8-21-2014 12:00 AM

Identification and characterization of cysteine protease genes in tobacco for use in recombinant protein production

Kishor Duwadi, *The University of Western Ontario*

Supervisor: Dr. Sangeeta Dhaubhadel, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

IDENTIFICATION AND CHARACTERIZATION OF *CYSTEINE PROTEASE* GENES IN TOBACCO FOR USE IN RECOMBINANT PROTEIN PRODUCTION

(Thesis format: Monograph)

by

Kishor Duwadi

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

Plants are an attractive host system for pharmaceutical protein production. Many therapeutic proteins have been produced and scaled up in plants at a low cost compared to the conventional microbial and animal based systems. The main technical challenge during this process is to produce sufficient level of proteins in plants. Low yield is generally caused by proteolytic degradation during expression and downstream processing of recombinant proteins. The yield of a human therapeutic protein interleukin (IL) -10 produced in transgenic tobacco leaves was found to be below the critical level, and is potentially due to degradation by tobacco cysteine proteases (CysPs). A total of 60 putative *CysP* genes were identified in tobacco and based on their expression in the leaf tissue, 10 candidate *CysP*s (*CysP1-CysP10*) were selected for further characterization. The effects of silencing and overexpressing these *CysP*s on IL-10 accumulation were examined in tobacco. It was found that the recombinant protein yield in tobacco could be increased by stably silencing *CysP6*. Transient expression of *CysP6* silencing construct also showed an increase in IL-10 accumulation in comparison to the controls. Moreover, CysP6 is localized to the Endoplasmic Reticulum (ER), suggesting that ER is the potential site of IL-10 degradation. Overall results suggest that *CysP6 is* involved in determining the yield of recombinant IL-10 in tobacco.

Key Words

Cysteine protease, tobacco, Interleukin-10, protein degradation, endoplasmic reticulum, silencing, ELISA

Acknowledgements

I express my gratitude to my supervisor Dr. Sangeeta Dhaubhadel for her invaluable assistance and guidance throughout my Master's program. She has always encouraged me to independently plan and conduct my experiments, and has provided useful comments on necessities. Furthermore, I thank my co-supervisor Dr. Denis Maxwell and my advisors Dr. Rima Menassa and Dr. Susanne Kohalmi for their helpful suggestions and advice during my graduate studies. Also, I thank current and previous members of Dr. Dhaubhadel Lab, Ling Chen, Dr. Jaeju Yu, Mehran Dastmalchi, Arun Kumaran Anguraj Vadivel, Hemanta Raj Mainali, Caroline Sepiol and Pravesh Lama for creating a wonderful working atmosphere and providing critical feedbacks on the project. I express gratitude to my loved ones, who have all supported me throughout the entire process. I will be forever grateful for your love and support.

Table of Contents

Abstract ..... **ii**

Acknowledgements ..... **iii**

Table of Contents ..... **iv**

List of Tables ..... **viii**

List of Figures ..... **ix**

List of Abbreviations ..... **x**

Chapter 1: Introduction **1**

1.1 Recombinant protein production systems 1

1.2 Recombinant protein expression in plants 3

1.3 Stable and transient protein expression systems 5

1.4 Challenges of recombinant protein production in plants 8

1.5 Strategies in improving protein yield 10

1.6 Plant proteases 11

1.7 CysPs 12

1.8 Interleukin-10 16

1.9 Hypothesis 16

1.10 Objectives 17

Chapter 2: Materials and Methods **18**

2.1 Plant materials and growth conditions 18

2.2 Bacterial strains 18

`2.3 *In silico* analysis 19

2.4 Cloning 19

2.4.1 Cloning for silencing constructs 19

2.4.2 Cloning of *CysP* genes for overexpression 21

2.4.3 Cloning for subcellular localization 25

2.5 Tobacco transformations 26

2.5.1 Stable transformation and generation of transgenic lines 26

2.5.2 Transient expression 27

2.6 Confocal Microscopy 27

2.7 Sample collection for protein and RNA extractions 28

2.8 Protein extraction 28

2.9 Protein assay 28

2.10 Enzyme-linked immunosorbent assay (ELISA) 29

2.11 Genomic DNA extraction and genotyping of T_1 transgenic lines 30

2.12 RNA isolation and RT-PCR analysis 30

Chapter 3: Results 32

3.1 Tobacco genome contains at least 60 putative CysPs 32

3.2. Catalytic triad and non-contigious ERFNIN motif are conserved in candidate CysPs 35

3.3 Generation of candidate *CysP* RNAi constructs 39

3.4 RNAi silencing of candidate *CysP* genes and generation of stable *CysP* silenced tobacco lines 41

3.5 Silencing of *CysP6* increases IL-10 accumulation in independent T_0 tobacco lines44

3.6 IL-10 accumulation increases with plant maturity in tobacco 48

3.7 Transient silencing of *CysP6* increases IL-10 accumulation 48

3.8 Transient overexpression of *CysP6* does not reduce IL-10 accumulation 51

3.9 CysP6 localizes in the ER 53

3.10 Generation of *CysP* silenced T_1 tobacco lines 55

3.11 IL-10 accumulation in *CysP* silenced T_1 tobacco lines 58

3.12 *CysP6* silenced T_1 lines display reduced expression of *CysP6* 58

Chapter 4: Discussion 61

4.1 Identification of CysPs involved in IL-10 proteolysis 61

4.2 ERFNIN motif and catalytic triad is conserved in papain family tobacco CysPs... 64

4.3 *CysP6* is involved in reducing IL-10 accumulation in tobacco 64

4.4 IL-10 accumulation in tobacco increases with plant maturity 65

4.5 Transient knockdown of *CysP6* in tobacco leaf 66

4.6 CysP6 localizes to the ER 67

4.7 *CysP6* silenced lines show reduced *CysP6* expression in T_1 generation 68

Chapter 5: Perspectives and Future work.....70

References.....72

CURRICULUM VITAE.....85

List of Tables

Table 2.1 List of primers used for silencing and overxpression of *CysP*s…….....……...22

Table 3.1 List of putative *CysP* genes in tobacco…….....……33

Table 3.2 Comparison of published *CysP* sequences with tentative contig sequences from tobacco EST database.....…….....34

Table 3.3 Detail domain information of candidate CysPs…….....…37

Table 3.4 Number of independent *CysP* silenced T_0 tobacco lines....…….....43

Table 3.5 Numbers of *CysP* silenced T_1 tobacco lines…….....57

List of Figures

Figure 1.1 Estimated cost of purified IgA antibody produced by different expression systems……...…….4

Figure 1.2 Structure and substrate binding sites of papain……15

Figure 2.1 Gateway silencing and overexpression vectors for use in plants…….……23

Figure 2.2 Cloning and screening strategy for *CysP* gene silencing…….……24

Figure 3.1 Schematic representation of different domains in papain family CysPs…..…36

Figure 3.2 Multiple sequence alignment of papain family CysPs…….38

Figure 3.3 Strategy for *CysP* silencing……..……40

Figure 3.4 Generation of stable *CysP* silenced tobacco lines……42

Figure 3.5 Detection of IL-10 by dsELISA...……....45

Figure 3.6 IL-10 accumulation in *CysP6* silenced T_0 tobacco lines…….……....46

Figure 3.7 IL-10 accumulation in *CysP* silenced T0 tobacco lines……...47

Figure 3.8 IL-10 accumulation in tobacco plants of different ages……..…….....49

Figure 3.9 Transient silencing of *CysP6* in G7 tobacco ……...……50

Figure 3.10 Transient overexpression of *CysPs* in G7 tobacco...……..52

Figure 3.11 ER localization of CysP6……54

Figure 3.12 Generation of CysP silenced T1 tobacco lines……...……56

Figure 3.13 IL-10 accumulation in CysP silenced T1 tobacco lines…….……59

Figure 3.14 *CysP6* expression and IL-10 accumulation in *CysP6* silenced T_1 lines.....60

List of Abbreviations

ABTS	2, 2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)
BSA	bovine serum albumin
BLAST	basic local alignment search tool
ccdb	cytotoxic protein CcdB
cDNA	complementary DNA
CFP	cyan fluroscent protein
CTAB	cetyl-trimethylammonium bromide
CysP	cysteine protease
DFCI	Dana-Farber cancer institute
DNA	deoxyribonucleic acid
DPI	days post-infiltration
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
eGFP	enhanced green fluroscent protein
ER	endoplasmic reticulum
EST	expressed sequence tag
F/R	forward/reverse
GRN	granulin like repeats

h	hours
ha	hectares
hGM-CSF	human granulocyte macrophage-colony stimulating factor
HRP	horse radish peroxidase
Ig	immunoglobulin
IL	interleukin
In	intron
NAA	naphthalene acetic acid
OD	optical density
OTU	otubain
OE	overexpressed
p35S	cauliflower mosaic virus 35S promoter sequence
PBS	phosphate buffer saline
PMSF	phenylmethanesulfonylfluoride
PTM	post translational modification
PVPP	polyvinylpolypyrrolidone
RNA	ribonucleic acid
RNAi	RNA interference
RISC	RNA induced silencing complex

RT-PCR	reverse transcription-polymerase chain reaction
Si	silenced
T35S	cauliflower mosaic virus 35S terminator sequence
TAE	tris-acetate-EDTA
TBE	tris-borate-EDTA
TC	tentative contig
TGI	tobacco genome index
TSP	total soluble protein
UV	ultraviolet
UTR	untranslated region
Vec	vector
v/v	volume/volume
w/v	weight/volume
YFP	yellow fluroscent protein

*SI units are not listed

Chapter 1: Introduction

1.1 Recombinant protein production systems

Recombinant proteins are protein molecules such as vaccines, antibodies, enzymes, cytokines and hormones which are produced in a system other than the native biological hosts. Such proteins are widely used in prophylaxis or treatment of diseases, agricultural crop improvement and industrial applications. Recombinant therapeutics and neutraceutical proteins have a strong market demand and their production account for multi-billion dollar market volume in recent days (Durocher and Butler, 2009). Availability of the engineered bacterial strains and/or transgenic plants and animals has provided rapid advancement in large scale production of recombinant proteins. This advancement has resulted in increased accessibility to products that were not available or less safe to produce 20 years back. The impact of recombinant protein has been further extended by the development of industrial enzymes, diagnostic kits and biological control agents and continues to grow as applications in the field of detergent production and food processing have met with notable successes (Palomares et al., 2004).

Recombinant proteins are produced in a variety of production systems ranging from prokaryotic systems such as bacteria to eukaryotic systems such as yeast, filamentous fungi, plants and animals. *Escherichia coli* and *Bacillus subtillis* are two widely used bacterial hosts for producing therapeutic proteins and industrial enzymes. A prominent therapeutic protein, human insulin, and a food processing enzyme, α-amylase, are produced in bacterial host system (Ma et al., 2003). Though well utilized for large scale production of many pharmaceutical proteins, bacterial systems are often disadvantageous because of the inability to synthesize and modify complex eukaryotic proteins. Proteins that require post-translational modifications (PTM) like glycosylation, carboxylation and phosphorylation cannot be processed in bacteria. PTM profoundly affects the biological activity of recombinant proteins and other properties relevant to their therapeutic application. Nearly 70% of the recombinant proteins that are in clinical trials at present, require glycosylation in their native form (Durocher and Butler, 2009).

In addition, recombinant proteins can be toxic to bacteria, form protein inclusion bodies during culture or get degraded by extracellular bacterial proteases.

Animal cell cultures and yeasts are currently widely employed for the production of many complex proteins because of their ability to perform PTM. However, recombinant proteins produced from animal cell cultures are expensive and carry an inherent risk from contaminants of animal origin (Lienard et al., 2007). These contaminants include blood-borne pathogens such as animal viruses or prions and thus, the produced proteins require cautious downstream processing and quality control. Transgenic animals are also commonly used to produce recombinant proteins, which carry similar risks as cell culture systems. The risks from transgenic animal systems are generally due to recombinant proteins which originate from the sources such as milk, blood and egg white (Wang et al., 2013). Yeasts, on the other hand, have a potential of synthesizing large amounts of mature recombinant proteins. Being an eukaryote, yeasts are expected to have some advantages over bacteria. However, in many cases yeasts were shown to inappropriately fold and secrete monoclonal antibodies (Houdebine, 2009) and some yeasts did not glycosylate proteins or added sugars that were not found in native proteins (Hamilton et al., 2003).

Transgenic plants are one of the most economical systems for the production of recombinant proteins of industrial and therapeutic use (Ma et al., 2003). In addition, plants have the ability to perform the majority of PTM's required for correct folding and assembly of eukaryotic multimeric proteins. The process of transgenesis is relatively easy in plants and foreign proteins can be stored in leaves, root and seeds. With the abundant biomass, the amount of protein that can be produced in plant is virtually unlimited. Moreover, agriculture offers great flexibility for scaling up. Combined with these major qualities and with genetic tools available for continuous improvement, plants offer a unique and safer platform than any other existing host systems for recombinant protein production.

1.2 Recombinant protein expression in plants

Plants are attractive biofactories for recombinant protein expression. They offer several advantages such as reduction in cost of production and increased scalability over microbial and mammalian cell culture systems. The low cost of large scale production is due to significant reduction in capital investment and running cost, as plants do not require huge fermenters as cell culture based systems or highly skilled personnel to operate them. It is estimated that recombinant proteins can be produced in plants at 2-10% of the cost required to produce them in microbial systems and at 0.1% of the cost required to produce in animal cell cultures, although the cost of production may differ depending on the specific protein (Kusnadi et al., 1997). A comparative cost evaluation of producing monoclonal antibody Immunoglobulin (Ig) A in different host systems is shown in Figure 1.1. It was found that, producing purified IgA using either green plant biomass or grain significantly undercuts the cost of production using either animal cell culture or transgenic animals (Daniell et al., 2001).

Recombinant proteins produced in plants are correctly assembled, functionally matured and glycosylated. Plants are capable of adding high-mannose type N-glycans, which are typical of mammalian glycoproteins (Cabanes-Macheteau et al., 1999; Bakker et al., 2001). Some concerns have been raised regarding PTM occurring in plants, especially immunogenicity caused by plant-specific glycosylations (Cabanes-Macheteau et al., 1999). However, remarkable achievements have been made in modifying plant-specific glycans and removing enzymatic pathways which generate immunogenic residues on glycoproteins (Paccalet et al., 2007; Decker and Reski, 2008; Sourrouille et al., 2008; Strasser et al., 2008). Moreover, despite differences in N-glycan structures, antibodies produced in plants have shown similar stability and antigen binding capacity as their homologs produced in animal cells (Khoudi et al., 1999; Bouquin et al., 2002). Besides low cost of production and efficient PTM's of recombinant proteins, other potential advantages offered by plants include reduced risk of contamination from animal-borne pathogens and high level production of proteins using improved molecular techniques that are easily adapted by plants (Kaldis et al., 2013).

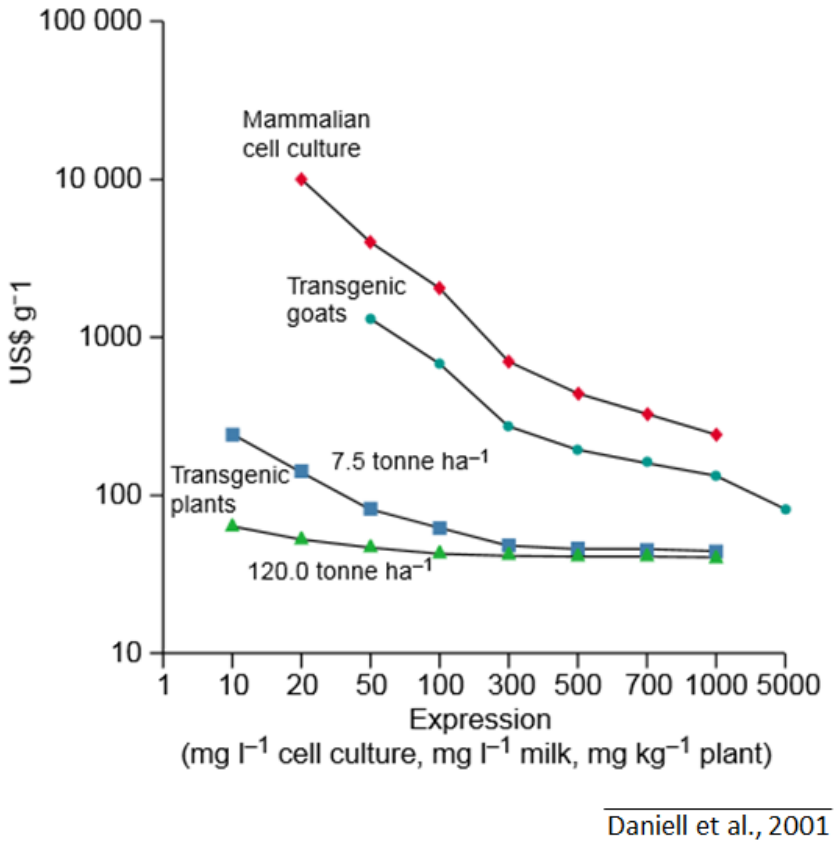


Figure 1.1: Estimated cost of purified IgA antibody produced by different expression systems. The cost is plotted as a function of starting product concentration before recovery and purification. Costs for mammalian cell culture were derived from industry cost and purification facilities. Costs for transgenic goats were obtained from Genezyme Transgenics (Farmingham, MA, USA). Costs for plants compare green biomass (120.0 tonne ha^{-1}) and seed production (7.5 tonne ha^{-1}). Cost differences reflect mainly the different costs of production: the cost of purification and the extent of product loss during downstream processing were assumed to be the same for all systems. Adapted from Daniell et al., (2001).

Different plants have been used for recombinant protein production over the past two decades. This includes a wide range of food and non-food crops like tobacco, alfalfa, cereals (rice and maize), legumes (soybean) and fruits and vegetables (carrot, tomato, potato) (Ma et al., 2003). Tobacco (*Nicotiana tabacum*) is the most widely used plant host and has an established history in recombinant protein production. Being an easily available, non-food crop, tobacco offers a competitive edge over any other plant host in terms of large scale production. The primary advantage of tobacco is the presence of well-established genetic tools for transfer and expression of foreign genes. Furthermore, tobacco provides a high yield of protein per green leaf biomass, can be harvested 3-4 times in a year and has a prolific seed production ability (Ma et al., 2003). Since tobacco is not a food or feed crop, the risk of food chain contamination via transgene escape is also limited. The first pharmaceutical protein produced in plant, human growth hormone (hGH), was expressed in the tobacco leaves (Barta et al., 1986). To increase the stability of *hGH* transcript and achieve an efficient translation into protein, the expression of *hGH* was driven by the promoter of *NOPALINE SYNTHASE* gene and polyadenylation sites were added to the *hGH* gene sequence. Since then, many recombinant proteins like monoclonal antibodies (IgG, IgA and IgM) and recombinant subunit vaccines (hepatitis B, *E. coli* enterotoxin and Cholera toxin B) have been successfully produced in tobacco and are in different stages of the clinical trials (Ma et al., 2003; Lienard et al., 2007). Other biopharmaceutical proteins produced in tobacco include plasma proteins, animal hormones and cytokines, which are used to improve blood and immune functions (Lienard et al., 2007). Some of the examples of such biopharmaceuticals are human albumin, erythropoietin and interleukins. As tobacco has been criticized for its high toxic alkaloid level, low alkaloid tobacco cultivar (cv. 81V9) has been established and can be efficiently used for production of recombinant proteins (Menassa et al., 2001).

1.3 Stable and transient protein expression systems

There are two main methods used for recombinant protein expression in plants: stable and transient transformations. Stable transformation of plants allows heritable integration of a transgene into the plant genetic material and consequently the acquired character is passed on to next generation. With transient transformation, the expression of

transgene is short-lived and the acquired trait is not transmissible to the next generation. Both strategies have been used widely for research and commercial production of recombinant proteins (Lienard et al., 2007).

