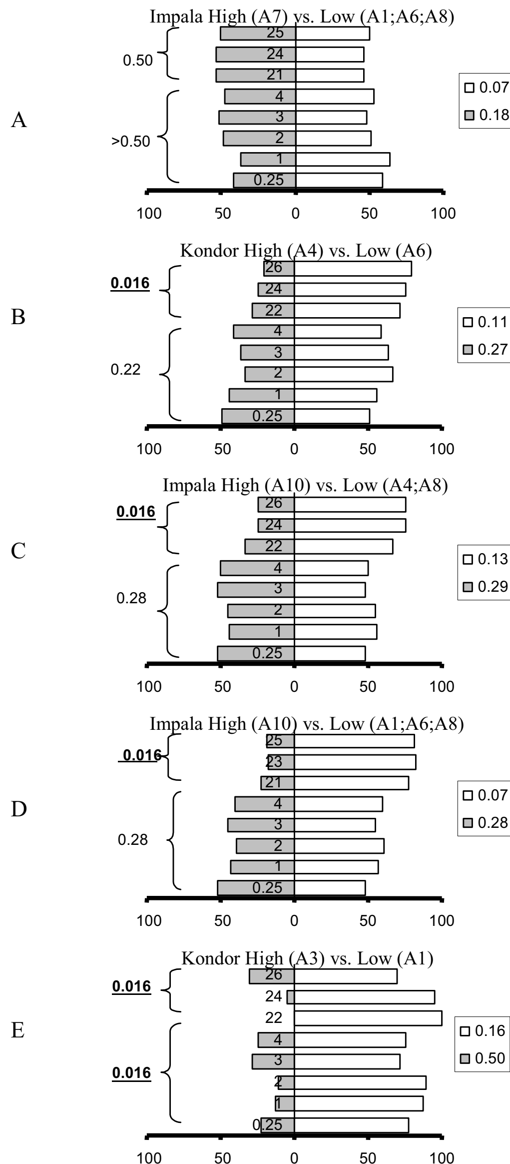


Figure 4. Results of five different thrips dual choice assays (A-E) on transgenic lines expressing stefin A. The plant lines and potato varieties are indicated on the top of each graph. On the X-axis the percentage of thrips on the low (white) and high (grey) expressing plants are plotted against time (hours) on the Y-axis. On the right, the levels of expression measured by dot blot analysis as a percentage of total soluble protein are shown. The calculated p-values of the one-sided Wilcoxon signed rank test for the periods early (0 to 6 hours after the start of the experiment) and late (21 to 26 hours after the start of the experiment) are given on the left of each panel. In bold and underlined are the statistically significant values $p < 0.05$.

No significant thrips food preferences were measured when the expression level differences of stefin A were 0.11% TSP or less (Figure 6).

Similar correlations, but at much higher expression level differences, were observed for equistatin (Fig. 5). Equistatin appeared to require approximately 20-fold higher levels of accumulated inhibitor (7.3% TSP expression level difference) compared to stefin A (0.34% TSP difference) to induce a highly significant effect on WFT. The graph in Figure 6A shows that for equistatin the threshold level of detection was about 1% TSP difference. Below this threshold, WFT females did not show a preference for the discs with low expression.

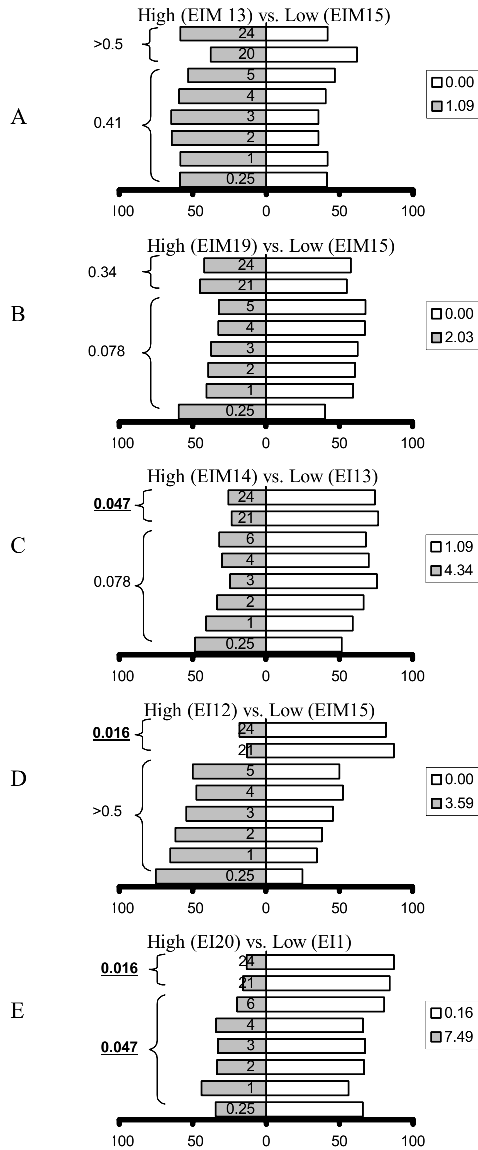


Figure 5. Results of five different thrips dual choice assays (A-E) on transgenic lines expressing the *EI* (AF184891) and *EIM* (AY166597) genes both of which encode an identical equistatin protein. For details refer to the legend of Figure 4.

The accumulated proteins kininogen domain 3 and cystatin C in potato did not show any effects in WFT choice assays in the tested concentration range with expression level differences of up to 1.6% TSP for cystatin C and 1.4% TSP for kininogen domain 3 (Figure 6B).

Thus, in transgenic potato plants stefin A is the most deterrent inhibitor for thrips females, followed by equistatin, while kininogen domain 3 and cystatin C do not deter WFT.

DISCUSSION

Western flower thrips is a serious polyphagous pest both in the field and in greenhouses on many different crops and has developed resistance to many synthetic

insecticides (Jensen, 2000). Identification of compounds or proteins that, in addition to being toxic/antimetabolic, will change the behaviour of an adult insect by causing it to leave the plant in search for better substrates are of great applied and scientific interest. Such deterrent or repellent effects potentially reduce the likelihood of pests developing resistance. Methyl jasmonate (MJ) induced defence responses in plants are known to provoke changes in food preferences of insect herbivores (Thaler et al., 2001; van Dam et al., 2000). However, the complexity of the induced responses such as digestive inhibitors (e.g. protease inhibitors), enzymes (e.g. polyphenol oxidases), and a range of secondary metabolites (e.g. nicotine), makes the precise assignment of plant compounds involved in provoking the behavioural responses problematic.

In this study we demonstrate that transgenic plants overexpressing cysteine protease inhibitors are strongly deterrent to Western flower thrips (WFT), and can rapidly affect behaviour less than an hour after feeding. Very strong deterrent effects were found for stefin A, followed by equistatin at 10-20 fold higher expression levels. Kininogen domain 3 and cystatin C did not yield any deterrent effect.

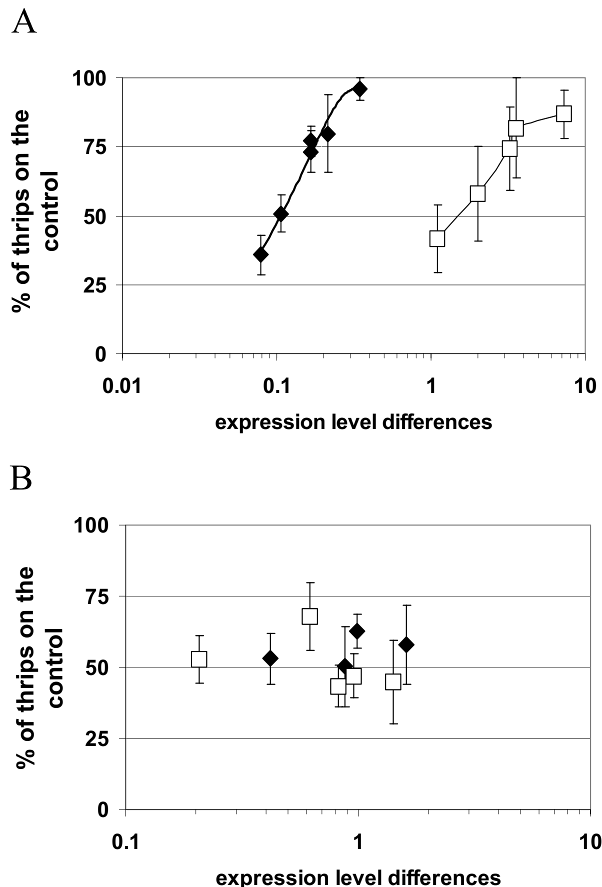


Figure 6. Results summary of all conducted thrips choice assays. Choice data obtained after 24 hours were used. They were correlated to the expression level differences between tested plants (%TSP). Legends: Panel **A.** (◆)- stefin A, (□)- equistatin; Panel **B.** (◆)- cystatin C; (□)- kininogen domain 3.

This difference with the other two inhibitors did not correlate to the *in-vitro* inhibition of WFT proteases. *In-vitro* equistatin was the most potent protease inhibitor (>90% inhibition at 10 nM), followed by kininogen domain 3, stefin A, cystatin C (> 90% at 100nM), and potato cystatin (~60% at 100nM).

A straightforward explanation for the lack of correlation between the potency of the inhibitor and the degree of deterrence is not possible with the present data. Due to the small size of the insects, the *in-vitro* studies were carried out with extracts of complete adult insects and not with dissected guts. Insect haemolymph is known to contain inhibitors of gut proteases and the surplus proteolytic activity in the extract may, therefore, not correctly represent the actual proteolytic activity in the gut. The ranking of the effectiveness of the inhibitors on the basis of the *in-vitro* experiments may, therefore, be wrong. Also, it is not known whether WFT will adapt with the induction of a novel complement of insensitive proteases in response to these inhibitors. It may be that kininogen domain 3 and cystatin C are not effective at inhibiting those enzymes. Finally, it may be that the high level of expression required for deterrent effects of equistatin, is related to the degradation of the protein *in planta* or *in insecta*. All expressed inhibitors, except stefin A and potato cystatin showed distinct bands of degradation when the transgenic plants were analysed by western blot. It was also shown that the degradation of PIs in insect guts could potentially affect the efficacy of PIs (Jongsma and Bolter, 1997; Michaud et al., 1995b; Michaud et al., 1996). When the integrity of the expressed inhibitor is already affected *in-planta,* higher expression levels may be necessary to obtain a deterrent effect, like in the case of equistatin, or the anti-WFT effects may even be lost like possibly in the case of cystatin C and kininogen domain 3. A quantitative comparison in feeding studies with purified proteins will help to resolve whether *in-planta* or *in-insecta* degradation of the inhibitors plays a role in determining the deterrence for thrips and whether the insects induce inhibitor-insensitive enzymes.

The effects of cysteine protease inhibitors on WFT food preferences may have ecological implications. The active search of WFT adults for an optimal food substrate devoid of high levels of cysteine protease inhibitors makes sense considering that earlier experiments demonstrated a high fitness cost (Annadana et al., 2002b). Exposure to potato cystatin and equistatin reduced fecundity of thrips in a few days by at least 50%. Flower organs are known to be very rich in protease inhibitors (Atkinson et al., 1993; van Dam et al., 2001). This high constitutive level may help to deter insects, which specifically attack the reproductive organs such as WFT. The availability of *Nicotiana alata* populations which differ in their ability to express protease inhibitors may enable studies that quantify the effects of insect deterrence on seed yield (Glawe et al., 2003a).

The mechanism which senses protease inhibitors, and affects behaviour of WFT, will most likely utilise a completely different signalling pathway compared to the well known olfactory and gustatory signalling pathways. Effects of protease inhibitors on behaviour have also been observed with other inhibitor classes and other organisms. In the pea aphid *Acyrthosiphon pisum* host plant derived chymotrypsin inhibitors were found to induce a behaviour of restlessness (Rahbe et al., 2003). In an experiment with human subjects, that were given a soup containing a high dose of trypsin inhibitors prior to their lunch, the calory intake during the rest of the lunch was strongly reduced (Hill et al., 1990). In mammals the inhibitors are known to induce the release of cholecystokinin, which stimulates pancreatic secretions and provokes a feeling of satiety (Liddle, 1995). In insects other hormones have been found to modulate protease secretion (Borovsky et al., 1990; Tortiglione et al., 2002). We expect similar hormones to determine the behaviour of WFT on inhibitor containing substrates.

Transgenic crop protection strategies favour a combination of indirect deterrent and direct toxic effects on both adults and larvae. In that way, the probability of resistance development in insect populations is reduced compared to a situation where, for example, the adult moth is unable to identify a Bt toxin expressing plant as a poor substrate for its offspring. Non-choice greenhouse experiments on transgenic potato plants expressing stefin A and a multidomain combination of different cysteine protease inhibitors have been carried out separately (Outchkourov *et al.,* submitted). They confirm the results presented here and of Annadana *et al.* (2002) with purified proteins, that cysteine protease inhibitors will strongly affect population development by a combination of deterrence and a reduction of WFT fecundity. Thus, cysteine protease inhibitor genes appear to be interesting novel candidates to control WFT in field and greenhouse crops like cotton and ornamentals suffering from this pest.

CHAPTER 5

EXPRESSION OF SEA ANEMONE EQUISTATIN IN POTATO: EFFECTS OF PLANT PROTEASES ON HETEROLOGOUS PROTEIN PRODUCTION

Nikolay S. Outchkourov[1*], Boris Rogelj[2*], Borut Strukelj[2] & Maarten A. Jongsma[1]

[1] Plant Research International, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands

[2] Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

* Both authors equally contributed to the provided research

Plant Physiology (2003) 133: 379-390

ABSTRACT

Plants are increasingly used as production platforms of various heterologous proteins, but rapid protein turnover can seriously limit the steady state expression level. Little is known about specific plant proteases involved in this process. In an attempt to obtain potato *Solanum tuberosum* cv. Desirée plants resistant to Colorado potato beetle larvae, the protease inhibitor equistatin was expressed under the control of strong, light-inducible and constitutive promoters, and targeted to the secretory pathway with and without endoplasmic reticulum retention signal. All constructs yielded similar stepwise protein degradation patterns, which considerably reduced the amount of active inhibitor *in-planta* and resulted in insufficient levels for resistance against Colorado potato beetle larvae. Affinity purification of the degradation products and N-terminal sequencing allowed the identification of the amino acid P_1-positions (Asn^{13}, Lys^{56}, Asn^{82} and Arg^{151}) that were cleaved *in-planta*. The proteases involved in the equistatin degradation were characterized with synthetic substrates and inhibitors. Kininogen domain 3 completely inhibited equistatin degradation *in-vitro*. The results indicate that Arg/Lys-specific and legumain-type Asn-specific cysteine proteases seriously impede the functional accumulation of recombinant equistatin *in-planta*. General strategies to improve the resistance to proteases of heterologous proteins in plants are proposed.

INTRODUCTION

The expression of many traits in transgenic plants can be severely hampered by the individual characteristics of foreign genes and proteins not adapted to the specific subcellular environment of the new host. Significant progress has been made in optimizing the rate of transcription and translation in plant cells (Dai et al., 2000; Koziel et al., 1996; Outchkourov et al., 2003a). However, proteolytic degradation of heterologously expressed proteins is still a limiting factor in the accumulation of many foreign proteins in plants (Dolja et al., 1998; Sharp and Doran, 2001a; Sharp and Doran, 2001b; Stevens et al., 2000). A generally adopted approach to increase heterologous protein accumulation levels in plants is to change the compartmentalization of the expressed proteins by targeting to and retention in the endoplasmic reticulum (ER) (Schouten et al., 1996; Wandelt et al., 1992) or chloroplasts (Wong et al., 1992). The targeting to the endomembrane system of plants, however, leads to an unpredictable final sub-cellular location of the expressed protein even when the ER retention signal is added (Vitale and Galili, 2001).

In different plant cellular compartments a wide variety of proteases are involved in the processing and degradation of proteins (Vierstra, 1993; Vierstra, 1996). The levels of these plant proteases are affected by many developmental factors like seed maturation and germination (Bottari et al., 1996) or leaf senescence (Ueda et al., 2000), combined with environmental conditions such as carbon starvation (Brouquisse et al., 1998) or infection and wounding (Guevara et al., 2002). In general subtilisin/kexin-like serine and metallo-proteases in the secretory pathway are known to process pro-proteins(Raharjo Sunu et al., 2001; Viale et al., 1999), cysteine proteases are known to process misfolded proteins (Otsu et al., 1995) or to mobilize

seed storage proteins (Schlereth et al., 2001). The legumain-type cysteine proteases enter the secretory pathway as inactive precursors and are activated in protein bodies (protein storage vacuoles) and the cell wall (Fischer et al., 2000; Hiraiwa et al., 1999; Kinoshita et al., 1999; Muntz and Shutov, 2002). Together with papain-like cysteine proteases they were shown to be involved in the degradation of plant storage proteins (Fischer et al., 2000; Schlereth et al., 2001).They function in processing of storage protein precursors during seed development and storage protein mobilization during seed germination (Hara Nishimura, 1993; Hara Nishimura et al., 1991; Hara Nishimura et al., 1995; Hara Nishimura et al., 1998; Hara Nishimura et al., 1993; Rotari et al., 2001). Legumain type proteases are also known to be up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions (Kinoshita et al., 1999). Several cysteine proteinases, pathogen-related proteins, chitinase and wound-inducible proteinase inhibitors are proposed as putative targets for activation by the action of legumain, Asn-specific proteases(Kinoshita et al., 1999).

The expression of heterologous protease inhibitors in plants has been used as an approach to induce plant resistance to insects (Ryan, 1990). Many host plant-derived inhibitors are unsuitable, however, due to the fact that many insects have evolved resistance (Jongsma and Bolter, 1997). Extensive screening of non-host plant derived inhibitors has resulted in novel candidate proteins which *in-vitro* and *in-vivo* demonstrated promising levels of resistance against various insect pests (Gruden et al., 1998; Harsulkar Abhay et al., 1999). The expression of these proteins *in-planta*, however, brings them into a subcellular environment which may be far from optimal for the stable accumulation of the protein, while the effectiveness of protease inhibitors against insects depends on an expression level of around 1% of total soluble protein(Jongsma and Bolter, 1997; Ryan, 1990).

Equistatin is a protease inhibitor from the sea anemone *Actinia equina* and consists of three thyroglobulin type-1 domains. The first, N-terminal domain, acts as a cysteine protease inhibitor (K_i for papain of 0.57 ± 0.04 nM), the second as an aspartic protease inhibitor (K_i for cathepsin D of 0.3 ± 0.15 nM), and the function of the third domain is not yet known (Lenarcic and Turk, 1999; Strukelj et al., 2000). The protein was previously found to be highly active against the gut proteases of Colorado potato beetle and on artificial diets the protein caused complete growth inhibition and mortality among the larvae (Gruden et al., 1998). The coding part of the equistatin gene was modified for maximal expression in potato by site-directed mutagenesis (accession number AY166597) and functional protein was expressed at levels of grams per liter in the yeast *Pichia pastoris* (Outchkourov et al., 2002). In potato leaves accumulation of only partly functional equistatin up to levels of 7% of total soluble protein was achieved (Outchkourov et al., 2003a). For functional expression it is necessary, however, that the expressed protein is properly folded and essentially undegraded.

In this study, equistatin expressed in the ER of potato was found to be cleaved and further degraded by endogenous plant proteases. Due to the light-regulated rubisco promoter this resulted in a cyclic presence of higher and lower levels of equistatin which were differentially degraded. As a result insufficient levels of functional protein were maintained for resistance against Colorado potato beetle. The cleavage sites were determined by sequencing of the peptide fragments. The plant proteases responsible for equistatin degradation were partly characterized with specific

substrates and inhibitors. The results may have significant implications in engineering the stability of heterologously expressed proteins in plants.

MATERIALS AND METODS

Plant material and insects

Potato plants (*Solanum tuberosum*) of the cultivar 'Desirée' were used in all experiments. Transgenic potato plants were grown in the greenhouse with supplementary high-pressure sodium light under 16/8 hours light/dark rhythm and temperature regime 21/18 degrees. CPB larvae were obtained from an established colony at the department of Entomology, Wageningen University and Research Center.

Preparation of constructs for plant transformation

Four plant transformation cassettes were prepared to express the equistatin gene (AY166597). The coding part of equistatin was excised out of the plasmid pEIM8 (Rogelj et al. unpublished) with NcoI-BglII restriction enzymes and then inserted into pUCLhca.3.St1 (Nap et al., 1993), pUCRbcS1 (Outchkourov et al., 2003a) and pUCAP35S (van Engelen et al., 1995) to create pUCLhca.3.St1EIM, pUCRbcS1EIM, pUCAP35S. pUCRbcS1/EIM/KDEL was constructed by PCR amplification of the equistatin gene using pUCLhca.3.St1/EIM as the template and the primers Cabd (5'-cttttgtatttaatttattcttg-3') and EIMKDEL (5'-ttttttagatctttaaagttcatccttagcgcatgtgggacgtttgaatc-3') to add the KDEL amino acid sequence to the C-terminal end of equistatin. The obtained fragment was cut with NcoI-BglII restriction enzymes and inserted into the linearised pUCRbcS1 vector. These four expression cassettes containing a promoter, equistatin and a terminator were then excised and inserted into the multiple cloning site of the pBINPLUS vector (van Engelen et al., 1995) to create pLhca.3.St1/EIM, pRbcS1/EIM, pRbcS1/EIMKDEL and p35S/EIM (Fig. 1).

Potato transformation

All pBINPLUS constructs and empty pBINPLUS vector as control were introduced into *A. tumefaciens* AGL0 by electroporation. Potato plants were transformed using the stem segment transformation method (Visser et al., 1989).

Northern blot

Total RNA was isolated from plant tissue according to (Verwoerd et al., 1989). Samples of 10 μg total RNA were denatured and separated on a 1.2% agarose gel as previously described (Kevil et al., 1997). As a positive control, RNA from the equistatin gene synthesised with Riboprobe® in vitro Transcription Systems

(Promega) was used. Band intensities were visualized by autoradiography using a BAS-2000 Phospho-imaging scanner (Fuji).

Dot blot imunological detection and quantification of equistatin in transgenic potato plants

Unless otherwise indicated leaf samples were colected between 9:30 and 10:00 am. Leaf discs were punched into separate wells of a 96-well microtiter plate and to each well 200 μL of protein extraction buffer (80 mM Tris-Cl, pH 7.6, containing 25 mM diethyldithiocarbamate, and 50 mM Na_2EDTA) was added. The samples were then three times frozen to -20°C and thawed to room temperature. The extract was pipetted to a new microtiter plate, separating it sufficiently from the insoluble leaf material. For dot blot analysis 2.5 μg of protein extract obtained with the freeze-thaw extraction procedure were transferred to Trans Blot (BioRad) nitrocellulose membranes using the SRC 96 D Dot blot apparatus (Schleicher & Schuell, Germany). Subsequently the membranes were blocked in TTBS (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween 20) containing 2% non-fat milk powder for 1 hour and then incubated with rabbit anti-EI antibodies (Eurogentec, Seraing, Belgium). The blots were subsequently washed and incubated with goat anti-rabbit IgG conjugated with horseradish peroxidase (Jackson Immuno Research, USA). The membranes were visualized with Lumi-Light Western blotting substrate and scanned in the Lumi-Imager F1^{TM} under the control of Lumi-Analist^{TM} software (Roche). Purified equistatin as a quantitative protein reference was obtained from the study of Strukelj et al. (2000). Based on dot blot analysis and polyclonal sera the total amount of equistatin present in extracts was determined. This method measured both intact and degraded non-functional protein.

Western blot analysis

A piece of a fully expanded young leaf (200-300mg) was placed in a 1.5 ml eppendorf tube and ground in liquid nitrogen to a fine powder. The powder was resuspended in 300 μL extraction buffer (100mM Tris-HCl, pH 7.6, 25 mM Na-diethyldithiocarbamate, 50 mM EDTA, 10 % PVPP, and 1 tablet per 25 mL of protease inhibitors cocktail Complete (Roche)) and this crude extract was twice centrifuged for 5 min at 14000 rpm and 4°C. Each time the supernatant was replaced into a new tube. SDS-PAGE was performed using a 15% precast resolving gel (BioRad, USA) on a mini-Protean II slab cell apparatus (BioRad). The gels were run according to the manufacturer's instructions. The separated protein samples from the SDS-PAGE gels were transferred to Trans Blot (BioRad) nitrocellulose membranes using the mini-Protean II electrotransfer apparatus (BioRad). Blocking of the membranes and the detection procedure was the same as described for the dot blots.

Assay of inhibitory activity

Plant protein extracts were tested for the presence of equistatin by the inhibition of papain (2x crystallised, Sigma, USA). Buffer solution 50 μL (50 mM MES pH 6.5

with 5 mM cysteine and 0.1 mg/ml BSA fraction V) was mixed with 20 μL of 150 nM papain in the wells of a microtiter plate. Increasing quantities (0-80 μL) of total soluble leaf protein (0.05 μg/μL) from the transgenic line KDEL22 were mixed with decreasing quantities (80-0 μL) of total soluble leaf proteins (0.05 μg/μL) from control plants transformed with the empty pBINPLUS vector in order to maintain a constant amount of plant protein (4 μg) in each measurement. After 30 min pre-incubation at room temperature the reaction was started by adding 50 μL of substrate consisting of 0.9 mg/mL Z-Phe-Arg-pNA (Bachem, Switzerland) in 50 mM MES pH 6.5 buffer (15 mg/mL stock in MeOH). The reaction was spectrophotometrically recorded at 405 nm on a Benchtop microtiterplate reader (Bio-Rad).

Purification of the equistatin degradation products from transgenic potato plants

Papain and anti-equistatin antiserum were coupled to CNBr-activated Sepharose 4B (Amersham Pharmacia Biotech) according to the manufacturer requirements. An additional carboxymethylation step was conducted for papain after the coupling to the CNBr-activated Sepharose 4B. The papain sepharose was first activated by addition of 5 mM L-cysteine in 0.1 M Na-phosphate pH 7.0 for 15 min at room temperature. Then the excess buffer was removed on a sintered glass filter (porosity G3) and the gel was placed in carboxymethylation buffer (0.1 M iodoacetate, 0.1 M Na-phosphate pH 7.0) at 37° C for one hour. The gel was then washed with a 0.1 M Na-phosphate pH 7.0 buffer. The papain and antisera-Sepharose affinity gels were packed in (1 x 6 cm) columns.

Fully expanded young leaves were collected and ground in liquid nitrogen to a fine powder. The powder was resuspended in 1:1 ratio of extraction buffer (100 mM Tris-HCl, pH 7.6, 25 mM Na-diethyldithiocarbamate, 50 mM EDTA, and 10 % PVPP). This crude extract was twice centrifuged for 5 min at 14000 rpm and 4°C. The supernatant was then heated for 3 min at 80°C and centrifuged again. This material was filtered through 22 μm filters and used as input for the immunoaffinity and papain affinity columns. The proteins in above buffers were loaded on either column at a speed of 1 ml/min. The immuno-affinity column was washed with 1M NaCl; 50 mM Tris pH 7.5, and the papain column with 1M NaCl; 50 mM MES pH 6.5. Bound proteins were eluted using 20 mM Tris pH 10.7 that contained 20 % DMSO. Fractions (1 ml) were collected and the immediately neutralised with 0.2 ml 1 M Na-phosphate pH 7.0. The fractions were analysed by Western blot analysis. The ones that contained the equistatin bands were separated on 15 % SDS-PAGE and blotted on PVDF membrane (Millipore). The membranes were washed once with double distilled water, then saturated with 100% methanol for a few seconds and finally stained with 0.1 % Coomassie Blue R-250 in 40 % methanol and 1% acetic acid. The bands were visualised after de-staining with 50% methanol. The bands were cut using surgical blades and placed in separate 1.5 ml Eppendorf tubes and send for N-terminal sequencing to ProSeq, Protein Microsequencing (3 King George Drive, Boxford, MA 01921 USA).

pH optimum of the proteolytic activities in potato leaf extracts

Fully expanded leaves were ground using a pre-chilled mortar in the presence (2:1, w/v) of double distilled water that contains 10 % PVPP. This crude extract was twice centrifuged for 5 min at 14000 rpm and 4°C, and each time the supernatant was replaced into a new tube. Buffers (50 mM MES-NaOH pH of 5.6; 6.0; 6.5 respectively, and 50 mM Tris-HCL with a pH of 7.6) were prepared and mixed in a ratio 1:1 (v/v) with the above prepared supernatant of the plant extract. These mixtures were centrifuged again to remove the insoluble materials and the supernatant was used to carry out the enzymatic analysis. In a 96-well microtiter plate, plant extracts that contain 60 µg of protein were mixed with buffers with a corresponding pH to final volume of 150 µL. The reaction was started by adding 50 µL of each of the two substrates consisting of 0.9 mg/mL Bz-Asn-pNA and Z-Arg-Arg-pNA (Bachem, Switzerland) in water (15 mg/mL stock in DMSO for Bz-Asn-pNA and in MeOH for Z-Arg-Arg-pNA). The reaction was spectrophotometrically recorded at 405 nm on a Benchtop microtiterplate reader (Bio-Rad). Equistatin was also incubated in the above prepared plant extracts in the different pH buffers and the degradation of the intact protein was monitored by western blot analysis.

Inhibitors of equistatin degradation

Leaf extract of potato cv. Desirée (60 µg) was incubated with equistatin (125 ng) for 1 hour at pH 5.6, plus or minus inhibitors (10 µM Kininogen domain 3, StefinA, Cystatin C, and Potato cystatin) and subsequently the remaining equistatin was analysed by Western blot. In parallel plant extract (60 µg, pH 5.6) was mixed with a MES-NaOH pH 5.6 buffer to a final volume of 130 µL in wells of a microtiter plate, then 20 µL of each inhibitor (100 µM stock) and 50 µL of each of the two substrates consisting of 0.9 mg/mL Bz-Asn-pNA and Z-Arg-Arg-pNA (Bachem, Switzerland) in water (15 mg/mL stock in DMSO for Bz-Asn-pNA and in MeOH for Z-Arg-Arg-pNA). The reaction was spectrophotometrically recorded at 405 nm on a Benchtop microtiterplate reader (Bio-Rad). Purified inhibitors: Kininogen domain 3, StefinA, and Cystatin C were provided by Dr. M. Abrahamson, Department of Clinical Chemistry, University of Lund, Sweden. Potato cystatin as a quantitative protein reference was obtained from the study of (Annadana et al., 2003).

Insect feeding trials

Young leaves about 5 cm long and having 0.5-1cm of stem were cut off from chosen transformed and control lines and inserted into an Eppendorf tube which contained 0.3% agar in water. Each leaf was placed on filter paper in a closed Petridish and inoculated with six newly hatched CPB larvae. The bioassays were conducted under the following conditions: 16/8 h light/darkness, 25°C and 70% relative humidity. Leaves were changed every day at 10 am. Surviving larvae were counted every day and from the third day onward they were also weighed to measure growth rates. Most of the high expressing transgenic plant lines were screened in eight different bioassays.

RESULTS

Equistatin expression analysis

Potato was transformed with three different constructs that contained the equistatin gene optimized for expression in plants under the control of the CaMV35S promoter, the Lhca.3.St1 promoter from potato and the RbcS1 rubisco small subunit promoter from chrysanthemum (Fig. 1).

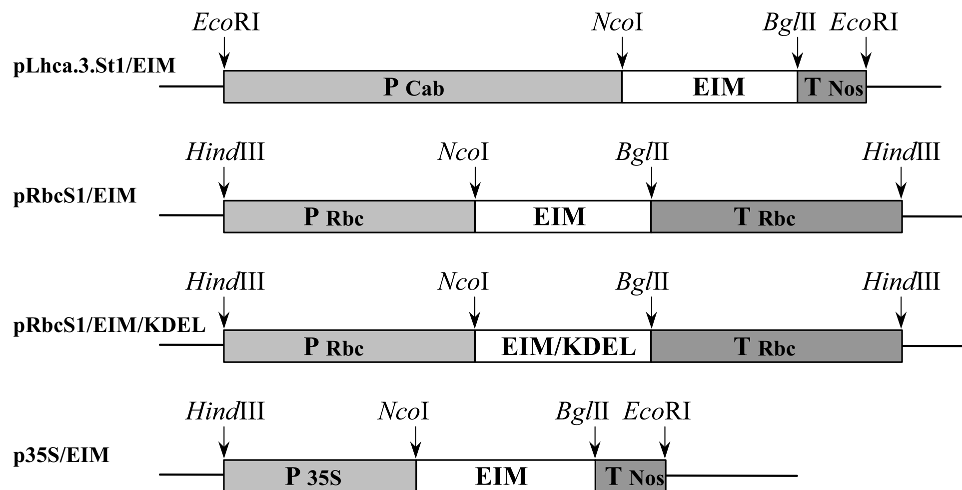


Figure 1. Schematic representation of the constructs used in this study. P Lhca.3.St1 is the light-inducible promoter of chlorophyll a/b; P RbcS1 is the promoter of the small subunit of rubisco; P 35S is the double CaMV35S promoter; T RbcS1 is the terminator of small subunit of rubisco; and T Nos is the terminator of the nopaline synthase gene. EIM is the modified equistatin coding part. EIM/KDEL is the modified equistatin coding part plus four aminoacids at the C-terminus, KDEL respectively.

In order to enhance the protein accumulation an additional construct was prepared in which the equistatin gene was also given an ER retention signal (KDEL) and placed under the control of the RbcS1 promoter. The expression of at least 30 transgenic plants per construct was analyzed by immunological detection, and on some of the pLhca.3.St1/EIM transformants Northern blots were carried out. The different constructs resulted in strongly different levels of expression. The highest population mean of expression for the equistatin gene without KDEL was obtained with the RbcS1 promoter (0.36% of total soluble protein), followed by the Lhca.3.St1 promoter (0.10%) and the 35S promoter with very low levels of less than 0.01% (Table 1). The addition of the four amino acid ER retention signal (KDEL) at the C-terminus of equistatin under the control of the RbcS1 promoter improved the

expression 5-fold, raising the mean expression level to 1.9% (Table 1). A large proportion of the pRbcS1/EIM/KDEL lines expressed equistatin at levels between 1 and 7% of total soluble protein.