Stable protein expression can be achieved by transformation of nuclear and plastid (chloroplast) genomes. This strategy has been used for a great variety of plant species such as rice (He et al., 2011), potato (Scheller et al., 2001), tomato (Pan et al., 2008) and tobacco (Staub et al., 2000). Stable nuclear transformants can be generated by transformation using the gram-negative bacterium, *Agrobacterium tumefaciens*. Using well-established transformation protocols, exogenous DNA can be inserted into a Tumor inducing (Ti) plasmid of *Agrobacterium* and delivered into the plant tissues for production of the desired protein (McCormick et al., 1986; Lienard et al., 2007). Alternatively, metal particles coated with exogenous naked DNA can also be shot into the nuclear genome using a gene gun (Christou, 1995). Nuclear transformation allows for regeneration of transgenic lines, followed by extraction and purification of proteins from the transgenic tissues in multiple generations.

Stable transformation of the chloroplast genome can be accomplished using the gene gun method. The advantages of chloroplast transformation include: a higher transgene copy number due to the presence of high number of chloroplast in photosynthetic cells and maternal inheritance of the chloroplast DNA which reduces risk of transgene escape via pollen. Furthermore, there is limited toxicity to the plant cell as proteins are produced and localized in the chloroplasts. In recent years, chloroplast transformation has been reported to result in a remarkable yield of heterologous proteins (Bock and Warzecha, 2010). However, chloroplasts are unable to perform typical eukaryotic PTM, such as protein glycosylation. With a limited protein maturation capacity, a chloroplast looks particularly well adapted for production of simple eukaryotic proteins (Jobling et al., 2003). Chloroplast transformation has been successfully attempted in very few plant species, primarily tobacco.

Transient expression of proteins in plants has been successfully performed using viral and non-viral expression vectors. Viral vectors are modified viral genomes which serve as a vehicle to carry foreign genes and the genes are expressed in plants as the by-product of viral replication. Recombinant proteins that have been produced using classical plant virus expression systems utilize gene fusion with different viral coat proteins (CP), or replacement of *CP* genes with the gene of interest, and are transferred by viral infection to the plant (Kohl et al., 2006; Avesani et al., 2007). The second generation viral vectors utilize *Agrobacterium* for transferring a proviral replicon harbouring the gene of interest into the plant cells. The proviral replicon is constructed in the transfer (T) - DNA region of the Ti plasmid, which is then transferred into plant host such as *Nicotiana benthamiana* (Santi et al., 2006; Werner et al., 2011). The protein is transiently expressed and harvested in approximately 2-14 days post-infiltration. Though considered ideal for expression of vaccine antigens, viral vector systems have biosafety issues regarding handling and use of virus originated material, and a cost disadvantage in comparison to the stable protein expression.

Non-viral expression vectors carrying a foreign gene of interest have also been used for transient expression of proteins in plants. It is commonly done by vacuum infiltration of recombinant *Agrobacterium* culture into the host plant (Vaquero et al., 1999). *N. benthamiana* is the most widely used host for this purpose, as its leaves are well suited for infiltration of the bacterial culture and results in transient transformation of many cells. A high level of protein can be transiently produced by large scale infiltration of *N. benthamiana* and commercially it has been brought to use by different companies (D'Aoust et al., 2008). However, using transient expression technique, higher yield of proteins are only achieved for a short period of time. In addition, low biomass of the plant hosts and lack of consistency between different batches of proteins are the major concerns in commercial production of recombinant proteins using transient expression systems.

1.4 Challenges of recombinant protein production in plants

Recognized as an efficient and most economical system, plant bioreactors have emerged as a reliable source for recombinant protein production. However, plant bioreactors have two major concerns that limit the production of recombinant proteins. First, lower accumulation levels of the proteins of interest, and second, lack of the efficient protein purification methods. Recovery of proteins from plant biomass requires multistep purification using chromatography, adsorption, precipitation and filtration, and thus, is a delicate and costly process. In some cost analysis cases, it has been found that purification steps account for more of the costs than production in a plant host (Evangelista et al., 1998). However, pricing of the recombinant proteins is mostly dependent on total yield provided by the production systems and final accumulation of the proteins in their bioactive forms. Lower accumulation levels are generally observed, which is mainly due to the lower level of transgene expression and protein instability found during the expression and recovery of recombinant proteins.

Transgene expression in plants is affected by various factors and involves regulation at different levels such as transcription, translation and post-translation. As a result, nascent and steady state transcript levels, translational turnovers, and the final levels of mature proteins do not co-relate to each other during recombinant protein production (Chen et al., 2013). Different strategies have been employed to improve transgene expressions in plants such as, using tissue specific promoters (Christensen and Quail, 1996), suppressing a commonly occurring transgene silencing (Lombardi et al., 2009), intron induced splicing of transcript messenger (m)RNA (Vain et al., 1996) and modifying the 3’ and 5’ untranslated regions (UTR) of the gene for producing higher transcript levels (Strizhov et al., 1996). Using these approaches, significant improvements have been made in obtaining a higher recombinant protein yield, however, in many cases the levels of recombinant protein accumulation still remain low at <1% of the total soluble protein (TSP) (Doran, 2006). One of the major reasons for low yield is the degradation of foreign protein by plant proteases (Faye et al., 2005; Doran, 2006). Therefore, the low levels of protein accumulations result from a process of synthesis and degradation occurring during protein expression in different systems.

Protein degradations in the plant systems are known to occur by both intracellular and extracellular ways. Inside the cell, degradations can occur after synthesis, assembly and in some cases during PTM in the Endoplasmic Reticulum (ER) and Golgi apparatus. Outside the cell, proteins are degraded by extracellular proteases in the apoplastic space or in the culture medium when they are synthesized as secreted proteins (Hehle et al., 2011). Protein fragments have been found to form in many plant hosts expressing recombinant proteins, such as in tobacco (De Neve et al., 1993; van Engelen et al., 1994; Schiermeyer et al., 2005), *Arabidopsis* (De Neve et al., 1993), corn (Russell, 1999; Russell et al., 2005), potato (Outchkourov et al., 2003) and alfalfa (Khoudi et al., 1999). The evidences for foreign protein degradation in these plants were obtained from Western blots using proteins present in the plant extracts. Cleavage products were also identified for proteins such as human growth hormone (Russell et al., 2005), plasminogen activator (Schiermeyer et al., 2005) and monoclonal antibody IgG1 (Sharp and Doran, 2001) using peptide sequencing methods.

Since the yield of recombinant protein is severely affected by proteolytic degradation, there is a large interest in identifying and addressing the cause of proteolysis. Analyzing proteolysis events and characterizing protein fragments have revealed information regarding cleavage sites and mechanism of some recombinant protein degradation processes (Outchkourov et al., 2003). In one of the early studies, degradation of IgG1 monoclonal antibody in tobacco hairy root and suspended cell cultures was found to be caused by extracellular proteases (Sharp and Doran, 2001). Analysis of protein fragments by western blot suggested that IgG1 was degraded not during extraction but during synthesis and protein maturation along the secretory pathway of plant cell. Later, identification of a specific class of plant protease responsible for the proteolysis was successfully performed in many systems producing recombinant proteins. It was found that *in planta* degradation of MGR48 monoclonal antibody, an IgG1 type immunoglobulin from mouse, is catalyzed by acidic proteases in the apoplastic space (Stevens et al., 2000). The degradation was more prominent in older leaves than in younger leaves while a comparison was done by in-vitro incubation of recombinant IgG1 with crude leaf extracts. This study indicated that the proteases associated with aging or

senescence, which also include the members of enzyme class cysteine proteases, are involved in IgG1 degradation. Cysteine proteases were also identified as the enzymes responsible for the degradation of recombinant sea anemone protein equistatin, during its expression in potato tubers (Outchkourov et al., 2003). More recently a rice cysteine protease, RETARDED PALEA1 (REP1), was also specifically found to degrade a human cytokine (human granulocyte macrophage-colony stimulating factor) while the protein was produced by secretion in rice cell suspension culture (Kim et al., 2008).

1.5 Strategies in improving protein yield

Although degradation of foreign protein in plants depend on different factors, it has been found that proteolysis occurs to a greater extent in tobacco than any other plant species such as *Arabidopsis* (De Neve et al., 1993) and alfalfa (Khoudi et al., 1999). Since tobacco is used to produce many recombinant proteins, it is important to minimize the protein degradation to achieve high levels of functional proteins. In whole tobacco plant, degradation of recombinant protein has been correlated to leaf maturation and senescence process (Stevens et al., 2000). Different strategies have been employed to influence the cell protease activity as a tool to improve the final protein yields. Targeting recombinant proteins to accumulate in the ER is one of the strategies used to minimize the exposure of proteins to cytosolic and apoplastic proteases (Menassa et al., 2001). Retention of proteins in the ER is helpful in their proper folding and maturation as ER harbours molecular chaperones and stabilizing agents. Strategies that directly affect protease activity have also been applied such as expressing companion protease inhibitior, which was successful in increasing the recombinant protein yield by 35-40% (Benchabane et al., 2009; Goulet et al., 2010). An efficient way to minimize protease activity was also suggested which involves identifying the responsible proteases and downregulating their expression that would allow for a sufficient accumulation of the target recombinant proteins (Kim et al., 2008). Alternatively, protease cleavage sites can also be identified and amino acids in the recombinant protein can be altered to minimize degradation. However, this approach first requires the identification of specific protease and cleavage sites in recombinant proteins, before the engineering of protease cleavage sites can be done.

In addition to reducing protein degradation, some other strategies have also been employed which are targeted to improve the recovery of recombinant proteins in plants. These strategies include: localization of proteins to a separable component such as to the embryo or oil bodies for an easier extraction, induction of protein bodies or secretion into cell suspension media for an easier purification, and production as fusion proteins to increasing the stability of the produced proteins (Gutierrez et al., 2013). Despite improvements in protein recovery such strategies often encounter problems; some of which have been reported during production of biopharmaceuticals as protein bodies or fusion proteins (Menassa et al., 2007; Kaldis et al., 2013). Biological inactivity of the fusion proteins were through animal feeding and cell lines functional assays.

1.6 Plant proteases

Plant proteases are a special class of enzymes involved in the degradation of native or foreign proteins. Plants contain four different class of proteases which are classified according to their catalytic site residues (van der Hoorn, 2008). These include: serine proteases, aspartic acid proteases, metalloproteases and cysteine proteases (CysPs). Serine proteases use the active site Ser as a nucleophile during catalytic reactions. With over 200 members among 14 families, serine proteases are the largest class of proteolytic enzymes in plants (van der Hoorn, 2008). The roles of serine proteases have been functionally characterized in proteolytic activation of enzymes (ABNORMAL LEAF SHAPE 1, ALE1 for the activation of the enzyme involved in embryogenesis in *Arabidopsis*), regulation of biosynthetic reactions (SINAPOYLGLUCOSE ACCUMULATOR 1 and 2, SNG1 and SNG2 in regulation of ester biosynthesis in *Arabidopsis*), as well as in tissue specific roles (STOMATAL DENSITY AND DISTRIBUTION 1, SDD1 for regulation of positional stomata development within leaf epidermis in *Arabidopsis*) (Lehfeldt et al., 2000; Tanaka et al., 2001; Von Groll et al., 2002).

Metalloproteases contain catalytic metal ions that activate water molecules for the proteolytic reactions. Over 100 plant metalloproteases have been identified which belong to the 19 different families (van der Hoorn, 2008). Metalloproteases are involved in

different aspects of plant growth and responses such as nodulation, plastid differentiation, regulation of shoot and root meristem size and thermo-tolerance (van der Hoorn, 2008). Aspartic acid proteases are also the other class of proteases that support water molecules for proteolysis. Two Asp residues are involved in binding to the water molecule which forms the active site in the enzyme. Aspartic acid proteases are the second most abundant plant proteases, with members divided into 3 large families (van der Hoorn, 2008). These proteases are similar to pepsin, their animal counterpart, and some of their members are involved in gametophyte survival and disease resistance signalling (e.g. *Arabidopsis* Promotion of cell survival-PCS1 and Constitutive disease resistance 1-CDR1) (Xia et al., 2004; Ge et al., 2005).

1.7 CysPs

CysPs are ubiquitous proteins found in organisms ranging from bacteria, fungi, viruses to plants and animals. CysPs involve a cysteine sulfhydryl group in their catalytic reactions and their optimum catalytic activity is seen in a pH range of 4-6.5 (Grzonka et al., 2001). CysPs are involved in responses to many environmental and developmental cues and represent one of the thoroughly investigated protease class in plants. Till now, 101 families of CysPs have been discovered which are listed in the MEROPS database and almost half of them are present in viruses (http://merops.sanger.ac.uk/cgi-in/family_index?type=P#C). Many of these enzymes are found in bacteria (family C10, streptopain in *Streptococcus pyogenes*; family C25, gingipain R in *Polyhyromonas gingivalis*) and fungi (family C1B, bleomycin hydrolase in yeast). In mammals, two main groups of CysPs, calcium-dependent cytosolic calpains (family C2), and lysosomal cathepsins (family C1A like, subfamily cathepsin B, C, F, H, K, L, M, N, S, T, V and W) have been found and are known to maintain the homeostasis of the cell's metabolism (Grzonka et al., 2001). Another representative CysP group in mammals are caspases (family C14), which play an important role in programmed cell death and hypersensitive responses controlling the so-called apoptosis cascades (Woltering, 2004).

In plants, most of the CysPs discovered till date belong to the papain (C1A) and legumain (C13) families, but caspases and calpains have also been found in plants. In

addition, otubains (family C12) have also been found in plants whose roles are similar to their animal counterpart (Vierstra, 2003). Among all these categories of CysPs, the papain family is the most widely characterized group and comprises of all peptidases which are structurally related to the protein, papain. The proteins in this family are recognized by the presence of two structural domains (Grzonka et al., 2001). As shown in Figure 1.2A, the catalytic pocket of papain is present in the cleft which is formed between the two domains. Figure 1.2B shows that, the substrate pocket of papain consists of at least 7 subsites which bind with 7 amino acid residues of the substrate for catalysis (Schechter and Berger, 2012). However, previous studies have concluded that only 5 subsites of the enzyme (S_1,S_2,S_3, S_1' and S_2') are necessary for substrate binding (Turk et al., 1998). The enzymatic activity of papain is dependent on the presence of a Cys, His and Asn residue, which form a catalytic triad of the enzyme. Asn in the catalytic triad is present as the conserved amino acid residue across all papain family CysPs and is important for the proper orientation of the His side chain (Beers et al., 2004). It has been found that, the members of the papain CysPs preferentially cleave peptide bonds with Arg in P1 position or Phe at the P2 position, in a pH interval of 3.5-8 (Grudkowska and Zagdanska, 2004).

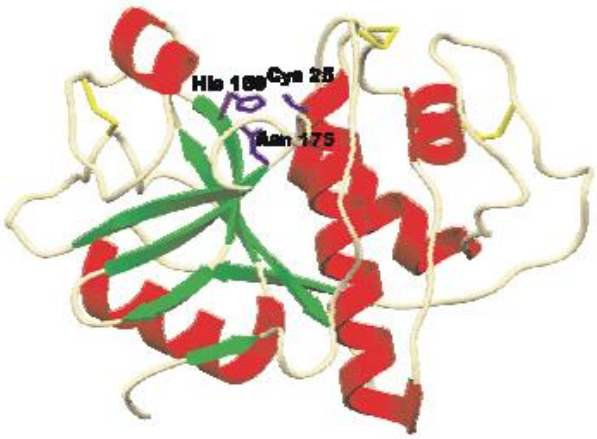
The otubain (OTU) is the most recently characterized family of CysPs in plants. Little is known regarding the functional roles of plant otubains but they are thought to be involved in many ubiquitin-related cellular processes (Aquea et al., 2008). An otubain CysP, OTUBAIN-LIKE HISTONE DEUBIQUITINASE-1 in *Arabidopsis*, indicates its role in transcriptional repression via histone deubiquitination (Krichevsky et al., 2011). Otubain-like CysP was also discovered in pine and was found to be expressed during early somatic embryogenesis of pine trees, but its biological function remains unknown (Aquea et al., 2008).

The main physiological role of CysPs is metabolic degradation of peptides and proteins. Mammalian CysPs have been associated with onset and development of many diseases that involve abnormal protein turnover (Twyman et al., 2003). In plants, CysPs play key roles in environmental stress response, nutrient remobilization and cellular housekeeping, which account for nearly 30% of the proteolytic activity in mature tissues

(Ueda et al., 2000). In *Arabidopsis* alone the genome codes for 32 papain-type CysPs, which are classified into eight main groups based on their function and sequence similarity to other known CysPs (Grudkowska and Zagdanska, 2004). These CysP groups include, senescence- and stress induced- aleurain, mammalian cathepsin B-like, pineapple bromelain-like, telo, KDEL and kiwi actinidain-like CysPs. The *Arabidopsis* CysPs are known to carry out multifarious roles in different tissues and at different stages of plant growth (Lee et al., 2004; Zhang et al., 2009). A highly senescence specific SENESCENSE-ASSOCIATED GENES-12 was identified in *Arabidopsis*, which was found to have an increasing expression during leaf senescence (Gan and Amasino, 1995). Three different KDEL-type CysP, C-TERMINALLY ENCODED PEPTIDE (CEP)-1, 2 and 3 were also identified in *Arabidopsis*, and were found to be involved in the different stages of plant growth (Gietl and Schmid, 2001). It was suggested that CEP2 could be involved in developmental programmed cell death during the tracheary element differentiation and cell elongation processes in the tissues. CEP1 and CEP3 were expressed respectively in unpollinated ovules and mature siliques and represented increase in CysP activity during the different developmental stages of plant. Some CysPs in *Arabidopsis* have also been shown to be induced during biotic stresses. One of such example is RESPONSIVE TO DEHYDRATION (RD) 19, which was induced during infection with pathogenic bacteria *Ralstona solanacearum*, and was found to be involved in disease resistance against the pathogen (Bernoux et al., 2008).

Though several CysPs have been characterized in *Arabidopsis*, not much is known regarding the role of CysPs in tobacco. NICOTIANA TABACUM CYSTEINE PROTEASE (NtCP) 23 is one of the earlier identified CysPs in tobacco and is involved in amino acid remobilization in senescing leaves (Ueda et al., 2000). Few other tobacco CysPs have also been characterized which are implicated in stress response, protein degradation during programmed cell death (Zhao et al., 2013) and pollen grain development (Zhang et al., 2009).

A



B

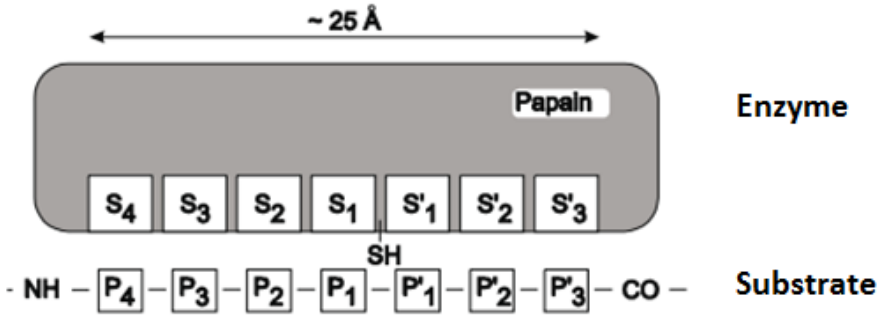


Figure 1.2 Structure and substrate binding sites of papain A) Papain has two structural domains with five alpha helices surrounding seven beta barrel sheets. The active site is present the cleft between the two domains and consists of the catalytic triad of Cys25, His159 and Asn175 residues (Grzonka et al., 2001). B) Substrate subsites of papain. The substrate pocket is approximately 25 Å in length. The Letters S_n and S'_n represent subsites, where n=1, 2, 3, 4 and SH represents the sulfhydryl group. P_n represents individual amino acid residues of the substrate that binds with the respective subsites in papain (Schechter and Berger, 2012)

1.8 Interleukin-10

Interleukin (IL)-10 is an immune-regulatory cytokine produced by immune cells such as T cells, natural killer cells and macrophages. The role of IL-10 has been investigated in the limitation and termination of immune responses during biological therapy of different malignant, autoimmune and chronic inflammatory diseases (Asadullah et al., 2003). IL-10 has been used as a model recombinant protein to study the factors that control the synthesis and accumulation of recombinant proteins in plants (Chen et al., 2013). For large scale production of IL-10, the low alkaloid variety of tobacco-cv. 81V9 has been previously used (Menassa et al., 2001). Recombinant IL-10 was produced using stable expression in a higher amount than the transient expression and was found to be biologically active during murine assays (Menassa et al., 2007). To achieve higher levels of IL-10, subcellular targeting was utilized to accumulate IL-10 in the ER by fusing ER retention signal KDEL to its C-terminal end. However, the accumulation of IL-10 during initial stable expression is reported to be less than 1% of the total soluble protein, a level which is not sufficient for the viable commercial production of proteins. Some strategies have been applied to increase the yield of IL-10, mostly through the tobacco cell-culture based system (Kaldis et al., 2013). However, as stable protein expression can offers several advantages (as mentioned in section 1.3), improving IL-10 accumulation in the stable transgenic tobacco line can serve as a better source for production and supply of recombinant IL-10.

1.9 Hypothesis

Though tobacco allows high level production of recombinant proteins, accumulation of IL-10 was lower than the critical level of 1% of total soluble protein in stable tobacco lines. Since CysPs are known to degrade heterologously expressed proteins in plants I hypothesize that tobacco CysP (s) affects IL-10 accumulation. Altering specific *CysP* gene product(s) can minimize proteolytic degradation of IL-10 and increase its accumulation in tobacco leaves.

1.10 Objectives

The whole genome sequence of tobacco is not available yet and only a few tobacco *CysP* sequences have been reported. However, the availability of an expressed sequence tag (EST) database of *N. tabacum* allows for the analysis of tobacco *CysP* sequences that have been deposited by the researchers worldwide (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=tobacco). The database also provides information regarding tissue specific expressions of different *CysP*s sequences.

The objectives of the present research are:

- To identify all expressed *CysP* genes in tobacco.

  This will be addressed by an extensive search for *CysP* genes in the tobacco EST database and *in silico* analysis of predicted amino acid sequences for their domain information.

- To select candidate *CysP* genes that may have an effect on IL-10 accumulation.

  Selection of *CysPs* will be done based on criteria such as expression in leaf tissue and availability of full length sequences in the database. Expression levels of candidate *CysPs* will be altered by RNAi silencing. The effect of RNAi silencing of candidate CysPs on IL-10 accumulation will be compared in silenced lines and controls.

- To characterize IL-10 specific *CysP* gene(s).

  Transient silencing and overexpression of *CysP* will be done to study the results obtained from the stable *CysP* silenced lines. The potential site of IL-10 degradation will be studied by subcellular localization of candidate CysPs. Because IL-10 was targeted to the ER, it is imperative to check the localization of candidate *CysP* potential of degrading IL-10.

Chapter 2: Materials and Methods

2.1 Plant materials and growth conditions

Seeds of tobacco (*N. tabacum*) cv. 81V9 and the transgenic line G7 (hereby called as IL-10 control) were obtained from Dr. Rima Menassa (Agriculture and Agri-Food Canada, London). The IL-10 control tobacco line overexpresses IL-10 and accumulates in the ER (Menassa et al., 2001). The tobacco seeds were grown under sterile condition in magenta boxes at 24°C and 16 h daylight for leaf disc transformation. For sterilization, seeds were treated with 70% ethanol for 30 sec and air dried. Seeds were sown on Murashige and Skoog (MS) basal medium (*Phyto*Technology Laboratories, USA), supplemented with 1% (w/v) phytoagar (*Phyto*Technology Laboratories, USA), 4% (w/v) sucrose (Sigma-Aldrich, Germany), phytohormones and the antibiotic, kanamycin (50 µg/ mL).

Transgenic tobacco plants were grown on PRO-MIX BX MYCORRHIZAE^{TM} soil (Rivière-du-Loup, Canada) at 24°C, 60% relative humidity and 16 h daylight in the greenhouse. For *N. benthamiana*, seeds were sprinkled into wet PRO-MIX BX MYCORRHIZAE^{TM} soil (Rivière-du-Loup, Canada) and pots were covered with a transparent plastic cover and grown at 25°C/ (day/night), 50% humidity and 16 h daylight in the growth chamber. Two-week-old seedlings were transferred to new pots and watered regularly. A nutrient mixture of nitrogen, phosphorus, and potassium (20-20-20) was applied once a week for both types of tobacco plants grown in growth chamber and greenhouse.

2.2 Bacterial strains

E. coli DH5α and DB3.1 strains (Invitrogen, USA) were used for cloning purposes. For stable transformation in tobacco and transient infiltration in *N. benthamiana*, *Agrobacterium* strain GV3101 was used. All bacterial transformations were carried out by electroporation in a Gene Pulser® Cuvette (Bio-Rad Laboratories, Inc., USA) using MicroPulserTM (Bio-Rad Laboratories, Inc., USA). The electroporation was

done for 6 millisec at an output voltage of 1.79 kV and 2.18 kV for DH5α and GV3101 respectively.

`2.3 *In silico* analysis

To identify *CysP* genes present in tobacco, the DFCI tobacco gene index (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=tobacco) was searched using 'cysteine protease' as keyword. This database contains ESTs obtained from cDNA libraries constructed using different tissues or plants grown under different conditions or exposed to different stresses. Complete and partial sequence availability of all putative *CysP*s was checked using ExPASy software (http://web.expasy.org/translate/). Candidate genes were selected according to their tissue-specific expression and availability of complete sequences in the EST database. Predicted amino acid sequences of candidate CysPs were aligned using CLUSTALW and the alignment file was exported to BOXSHADE 3.21. The domain organizations of candidate CysPs were predicted using Pfam 27.0 (http://pfam.xfam.org/). The subcellular localization of putative CysPs was predicted using PSORT program (http://psort.hgc.jp/form.html).

2.4 Cloning

To clone *CysP* genes from tobacco, candidate genes were amplified by polymerase chain reaction (PCR) using leaf cDNA as template and gene specific primers which contained *attB* recombination sites. Primers and PCR conditions used for amplifying *CysP* gene fragments for silencing and full open reading frames for overexpression are shown in Table 2.1. All the primers contain th *attB1* adaptor sequence (5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC T-3' for forward primers) and the *attB2* adapter sequence (5'-GGGG AC CAC TTT GTA CAA GAA AGC TGG GT-3' for reverse primers), preceding their 5'sequences for cloning in gateway vectors.

2.4.1 Cloning for silencing constructs

To create *CysP* silencing constructs, a gene-specific region (178-345 base pairs) for each candidate *CysP* was selected and amplified using gene-specific primers (Table

2.1). The PCR products were separated on 1.5% agarose gels. A single band of *CysP* amplicon was excised from the gel and purified using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic Inc., Canada). The purified gene fragments were recombined into gateway entry vector pDONR/Zeo using the BP clonase reaction mix (Invitrogen, USA), and the BP reaction products were transformed into *E. coli* DH5α by electroporation. The transformed cells were grown in low salt Lysogeny Broth (LB) agar containing 50 µg/mL of zeocin. Colonies that contained the recombinant plasmid (pDONRZ-CysP) were screened using *CysP* gene specific primers. The positive *E. coli* colonies that contained recombinant plasmid were picked and grown overnight in 3 mL of LB at 37^0C. Plasmid DNA was extracted from overnight culture using EZ-10 Spin Column Plasmid DNA Kit (Bio Basic Inc., Canada). The presence of *CysP* sequences in the recombinant plasmids was verified by sequencing using vector-specific M13 forward and reverse primers. After sequence verification, the pDONRZ-CysP plasmids were recombined into RNAi destination vector pB7GWIWG2(II),0 or pK7GWIWG2D(II),0 using the LR clonase reaction mix (Invitrogen, USA). The destination vector pB7GWIWG2(II),0 contains spectinomycin resistance gene for selection in bacteria, and a *Bar* gene which confers herbicide BASTA resistance for the selection of transformed plant cells (Figure 2.1A). Vector pK7GWIWG2.D contains kanamycin resistance for selection in bacteria and an enhanced green fluorescent protein (eGFP) reporter gene, which was used for visual screening of transgenic plants using UV fluorescent microscope (Figure 2.1B).

The LR reaction products were transformed into *E.coli* DH5α by electroporation and plated on a LB agar plate containing 50 µg/mL spectinomycin. A schematic representation of BP and LR recombination is shown in figure 2.2A. After LR recombination, positive *E.coli* colonies were screened using colony PCR with the primer combination as shown in Figure 2.2B. The *CysP* fragments were inserted in the opposite orientation after recombination with the destination vector. The orientation of each candidate *CysP* gene in the destination vector was confirmed using a gene-specific forward primer (CysPF) and the outward chloramphenicol resistance (*Cmr*) forward and reverse primers (CmrF: 5'- CGA TTC AGG TTC ATC ATG CCG TCT-3' and CmrR:

5'-TGA GCA ACT GAC TGA AAT GCC TCC-3'). Positive *E.coli* colonies that contained recombinant pK7GWIWG2D(II),0-CysP or pB7GWIWG2(II),0-CysP plasmids were picked and grown overnight in 3 mL of LB at 37 ^{0}C. Plasmids DNA were extracted from the overnight culture using EZ-10 Spin Column Plasmid DNA Kit. For the purpose of introducing recombinant pK7GWIWG2D(II),0-CysP and pB7GWIWG2(II),0-CysP plasmids into tobacco, the plasmids were transformed into *Agrobacterium* using electroporation and grown in LB agar containing rifampicin (10 µg/mL), gentamycin (50 µg/mL) and kanamycin (50 µg/mL for pK7GWIWG2D(II),0-CysP plasmids) or spectonomycin (50 µg/mL for pB7GWIWG2(II),0-CysP plasmids).

2.4.2 Cloning of *CysP* genes for overexpression

To identify *CysP* genes that act on factor (s) which directly or indirectly influence IL-10 accumulation, full-length coding regions of *CysP* genes were amplified using gene specific primers (Table 2.1), purified using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic Inc., Canada) and recombined into pDONR/Zeo vector as described in section 2.4.1. Colonies that contain the pDONRZ-CysP plasmids were screened by PCR using gene specific primers. Positive colonies that contained the right size of full length *CysP* insert were selected and cultured overnight at 37^{0}C in LB medium containing zeocin (50 µg/mL). Plasmid preparation was done from the overnight culture using the EZ-10 Spin Column Plasmid DNA Kit (Bio Basic Inc., Canada) and sequences were verified. After sequence verification recombinant pDONRZ-CysP plasmids were recombined with overexpression vector pB7WG2D,1 using the LR Clonase reaction mix and transformed into *E.coli* DH5α (Invitrogen, USA). This destination vector contains a cauliflower mosaic virus 35S promoter, which drives constitutive expression of *CysP* genes in the plant and *Bar* gene for selection of transformed plant cells (Figure 2.1C). *E.coli* colonies containing the recombinant plasmids (pB7WG2D,1-CysP) were screened using gene specific primers by PCR. Positive colonies were cultured overnight at 37^{0}C in LB medium containing spectinomycin (50 µg/mL).

Table 2.1: List of primers used for silencing and overexpression of *CysP*s.

Primer name	Sequence (5'-3')	Primer annealing temperature (^{0}C)	Amplicon size (base pairs)
Silencing Primers			
CysP1F	CCA CCA TAA ACA ATC ATC CAA TCC GT	52	280
CysP1-5R	TAC CTC TTC CCA TAC CTG TGA GC		
CysP2F	TCC GTA CTT TTC TAG GAG CTT CAC	52	242
CysP2R	TGA CTA GTG TCA CAG TCG ACA AG		
CysP3F	CAA CTC ATA GCT CCA CTC TCA CCA	52	178
CysP3R	TCC ATG TTC GAG TAG CCA TGA CTC		
CysP4F	GGC TAG CTA GTG GTT CAT AAT GGC	60	280
CysP4R	CCA CGC CGA GTT TAT ACC CTT		
CysP5F	ATC CCT TCC CCA CAT TCT CTG A	52	249
CysP6F1	GCG ATA CTT CCT ACA ATG ACG GTT G	54	345
CysP6R1	CAT CTC ACT ACC ATA TCC ACA GCA TCA		
CysP7F1	CGT GGG ATA CTC TTC AGA TGA CTT GAC	52	204
CysP7R1	TCA GCA AAC TCA TTT AGA CCA AGC C		
CysP8F1	GAA TGG AAA CAG AAG CAG GGG AAA G	53	281
CysP8R1	GGA GGA GCA TCA CAA TAA ACT GGA G		
CysP 9F	GCA TGT GGC TGA TGC AGG A	50	255
CysP 9R	CGT GTT CTT CTA AGG GCA		
CysP10F1	GTC CGT AAA ACC CAA AAC CAC AC	52	382
CysP10R1	GCT TCT CCA TCT CTC ATA CAA CTC C		
Overexpression Primers			
CysP1-OE-F	ATG TCT CGT TTC TCA CTC CTA TTG GC	51	1080
CysP1-OE-R	GAA CGG AGC CCG AAT GCT GTT GTA		
CysP2-OE-F	CGA GGA GCC TTG ACG AGA AAG ACA A	52	947
CysP2-OE-R	CTT AGA GTT CAT CCT TAG GTG CCG T		
CysP5-OE-F	GTT CCA ATG TCT CGT TTC ACA CTC C	52	1087
CysP5-OE-R	AGG CAA CGA CAG GGT AGG AT		
CysP6-OE-F	GGC ATG GCA ACT CTT AGC TTT ACT C	52	1401
CysP6-OE-R	TCA AGA ACT GCT CTT CTT TCC TCC		

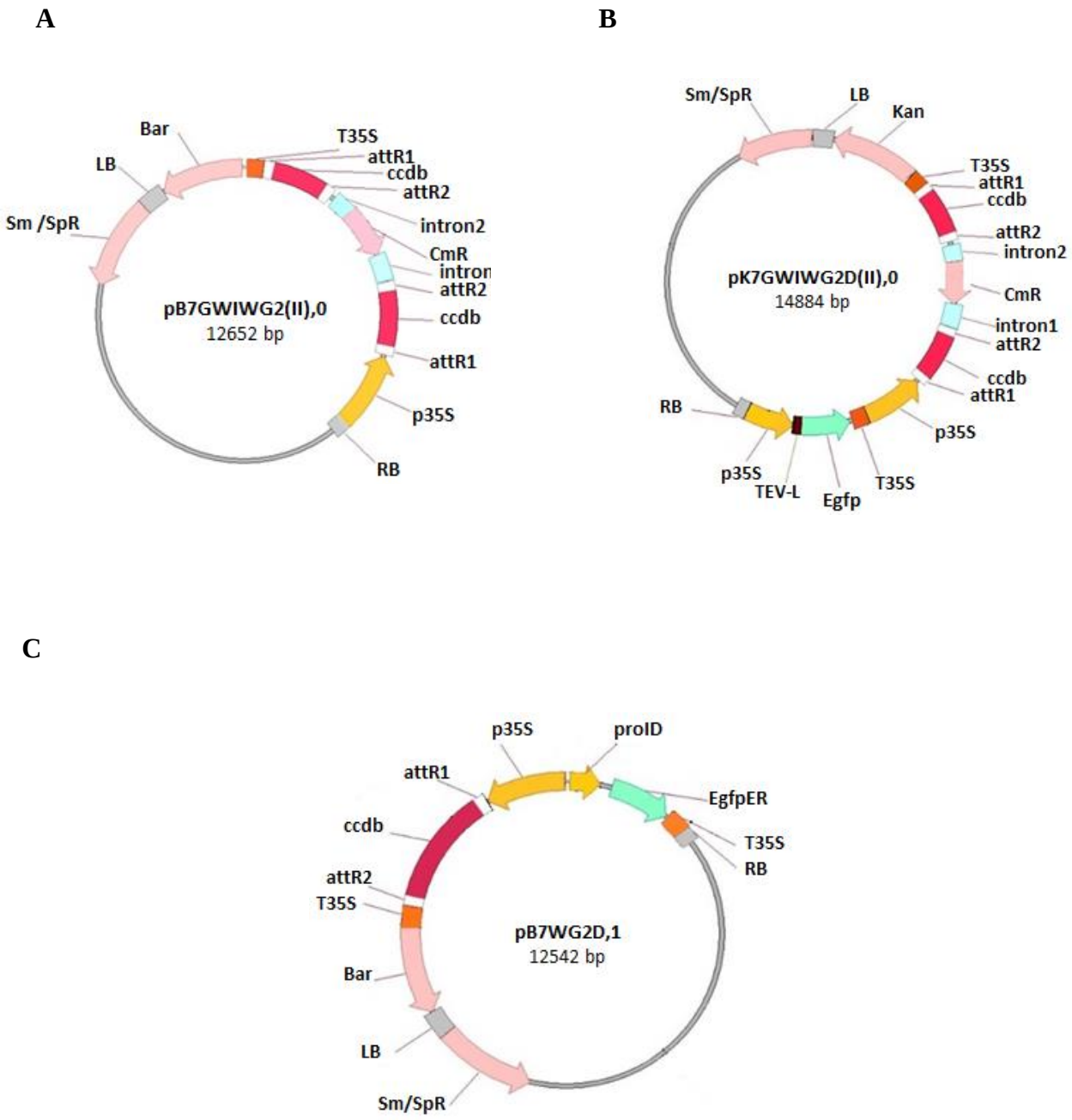
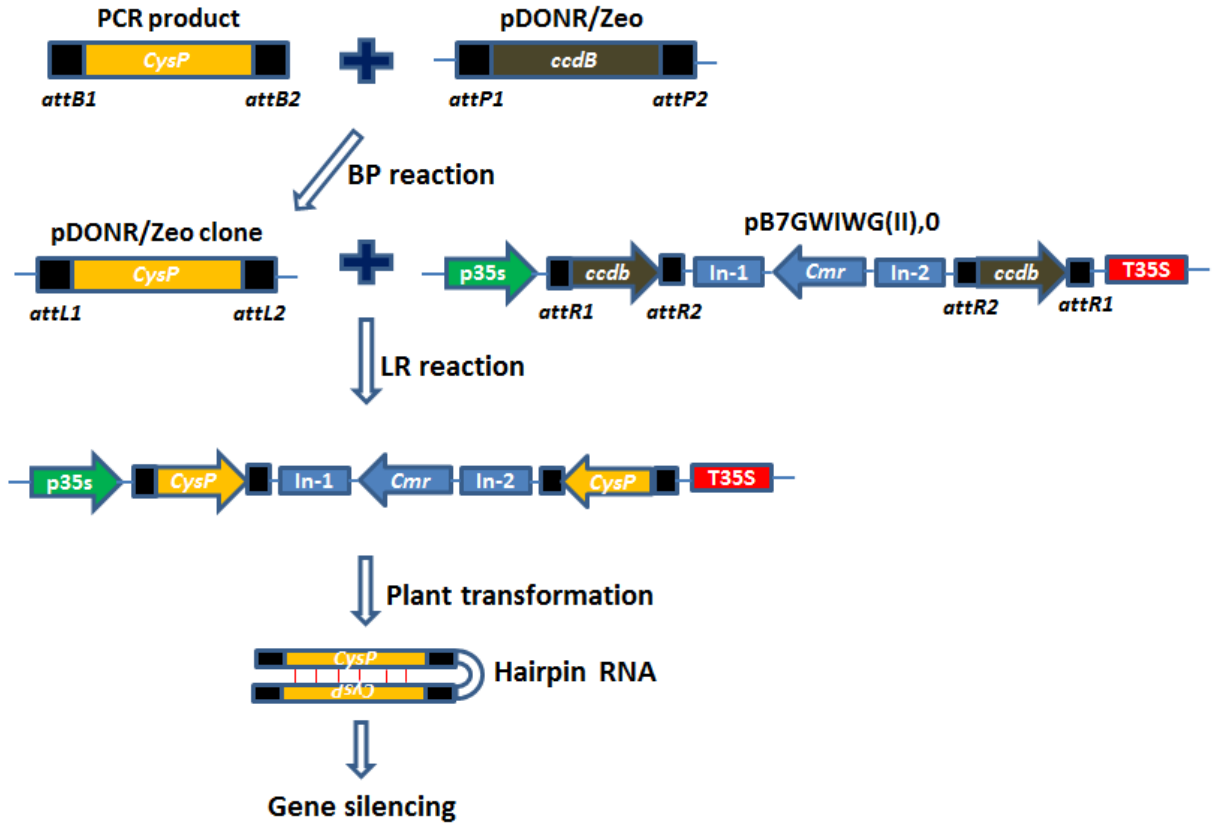