Table I. Expression levels of the different constructs measured by dot blot imunological detection on samples collected between 9:30 and 10:00 a.m.

Constructs	Plants	Mean expression (% of total protein)
pLhca.3.St1/EIM	30	~0.10
pRbcS1/EIM	30	0.36 (± 0.12)*
pRbcS1/EIM KDEL	30	1.9 (± 0.39)*
p35/EIM	30	< 0.01

* Standard errors of the mean. The low signal of the plants generated with pLhca.3.St1/EIM, and p35/EIM constructs did not allow precise estimation of the mean value.

Considering that the light regulated promoters Lhca.3.St1 and RbcS1 performed 10-40 fold better than the constitutive 35S promoter it seemed that possibly also the known circadian expression of these promoters might affect our results. Twenty-four hour time-course samples of both RNA and protein were taken to investigate the relationship between equistatin synthesis and degradation in the plants transformed with the pLhca.3.St1/EIM construct (Fig. 2). The Northern blot confirmed that the Lhca.3.St1 promoter is transcriptionally active during the morning and nearly silent during the night in agreement with (Nap et al., 1993). At the protein level (Fig. 2B) intact equistatin (top band) only represented about 1/3 of the signal even at 11 a.m. at the peak of transcription (9-11 a.m.). Most of the protein was degraded into increasingly smaller fragments. After 6 p.m. only the lowest 5 kD bands were visible, and with very low intensity after 7 hours of darkness (1 a.m.). The defined sets of bands indicated that degradation is initiated by proteases that first cleave the peptide chain of equistatin at specific sites. This degradation did not occur during the extraction procedure as, firstly, a cocktail of synthetic inhibitors including E-64 was added to the extraction buffer to prevent degradation. We knew from earlier experiments that E-64 could prevent equistatin degradation in extracts. Secondly, the patterns of degradation were clearly different depending on the time of the day, the promoter and the targeting used (Fig 2 and 3). Thirdly, all samples were prepared and extracted in the same way and at the same time. It would be unusual for a protease to be present at high concentrations late at night and to be gone at dawn. Instead, it is shown that the high concentrations of intact protein in the morning derive from the strong transcription of the gene at that time, which disappears during the rest of the day.

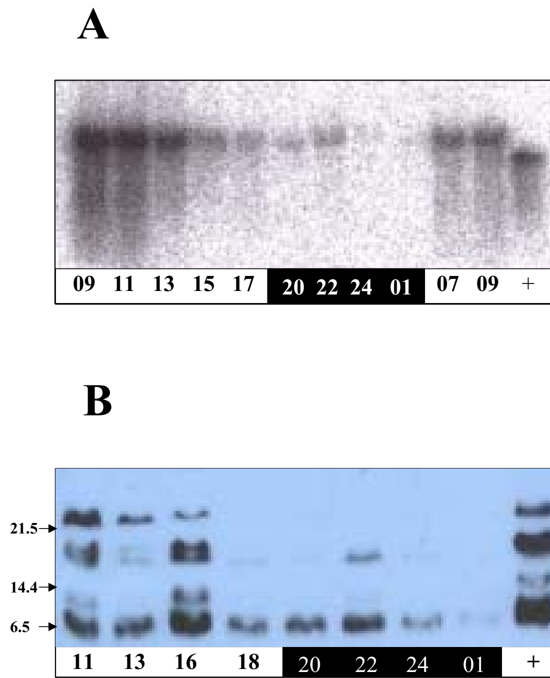


Figure 2. A. Northern blot analysis of the equistatin transcript expressed in potato cv. Desirée transformed with pLhca.3.St1/EIM construct. Numbers represent the hours of the day, black and white shading represent the dark and the light part of the photoperiod. (+) Indicates the positive control mRNA used in the experiment. **B.** Western blot analysis of the equistatin protein accumulated in potato cv. Desirée transformed with pLhca.3.St1/EIM construct. Numbers represent the hours of the day, black and white shading represent the dark and the light part of the photoperiod. (+) Indicates the positive control protein, produced in *E. coli* and used in the experiment.

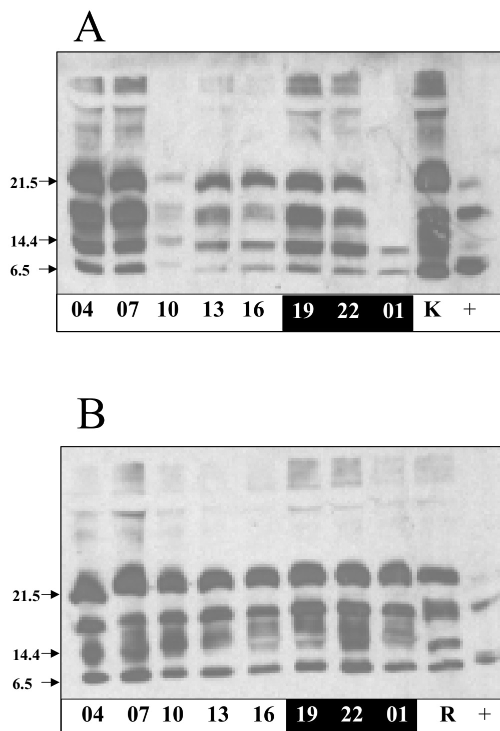
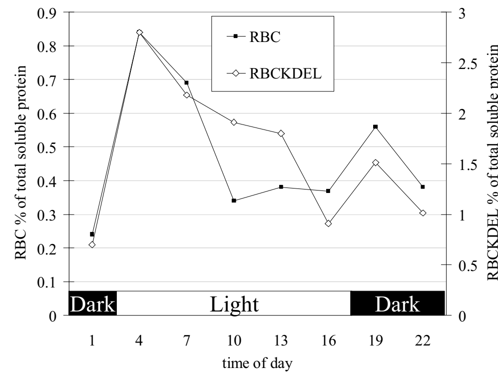


Figure 3. Western blot analysis of the equistatin protein accumulated in potato cv. Desirée transformed with **(A)** pRbcS1/EIM and **(B)** pRbcS1/EIMKDEL construct. Numbers represent the hours of the day, black and white shading represent the dark and the light part of the photoperiod. (+) Indicates the positive control protein, produced in *E. coli* and used in the experiment. K is protein extract from KDEL plant and R is protein extract from non KDEL plant, both taken at 7 a.m. in the morning.

Western blot analysis of the plants that expressed the pRbcS1/EIM construct (Fig. 3A) revealed a diurnal rhythm of protein expression and a pattern of equistatin degradation similar to the plants expressing pLhca.3.St1/EIM. A notable difference is that a higher level of transcription is apparently maintained during daylight and early night resulting in the presence of undegraded protein during the whole day and the first part of the night. Only after 7 hours darkness all intact bands have disappeared and only degraded protein remains.

The effect of adding a KDEL retention signal was a 5-fold improvement of the amount of accumulated protein (degraded and undegraded), but it was not yet clear whether qualitative changes had occurred in the type of degradation of the protein. The differences in degradation of equistatin with and without the KDEL signal can be observed by comparing Figure 3A and 3B. Four bands were observed for the –KDEL plants: 21.5 kD (intact protein), ~18 kD, ~ 12 kD and ~ 6 kD. The +KDEL construct resulted in a slightly different degradation pattern: instead of the defined ~12 kD band a number of bands with slower mobility appeared. The western blots on figure 3 were, however, overexposed to show all degradation bands. To precisely quantify the qualitative diurnal variation of accumulated equistatin protein two plant lines of the pRbcS1/EIM construct, lines 24 and 36, and two lines of the pRbcS1/EIMKDEL construct, lines 21 and 28, were selected for a precise quantification. Time course samples were taken every 3 hours during a 24 hour period and the precise amount of equistatin protein was quantified by dot blot immunological analysis. In this way all immuno-reactive equistatin, degraded and undegraded, was measured and compared. The results for the two different constructs were compared by aligning the two data sets on the basis of the highest protein accumulation point at 4 a.m. (Fig. 4 A). The two data sets are following a very similar diurnal fluctuation despite the fact that the average protein expression level was 5-fold higher in the presence of KDEL. The amount of accumulated protein in both cases peaks to a 4-fold higher level 1 hour after light is switched on, then gradually decreases during the day to a 3-fold lower level at 4 pm. One hour after the light is switched off there is a smaller secondary peak, finally leading to the lowest accumulation level at 1 am after a 7 hour dark period. The increased equistatin accumulation in KDEL plants, therefore, seems to be due to an overall slow down of degradation of the protein as it is unlikely that transcription or translation itself would be affected by the downstream KDEL-element. The graphs suggest that the sampling moment at 10 am to determine the population average may have affected the ratio found of a 5-fold improvement due to KDEL and would have been less pronounced at other moments during the day. The cleavage and degradation of equistatin was expected to directly affect the amount of active protein. This was investigated by taking plant material harvested at 10 a.m. and titrating it against papain (Figure 4 B). The selected line KDEL 22 was found to express 1.6% of active protein, while according to the immunological dot blot detection it expressed equistatin at 5.2%. This indicated that a large fraction of the expressed equistatin was probably inactivated by the specific cleavages observed on the western blots.

A



B

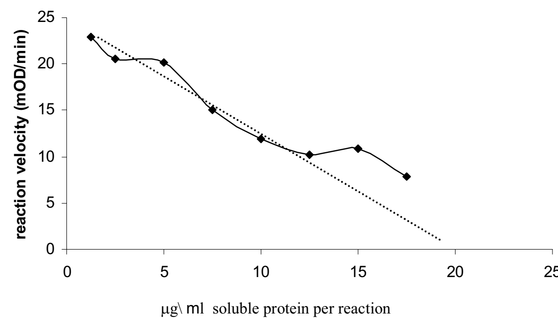


Figure 4. A. Diurnal equistatin protein levels quantification based on dot blot analysis. Mean values found on plant lines pRbc1/EIM 24 and 36, indicated as RBC are shown against mean values of pRbc1/EIM/KDEL lines 21 and 28, indicated as RBCKDEL. **B.** Titration of 15 nM papain activity against potato leaf extracts from pRbc1/EIM/KDEL22 using Z-Phe-Arg-pNA as a substrate as described in Materials and Methods. The X-axis indicates the concentration of total soluble protein fron the transgenic line required to inhibit 15 nM papain; At the X-axis intercept (doted line) all papain is inhibited by 15 nM equistatin (=0.32 µg/mL), which converts to 1.6% (w/w) of total soluble protein.

Isolation and characterization of *in-planta* equistatin degradation products

The defined set of equistatin degradation bands *in-planta* indicated that specific proteases initiate the degradation of equistatin at specific positions in the polypeptide chain. To investigate the sites of cleavage we isolated the protein and its degradation products by means of immuno-affinity and papain affinity chromatography. The isolated products were separated on gel, blotted to PVDF membrane and N-terminally sequenced (Fig. 5). By immuno-affinity only the immunologically weaker bands could be recovered. The intact equistatin band remained on the column probably due to the multiple clonal IgG proteins present in the serum which co-operatively bind many different epitopes on the intact protein. The papain affinity column was expected to result in the purification of those equistatin fragments, which still

contained a functional first domain, as only that domain inhibits papain. Interestingly, the papain column led to the purification of the intact band (22 kD), the largest degraded fragment (~16 kD) and a small fragment (~5 kD), but not of any of the weaker other bands.

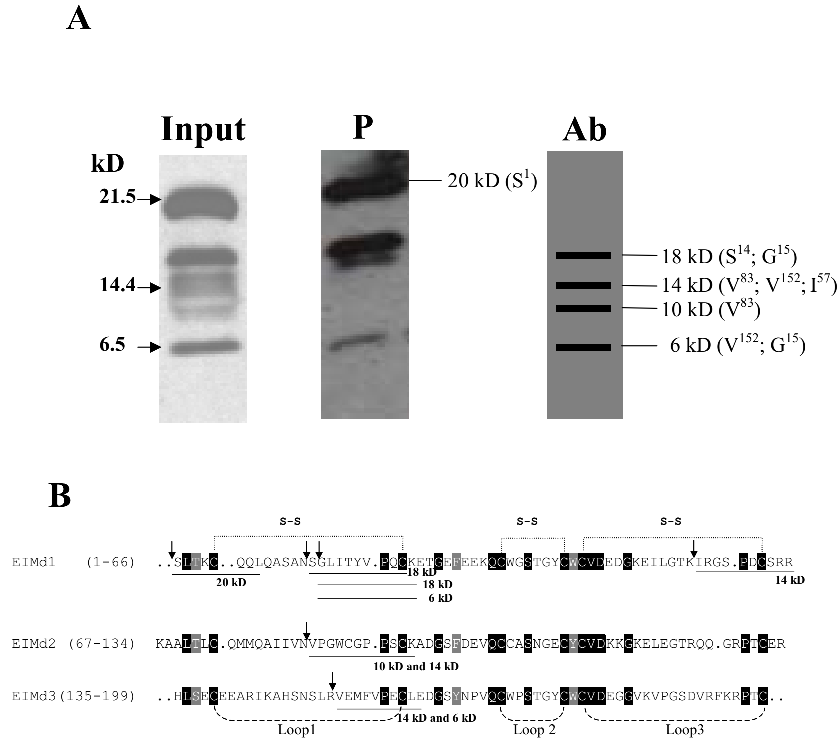


Figure 5. Isolation from plants and N-terminal sequencing of the equistatin degradation products. **A**. Purification on either papain sepharose (P) or antibody affinity column (Ab) as described in Materials and Methods. For the P column the fractions containing equistatin were analyzed by western blot and are shown. For the Ab column horizontal bars indicate the sequenced bands found on the Coomassie stained membrane. Molecular mass and the N-terminal aminoacids of the different peptides are indicated. **B.** An alignment of the tree domains of equistatin is shown. The positions of the disulfide bridges and protein loops as deduced from the solved structure of the homologous thyropin p41 (Gunčar et al., 1999) are marked above or below the alignment. The N-terminal sequences of the *in-planta* degradation bands are shown underlined.

All bands from the immuno-affinity column and the intact band of the papain column were sequenced by Edman degradation. We were unsuccessful in sequencing the degraded papain purified bands, because the papain column yielded a smear of proteins in which only the undegraded band could be identified for sequencing. Some of the bands yielded different peptide fragments, probably due to complex trimming of the polypeptide both at the N- and C-terminus resulting in similar sized peptides. In addition, we obtained mixed N-terminal sequences for the cleavage in loop 1 of domain 1 (1.1), two from the fragment with a size of 18 kD (N-terminus: S^{14}, G^{15}), and one from the fragment of 6 kD (N-terminus: G^{15}) (Figure 5B). The fragments differ by one amino acid which suggests N-terminal trimming by aminopeptidase activity. Also, the first loops of the other equistatin domains (2.1; 3.1) were cleaved at nearly identical topological positions (N-terminus: V^{83}, V^{152}) (Figure 5). A fourth

cleavage site was identified in the third loop of the first domain (1.3). The specificity of proteases is usually mainly determined by the P_1 aminoacid, N-terminal to the hydrolyzed peptide bond. These were characterized by two asparagines Asn^{13} (1.1) and Asn^{82} (2.1), one lysine Lys^{56} (1.3) and one arginine Arg^{151} (3.1).

The fast shift in size from intact protein to the first major degradation band with an N-terminus at Asn^{13} (1.1) indicates that probably the Asn^{13} cleavage occurs at a similar rate to a cleavage at possibly Arg^{151} which can explain the ~18 kD peptide product. The cleavage Lys^{56} (1.3) we considered as less efficient *in-planta* than the other three, because the corresponding peptide was found only in minor quantities among two other peptides, while the corresponding position of the band was weaker when analyzed by western blot.

Proteolytic activities in potato plant extract

Equistatin was incubated in potato leaf extract in buffers of different pH. In addition, we used an asparagine-specific substrate Bz-Asn-pNA, which is specific for legumain-type proteases, and an arginine-specific substrate Z-Arg-Arg-pNA, which is known to be hydrolyzed by both cysteine (papain-like) and serine proteases (kexin/subtilisin-like). The two substrates were chosen on the basis of the *in-planta* cleavage points behind Asn and Arg/Lys P_1 residues. The results are presented in Figure 6.In our *in-vitro* experiment we could only observe the disappearance of the intact equistatin band and not see the equistatin degradation products as in the *in-planta* degradation. Apparently, due to tissue grinding many plant peptidases were released from their subcellular compartment and degraded the equistatin degradation products faster than they could be observed. It was found that the optimal pH for degradation of equistatin in potato leaf extract was pH 5.6 or below (not determined). This agrees with the optimum of proteolytic activity known for legumain proteases, which was confirmed with the Bz-Asn-pNA substrate. In contrast, the Z-Arg-Arg-pNA substrate showed an optimum at pH 7.8 which did not coincide with the optimum of degradation of equistatin. This suggested that legumain-like proteases may be crucial in the initiation of degradation of equistatin *in-planta*.

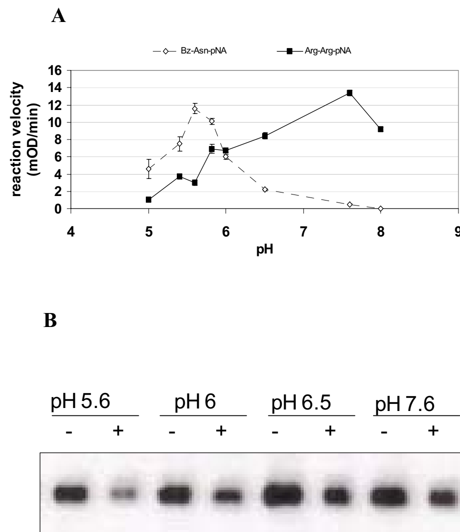


Figure 6. A. On the top proteolytic activities of potato leaf extract cv. Desirée (60 μg protein) towards two different substrates Bz-Asn-pNA and Z-Arg-Arg-pNA at various pH. **B.** On the bottom leaf extract of potato cv. Desirée (60 μg) was incubated with equistatin (125 ng) for 1 hour at various pH, and subsequently analysed for remaining equistatin by Western blot. (-) negative control preheated potato extract in the corresponding buffer.

Proteinaceous inhibitors of equistatin degradation

The pH optimum of equistatin degradation below pH 6 and the fast Arg- and Asn-specific cleavages in the protein suggested the involvement of cysteine proteases. Papain-like cysteine proteases can be inhibited by all types of cystatins, but plant legumains are inhibited only by type II (egg-white cystatin) and type III cystatins (kininogen)(Abe et al., 1993; Rotari et al., 2001). The inhibition of porcine legumain by human cystatin C (similar to egg-white cystatin) was found to be due to a second reactive site situated at the “back” of the inhibitor. Thus, cystatin C and some other cystatins possess two separate inhibitor active sites, that enable them to inhibit both legumain and papain proteases simultaneously (Alvarez Fernandez et al., 1999; Chen Jinq et al., 1997).

To test whether in a plant extract the equistatin degradation could be selectively inhibited by different cystatins we tested representative members of all known cystatin types: stefin A - type I; cystatin C – type II; kininogen D3 – type III; potato cystatin – phytocystatin (Fig. 7). The two inhibitors without cysteine bridges and no known legumain inhibitory activity, stefin A and potato cystatin, inhibited the equistatin degradation by only 10-12%, relative to the control without inhibitor. In contrast cystatin C inhibited equistatin degradation up to 60% and full inhibition was obtained using kininogen domain III as inhibitor. In line with our previous experiments when the synthetic substrate Bz-Asn-pNA was added to the plant extract the highest inhibition of the proteolysis was achieved by using kininogen D3 and cystatin C. In contrast, when the Z-Arg-Arg-pNA substrate was used, stefin A was the strongest inhibitor, but this had no significant effect on preventing equistatin degradation. This suggests that, at least in a crude plant extract, legumain-like plant proteases initiate the degradation of equistatin, whereas the papain-like cysteine proteases contribute only about 10% of equistatin degradation.

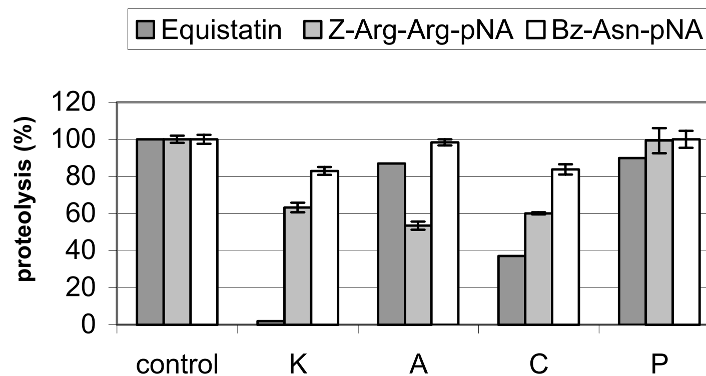


Figure 7. Hydrolysis of a leaf extract at pH 5.6 was measured using Z-Arg-Arg-pNA and Bz-Asn-pNA, and equistatin as substrates in the presence/absence of the indicated inhibitors. **control**-no inhibitor; **K**- Kininogen domain 3; **A**- StefinA; **C**- Cystatin C; **P**- Potato cystatin. Error bars represent the standard error of the mean from four simultaneous measurements. The proteolysis of equistatin was measured by Western blot analysis. The amount of disappearance of intact protein in the absence of inhibitors (equistatin in preheated extract – equistatin in plant extract) was taken as 100% proteolysis.

Insect feeding trials

The efficacy of equistatin to give rise to resistance against CPB larvae when expressed in potato was checked in leaf feeding trials. In total eight different bioassays were conducted. Figure 8 shows a representative bioassay. Newly hatched larvae were fed with leaves from several lines originating from pRbcS1/EIM and pRbcS1/EIM/KDEL, and using pBINPLUS transformed lines as a control. Line KDEL22, which at 10 a.m. showed an expression level of 5.2 % total protein (dot-blot) and 1.6% functional protein (papain titration), was found to suppress larval growth of CPB by 38%. Larval development was not altered when the larvae were fed KDEL23 plants which expressed equistatin at 10 a.m. at 3.3% total protein (dot blot). In the given experiment no significant differences were observed in larval mortality. However, between the eight different experiments the effects on larval growth and mortality tended to vary strongly (from high mortality to no effect on the larvae fed on transgenic plants). This may have been due to the observed fluctuation in the levels of expressed functional protein, which between assays are probably also affected by developmental (plant age), stress, and seasonal (light intensity) factors.

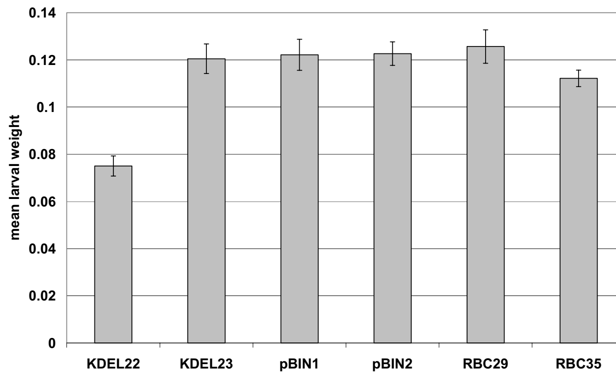


Figure 8. Representative CPB beetles bioassay. Mean larval weight (n=6) after three days of feeding. Expression levels KDEL22 (5.2%); KDEL23 (3.3%); pBINPLUS1 and 2 (0%); RBC29 (0.17%); RBC35 (0.58%).

DISCUSSION

Equistatin degradation in potato cv. Desirée

In this study we aimed to obtain potato plants resistant to CPB larvae via stable and high level expression of recombinant equistatin in the leaves of potato cv. Desirée plants. Three different promoters (CaMV35S, Lhca.3.St1 and RbcS1) were analyzed for their ability to drive maximal expression of equistatin. The *rbcS1* promoter from chrysanthemum (Outchkourov et al., 2003a) yielded the highest average expression level (0.36%), and this was further improved 5-fold by the addition of the KDEL retention signal (1.9%). Some plant lines were found to express equistatin at levels of up to 7% of total soluble protein as measured by dot blot detection. Protein degradation and the circadian expression of the RbcS1 promoter, however, affected the functionality and accumulation level of the accumulated protein. The distinct bands of degradation indicated that equistatin was cleaved at specific positions *in-planta*. Two thirds of the protein were found to be degraded into an inactive form at 10 a.m. and during the night only 20% of the initial protein accumulated in the morning remained. Thus, during the night the level of functional protein in the best plant was probably at least 15-fold lower: around 0.5 %. Probably as a consequence of the degradation of equistatin in plants we did not consistently observe resistance to CPB larvae in these plants. Apparently, the level of expression of functionally active equistatin was too strongly affected by the endogenous proteases of potato.

Identification of the plant proteases involved in equistatin proteolysis

The cleavage sites of the equistatin degrading proteases in plants were investigated by affinity purification and N-terminal sequencing of the degraded equistatin fragments. (Lenarcic et al., 1997) isolated intact equistatin protein from its natural host *Actinia equina*. (Lenarcic and Turk, 1999) showed that *in-vitro* equistatin is susceptible to cleavage by the serine protease trypsin at Lys^{67}-Ala^{68} and Arg^{151}-Val^{152} producing fragments of 7 and 14 kDa. In *E. coli* equistatin was found to be cleaved at Val^{152} –

Glu^{153} (Strukelj et al., 2000). *In-planta* four putative cleavage positions were identified: Asn^{13}, Ser^{14}, Lys^{56}, Asn^{82}, and Arg^{151}. The different cleavage positions (Asn13, Ser14, Lys56, Asn82) clearly indicate the involvement of plant-specific proteases since they do not occur in any of the other hosts. *In-planta* we suspect that in the case of the linked Asn^{13} and Ser^{14} cleavages the primary cleavage site is at Asn^{13} rather than Ser^{14} and that the cleavage at Ser^{14} is the result of aminopeptidase activity which removed the Ser^{14} residue. An asparagine residue (Asn^{148}) is also found in domain 3 in a corresponding position to the cleaved Asn-residues in the other two domains. From the sequence alignment it seems likely that Asn^{148} in domain 3 is the actual endopeptidase clevage site, but due to aminopeptidase activity we have possibly misidentified Arg^{151} as the primary P1 residue (Figure 5B).

Furthermore, equistatin was most efficiently degraded at pH 5.6 suggesting that Asn-specific proteases may initiate the degradation of the protein. The only known family of Asn-specific proteases in plants is known as the legumain family (Muntz and Shutov, 2002) and the pH 5.6 optimum is in agreement with the characteristics of this enzyme family (Rotari et al., 2001).

Until now there are no reported legumain inhibitors from plants. In contrast, cystatins derived from the animal kingdom have been shown to inhibit some legumains from both plants and animals It was demonstrated that the addition of domain 3 of human kininogen (type III) at 10 μM concentration could completely inhibit the degradation of equistatin *in-vitro* in potato cv. Desirée leaf extract. The known inhibitor of the mammalian legumains human cystatin C (type II) inhibited equistatin degradation up to 60 % at the same concentration. In contrast, there was very little effect (~10%) of stefin A (type I) and potato cystatin (phytocystatin). Legumains from Jack Bean seeds (Abe et al., 1993) were also found to be sensitive to inhibition by kininogen and not by stefin A. Using synthetic substrates potato Asn-specific activity was also partially inhibited by kininogen and cystatin C, and not by stefin A and potato cystatin. Arg-specific activity was inhibited most strongly by stefin A, less by kininogen and cystatin C, and not by potato cystatin. The combination of results on the pH optimum of degradation, the cleavage sites, and the specific inhibitors of degradation of equistatin and synthetic substrates, therefore, indicate that the third domain of kininogen is preventing equistatin degradation *in vitro* by inhibiting both potato legumains and Arg-specific proteases. The activity of kininogen against potato legumain is confirmed by the presence on domain III of a potential legumain binding back-side loop similar to the one on type 2 cystatins, which is not present on domain I/II of kininogen (Alvarez Fernandez et al., 1999). In the study of Alvarez Fernandez et al. (1999) kininogen was not active towards pig legumain, but structural differences between pig and potato/kidney bean legumain may be involved in the different sensitivity to kininogen (Rotari et al., 2001).

Plant endosomal proteases as an impediment for protein production in plants

The function of the ER as the initial compartment of the secretory pathway is to ensure proper folding, post-translational modifications (peptide processing, glycosylation, disulphide formation), and oligomerization of the secretory proteins. For functional expression many heterologous proteins depend on the secretory

pathway, yet it is evident that the protein accumulation levels may be seriously affected by the proteases resident in the endososomal system of potato.

The accumulation of other proteins like antibodies, targeted to the ER, is known to be severely affected by plant proteases (Stevens et al., 2000). The study of (Sharp and Doran, 2001a) indicate that most likely the degradation of a murine IgG1 is initiated along the secretory pathway outside the ER, and further degraded in the extra-cellular space. In our particular case it is likely that equistatin ends up in the vacuole or the cell wall, where it is further degraded. Legumain activity resides both in the vacuole and the cell wall (Muntz and Shutov, 2002). The KDEL signal seems to only partly retain equistatin in the ER since nearly identical patterns of degradation were found in both plants expressing equistatin with and without KDEL signal. This inability of the KDEL signal to retain proteins completely in the ER has been described before with proteins ending up in either the vacuole (Toyooka et al., 2000) or the cell wall (Inohara et al., 1989).

Furthermore, the significance of proteolytic degradation in plants is highlighted by a number of studies that describe different proteolytic fragments in different plant species and plant cell cultures (De Neve et al., 1993; Hiatt et al., 1989; Dolja et al., 1998; Khoudi et al., 1999). Thus, plant proteases can be a serious hurdle for the production of functional intact protein in plants and strategies to effectively avoid this degradation are poorly developed.

Strategies to reduce the degradation of heterologous proteins in plants

Reduction of the degradation of ER-targeted equistatin in potato, and of proteins in the secretory pathway in general, may be achieved in two different ways. Replacement of the amino acids presently cleaved by functionally equivalent amino acids, which preserve the biological activity of the protein, but prevent cleavage, may allow the design of stable molecules for expression in plants. Alternatively, the co-expression of protease inhibitors, that prevent protein degradation *in-vitro,* may also stabilize the protein *in-planta*. A rational mutagenesis approach will only be possible with proteins for which precise information on susceptible cleavage sites is available. In most cases protein degradation occurs too rapidly to identify the initial cleavage points. As a general approach the co-expression of cysteine protease inhibitors like kininogen domain III may, therefore, more effectively stabilize foreign proteins in plants. Future studies are needed to investigate the feasibility of such approaches.

CHAPTER 6

ENGINEERING PROTEIN STABILITY OF EQUISTATIN IN POTATO PLANTS

Nikolay S. Outchkourov[1,3], Boris Rogelj[2], Stanimir Dulev[3], Dusan Turk[2], Maarten A. Jongsma[1]

[1]Plant Research International, PO Box 16, NL-6700 AA Wageningen, The Netherlands

[2]Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

[3]Department of Plant Physiology and Molecular Biology, University of Plovdiv, Bulgaria

Submitted for publication

ABSTRACT

Equistatin is a protease inhibitor with potential to control a wide array of insects. Functional accumulation of equistatin in transgenic potato plants (*Solanum tuberosum* cv. Desirée) is, however, hampered by endogenous proteolytic activities. Four different proteolytic cleavage sites along the polypeptide chain of equistatin were found to initiate the degradation of the protein. In this study, an attempt was made to stabilize equistatin by substituting amino acids at the positions sensitive for proteolysis. In total eight different variants of equistatin with various combinations of amino acid substitutions were produced. The protein variants expressed in *Pichia pastoris* retained the ability to strongly inhibit papain (K_i from 0.68 to 2.6 nM) and cathepsin D. Expression of the equistatin variants in potato revealed only partial stabilization. Some protein fragments accumulated as proteolytically resistant bands, yet, on the overall the stability of the protein against proteolytic degradation had deteriorated. Alternative approaches to stabilize equistatin are discussed.

INTRODUCTION

In plants endogenous proteolytic activity which degrades heterologous proteins still represents a major hurdle for the development of economically feasible plant expression systems. Understanding the nature of the proteases and sensitivities of foreign proteins in plants is crucial in designing strategies for proteolytic stabilisation of foreign proteins in plants.

Equistatin is an inhibitor of cysteine and aspartic proteases isolated from the sea anemone *Actinia equina* (Lenarcic and Turk, 1999). It was shown to be an efficient inhibitor of the digestive proteases of Colorado potato beetle (*Leptinotarsa decemlineata Say* (*Coleoptera, Chrysomelidae*)) and western flower thrips (*Frankliniella occidentalis*), and its insecticidal properties were demonstrated in artificial diet assays (Annadana et al., 2002b; Gruden et al., 1998). The gene sequence of equistatin was modified for optimal expression in plants and expressed in potato cv. Line V and Desirée plants up to very high levels (Outchkourov et al., 2003a; Outchkourov et al., 2003b). The accumulation of functionally active inhibitor was, however, hampered by the rapid proteolytic cleavages *in-planta,* and this reduced the insecticidal properties of this protein (Outchkourov et al., 2003b). The amino acids sensitive to cleavage were identified after isolation of the equistatin degradation products and Edman degradation to obtain the N-terminal amino acids (Outchkourov et al., 2003b). Evidence was presented which indicated that the proteases involved in equistatin degradation *in-planta* were mainly legumain-type cysteine proteases as well as some Arg/Leu- specific proteases.

A general strategy for stabilization of proteins is to substitute amino acid residues in or nearby the cleaved peptide sequence, and, thus, to remove or protect the protease-sensitive site (Audtho et al., 1999; Frenken et al., 1993; Kapust et al., 2001; Markert et al., 2001; Varallyay et al., 1998). Such rational design to achieve stabilization of the protein requires that the protein folding and activity remains intact.

In this study, amino acids at the positions that were previously found to be cleaved *in-planta* (Outchkourov et al., 2003b) were substituted to increase stability against proteolytic degradation. The stability *in-planta* was assessed by taking time course samples and making use of the circadian rhythm of the *rbcS1* promoter. The feasibility of rational design for stabilization is addressed, and more random approaches in terms of residue substitutions are proposed.