Figure 2.1: Gateway silencing and overexpression vectors for use in plants. A) RNAi silencing vector, pB7GWIWG2(II),0, B) RNAi silencing vector, pK7GWIWG2D(II),0 and C) Overexpression vector, pB7WG2D,1. All the vectors contain *attR* sites and a *ccdb* casettee for selection of successful recombination events. Vector maps adapted and modified from: http://gateway.psb.ugent.be/search

A



B

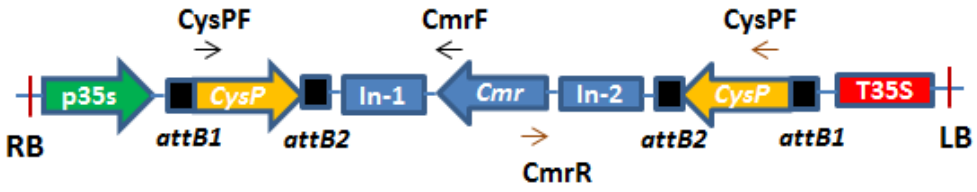


Figure 2.2: Cloning and screening strategy for *CysP* gene silencing. a) Schematic illustration of gateway cloning for silencing of tobacco *CysP* genes. *CysP* PCR products were cloned into entry vector pDONR/Zeo by BP recombination and RNAi silencing vector pB7GWIWG(II),0 or pK7GWIWG2D(II),0 by LR recombination. *CysP* gene fragments are placed in the opposite orientation after recombination event in the silencing vector. b) Screening strategy for CysP gene inserts in RNAi vector. Gene specific forward primer (CysPF) and chloramphenicol resistance forward and reverse primers (CmrF and CmrR) are used to screen the double inserts of CysPs.

Plasmid preparation was done from the overnight culture using EZ-10 Spin Column Plasmid DNA Kit and transformed into *Agrobacterium* followed by plating in LB agar containing rifampicin (10 µg/mL), gentamycin (50µg/mL) and spectinomycin (50 µg/mL). The *Agrobacteria* containing overexpression constructs were used for transient infiltration in tobacco plants.

2.4.3 Cloning for subcellular localization

To study the localization of a candidate CysP, translational fusion was created with a yellow fluorescent protein (YFP) using the gateway cloning method. Full-length *CysP6* excluding the stop codon was amplified by PCR and recombined to a gateway entry vector pDONR/Zeo by the BP reaction (Invitrogen, USA). The reaction was used to transform *E.coli* DH5α by electroporation and positive clones were screened using gene-specific primers. Recombinant pDONRZ-CP6 plasmid was extracted from an overnight culture grown in a 3 mL LB containing kanamycin (50 µg/ml) at 37^0C. The pDONRZ-CysP6 construct was sequenced using *CysP6* sequencing primers and recombined with the destination vector pEarlyGate101 to create a *CysP6-YFP* fusion using LR clonase. The LR reaction product was transformed into *E.coli* DH5α using electroporation and plated on the LB agar containing 50 µg/mL kanamycin, followed by screening using *CysP6* gene-specific primers. A positive *E.coli* colony containing recombinant plasmid pEG101-CysP6-YFP plasmid was picked and grown overnight in a 3 mL LB at 37^0C. Recombinant plasmid was extracted using EZ-10 Spin Column Plasmid DNA Kit and transformed into *Agrobacterium* by electroporation.

A cyan fluorescent protein (CFP) targeted to localize in the ER was used as a control to verify the probable localization of CP6-YFP in the ER. The control construct was obtained from Arabidopsis Biological Resource Center (ABRC, clone name: ER-CK) and contains an ER retention signal sequence fused to C-terminal region of CFP (Nelson et al., 2007).

2.5 Tobacco transformations

2.5.1 Stable transformation and generation of transgenic lines

To generate stable tobacco transgenic lines, the *CysP* silencing constructs were introduced into leaf cells by *Agrobacterium* mediated transformation. *Agrobacterium* with *CysP* silencing construct (as obtained in section 2.4.1) was grown in 3 mL of infiltration culture medium containing 10 mM morpholino-ethanesulfonic acid [MES] pH 5.6, 100 µM acetosyringone supplemented with rifampicin (10 µg/mL), gentamycin (50 µg/mL) and kanamycin (50 µg/mL for pK7GWIWG2D(II),0-CysP) and spectinomycin (50 µg/mL for pB7GWIWG2(II),0-CysP). The 3 mL culture was grown overnight at 28 ^{0}C and transferred to a fresh 40 mL infiltration culture medium later. The culture was allowed to grow at 28 ^{0}C until it reached an optical density (OD600) of 0.5-0.8. After the optimum OD was obtained, the culture was transferred to a 50 mL conical tube and centrifuged at 3000g for 30 min. The pellet was resuspended in Gamborg's solution (3.2 g/L Gamborg's B5 and vitamins, 20 g/L sucrose, 10 mM MES pH 5.6, and 200 µM acetosyringone) to a final OD600 of 1 and incubated at room temperature for 2h with gentle agitation to activate the virulence gene required for transformation.

Transformation of tobacco leaves was performed using the leaf disc transformation method in a sterile laminar flow hood. Five mm^2 tobacco leaf **e**xplants were submerged in *Agrobacterium* suspensions, briefly dried on a filter paper, and placed on plates containing MS medium supplemented with naphthalene acetic acid (NAA- 1 mg/L) and benzyl aminopurine (98 µg/L) (MST-agar medium). After 2-3 days of co-culture with *Agrobacterium*, the leaf explants were transferred to a new MST-agar medium containing Timentin (500 mg/L) and BASTA (4 mg/L). Timentin helps to kill live *Agrobacterium* cells after the transformation. The explants were then transferred to a BASTA only MST medium after 3 days and the medium was changed every week until calli developed from the transformed plant cells. The regenerated calli differentiated into independent shoots, which were cut and transferred to a rooting MST-agar medium after 6 weeks. The rooting medium was supplemented with NAA (1 mg/L) only. *CysP* silenced shoots were transferred to the soil after sufficient roots were grown in the

rooting medium. Individual lines were grown in different pots and transferred to the greenhouse. The plants were grown together with IL-10 control plants and used for the evaluation of IL-10 levels.

2.5.2 Transient expression

To study the subcellular localization of CysP6 and the effect of overexpressing some *CysPs* on IL-10 accumulation, transient transformation was used. For the subcellular localization study, 4-6 weeks old tobacco leaves (tobacco, in this case means *N. benthamiana*) were infiltrated with the transformed *Agrobacterium* suspension culture as described in the section 2.5.1. For co-localization with the ER control, *Agrobacterium* cultures containing pEG101-CP6-YFP and pEG101-ERCFP constructs were mixed in a ratio of 1:1 and co-infiltrated in the tobacco leaves. Infiltration was done using a 1 mL needle-less syringe by pressing the plunger against the lower leaf epidermis. The plant was left at room temperature and the subcellular localization was visualized after 48 h using confocal microscopy.

For the transient overexpression of *CysP*s, *Agrobacterium* cells containing the *CysP* overexpression constructs were infiltrated into tobacco, as described in section 2.5.1 for the transfer of transgenes into plant, and with the help of needle-less syringes. The tissues were collected 4 days post infiltration in order to analyze the effects on IL-10 accumulation levels.

2.6 Confocal Microscopy

Transient expression of CP6-YFP and CFP-KDEL was visualized using a Leica TCS SP2 inverted microscope. A 63X water immersion objective lens was used at an excitation wavelength of 514 nm for YFP and 434 nm for CFP. The emission wavelengths for YFP and CFP were between 530-560 nm and 460-490 nm, respectively.

2.7 Sample collection for protein and RNA extractions

Leaf tissues were sampled when tobacco plants reached a height of ~1.5 ft. Six to eight leaf discs were taken from the $3^{rd,}$ 4^{th} and 5^{th} leaves from the top. Leaf tissues were collected in 2 mL Eppendorf tubes that contained 2 small copper beads (BioSpec Products Inc, USA). Once the leaf samples were collected and the weight recorded, they were immediately frozen in liquid N_2 and stored at -80 ^{0}C freezer.

2.8 Protein extraction

For protein extraction, tubes containing leaf samples were placed in precooled homogenization blocks, and homogenized using a mixer mill at 25 Hertz for 1 min. Protein extraction buffer (400 µL) containing 1X Phosphate buffer saline (PBS, pH 7.4), 0.1% TWEEN-20, 1mM EDTA, 100 mM sodium ascorbate, 2% PVPP, 1 mM PMSF and 100 µg leupeptin was added to the homogenized tissue. PVPP is insoluble in water, hence required at least 2-4 h of hydration before use. The lysates were centrifuged at 14,000 *g* at 4^0C for 10 min in a microcentrifuge, and the supernatants were collected.

2.9 Protein assay

Total soluble protein was measured using the Bradford dye reagent (Bio-Rad Laboratories, Inc., USA). Protein assays were carried out according to Bradford's method (Bradford, 1976). Bovine serum albumin (BSA) solutions of concentrations 0, 40, 100, 200, 400, 600 and 800 µg/mL were used for creating a standard curve. Standards and the TSP samples (10 µL each) were added to a 96-well plate, and 200 µL of Bradford dye (diluted 1:5 in ultrapure water) was added to each sample. The absorbance was measured at 595 nm after 30 sec using a BioTek Synergy2 microplate reader (BioTek Instruments, Inc., USA). Three technical replicates were used for each sample and standard. The absorbance of the BSA standards was plotted against their concentrations to generate a standard curve whose linear range was used to determine the concentrations of proteins in unknown samples. The extraction buffer was used as blank to subtract the background.

2.10 Enzyme-linked immunosorbent assay (ELISA)

To measure IL-10 levels in *CysP* silenced and overexpressed tobacco lines, ELISA was used. An enhanced protein binding ELISA plate (Nunc-Immuno, Maxi-Sorp plates, VWR International, USA) was coated with 100 ng of purified rat anti-human IL-10 capture antibody per well and incubated at 4^0C overnight. The antibody was diluted in in 0.1 M Na_2HPO_4 pH 9.0. Following the incubation, the plate was washed two to three times with wash buffer (PBS and 0.05% (v/v) Tween-20) and saturated with 200 µL of blocking buffer (1% BSA in PBS) to avoid non-specific binding in the wells.

The extracted protein samples were diluted by 10 times in the blocking buffer and 100 µL of diluted samples were added to each well of the capture antibody-coated ELISA plate. Pure protein for human IL-10 (BD Pharmingen, Canada) was used as the standard at the concentrations of 2, 1, 0.5, 0.25, 0.125 and 0.0625 ng/mL. Once the standards (10 µL) and samples were loaded, the plate was incubated at room temperature for 4 h and then washed two to three times with the wash buffer. Three technical replicates were used for all samples and standards. Following the wash, biotinylated anti-IL-10 antibody (BD Pharmingen, Canada, 100 ng/well; 100 µl of 0.5 µg/ml diluted antibody solution in blocking buffer) was added to each well using multichannel pipette. The plate was incubated for 1 h at room temperature with gentle shaking, followed by three washes with the wash buffer. After the wash, secondary antibody conjugated to Horse Radish Peroxidase (HRP) (BD Pharmingen, Canada, 100 ng/well diluted in blocking buffer) was added to each well. The plate was kept at room temperature for 30 min followed by four washes with the wash buffer before adding the substrate. Hundred microliter of 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS dissolved in 300 ng/mL of 0.1M citric acid) containing 30% H_2O_2 was added to each well, incubated at room temperature for 3.5 min, and the absorbance was measured at 405 nm using microplate reader (Bio TeK Instruments, Inc., USA). The absorbance of blocking buffer was used as blank. IL-10 standard curve was generated by plotting absorbances of the standards against their respective concentration and the graph was used to determine IL-10 level in the unknown samples.

2.11 Genomic DNA extraction and genotyping of T_1 transgenic lines

To screen transgenic lines, T_1 seeds obtained from the *CysP* silenced T_0 tobacco lines were grown in MST-agar containing BASTA (4 mg/L). The BASTA resistant seedlings were transferred to small pots. Seedlings were grown for 2 weeks and leaf tissue was sampled from each independent transgenic line for genomic DNA extraction.

Genomic DNA was extracted from tobacco leaves using the cetyl-trimethylammonium bromide (CTAB) method (Stewart and Via, 1993). Approximately 5 to 8 leaf discs were lysed in a 2 mL microfuge tube using 500 µL of extraction buffer (1.4 M NaCl, 2% [w/v] CTAB, 50 mM Tris-HCl pH 8.0, 10 mM EDTA and 0.1 mg/mL RNase added just prior to use) and incubated at 65^0C for 20 min. The samples were allowed to cool down to room temperature and 500 µL of chloroform was added. The mixtures were vortexed thoroughly and centrifuged for 3 min at 14000 *g*. The aqueous upper phase was transferred to a clean microfuge tube to which 0.6 volume of isopropanol was added and mixed thoroughly by vortexing. The samples were then centrifuged for 10 min at 14000 *g*, and the pellets washed with 1 mL 70% (w/v) ethanol. The pellets were air dried and dissolved in 20-50 µL of sterile MilliQ water.

Genotyping was carried out to identify transgenic T_1 plants using genomic DNA as template and chloramphenicol resistance primers CmrFw: 5'-GGAGGCATTTCAGTCAGTTGCTCA- 3' and CmrRv: 5'-AGACGGCATGATGAACCTGAATCG-3' (Figure 2.1A).

2.12 RNA isolation and RT-PCR analysis

To check silencing or overexpression of *CysP*s in tobacco, total RNA was extracted from tobacco leaves using a RNeasy Plant Minikit (Qiagen Inc., USA) following manufacturer's instructions with slight modifications. On column DNase1 (RQ1 RNase-Free DNase, Promega, USA) digestion was performed in order to get rid of the any genomic DNA contamination during extraction. RNA was quantified using a NanoDrop spectrophotometer (Thermo SCIENTIFIC, USA). The integrity of RNA was checked by running 300-500 ng of RNA on 1.2% (w/v) agarose gel in 1X TAE buffer (40

mM Tris, 40 mM acetate, 1 mM EDTA, pH 8.2). cDNA was synthesized from 1 µg of RNA using a QuantiTect® Reverse transcription Kit (Qiagen Inc., USA). To check the transcript level of *CysP6* in transient silencing or overexpression experiments, gene specific primers (Table 2.1, CysP6-OE-F/R) were used for 36 cycles of PCR. Ten microliters of amplified product was run on a 1.5% agarose gel.

A quantitative RT-PCR (qRT-PCR) was performed to evaluate the level of *CysP6* in T_1 stable *CysP6* silenced plants. For this procedure, SsoFastTM EvaGreen® Supermix (Bio-Rad Laboratories, Inc., USA) was used with CFX96 real-time PCR detection system (Bio-Rad Laboratories, Inc., USA). *CysP6* transcript accumulation was checked using the following primers, qCP6F: 5'- CACCATATCCCTACTTCTCCTCCTC -3' and qCP6R: 5'- CCATGTTCGACTAGCCATGACTC-3'. The amplicon was cloned into pGEMT vector (Promega, USA) and its sequence was verified. *N. tabacum Actin* (*NtActin*, Gene bank ID: AB158612.1) was used as a reference gene. The primers used for reference gene amplification were, qNTActinF: 5'-GGTTGGTATGGGTCAAAAGGATCA-3' and qNTActinR: 5'-GGAGCAACACGCAACTCATTG-3'. The melting curve analysis was performed with pDONR-CysP6 plasmid as a template.

Chapter 3: Results

3.1 Tobacco genome contains at least 60 putative CysPs

To identify *CysP* genes in tobacco, an *in silico* search was done using the DFCI tobacco gene index (TGI) database. The TGI consists of 324,058 ESTs. A total of 55 putative *CysP* genes were identified by the keyword search 'cysteine protease'. Out of 55 putative *CysP* gene sequences, 32 were tentative contig sequences which were created by assembling ESTs into virtual transcripts. The rest were 23 singleton sequences, each of which represented a unique EST in the database. These putative *CysPs* were derived from tobacco cDNA libraries constructed from a variety of tissues including leaf, flower, root, whole seeding or cultured tobacco (cv Bright Yellow, BY-2) cells. Table 3.1 provides a list of putative *CysP* genes as represented by their tentative contig or singleton sequence identification numbers, tissue specific expressions and predicted protein sizes for those where complete coding regions were available. Of the 55 putative *CysP* genes, 10 genes were selected as candidates for further study based on complete sequence availability and expression in leaf tissue. Leaf specific CysPs were of main interest, as the principal objective of this research was to identify potential CysPs involved in determining IL-10 yield in leaf. The candidate *CysPs* were named as *CysP1* to *CysP10*.

In addition to the 55 putative *CysPs* identified through a database search, five other *CysP* sequences were found through a literature search. Their accession numbers and the paralog tentative contig sequences are shown in Table 3.2. Among them, *NtCP1* was expressed only in senescing leaves and *NtCP2* was expressed in mature tobacco leaves. *NtCP56* was strongly expressed in anthers. *CyP7* and *CyP8* were expressed in leaf tissues. A very high degree of amino acid identity was found between NtCP56 and CysP10 (99.17%), which indicates that they might perform similar functions in tobacco.

Table 3.1 List of putative *CysP* genes in tobacco.

TC sequence	Most expressed in[a]	Candidate *CysP*	Predicted Protein size (AA)[b]	Singleton Sequence	Most expressed in	Predicted Protein size (AA)[b]
TC122594	R,L	***CysP1 (NtCP23)***	360	BP132783	BY-2	_
TC123875	R		_	BP529990	BY-2	_
TC124565	L,F,R	***CysP2***	240	CN498801	L	_
TC124614	R,L		_	DV159633	Seed	_
TC124631	L,F,R		_	DW001815	L,R	_
TC126531	L,F,R	***CysP3***	_	AM837057	L	_
TC128311	L,F,R		_	AM829003	L	_
TC129167	L,F,R		_	AM831937	L	_
TC129618	L,F,R, BY-2	***CysP5***	360	AM791536	L	_
TC130400	R		_	AM787277	L	_
TC130996	L,F,R		_	AM846161	W	_
TC132948	L,F,R	***CysP6 (NtCP6)***	466	AM834885	W	_
TC133645	L,F,R	***CysP4***	349	AM836936	W	_
TC133786	BY-2		_	AM801664	L	_
TC135985	L,R		_	AM787303	W	_
TC139681	L,F,R		_	AM823454	W	_
TC140530	L,R,W		_	AM801795	W	_
TC141516	L,F,R	***CysP7***	355	AM844688	W	_
TC143494	L,F,R	***CysP8***	434	FG634183	L	_
TC143787	L,F,R	***CysP9***	301	FG644916	L	_
TC144748	L, R,W		_	FG167001	R	_
TC145665	L,R		_	FG172560	R	_
TC151953	W		_	FG190220	L,F,R	_
TC152484	L,F,R,W		_			
TC154564	L,R		_			
TC155175	L,F,R,W		_			
TC162422	L		_			
TC163159	L,W		_			
TC164523	L		_			
TC164665	L,F,R		_			
TC166013	Proembryo		_			
TC166795	L,F,R	***CysP10***	361			

[a]L, leaf; F, flower; R, root; W, whole plant, BY-2, *Nicotiana tabacum cv.* Bright Yellow cell line; [b]AA, amino acid; -, partial *CysP* sequences

Table 3.2 Comparison of published tobacco *CysP* sequences with tentative contig sequences from tobacco EST database.

***CysP* genes**	**Accession #**	**Closest tentative contig[a]**	**Amino Acid identity (%)**	**Reference**
NtCP1	AY881011	TC133645 (*CysP4*)	94.84	(Beyene et al., 2006)
NtCP2	AY881010.1	TC124565 (*CysP2*)	96.6	(Beyene et al., 2006)
CyP7	Z13959.1	TC131927	97.8	(Linthorst et al., 1993)
CyP8	Z13964.1	TC131927	97.5	(Linthorst et al., 1993)
NtCP56	EU429306.1	TC166795 (*CysP10*)	99.17	(Zhang et al., 2009)

[a]highlighted tentative contigs, candidate *CysPs*

3.2. Catalytic triad and non-contigious ERFNIN motif are conserved in candidate CysPs

CysP1- CysP8 and CysP10 have a common papain domain that has been shown to be responsible for endo and exo-peptidase activity of CysPs. Proteins containing such domains are classified under the papain or peptidase_C1 family and are synthesized as precursor CysPs (Wiederanders et al., 2003) .

Precursor CysPs have an additional propeptide sequence preceeding their papain domain. As shown in Figure 3.1, all of these CysPs have a propeptide sequence which is also called as I29 or the cathepsin propeptide inhibitor domain. For enzyme activation, papain family CysPs require a cleavage of the propeptide sequence (Wiederanders et al., 2003). CysP6 and CysP8 have an additional granulin-like repeat that is known to play a role in targeting and activity regulation of certain CysPs (Tolkatchev et al., 2001). CysP9 may belong to the otubain or peptidase_C65 family since it consists of a single large Ovarian Tumor (OTU) domain, and may have similar functions to that of deubiquitinating enzymes of the peptidase_C65 family. Detailed information regarding putative domain lengths of all candidate CysPs is shown in detail in Table 3.3**.**

To identify catalytic residues in the CysPs, the deduced amino acid sequences of those belonging to the papain family were aligned, along with a previously characterized tobacco CysP, NtCP23 (Figure 3.2). A catalytic triad comprised of Cys25, His159 and Asn175 (papain numbering) was found in the papain domain of all aligned CysPs. The catalytic site of these CysPs is a thiolate-imidazolium ion pair between Cys25 and His159, which forms a reactive cysteine nucleophile (Grzonka et al., 2001). The alignment shown in Figure 3.2 excludes CysP3 as it lacked several N-terminal amino acid residues, and critical residues Asn and His in the catalytic triad. A highly conserved block of amino acid interspersed with variable residues, **E**X_3**R**X_3**F**X_2**N**X_3**I/V**X_3**N (**ERFNIN motif), is present in the propeptide region of all papain family CysPs (Karrer et al., 1993). The candidate sequence alignment also revealed such general features in the papain family CysPs (Figure 3.2).

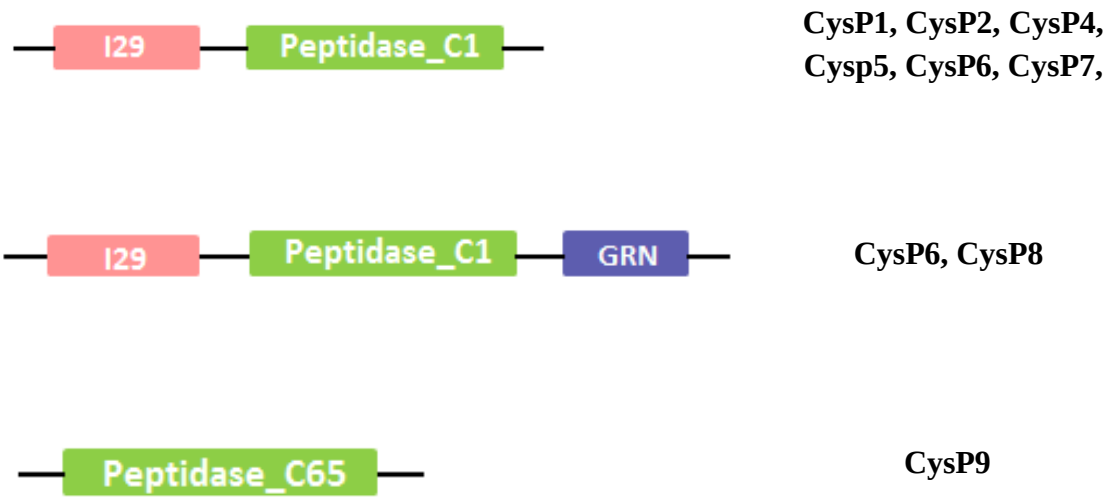


Figure 3.1: Schematic representation of different domains in papain family tobacco CysPs. Candidate CysPs having specific domain types are indicated on the right side. The diagram is not drawn to scale. I29, cathepsin propeptide inhibitor domain; Peptidase_C1, papain family domain; GRN, granulin like repeats; Peptidase_C65, otubain family domain

Table 3.3 Detail domain information of the candidate CysPs.

Candidate CysPs	Cathepsin propeptide inhibitor domain	Papain family CysP domain	Granulin like repeats	Otubain domain
CysP1(NtCP23)	61-117	143-358	X	X
CysP2	Cleaved	5-221	X	X
CysP3	53-110	141-246*	X	X
CysP4	42-99	131-348	X	X
CysP5	61-117	143-358	X	X
CysP6(NtCP6)	50-107	138-353	387-435	X
CysP7	49-105	137-352	X	X
CysP8	1-41	75-292	339-387	X
CysP9	X	X	X	45-300
CysP10	38-93	126-342	X	X

Numbers in the domains represent amino acids.

*complete coding region sequence was not available.

X, absent in respective CysPs.

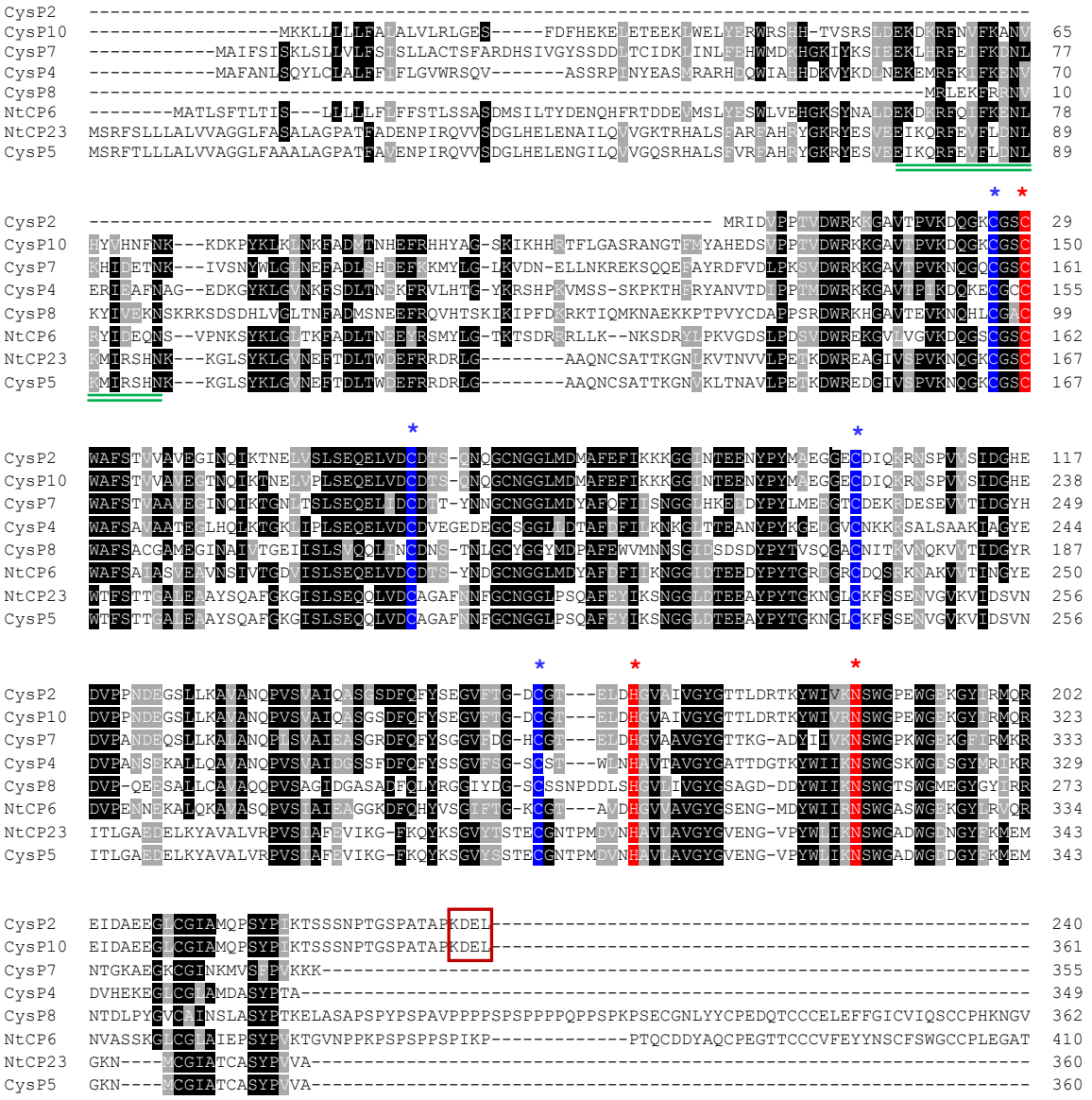


Figure 3.2: Multiple sequence alignment of papain family CysPs. Predicted amino acid sequences of NtCP23 (CysP1), CysP2, CysP4, CysP5, NtCP6 (CysP6), CysP7, CysP8 and CysP10 were aligned using CLUSTALW and shaded using BOXSHADE 3.21. Identical and conserved amino acids are shaded in dark and light grey, respectively. (*) indicates catalytic residues cysteine (C), histidine (H) and asparagine (N). (*) indicates the disulfide bridge-forming cysteine residues. Red box indicate endoplasmic reticulum retention signal KDEL in CysP2 and CysP10. The conserved ERFNIN motif in the propeptide region is double-underlined.

3.3 Generation of candidate *CysP* RNAi constructs

To identify *CysPs* that influence IL-10 accumulation in tobacco, RNAi silencing of selected *CysPs* was performed. Candidate genes *CysP1* to *CysP5* were cloned into RNAi vector, pK7GWIWG2D(II),0 and *CysP6* to *CysP10* in pB7GWIWG(II),0 as described in section 2.4.1 and Figure 2.2A. Gene specific unique sequences were cloned in opposite orientation in the destination vector to create RNAi constructs. The RNAi constructs are transcribed to the hairpin RNA precursors which are processed into the small interfering RNA's (siRNA) by the homologues of the enzyme dicer and incorporated into RNA-induced silencing complex (RISC). Silencing is achieved either by sequence specific degradation of complementary target sequences or by repression of *CysP* messenger RNA by the RISC (Meister and Tuschl, 2004).

Unique regions were identified for each of the *CysP* sequences to facilitate targeted gene silencing. RNAi vectors with the specific *CysP* inserts were screened using two different primer combinations by PCR, which gave the amplicons of different sizes as shown in Figure 3.3. The first amplicon was produced by the gene specific *CysP* forward primer (CysPF) and *Cmr* forward primer (CmrF), and contained the larger intron (In-1, 376 bp) region. The second amplicon was produced by CysPF and the *Cmr* reverse primer (*CmrR*), and contained the smaller intron (In-2, 275 bp) region. The silencing constructs were transformed into *Agrobacterium* in-order to deliver them into tobacco leaf epidermal cells.

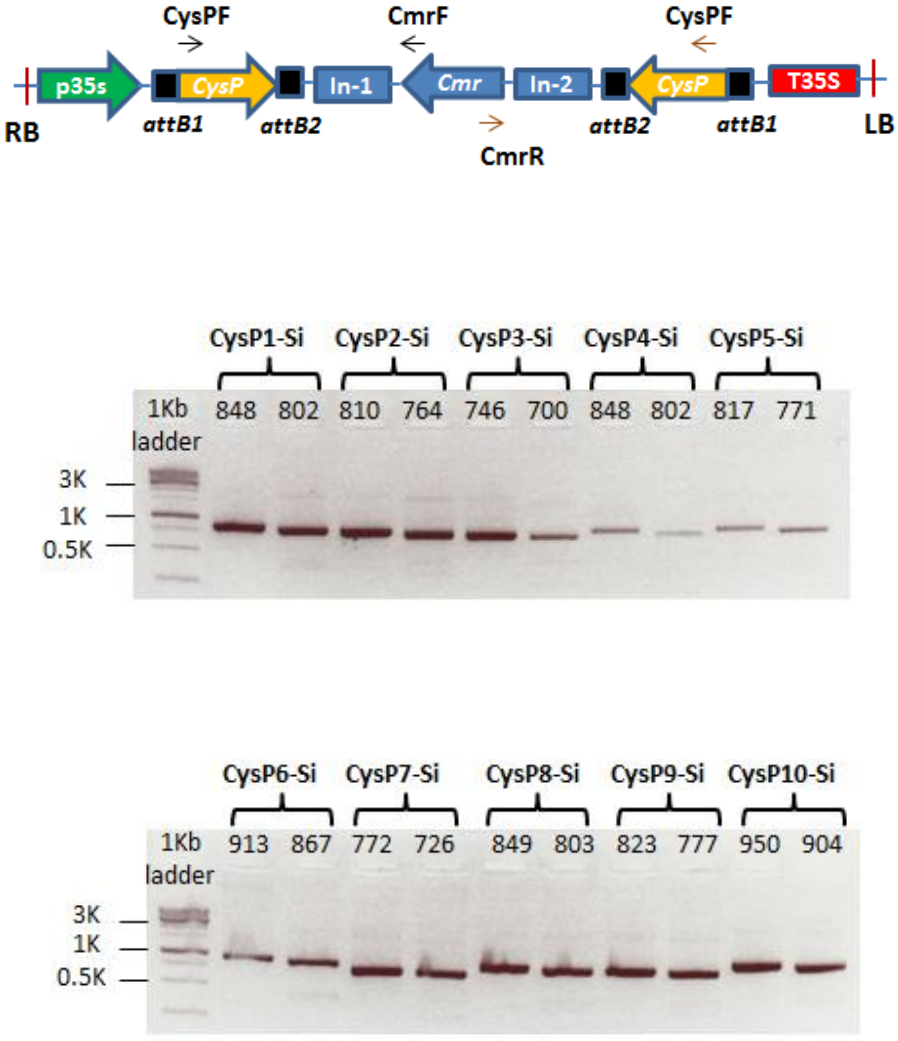


Figure 3.3: Strategy for *CysP* silencing. Screening of *CysP* double inserts in the silencing vector. Primer combinations used for screening are shown by arrowheads in the RNAi construct. Gene specific forward primers (CysPF) and chloramphenicol resistance forward and reverse primers (CmrF and CmrR) produce the *CysP* amplicons of specific sizes. Agarose gel images demonstrate the amplicon size differences for specific *CysP* inserts. The numbers on the top of each lane correspond to the amplicon size.

3.4 RNAi silencing of candidate *CysP* genes and generation of stable *CysP* silenced tobacco lines

To generate *CysP* silenced tobacco lines, *Agrobacterium* cultures harbouring *CysP* RNAi constructs were used to transform tobacco G7 line. The G7 line is a useful control for this experiment as in the untransformed state it constitutively overexpresses IL-10. Figure 3.4 describes the process of generating *CysP* silenced tobacco plants using leaf disc transformation and the approximate time it required to proceed from one stage to the other. The leaf disc and *Agrobacterium* were co-cultured for 2-3 days and transferred to callus induction medium containing herbicide BASTA for selection. After nearly 1.5 weeks, calli were induced from the transformed tobacco cells, leaving the untransformed cells dead in the tissue culture medium. The induced calli were transferred to the differentiation medium where they gave rise to small shoots in 4-6 weeks. A single shoot was cut from individual callus and transferred to a new plate which later giving rise to an independent transgenic line. In approximately 16 weeks from the initial transformation, the independent transgenic lines were ready for an evaluation of *CysP* transcript and IL-10 levels.

In addition to the targeted silencing of specific candidate *CysPs*, a conserved region of *CysP1* to *CysP5* genes was also chosen for silencing and two silenced lines (*CysPAll-Si*) were generated. A total of 82 *CysP* silenced T_0 lines were generated for all candidate genes (Table 3.4). Since no RNAi line could be generated for *CysP5*, further work on that candidate gene was not carried out.

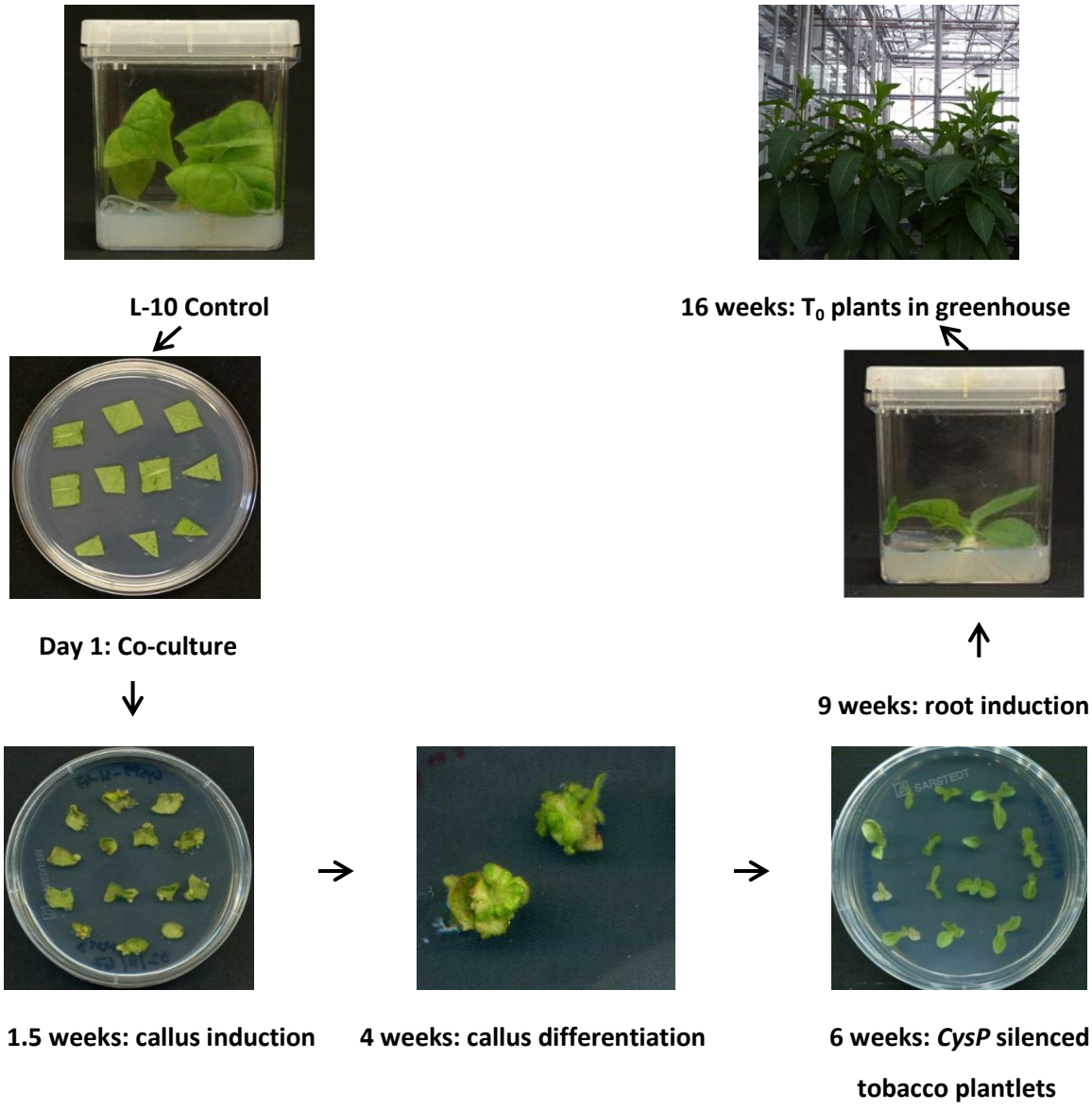


Figure 3.4: Generation of stable *CysP* silenced tobacco lines. G7 tobacco lines overexpressing IL-10 protein were grown in magenta boxes. Leaf discs were cut and co-cultured with A*grobacterium* cultures containing candidate *CysP* silencing constructs. Calli were regenerated from transformed cells and subcultured in differentiation medium containing BASTA (4 mg/L). Independent *CysP* silenced transgenic lines were obtained from the individual callus and transferred to the rooting media before transferring them to greenhouse.