MATERIALS AND METHODS

Site directed mutagenesis

The Chameleon® Double-Stranded, Site-Directed Mutagenesis Kit was used for the introduction of point mutations within the coding sequence of the equistatin gene (Gene bank accession number AY166597) cloned in the pBluescript SK+ vector (Stratagene) using the following 5'-phosphorylated oligonucleotides (Eurogentec, Seraing, Belgium):

Dom1.1N/P 5'-CCTATCAGACCACTTGGAGCCGAAGCCTGCAGCTGTTGGC-3'; Dom1.1Ii 5'- GCATTGTGGTACATAAGTACCTGGATGAACAGCTGGGATATGAGACTGCAGCTGTTGGCATTTCG-3'; Dom1.3K-E 5'-GGAGATCCACGGATCTCTGTACCTAGGATCTTCTTTCCATCTTCATCC-3'; Dom1.3Ii 5'-CTGCAGTCTGGAGATCCACGAGACCTAGTGTTTGGAATCTCAGTTCCATCTTCATCCACACACC-3'; Dom2.1V/P 5'-CACCAACCAGGGACATTTGGAATGATGGCTTGCATC-3'; Dom2.1 N/P 5'-CACCAACCAGGGACTGGTACAATGATGGCTTGCATC-3'; Dom2.1 N/A 5'-CACCAACCAGGGACTGCTACAATGATGGCTTGCATC'-3'; Dom3.1R/A 5'-GGCACGAACATCTCAACAGCAAGACTGTTTGAATGTGCC3'; Dom3.1S/VNNV 5'-CTCAACACGAAGACTGTTAACATTGTTCACATGTGCCTTGATACGAGCTTCCTCGCATTCGCTTAGG

The mutations were verified by nucleotide sequencing.

Expression purification from *Pichia pastoris* and analysis of the inhibitory properties

After verification of the desired mutations, the coding regions of EIS-1 to EIS-8 (Figure 1) were amplified by PCR and cloned into the *Xho*I-*Not*I site of the pPIC9 vector, transformed to *P. pastoris* strain GS115, and expressed essentially as described for the EIM gene in (Outchkourov et al., 2002). The purification was done as described in (Outchkourov et al., 2002) with the difference of using smaller 12 ml Phenyl-Sepharose columns and no further purification step. EIS1 - 6 purified proteins were used for a K_i determination versus papain (Sigma, St Louis, MO, USA) as described in (Beekwilder et al., 2000; Rogelj et al., 2000). EIS1 and EIS8 at 40 nM concentration were tested for inhibition of 10 nM Cathepsin D (Sigma) using H-Pro-Thr-Glu-Phe-Phe-(NO_2)-Arg-Leu-OH (Novabiochem, Laufelfingen, Germany) as described in (Lenarcic and Turk, 1999).

Expression in plants

The coding sequences of the eight constructs EIS 1-8 were amplified using primers pUC/M13 Forward (5'- CTAAAACGACGGCCAGTG-3') and EIMKDEL (5'-TTTTTTAGATCTTTAAAGTTCATCCTTAGCGCATGTGGGACGTTTGAATC-3') to add the KDEL amino

acid sequence to the C-terminal end of EIS1-8. The DNA fragments were digested with NcoI-BglII and cloned in the pUCRBC vector (Outchkourov et al., 2003a). Subcloning into pBINPLUS, and electroporation of *Agrobacterium tumefaciens* (AGL0) (Lazo et al., 1991) was as described for the EI and EIM genes in (Outchkourov et al., 2003a). Plant transformation with the wild type and the EIS 1-8 constructs was conducted with potato cv. Impala using tuber discs as a source of explants (Hoekema et al., 1989). In addition, the wild type equistatin gene was transformed to potato cv. Kondor. The expression levels of the equistatin variants EIS1-8 in transgenic potato plants were quantified using dot blot and Western blot methods as described in (Outchkourov et al., 2003a).

RESULTS AND DISCUSSION

Strategy for stabilization

Equistatin is a protein which consists of three thyroglobulin type-1 domains. The first N-terminal domain acts as an inhibitor of cysteine proteases (K_i for papain of 0.57 ± 0.04 nM), the second as an inhibitor of aspartic proteases (K_i for cathepsin D of 0.3 ± 0.15 nM), and the function of the third domain remains to be elucidated (Lenarcic and Turk, 1999). Recently, the structure of p41, a similar thyroglobulin type-1 protein (also called thyropin (Lenarcic and Bevec, 1998)), was solved in complex with Cathepsin L (Guncar et al., 1999). The thyropin domain of p41 was shown to possess a wedge shape, created by three structural loops. In terms of its structural topology the folds are reminiscent of the inhibitory wedge of cystatins. Each of the three loops was shown to interact with Cathepsin L. Residues on the first loop are bound to the surface of cathepsin L and contribute the largest number of interaction points per loop. The second loop was shown to be tightly packed into the active site, in the vicinity of the catalytic residues. The third loop was shown to be placed higher above the active site than the other two. Figure 1 shows an alignment of the p41 protein and the three domains of equistatin. The inhibitory mechanism of domain 2, which inhibits cathepsin D, is not clear. Also, the interactions with specific insect cysteine and aspartic proteases of both domains 1 and 2 are unknown. Therefore, the choices of amino acid residue substitutions with the aim to improve the stability without destroying the inhibitory characteristics is rather empirical, and have to be experimentally validated.

Within the putative first loops of each of the three thyropin domains of equistatin cleavage sites were identified (Figure 1). Amino acid residue substitutions were based on considerations to optimally preserve the structural (and therefore inhibitory) properties of equistatin. In domain 1 the cleaved Asn13 (first loop)(Outchkourov et al., 2003b), was removed by two different substitution strategies. In the first strategy Asn13 was exchanged for a Pro-residue. Proline is an imino rather than amino acid that is known to restrict the conformational freedom of the backbone of polypeptide chains (Schimmel and Flory, 1968). In addition, Pro-residues are known to be unfavourable in the P_1 as well as P_1'-position for a number of different proteases (Bromme et al., 1986). The combination of both considerations, the so-called proline concept, has been realised in a number of cases (Frenken et al., 1993), and improved protein stability has been achieved. Furthermore, a Pro-residue

was found in other thyropins (p41) in a homologous position, indicating that the rigidity of the Pro-residue would probably not destroy the capacity of equistatin domain 1 to inhibit cysteine proteases. On the other hand, the Pro-residue in the p41 protein was found to be positioned N-terminally to the amino acids interacting with cathepsin L, at the edge of loop one (Guncar et al., 1999). Merging combinations of amino acids from different homologous proteins in this way might have destructive effects on protein folding. To reduce the risk of designing a dysfunctional inhibitor we also exchanged the complete edge of the putative loop of equistatin with amino acids found on the edge of the homologous p41 protein. Thus, the entire peptide ASANSGLI found in equistatin was converted to SHIPAVHP (Figure 1A).

In the third loop of domain 1 a second cleavage site was previously found at Lys56 (Outchkourov et al., 2002). We exchanged this Lys56 amino acid residue with Glu, thereby substituting the charge from positive to negative and expecting that this amino acid would not likely be digested by the same protease. For example trypsin is highly selective for Arg and Lys residues, because they are positively charged (Huber and Bode, 1978). However, the problem here was that the Lys56 residue likely interacts electrostatically with Glu51, like Glu243 with Arg248 in p41 (Guncar et al., 1999) (Figure 1A). The Glu243 to Arg248 electrostatic interaction is known to shield the Trp235 in p41, part of the conservative Cys-Trp-Cys-Val and a main feature of the tyroglobulin type-1 related proteins (Guncar et al., 1999). The position of both charged residues (like Glu243 and Arg248 in p41; Lys56 and Glu51 in equistatin domain 1) is conserved in most of the thyroglobulin type-1 domains (Figure 1A) (Guncar et al., 1999) except that the position of the negative and positive residues seems to be interchanged in some cases. As a result of this analysis, in addition to the Lys56 to Glu substitution, we also exchanged Glu51 to Lys in order to recreate the structurally important electrostatic interaction (Figure 1A). In addition, we conducted a different strategy for the third loop of domain 1 in which we exchanged a number of residues putatively situated at the edge of loop 3 of equistatin with residues from the edge of loop 3 of p41 (EILGTKI to TIPNTRS).

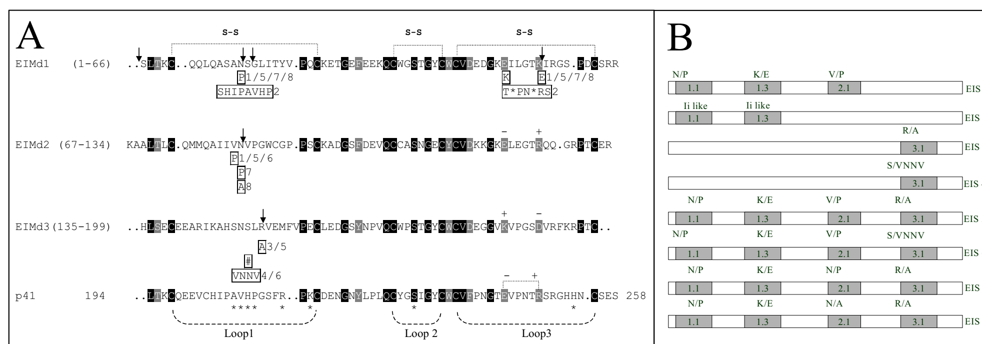


Figure 1. A. Sequence alignment of the three equistatin domains (EIMd1-3) with the homologous p41 protein (Guncar et al., 1999). Arrows indicate the proteolytic cleavage sites found in the study of (Outchkourov et al., 2003b). Amino acid substitutions are indicated in boxes, and the number of the corresponding new EIS (equistatin stabilized) construct in which the mutations are incorporated is indicated. **B.** Schematic representation of the EIS constructs used in this study, numbers indicate the position of the amino acid substitutions in terms of domain and loop. The substitutions are indicated using the single amino acid letter code

In domain 2 of equistatin the cleavage at the Asn82 residue was approached with three different strategies for stabilisation. Considering that dominhibitor of aspartic proteases (Cathepsin D) which lacks any structural information on the protein interaction, any residue substitution in the polypeptide chain could potentially abolish the activity. Substitutions of no more than a single residue were, therefore, considered. We substituted the P_1-Asn82 residue for Pro or Ala, but also the Val81 residue for Pro. The last exchange was done because we did not know the function of Asn82 in inhibition, and reasoned that increasing the rigidity of the loop could make it a less favoured proteolytic substrate.

Within the first loop of domain 3 of equistatin a fourth cleavage site was found at Arg151. The function of the third domain of equistatin is not known and likely not relevant as an inhibitor of insect proteases. Therefore, the Arg-residue was exchanged with the most neutral amino acid Ala. However, also, as a second strategy the loop N-terminal to Arg151 was extended with the addition of the VNNV in order to extend the alpha-helix and thus spatially protect Arg151 from proteases.

Expression in Pichia pastoris and determination of the inhibitory properties

The various mutations were combined in 8 different ways resulting in equistatin variants EIS1-8 (Figure 1B). The genes were first expressed in *Pichia pastoris* and purified according to a previously developed procedure (Outchkourov et al., 2002) (Figure 2).

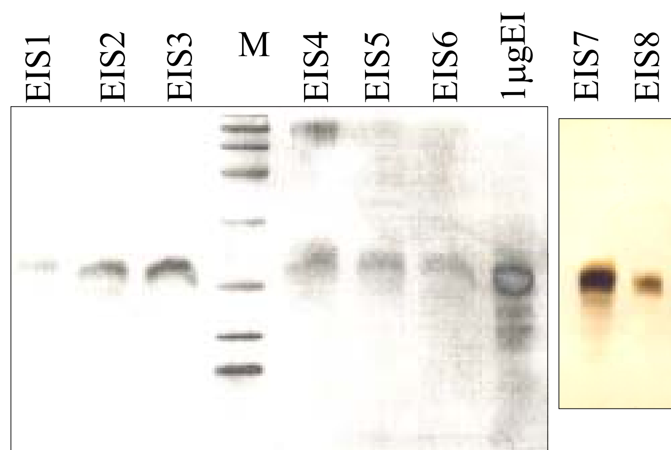


Figure 2. Purified equistatin variants expressed in *P. pastoris*. On the left, Coomassie-stained SDS-PAGE of the recombinant EIS 1-6 constructs together with the wild type equistatin protein. On the right, silver stained SDS-PAGE of the EIS 7 and EIS 8 proteins.

The EIS 1-6 proteins were subsequently titrated against papain in order to determine the constant of inhibition (K_i) of the engineered recombinant proteins (Figure 3). For all inhibitors the inhibitory constants were found to be close to the ones obtained for the wild type protein (K_i = 0.57 ± 0.04 nM). Both EIS3 and EIS4 were used as internal controls, since they did not contain amino acid substitutions in domain 1, which is known to be involved in the inhibition of cysteine proteases (EIS3, K_i = 0.93 ± 0.14 nM; EIS4, K_i = 1.84 ± 0.61 nM). The titration of most of the other equistatin variants (EIS1; EIS5; EIS6) resulted in K_i values that did not differ significantly from the values obtained for EIS 3 and EIS 4 indicating that the amino acid substitutions

Asn13 to Pro and Lys56 to Glu did not change the inhibitory properties against papain. In contrast in EIS2, the larger number of amino acid substitutions derived from the homologous p41 protein resulted in a significant increase in the K_i value to 2.6 ± 0.59 nM, which is about 2-fold higher than the average of the others. This small reduction in affinity of the inhibitor for the target proteases is likely attributable to the larger number of amino acid substitutions and also fits the fact that p41 is a slightly weaker inhibitor of papain (K_i = 1.4 nM) (Bevec et al., 1996; Lenarcic and Turk, 1999).

Cathepsin D inhibition was tested against the EIS1 and EIS8 recombinant proteins, without determining the K_i values (Figure 4). In EIS1 Val81 was changed to Pro, and in EIS8 Asn82 was exchanged to Ala. Both inhibitors retained the ability to strongly inhibit the proteolysis of Cathepsin D, which indicated that the protein had retained its functionality. EIS7 was not tested towards cathepsin D, but since the two other variants EIS1 and EIS8 have similar substitutions to EIS7 it is likely that EIS7 inhibits cathepsin D as well. Overall, therefore, all EIS proteins can be regarded to have retained their properties as tight functional inhibitors of both papain and cathepsin D.

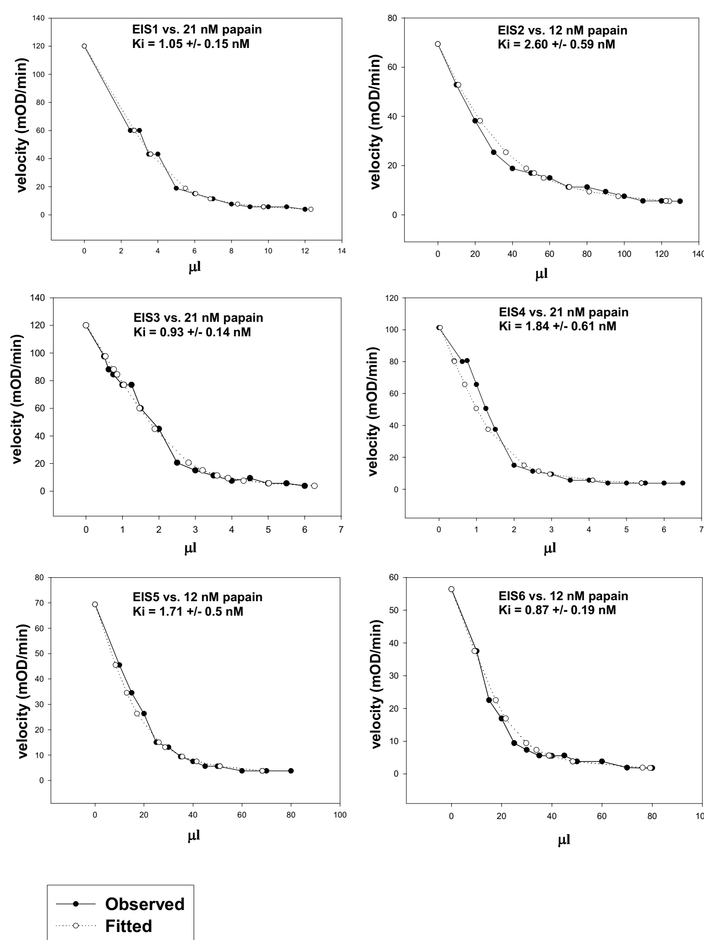


Figure 3. Titration of the inhibitory activities against papain of the EIS 1-6 equistatin mutant proteins using Z-Phe-Arg as substrate. The concentration of papain and experimentally determined K_i values are shown for each of the experiments.

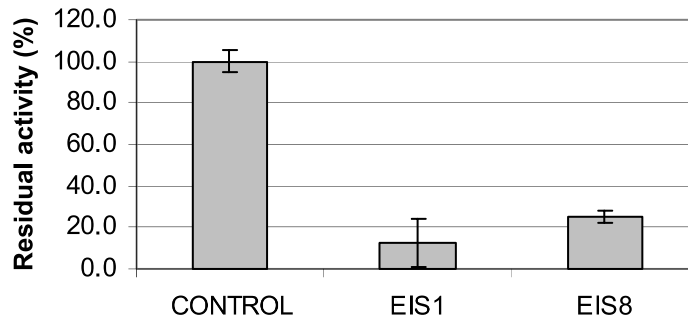
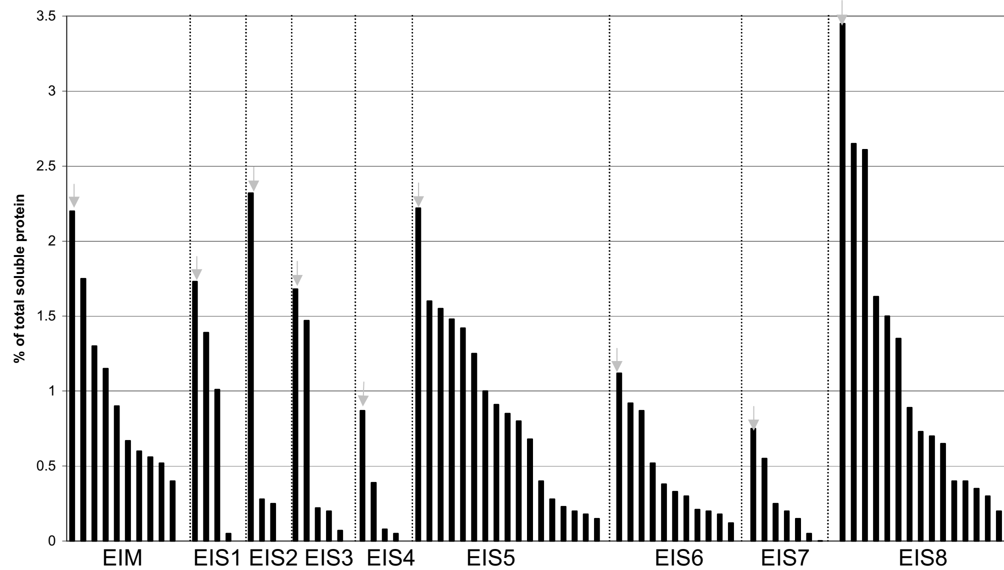


Figure 4. Inhibitory activities of the EIS 1 and EIS 8 proteins at equimolar concentrations of 10 nM against cathepsin D using H-Pro-Thr-Glu-Phe-Phe-(NO_2)-Arg-Leu-OH as a substrate.

Expression in plants

The constructs containing the different EIS1-8 variants plus EIM as a control were introduced into potato plants *Solanum tuberosum* cv. Impala. The EIM gene was in addition introduced into the potato cv. Kondor. The expression levels of all transgenic potato plants (10-30 plants per construct) were screened in first round by dot blot analysis (Figure 5) in order to select a few high-expressing plants for each construct. The selected plants were subsequently analysed by western blot analysis (Figure 6). Previous western blots of wild type equistatin expressed in potato cv. Desirée always yielded a distinct protein degradation pattern (Outchkourov et al., 2003b). In the cultivar Impala, which was the cultivar chosen for these experiments, the degradation products of equistatin were not observed, however. In Impala the only visible band was the top intact 22 kD band. This band was strongly reduced in intensity at the end of the day, when the rubisco promoter was much less active. In Impala, therefore, the exo- and endopeptidase proteolytic activities appear to degrade the primary cleavage products so fast that they do not remain visible as degraded bands. Genotype-linked factors can probably explain these differences between cultivars. Furthermore, the high levels of equistatin accumulation (up to 7% of total soluble protein) achieved in Line V and Desirée (Outchkourov et al., 2003a; Outchkourov et al., 2003b), could not be achieved in the cultivars Impala and Kondor. This is in support of the suggestion that proteolytic activities in the cultivars Impala and Kondor are higher than in the cultivars we employed previously (Line V and Desirée).

A



B

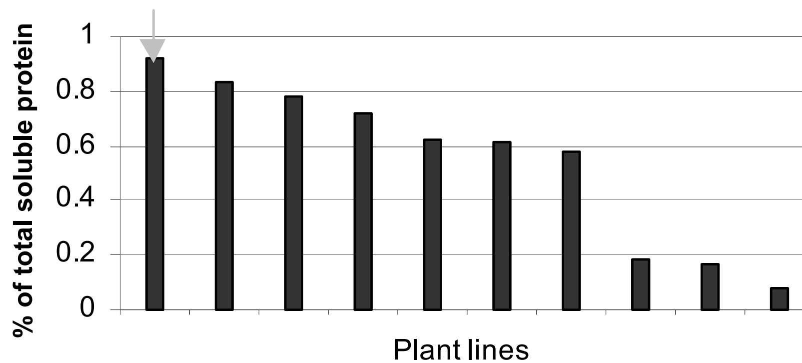


Figure 5. A. Expression levels of transgenic potato cv. Impala containing the EIM, EIS 1, EIS2, EIS 3, EIS 4, EIS 5, EIS 6, EIS 7, and EIS 8 constructs. **B.** expression levels in potato cv. Kondor transformed with the EIM construct. Arrows indicate the plant lines selected for western blot.

In Impala the gene variants EIS1, EIS5 and EIS6, all of which contained mutations in domain-loop 1.1 and 1.3 and 2.1 (plus mutations in 3.1 for EIS 5 and EIS 6) did not show the intact band, in contrast to the original EIM gene (Figure 6). This suggested

that some of the mutations actually *increased* the susceptibility of equistatin to proteolytic degradation.

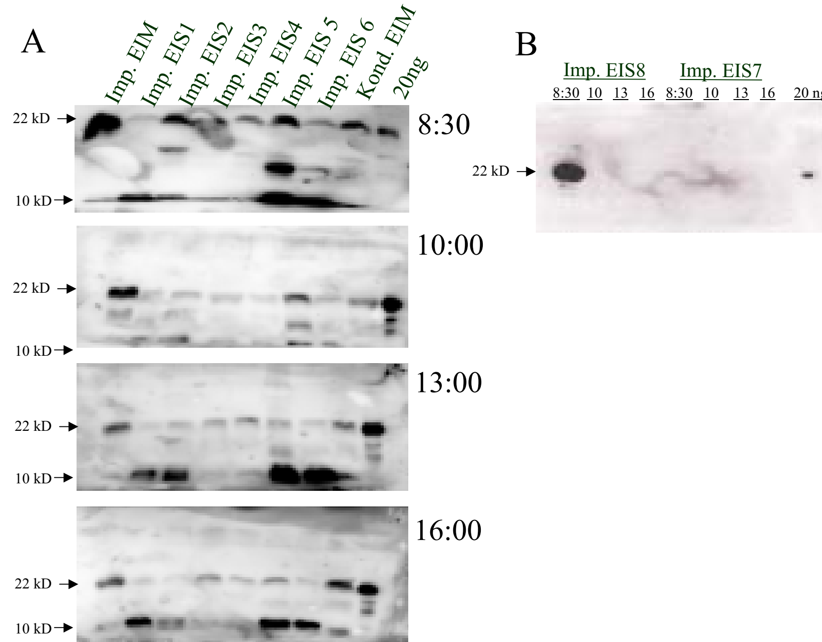


Figure 6. Western blot analysis of the plant lines expressing the equistatin variants. Sampling times are indicated. Imp. = potato cultivar Impala; Kon. = potato cultivar Kondor. A. Analysis of EIS constructs 1-6. B. Analysis of EIS 7-8.

Remarkably, however, the lower thick protein band of ~10kD in EIS1, EIS5 and EIS6 was much more stable than any part of the wild type protein EIM. This is evidenced by the fact that the intensity of that band remained equally strong at the various time points also when the promoter activity was low (16:00 hours)(Figure 6A). This demonstrated that part of domain 1 or 2 was more stable, even though it remains unclear which mutations were favourable. It meant that the basic strategy we followed was to some extent successful, since the constructs EIS1, EIS5 and EIS6 contained identical mutations in domains 1 and 2, while the accumulation of a similar proteolytically resistant band was observed. In the constructs EIS5 and EIS6 at 8:30 a.m. the samples have also accumulated a band with molecular size of ~16kD (thicker in the case of EIS5) which was not present in EIS1, but disappeared in the other time course samples. This indicates that also the mutations in domain-loop 3.1 improved the stability of that domain somewhat, especially in the case of EIS5. Overall, the results demonstrate the possibility to stabilise parts of the protein by changing amino acids in a way that takes into account information on the structure of the protein. The proper substitution, however, seems to be difficult to predict. EIS 7 plants had the lowest levels of expression as measured by dot blot, and the protein was even undetectable in Western blots. The difference between the detection between western and dot blots is explained by the 100-fold differential sensitivity of the antisera to native (Dot blot) and denatured (Western blot) protein. Also EIS 8 was not detectable at time points later than 8:30 AM (Figure 6). The constructs EIS7 and EIS8 differed slightly from the EIS5 construct, in the sense that both EIS7 and 8 contained two residues differences in domain 2 compared to EIS5 (Figure 1). These slight

differences were sufficient to completely destroy the protein *in planta* (Figure 6). Possibly, therefore, the stable ~10kD band in the EIS1, EIS5 and EIS6 constructs was due to the identical mutation between all three constructs in domain 2 which was destroyed in constructs EIS7 and EIS8.

An alternative explanation for the very fast turnover rate of EIS7 (and -8) might be that the amino acid substitutions along the polypeptide chain created a sequence motif that targets the entire protein for degradation. Many short-lived proteins are targeted for degradation by certain sequence elements, known as degradation signals (Sadis et al., 1995). The heterogeneity of these signals makes their functional identification a rather difficult task. These degradation signals target the entire polypeptide chain into the ubiquitin-dependent degradation pathway. The very fast disappearance of EIS 7 and EIS 8, without a visible pattern of degradation might indicate that these two proteins are targeted for degradation through this pathway. Some support for this hypothesis may be found in the study of (Ramos et al., 2001), where the 13- amino acid domain II consensus sequence (QVVGWPPVRSYRK) was found to be sufficient to target recombinant firefly luciferase for degradation in plants. The core amino acids GWPPV in this sequence are required but not sufficient, since the alanine substitutions in the other parts of the consensus partly increase the stability of the protein. The sequence where the differences between the EIS 5, EIS 7 and EIS 8 were made was QAIIPNVPGWCGPP for the EIS 5 (most stable), QAIIVPVPGWCGPP for the EIS 7 (least stable), and QAIIVAVPGWCGPP for the EIS 8 (moderate relative stability). The differences in the last three are underlined. The sequence motifs in the EIS 5, EIS 7, and EIS 8 are vaguely similar (bold residues) to the motif that was found by (Ramos et al., 2001) to target proteins for degradation, and could provide an alternative explanation for the rapid degradation.

Conclusion and recommendations

The strategy to substitute amino acids to improve the proteolytic stability has been used in a number of cases (Audtho et al., 1999; Frenken et al., 1993; Kapust et al., 2001; Markert et al., 2001; Varallyay et al., 1998). Until now this strategy had never used been to improve the stability against degradation of heterologously expressed proteins in transgenic plants. Using this strategy we achieved the accumulation of fragments which were more resistant to proteolysis than the wild type protein. This indicates that our strategy could yield increased protein stability against proteolytic degradation if we succeed in making the right substitutions in the right amino acids. Most of the substitutions appear to have had a negative impact on protein stability. A more efficient approach to this problem is, therefore, needed. In our case, there were multiple unknown proteases cleaving the protein at multiple partially known positions. In such a situation creating a stable protein is a numbers game, and more random strategies efficient at selecting the stable, functionally active mutant proteins may be more successful.

CHAPTER 7

ENGINEERED MULTI-DOMAIN CYSTEINE PROTEASE INHIBITORS YIELD RESISTANCE AGAINST WESTERN FLOWER THRIPS (*FRANKLINIELLA OCCIDENTALIS)* IN GREENHOUSE TRIALS

Nikolay S. Outchkourov[1], Willem Jan de Kogel[1], Gerrie L. Wiegers[1], Magnus Abrahamson[2], Maarten A. Jongsma[1]

[1]Plant Research International, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands

[2]Department of Clinical Chemistry, Institute of Laboratory Medicine, Lund University Hospital, S-221 85 Lund, Sweden

Submitted for publication

ABSTRACT

Western flower thrips, *Frankliniella occidentalis* (Pergande) (Thysanoptera: Thripidae) causes very large economic damage on a variety of field and greenhouse crops. In this study plant resistance against thrips was introduced into transgenic potato plants through the expression of engineered multi-domain protease inhibitors consisting of representative classes of cysteine and aspartic protease inhibitors: kininogen 3rd domain (K), stefin A (A), cystatin C (C), potato cystatin (P) and equistatin (EIM). Four (K-A-C-P) to five (EIM-K-A-C-P) inhibitors were fused into one single reading frame, and were shown to fold into functional protein in *P. pastoris.* In a bioassay using 3 µM purified EIMKACP protein (15 µM papain inhibitor activity) the oviposition rate of thrips was reduced by 70%. At a similar activity concentration the individual internal domains equistatin and stefin A reduced oviposition only by 47% and 55% respectively. The multidomain proteins were expressed in potato and found to be less susceptible to degradation by plant proteases than the individual domains. In a time span of 14-16 days transgenic potato plants expressing EIMKACP and KACP at a similar concentration reduced the number of larvae and adults to less than 20% of the control. Leaf damage on protected plants was minimal. Engineered multi-domain cysteine protease inhibitors, thus, provide a novel way of controlling western flower thrips in greenhouse and field crops.

INTRODUCTION

Western flower thrips (WFT), *Frankliniella occidentalis* (Pergande) (Thysanoptera: Thripidae) is a highly polyphagous insect (Jensen, 2000). It causes considerable economic losses on a large number of field grown vegetable, fruit, ornamental and plantation crops (Lewis, 1998). In addition WFT is the most prevalent pest in greenhouses throughout the world (Parrella, 1995; Shipp et al., 1998b). The small size of WFT (1-2 mm), its sexual habit (arrhenotoky), and short generation time, combined with its polyphagous nature (Loomans et al., 1995), its resistance to many pesticides (Jensen, 2000) and feeding habit in the inner whorls of flowers and buds make this insect extremely difficult to control.

Proteinase inhibitors (PIs) are proteins that form complexes with proteases and inhibit their proteolytic activity. Plants utilize PIs in order to moderate the adverse effects from attacking herbivores, or pathogens (Atkinson et al., 1995; Green and Ryan, 1972; Jongsma et al., 1994; Koiwa et al., 1997; Lorito et al., 1994). Insect herbivores have evolved many adaptive mechanisms that allow them to circumvent plant defences in general (Karban and Baldwin, 1997; Kessler and Baldwin, 2002). Mostly, adaptation to the presence of PIs involves the induction of novel proteases insensitive to PIs in the diet (Bown et al., 1997; Jongsma et al., 1995; Volpicella et al., 2003), or the expression of proteases that degrade the PIs (Giri et al., 1998; Michaud et al., 1995b). Pyramiding different protease inhibitors from non-host origin in transgenic plants that target the entire complex of insect proteases has been proposed as a method to combat the adaptive process in the insect pests (Jongsma and Bolter, 1997; Jongsma et al., 1996).

We previously established that the proteolytic activity of WFT extracts has an optimum at pH 3.5 and is inhibited by protease inhibitors specific for cysteine proteases. The conclusion that cysteine proteases predominate in the WFT digestive tract was supported by a strong reduction in WFT oviposition rate, when purified potato cystatin and equistatin were fed to adult females (Annadana et al., 2002b). Subsequently, four different cystatin members were expressed in potato plants, and the effects on thrips food preferences were studied (Outchkourov et al.(chapter 4)). The results indicated that stefin A is highly deterrent to thrips, followed by equistatin; while cystatin C and kininogen D3 were not deterrent. The deterrent effect of stefin A, in contrast to the other three inhibitors, could not be explained by better inhibition of specific proteases in thrips guts. Instead, the functional stability of the expressed inhibitor *in-planta* seemed critical in yielding thrips deterrent effects. Previously, kininogen D3 and cystatin C were found to inhibit the degradation of equistatin in potato plant extracts (Outchkourov et al., 2003b). The co-expression of all inhibitors together was, therefore, an attractive possible approach to stabilise them *in-planta*. At the same time, this would target a broad spectrum of thrips proteolytic activities and possibly result in broader and more durable pest control.

Functional co-expression of five different protease inhibitors in plants is technically challenging if each one is expressed separately. We, therefore, engineered a single open reading frame that encoded four or five inhibitors in two multi-domain proteins coined KACP and EIMKACP, and expressed them in yeast and potato. The inhibitory activity and anti-thrips effects were first assessed *in vitro* with purified protein, and, subsequently, with transgenic potato plants in greenhouse trials.

The results described here demonstrate the potential of the multi-domain technology to engineer insect resistance traits, but also to reduce unwanted proteolysis of heterologous proteins in transgenic plants.

MATERIALS AND METHODS

Oligonucleotides

EIMKACP(N)
5'- AAAAACTCGAGAAAAGAGAGGCTGAAGCTAGTCCCATGGATCCAAGTCTAACG
EIMKACP (C)
5'- AAAAAGCGGCCGCTTAGCATGTGGGACGTTTGAATCTG

Preparation of multi-domain inhibitors EIMKACP and KACP

For the construction of the multidomain constructs, previously developed single domain constructs were used: pUCRBC-EIM-SP, pUCRBC-**K**, pUCRBC-**A** pUCRBC-**C**, pUCRBC-**P** (Outchkourov et al.). The Stefin A fragment was excised from pUCRBC-**A** after *Bam*HI-*Bcl*I digestion and inserted into the *Bcl*I site of pUCRBC-**K** to create pUCRBC-**KA**. Then the cystatin C fragment was excised from pUCRBC-**C** after *Bam*HI-*Bcl*I digestion and inserted into the *Bcl*I site of pUCRBC-**KA** to create pUCRBC-**KAC**. Then the potato cystatin fragment was excised from pUCRBC-**P** after *Bam*HI-*Bcl*I digestion and inserted into the *Bcl*I site of pUCRBC-**KAC** to create pUCRBC-**KACP**. The pUCRBC-**KACP** was then digested with

*Bam*HI-*Bcl*I to remove the four domains KACP and the KACP fragment was cloned into the *Bcl*I site of the equistatin expression construct pUCRBC-EIM-SP (Outchkourov et al.) to create pUCRBC-EIM**KACP**. The two expression cassettes containing rbcS1 promoter-KACP or EIMKACP-rbcS1 terminator were excised from the pUC-based vectors by digestion with *Asc*I-*Eco*RI and inserted into the multiple cloning site of the pBINPLUS vector (van Engelen et al., 1995) that was digested with the same enzymes to create pRBC**KACP** or pRBC**EIMKACP**.

Transformation of potato

All pBINPLUS constructs and empty pBINPLUS (as control) were mobilised into *A. tumefaciens* AGL0 (Lazo et al., 1991) by electroporation. Potato plants (*Solanum tuberosum cv.* Impala) were transformed according to (Hoekema et al., 1989) using tuber discs as a source of explants.

Expression in *P. pastoris* and subsequent protein purification and characterization

The coding region of EIMKACP was amplified by PCR with the primer pairs EIMKACP(N) and EIMKACP(C) using PWO polymerase (Roche). The PCR product was purified on agarose gel and digested with *Xho*I-*Not*I restriction endonucleases. The digested fragment was gel-purified again and inserted into the pPIC9 vector digested with the same enzymes. The constructed vector was linearised with *Sal*I and transformed into *P. pastoris* strain GS115 (*his4*) (Invitrogen, Carlsbad, USA) by electroporation using a Gene Pulser electroporator (Bio-Rad, USA). Transformants were selected on MD agar plates (1.34% yeast nitrogen base, 4×10^{-5}% biotin, 1% D-glucose, 1.5% agar). Six his^+ colonies were analysed for expression. Shake flask expression studies were performed by pre-growing cultures in a shaking incubator at 250 rpm for 48 hours at 30°C in 1 litter baffled shake flasks containing 100 ml BMG medium (1.34% yeast nitrogen base, 4×10^{-5}% biotin, 1% glycerol, 0.1 M K-phosphate, pH 6.0). The cultures were centrifuged and the pellet containing the cells was resuspended in 100 ml BMM (1.34% yeast nitrogen base, 4×10^{-5}% biotin, 0.5% methanol, 0.1 M K-phosphate, pH 6.0) and grown under the same conditions for 48 hours. SDS-PAGE and western blot analysis were used to monitor the accumulation of the EIMKACP protein in the supernatants. A selected, high expressing clone was used for larger scale EIMKACP protein production in a bioreactor. The bioreactor production was carried out as previously described for equistatin (Outchkourov et al., 2002). At the end of the bioreactor run the cells were removed from the culture by two centrifugations at 28,000 x g. Cell-free supernatant was filtered through a 0.45 μm filter, and $(NH_4)_2SO_4$ was added to a final concentration of 0.5 M. This supernatant was applied to a 200 ml Phenyl Sepharose column. The column was washed twice with two column volumes of 0.5 M and 0.25 M $(NH_4)_2SO_4$ respectively and bound EIMKACP was eluted with demineralized water. The conductivity of 20 ml fractions was monitored. A further purification step using anion exchange chromatography was performed on a HiTrap Q Sepharose 5 ml column. The pH of the water fractions from the Phenyl Sepharose column was adjusted to 8.5 by adding 2 M Tris to a final concentration of 16 mM and this solution was applied to a 5 ml HiTrap Q-Sepharose

column. After washing with five column volumes of 20 mM Tris pH 8.5, stepwise elution was performed with 50 mM, 100 mM, 150 mM and 200 mM NaCl, respectively in 20 mM Tris pH 8.5. The fractions were analysed on 10% SDS-PAGE, and the one containing the undegraded molecular size band were pooled. The buffer was exchanged to water on Vivacell 70, 5.000 MWCO filters (Vivascience, Hanover, Germany). Quantitation of the inhibitor activity against papain and cathepsin D was performed as described in (Rogelj 2000).

Artificial diet bioassay of WFT

Adult female thrips of mixed ages were collected with an aspirator from a population of western flower thrips (WFT) *Frankliniella occidentalis* reared on flowering chrysanthemum (*Chrysanthemum morifolium* Ramat.) cv. Sunny Casa in a greenhouse at 25 °C. Purified inhibitors equistatin EIM (Outchkourov et al., 2002), Stefin A (Kristina Gruden; Jozef Stefan Institute, Slovenia) and EIMKACP (this study) were used for subsequent bioassay. The experiment was conducted in perspex ring cages as described in (Annadana et al., 2002b). Nine repetitions were used for the control without inhibitor, and six for each of the treatments with inhibitor. The adult females (10) were allowed to adapt to the diet for five days and the subsequent addition of the inhibitors continued for a period of eight days. For both the pre-treatment and treatment periods systematic differences in the mean numbers of eggs observed per cage per day between different experimental treatment sets were investigated using one-way ANOVA. Differences in average egg production between the pre-treatment and treatment period were not analysed, since profiles of the control (see results) indicate that the differences depend on the length of pre and post treatment period and are quite variable in time. In the ANOVA differences between treatments were assessed using an F test for the variance ratio. If the F-test is significant, the null hypothesis H_0:' mean number of eggs of control not different from inhibitor treatment' was tested against the alternative hypothesis H_1: 'mean number of eggs of control larger than mean number at applied dose' was assessed using a one-sided t-tests at size α=0.05. The same procedure was followed for differences between means at the high and low inhibitor dose.

Dot blot analysis

Leaf discs were punched into separate wells of a 96-well microtiter plate and to each well 200 μl of protein extraction buffer (80 mM Tris-Cl, pH 7.6, containing 25 mM diethyldithiocarbamate, and 50 mM Na_2EDTA) was added. The samples were then three times frozen to -20°C and thawed to room temperature. The extract was pipetted to a new microtiter plate, separating it sufficiently from the insoluble leaf material. For dot blot analysis 2.5 μg of protein extract obtained with the freeze-thaw extraction procedure were transferred to Trans Blot (BioRad) nitrocellulose membranes using the SRC 96 D Dot blot apparatus (Schleicher & Schuell, Dassel, Germany). Subsequently, the membranes were blocked in TTBS (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween 20) containing 2% non-fat milk powder for 1 hour, and then incubated with rabbit anti-K or anti-C (Abrahamson et al., 1986). The blots were subsequently washed and incubated with goat anti-rabbit IgG conjugated with

horseradish peroxidase (Jackson Immuno Research, USA). The membranes were visualized with Lumi-Light Western blotting substrate and scanned in the Lumi-Imager F1^{TM} under the control of Lumi-Analist^{TM} software (Roche).

Western blot analysis

A piece from a fully expanded young leaf (200-300 mg) was placed in 1.5 ml eppendorf tube and ground in liquid nitrogen to a fine powder. The powder was resuspended in 300 µl extraction buffer (100 mM Tris-HCl, pH 7.6, 25 mM Na-diethyldithiocarbamate, 50 mM EDTA, and 10 % PVPP) and this crude extract was twice centrifuged for 5 min at 14000 rpm and 4°C. Each time the supernatant was replaced into a new tube. SDS-PAGE was performed using a 12% precast resolving gel (BioRad, Hercules, CA, USA) on a mini-Protean II slab cell apparatus (BioRad Hercules, CA, USA). The gels were run according to the manufacturer's instructions. The separated protein samples from the SDS-PAGE gels were transferred to Trans Blot (BioRad) nitrocellulose membranes using the mini-Protean II electrotransfer apparatus (BioRad). Blocking of the membranes and the visualisation was performed the same way as described for Dot blot.

Greenhouse bioassay of WFT

Rooted cuttings from four *in-vitro* propagated transgenic potato lines (KACP11; KACP23; EIMKACP6; A9) were adapted to soil in 2 litre plastic pots. The plants were grown in the greenhouse under a temperature regime of 20/18 °C and a photoperiod of 16h/8h day/night. Three days before the bioassay the pots together with the plants were placed into large Perspex cylinders (80 cm tall, 20 cm diameter) closed on top with nylon gauze of 120 µm mesh size. Two 15cm diameter holes on the sides of the cylinder closed with the same gauze ensured sufficient aeration of the plants. Ten adult female thrips derived from a population of WFT maintained on flowering chrysanthemum (*Dendranthema grandiflora*) cv. Sunny Casa in a greenhouse were inoculated on each plant. At least five transgenic plants per line were tested. The experiment was repeated twice for periods of 14 and 16 days. At the end of the experiment the plastic cylinders were removed carefully from the plants, the plants were cut close to the soil and the aerial part of the plant was submerged into a 5 litre solution of 70 % ethanol. After about 30 seconds each plant was shaken and removed from the 70% ethanol solution. The ethanol solution was subsequently passed through filter paper. The numbers of adult and larval thrips retained on the filter were counted separately under a binocular. Differences in the number of thrips between the different treatments compared to the control were investigated by using one-sided t-tests at size α=0.05.

RESULTS

Construction and characterization of the new multi-domain EIMKACP inhibitor

In an attempt to mimic natural multidomain inhibitors we fused the cysteine and aspartic protease inhibitor equistatin (EIM) to four single domain cysteine protease

inhibitors (K, A, C, and P), thereby, creating an open reading frame of 708 amino acids (EIM-K-A-C-P). Equistatin naturally consists of three domains: domain I inhibits cysteine proteases, domain II inhibits aspartic proteases, and domain III is of unknown function (Lenarcic and Turk, 1999; Strukelj et al., 2000). Kininogen domain 3 (K), stefin A (A), cystatin C (C) and potato cystatin (P) belong to the cystatin super family and occur either as independent proteins (A and C), or as parts of larger multidomain proteins (K and P). They inhibit papain-like cysteine proteases (Annadana et al., 2003; Rawlings and Barrett, 1990). In addition, cystatin C and kininogen possess a second reactive site at the "back" of the protein, which is involved in the inhibition of asparaginyl-specific cysteine proteases called legumains (Abe et al., 1993; Alvarez Fernandez et al., 1999; Outchkourov et al., 2003b). This structural feature allows simultaneous inhibition of both papain- and legumain-type proteases (Alvarez Fernandez, 1999). Thus, in total, EIMKACP possesses 8 known inhibitor reactive sites: 5 inhibitors of papain-like cysteine proteases, 2 inhibitors of legumain-like cysteine proteases and 1 inhibitor of the aspartic protease cathepsin D. The different proteins in EIMKACP were linked by five glycine residues to allow flexible movement of the domains with respect to each other. When expressed and purified from *Pichia pastoris* the recombinant EIMKACP multi-domain protein was tested for inhibition against cysteine papain) and aspartic cathepsin D) proteases (Fig. 1).

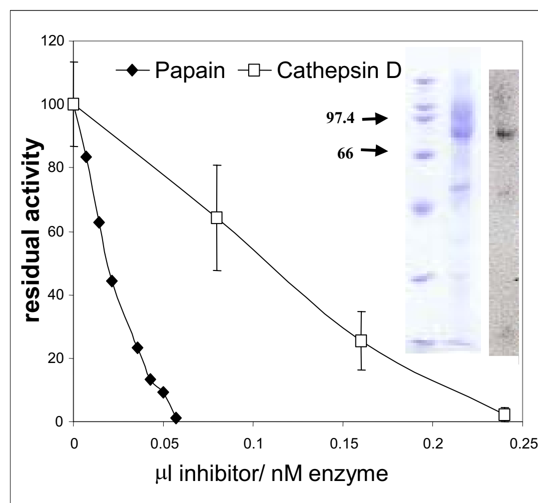


Figure 1 Titration of partially purified EIMKACP protein against papain and Cathepsin D as described in Materials and Methods. The inset shows a coomassie stained gel (1) and Western blot analysis (2) as evidence of the quality of the purified protein.

A plot of the results of both experiments after normalizing the concentrations of active protease in each experiment, revealed that five-fold less EIMKACP protein was needed to inhibit the same concentration of papain compared to cathepsin D (Fig. 1). Thus, as expected the EIMKACP protein possesses five active sites involved in the inhibition of papain-like cysteine proteases versus only one active site involved in the inhibition of the aspartic proteases (Cathepsin D). Apparently, also when expressed as part of a single reading frame, they are properly folded and can act independently of each other.

Effects of multidomain and single domain inhibitors on thrips fecundity

The effects on thrips fecundity of the purified EIM, Stefin A and EIMKACP proteins was analysed as described in Materials and Methods in an artificial diet bioassay (Fig. 2). During the pre-treatment period (day 1-5) thrips were reared on pollen and water. In the treatment period (day 6-13) the water was supplemented with equistatin, stefin A, and EIMKACP each at two different inhibitory concentrations. Analysis of variance of mean egg production for all of the experimental groups in the pretreatment period established a p-value (F-test) of 0.543, which is evidence for the absence of differences prior to treatment. In the treatment period the same analysis revealed a p-value (F-test) of 0.005, which demonstrates strong evidence for differences between treatments.

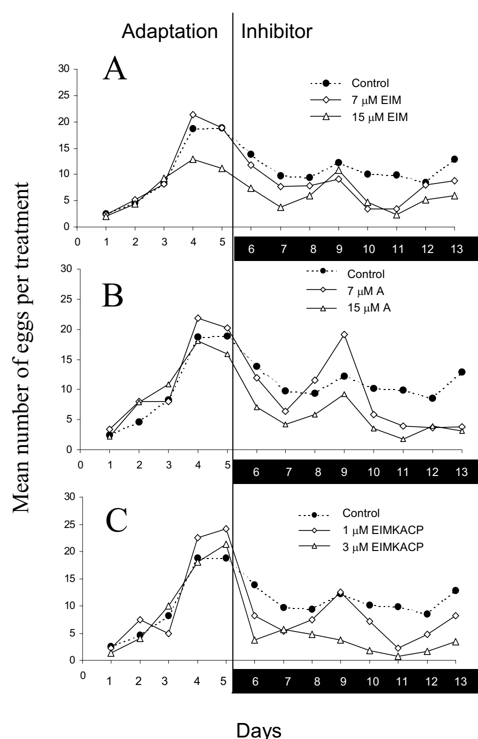


Figure 2 Effect of dietary cysteine protease inhibitors on the oviposistion rate of WFT. Day 1-5 (pre-treatment period) WFT were fed on pollen grains as a food source and water. Day 6-13 (treatment period) the water of WFT was supplemented with protease inhibitors. **A**- effects of 7 and 15 μM equistatin (EIM) in comparison to the control; **B**- effects of 7 and 15 μM stefin A (A) in comparison to the control; **C**- effects of 1 and 3 μM EIMKACP multi-domain protein in comparison to the control. The table summarises the results of the treatment and provides the statistical analysis.

	Concentration	Mean number of eggs (10 females/day)	One-side t-tests of pair-wise comparisons *			
Control	-	10.828	A			
EIM	7 μM	7.542		B	C	
	15 μM	5.771		B	C	D
A	7 μM	8.300	A	B		
	15 μM	4.833			C	D
EIMKACP	1 μM (5μM)	7.000		B	C	
	3 μM (15μM)	3.200				D

* Means with letters in common within a row do not differ significantly at size α=0.05

WFT females fed on control (water) produced on average 10.8 eggs/10 females/day during the post treatment period (=100%). Treatments with 7 and 15 μM EIM supplement to the water produced less 7.5 (70%) and 5.8 (53%) eggs/10 females/day. Similar values (8.3 (77%) and 4.8 (45%) eggs/10 females/days) were recorded when

stefin A solutions were added at the same 7 and 15 μM concentrations. In contrast, the addition of EIMKACP protein at 1 and 3 μM concentrations (papain inhibitor equivalents of 5 and 15 μM) resulted in 7 (65%) and 3.2 (30%) eggs/10 females/day. Pair-wise differences between treatments in the post treatment period were investigated by t-tests (Fig. 2). When higher inhibitor concentrations (papain inhibitor equivalents of 15 μM) were applied significant differences compared to the control were observed for all of the tested inhibitors (Figure 2). At the lower concentrations (papain inhibitor equivalents 7 μM EIM and Stefin A, and 5 μM EIMKACP) significant suppression of the egg production was found for EIM and EIMKACP, but not for stefin A. Differences between the higher (15 μM inhibitory equivalent) and the lower concentration (5 or 7 μM) were found to be significant for EIMKACP and stefin A but not for EIM. Overall, the results of the bioassay confirmed the strong ability of cysteine protease inhibitors to suppress fecundity of WFT. The observed reduction of the number of eggs correlated to the concentration of the added inhibitor. The multidomain EIMKACP protein was found to act synergistically compared to two of its internal domains when given alone (EIM and A).

Expression of multi-domain protease inhibitors in plants.

In order to ensure high expression levels in potato plants the multi-domain protease inhibitors EIMKACP and KACP were expressed using the powerful *rbcS1* expression cassette (Outchkourov et al., 2003a). In addition, the constructs contained a sequence at the N-terminus encoding the leader peptide of equistatin to ensure targeting into the secretory pathway, and a four amino acid sequence at the C-terminus (KDEL) for retention in the endoplasmatic reticulum.

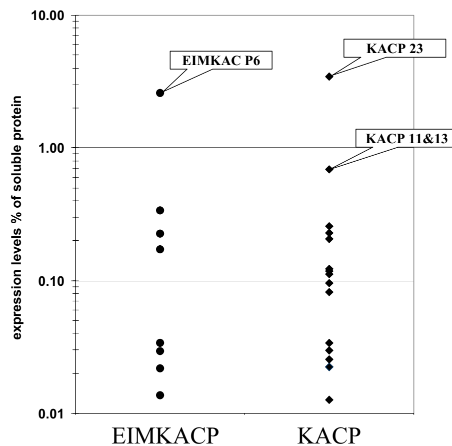


Figure 3 Distribution of expression levels found on potato plants (Impala) transformed with EIMKACP and KACP multidomain proteins. Only the plants with detectable expression (> 0.01%) are shown.

As indicated by dot blot analysis about half of the plants accumulated the proteins at detectable levels above 0.01% (Fig. 3), and only 1-2 among 20-30 transgenics yielded sufficiently high levels of > 0.5%: EIMKACP line 6 accumulated the heterologous protein at 2.6% of total soluble protein, KACP11 accumulated levels of 0.68%, and

KACP23 at levels of 3.5%. The integrity of the expressed EIMKACP and KACP proteins in the EIMKACP6 and KACP23 plant lines was further checked by Western blot using antibodies raised against cystatin C or kininogen. It was found that both proteins EIMKACP and KACP accumulated at the expected molecular weight without visible traces of degradation (Figure 4B). In contrast, the expression of the separate, individual protease inhibitors: equistatin, kininogen domain 3, and cystatin C (Outchkourov et al.) (Figure 4B) resulted in a number of degradation bands. Considering that a combination of sera raised against cystatin C and kininogen were used to visualize the multidomain proteins, degradation bands caused by the usual cleavages in these proteins and equistatin would surely have resulted in multiple immunologically reactive bands if the domains would have been cleaved in a way similar to the single domains. Thus, the multidomain proteins both seemed to be protected against the hydrolysis by plant proteases in contrast to the single domain inhibitors. Apparently, the inhibitors cross-protected each other against degradation.

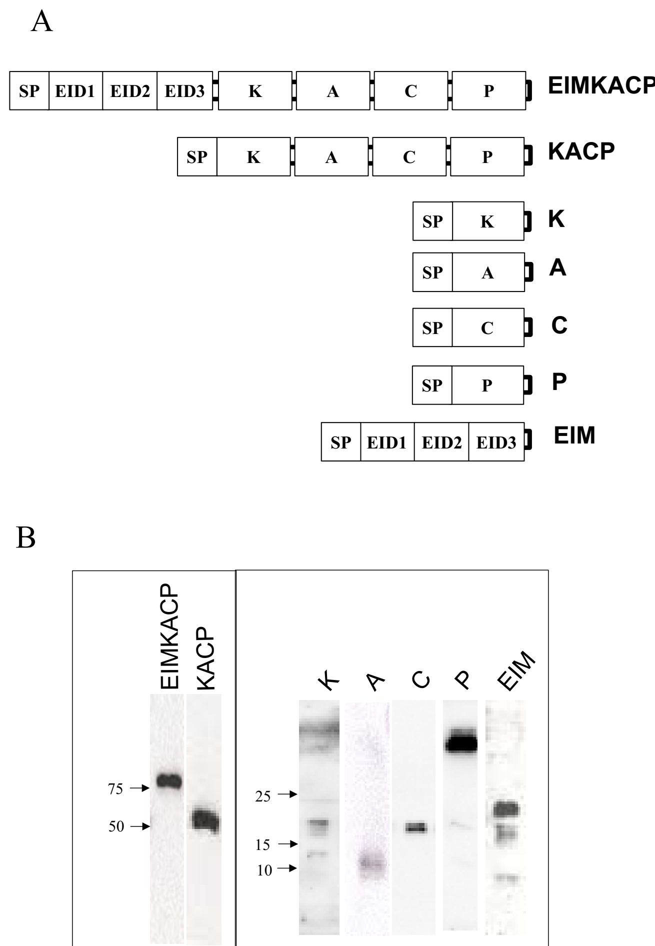


Figure 4 **A** Schematic representation of the different single and multi-domain constructs. **B** Western blot analysis of the transgenic plants expressing the single and the multidomain proteins. On the left, analysis of the EIMKACP6 and KACP23 transgenic lines expressing the EIMKACP and KACP proteins first separated on 10% SDS-PAGE. The results on the left are from (Outchkourov et al.)(first separated on 15% SDS-PAGE). Arrows indicate the approximate positions of the molecular weight standards in kilodaltons. Abbreviations: EIM- , equistatin optimized for expression in potato; P- potato cystatin; C- cystatin C; A- Stefin A; K- kininogen Domain 3; KACP and EIMKACP- multidomain proteins that contain the above single domain inhibitors.

Non-choice greenhouse trials

Previous experiments have shown that the effects of cysteine protease inhibitors on thrips are primarily to reduce the number of offspring and to act as a deterrent to the adults (Annadana et al., 2002b; Outchkourov et al. 2003(chapter 4)). To be successful as a resistance strategy, however, only non-choice whole plant bioassays can properly demonstrate the potential of these genes for crop protection. To assess the level of thrips resistance in the transgenic lines two independent thrips greenhouse trials were conducted as described in Materials and Methods (Figure 5).

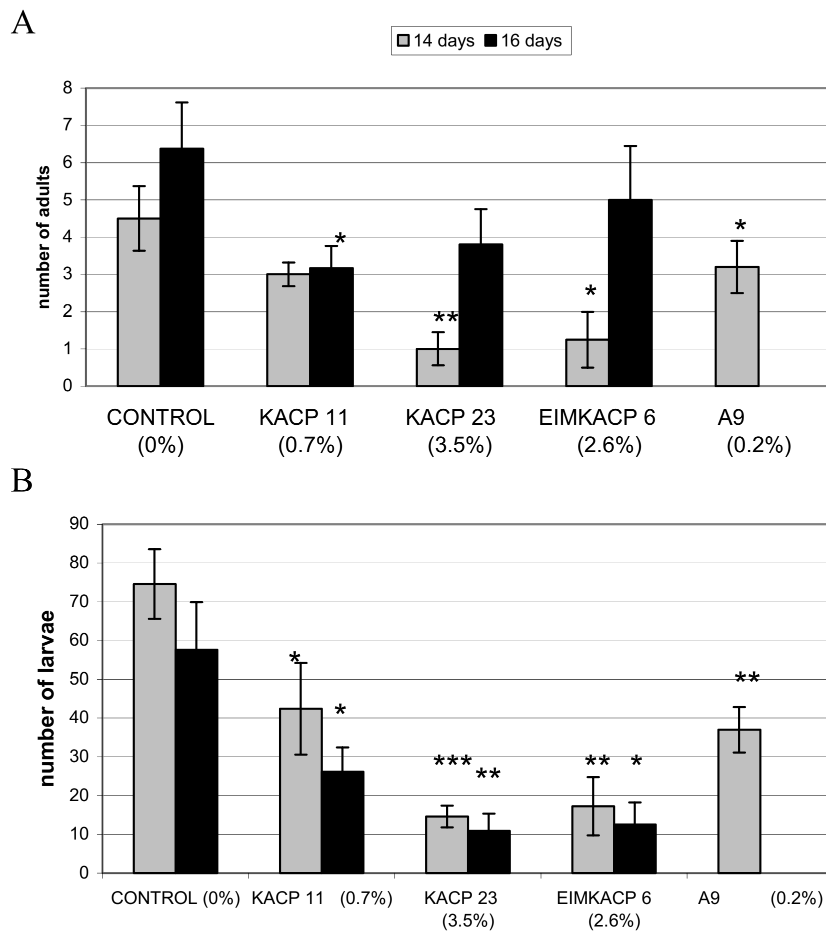


Figure 5 Effect of the expression of different cysteine proteinase inhibitors on thrips reproduction. Ten female thrips were initially inoculated per potato plant. The data is from two separate experiments lasting 14 and 16 days. **A-** number of adults**, B-** number of larvae. Stars indicate the level of significance: *** Extremely significant (t-test < 0.001); ** Very significant (t-test 0.001 to 0.01); * Significant (t-test 0.01 to 0.05).

Lines EIMKACP6 accumulating EIMKACP at 2.6% of total soluble protein, KACP11 at 0.68%, and KACP23 at 3.5% were selected for the experiment. In

addition, the plant line expressing stefin A (0.2% TSP), which was known to cause strong deterrence of thrips, was included in one of the two experiments (Outchkourov et al. (chapter 4)). The levels of heterologous protein accumulation were rechecked by dot blot analysis on the clonally propagated plants (greenhouse cuttings) prior the first thrips experiment. The variations in the levels found on clonally propagated lines were found to deviate not more than 15% from the initially established levels (data not shown). The experiments were designed to last 14 and 16 days in order to allow enough time for thrips to adapt and produce offspring. Longer periods would limit plant growth in the narrow cylinders and result in condensation of moisture where leaves would contact the cylinder wall and increase insect mortality. Also, this period allowed counting of all larvae before pupation in the soil. At the end of the experiment the number of thrips adults and larvae were counted separately. Adult thrips reared on transgenic plants expressing inhibitors produced fewer offspring and had reduced survival rates compared to the ones reared on control plants (Fig. 5A). In the first experiment lasting 14 days on average about 1 out of 10 females inoculated on the KACP23 and EIMKACP6 lines survived compared to 4 out of 10 on the controls. KACP23 and EIMKACP6 were the lines with highest levels of accumulation of heterologous inhibitor. Intermediate numbers of adults were found on the KACP11 and A9 plant lines that expressed inhibitors at lower levels. In the second experiment lasting 16 days the number of adults was again reduced on the plants expressing the cysteine protease inhibitors, however, this time there was no correlation to the initially measured expression levels.

The two experiments lasting 14 and 16 days produced on average similar number of larvae on all of the tested plant lines. On the controls they were 74 ± 9 and 57 ± 12 larvae for the 14 and 16 day experiments (Figure 5B). Significantly less (~50%) larvae in comparison to the controls were produced on the plant lines KACP11 and A9. The strongest reductions in the number of larvae (about 80%) were obtained with the plant lines KACP23 and EIMKACP6, both of which accumulated the inhibitors at the highest levels of 3.5% and 2.6% of total soluble protein. The actual numbers of larvae (14 and 12 for the two experiments on KACP23 and 17 and 15 for the two experiments on EIMKACP6) was in the range of the 10 initially inoculated adults, so that the population increase was significantly impaired on the plant lines EIMKACP6 and KACP23.

DISCUSSION

In this study representative members of the main known types of cysteine protease inhibitors: type 1 cystatin (stefin A), type 2 cystatin (cystatin C), type 3 cystatin (kininogen, domain 3) (Abrahamson et al., 1986; Alvarez Fernandez et al., 1999), phytocystatin (potato cystatin) (Annadana et al., 2002b), and thyropins (equistatin) (Lenarcic and Bevec, 1998), were fused into novel multi-domain protease inhibitors. To allow flexibility of the different domains, like in the case of the natural multidomain inhibitors (Barrette-Ng et al., 2003), five glycine residues were used as linkers in between the different domains. Based on our previous studies the inhibitors were expected to inhibit more than 95% of proteolytic activity in thrips guts (Outchkourov et al. 2003 (chapter 4)) and to possibly inhibit each other's degradation *in-planta* (Outchkourov et al., 2003b). Two combinations of multi-domain proteins

were prepared: EIMKACP, which contained all of the above inhibitors, and KACP, which contained all except equistatin. All domains of the EIMKACP protein produced in *Pichia pastoris* were found to be capable of simultaneous binding and inhibition of proteases. The effect of EIMKACP protein on fecundity of WFT was subsequently analysed in a bioassay. Previously we tested 30 μM concentrations of potato cystatin and equistatin (Annadana et al., 2002b) and found for both inhibitors approximately 50% reduction in the fecundity of WFT after 5 days. Here, we established that cumulative fecundity over eight days at 15 μM equistatin or stefin A also reduced fecundity by approximately 50% and that at 7 μM this value was around 25%. The EIMKACP protein at similar papain inhibitory concentrations (five fold lower molar concentration) reduced fecundity of WFT more strongly by 70% (15 μM) and 35% (5 μM). There may be several reasons why the multidomain inhibitor is more effective than the individual component domains. The component inhibitors were tested *in vitro* against thrips extracts (Outchkourov et al. (chapter 4)). There, addition of equistatin resulted in the strongest level of inhibition of WFT proteases, less efficient were kininogen domain 3, stefin A, and cystatin C (all required a 10-fold excess), while the weakest inhibition (100-fold excess required) was exhibited by potato cystatin. This demonstrates that thrips proteases are *in vitro* differentially sensitive to this set of inhibitors, but, surprisingly, equistatin is not performing any better than potato cystatin *in vivo*. Possibly, the insects overexpress less sensitive proteases when only one type of inhibitor is added to the diet or they degrade the inhibitor with other enzymes (Jongsma et al., 1996). The addition of a complex of inhibitors with different specificities may prevent the success of this type of adaptive response. Furthermore, multidomain inhibitors might have a smaller apparent dissociation constant than the individual domains due to the fact that the probability of rebinding to a neighbouring inhibitor domain is much larger compared to a situation where such domains are not physically linked.

Translating the artificial diet data to transgenic plants a concentration of 1 μM of purified EIMKACP or KACP proteins corresponds to about 0.5% of total soluble plant protein. This is near or just above the threshold level for effects on thrips. Therefore, only potato plant lines expressing EIMKACP and KACP at levels above 0.5% of total soluble protein (EIMKACP6, KACP23 and KACP11) were tested for increased WFT resistance. Bioassays with WFT on intact plants revealed that the population development of WFT is up to 80% suppressed on transgenic plants expressing the new cysteine protease inhibitors, and that the effects correlated to the levels of accumulated inhibitor. The small number of useful high expressing lines per construct did not allow a comparison of the constructs among each other. Based on our previous experiments using purified proteins we expect that the effects of the transgenic plants on the population development are largely due to the reduction in the WFT oviposition rate. Potential effects on larval development and survival could not be established in this type of experiment where larvae emerge continuously from eggs produced during the two week experiment. However, in the two experiments on the transgenic plants expressing EIMKACP, KACP and A9 significant reductions on the number of thrips adults were also observed, though they were less consistent as the reduction of the number of larvae. It is not clear whether the reduction of the number of adults were due to higher mortality or due to more successful escape events from the cage. Annadana et al. (2002) demonstrated on artificial diet that at least after five days adult females did not suffer increased mortality rates, while Outchkourov et al

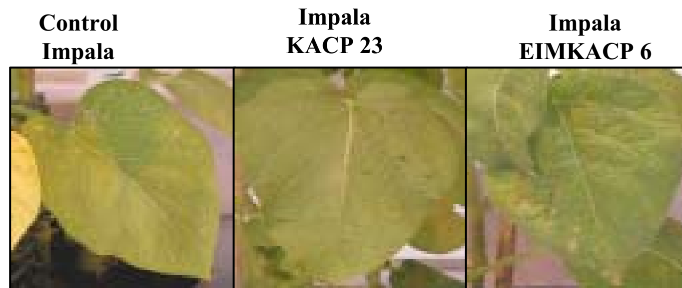
demonstrated that the single domain inhibitors stefin A and equistatin expressed in potato caused very strong deterrence of the adults. It is, therefore, likely that the transgenic plants besides directly reducing the fecundity of the insects due to reduced availability of protein, also induced a behaviour which resulted in fewer adults on the plants, and, therefore, fewer offspring. Further, and more controlled, experiments on the transgenic plants are now in progress to investigate this in more detail.