Table 3.4: Numbers of independent *CysP* silenced T_0 tobacco lines.

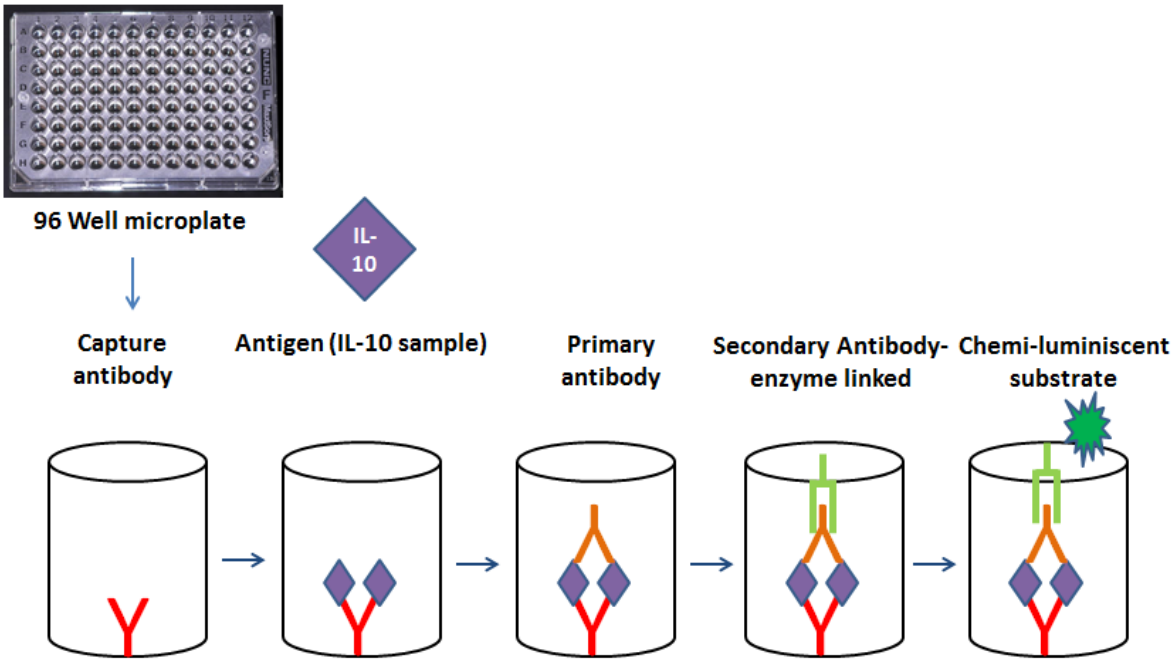
Candidate *CysP* gene	*CysP* silenced lines (T_0)	Candidate *CysP* gene	CysP silenced lines (T_0)
CysP1	11	*CysP6*	16
CysP2	6	*CysP7*	9
CysP3	8	*CysP8*	9
CysP4	5	*CysP9*	12
CysPAll-Si	2	*CysP10*	4

3.5 Silencing of *CysP6* increases IL-10 accumulation in independent T_0 tobacco lines

To determine the effect of *CysP* silencing in the accumulation of recombinant IL-10, the IL-10 levels were measured using a double sandwich (dsELISA) in T_0 tobacco lines, all of which were generated from the independent transformation events. The principle of dsELISA is shown in Figure 3.5. The dsELISA results revealed that IL-10 levels were higher in six out of thirteen *CysP6* silenced lines compared to the controls (Figure 3.6). Three other lines measured for IL-10 (CysP6Si-4, 7 and 13) displayed almost similar levels of IL-10 accumulation in comparison to the control. IL-10 level was increased by nearly 1.6- fold in CysP6Si-15, the highest IL-10 accumulating line. Results indicate that silencing of *CysP6* can affect IL-10 accumulation in T_0 lines of G7 tobacco.

IL-10 accumulation levels were also checked for *CysP1, CysP2, CysP3, CysP4, CysP7, CysP8, CysP9* and *CysP10* silenced independent T_0 lines. As shown in Figure 3.7, there was no increase in IL-10 accumulation among any of these lines when compared to the levels in IL-10 control plants except, for *CysP8* silenced line (CysP8Si-3). These *CysP* silenced lines were generated over different periods of time. Therefore, for an appropriate comparison between the IL-10 levels, IL-10 control plants were also grown at the same time with *CysP* silenced plants. The accumulation of IL-10 varied among the control plants, even though the plants were grown under similar conditions. Thus, IL-10 accumulations for different T_0 lines are represented as normalized fold levels. The normalized fold levels are the ratios of IL-10 levels present in the silenced lines to the levels present in the IL-10 control plants.

A



B

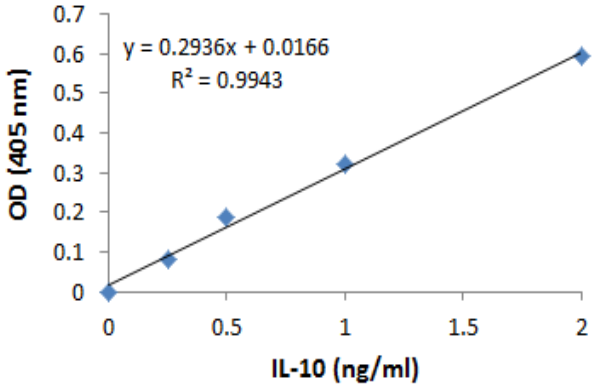


Figure 3.5: Detection of IL-10 by dsELISA. A) Schematic illustration of IL-10 quantification. The 96-well ELISA plate was coated with anti- IL-10 capture antibody and 0- 2 ng/ml of IL-10 standard was added. Primary and secondary antibodies were added allowing triplicate determination to be made in each channel. **B)** Standard curve for the IL-10 concentration versus luminescence intensity (OD405). Luminescence increased in concentration dependent fashion.

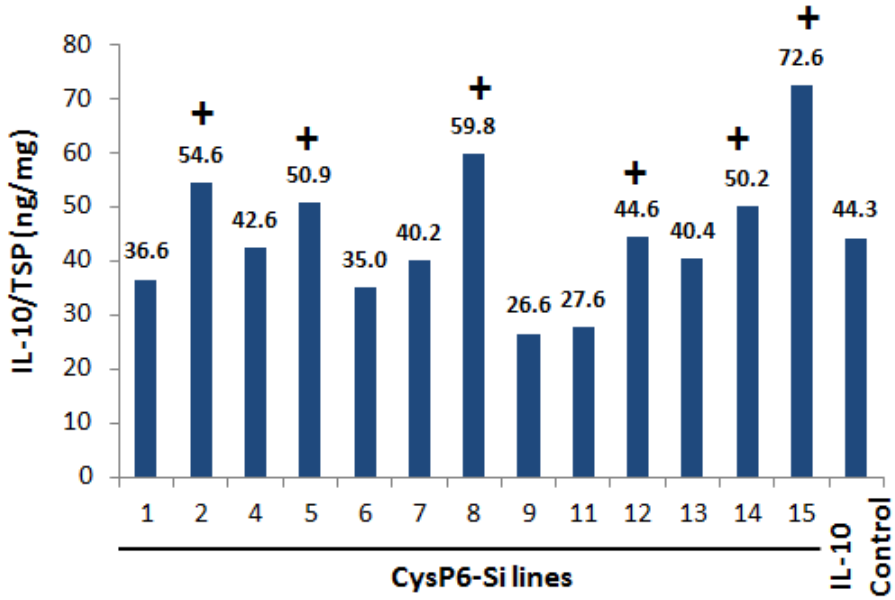


Figure 3.6: IL-10 accumulation in *CysP6* silenced T_0 tobacco lines. Measurement of IL-10 level in independent T_0 *CysP6* silenced lines by dsELISA. Data label above the bar represents IL-10 level of individual *CysP6* silenced lines expressed as ng/mg of total soluble protein. + indicates the *CysP6* silenced lines with higher level of IL-10 accumulation in comparison to the IL-10 control. TSP, total soluble protein; Si, silenced

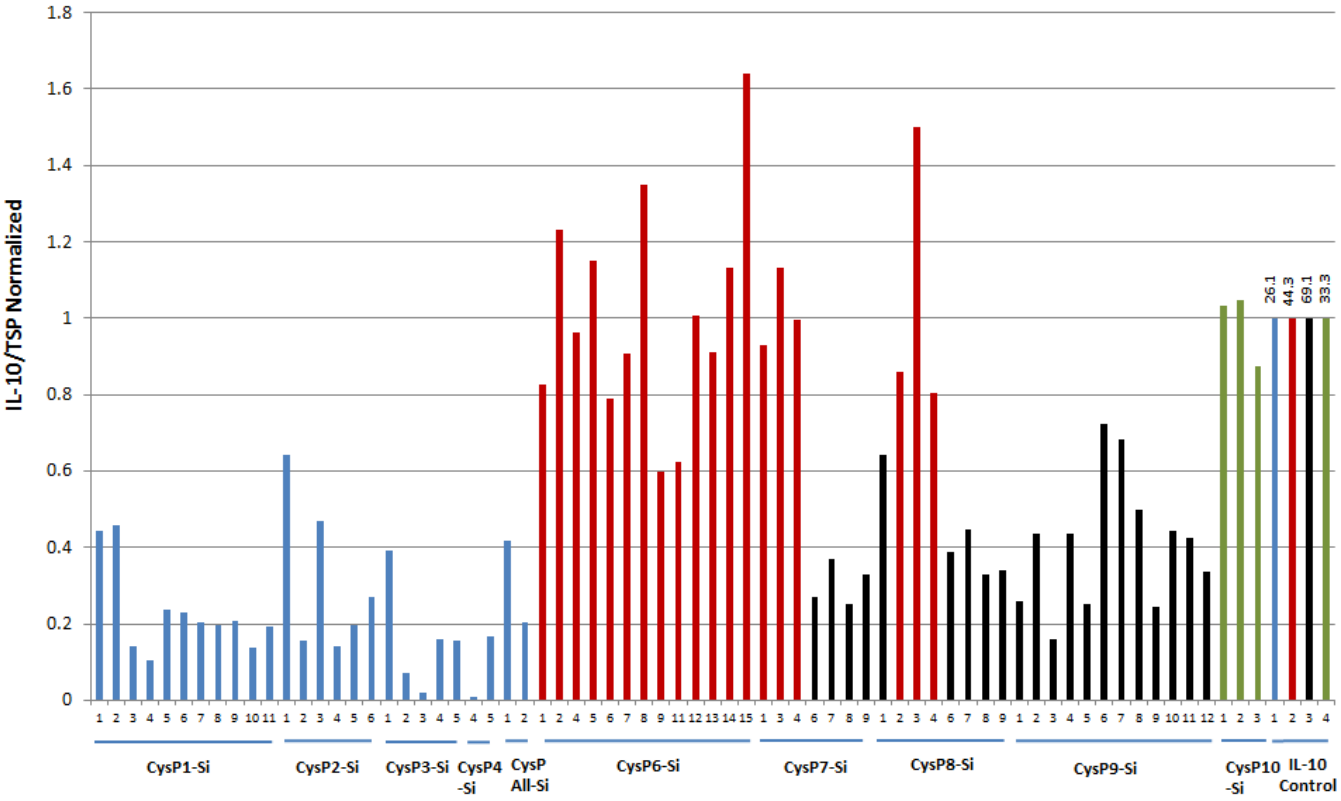


Figure 3.7: IL-10 accumulation in *CysP* silenced T_0 tobacco lines. IL-10 accumulation remains lower in comparison to the controls in T_0 generation of *CysP1-CysP10* (except *CysP6*) silenced lines. Y-axis represents the normalized value of IL-10/TSP for different independent transgenic lines grown at different time period. Blue, red, black and green bars are normalized against their respective colour coded IL-10 level in IL-10 control plant. Numbers above the bar represent actual IL-10 level in ng/mg of total soluble protein in control plants. Si, silenced

3.6 IL-10 accumulation increases with plant maturity in tobacco

To evaluate if the degree of plant maturity affects the production of recombinant protein, IL-10 levels in *CysP6* silenced lines and IL-10 controls were checked at two different time points. The plants were grown simultaneously in the greenhouse and IL-10 ELISA was done at 4 weeks and 7 weeks after transfer to the greenhouse. The plants were approximately 0.5 ft. tall at 4 weeks of age and approximately 1.5 ft. tall at 7 weeks of age. As shown in Figure 3.8, the levels of IL-10 accumulation in the leaf tissue increased as plants progressed to maturity in both *CysP6* silenced and IL-10 control plants. Based on this result, to maintain a consistency in experimental approaches, further experiments were performed using 7 week-old plants. Moreover, no abnormal growth or any morphological differences were observed for the *CysP* silenced lines compared to IL-10 controls.

3.7 Transient silencing of *CysP6* increases IL-10 accumulation

To further examine the effect of *CysP6* silencing on IL-10 accumulation levels, transient expression experiments were performed in tobacco, and transcript and protein accumulation were studied. *CysP6* silencing or vector-only (control) constructs were transiently delivered to a total of seven, 1.5 ft. tall IL-10 control plants using agro-infiltration of leaf tissue (Figure 3.9A). After 4 days post infiltration, tissues were collected for both protein and transcript analysis. Unlike in vector-only plants where the *CysP6* transcript was detected by RT-PCR, the *CysP6* transcript was undetectable in all 7 biological replicates infiltrated with the silencing construct (Figure 3.9B**)**. This suggests that *CysP6* was efficiently silenced by transient infiltration of the silencing construct. The same *CysP6* silencing construct was used for the generation of stable *CysP6* silenced transgenic tobacco lines as mentioned previously in section 3.5.

To check if IL-10 accumulation changed in response to decreased *CysP6* transcript, IL-10 levels were measured in leaf tissue collected at the same time along with tissue that was used for RT-PCR.

A

B

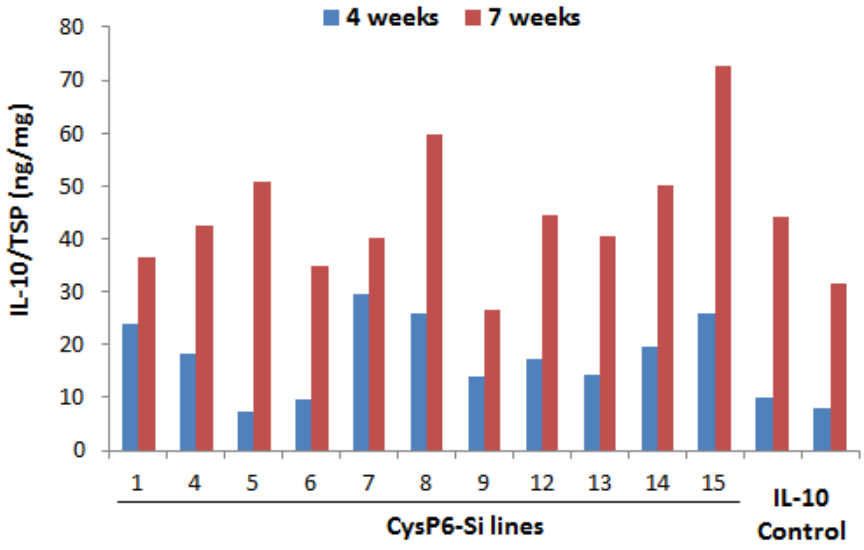


Fig 3.8: IL-10 accumulation in tobacco plants of different ages. **A)** A photograph showing 7 weeks old *CysP* silenced tobacco plants in the greenhouse. **B)** IL-10 accumulation in 4 weeks and 7 weeks old *CysP6* silenced tobacco lines. IL-10 accumulation increased consistently in *CysP6* silenced lines and IL-10 controls over time. The numbers indicate different independent *CysP6* silenced lines. TSP, total soluble protein; Si, silenced.