Pyramiding different protease inhibitor genes is of value not only for engineering resistance against insects, but has already proven useful in controlling plant parasitic nematodes (Urwin et al., 1993). In addition there are many examples of natural defensive multi-domain inhibitors, in which single proteins can bind multiple proteases, such as for example the 8-domain multicystatin from potato (Waldron et al., 1993b; Walsh and Strickland, 1993), the 6 domain PI2 from tobacco (Atkinson et al., 1993), and many more. This apparent multiple inhibitory function on a single reading frame may be advantageous since it increases the inhibitory potential and spectrum of a given protein, while it is saves resources on genetic control elements. Considering that even single domain proteins form multimeric complexes, however, strongly points to a functional role for this multiciplicity. Insect proteases were shown to occur as complexes in several cases (Terra and Ferreira, 1994). It has been suggested that the effect of combining multidomain or multimeric inhibitors with multimeric enzymes in equimolar amounts will be very similar to that of polyclonal sera and an antigen with multiple epitopes. A large protein complex will be formed that precipitates out of solution and low affinity binding is compensated by the presence of multiple binding sites (Jongsma and Bolter, 1997). The benefit of an artificial multi-domain protein for an engineer is that the expression construct remains simple even when several proteins must be co-expressed. The protein can be custom made for the specific requirements of the pest involved, although there does seem to be a cost in terms of the expression levels that can be attained with larger multidomain proteins, as we observed in our experiments. Only 2-3 plant lines accumulated the inhibitor to levels above 1% of total soluble protein out of population of 20-30 transgenic plants, in contrast to previous studies, using the same *rbcS1* promoter, but with equistatin alone, where about 50% of the generated transgenic plants expressed equistatin to levels above 1% (Outchkourov et al., 2003a). Thus, careful selection of a limited number of inhibitors that do not reduce significantly the expression levels could be the key to engineer feasible resistance traits using multi-domain inhibitors. One such additional domain could be the mustard trypsin inhibitor, which was recently optimised to fight aphids (Ceci et al., 2003). Thus, a resistance strategy based on the concept of multi-domain technology as found in nature as, might be useful to control many other pests. We envision that by including more serine and aspartic protease inhibitor domains super-multidomain protease inhibitors can be created which will target entire pest complexes of specific crops.

. **A**

Figure 6 A Picture of the set up of the whole plant greenhouse trial. **B**. Picture of the leaves of potato plants after the whole plant assay lasting 16 days. Representative leafs per each of the plant lines are

B



The application of this concept will need to solve some potential problems that lie ahead. We noticed that our multidomain inhibitors cross stabilized each other against degradation by plant proteases. This trait is potentially useful as plant proteases significantly reduce the yields of many economically interesting proteins that could be harvested from plants (Cunningham and Porter, 1998; Outchkourov et al., 2003b; Stevens et al., 2000). The inhibition of endogenous proteases did not result in a phenotype in the case of potato, but this may be different in other plant species. Clearly, with novel multi-domain proteins to be developed a major concern will be to prevent any potential interference with basic cellular functions by judicious choice of inhibitors and intelligent sub-cellular targeting or pest-induced expression.

CHAPTER 8

CHARACTERISATION OF TWO *DROSOPHILA MELANOGASTER* GENES ENCODING MULTIPLE PUTATIVE PROTEASE INHIBITOR DOMAINS

Nikolay S. Outchkourov[1], Stanimir Dulev[2], Ralph H.G.M. Litjens, Maarten A. Jongsma[1]

[1]Plant Research International, PO Box 16, NL-6700 AA Wageningen, The Netherlands

[2]Department of Plant Physiology and Molecular Biology, University of Plovdiv, Bulgaria

Prepared for publication

ABSTRACT

To obtain natural multidomain protease inhibitors with potential for insect pest control two *Drosophila melanogaster* genes *CG2264* and *CG5639* that contain multiple (3 and 12) modules similar to known protease inhibitors were cloned, expressed in yeast and partly characterized. The *CG2264* gene represents the fruit fly orthologue of the group of SPARC-related proteins, important in the regulation of cell-to-matrix interactions, and best studied in mammals. The mRNA of *CG2264* was found to contain alternative splice forms, not yet annotated in the sequence databases. In *Pichia pastoris* the sequence encoding the full-length 67 kD protein, resulted in accumulation of a 75 kD protein band on SDS-PAGE probably as a result of glycosylation. The protein was found to inhibit three different types of serine proteases (trypsin, chymotrypsin and subtilisin), and to inhibit the hydrolysis of two synthetic protease substrates, Z-Arg-Arg-pNA and Suc-Ala-Ala-Pro-Leu-pNA, in total protein extract from *D. melanogaster*. The *CG5639* protein domains are organized in a manner similar to lp11788p from *D. melanogaster* and c08g9.2 from *C. elegans* and the multiple thyroglobulin type-1 domains are reminiscent of the mammalian thyroglobulin precursor protein. The production of CG5639 in *P. pastoris* was hampered, probably due to the large size (166 kD) and the C-terminal trans-membrane domain, and resulted in the accumulation of smaller protein fragments which inhibited the proteolysis of three different synthetic substrates (Z-Arg-Arg-pNA, Z-Phe-Arg-pNA, and Suc-Ala-Ala-Pro-Leu-pNA) in total protein extract from *D. melanogaster*. The mRNA levels for both fly genes were found to vary considerably during fly development. At least one of the splice forms of *CG2264* shows a female-specific expression pattern in adult fruit flies and *CG5639* is mainly expressed during pupal development.

INTRODUCTION

Recently, the availability of the full sequence of the *Drosophila melanogaster* genome revealed the existence of several putative extracellular matrix proteins composed of an array of different protease inhibitor domains (Gelbart et al., 2002; Trexler et al., 2001). Two of them CG2264 and CG5639 contain multiple thyroglobulin type-1 domains that potentially inhibit cysteine and aspartic proteases (Lenarcic and Bevec, 1998). Thyroglobulin type-1 domains contain six cysteine residues including a characteristic CWCV motive (Molina et al., 1996). The structure of one thyropin (p41) in complex with Cathepsin L has been solved (Guncar et al., 1999). The inhibitory properties of p41 suggest a role as regulator of the cathepsin L activity (Rodriguez and Diment, 1995), yet also a role as a chaperone that helps to maintain a pool of mature enzyme when needed has been proposed (Lennon et al., 2001). CG2264 contains, in addition, a Kazal domain at the N-terminus and two EF-hand domains. The overall structure of CG2264 resembles SPARC-like proteins, such as Testican or SMOC-1, and thus may represent the fruit fly orthologue. The CG5639 protein is much larger, and contains a much greater number and diversity of protease inhibitor modules (one Kunitz; two WAP; four antistatin and five thyroglobulin

domains). The CG5639 protein is organized in a manner similar to lp11788p from *D. melanogaster* and c08g9.2 from *C. elegans* both of which are predicted from the genomic sequences yet, neither of has been functionally or genetically characterized in any way.

The group of proteins which contains a Kazal inhibitory domain (Falquet et al., 2002) is known to control cell growth and adhesion and to participate in functions that regulate processes of tissue remodelling. Examples of these are agrins (Ferns and Hall, 1992), BM-40 also called SPARC (Brekken and Sage, 2000; Brekken and Sage, 2001; Motamed, 1999), follistatin (Welt, 2002; Welt et al., 1997; Zwijsen et al., 1994), SC1/hevin (Hodges and Miginiac Maslow, 1993; Johnston et al., 1990), QR1 (Guermah et al., 1991, testican (Alliel et al., 1993, SMOC-1 (Vannahme et al., 2002), and smap-2 (Nishimoto et al., 2002). Many of these proteins, in addition to the Kazal domain, also contain a C-terminal calcium-binding domain that consists of two so-called EF-hand calcium-binding motifs. The structure of these EF-hand domains in BM-40 (SPARC) has been solved (Hohenester et al., 1996; Hohenester et al., 1997) and the mode of interaction of the protein with calcium has been studied in detail (Busch et al., 2000).

The CG5639 protein contains 5 thyroglobulin type-1, 4 antistatin, 2 WAP, and 1 Kunitz domains. WAP-domains (Whey Acidic Protein) are found in a group of proteins containing eight characteristically spaced cysteine residues, which are involved in disulphide bond formation, and have been termed “4-disulphide core” proteins (Hennighausen and Sippel, 1982). Many members of this family have an inhibitory activity against serine proteases (Grutter et al., 1988; Tsunemi et al., 1996; Wiedow et al., 1990). Antistatin and Kunitz domains also represent two classes of serine protease inhibitors characterised by well-conserved patterns of cysteine residues. Many of these proteins are anti-coagulants (Grutter et al., 1988; Mittl et al., 1997; Salier, 1990; Tsunemi et al., 1996; Wiedow et al., 1990).

Broad inhibitory function is desirable for variety of biotechnological applications. The aim of this study was to characterise the *Drosophila melanogaster* genes *CG2264* and *CG5639* at the gene and protein level. The cDNAs of the two genes were cloned and expressed in the yeast *Pichia pastoris* and the proteins were tested for protease inhibitor activities. In addition, the level of mRNA accumulation throughout sex-specific development was quantified. Our results serve as a primary characterisation of these genes. They demonstrate inhibitory activities of CG2264 and CG5639 proteins against *D. melanogaster* proteolytic enzymes, and indicate different developmental roles during sexual development and pupation. Our study may serve to initiate studies of the basic biological functions of SPARC- and thyroglobulin-like proteins using *Drosophila melanogaster* as a model organism.

MATERIALS AND METHODS

Oligonucleotides

The following primers (Eurogentec, Seraing, Belgium) were used for cloning and sequencing of the genes *CG2264* and *CG5639*. Restriction sites are underlined.

PPIC9(CG2264)N 5’- AAAAACCTAGGGACGATCAGTTGATGAAGCTCC
PPIC9(CG2264)C 5’- AAAAAGAGCTCTTAAGCGGCCGCATCTTTCAGATATCGCTCC
pPIC9(CG5639)N 5’- AAAAAGAATTCAAGCGATCAAATGCCAATCTTACG

pPIC9(CG5639)C 5'- AAAAAAGATCTTTAAGCGGCCGCCGCGGCGAACCCCTTCTTGC
CG5639ATG 5'- AAAAACCATGGCGCACTATGATCAACG
CG2264ATG 5'- AAAAACCATGGTGGCGAGATGTGTTTTGTTCG
CG2264DNa 5'- CAGTTCACGAACGCATATCC
pUC/M13 F 5'- CTAAAACGACGGCCAGTG
pUC/M13 R 5'- GGAAACAGCTATGACCATG

cDNA cloning of the two *D. melanogaster* genes

Total mRNA from mixed stages of *D. melanogaster* was isolated using the GenoPrep mRNA kit (Qiagen) and first strand cDNA was synthesised using SuperScript II reverse transcriptase (Gibco). First strand total cDNA was used as a template for the PCR reaction, using PWO polymerase (Roche) and two combinations of primers for each gene. For *CG2264* gene primers CG2264ATG/pPIC9(CG2264)C were used to amplify the *entire* predicted coding region including the N-terminal targeting sequence and primers pPIC9(CG2264)N/pPIC9(CG2264)C were used to amplify the sequence encoding only the predicted *CG2264* mature peptide. Similarly for *CG5639* gene primers CG5639ATG/pPIC9(CG5639)C were used to amplify the entire predicted coding region including the N-terminal targeting sequence and primers pPIC9(CG5639)N/pPIC9(CG5639)C were used to amplify the sequence of *CG5639* encoding only the predicted mature peptide. The PCR reactions were according to the PWO polymerase instructions and the specific requirements of the primers. The PCR products of the two full-length genes including the N-terminal leader sequences were purified from gel. The two major bands obtained for *CG2264* were purified separately and named *CG2264* for the bigger product with the expected size and *CG2264A* for the smaller product. Subsequently, terminal A-nucleotides were added using Taq polymerase and the fragments were cloned into the pGEM T-easy vector (Stratagene). The resulting plasmids, containing *CG2264*, *CG2264A* and *CG5639* insert were named pGEMCG2264, pGEMCG2264A and pGEMCG5639. The inserts of pGEMCG2264, pGEMCG2264A were fully sequenced with pUC/M13 forward and reverse sequencing primers and the gene-specific CG2264DNa primers. The insert of pGEMCG5639 was verified only at the 5' and 3' terminal ends by sequencing ~500-bp.

Analysis of the CG2264 and CG5639 gene expression by quantitative real time polymerase chain reaction (RT-PCR)

The fluorescent intercalating dye SYBR-Green was used to monitor the RT-PCR product accumulation on the ABI PRISM 7700 sequence detection system (Applied Biosystems). The specificity of the amplified products was checked on an agarose gel. A comparative C_T method was used to calculate the relative mRNA levels (Applied Biosystems), using 18S rRNA as a standard reference.

The primer combinations were for 18S rRNA (18S F [5'-TCCGGGTAAACCGCTGAAC-3'] and 18S R [5'-CAAAGGGCAGGGACGTAATC-3']); for *CG 2264*(I)(64 F [5'- AGGCTCAGAGCTTTTTCCGC-3'] and 64 R [5'-TTCAGCTCCCGAATCTCCTG-3']); for the alternatively spliced exon of *CG 2264* (II) (AF [5'-ACCGAGGTCACACCAATACCA-3'] and AR [5'-TTCACCTCCGTAGCCCGTT-3']); for *CG 5639* (39F [5'- ATGTTCCAGGAACGGACTCG-

3'] and 39R [5'- CCTGCTAATCATGTTCTCCCATC-3']). All primers for quantitative PCR were from AppliedBiosystems (Warrington, UK).

Preparation of *Pichia pastoris* expression constructs

The protein encoding regions encoding the mature peptides of the *CG2264* and *CG5639* genes were generated by PCR using the pGEMCG2264 and pGEMCG5639 vectors as templates and primer pairs pPIC9(CG2264)C / pPIC9(CG2264)N for *CG2264*, and pPIC9(CG5639)C / pPIC9(CG5639)N for *CG5639* as described in the previous section. The resulting fragments were digested with *Avr*II/*Not*I for *CG2264* and *Eco*RI/*Not*I for *CG5639* and introduced into a previously modified pPIC9 vector, pPIC9HisTagM (unpublished), that had a *Sna*BI site removed and 6 terminal histidine residues added behind the NotI site. Approximately 15 µg of each recombinant pPIC9HisTagM construct with insert, digested with *Sal*I for *CG2264*, *CG2264A* and *Stu*I for *CG5639* were electroporated into *P. pastoris* GS115 (*his4*) cells (Invitrogen) using a Bio-Rad Gene Pulser electroporator. Transformants were selected on MD medium plates (Invitrogen).

Expression the recombinant CG2264, and CG5639 proteins in *Pichia pastoris*

Single colonies of each construct were picked and inoculated into 50 ml BMGY medium (1% yeast extract, 2% peptone, 1.34% yeast nitrogen base, 4×10^{-5}% biotin, 1% glycerol, 0.1 M K-phosphate, pH 6.0) in 500 ml shake flasks. Cultures were grown two times overnight at 30° C on plate agitation at 250 rpm. Cells were collected by centrifugation, resuspended in BMM (1.34% yeast nitrogen base, 4 X 10^{-5}% biotin, 0.5% methanol, 0.1 M K-phosphate, pH 6.0) medium and the cultures were incubated in the same conditions as above and supplemented daily with 250 µl of 100% methanol for 96 hours. Cells of both cultures were removed by centrifugation at 1500 g for 10 min and the cell free supernatants were analysed by SDS-PAGE, and by western blotting. The procedure for the western blot was as described in (Outchkourov et al., 2002) with the difference of using anti-His mouse antiserum as primary, and anti-mouse antiserum conjugated to alkaline phosphatase as a secondary serum (Jackson Immuno Research, USA). The supernatants of the above *P. pastoris* transformants were concentrated 10-fold and the buffer was exchanged to water on Vivacell 70, 50.000 MWCO filters (Vivascience).

Assays of the protease inhibitory activities

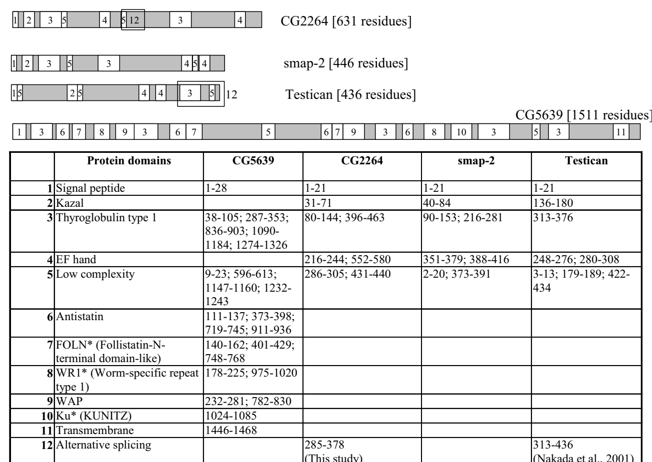
Total protein extracts from *D. melanogaster* were prepared in 1 mM Tris–HCl, pH 7.0 (extraction buffer), by grinding adult flies (~200 ul, stored at –20° C) using a plastic pestle in 1.5 ml microfuge tubes. The extract was clarified by centrifugation at 12,000 rpm for 10 minutes at 4° C in an Eppendorf centrifuge 5417R. The supernatants were pooled and purified on a Sephadex G-25 10 ml column (Pharmacia). The resulting supernatant was used as proteinase solution. The concentration of protein was determined according to the method of (Bradford, 1976), using bovine serum albumin

(BSA) as reference. Protease activity towards the following substrates, Z-Arg-Arg-pNA, Z-Phe-Arg-pNA, Suc-Ala-Ala-Pro-Phe-pNA, Suc-Ala-Ala-Ala-pNA, and Suc-Ala-Ala-Pro-Leu-pNA were measured spectrophotometrically at various pH values. All substrates were dissolved in dimethylsulphoxide at a concentration of 15 mg/ml. Drosophila total protein extract (20 µl of 5µg/µl protein) was mixed with 130 µl of one of the reaction buffers (50 mM MES pH 4.5; 50 mM MES pH 6.5; 50 mM Tris pH 7.5; 50 mM Tris pH 8.5; 50 mM Tris pH 9.5) and the reactions were started by adding 50 µl 0.3 mg/ml of the substrates diluted in water. Microtiter plates were used to follow the pNA release by measuring the OD 405 nm at 30 seconds intervals for 2 hours at 37° C on a Benchmark microplate reader (Bio-Rad).Identical substrates were used to further characterize the protease inhibitor activities of the recombinant proteins against Drosophila protein extract. The extract (20 µl) was pre-incubated with or without 20 µl the concentrated supernatants of CG2264 or CG5639 for 15 min at room temperature. The reactions were started by adding 50 µl substrate in optimal buffer in a total volume of 200 µl. Radial diffusion assays were conducted essentially as described in (Jongsma et al., 1993) using 4.2 nM chymotrypsin, 4.2 nM subtilisin, and 42 nM trypsin. Prior to photography, the plates were stored in water for 48 hours at 4 ° C in order to improve the contrast of the halos.

RESULTS

cDNA cloning of CG2264 and CG5639

As a result of the genome sequencing project of *Drosophila melanogaster,* the genes *CG2264* and *CG5639* from *Drosophila melanogaster* were discovered and annotated in FlyBase (Figure 1) (Adams et al., 2000; Gelbart et al., 2002).



	Protein domains	CG5639	CG2264	smap-2	Testican
1	Signal peptide	1-28	1-21	1-21	1-21
2	Kazal		31-71	40-84	136-180
3	Thyroglobulin type 1	38-105; 287-353; 836-903; 1090-1184; 1274-1326	80-144; 396-463	90-153; 216-281	313-376
4	EF hand		216-244; 552-580	351-379; 388-416	248-276; 280-308
5	Low complexity	9-23; 596-613; 1147-1160; 1232-1243	286-305; 431-440	2-20; 373-391	3-13; 179-189; 422-434
6	Antistatin	111-137; 373-398; 719-745; 911-936			
7	FOLN* (Follistatin-N-terminal domain-like)	140-162; 401-429; 748-768			
8	WR1* (Worm-specific repeat type 1)	178-225; 975-1020			
9	WAP	232-281; 782-830			
10	Ku* (KUNITZ)	1024-1085			
11	Transmembrane	1446-1468			
12	Alternative splicing		285-378 (This study)		313-436 (Nakada et al., 2001)

Unless indicated the domain names are according to the PFAM database http://www.sanger.ac.uk/Software/Pfam/
* The domains names are according to the SMART database http://smart.embl-heidelberg.de/

Figure 1. Schematic representation of the protein domains of the CG2264 and CG5639 *Drosophila melanogaster* proteins in comparison to the mammalian smap-2 and testican.

Based on the sequence information found in the FlyBase database two combinations of primers were designed for each gene: One to amplify the entire coding part, starting from the predicted ATG codon; the other to amplify the part of the gene encoding the mature peptide, starting from the cleavage site of the leader peptide. The two combinations of primers amplifying the coding part, with and without the N-terminal leader peptide for the *CG5639* yielded single and specific bands with the expected sizes (4.4 and 4.5 kb) on cDNA template derived from adult flies (Figure 3A). The amplification of the *CG2264* sequence with primers designed to amplify the entire coding part (starting from the ATG codon) resulted in six different amplification bands (Figure 3A). With the primer combination that was designed to amplify the mature protein without the leader peptide of *CG2264,* six bands similar in size and intensity were obtained.

A

```
CG2264(NP_724874) 31 CAAKQGE----------CD---DNEGPVCGTDGQTYPTRCHLLRAQCG------GHQVSLKYSGSC 77
Bm-40 (NP_033268) 29 CQDPTS-----------CPAPIGEFEKVCSNDNKTFDSSCHFFATKCTLEGTKKGHKLHLDYIGPC 83
Try_inh (700704A)  9 CTSEVSG----------CP---KIYNPVCGTDGITYSNECVL----CSENK-KRQTPVLIQKSGPC 56
SC-1  (NP_037078)426 CQDPET-----------CPP-AKILDQACGTDNQTYASSCHLFATKCMLEGTKKGHQLQLDYFGAC 479
Agrin    (P25304)515 CPR--------------CEH--PPPGPVCGSDGVTYLSACELREAAC-----QQQVQIEEAHAGPC 559
Follist. (Q90844)118 CAPD-------------CSN-ITWKGPVCGLDGKTYRNECALLKARC-----KEQPELEVQYQGRC 164
OVO      (P01003) 70 CSRYPNTTSEDGKVTILCT---KDFSFVCGTDGVTYDNECML----CAHNV-VQGTSVGKKHDGEC 127
OVO      (P01005)  9 CSSYANTTSEDGKVMVLCN---RAFNPVCGTDGVTYDNECLL----CAHKV-EQGASVDKRHDGGC 151
```

B

```
CG5639 (NP_651570) 232 KPGICPAPDHSQYTERTGYMCGSPCSHDLECRNMEKCCFTKGCQFNCQQP 281
CG5639 (NP_651570) 782 KPGQCPYLVPPGPDNLDANTCAYECRTDAHCDGARRCCSNG.CGTQCVDP 830
ELAFIN (NP_002629)  72 KPGSCPIIL----IRCAMLNPPNRCLKDTDCPGIKKCCEGS-CGMACFVP 116
ALK1   (NP_003055)  31 KAGVCPPKK----SAQCLRYKKPECQSDWQCPGKKRCCPDT-CGIKCLDP 75
ALK1   (NP_003055)  85 KPGKCPVTY----GQCLMLNPPNFCEMDGQCKRDLKCCMGM-CGKSCVSP 129
```

C

```
CG2264      80 -----CLEAV----KFARRQQERDP-GY-FVPRCR-KDGNFAAMQCYGNNG-------CWCSDSQ-GRPIADDNKQFRRKGKLRC-- 114
CG2264     396 -----CWMDQSVTLEEQGHGGKSV---L-FVPQCLP-DGRYQRIQCYSSTSTSY----CWCVNEDTGKSIPGTSVKNKR---PQC-- 463
CG5639      38 -----CQHLRRSESRRAKALE--GS-SV-RVPRCQ-KNGDFDAIQCQDEKHGRD----CWCVDDY-GVELPGSRNETRTG--VVC-- 105
CG5639     287 -----CHQAKA----LADILSINEREGRGYVPECNGPGGQFSPRQC--SRNGLV----CWCVDPRTGHKIKETMG---AANNVNC-- 353
CG5639     836 -----CQHLQAIQLHQSSELGIPAR-QM-AVAQCDPNNGKWNQVQCSPDGH-------CWCVDDQ-GKILPGTRVKSPAT--PKC-- 903
CG5639    1090 -----CERLK---LKNNLAAQRTGHSSVWFQPRCDPVTGHWSPVQCLGKQP/26/PGVCWCADKK-GAPLKGTLTR---ESEPIC-- 1184
CG5639    1274 -----C---RA----LSKTAP--------FPVSCDEAGA-FRPLQC----NGRS----CWCVDAAGNQLQSTHVF---GAGDRRC-- 1326
Ttestican 310 G--LPCQNEMNRIQKLSKGKSL---LGA-FIPRCN-EEGYYKATQCHGSTG--Q----CWCVDKY-GNELAGSRKQ---GAVS-CEE 327
ECI          1 K--TPCE--------LARDAATHGPIGG-FIPTCD-YNGQYTPEQCWGSTG--Y----CWCVNSS-GQKLPGTDTPP--GSASNC-- 64
p41        194 --LTKCQEEVCHI------PAVHP--GS-FRPKCD-ENGNYLPLQCYGSIG--Y----CWCVFPN-GTEVPNTRSR----GHHNCSE 258
EI           1 S-LTKCQQLQAS--------ANSGLIGT-YVPQCK-ETGEFEEKQCWGSTG--Y----CWCVDED-GKEILGTKIR----GSPDCSR 66
EI          67 AALTLCQMMQAII------VNVPGWCGP---PSCK-ADGSFDEVQCCASNG--E----CYCVDKK-GKELEGTRQQ----GRPTCER 134
EI         135 S---ECEE--ARIK--AHSNSLR--VEM-FVPECL-EDGSYNPVQCWPSTG--Y----CWCVDEG-GVKVPGSDVRF--KRPT-C-- 199
```

```
Testican (AAH30691)
ECI (2207269A)
P41 (P04233)
EI (AAF02722)
```

Figure 2. Alignment of the Kazal **(A)**, WAP **(B)**, and thyroglobulin type-1 **(C)** domains found on CG2264 and CG5639, with other proteins. GenBank accession numbers are provided between brackets.

The mature protein bands were slightly shorter corresponding to the size of the missing leader peptide DNA sequence. The top band 1 could be a contamination of genomic DNA in our mRNA preparation as judged by the size of the band. The other bands, however, are likely to be products of alternative splicing of *CG2264* within the protein-coding region, since the primers were designed at the borders at the protein-

encoding region. In FlyBase alternative mRNAs for *CG2264* are predicted only outside the protein-coding region (Figure 3B).
Sequencing of the two most intense bands with an approximate size of 1.8 kb confirmed that two of the five remaining bands are indeed products of alternative splicing (Figure 3B). The higher one of the two bands represented *CG2264* gene as predicted in FlyBase. The lower one of the two was lacking one of the exons (exon 7, Figure 3B), and was named *CG2264A*. The PCR program was optimized to amplify fragments with an expected size of about 2 kb, and the much higher intensity of the two bands, therefore, does not completely reliably indicate the relative abundance relative to longer transcripts.

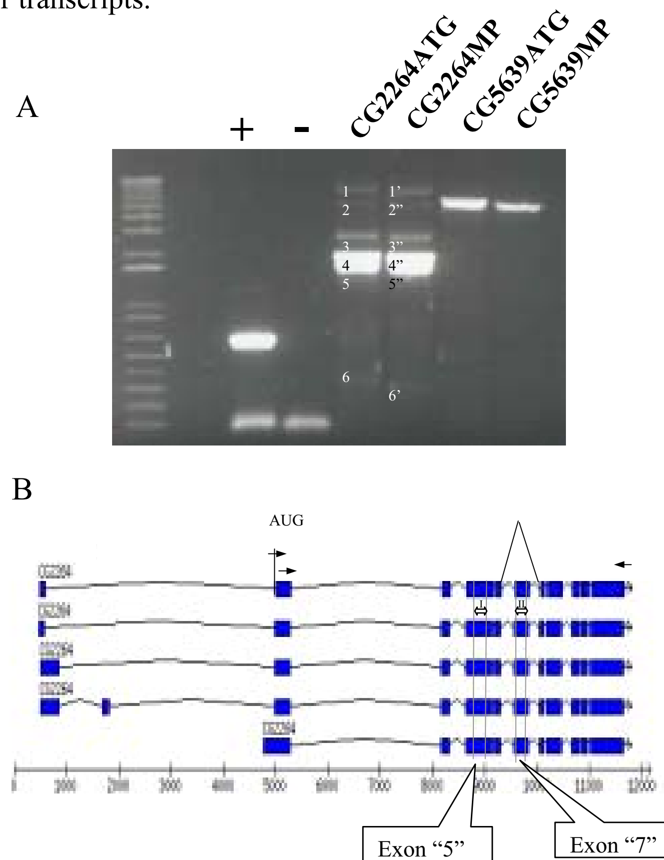


Figure 3. A. DNA gel electrophoresis of products of PCR amplification of on *D. melanogaster* cDNA as a template.. CG2264ATG: a primer combination designed to amplify the entire protein encoding region of *CG2264*. CG2264MP: primer combination designed to amplify the mature peptide of *CG2264*. CG5639ATG: primer combination designed to amplify the entire protein encoding region of the *CG5639*. CG5639MP: primer combination designed to amplify the mature peptide of the *CG5639*. Visible bands for primers combinations for the *CG2264* are shown with numbers. **B.** Schematic representation of the *CG2264* mRNA transcripts as predicted in the FlyBase. Dark squares indicate the positions of the exons, and bars indicate the positions of the introns, the alternative splice form of the *CG2264* as found in this study is indicated by double bars. The positions of the primers for cloning of the cDNA and the start codon AUG are indicated. The amplicons I and II that were used for a quantitative SYBR-Green RT-PCR are shown with bi-directional arrows.

Prediction of the potential biological function of the alternative splicing based on the protein sequence is not possible as exon 7 encodes part of a low complexity region

(multiple repeated aminoacid residues), which has no significant homology to known proteins (Figure 1).

Analysis of the CG2264 and CG5639 gene expression

Real-time RT-PCR based on the SYBR-Green dye was used to examine the transcript variations for both *CG2264* and *CG5639* throughout the development of *Drosophila melanogaster*. Larval, pupal as well as adult developmental stages were studied. Two pupal stages were examined which could be separated on the basis of flotation after submergence in water. Submersed pupae corresponded to early pupae and flotating pupae to older ones. Four variants of the adult stage were used: less than 24 hours after pupation, more than 24 hours after pupation, males, and females. Larval stages and 18S rRNA were used for normalization of the results. Two new splice forms were found in this study and two primer pairs were designed in order to quantify the two mRNA transcripts (Figure 3). The RT-PCR experiment which quantified all transcripts containing exon 7 (primer combination CG2264(II) figure 3B) resulted in transcript levels which were relatively constant between the different developmental stages that were tested (< 2-fold variations) (Figure 4).

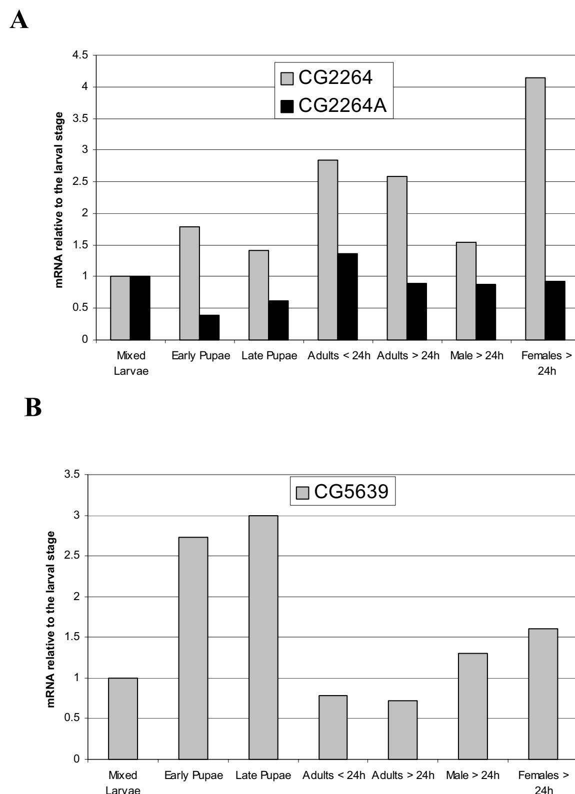


Figure 4. Analysis of the relative mRNA accumulation of the CG2264 and CG2264A, shown in **A** and CG5639 transcripts shown in **B** in the development of *Drosophila melanogaster*. SYBR-Green RT-PCR was used. Values were normalized to the 18S rRNA, and then to the larval stage. As CG2264 was found to be alternatively spliced two combinations of primers were used I and II respectively for each of the splice forms as shown on figure 3B.

In contrast, the primer pair which amplified part of exon 5 of gene CG2264(I) which is shared by both major transcripts (Figure 3B) showed significant variations in transcript levels between the two sexes at various developmental stages (Figure 4).

The lowest relative transcript levels were observed in the larval stage, followed by a slight increase (factor 1.5) of the CG2264(I) mRNA/s for both pupal stages, and a maximum level in the adults (factor 2.5). In the adults, the females contributed most to the increase, as the females contained 2.5 fold more CG2264(I) mRNA compared to males. Apparently, the CG2264 protein containing exon 7 is not regulated throughout the development and does not have a sex-specific role like CG2264A lacking exon 7, which is highly up regulated in adult females.

In contrast, the mRNA accumulation for *CG5639* was maximal in both pupal stages and about 3-fold higher compared to all other tested stages (Figure 4). Apparently, CG5639 is expressed mostly in the pupal stages where tissue remodelling processes dominate.

Expression in *Pichia pastoris* and characterization

The coding regions for the mature peptides of both CG2264 and CG5639 proteins were fused to the signal peptide of the α-mating factor from *Saccharomyces cerevisiae* in order to secrete both mature proteins in the culture supernatant of *P. pastoris*. When analysed on 10% SDS-PAGE the supernatant of *CG2264 P.pastoris* transformants contained a band with an apparent molecular mass of ≈75 kD, that was slightly larger than the expected 67 kD (Figure 5). The identity of the protein was confirmed by Western blot with serum specific for the His-tag (Figure 6). The supernatant from *CG5639* transformants on Coomassie stained gel showed a number of protein bands of ~40 , ~37 and ~15 kD that were much smaller than the expected size of 166 kD. On Western blot with His-tag antisera none of the bands reacted (Figures 5 and 6).