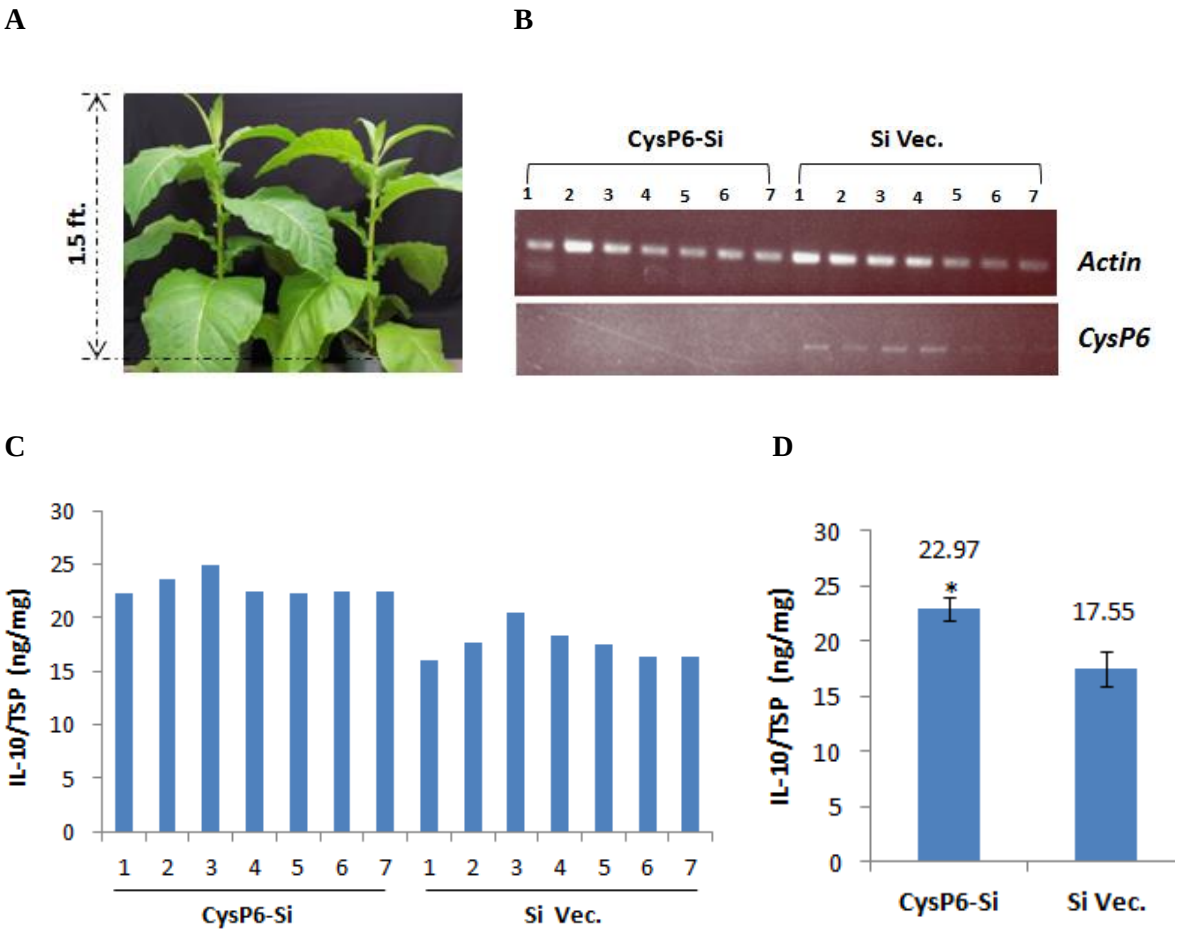


Fig 3.9: Transient silencing of *CysP6* in G7 tobacco. A) 7 Week-old G7 tobacco used for transient *CysP6* silencing experiment. Leaves (4th, 5th and 6th from the top) were infiltrated and tissue was used to determine *CysP6* transcript and IL-10 accumulation. **B)** *CysP6* expression in *CysP6* silencing and vector-only (Si Vec.) infiltrated tissue, 4 days post infection. The top panel corresponds to *Actin* (accession no., #AB158612.1) as a reference gene and the bottom panel corresponds to *CysP6* transcript in the biological replicates (1 to 7). **C)** IL-10 accumulation levels in *CysP6* silencing and vector-only infiltrated tissues. Numbers on the X-axis represent the biological replicates. **D)** Change in IL-10 accumulation due to transient *CysP6* silencing. The IL-10/TSP values are the averages of 7 biological replicates and asterisk (*) represents a significant difference in IL-10 levels between CysP6-Si and control tissues using student *t*-test at 95% confidence level. Si, Silenced; Vec., Vector.

IL-10 levels were found to be higher in all seven biological replicates which were infiltrated with the silencing construct, whereas tissue infiltrated with the vector-only control accumulated lower levels of IL-10 (Figures 3.9C). The difference in IL-10 accumulation between CysP6 silenced and control tissues was found to be statistically significant (Figure 3.9D). These results also indicate that reduction in CysP6 expression results in increased IL-10 accumulation.

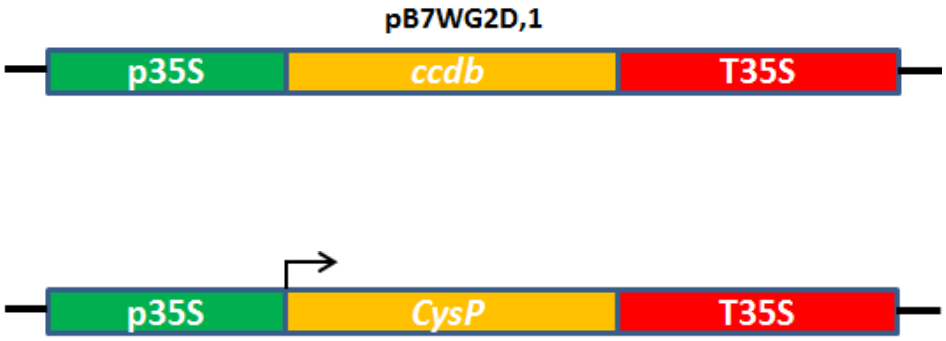
3.8 Transient overexpression of *CysP6* does not reduce IL-10 accumulation

Since silencing of *CysP6* increased IL-10 accumulation in both stable and transient expression systems, overexpression of *CysP6* was carried out to see if it would reduce IL-10 accumulation. IL-10 control plants were transiently infiltrated with *CysP6* over-expression or vector-only constructs and IL-10 level was determined 4 days post infiltration.

As shown in Figure 3.10, it was found that IL-10 accumulation in *CysP6* overexpressed tissue was not lower in comparison to the control. Even though a slight reduction in the levels of IL-10 was observed for *CysP6* overexpressed tissues, the difference was not significant.

In the previous experiment with T_0 stable lines, IL-10 accumulation in *CysP* silenced plants (except for CysP6-Si, lines) was either unaffected or in some cases lower than in IL-10 controls (Figure 3.7). To check if overexpressing these CysPs could affect IL-10 accumulation, full length *CysP* genes (*CysP1, CysP2* and *CysP5*) were successfully cloned and transiently overexpressed in IL-10 control plants. Similar methodologies and conditions were followed as in the transient silencing experiment (section 3.7). Transiently overexpressing these *CysPs* did not change IL-10 accumulation, as no significant difference was observed between IL-10 levels in the overexpression or vector-only infiltrated tissues (Figure 3.10).

A



B

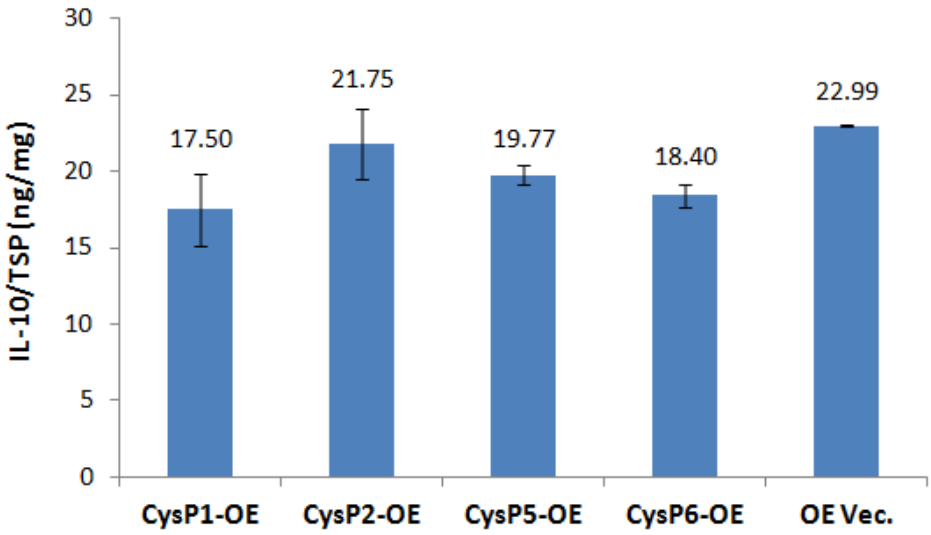


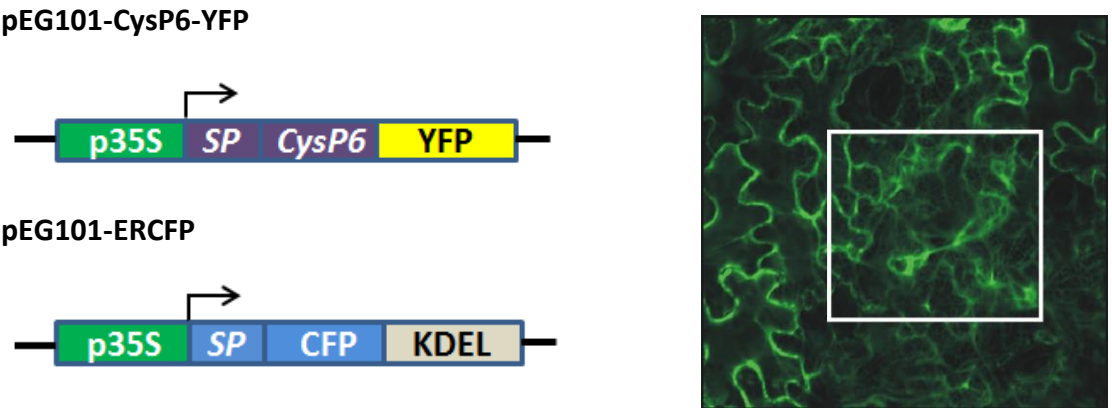
Figure 3.10: Transient overexpression of *CysP*s in G7 tobacco. A) Schemetic representation of overexpression vector with or without *CysP* gene. Open reading frames of *CysP1*, *CysP2*, *CysP5* and *CysP6* were cloned into expression vector, pB7WG2D,1. **B)** IL-10 level in *CysP* overexpression (CysP-OE) and vector-only (OE Vec.) samples. *Agrobacterium* containing specific CysP expression constructs were infiltrated in G7 tobacco plant and IL-10 level was measured 4 days post infiltration. Data labels above the bar represent IL-10 accumulation levels. The data were the averages of two biological replicates for each construct. OE, overexpression; Vec. vector.

3.9 CysP6 localizes in the ER

To sufficiently produce IL-10 in tobacco, the protein was targeted to the secretory pathway and retained within the ER. To study the possibility IL-10 being degraded by CysP6 using the secretory pathway, the intracellular localization of CysP6 was performed.

A translational fusion of CysP6 and YFP was created and transiently expressed in *N. benthamiana* using the methods described in the section 2.5. Protein accumulation was monitored via confocal microscopy. Confocal microscopy is an optical imaging technique, which with the help of fluorescent probes, enables visualization of the intracellular structures and processes occuring inside the cells. Several net-like structures were seen in CysP6-YFP that resembled the localization pattern for the ER (Figure 3.11A). To confirm if the fluorescence in net-like structures was indeed ER, co-localization of CysP6-YFP fusion protein with ER localizing control (ER-CFP) was performed. ER-CFP construct contained both ER targeting at the N-terminal and retention signal sequences (KDEL) fused to the Cterminal end of CFP. Co-expression of CysP6-YFP and ER-CFP showed overlap of CFP signal with that of YFP from CysP6-YFP (Figure 3.11B), thus confirming that CysP6 localizes in the ER in tobacco leaf cells.

A



B

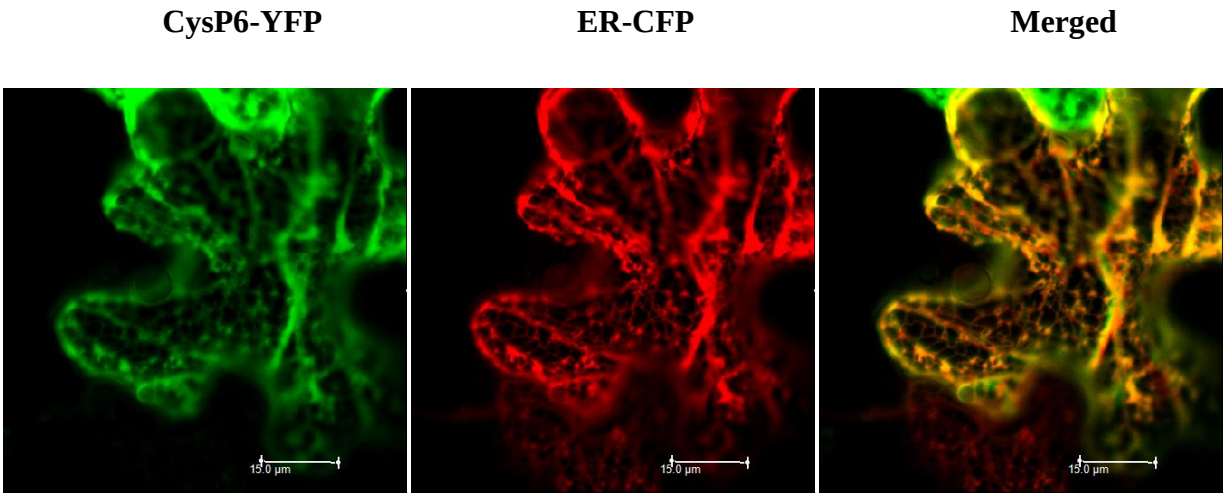


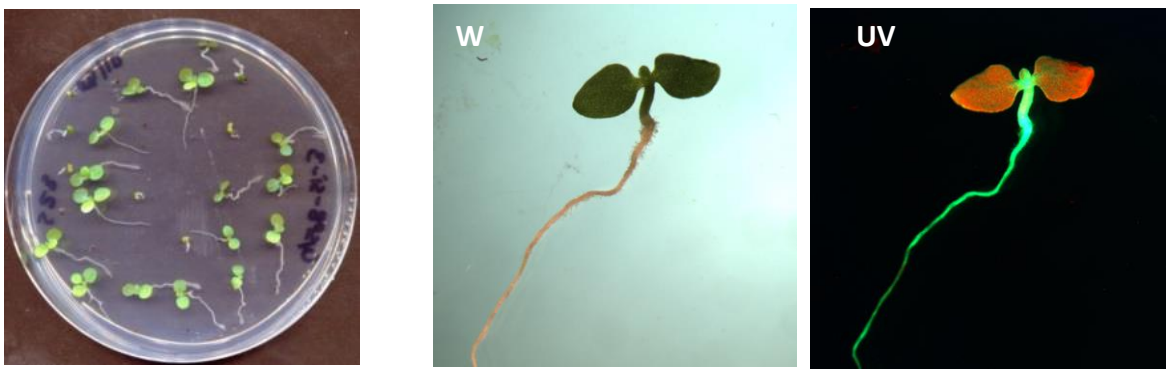
Figure 3.11: ER localization of CysP6. A) *Agrobacterium tumefaciens* carrying pCysP6-YFP and pER-CFP plasmids were infiltrated into *N. benthamiana* and visualized using confocal microscopy. ER localization of CysP6-YFP is seen in tobacco leaf cells. White box indicates the region with network like structures. **B)** Co-expression of CysP6-YFP and ER-CFP showing the overlapping of YFP and CFP signals. SP, signal peptide, Scale bars indicate15 µm. Images shown in A and B are the representatives.

3.10 Generation of *CysP* silenced T_1 tobacco lines

From the ELISA analyses of T_0 plants it had been found that silencing of *CysP6* increases IL-10 accumulation. Moreover, silencing of other candidate *CysPs* lowers or doesn't affect IL-10 levels in the T_0 plants. To study if the effect of *CysP* silencing on IL-10 accumulation can be stable over generations, T_1 plants were generated using seeds obtained from the T_0 *CysP* silenced plants. As plants segregate for transgenes in successive generations, seeds from *CysP6-CysP10* silenced T_0 lines were screened in BASTA containing medium and genotyping was done to confirm the presence of transgene in T_1 plants (Figure 3.12A and 3.12C). For seeds from *CysP1-CysP5* silenced lines, since the transgenic plants inherit *eGFP* from the RNAi vector (see 2.4.1 of method section), selection was done by observing GFP fluorescence under the UV microscope. As shown in Figure 3.12B, a transgenic line is recognized by visualizing green fluorescence under the UV.

The T_1 transgenic lines were also confirmed by genotyping (3.12C). Genomic DNA was extracted from at least 2 biological replicates of each *CysP* silenced T_1 line and primers specific to *Cmr* gene present in the *CysP* silencing construct were used for PCR. Transgenes were present in a total of 64 T_1 lines. The numbers of *CysP* silenced T_1 lines generated for each candidate *CysP* gene is shown in Table 3.5. All the T_1 lines were grown in the greenhouse and used for IL-10 ELISA analysis when the plants were of mature height (nearly 1.5 ft. tall).

A **B**



C

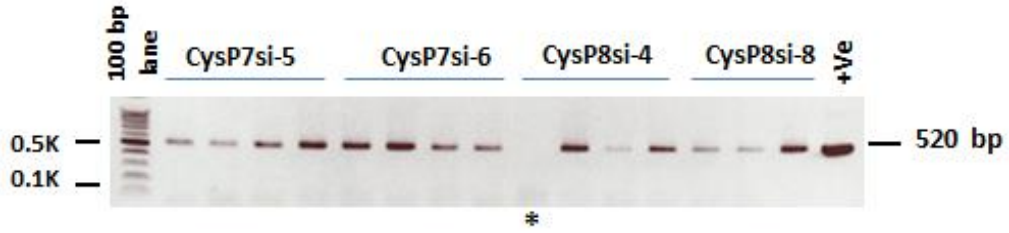


Figure 3.12: Generation of *CysP* silenced T_1 tobacco lines. A) Screening of T_0 seeds from *CysP* silenced tobacco lines. Seeds from *CysP6-CysP10* silenced lines were grown on MS basal medium supplemented with BASTA. B) Screening of T_0 seeds from *CysP1-CysP5* silenced tobacco lines. Seeds were grown in MS basal medium and transgenic plants were selected by observing GFP fluorescence (green colour) under UV. Images show a *CysP* silenced transgenic line under white light (W) and UV. C) Genotyping of T_1 plants By PCR to validate the presence of *Cmr* gene. A representative gel image showing amplicon of 520 bp corresponding to *Cmr* positive lines (contain RNAi constructs). Empty lane indicated by the asterisk sign corresponds to a *Cmr* –ve line (doesn't contain RNAi construct). Positive PCR control, pB7GWIWG2(II),0 plasmid.

Table 3.5: Numbers of *CysP* silenced T_1 tobacco lines.

Candidate *CysP* gene	*CysP* silenced lines (T_1)	Candidate *CysP* gene	*CysP* silenced lines (T_1)
CysP1	9	*CysP6*	12
CysP2	4	*CysP7*	9
CysP3	8	*CysP8*	7
CysP4	3	*CysP9*	12
CysPAll-Si	1	*CysP10*	-

3.11 IL-10 accumulation in *CysP* silenced T_1 tobacco lines

To study the stability of IL-10 accumulation in the *CysP* silenced tobacco plants, IL-10 levels were measured for a total of 30 different T_1 lines. The production of IL-10 was stable over successive generations of tobacco (Figure 3.13). *CysP* silenced T_1 lines (CysP1-Si, CysP2-Si, CysP3-Si, CysP4-Si) all showed lower levels of IL-10 in comparison to their respective IL-10 control plants (represented by same color bars in Figure 3.13). Lower IL-10 accumulation was also seen for these lines in the T_0 generation (Figure 3.7). In a similar pattern to that seen in the T_0 generation, *CysP6* silenced T_1 lines showed nearly equal (line 2 and 15) or higher (line 8 and 9) levels of IL-10 accumulation in comparison to respective control plant. This suggests that, the effect of *CysP* silencing on IL-10 accumulation in tobacco remains similar over the next generation.

3.12 *CysP6* silenced T_1 lines display reduced expression of *CysP6*

To confirm that the increased IL-10 accumulation in CysP6-Si lines resulted from reduced *CysP6* levels, transcript abundance was quantitated using qRT-PCR. As shown in Figure 3.14, *CysP6* expression was lower in all CysP6-Si lines in comparison to the IL-10 controls. A comparison of *CysP*6 silencing and IL-10 accumulation was done between *CysP6* silenced lines and control and the correlation between silencing of *CysP6* and resulting IL-10 accumulation was studied. *CysP6* transcript was higher (IL-10 control), IL-10 accumulation was lower, and as *Cys6* transcript was lower (CysP6Si-9, CysP6Si-8), IL-10 accumulation was higher. However, a high error was seen in *CysP6* expression among CysP6Si-9 and CysP6Si-15 lines. Larger numbers of replicates are required in-order to relate the degree of *CysP6* silencing and accumulation of recombinant protein in transgenic plants.