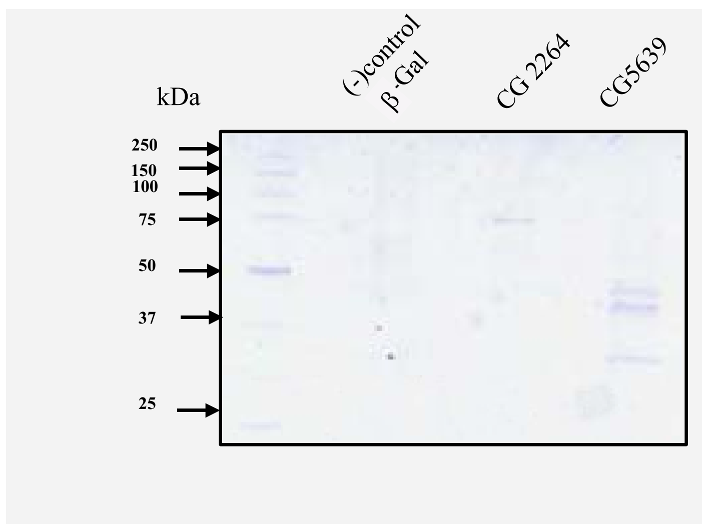


Figure 5. Coomassie stained SDS-PAGE analysis of *P. pastoris* supernatants obtained from various transformants: two of them aimed for extracellular secretion of CG2264, CG5639, and one for intracellular accumulation of β-galactosidase as negative control.

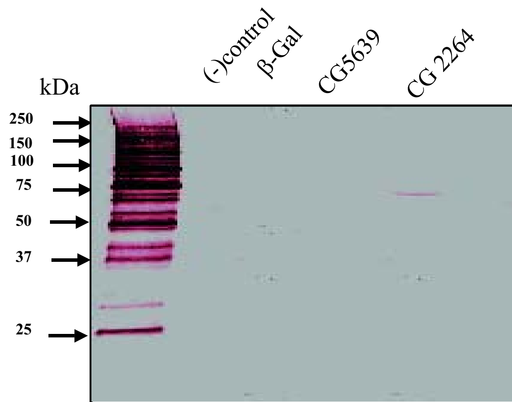


Figure 6. Western blot analysis of *P. pastoris* supernatants obtained from various transformants: two of them aimed for extracellular secretion of CG2264, CG5639, and one for intracellular accumulation of β-galactosidase as negative control.

Analysis of the inhibitory activities

The concentrated *P. pastoris* supernatants were tested for inhibitory activity against serine proteases using a radial diffusion assay as previously described (Jongsma et al., 1993). Inhibition against three different enzymes was tested. As shown in Figure 7 the addition of the CG2264 protein resulted in inhibition of three different proteases of the serine type: bovine trypsin, bovine chymotrypsin, and subtilisin from *Bacillus subtilis*.

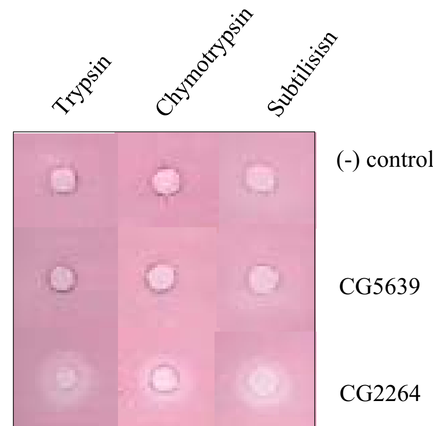


Figure 7. Radial diffusion assay of the *P. pastoris* supernatants versus three different serine proteases.

To identify endogenous *Drosophila* proteases that are potentially inhibited by the CG2264 and CG5639 proteins, we first analysed the proteolytic pH optima of total *D. melanogaster* extracts using six different synthetic substrates (Z-Arg-Arg-pNA; Z-Phe-Arg-pNA; Suc-Ala-Ala-Pro-Phe-pNA; Suc-Ala-Ala-Ala-pNA; Suc-Ala-Ala-Pro-Leu-pNA; Leu-pNA). The results of the optimal pH activity towards each one of the substrates are shown in Figure 8A. The Z-Arg-Arg-pNA substrate was most optimally hydrolysed at pH 9.5 which indicates serine protease activities. The similar

Z-Phe-Arg-pNa substrate was most optimally hydrolysed at a slightly alkaline pH of 7.5-8.5. The Suc-Ala-Ala-Pro-Phe-pNA substrate peaked at pH of 6.5. The hydrolysis of Suc-Ala-Ala-Ala-pNA was low in all of the tested pHs, and it seemed to have two peaks at pH of 6.5 and 8.5. In contrast the activity against Suc-Ala-Ala-Pro-Leu-pNa was clearly stronger and peaked at pH of 7.5. The activity against the aminopeptidase substrate Leu-pNA resulted in the strongest velocity and peaked around the neutral pH.

A

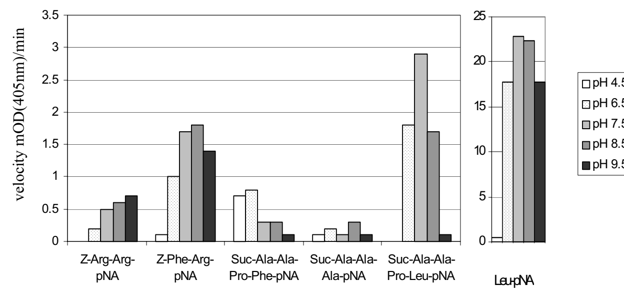
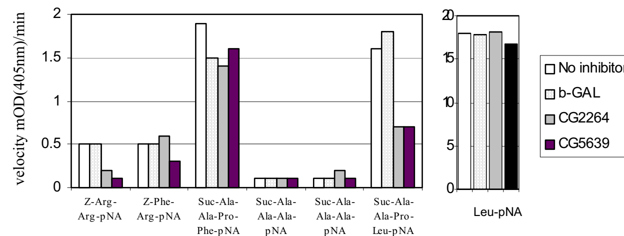


Figure 8. A. Proteolytic activities of total Drosophila extracts in various pH for each of the synthetic substrates. **B.** Inhibition of the proteolitic activity in Drosophila total extract by the *P. pastoris* supernatants using various synthetic substrates at optimal pH.

B



Subsequently, each of the substrates was tested at its pH optimum in the presence of the concentrated supernatants of the *P. pastoris CG2264* and *CG5639* transformants. The hydrolysis of Z-Arg-Arg-pNA, optimally hydrolyzed at pH 9.5, was inhibited by the addition of the CG2264. This correlated with the inhibition of purified trypsin in the halo assay, since the Z-Arg-Arg-pNA is a trypsin specific substrate, and the pH of 9.5 is optimal for many insect trypsins. In contrast, the supernatant of CG5639, which did not inhibit bovine trypsin, did also inhibit the Drosophila Z-Arg-Arg-pNA specific activity. Also the chymotrypsin/elastase substrate Suc-Ala-Ala-Pro-Leu-pNA resulted in significant levels of inhibition, and again the hydrolysis was inhibited by both supernatants of CG2264 and CG5639 (Figure 8B).

DISCUSSION

In this study we aimed to clone and characterize the multidomain genes *CG2264* and *CG5639* from *Drosophila melanogaster* that have been annotated in the public databases, but not yet characterized at neither the genetic nor the protein level.

Based on the domain organization, the CG2264 protein represents the fly orthologue of SPARC- or BM-40 related proteins, and is most similar to smap-2 and testican (Figure 1). For *CG2264* we found alternatively spliced mRNAs within the protein coding region. Also for testican different mRNA transcripts were found to be produced in mammals (Nakada et al., 2001). The predicted start codon (AUG) of the *CG2264* mRNA is situated exactly at the beginning of an exon (Figure 3B). It is possible that the present secretory targeting sequence could be exchanged by means of alternative splicing, and, thus, also change the protein localisation, as we observed at least 4-5 different splice forms which are not annotated in the databases. This was described for the protease inhibitor gene KAZ1 from Drosophila that contains a kazal domain at the N-terminus. The multiple spliced forms, exchange protein targeting sequences and result in multiple sub-cellular destinations (secreted, cytoplasmic and mitochondrial) (Niimi et al., 1999).

Those mRNA species of CG2264 that contain exon 7 were found to be relatively constantly expressed during fly development. In contrast, those mRNA species containing exon 5 (CG2264, CG2264A and possibly others) were increased during development: it is present less in larvae, more in pupae, and maximum in adults. This increase seems to be contributed mainly by the female developmental pathway, as adult females produced 2.5-fold more exon 5 containing mRNA than the males.

The mRNA of CG5639 was found as a single band in mixed adult populations and the differences in the regulation throughout the development peaked during the pupation.

Both *CG2264* and *CG5639* were found to be expressed in all tested developmental stages: larvae, pupae, and adults. Studies in mammals focussed in more detail on the characterization of the tissue specificity of expression of similar multidomain genes such as follistatin and SMOC-1 (Marr et al., 2000; Marr et al., 1997; Nakada et al., 2001; Trexler et al., 2001; Vannahme et al., 2002). From our finding that mRNA species lacking domain 7 of *CG2264* are highly expressed in females, we expect that *CG2264* is abundantly expressed in the female reproductive system in analogy to follistatin and SMOC-1 (Schneyer et al., 2000; Vannahme et al., 2002). In contrast, the mRNA transcript of *CG5639* was most abundantly present in pupae, indicating a possible role during fly metamorphosis.

The CG2264 recombinant protein was possibly post-translationally modified in *P. pastoris* or possessed a deviant mobility during SDS-PAGE analysis, as shown from the apparent mass of 75 kD on SDS-PAGE instead of the predicted 67 kD. The lower mobility of the expressed protein may be due to *P. pastoris* glycosylation, since there are six putative N-linked glycosylation sites within the coding sequence of the protein. Similarly, testican and SMOC-1 were found to be glycosylated *in vivo* and appeared with an increased molecular weight (Basalamah et al., 2001; Vannahme et al., 2002).

The expression in yeast of the *CG5639* gene resulted in the accumulation of smaller protein products, and no anti-His antibody cross-reaction on western blots.

The absence in the negative control of similar bands on the coomassie stained gel indicated that the expressed protein is degraded during the fermentation procedure. The histidine-tag could not be detected, but may have been cleaved off as a smaller peptide. Alternatively, the protein may not be secreted as a fully intact protein from *P. pastoris*. The protein contains a predicted transmembrane region located at the C-terminus which also harboured the His-tag (Figure 1). Possibly, only the products of limited proteolysis accumulated in the supernatant.

Inhibitory properties of the concentrated, unpurified yeast supernatants were tested against purified non-Drosophila serine proteases and crude *D. melanogaster* fly extracts as a source of proteolytic activity. CG2264 protein in contrast to CG5639 protein fragments was shown to inhibit three different types of mammalian serine proteases: trypsin, chymotrypsin and subtilisin (Figure 7). The inhibition of those enzymes is most likely attributable to the Kazal-like domain that is part of the CG2264 protein. It is the only domain found on CG2264, expected to inhibit proteases of the serine type. Although Kazal-like inhibitors are known to have broad inhibitory activity towards proteases of the serine type (Lu et al., 2001), the SPARC/BM-40 proteins, which contain a Kazal domain, were found to be inactive against serine proteases (Hohenester et al., 1997). This study, therefore, represents the first instance in which there is evidence of protease inhibitory activities of SPARC-related proteins towards serine proteases.

When tested against the endogenous proteolytic activity present in Drosophila extracts, CG2264 was active against Z-Arg-Arg-pNA hydrolyzing activity at pH 9.5. Also, the supernatant of CG5639 protein inhibited the hydrolysis of Z-Arg-Arg-pNA at the same pH, while the supernatant of the yeast negative control did not. This indicates that CG2264 and likely one or some of the protein fragments that are part of CG5639 are specific inhibitors of endogenous *D. melanogaster* serine trypsin-like proteases. Also for the substrate Suc-Ala-Ala-Pro-Leu-pNA at pH 7.5 it was found that the supernatants of both CG2264 and CG5639 inhibited this activity, in contrast to the supernatant from the control yeast strain. The substrate Suc-Ala-Ala-Pro-Leu-pNA is commonly used for characterization of serine proteases called elastases or chymotrypsin. Therefore, it is possible that the inhibition of endogenous enzymes with activity against the Suc-Ala-Ala-Pro-Leu-pNA is one of the functions of the CG2264 and CG5639 proteins. The type of *D. melanogaster* proteases that digest the Suc-Ala-Ala-Pro-Leu-pNA substrate remain to be elucidated. However, some of the proteins similar to CG2264, like the testican splice variants Testican-3 and N-Tes, were found to suppress (possibly by direct inhibition), the matrix metalloproteinase MT1(3)-MMP, which is an activator of pro-MMP2 (gelatinase A) (Nakada et al., 2001). MT-MMPs are known to be highly expressed on cancer cell membranes, and besides a function in activation of pro-MMP2, they have been shown to exhibit gelatinolytic activity themselves (Ohuchi et al., 1997). In contrast, SPARC has been shown to activate rather than suppress MMP-2 in invasive cancer cell lines (Gilles et al., 1998). It would be interesting to test whether the Suc-Ala-Ala-Pro-Leu-pNA substrate is digestible by the *D. melanogaster* MMPs. Expression of the separate domains of the CG2264 and CG5639 proteins could lead to the identification of the functional protease inhibitors domains within these proteins. It is possible that new MMPs inhibitors with potential anticancer therapy could be identified. As discussed in the introduction, previous studies on similar multidomain proteins have clearly shown their important function in mammals. They control cell growth and adhesion,

participate in functions that regulate processes of tissue remodelling, are involved in cancer development (Gilles et al., 1998), and mutations are often linked to hereditary diseases (Maegert et al., 2002; Mueller et al., 2002). The characterisation of similar genes in model organisms like fruit fly could yield a better understanding of the basic biological functions of these proteins. Furthermore, the availability of genetic knockouts and their possible complementation with mammalian genes in *D. melanogaster* could serve as a basis for future studies.

The expected inhibitory profile of both proteins was rather limited since no inhibition of cysteine and aspartic proteases such as papain and cathepsin D could be demonstrated, besides the presence of multiple thyroglobulin type-1 domains (2 for CG2264; and 5 for CG5639). Only inhibitions of three serine proteases were observed by the addition of CG2264, which was likely due to its Kazal domain. This indicates that the thyroglobulin domains of CG2264 and CG5639 are inactive in inhibiting the above two proteases (papain and cathepsin D). The above two proteins seem, therefore, of limited value for the development of biotechnological applications, where inhibition of a broad range of cysteine and serine gut proteases is required.

CHAPTER 9

GENERAL DISCUSSION

***"Evolution sparks an arms race between human chemical control and pest or disease agents, dramatically increasing costs that are eventually paid by consumers"* (Palumbi, 2001)**

ABSTRACT

Sustainable agriculture depends among other things on our ability to avoid or delay the development of resistance in insects against the pest control principles we deploy. Inspiration can be derived from principles already operational in nature like plant protective protease inhibitors. Insects have developed various forms of resistance to these natural inhibitors. Selection of heterologous inhibitors with desired properties, and engineering their stable accumulation in plants are critical in obtaining transgenic plants with significantly improved defences against insects. In this thesis, combinations of non-plant protease inhibitors were challenged against the adaptive potential of Western Flower Thrips (WFT). Stable, high level expression of selected inhibitors in plants was achieved by improving the promoter, the gene and the sub-cellular localization. The plants demonstrated resistance to WFT. The principle employed does not kill the insects, but rather strongly reduces their fecundity and spoils their appetite. The sustainability of such a trait will be a relevant direction of future research.

Protease inhibitors as part of integrated pest management programs

Due to the increasing world population there is a demand for increasing agricultural production from a decreasing acreage of arable land. At the same time there is an increasing concern for a more safe and environmentally friendly agriculture. Insect pests evolve resistance to pesticides in general within a decade, and some insects are so resistant to several pesticides that they are difficult or impossible to control (Palumbi, 2001, for review). Transgenic plants expressing the toxic proteins from *Bacillus thuringiensis* (Bt-toxins) seem to offer new, much needed alternatives (Cannon, 2000; Peferoen, 1997). Yet due to the strong and specific selection pressures the pests may adapt to the engineered trait or secondary pests can become equally problematic (Ferre and Van Rie, 2002; Gahan et al., 2001; Greene et al., 1999; Turnipseed et al., 1995). The integration of multiple levels of insect resistance traits into transgenic crops is desirable, since it is expected to slow down the evolution of insect resistance and offer the holy grail of sustainable solutions despite the practice of monocultures.

Protease inhibitors have probably evolved as a means of natural protection against insect herbivores for millions of years, and only recently we have begun to understand their ecological and biological roles. Protease inhibitors seem to have

evolved as relatively mild toxins affecting growth, development, and fecundity but without being acutely toxic (Annadana et al., 2002b; Ashouri et al., 1998; De Leo and Gallerani, 2002; Heath et al., 1997; Spates and Harris, 1984). Possible advantages of these milder and chronic effects might be understood in terms of a slower evolutionary adaptation process in herbivores, due to reduced selection pressure.

The larval stage of most insect pests is the one that is most destructive. The larvae originate from eggs deposited after careful selection of host plants by their mothers. Novel compounds that deter female pests from host plants are therefore from great scientific and applied interest. They represent the opportunity to study new plant insect interactions, and at the same time to engineer applications of insect resistance in which the direct genetic selection of resistant larvae is avoided. Induced defence responses in plants, in general, have been found to deter insect herbivores from their host plants (Thaler et al., 2001; van Dam et al., 2000). Deterrence could function effectively by remitting the attacking herbivore to the competitor plant (Tuomi et al., 1994; van Dam and Baldwin, 1998; van Dam et al., 2000).

The results described in Chapter 4 demonstrate that adult females of the generalist herbivore Western flower thrips *Frankliniella occidentalis* (Pergande) are deterred by host plants with high levels of cysteine protease inhibitors. In Chapter 7 it is demonstrated that in a non-choice situation the population development of WFT on plants with high levels of cysteine protease inhibitors is up to 80% reduced. Thus, apparently some insect herbivores carefully select host plants with low levels of inhibitors on which the development of the future population would be optimal.

Protease inhibitors, in this way, introduce a multi-level form of resistance which might be well implemented in integrated pest management strategies that aim for lower needs of insecticides and better use of biological control strategies.

Important technological considerations for engineering applications of protease inhibitors in plants

High-level, stable protein expression in transgenic plants is the key to the success of many traits. Successful applications of protease inhibitors for pest control particularly depend on expression levels of selected inhibitors of at concentrations of around 0.5-1% of total soluble protein (De Leo et al., 1998; De Leo and Gallerani, 2002; Jongsma, 1995; Jongsma et al., 1996). In addition, protease inhibitors in plants need to be properly folded by targeting to the appropriate sub-cellular environment and must resist the degradation by plant proteases present in those compartments. While there is a large amount of knowledge concerning protein catabolism, the reasons for success or failure to express proteins in plants mostly remain obscure. Until recently mammalian inhibitors had not been expressed in plants. Christeller et al., (2002), however, found that the expression of bovine spleen trypsin inhibitor in tobacco varied about 20-fold depending on modifications of the cDNA (3’and 5’changes and minor codon changes). This showed that those changes could have drastic effects at the gene expression level. In order to exploit the protein synthesis machinery of the living cell knowledge of every step in the production process is crucial. Prediction of the fate of heterologous proteins in a new cellular environment will, however, always remain difficult. Thus, the technology of production of heterologous proteins still needs major improvements. Problems like gene silencing, *in-vivo* proteolysis by uncontrolled plant proteases or degradation by post-translational ubiquitination and

possibly other epigenetic mechanisms still hamper heterologous protein production. These problems could be tackled with an experimental, approach in which the problem is studied in detail and possible engineered solutions are experimentally evaluated and modelled.

Heterologous protein production in *Pichia pastoris*

In order to engineer insect resistance traits into crop plants it is necessary to select genes with desired insecticidal properties, and for this high level protein production in heterologous expression systems like yeast is a crucial prerequisite. Only then sufficient quantities of purified protein for insect bioassays can be obtained. *Pichia pastoris* is widely used for heterologous protein production from eukaryotic genes (Lin Cereghino et al., 2001, for review). Different proteins are produced at very different levels. While it is known, that some foreign DNA sequences in yeast can produce signals for premature polyadenylation resulting in short mRNA transcripts (Romanos et al., 1992, for review), other factors affecting the suitability of the foreign DNA sequences to be expressed in yeast have been poorly studied. In Chapter 2, various DNA sequences, encoding an identical protein were evaluated for their ability to express the protein at high levels. Most importantly a sequence motif (TACGTAGAATTC), present in the coding sequence of equistatin gene, which was identified responsible for dramatically, reduced protein production, without affecting the length of the mRNA transcript. Part of this sequence motif “TACGTA” corresponded to an efficiency element, that is a part of the yeast polyadenylation machinery normally present in the 3’ non-coding region of yeast genes (Guo and Sherman, 1996; van Helden et al., 2000). In our case, however, the presence in the coding sequence hampered significantly (> 10-fold) the protein production. The most optimal gene sequence accumulated mRNA of equistatin at levels equivalent to the endogenous AOX1 transcript levels. Both genes were under control of two AOX1 promoters present at different loci in the genome. This indicated optimal transcription and transcript stability and demonstrated that an important element governing the transcription process had been discovered. Overall protein production was improved 20-fold, and resulted in the accumulation of 1.66g/L of active inhibitor in supernatant of the *P. pastoris*.

An improvement of the expression of heterologous proteins in plants

A large number of promoters in plants have been described in the literature, yet expression levels of different transgenes in the range of the most abundant endogenous protein ribulose-1,5-bisphosphate carboxylase (rubisco) have not yet been achieved. To ensure high levels of protein accumulation in transgenic plants we cloned the gene regulatory sequences of an abundantly expressed member of the rubisco small subunit gene family from chrysanthemum (Chapter 3). Usually the *rbcS* genes are members of multigene families within a plant and the different members are expressed at different levels (Dean et al., 1989c). R*bcS* promoters have been extensively studied, and it was demonstrated that both the 5’and 3’ regulatory sequences are important in determining the transcript level (Dean et al., 1989a; Dean et al., 1989b). All biotechnological studies that aim to engineer high levels of

heterologous protein accumulation, however, have used *rbcS* promoters and terminators from different origin. We engineered for the first time a promoter and terminator of the same gene into an expression vector for heterologous protein production. It was subsequently demonstrated, that the new *rbcS1* expression cassette yielded expression levels in Solanaceous plants of 3-10% of total soluble protein for various proteins. It is not clear whether these unusually high levels are due to the combination of the promoter with its own terminator or whether the promoter simply drives unusually high levels of transcription. Many of the abundantly expressed *rbcS* promoters possess homology to the *rbcS1* promoter from chrysanthemum, which favours the first possibility. Future studies should dissect these questions by cloning homologous promoters from other plant species and study expression with and without their terminator. A number of new conserved sequence elements shared with other *rbcS* promoters were found in the proximal part of the *rbcS1* promoter, and could serve as a basis for future studies in which the function of these conservative elements are investigated by mutational analysis.

The role of endogenous proteases as an impediment for recombinant protein production in plants

Proteolytic degradation in plants as a problem is highlighted by a number of studies that describe different proteolytic fragments in different plant species and plant cell cultures (De Neve et al., 1993; Dolja et al., 1998; Hiatt et al., 1989; Khoudi et al., 1999). Up to now the proteases involved in the degradation of foreign proteins in plants have not been studied. An understanding of protease activity and the susceptibility of foreign proteins to protease attack is essential for the development of economically-feasible transgenic plants not only for insect resistance, but also for general large-scale protein production such as antibodies in plants. The use of the *rbcS1* promoter and terminator ensured high levels of transcription for a number of different protease inhibitors. Only one out of five inhibitors, when targeted to the secretory pathway of potato leaf cells appeared to accumulate as a single band. The rest seemed to be proteolytically degraded (Chapter 4). The plant proteases involved in the degradation of recombinant equistatin were studied in detail (Chapter 5), and it was found that the protein is degraded *in-planta* mainly due to cysteine proteases specific for the Asn-residue and and to a lesser extent by Arg/Lys specific proteases. These results may be relevant to the considerable reduction in the level of functionally intact antibodies in aging tobacco plants (Stevens et al., 2000), yet it seems that general strategies to effectively avoid this degradation are presently poorly developed.

Strategies for stabilization of heterologous proteins in transgenic plants

Proteolytic degradation of proteins generally begins with the hydrolysis of internal peptide bonds. Two factors are needed for the initial cleavages: (i) consensus residues that are exposed and accessible at the surface of the protein, (ii) proteases that recognize and cleave the exposed motif. The initial cleavages usually destabilize the protein structure and expose new and internal cleavage sites, which are then further

degraded. To reduce the hydrolysis of recombinant equistatin in plants two different strategies were attempted (Chapters 6 and 7).

The first one depended on rational design to stabilize the protein. Residues sensitive to cleavage *in-planta* were exchanged with ones that were likely to be less sensitive to cleavage. At the same time structural considerations (modeling) were used to guide the choice of mutations. This general strategy has been successful in a number of different cases in organisms other than plants (Audtho et al., 1999; Frenken et al., 1993; Kapust et al., 2001; Markert et al., 2001; Varallyay et al., 1998). In Chapter 6 this strategy is attempted for the first time in plants. The results indicate that the overall strategy, as described in chapter 6, was only partly successful. Some of the amino acid substitutions improved resistance to proteases as deduced from the accumulation of proteolytically resistant fragments, yet in overall the strategy did not yield the expected protease resistant full-length protein. It seems that with more effort in finding the right amino acids substitutions the strategy might work. Therefore, more random approaches to seek for appropriate substitutions and functional activity would be preferred

The second strategy focused on the fact that the addition of kininogen domain 3 and cystatin C inhibited the plant proteases involved in the degradation of equistatin in potato plant extract (Chapter 5). With this in mind, but also with the aim to achieve concerted accumulation of multiple protease inhibitors in plants we designed a synthetic multi-domain protein in which representative members of different cysteine protease inhibitors were expressed as a single open reading frame. As reviewed in the introduction of the thesis nature has offered us numerous examples of different multi-domain inhibitors. Synthetic multi-domain inhibitors offer the advantage of being "custom" made: Inhibitors with desired properties can be selected and fused together. It was expected that multi-domain inhibitors would have the cumulative properties of the single domain inhibitors, although there were only few studies evaluating the functionality of fused inhibitors.

In this study engineered multi-domain protease inhibitors consisted of representative classes of cysteine and aspartic protease inhibitors: kininogen 3rd domain (K), stefin A (A), cystatin C (C), potato cystatin (P) and equistatin (EIM). Four (K-A-C-P) to five (EIM-K-A-C-P) inhibitors were fused into one single reading frame, and were shown to fold into functional protein in *P. pastoris*.

Thus, a broad range of different known types of cysteine protease inhibitors was covered. All the selected inhibitors were strong inhibitors of Western flower thrips proteases (Chapter 4), and inhibited each others degradation in plant extract (Chapter 5). Thus, a protein was constructed which targeted a maximal number of insect proteases while at the same time stabilized itself against proteolytic degradation, and eventually yielded in a novel insect resistance strategy (Chapter 7). Apart from being an effective thrips resistance factor this novel protein might have additional applications to prevent the degradation of co-expressed proteins. The potential of natural multidomain inhibitors consisting of putative cysteine and aspartic protease inhibitors for similar applications was studied in Chapter 8, where two multidomain proteins CG2264 and CG5639 were obtained from *Drosophila melanogaster* cDNA, and partly characterized. Only inhibition of proteases of serine type was observed and no inhibition of cathepsin D and papain, despite the presence of multiple thyroglobulin type-1 domains. Engineered multidomain proteins may

therefore prove to be more useful for biotechnological applications since they carry the desired custom made properties.

Issues with regard to genetically modified crop plants

The successful release of crop plants produced by genetic engineering depends on a range of scientific and non-scientific issues such as laboratory evaluation, field trials, regulatory approval, proprietary rights and public perception. Debates on the potential dangers of genetically modified (GM) crops still continue, and one of the most controversial issues has been the potential risk for human health and the environment. Transgenic crops are grown at a large commercial scale since 1996, where 1.7 million hectares of farmland were planted. Five years later, in 2001 the GM crop acreage increased to an amazing 52.6 million hectares that is 526 km x 1000 km (3 x the size of UK) (James, 2002)! The impact of GM crops in the US is very significant. For an acreage of 34.8 million hectares in 2001, the use of pesticides was reduced with 20.7 metric tons and resulted in a net economical gain of 1.4 billion US dollars (Gianessi et al., 2002). Until now, instances of GM crops with proven, negative effects on the environment or human health have not been reported. Indeed transgenic crops have offered, in the short time frame from 1996 until now, an environmentally friendly and cost effective alternative to the use of pesticides (Shelton et al., 2002, for review). In addition to the general acceptance of GM crops in USA, China and Argentina many developing countries are currently in the process of assessing the costs and benefits of GM crops. Field trials in farms of different states of India demonstrate not only greatly reduced use of pesticides but also substantially reduced pest damage that increased yields up to 80% (Qaim and Zilberman, 2003). In GM crops due to the reduced use of pesticides additional methods such as biological control get a chance to work as a part of the integrated pest management strategy. Despite all these positive economic and environmental aspects the public opinion, especially in Europe, is often strongly opposed. The current GM crops are designed to improve agronomic traits and lack direct benefits for the consumer. Thus, many consumers prefer to consume GM free products simply because they are more “natural”. At the same time, GM crops that have consumer traits like improved nutritional value and benefits for human health have been already developed (Engel et al., 2002; Yan and Kerr, 2002, for reviews). It is possible that when they appear on the market the public opinion in the developed countries might change.

In this thesis we have expressed a number of different protease inhibitors in potato plants and evaluated the effect on western flower thrips and colorado potato beetle as insect pests. The origin of the different inhibitors varied since the selection of the inhibitors was based on availability of coding sequences of different types of cysteine proteinase inhibitors. Type 1 cystatin (human stefin A), type 2 cystatin (human cystatin C), type 3 cystatin (human kininogen, domain 3), phytocystatin (potato cystatin), and thyropin (sea anemone, equistatin) were used. Although potato is a good host for thrips in the greenhouse, it rarely suffers economically significant damage in the field. Thus, this study has been carried out with potato as a model plant. Considering the current debate around GM plants we do not expect that the public will be enthusiastic to consume crops containing human protease inhibitors, even though we expect no harm from it. Thus, this particular multidomain construct may not find practical applications in food crops. However, similar inhibitors can be

constructed with similar genes that are found in farm animals which humans traditionally and without any harm have consumed. Such constructs should pose the least problems on thrips predated, non-food crops like cotton and ornamentals, and could be implemented in resistance management traits in addition to the Bt toxins.

Future outlook

The aim of a sustainable agriculture depends among other things on our ability to avoid or delay the development of resistance in insects against the pest control principles we deploy. To achieve sustainable solutions different technological approaches need to be designed into genetically modified crops. As demonstrated in this thesis inspiration can be derived from principles already operational in nature like protective protease inhibitors. At the same time public perception and regulation of genetically modified crops in general are just grown up from its infancy. Future research is needed and required by the regulatory authorities to establish the risks and benefits of crops expressing protease inhibitor mediated insect resistance traits.

ACKNOWLEDGEMENTS

More than five years after I joined the group of Maarten Jongsma I have the pleasure to acknowledge the people that contributed personally, morally and professionally to my development and experience as a PhD student and helped to make progress with my thesis.

First things first, so chronologically speaking I should say thanks to Prof. Dukiandjiev and Prof. Minkov from University of Plovdiv, Bulgaria for granting me a Tempus fellowship by which I more or less accidentally got involved in work on "Protease inhibitors" as part of a research group led by Dr. Maarten Jongsma.
Dr. Jongsma, as my direct supervisor contributed the most to my personal development at PRI (CPRO), and to the thesis itself. His ability to understand a given problem, his straightforward pragmatic approach, and excellent guidance is highly acknowledged. I learned a lot from Maarten in every aspect and at every stage of my period as a PhD student. Maarten, many thanks for that!

Within the "Protease inhibitors" group at the beginning (1998), the personal help of Jules Beekwilder, Paul Bijkerk, and Boris Rogelj is also highly acknowledged. Jules contributed significantly throughout my term there with helpful discussion and practical advice. My first day of (CPRO) lab experience was guided by him, and lead to a midi-prep isolation of pBINPLUS vector, which I successfully used for all plant transformations later on. Paul Bijkerk helped me a lot during the initial months in finding stuff around the institute, operating different equipment, and introducing me to various lab protocols. Boris, at that time a PhD student, often liked to travel so that the experiments he had started, often had to be completed by me. Many of the constructs and experiments in this thesis were done together with Boris. Paul, Jules and Boris nicely combined the lab work with social beer drinking experiences and generated unforgettable memories of my initial period in Wageningen. Furthermore, many thanks to Bert Schipper who was my bench neighbour for quite some time. I learned a lot from him and not only in technical terms. He helped me enormously with the "practical protein work" mentioned in this thesis. But it was also a pleasure to work next to him as a serene colleague ready to offer help whenever needed. Ralph Litjens contributed to the thesis by keeping the Drosophila population healthy and in good order. He together with the other group members: Anja Kuipers, Mariatereasa Volpicella and Seetharam Annadana contributed with stimulating discussions and generating the pleasant working atmosphere. The cooking and social skills of Seetharam generated numerous events that made the life outside the lab space nice and colourful.

Seetharam as a PhD student laid the foundations for the protease inhibitors-thrips related work of the group. On the thrips side the invaluable contribution of the people from the Business Unit Biointeractions: Willem Jan De Kogel, Antje Schuurman-de Bruin, and Gerrie Wiegers in setting up and helping with the thrips bioassays is highly appreciated. Kristina Gruden from Slovenia also helped with the thesis by pioneering the selection of suitable protease inhibitors for the control of thrips and Colorado potato beetle.

At the cluster level, Ruud de Maagd contributed mostly by helpful and straight-to-the-problem criticism that helped me to reconsider interpretations of

scientific results and to be extremely careful in any statement concerning results. In addition, he organised numerous social events, dominated by barbeques and visits to the Balkan restaurant that helped to reduce the working stress quite significantly. Harro Bouwmeester and Francel Verstappen also contributed with discussion and by providing specific compounds. Petra Bakker contributed to the nice peaceful working lab atmosphere with her specific appearance and personality. For example: If you load a gel next to Petra she would usually mention “gezellig, he!”. Mieke Hendriks is acknowledged for her constant smile and ability to remain cheerful under all circumstances, thus generating a pleasant lab environment, while keeping strict control over the responsibilities of everyone participating in the lab. Jos Molthof should be mentioned for contributing with a number of practical advises to the thesis. Stanimir Dulev a student that joined later the group contributed to the thesis with a number of different experiments.

The friendship, and support of Najely, Stefan, Raffa, Asaph, Cinzia, Wagner, Salvador, Samir, Andy, Poli, Segei, Ilia, Tetty, Alexandra, Alex, Silvia, Paulina, Iris, Violeta, Katja, Elena and Aernoud is highly appreciated. Without Raffa the thesis defence would have been somewhat later, thanks for that.

Special gratitude goes to my wife Rumyana for crucial and highly valuable contribution to this thesis. She has been a constant supporter with whom I shared successes and disappointments. She took care of my dearest daughter Sophia in many evenings and weekends in which I was working on my thesis while during the week I had already started my new post doctoral job on malaria.

Last but not least, I devote this thesis to my parents for their constant support and dedication.

Посвещава се на моите родители заради тяхната постоянна подкрепа и преданост.

CURRICULUM VITAE

Nikolay Stoyanov Outchkourov was born on 5th of February 1972 in Plovdiv, Bulgaria. In 1990 he finished secondary school in Sofia, Bulgaria. In September 1992, after a year and a half of obligatory military service, he started courses in Biology at the University of Plovdiv, Bulgaria. He specialised as a teacher of biology, and subsequently obtained an MSc degree in Plant biotechnology in 1999. Part of his practical training for the MSc degree was conducted at Plant Research International, Wageningen, The Netherlands in the group of Maarten Jongsma. After his graduation he continued the same line of research at Plant Research International and was enrolled as a PhD student. Since February 2003, he joined as a postdoc the Malaria Vaccine Development team at the Catholic University of Nijmegen.

LIST OF PUBLICATIONS

Outchkourov, N.S., Stiekema, W.J. and Jongsma, M.A. (2002) Optimization of the expression of equistatin in Pichia pastoris. *Protein Express. Purif.*, **24**, 18-24.

Outchkourov, N.S., Peters, J., de Jong, J., Rademakers, W. and Jongsma, M.A. (2003) The promoter-terminator of chrysanthemum rbcS1 directs very high foreign gene expression levels in dicotyledonous plants. *Planta*, **216**, 1003-1112.

Outchkourov, N.S., Rogelj, B., Strukelj, B. And Jongsma, M.A. (2003) Expression of sea anemone equistatin in potato: effects of plant proteases on heterologous protein production. *Plant Physiol.*, **133**, 379-390.

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. *J. Insect Physiol.*, **48**, 701-706.

Annadana, S., Schipper, A., Beekwilder, M.J., Outchkourov, N.S., Udayakumar, M. and Jongsma, M.A. (2003) Cloning, functional expression in *Pichia pastoris*, and purification of potato cystatin and multicystatin. *J. Biosci. Bioengineer.,* **95,** 118-123

Outchkourov, N.S., de Kogel, W.J., Schuurman-de Bruin, A., Abrahamson, M. and Jongsma, M.A. Specific cysteine protease inhibitors act as deterrents of Western flower thrips *Frankliniella occidentalis* (Pergande) in transgenic potato. s*ubmitted for publication*.

Outchkourov, N.S., de Kogel, W.J., Wiegers, G.L., Abrahamson, M. and Jongsma, M.A. Engineered multi-domain cysteine protease inhibitors yield resistance against western flower thrips (*Frankliniella Occidentalis*) in greenhouse trials. *submitted for publication.*

REFERENCES

Abe, H., Yamaguchi_Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D. and Shinozaki, K. (1997) Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell*, 9, 1859-1868.

Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I. and Ishii, S. (1993) Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis. *J. Biol. Chem.*, 15, 3525-3529.

Abrahamson, M., Dalbøge, H., Olafsson, I., Carlsen, S. and Grubb, A. (1988) Efficient production of native, biologically active human cystatin C by Escherichia coli. *FEBS Lett.*, 236, 14-18.

Abrahamson, M., Salvesen, G., Barrett, A.J. and Grubb, A. (1986) Isolation of six cysteine proteinase inhibitors from human urine. Their physicochemical and enzyme kinetic properties and concentrations in biological fluids. *J. Biol. Chem.*, 261, 11282-11289.

Adams, M.D., et.al. (2000) The Genome Sequence of Drosophila melanogaster. *Science*, 287, 2185-2195.

Al-Khunaizi, M., Luke, C.J., Askew, Y.S., Pak, S.C., Askew, D.J., Cataltepe, S., Miller, D., Mills, D.R., Tsu, C., Bromme, D., Irving, J.A., Whisstock, J.C. and Silverman, G.A. (2002) The serpin SQN-5 is a dual mechanistic-class inhibitor of serine and cysteine proteinases. *Biochemistry*, 5, 3189-3199.

Alliel, P.M., Perin, J.P., Jolles, P. and Bonnet, F.J. (1993) Testican, a multidomain testicular proteoglycan resembling modulators of cell social behavior. *Eur. J. Biochem.*, 214, 347-350.

Alvarez Fernandez, M., Barrett Alan, J., Gerhartz, B., Dando Pam, M., Ni, J. and Abrahamson, M. (1999) Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J. Biol. Chem.*, 274, 19195-19203.

Annadana, S., Mlynarova, L., Udayakumar, M., de Jong, J. and Nap, J.P. (2002a) The potato Lhca3.St.1 promoter confers high and stable transgene expression in chrysanthemum, in contrast to CaMV-based promoters. *Mol Breed.*, 8, 335-344.

Annadana, S., Peters, J., Gruden, K., Schipper, A., Outchkourov, N.S., Beekwilder, M.J., Udayakumar, M. and Jongsma, M.A. (2002b) Effects of cysteine protease inhibitors on oviposition rate of the western flower thrips, Frankliniella occidentalis. *J. Insect Physiol.*, 48, 701-706.

Annadana, S., Schipper, A., Beekwilder, M.J., Outchkourov, N.S., Udayakumar, M. and Jongsma, M.A. (2003) Cloning, functional expression in *Pichia pastoris*, and purification of potato cystatin and multicystatin. *J. Biosci. Bioengineer.,* 95, 118-123

Ashouri, A., Overney, S., Michaud, D. and Cloutier, C. (1998) Fitness and feeding are affected in the two-spotted stinkbug, Perillus bioculatus, by the cysteine proteinase inhibitor, oryzacystatin I. *Arch. Insect Biochem. Physiol.*, 38, 74-83.

Atkinson, A.H., Heath, R.L., Simpson, R.J., Clarke, A.E. and Anderson, M.A. (1993) Proteinase inhibitors in Nicotiana alata stigmas are derived from a precursor protein which is processed into five homologous inhibitors. *Plant Cell*, 5, 203-213.

Atkinson, H.J., Blundy, K.S., Clarke, M.C., Hansen, E., Harper, G., Koritsas, V., McPherson, M.J., O'Reilly, D., Scollan, C., Turnbull, S.R. and Urwin, P.E. (1984) Novel plant defences against nematodes. *NATO ASI ser, Ser A, Life sci. New York : Plenum*, 268, 197-210.

Atkinson, H.J., Urwin, P.E., Hansen, E. and McPherson, M.J. (1995) Designs for engineered resistance to root-parasitic nematodes. *Trends Biotechnol.*, 13, 369-374.

Audtho, M., Valaitis Algimantas, P., Alzate, O. and Dean Donald, H. (1999) Production of chymotrypsin-resistant Bacillus thuringiensis Cry2Aa1 delta-endotoxin by protein engineering. *Appl. Environm. Microbiol.*, 65, 4601-4605.

Auerswald, E.A., Rössler, D., Mentele, R. and Assfalg-Machleidt, I. (1993) Cloning, expression and characterization of human kininogen doain 3. *FEBS Lett.*, 321, 93-97

Barret, A.J., Rawlings, N.D. and Woessner, J.F. (1998) Handbook of Proteolytic Enzymes. *(Academic press, 1998)*

Barrette-Ng, I.H., Ng, K.K., Cherney, M.M., Pearce, G., Ghani, U., Ryan, C.A. and James, M.N. (2003) Unbound form of tomato inhibitor-II reveals interdomain flexibility and conformational variability in the reactive site loops. *J. Biol. Chem.*, In press.

Basalamah, M.A., Marr, H.S., Duncan, A.W. and S., E.C.J. (2001) Testican in Human Blood. *Biochem. Biophys. Res. Commun.*, 283, 1083-1090.

Beekwilder, J., Schipper, A., Bakker, P., Bosch, D. and Jongsma, M. (2000) Characterization of potato proteinase inhibitor II reactive site mutants. *Eur. J. Biochem.*, 267, 1975-1984.

Berry, J.O., Nikolau, B.J., Carr, J.P. and Klessig, D.F. (1986) Translational regulation of light-induced ribulose 1,5-bisphosphate carboxylase gene expression in amaranth. *Mol. Cell. Biol.*, 6, 2347-2353.

Bevec, T., Stoka, V., Pungercic, G., Dolenc, I. and Turk, V. (1996) Major histocompatability complexclass II-asosiated p41 invarian chain fragment is a strong inhibitor of lysosomal cathepsin L. *J. Exp. Med*, 183, 1331-1338.

Bode, W. and Hubert, R. (2000) Structural basis of the endoproteinase-protein inhibitor interaction. *Biochem. Biophys. Acta*, 1477, 241-252.

Bolter, C.J. and Jongsma, M.A. (1995) Colorado potato beetles (Leptinotarsa decemlineata) adapt to proteinase inhibitors induced in potato leaves by methyl jasmonate. *J. Insect. Physiol.*, 41, 1071-1078.

Borovsky, D., Carlson, D.A., Griffin, P.R., Shabanowitz, J. and Hunt, D.F. (1990) Mosquito oostatic factor: a novel decapeptide modulating trypsin-like enzyme biosynthesis in the midgut. *FASEB J*, 4, 3015-3020.

Bottari, A., Capocchi, A., Galleschi, L., Jopova, A. and Saviozzi, F. (1996) Asparaginyl endopeptidase during maturation and germination of durum wheat. *Physiol. plantarum*, 97, 475-480.

Bown, D., Wilkinson, H., Davison, G., Outchkourov, N.S., Jongsma, M.A. and Gatehouse, J.A. Effects of equistatin against corn rootworm in enzyme assays and artificial diet (in preparation).

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. Biochem. Mol. Biol.*, 27, 625-638.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 7, 72248-72254.

Brekken, R.A. and Sage, E.H. (2000) SPARC, a matricellular protein: At the crossroads of cell-matrix. *Matrix Biol.*, 19, 569-580.

Brekken, R.A. and Sage, E.H. (2001) SPARC, a matricellular protein: At the crossroads of cell-matrix communication. *Matrix Biology*, 19, 816-827.

Broadway, R.M. (1996) Dietary protease inhibitors alter compliment of midgut proteases. *Arch. Insect Biochem Physiol.*, 32, 39-53.

Bromme, D., Peters, K., Fink, S. and Fittkau, S. (1986) Enzyme-substrate interactions in the hydrolysis of peptide substrates by thermitase, subtilisin BPN', and proteinase K. *Arch. Biochem. Biophys.*, 1, 439-446.

Brouquisse, R., Gaudillere, J.P. and Raymond, P. (1998) Induction of a carbon-starvation-related proteolysis in whole maize plants submitted to light/dark cycles and to extended darkness. *Plant Physiol.*, 117, 1281-1291.

Busch, E., Hohenester, E., Timpl, R., Paulsson, M. and Maurer, P. (2000) Calcium affinity, cooperativity, and domain interactions of extracellular EF-hands present in BM-40. *J. Biol. Chem.*, 275, 25508-25515.

Cannon, R.J.C. (2000) Bt transgenic crops: risks and benefits. *Integr. Pest Manag. Rev.*, 5, 151-173.

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

Chen Jinq, M., Dando Pam, M., Rawlings Neil, D., Brown Molly, A., Young Nina, E., Stevens Richard, A., Hewitt, E., Watts, C. and Barrett Alan, J. (1997) Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. *J Biol. Chem.*, 272, 8090-8098.

Chen, S. and Hyman, L.E. (1998) A specific RNA-protein interaction at yeast polyadenylation efficiency elements. *Nucleic Acids Res.*, 26, 4965-4974.

Christeller John, T., Burgess Elisabeth, P.J., Mett, V., Gatehouse Heather, S., Markwick Ngaire, P., Murray, C., Malone Louise, A., Wright Michelle, A., Philip Bruce, A., Watt, D., Gatehouse Laurence, N., Lovei Gabor, L., Shannon April, L., Phung Margaret, M., Watson Lynn, M. and Laing William, A. (2002) The expression of a mammalian proteinase inhibitor, bovine spleen trypsin inhibitor in tobacco and its effects on Helicoverpa armigera larvae. *Transgen. Res.,* 11, 161-173.

Christiansen, P., Gibson, J.M., Moore, A., Pedersen, C., Tabe, L. and Larkin, P.J. (2000) Transgenic Trifolium repens with foliage accumulating the high sulphur protein, sunflower seed albumin. *Transgen. Res.*, 9, 103-113.

Cunningham, C. and Porter, A.J.R. (1998) Recombinant proteins from plants : production and isolation of clinically useful compounds. *Totowa, N.J. : Humana Press, c1998. xi*, 308.

Dai, Z., Hooker, B.S., Anderson, D.B. and Thomas, S.R. (2000) Improved plant-based production of E1 endoglucanase using potato: expression optimization and tissue targeting. *Mol. Breed.*, p. 1277-1285.

De Leo, F., Bonade Bottino, M.A., Ceci, L.R., Gallerani, R. and Jouanin, L. (1998) Opposite effects on Spodoptera littoralis larvae of high expression level on a trypsin proteinase inhibitor in transgenic plants. *Plant Physiol.*, 118, 997-1004.

De Leo, F. and Gallerani, R. (2002) The mustard trypsin inhibitor 2 affects the fertility of Spodoptera littoralis larvae fed on transgenic plants. *Insect Biochem. Mol. Biol.*, 32, 489-496.

De Leo, F., Volpicella, M., Licciulli, F., Liuni, S., Gallerani, R. and Ceci, L.R. (2002) PLANT-PIs: A database for plant protease inhibitors and their genes. *Nucleic Acids Res.*, 30, 347-348.

De Neve, M., De Loose, M., Jacobs, A., Van Houdt, H., Kaluza, B., Weidle, U., Van Montagu, M. and Depicker, A. (1993) Assembly of an antibody and its derived antibody fragment in Nicotiana and Arabidopsis. *Transgen. Res.*, 2, 227-237.

Dean, C., Favreau, M., Bedbrook, J. and Dunsmuir, P. (1989a) Sequences 5' to translation start regulate expression of petunia rbcS genes. *Plant Cell*, 1, 209-215.

Dean, C., Favreau, M., Bond Nutter, D., Bedbrook, J. and Dunsmuir, P. (1989b) Sequences downstream of translation start regulate quantitative expression of two petunia rbcS genes. *Plant Cell*, 1, 201-208.

Dean, C., Favreau, M., Dunsmuir, P. and Bedbrook, J. (1987) Confirmation of the relative expression levels of the Petunia (Mitchell) rbcS genes. *Nucleic Acids Res.*, 15, 4655-4668.

Dean, C. and Leech, R.M. (1982) Genome Expression during Normal Leaf Development. *Plant Physiol.*, 69, 904-910.

Dean, C., Pichersky, E. and Dunsmuir, P. (1989c) Structure, evolution, and regulation of RbcS genes in higher plants. *Annu. Rev. Plant Phys. Plant Mol. Biol.*, 40, 415-439.

Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P. and Bedrook, J. (1985) Different expression of the eight genes of the petunia ribulose biphosphate carboxylase small subunit multi-gene family. *EMBO J.*, 4, 3055-3061.

Deraison, C., Outchkourov, N.S., Jongsma, M.A., Rahbe, Y. and Jouanin, L. Characterisation of digestive protease of myzus and impact of cysteine protease inhibitor equistatin in vitro on 3 aphids and in planta on Myzus (in preparation).

Dolja, V.V., Peremyslov, V.V., Keller, K.E., Martin, R.R. and Hong, J. (1998) Isolation and stability of histidine-tagged proteins produced in plants via potyvirus gene vectors. *Virology*, 252, 269-274.

Donald, R.G.K. and Cashmore, A.R. (1990) Mutation of either G box or l box sequences profoundly affects expression from the Arabidopsis rbcS-1A promoter. *EMBO J.*, 9, 1717-1726.

Edwards, G.A., Hepher, A., Clerck, S.P. and Boulter, D. (1991) Pea lectin is correctly processed, stable and active in leaves of transgenic potato plants. *Plant Mol. Biol.*, 17, 89-100.

Engel, K.-H., Frenzel, T. and Miller, A. (2002) Current and future benefits from the use of GM technology in food production. *Toxicology Lett.*, 127, 329-336.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J.A., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.*, 30, 235-238.

Ferns, M.J. and Hall, Z.W. (1992) How many agrins does it take to make a synapse? *Cell*, 10, 1-3.

Ferre, J. and Van Rie, J. (2002) Biochemistry and genetics of insect resistance to Bacillus thuringiensis. *Annu. Rev. Entomol.*, 47, 501-533.

Fischer, J., Becker, C., Hillmer, S., Horstmann, C., Neubohn, B., Schlereth, A., Senyuk, V., Shutov, A. and Muntz, K. (2000) The families of papain- and legumain-like cysteine proteinases from embryonic axes and cotyledons of Vicia seeds: developmental patterns, intracellular localization and functions in globulin proteolysis. *Plant Mol. Biol.*, 43, 83-101.

Fluhr, R., Moses, P., Morelli, G., Coruzzi, G. and Chua, N.-H. (1986) Expression dynamics of the pea rbcS multigene family and organ distribution of the transcripts. *EMBO J.*, 5, 2063-2071.

Freije, J.P., Balbín, M., Abrahamson, M., Velasco, G., Dalbøge, H., Grubb, A. and López-Otín, C. (1993) Human cystatin D: cDNA cloning, characterization of the E. coli expressed inhibitor, and identification of the native protein in saliva. *J. Biol. Chem.*, 268, 15737-15744.

Frenken, L.G.J., Egmond, M.R., Batenburg, A.M. and Verrips, C.T. (1993) Pseudomonas glumae lipase: increased proteolytic stability by protein engineering. *Protein Engineer.*, 6, 637-642.

Fritz, C.C., Wolter, F.P., Schenkemeyer, V., Herget, T. and Schreier, P.H. (1993) The gene family encoding the ribulose-(1,5)-bisphosphate carboxylase/oxygenase (Rubisco) small subunit of potato. *Gene*, 137, 271-274.

Gahan, L.J., Gould, F. and Heckel, D.G. (2001) Identification of a gene associated with Bt resistance in *Heliothis virescens*. *Science*, 293, 857-860.

Galesa, K., Strukelj, B., Bavec, S., Turk, V. and Lenarcic, B. (2000) Cloning and expression of functional equistatin. *Biol. Chem.,* 381, 85-88.

Gelbart, W., et al., (2002) The FlyBase database of the Drosophila genome projects and community literature. *Nucleic Acids Res.*, 30, 106-108.

Gianessi, L.P., Silvers, C.S., Sankula, S. and Carpenter, J.E. (2002) Plant biotechnology: Current and potential impact for improving pest mamagement in U.S. agriculture. An analysis of 40 case studies.

National Center for Food and Agricultural Policy, Washington DC, USA. pp.75 **http://www.ncfap.org/40CaseStudies.htm**.

Gilles, C., Bassuk, J.A., Pulyaeva, H., Sage, E.H., Foidart, J.M. and Thompson, E.W. (1998) SPARC/osteonectin induces matrix metalloproteinase 2 activation in human breast cancer cell lines. *Cancer Res.*, 58, 5529-5536.

Gilmartin, P.M., Sarokin, L., Memelink, J. and Chua, N.-H. (1990) Molecular Light Switches for Plant Genes. *Plant Cell*, 2, 369-378.

Girard, C., Picard-Nizou, A.L., Grallien, N., Zaccomer, B., Jouanin, L. and Pham-Delegue, M.H. (1998) Effects of protease inhibitor ingestion on survival, learning abilities and digestive proteinases of the honeybee. *Transgen. Res.*, 7, 1-8.

Giri, A.P., Harsulkar, A.M., Deshpande, V.V., Sainani, M.N., Gupta, V.S. and Ranjekar, P.K. (1998) Chickpea defensive proteinase inhibitors can be inactivated by podborer gut proteinases. *Plant Physiol.*, 116.

Gittins, J.R., Pellny, T.K., Hiles, E.R., Rosa, C., Biricolti, S. and James, D.J. (2000) Transgene expression driven by heterologous ribulose-1, 5-bisphosphate carboxylase/oxygenase small-subunit gene promoters in the vegetative tissues of apple (*Malus pumila* mill.). *Planta*, 210, 232-240.

Giuliano, G., Pichersky, E., Malik, V.S., Timko, M.P., Scolnik, P.A. and Cashmore, A.R. (1988) An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. Natl. Acad. Sci. U S A*, 85, 7089-7093.

Glawe, G.A., Zavala, J.A., Kessler, A., van Dam, N.M. and Baldwin, I.T. (2003) Ecological costs and benefits are correlated with trypsin protease inhibitor production in Nicotiana attenuata. *Ecology*, in press.

Grandbastien, M.A., Berry Lowe, S., Shirley, B.W. and Meagher, R.B. (1986) Two soybean ribulose-1,5-bisphosphate carboxylase small subunit genes share extensive homology even in distant flanking sequences. *Plant Mol. Biol.*, 7, 451-465.

Green, P.J., Kay, S.A. and Chua, N.-H. (1987) Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. *EMBO J.*, 6, 2543-2549.

Green, T.R. and Ryan, C.A. (1972) Wound-induced proteinase inhibitor in plant leaves: a possible defense mechanism against insects. *Science*, 175, 776-777.

Greene, J.K., Turnipseed, S.G., Sullivan, M.J. and Herzog, G.A. (1999) Boll damage by southern green stink bug (Hemiptera: Pentatomidae) and tarnished plant bug (Hemiptera:Miridae) caged on transgenic *Bacillus thuringiensis* cotton. *J. Econ. Entomol.*, 92, 941-944.

Gruden, K., Strukelj, B., Popovic, T., Lenarcic, B., Bevec, T., Brzin, J., Kregar, I., HerzogVelikonja, J., Stiekema, W.J., Bosch, D. and Jongsma, M.A. (1998) The cysteine protease activity of Colorado potato beetle (*Leptinotarsa decemlineata* Say) guts, which is insensitive to potato protease inhibitors, is inhibited by thyroglobulin type-1 domain inhibitors. *Insect Biochem. Mol. Biol.*, 28, 549-560.

Grutter, M.G., Fendrich, G., Huber, R. and Bode, W. (1988) The 2.5 A X-ray crystal structure of the acid-stable proteinase inhibitor from human mucous secretions analyzed in its complex with bovine alpha-chymotrypsin. *EMBO J.,* 7, 345-352.

Guermah, M., Crisanti, P., Laugier, D., Dezelee, P., Bidou, L., Pessac, B. and Calothy, G. (1991) Transcription of a quail gene expressed in embryonic retinal cells is shut off sharply at hatching. *Proc. Natl. Acad. Sci. U S A*, 88, 4503-4507.

Guevara, M.G., Oliva, C.R., Huarte, M. and Daleo, G.R. (2002) An aspartic protease with antimicrobial activity is induced after infection and wounding in itercellular fluids in potato tubers. *Eur. J. Plant Path.*, 108, 131-137.

Guncar, G., Pungercic, G., Klemencic, I., Turk, V. and Turk, D. (1999) Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J.*, 154, 793-803.

Guo, Z. and Sherman, F. (1995) 3'-end-forming signals of yeast mRNA. *Mol. Cell Biol.*, 15, 5983-5990.

Guo, Z. and Sherman, F. (1996) Signals sufficient for 3'-end formation of yeast mRNA. *Mol. Cell. Biol.*, 16, 2772-2776.

Hara Nishimura, I. (1993) Vacuolar processing enzyme [in Ricinus communis, rice, maize, soyabeans, pumpkins and spinach] responsible for proprotein processing of various seed proteins. *Report of the Soy Protein Research Committee Japan*, 14, 122-126.

Hara Nishimura, I., Inoue, K. and Nishimura, M. (1991) A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS Lett.,* 294, 89-93.

Hara Nishimura, I., Shimada, T., Hiraiwa, N. and Nishimura, M. (1995) Vacuolar processing enzyme responsible for maturation of seed proteins. *J. Plant Physiol.*, 145, 632-640.

Hara Nishimura, I., Shimada, T., Kuroyanagi, M. and Yamada, K. (1998) Vacuolar sorting machinery and processing mechanism for storage proteins. *Soy Protein Research, Japan*, 1, 6-12.

Hara Nishimura, I., Takeuchi, Y. and Nishimura, M. (1993) Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. *Plant Cell*, 5, 1651-1659.

Harpster, M.H., Townsend, J.A., Jones, J.D.G., Bedbrook, J. and Dunsmuir, P. (1988) Relative strengths of the 35S cauliflower mosaic virus, 1',2', and nopaline synthase promoters in transformed tobacco sugarbeet and oilseed rape callus tissue. *Mol. Gen. Genet.*, 212, 182-190.

Harsulkar Abhay, M., Giri Ashok, P., Patankar Aparna, G., Gupta Vidya, S., Sainani Mohini, N., Ranjekar Prabhakar, K. and Deshpande Vasanti, V. (1999) Successive use of non-host plant proteinase inhibitors required for effective inhibition of Helicoverpa armigera gut proteinases and larval growth. *Plant Physiol.*, 121, 497-506.

Hartley, B.S. (1960) Proteolitic enzymes. *Ann. Rev. Biochem*, 29, 45-72.

Heath, R.L., McDonald, G., Christeller, J.T., Lee, M., Bateman, K., West, J., Van Heeswijck, R. and Anderson, M.A. (1997) Proteinase inhibitors from Nicotiana alata enhance plant resistance to insect pests. *J. Insect Physiol.*, 43, 833-842.

Hennighausen, L.G. and Sippel, A.E. (1982) Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic Acids Res.*, 10, 2677-2684.

Herr, W. and Clarke, J. (1986) The SV40 Enhancer Is Composed of Multiple Functional Elements That Can Compensate For One Another. *Cell*, 45, 461-470.

Hiatt, A., Cafferkey, R. and Bowdish, K. (1989) Production of antibodies in transgenic plants. *Nature*, 342, 76-78.

Hilder, V.A. and Boulter, D. (1999) Genetic engineering of crop plants for insect resitance - a critical review. *Crop Protection*, 18, 177-191.

Hill, A.J., Peikin, S.R., Ryan, C.A. and Blundell, J.E. (1990) Oral administration of proteinase inhibitor II from potatoes reduces energy intake in man. *Physiology and Behaviour*, 48, 241-246.

Hiraiwa, N., Nishimura, M. and Hara Nishimura, I. (1999) Vacuolar processing enzyme is self-catalytically activated by sequential removal of the C-terminal and N-terminal propeptides. *FEBS Lett.*, 447, 213-216.

Hodges, M. and Miginiac Maslow, M. (1993) The in vitro effects of ATP and protein phosphorylation on the activity of ferredoxin: NADP oxidoreductase from spinach chloroplasts. *Plant Science.*, 90, 21-29.

Hoekema, A., Huisman, M.J., Molendijk, L., Elzen, P.J.M.v.d. and Cornelissen, B.J.C. (1989) The genetic engineering of two commercial potato cultivars for resistance to potato virus X. *Bio/Technol*, 7, 273-278.

Hohenester, E., Maurer, P., Hohenadl, C., Timpl, R., Jansonius Johan, N. and Engel, J. (1996) Structure of a novel extracellular Ca-2+-binding module in BM-40. *Nature Structural Biology*, 3, 67-73.

Hohenester, E., Maurer, P. and Timpl, R. (1997) Crystal structure of a pair of fillistatin-like and EF-hand calcium-binding domains in BM-40. *EMBO J.*, 16, 3778-3786.

Hollander, M. and Wolfe, D.A. (1973) *Nonparametric statistical methods. Wiley, New York*.

Huber, R. and Bode, W. (1978) Structural basis of the activation of trypsin. *Acc. Chem. Res.*, 11, 114-122.

Huntington, J.A., Read, R.J. and Carell, R.W. (2000) Structure of a serpin-protease complex shows inhibition by deformation. *Nature*, 407, 923-926.

Inohara, N., Shimomura, S., Fukui, T. and Futai, M. (1989) Auxin-binding protein located in the endoplasmatic reticulum of maize shoots: Molecular cloning and complete primary structure. *Proc. Natl. Acad. Sci. U S A*, 86, 3564-3568.

Ishimoto, M. and Chrispeels, M.J. (1996) Protective mechanism of the Mexican bean weevil against high levels of alpha-amylase inhibitor in the common bean. *Plant Physiol.*, 111, 393-401

Irniger, S. and Braus, G.H. (1994) Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3' end formation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* 91, 257-261.

James, C. (2002) Global Review of Commercilized Transgenic Crops: 2001. *ISAAA Briefs No. 26*.

Jensen, S.E. (2000) Insecticide resistance in the western flower thrips, Frankliniella occidentalis. *Integr. Pest Manag. Rev.*, 5, 131-146.

Johnston, I.G., Paladino, T., Gurd, J.W. and Brown, I.R. (1990) Molecular cloning of SC1: a putative brain extracellular matrix glycoprotein showing partial similarity to osteonectin/BM40/SPARC. *Neuron*, 2, 165-176.

Jong, J.d., Mertens, M.M.J. and Rademaker, W. (1994) Stable expression of the GUS reporter gene in chrysanthemum depends on binary plasmid T-DNA. *Plant Cell Rep.*, 14, 59-64.

Jongsma, M.A. (1995) The resistance of insects to plant protease inhibitors. *Ph.D. thesis*.

Jongsma, M.A., Bakker, P.L., Peters, J., Bosch, D. and W.J., S. (1995) Adaptation of Spodoptera exiqua larvae to plant protease inhibitors by induction of proteinase activity insensitive to inhibition. *Proc. Natl. Acad. Sci. U S A*, 92, 8041-8045.

Jongsma, M.A., Bakker, P.L. and Stiekema, W.J. (1993) Quantitative detemination of serine proteinase inhibitor activity using a radial diffusion assay. *Anal. Biochem.*, 212, 79-84.

Jongsma, M.A., Bakker, P.L., Visser, B. and Stiekema, W.J. (1994) Trypsin inhibitor activity in mature tobacco and tomato plants is mainly induced locally in response to insect attack, wounding and virus infection. *Planta*, 195, 29-35.

Jongsma, M.A. and Bolter, C. (1997) The adaptation of insects to plant protease inhibitors. *J insect physiol*, 43, 885-895.

Jongsma, M.A., Stiekema, W.J. and Bosch, D. (1996) Combatting inhibitor-insensitive proteases of insect pests. *Trends Biotechnol.*, 14, 331-333.

Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D. and D.S., W. (2001) Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Engineer.*, 14, 993-1000.

Karban, R. and Baldwin, I.-T. (1997) Induced responses to herbivory. *Chicago : University of Chicago Press, c1997. ix*, 319.

Kawalleck, P., Somssich, I.E., Feldbrugge, M., Hahlbrock, K. and Weisshaar, B. (1993) Polyubiquitin gene expression and structural properties of the ubi4-2 gene in Petroselinum crispum. *Plant Mol. Biol.*, 21, 673-684.

Kessler, A. and Baldwin, I.T. (2002) Plant Responses to Insect Herbivory: The Emerging Molecular Analysis. *Annu. Rev. Plant. Biol.*, 53, 299-328.

Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A. and Moore, C.L. (1997) Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev. ,* 11, 2545-2556.

Kevil, C.G., Walsh, L., Laroux, F.S., Kalogeris, T., Grisham, M.B. and Alexander, J.S. (1997) An improved, rapid Northern protocol. *Biochem. Biophys. Res. Commun.*, 238, 277-279.

Khoudi, H., Laberge, S., Ferullo, J.M., Bazin, R., Darveau, A., Castonguay, Y., Allard, G., Lemieux, R. and Vezina, L.P. (1999) Production of a diagnostic monoclonal antibody in perennial alfalfa plants. *Biotechnol. Bioeng.*, 64, 135-143.

Khoudi, H., Vezina, L.P., Mercier, J., Castonguay, Y., Allard, G. and Laberge, S. (1997) An alfalfa rubisco small subunit homologue shares cis-acting elements with the regulatory sequences of the RbcS-3A gene from pea. *Gene*, 15, 343-351.

Kinoshita, T., Yamada, K., Hiraiwa, N., Kondo, M., Nishimura, M. and Hara Nishimura, I. (1999) Vacuolar processing enzyme is up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions. *Plant J,.* 19, 43-53.

Knight, M.R. and Jenkins, G.I. (1992) Genes encoding the small subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase in *Phaseolus vulgaris* L.: nucleotide sequence of cDNA clones and initial studies of expression. *Plant Mol. Biol.,* 18, 567-579.

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

Kotula, L. and Curtis, P.J. (1991) Evaluation of foreign gene codon optimization in yeast: expression of a mouse IG kappa chain. *Biotechnology,* 9, 1386-1389.

Kouzuma, Y., Inanaga, H., Doi Kawano, K., Yamasaki, N., Kimura, M., Wu, J.W. and Haard, N.F. (2000) Molecular cloning and functional expression of cDNA encoding the cysteine proteinase inhibitor with three cystatin domains from sunflower seeds. *J. Biochem.*, 128.

Koziel, M.G., Carozzi, N.B. and Desai, N. (1996) Optimizing expression of transgenes with an emphasis on post-transcriptional events. *Plant Mol. Biol.*, 32, 393-405.

Ku, M.S.B., Agarie, S., Nomura, M., Fukayama, H., Tsuchida, H., Ono, K., Hirose, S., Toki, S., Miyao, M. and Matsuoka, M. (1999) High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants. *Nature Biotechnology*, 17, 76-80.

Kuhlemeier, C., Cuozzo, M., Green, P.J., Goyvaerts, E., Ward, K. and Chua, N.H. (1988) Localization and conditional redundancy of regulatory elements in rbcS-3A, a pea gene encoding the small subunit of ribulose-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. U S A*, 85, 4662-4666.

Laskowski, M.J. and Kato, I. (1980) Protein Inhibitors of proteinases. *Ann. Rev. Biochem.*, 49, 593-626.

Lazo, G.R., Stein, P.A. and Ludwig, R.A. (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. *Bio/Technology*, 9, 963-967.

Lekstrom-Himes, J. and Xanthopoulos, K.G. (1998) Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *J. Biol. Chem.*, 30, 28545-28548.

Lenarcic, B. and Bevec, T. (1998) Thyropins--new structurally related proteinase inhibitors. *Biol. Chem.*, 379, 105-111.

Lenarcic, B., Ritonja, A., Strukelj, B., Turk, B. and Turk, V. (1997) Equistatin, a new inhibitor of cysteine proteinase from Actinia equina, is structurally related to thyroglobulin type-1 domain. *J. Biol. Chem.*, 272, 13899-13903.

Lenarcic, B. and Turk, V. (1999) Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. *J. Biol. Chem.*, 274, 563-566.

Lennon, D.A.M., Roberts, R.A., Valentijn, K., Driessen, C., Overkleeft, H.S., Erickson, A., Peters, P.J., Bikoff, E., Ploegh, H.L. and Bryant, P.W. (2001) The p41 isoform of invariant chain is a chaperone for cathepsin L. *EMBO J.*, 20, 4055-4064.

Lewis, T. (1998) Pest thrips in perspective. *In Proceedings The 1998 Brighton Conference-Pest and Diseases, No. 2, 16-19 November 1998, Brighton UK, pp.385-390. Farnham, UK: British crop protection council*.

Liddle, R.A. (1995) Regulation of cholecystokinin secretion by intraluminal releasing factors. *Am. J. Physiol.*, 269, 319-327.

Lin Cereghino, G.P., Sunga, A.J., Lin Cereghino, J. and Cregg, J.M. (2001) Expression of foreign genes in the yeast *Pichia pastoris*. *Genet. Eng.*, 23, 157-169.

Loomans, A.J.M., Van Lentern, J.C., Tommasini, M.G., Maini, S. and Riudavets, J. (1995) Biological control of Western flower thrips pests. *Wageningen Agricultural Papers*, 1, 201pp.

Lorito, M., Broadway, R.M., Hayes, C.K., Woo, S.L., Noviello, C., Williams, D.L. and Harman, G.E. (1994) Proteinase inhibitors from plants as a novel class of fungicides. *Mol. Plant Microb. Iinteract.*, 7, 525-527.

Lu, S.M., et al., (2001) Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. *Proc. Natl. Acad. Sci. U S A*, 98, 1410-1415.

Lynch, R.E., Wiseman, B.R., Sumner, H.R., Plaisted, D. and Warnick, D. (1999) Management of corn earworm and fall armyworm (Lepidoptera: Noctuidae) injury on a sweet corn hybrid expressiong a cryIA(b) gene. *J. Eecon. Entomol.*, 92, 1217-1222.

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

Markert, Y., Koeditz, J., Mansfeld, J., Arnold, U. and Ulbrich, H.R. (2001) Increased proteolytic resistance of ribonuclease A by protein engineering. *Protein Engineer.*, 14, 791-796.

Markwick, N.P., Laing, W.A., Christeller, J.T., McHenry, J.Z. and Newton, M.R. (1998) Overproduction of digestive enzymes compensates for inhibitory effects of protease and alpha-amylase inhibitors fed to three species of leafrollers (Lepidoptera: Tortricidae). *J. Econ. Entomol.*, 91, 1265-1276.

Markwick, N.P., Laing, W.A., Christeller, J.T., Reid, S.J. and Newton, M.R. (1996) alpha-Amylase activities in larval midgut extracts from four species of Lepidoptera (Tortricidae and Gelechiidae): response to pH and to inhibitors from wheat, barley, kidney beans, and Streptomyces. *J. Econ. Entomol.*, 89, 39-45.

Marr, H.S., Basalamah, M.A., Bouldin, T.W., Duncan, A.W. and Edgel, C.J.S. (2000) Distribution of testican expression in human brain. *Cell Tissue Res.*, 302, 139-144.

Marr, H.S., Basalamah, M.A. and Edgel, C.J.S. (1997) Endothelial cell expression of testican mRNA. *Endothelium*, 5, 209-219.

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

McElroy, D., Blowers, A.D., Jenes, B. and Wu, R. (1991) Construction of expression vectors based on the rice actin 1 (Act1) 5' region for use in monocot transformation. *Mol. Gen. Genet.*, 231, 150-160.

Meagher, R.B., Berry-Lowe, S. and Rice, K. (1989) Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. *Genetics*, 123, 845-863.

Menkens, A.E., Schindler, U. and Cashmore, A.R. (1995) The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. *Trends Biochem. Sci.*, 20, 506-510.

Michaud, D., Bernier Vadnais, N., Overney, S. and Yelle, S. (1995a) Constitutive expression of digestive cysteine proteinase forms during development of the colorado potato beetle, Leptinotarsa decemlineata Say (Coleoptera: Chrysomelidae). *Insect Biochem. Mol. Biol.*, 25, 1041-1048.

Michaud, D., Cantin, L. and Vrain Thierry, C. (1995b) Carboxy-Terminal Truncation of Oryzacystatin II by Oryzacystatin-Insensitive Insect Digestive Proteinases. *Arch. Biochem. Biophys.*, 322, 469-474.

Michaud, D., NguyenQuoc, 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. *Arch. Insect Biochem. Physiol.*, 31, 451-464.

Miller, E.A., Lee, M.C., Atkinson, A.H. and Anderson, M.A. (2000) Identification of a novel four-domain member of the proteinase inhibitor II family from the stigmas of Nicotiana alata. *Plant Mol. Biol.*, 42, 329-333.

Minvielle Sebastia, L., Beyer, K., Krecic, A.M., Hector, R.E., Swanson, M.S. and Keller, W. (1998) Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. *EMBO J.*, 17, 7454-7468.

Mittl, P.R., Di Marco, S., Fendrich, G., Pohlig, G., Heim, J., Sommerhoff, C., Fritz, H., Priestle, J.P. and Grutter, M.G. (1997) A new structural class of serine protease inhibitors revealed by the structure of the hirustasin–kallikrein complex. *Structure*, 15, 253-264.

Mlynarova, L., Loonen, A., Heldens, J., Jensen, R.C., Keizer, P., Stiekema, W.J. and Nap, J.P. (1994) Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. *Plant Cell*, 6, 417-426.

Molina, F., Bouanani, M., Pau, B. and Granier, C. (1996) Characterization of the type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. *Eur. J. Biochem*, 240, 125-133.

Motamed, K. (1999) SPARC (osteonectin/BM-40). *Int. J. Biochem. Cell Biol.*, 31, 1363-1366.

Mueller, F.B., Hausser, I., Berg, D., Casper, C., Maiwald, R., Jung, A., Jung, H. and Korge, B.P. (2002) Genetic analysis of a severe case of Netherton syndrome and application for prenatal testing. *British Journal of Dermatology.*, 146, 495-499.

Muntz, K. and Shutov, A. (2002) Legumains and their functions in plants. *Trends Plant Sci.*, 7, 340-344.

Murai, T. (1990) Rearing method and biology of thrips parasitoid, Ceranisus menes. IOBC/WPRS Bulletin 13: 142-146. *IOBC/WPRS Bulletin*, 13, 142-146.

Nakada, M., Yamada, A., Takino, T., Miyamori, H., Takahashi, T., Yamashita, J. and Sato, H. (2001) Suppression of membrane-type 1 matrix metalloproteinase (MMP)-mediated MMP-2 activation and tumor invasion by testican 3 and its splicing variant gene product, N-Tes. *Cancer Res.,* 15, 8896-902.

Nap, J.P., van Spanje, M., Dirkse, W.G., Baarda, G., Mlynarova, L., Loonen, A., Grondhuis, P. and Stiekema, W.J. (1993) Activity of the promoter of the Lhca3.St.1 gene, encoding the potato apoprotein 2 of the light-harvesting complex of Photosystem I, in transgenic potato and tobacco plants. *Plant Mol. Biol.*, 23, 605-612.

Niimi, T., Hiroaki, Y., Goto, A., Beck, K. and Kitagawa, Y. (1999) A Drosophila gene encoding multiple splice variants of Kazal-type serine protease inhibitor-like proteins with potential destinations of mitochondria, cytosol and the secretory pathway. *Eur. J. Biochem.*, 266, 282-292.

Nishimoto, S., Hamajima, Y., Toda, Y., Toyoda, H., Kitamura, K. and Komurasaki, T. (2002) Identification of a novel smooth muscle associated protein, smap2, upregulated during neointima formation in a rat carotid endarterectomy model. *Biochim. Biophys. Acta.*, 1576, 225-230.

Nomura, M., Katayama, K., Nishimura, A., Ishida, Y., Ohta, S., Komari, T., Miyao-Tokutomi, M., Tajima, S. and Matsuoka, M. (2000) The promoter of rbcS in a C3 plant (rice) directs organ-specific, light-dependent expression in a C4 plant (maize), but does not confer bundle sheath cell-specific expression. *Plant Mol. Biol.*, 44, 99-106.

Oerke, E.C., Dehne, H.W., Schonbeck, F. and Weber, A. (1994) *Crop Production and Crop Protection: Estimated Losses in Major Food and Cash Crops*. Elsevier, Amsterdam.

Ohuchi, E., Imai, K., Fujii, Y., Sato, H., Seiki, M. and Okada, Y. (1997) Membrane type 1 matrix metalloproteinase digests interstitial collagens and other extracellular matrix molecules. *J. Biol. Chem.*, 272, 2446-2451.

Orozco-Cardenas, M., McGurl, B. and Ryan, C. (1993) Expression of an Antisense Prosystemin Gene in Tomato Plants Reduces Resistance Toward Manduca sexta Larvae. *Proc. Natl. Acad. Sci. U S A*, 90, 8273-8276.

Otsu, M., Urade, R., Kito, M., Omura, F. and Kikuchi, M. (1995) A possible role of ER-60 protease in the degradation of misfolded proteins in the endoplasmic reticulum. *J. Biol. Chem.*, 270, 14528-14961.

Outchkourov, N.S., de Kogel, W.J., Schuurman-de Bruin, A., Abrahamson, M. and Jongsma, M.A. Specific cysteine protease inhibitors act as deterrents of Western flower thrips *Frankliniella occidentalis* (Pergande) in transgenic potato. *submitted.*

Outchkourov, N.S., de Kogel, W.J., Wiegers, G.L., Abrahamson, M. and Jongsma, M.A. Engineered multi-domain cysteine protease inhibitors yield resistance against western flower thrips (*Frankliniella Occidentalis*) in greenhouse trials. *submitted.*

Outchkourov, N.S., Peters, J., de Jong, J., Rademakers, W. and Jongsma, M.A. (2003a) The promoter-terminator of chrysanthemum rbcS1 directs very high foreign gene expression levels in dicotyledonous plants. *Planta*, 216, 1003-1112.

Outchkourov, N.S., Rogelj, B., Dulev, S., Turk, D. and Jongsma, M.A. Engineering protein stability of equistatin in potato plants. *submitted.*

Outchkourov, N.S., Rogelj, B., Strukelj, B. and Jongsma, M.A. (2003b) Expression of sea anemone equistatin in potato: effects of plant proteases on heterologous protein production. *Plant Physiol.*, 133, 379-390

Outchkourov, N.S., Stiekema, W.J. and Jongsma, M.A. (2002) Optimization of the expression of equistatin in Pichia pastoris. *Protein Express. Purif.*, 24, 18-24.

Ow, D.W., Jacobs, J.D. and Howell, S.H. (1987) Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. *Proc. Natl. Acad. Sci. U S A*, 84, 4870-4874.

Palumbi, S.R. (2001) Humans are the world's greatest evolutionary force. *Science*, 293, 1786-1790.

Parrella, M.P. (1995) IPM-approachesand prospects, pp. 357-364. *In* B.L.Parker, M.Skinner and T.Lewis [eds.], Thrips biology and managenent. Plenum, New York.

Pearce, G., Ryan, C.A. and Liljegren, D. (1988) Proteinase inhibitor I and II in in fruit of wild tomato species: Transient components of a mechanisms for defence and seed dispercal. *Planta*, 175, 527-531.

Peferoen, M. (1997) Progress and prospects for field use of Bt genes in crops. *Trends in Biotechnology*, 15, 173-177.

Pesole, G., Liuni, S., Grillo, G., Liccuiulli, F., Mignone, F., Gissi, C. and Saccone, C. (2002) UTRdb and UTRsite: specialized databases of sequences and functional elements of 5'and 3'untranslated regions of eucariotic mRNAs. *Nucleic Acids Res.*, 30, 335-340.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Lucciulli, F. and Liuni, S. (2001) Structural and functional features of eucaryotic mRNA untranslated regions. *Gene*, 276, 73-81.

Pham, D.M.H., Girard, C., Le Metayer, M., L., P.N.A., Hennequet, C., Pons, O. and Jouanin, L. (2000) Long-term effects of soybean protease inhibitors on digestive enzymes, survival and learning abilities of honeybees. *Ent. Exp. App.*, 95, 21-29.

Pozueta Romero, J., Houlne, G., Canas, L., Schantz, R. and Chamarro, J. (2001) Enhanced regeneration of tomato and pepper seedling explants for Agrobacterium-mediated transformation. *Plant Cell Tissue and Organ Culture*, 67, 173-180.

Qaim, M. and Zilberman, D. (2003) Yield effects of genetically modified crops in developing countries. *Science,* 299, 900-902.

Raharjo Sunu, B., Emoto, N., Ikeda, K., Sato, R., Yokoyama, M. and Matsuo, M. (2001) Alternative splicing regulates the endoplasmic reticulum localization or secretion of soluble secreted endopeptidase. *J. Biol. Chem.*, 276, 25612-25620.

Rahbe, Y., Ferrasson, E., Rabesona, H. and Quillien, L. (2003) Toxicity to the pea aphid Acyrthosiphon pisum of anti-chymotrypsin isoforms and fragments of Bowman-Birk protease inhibitors from pea seeds. *Insect Biochem. Mol. Biol.*, 33, 299-306.

Ramos, J.A., Zenster, N., Leyser, O. and Callis, J. (2001) Rapid degradation of Auxin/Indoleacetic acid proteins requires conserved aminoacids of domain II and is proteasome dependant. *Plant cell*, 13, 2349-2360.

Rawlings, N.D. and Barrett, A.J. (1990) Evolution of proteins of the cystatin superfamily. *J. Mol. Evol*, 30, 60-71.

Robb, K.L. and Parrella, M.P. (1995) Insecticide resistance in western flower thrips . In B.L. Parker, M. Skinner and T. Lewis (eds) Thrips biology and management, pp. 365-370. New York: Plenum press.

Rodriguez, G.M. and Diment, S. (1995) Destructive proteolysis by cysteine proteases in antigen presentation of ovalbumin. *Eur. J. Immunol*, 25, 1823-1827.

Rogelj, B., Strukelj, B., Bosch, D. and Jongsma, M.A. (2000) Expression, purification, and characterization of equistatin in Pichia pastoris. *Protein Expr. Purif.,* 19, 329-334.

Romanos, M.A., Scorer, C.A. and Clare, J.J. (1992) Foreign gene expression in yeast: a review. *Yeast,* 8, 423-488.

Rose, A., Meier, I. and Wienand, U. (1999) The tamato I-box binding factor LeMYBI is a member of a novel class of Myb-like proteins. *Plant J.*, 20, 641-652.

Rotari, V.I., Dando, P.M. and Barrett, A.J. (2001) Legumain forms from plants and animals differ in their specificity. *Biol Chem*, 382, 953-959.

Royo, J., Leon, J., Vancanneyt, G., Albar, J.P., Rosahl, S., Ortego, F., Castanera, P. and Sanchez-Serrano, J.J. (1999) Antisense-mediated depletion of a potato lipoxygenase reduces wound induction of proteinase inhibitors and increases weight gain of insect pests. *Proc. Natl. Acad. Sci. U S A*, 96, 1146-1151.

Russo, P., Li, W.Z., Guo, Z. and Sherman, F. (1993) Signals that produce 3' termini in CYC1 mRNA of the yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.*, 13, 7836-7849.

Ryan, C.A. (1978) Proteinase inhibitors in plant leaves: a biochemical model for pest-induced natural plant protection. *Trends Biochem. Sci.,* 5, 148-150.

Ryan, C.A. (1990) Protease inhibitors in plants: genes for improving defences against insects and pathogens. *Annu. Rev. Phytopathol.*, 28, 425-449.

Sadis, S., Atienza, C.J. and Finley, D. (1995) Synthetic signals for ubiquitin-dependant proteolysis. *Mol. Cell. Biol*., 15, 4086-4094.

Salier, J.P. (1990) Inter-alpha-trypsin inhibitor: emergence of a family within the Kunitz-type protease inhibitor superfamily. *Trends Biochem Sci.*, 15, 435-439.

Schimmel, P.R. and Flory, P.J. (1968) Conformational energies and configurational statistics of copolypeptides containing L-proline. *J. Mol. Biol.*, 28, 105-120.

Schirra, H.J., Scanlon, M.J., Lee, M.C.S., Anderson, M.A. and Craik, D.J. (2001) The Solution Structure of C1-T1, a Two-domain Proteinase Inhibitor Derived from a Circular Precursor Protein from *Nicotiana alata*. *J. Mol. Biol.*, 306, 69-79.

Schlereth, A., Standhardt, D., Mock Hans, P. and Muentz, K. (2001) Stored cysteine proteinases start globulin mobilization in protein bodies of embryonic axes and cotyledons during vetch (*Vicia sativa* L.) seed germination. *Planta,* 212, 718-727.

Schmitt, M.E., Brown, T.A. and Trumpower, B.L. (1990) A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Res.*, 18, 3091-3092.

Schneyer, A.L., Fujiwara, T., Fox, J., Welt, C.K., Adams, J., Messerlian, G.M. and Taylor, A.E. (2000) Dynamic changes in the intrafollicular inhibin/activin/follistatin axis during human follicular development: Relationship to circulating hormone concentrations. *Journal of Clinical Endocrinology and Metabolism*, 85, 3319-3330.

Schouten, A., Roosien, J., Engelen, F.A. van., Jong, G.A.M. de., Borst Vrenssen, A.W.M., Zilverentant, J.F., Bosch, D., Stiekema, W.J., Gommers, F.J. and Schots, A. (1996) The C-terminal KDEL sequence increases the expression level of a single-chain antibody design to be targeted to both the cytosol and the secretory pathway in transgenic tobacco. *Plant Mol. Biol.*, 30, 781-793.

Schuler, T.H., Poppy, G.M., Kerry, B.R. and Denholm, I. (1998) Insect resistant transgenic plants. *Trends Biotechnol.*, 16, 168-175.

Scott, M.J., Huckaby, C.S., Kato, I., Kohr, W.J., Laskowski, M.J., Tsai, M.J. and O'Malley, B.W. (1987) Ovoinhibitor introns specify functional domains as in the related and linked ovomucoid gene. *J. Biol. Chem.*, 25, 5899-5907.

Seiichi, F., de Jong, J. and Rademaker, W. (1995) Efficient genetic transformation of chrysanthemum (*Dendranthema grandiflorum* (Ramat.) Kitamura) using stem segments. *Breeding Science*, 45, 179-184.

Sharp, J.M. and Doran, P.M. (2001a) Characterization of monoclonal antibody fragments produced by plant cells. *Biotechnol. Bioengineer*., 73, 338-346.

Sharp, J.M. and Doran, P.M. (2001b) Strategies for enhancing monoclonal antibody accumulation in plant cell and organ cultures. *Biotechnology progress*, 17, 979-992.

Shelton, A.M., Zhao, J.-Z. and Roush, R.T. (2002) Economic, Ecological, Food safety, and social consequences of the deployment of BT transgenic plants. *Annu. Rev. Entomol.*, 47, 845-881.

Shipp, J.L., Binns, M.R., Hao, X. and Wang, K. (1998a) Economic injury levels for western flower thrips (Thysanoptera: Thripidae) on greenhouse sweet pepper. *J. Econ. Entomol.*, 91, 671-677.

Shipp, J.L., Boland, G.J. and Shaw, L.A. (1991) Integrated pest management of disease and arthropod pests of greenhouse vegetable crops in Ontario: current status and future possibilities. *Can. J. Plant Sci.*, 71, 887-914.

Shipp, J.L., Hao, X., Papadopoulos, A.P. and Binns, M.R. (1998b) Impact of western flower thrips (Thysanoptera: Thripidae) on growth, photosynthesis and productivity of reenhouse sweet pepper. *Sci. Hortic.*, 72, 87-102.

Shipp, J.L., Wang, K. and Binns, M.R. (2000) Economic injury levels for western flower thrips (Thysanoptera: Thripidae) on greenhouse cucumber. *J. Econ. Entomol.*, 93, 1732-1740.

Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., Gettins, P.G.W., Irving, J.A., Lomas, D.A., Luke, C.J., Moyer, R.W., Pemberton, P.A., Remold-O'Donnell, E., Salvesen, G.S., Travis, J. and Whisstock, J.C. (2001) The Serpins Are an Expanding Superfamily of Structurally Similar but Functionally Diverse Proteins. EVOLUTION, MECHANISM OF INHIBITION, NOVEL FUNCTIONS, AND A REVISED NOMENCLATURE. *J. Biol. Chem.*, 276, 33293-33296.

Smith, C.M. (1999) Plant Resistance to Insects. In Rechcigl, J.a.R., N. (ed.), *Biological and Biotechnological Control of Insects*. Lewis Publishers, Boca Raton, pp. 171 - 205.

Song, H.K., Kim, Y.S., Yang, J.K., Moon, J., Lee, J.Y. and Suh, S.W. (1999) Crystal structure of a 16 kDa double-headed Bowman-Birk trypsin inhibitor from barley seeds at 1.9 angstrom resolution. *J. Mol. Biol.*, 293, 1133-1144..

Spates, G.E. and Harris, R.L. (1984) Reduction of fecundity, egg hatch, and survival in adult horn flies fed protease inhibitors. *Southwest Entomol. College Station, Tex. : Southwestern Entomological Society. Dec 1984*, 9, 399-403.

Sreekrishna, K., Brankamp, R.G., Kropp, K.E., Blankenship, D.T., Tsay, J.T., Smith, P.L., Wierschke, J.D., Subramaniam, A. and Birkenberger, L.A. (1997) Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene*, 190, 55-62.

Stevens, L.H., Stoopern, G.M., Elbers, I.J.W., Molthoff, J.W., Bakker, H.A.C., Lommen, A., Bosch, D. and Jordi, W. (2000) Effect of climate conditions and plant developmental stage on the stability of antibodies expressed in transgenic tobacco. *Plant Physiol*, 124, 173-182.

Strizhov, N., Keller, M., Mathur, J., Koncz Kalman, Z., Bosch, D., Prudovsky, E., Schell, J., Sneh, B., Koncz, C. and Zilberstein, A. (1996) A synthetic cryIC gene, encoding a Bacillus thuringiensis delta-endotoxin, confers Spodoptera resistance in alfalfa and tobacco. *Proc. Natl. Acad. Sci. U S A.*, 93, 15012-15017.

Strukelj, B., Lenarcic, B., Gruden, K., Pungercar, J., Rogelj, B., Turk, V., Bosch, D. and Jongsma, M.A. (2000) Equistatin, a protease inhibitor from the sea anemone actinia equina, is composed of three structural and functional domains. *Biochem. Biophys. Res. Commun.*, 269, 732-736.

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

Sugita, M., Manzara, T., Pichersky, E., Cashmore, A. & Gruissem, W. (1987) Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. *Mol. Gen. Genet.,* 209, 247-256.

Taoka, K., Kaya, H., Nakayama, T., Araki, T., Meshi, T. and Iwabuchi, M. (1999) Identification of three kinds of mutually related composite elements conferring S phase-specific transcriptional activation. *Plant J.*, 18, 611-623.

Terra, W.R. and Ferreira, C. (1994) Insect digestive enzymes: properties, compartmentalization and function. *Comp. Biochem. Physiol.*, 109, 1-62.

Thaler, J.S., Stout, M.J., Karban, R. and Duffey, S.S. (2001) Jasmonate-mediated induced plant resistance affects a community of herbivores. *Ecol. Entomol.*, 26, 312-324.

Tommasini, M.G. and Maini, S. (1995) *Frankiniella occidentalis* and other thrips harmful to vegetable and ornamental crops in Europe. in A.J.M. Loomans, J.C. van lenteren, M.G. Tommasini, S. Maini, and J. Riudavets (eds) Bilogical control of thrips pests. pp.1-42. Wageningen: Wageningen Agricultural University Papers.

Tortiglione, C., Fanti, P., Pennacchio, F., Malva, C., Breuer, M., De Loof, A., Monti, L.M., Tremblay, E. and Rao, R. (2002) The expression in tobacco plants of Aedes aegypti Trypsin Modulating Oostatic Factor (Aea-TMOF) alters growth and development of the tobacco budworm, *Heliothis virescens*. *Mol. Breed.*, 9, 159-169.

Toyooka, I., Okamoto, T. and Minamikawa, T. (2000) Mass Transport of proform of a KDEL-tailed cysteine proteinase (SH-EP) to protein storage vacuoles by endoplsmic reticulum derived vesicle is involved in protein mobilization in germinating seeds. *J.f Cell Biol.*, 148, 453-463.

Trexler, M., Banyai, L. and Patthy, L. (2001) A human protein containing multiple types of protease-inhibitory modules. *Proc. Natl. Acad. Sci. U S A*, 98, 3705-3709.

Tsunemi, M., Matsuura, Y., Sakakibara, S. and Katsube, Y. (1996) Crystal structure of an elastase-specific inhibitor elafin complexed with porcine pancreatic elastase determined at 1.9 A resolution. *Biochemistry*, 35, 11570-11576.

Tuomi, J., Augner, M. and Nilsson, P. (1994) A Dilemma of Plant Defences: Is it Really Worth Killing the Herbivore? *J. Theor. Biol.*, 170, 427-430.

Turnipseed, S.G., Sullivan, M.J., Mann, J.E. and Roof, M.E. (1995) Secondary pests in transgenic Bt cotton in South Carolina. *Proc Beltwide Cotton Conf. Memphis, Tenn. : National Cotton Council of America*, 2, 768-769.

Ueda, T., Seo, S., Ohashi, Y. and Hashimoto, J. (2000) Circadian and senescence-enhanced expression of a tobacco cysteine protease gene. *Plant Mol. Biol.*, 44, 649-657.

Urao, T., Yamaguchi_Shinozaki, K., Urao, S. and Shinozaki, K. (1993) An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell*, 5, 1529-1539.

Urwin, N.A. and Jenkins, G.I. (1997) A sucrose repression element in the Phaseolus vulgaris rbcS2 gene promoter resembles elements responsible for sugar stimulation of plant and mammalian genes. *Plant Mol. Biol.*, 35, 929-942.

Urwin, P.E., McPherson, M.J. and Atkinson, H.J. (1993) Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs. *Planta*, 204, 472-479.

van Dam, N.M. and Baldwin, I.T. (1998) Cost of jasmonate-induced responses in plants competing for limited resources. *Ecol. Lett.*, 1, 1-5.

van Dam, N.M., Hadwich, K. and Baldwin, I.T. (2000) Induced responses in Nicotiana attenuata affect behavior and growth of the specialist herbivore *Manduca sexta*. *Oecologia,* 122, 371-379.

van Dam, N.M., Horn, M., Mares, M. and Baldwin, I.T. (2001) Ontogeny constrains the systemic proteinase inhibitor response in Nicotiana attenuata. *J. Chem. Ecol.*, 27, 547-568.

van Engelen, F.A., Molthoff, J.W., Conner, A.J., Nap, J.P., Pereira, A. and Stiekema, W.J. (1995) pBINPLUS: an improved plant transformation vector based on pBIN19. *Transgen. Res.*, 4, 288-290.

van Helden, J., del Olmo, M. and Perez Ortin, J.E. (2000) Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic Acids Res.*, 28, 1000-1010.

Vannahme, C., Smyth, N., Miosge, N., Gosling, S., Frie, C., Paulsson, M., Maurer, P. and Hartmann, U. (2002) Characterization of SMOC-1, a novel modular calcium-binding protein in basement membranes. *J. Biol. Chem.*, 277, 37977-37986.

Varallyay, E., Pal, G., Patthy, A., Szilagyi, L. and Graf, L. (1998) Two mutations in rat trypsin confer resistance against autolysis. *Biochem. Biophys. Res. Comm.*, 243, 56-60.

Verwoerd, T.C., Dekker, B.M. and Hoekema, A. (1989) A small-scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Res.*, 17, 2362.

Viale, A., Ortola, C., Hervieu, G., Furuta, M., Barbero, P., Steiner Donald, F., Seidah Nabil, G. and Nahon Jean, L. (1999) Cellular localization and role of prohormone convertases in the processing of pro-melanin concentrating hormone in mammals. *J. Biol. Chem.*, 274, 6536-6545.

Vierstra, R.D. (1993) Protein degradation in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol*, 44, 385-410.

Vierstra, R.D. (1996) Proteolysis in plants: mechanisms and functions. *Plant Mol.Biol.*, 32, 275-302.

Visser, R.G.F., Jacobsen, E., Hesseling Meinders, A., Schans, M.J., Witholt, B. and Feenstra, W.J. (1989) Transformation of homozygous diploid potato with an Agrobacterium tumefaciens binary vector system by adventitious shoot regeneration on leaf and stem segments. *Plant Mol. Biol.*, 12, 329-337.

Vitale, A. and Galili, G. (2001) The endomembrane system and the problem of protein sorting. *Plant Physiol.*, 125, 115-118.

Volpicella, M., Ceci, L.R., Cordewener, J., America, T., Gallerani, R., Bode, W., Jongsma, M.A. and Beekwilder, J. (2003) Properties of purified gut trypsin from Helicoverpa zea , adapted to proteinase inhibitors. *Eur. J. Biochem.*, 270, 10-19.

Volpicella, M., Schipper, A., Jongsma, M.A., Spoto, N., Gallerani, R. and Ceci, L.R. (2000) Characterization of recombinant mustard trypsin inhibitor 2 (MTI2) expressed in Pichia pastoris. *FEBS Lett.*, 468, 137-141.

Waksman, G., Lebrun, M. and Freyssinet, G. (1987) Nucleotide sequence of a gene encoding sunflower ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcs). *Nucleic Acids Res.*, 15, 7181.

Waldron, C., Wegrich, L.M., Merlo, P.A.O., Walsh, T.A. and Strickland, J.A. (1993a) Characterization of a genomic sequence coding for potato multicystatin, and eight-domain cysteine proteinase inhibitor. *Plant Mol. Biol.,* 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 Physiol.,* 103*,* 1227-1234.

Wandelt, C.I., Khan, M.R.I., Craig, S., Schroeder, H.E., Spencer, D. and Higgins, T.J.V. (1992) Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenic plants. *Plant J.*, 2, 181-192.

Welt, C.K. (2002) The physiology and pathophysiology of inhibin, activin and follistatin in female reproduction. *Curr. Opin. Obstet. Gynecol.*, 14, 317-323.

Welt, C.K., Lambert, M.G., Zheng, W., Crowley, W.F.J. and Schneyer, A.L. (1997) Presence of activin, inhibin, and follistatin in epithelial ovarian carcinoma. *Journal of Clinical Endocrinology and Metabolism.,* 82, 3720-3727.

Wiedow, O., Schroder, J.M., Gregory, H., Young, J.A. and Christophers, E. (1990) Elafin: an elastase-specific inhibitor of human skin. Purification, characterization, and complete amino acid sequence [published erratum appears in *J. Biol. Chem.* 1991 Feb 15;266(5):3356]. *J. Biol. Chem.*, 256, 14791-14795.

Wijkamp, I., Almarza, N., Goldbach, R. and Peters, D. (1995) Distinct levels of specificity in thrips transmision of Tospoviruses. *Phytopathology*, 85, 1069-1074.

Wilbur, W.J. and Lipman, D.J. (1983) Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. U S A*, 80, 726-730.

Wolfson, J.L. and Murdock, L.L. (1990) Growth of Manduca sexta on wounded tomato plants: role of induced proteinase inhibitors. *Entomol. Exp. Appl.*, 54, 257-264.

Wong, E.Y., Hironaka, C.M. and Fischhoff, D.A. (1992) Arabidopsis thaliana small subunit leader and transit peptide enhance the expression of *Bacillus thurigiensis* proteins in transgenic plants. *Plant Mol. Biol.*, 20, 81-93.

Wu, J.W. and Haard, N.F. (2000) Purification and characterization of a cystatin from the leaves of methyl jasmonate treated tomato plants. *Comp. Biochem. Physiol.,* 3, 209-220.

Wu, Y., Llewellyn, D., Mathews, A. and Dennis Elizabeth, S. (1997) Adaptation of Helicoverpa armigera (Lepidoptera: Noctuidae) to a proteinase inhibitor expressed in transgenic tobacco. *Mol. Breed.*, 3, 371-380.

Yan, L. and Kerr, P.S. (2002) Genetically engineered crops: their potential use for improvement of human nutrition. *Nutr Rev.*, 60, 135-141.

Zwijsen, A., Blockx, H., Van Arnhem, W., Willems, J., Fransen, L., Devos, K., Raymackers, J., Van De Voorde, A. and Slegers, H. (1994) Characterization of a rat C-6 glioma-secreted follistatin-related protein (FRP): Cloning and sequence of the human homologue. *Eur. J. Biochem.*, 225, 937-946.