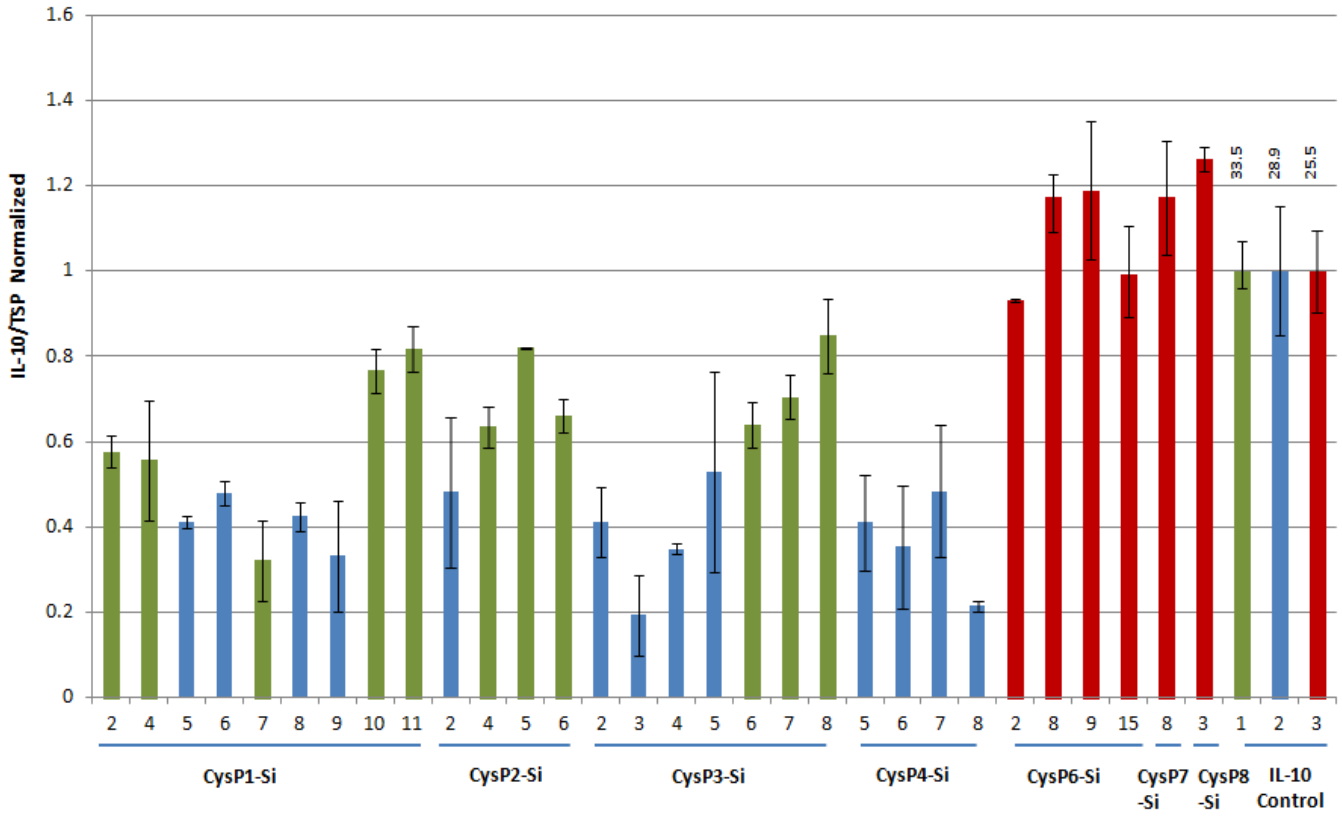


Figure 3.13: IL-10 accumulation in *CysP* silenced T_1 tobacco lines. Normalized IL-10 levels among different CysP-Si T_1 lines and controls. Green, blue and yellow colour bars are normalized to IL-10 levels in their respective respective control plants (IL-10 control, on the right). Numbers above the bar represent actual IL-10 level in ng/mg of total soluble protein. Error bars represent the mean of at two biological replicates for all the lines except CysP6-Si lines 8 and 15 (3 biological replicates).

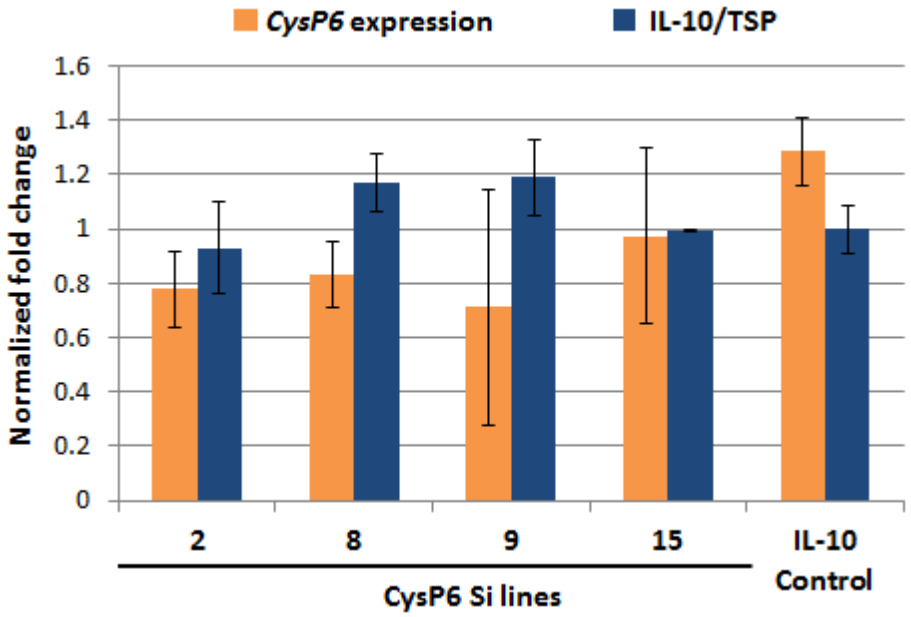


Figure 3.14: *CysP6* expression and IL-10 accumulation in *CysP6* silenced T_1 lines. A comparison between IL-10 accumulation and *CysP6* expression levels of *CysP6* silenced lines and controls. Blue bars represent IL-10/TSP normalized to IL-10 control and orange bars represent fold expression of *CysP6* normalized to *Actin*. Three biological replicates were used for all the lines and control except for lines 2 and 9 (2 biological replicates). Error bars represents the standard deviation.

Chapter 4: Discussion

CysPs are known to be the major proteases that affect the total yields of recombinant proteins in plants. However, none of the CysPs that are implicated in the regulation of recombinant protein accumulation have been reported so far in tobacco, despite the fact that tobacco is an ideal platform for protein production. In this study, the role of tobacco CysPs in regulating recombinant IL-10 accumulation in transgenic tobacco was investigated. The results obtained are discussed below.

4.1 Identification of CysPs involved in IL-10 proteolysis

To identify CysPs that potentially alter IL-10 accumulation in tobacco, all the expressed CysPs were identified by searching the TGI database. Since the tobacco genome sequence has not been published yet, the TGI database is the only major source of information regarding gene sequences and their tissue specific expression. At least 60 putative *CysPs* were identified in tobacco (Table 3.1 and Table 3.2). The total number of *CysPs* may change as more ESTs are deposited in the database or after the availability of whole genome sequence of tobacco. Out of 32 tentative contig sequences available for *CysP*s, 9 contained full open reading frame.

The *CysP* gene family in tobacco appears to be a large gene family, and is comparable to that of other plants such as *Arabidopsis* (38 *CysP* genes) and *Populus* (44 *CysP* genes) (Garcia-Lorenzo et al., 2006). Evolutionarily, tobacco (*N. tabacum*) originated from an interspecific cross of two wild forms, *N. sylvestris* and *N. tomentosiformis*. The interspecific cross between two diploid species with equal number of chromosomes (12 pairs each) resulted into a fertile amphiploid carrying a total of 24 pairs of chromosomes (Leitch et al., 2008). The larger number of *CysPs* in tobacco could have resulted from the preservation of *CysPs* from both of the parents. Many of these CysPs may have critical roles in normal plant growth and development and responses to biotic and abiotic factors that have been characterized in many plant species including tobacco (Grudkowska and Zagdanska, 2004). CysPs are known to initiate and mediate storage protein degradation in germinating cereal seeds (Kato and Minamikawa, 1996).

For example, rice REP-1 digests the rice seed storage protein glutein *in vitro*, and is regulated by giberellic acid in germinating seeds (Ho et al., 2000). The expression of tobacco the CysP, NtCP23, was enhanced during senescence and is similar to what is seen in other plant senescence associated CysPs and mammalian Cathepsin H (Ueda et al., 2000). Degradation of proteins during senescence is essential for amino acid remobilization. Similarly, CysPs have been implicated in cold, drought and wounding stress (Schaffer and Fischer, 1988; Koizumi et al., 1993; Linthorst et al., 1993) , as well as during programmed cell death (Ling et al., 2003; Zhao et al., 2013).

Out of 55 putative CysPs identified in the EST database, CysP1-CysP10 were selected for characterization because they were known to be expressed in leaves and had a complete sequence availability in the EST database. These CysPs were of primary interest as they may target IL-10 expressed in the tobacco leaves. As shown in Table 3.1, 41 out of 55 putative CysPs were found to be expressed in leaf, which was considered as either in leaf only or seedling or in pooled sample of leaf, flower and root tissues. Eight out of these 41 (*CysP1, CysP2, CysP4, CysP5, CysP6, CysP7, CysP8* and *CysP10*) consisted of complete open reading frames and their putative protein sequences contained peptidase_C1 domain (Figure 3.1). Peptidase_C1 is also called papain domain, as the name is given to all the CysPs containing Cys, His and Asn amino acid residues in their catalytic triad and was first characterized in Papaya proteinase 1 from *Carica papaya* (Grzonka et al., 2001). Many papain family CysPs have been characterized in plants and animals that are known to degrade indigenous and foreign proteins (Shenai et al., 2000; Outchkourov et al., 2003).

The papain family tobacco CysPs consist of 38-250 long propeptide sequence and are possibly synthesized as precursor CysPs (Grudkowska and Zagdanska, 2004) (Figure 3.1 and Table 3.3). Their papain domain comprised 220-260 amino acids and might represent the mature enzyme. Activation of papain CysPs is effected by intra- or inter-molecular proteolysis, where the propeptide sequence is cleaved off (Wiederanders et al., 2003). CysP6 and CysP8 contain an extra GRN motif, the role of which in plant is not well understood, but is implicated in protein-protein interactions, targeting and/or activity regulation inside the cell (Tolkatchev et al., 2001). In *Arabidopsis*, GRN is removed by

autocatalysis in RD21, a CysP involved in plant immunity and senescence. However, in an *in vitro* condition, removal of GRN is not necessarily required for enzyme activation (Gu et al., 2012).

CysP9 is an otubain CysP. CysP9 is predicted to contain a large OTU domain, and Asp, Cys and His residues in its putative catalytic triad. Proteins containing such domain are classified as deubiquitinating enzymes, and have been shown to play a role in removing ubiquitin from the proteins that are bound for degradation in proteasome-ubiquitin pathway (Balakirev et al., 2003), suggesting a similar role for CysP9 in tobacco.

As shown in Table 3.2, five additional full length *CysP* sequences were identified and characterized in tobacco, bringing the total to at least 60 *CysP*s. These sequences were not represented in the TGI database, suggesting the possibility of additional *CysP*s. The release of the whole genome sequence of tobacco will help in the genome-wide identification of *CysPs* genes. These five additional sequences showed a high degree of amino acid identity with the TC sequences in the TGI database, and were expressed mostly in tobacco leaf tissue. NtCP1 was expressed only in senescing leaf and is likely a senescence marker in tobacco plant (Beyene et al., 2006). NtCP2 is a CysP with KDEL signal in its C-terminal and was expressed in the mature tobacco leaves (Beyene et al., 2006). NtCP56 was strongly expressed in anthers and was involved in pollen grain development (Zhang et al., 2009). *CyP7* and *CyP8* were weakly expressed in leaf tissue and their transcript accumulation increased by six fold in response to wounding, indicating that they are possibly stress related CysPs (Linthorst et al., 1993). A very high degree of amino acid identity was found between NtCP56 and CysP10 (99.17%), with changes in three amino acid residues (Iso162 to Tyr, Ser172 to Pro, Lys306 to Arg). At the nucleotide level, *NtCP56* is 98.6% identical to *CysP10* with the sequence coverage of 66.6% of *CysP10*, suggesting that they are not the same *CysPs*. However, as NtCP56 and CysP10 share very high degree of amino acid identity, it can be speculated that these two CysPs are protein isoforms in tobacco. CyP7 and CyP8 showed similarity to TC131927, which is surprisingly not included in the list of 55 putative CysPs, but was found as the result of BLAST with the TGI database by using the respective nucleotide sequence as the query sequence.

4.2 ERFNIN motif and catalytic triad is conserved in papain family tobacco CysPs

Sequence alignment of the papain family candidate CysPs, which also included two previously characterized NtCP6 and NtCP23, showed a conserved ERFNIN motif in their propeptide sequence (Figure 3.2). The ERFNIN motif is shared by CysPs from diverse group of species, ranging from protozoa to mammals, and has been mostly reported in mammalian cathepsin H and L type proteins (Karrer et al., 1993). Mammalian cathepsins are known to carry out different intracellular and extracellular functions that require right balance of enzyme activation from precursor CysPs. Intracellularly, they function in apoptosis, antigen and protein processing .Whereas extracellularly, they contribute directly to the degradation of foreign proteins and the extracellular matrix (Dickinson, 2002). The signature ERFNIN motif is thought to be associated mostly with inhibiting protease activity before enzyme activation is required.

Also seen in the amino acid alignment, a conserved catalytic triad comprising of Cys25, His159 and Asn175 residues was found in all the candidate papain family CysPs. Enzymatic activity of CysP is dependent on Cys25 and His159 residues, which exist as ion-pairs in a pH interval of 3.5-8 (Grzonka et al., 2001). The role of Asn175 in the catalytic triad is not understood well, but is thought to be responsible in maintaining structural integrity of the enzyme, as it is conserved across all papain family CysPs (Vernet et al., 1995). Besides catalytic Cys residue, four other Cys residues were also identified to be conserved in the papain domain of all aligned CysPs. These residues are likely involved in disulfide bridge formation, thus in proper folding and activity of the enzyme.

4.3 *CysP6* is involved in reducing IL-10 accumulation in tobacco

Quantitative analysis of IL-10 levels in *CysP6* silenced tobacco lines showed increased accumulation compared to the IL-10 control (Figure 3.6). Some *CysP6* silenced lines showed no change in IL-10 accumulation while others accumulated lower level of IL-10 in comparison to the IL-10 control plant. Thus, the accumulation of recombinant protein varied across different *CysP6* silenced lines. The variation in the IL-10 levels

could have resulted from a position effect or differential levels of *CysP6* gene silencing (also seen in *CysP6* expression levels in different *CysP6* silenced T_1 lines, Figure 3.14). Such a position effect has been explained by researchers in several transgenic plants and animals (Matzke and Matzke, 1998).

Silencing indigenous tobacco putative *CysPs*; *CysP1*, *CysP2*, *CysP3*, *CysP4*, *CysP5*, *CysP7*, *CysP8*, *CysP9* and *CysP10* did not increase IL-10 accumulation (Figure 3.7). Surprisingly, in all of these *CysP* silenced T_0 independent transgenic lines, the level of IL-10 accumulation remained lower in comparison to that of the IL-10 control plant. Since silencing these *CysPs* didn't result in an increase in IL-10 accumulation, no further characterization was done with these respective transgenic lines. However, performing a detailed transcript analysis of all candidate CysPs can be done in-order to know if and to what extent these *CysPs* are silenced, and how does silencing relate to IL-10 accumulation seen in the individual silenced lines. More work is required to reach conclusions regarding their role in recombinant protein accumulation in tobacco.

4.4 IL-10 accumulation in tobacco increases with plant maturity

Inherent factors affecting synthesis and accumulation of IL-10 have been characterized in detail in *Arabidopsis* (Chen et al., 2013). To study if physical parameters such as age and height of the plant affect IL-10 accumulation, IL-10 content was measured at two different time points (Figure 3.8). In 7 week-old tobacco, the plants accumulated a higher amount of IL-10 in all the transgenic lines in comparison to the 4 weeks old plants, which suggests that IL-10 production increases with plant maturity.

For the production of IL-10 in tobacco, the recombinant protein requires harvesting before the flowering stage of the plant. Therefore, IL-10 levels in plants older than 7 weeks were not determined. Moreover, after flowering the expression of senescence associated genes could pose a problem in recombinant protein accumulation, as protease activities are high in senescessing organs. Higher protease activity during leaf senescence has been reported in many plants including tobacco and CysPs are known to be responsible for it (Ueda et al., 2000; Beyene et al., 2006).

To minimize the possible error that can arise due to differences in transgene expression between young and mature leaves of the same plant, IL-10 level for each line was averaged from fully expanded 4rd, 5th and 6th leaves from the top, thus excluding variability caused during sampling, if any.

4.5 Transient knockdown of *CysP6* in tobacco leaf

Identification of *CysP6* as a potential candidate involved in affecting recombinant protein accumulation is a first reported case in tobacco. To study further the effects of silencing *CysP6* on the accumulation of *CysP6* transcript and subsequent IL-10 levels in tobacco, a transient experiment was performed. Transient expression of the *CysP6* RNAi construct resulted in near to invisible levels of *CysP6* transcript and significantly increased the accumulation of IL-10 in tobacco (Figures 3.9B and D). This also confirmed the putative role of *CysP6* in determining the yield of foreign protein, which was also seen in stable silenced lines.

With the post-genomic era and a flooding of gene sequences, several ways have been explored to assign putative gene functions. Because of its efficiency and specificity, RNAi is one of the widely used techniques in determining gene function. Most of the RNAi approaches utilize generation of stable transgenic lines and observation of RNAi phenotypes related to the particular genes. However, generation of stable lines usually requires a long time, which may range from 2-6 months depending on the type of plants used (4 months in tobacco, see Figure 3.4). Using *Agrobacterium*-mediated transient expression of the silencing construct, gene functions can be determined in a short period of time. As seen in this study, transient silencing of *CysP6* could be achieved in nearly a week, which is very short compared to the time required to generate stable transgenic tobacco lines. Transient silencing of indigenous genes via agro-infiltration had been reported so far in plant species such as *Vitis vinifera* (Bertazzon et al., 2012), but not in tobacco. Particularly with the results seen here, transient RNAi can be efficiently utilized for gene functional assays in tobacco.

4.6 CysP6 localizes to the ER

CysP6 fused to YFP showed that it is localized to the ER. It was interesting to see that CysP6 is localized in the ER, without the presence of C-terminal ER retrieval signal sequence KDEL or HDEL in its C-terminal end (see Figure 3.2 for CysP6 sequence). However, CysP6 consists of a signal peptide with predicted cleavable position between 26^{th} (Ala) and 27^{th} (Ser) residues in its amino acid sequence. Localization of the proteins with the signal peptide can possibly be in the ER lumen, ER membrane, tonoplast, cell membrane or extracellular space and are collectively processed through the secretory pathway involving ER as the initial component. Several ER resident proteins that lack the retention signal are known to localize in the ER in a signal-independent fashion (Gu et al., 2012). *Arabidopsis* RD21 is one of such protein without KDEL sequence, and is localized to the vacuole and ER in response to osmotic stress and wounding (Gu et al., 2012). Some of these proteins retained in the ER also play a role in quality control through an interaction with the ER chaperones, whose function is to retain and subsequently degrade assembly defective proteins (Hammond and Helenius, 1995).

Accumulation of proteins targeted to the ER has been reported to be affected by plant proteases (Stevens et al., 2000). It was shown that IgG1 targeted to accumulate in the ER is possibly degraded by acidic proteases, most of which consisted of senescence associated CysPs (Stevens et al., 2000). A study by Sharp and Doran (2001) indicated that initiation of the antibody IgG1 degradation in a transgenic tobacco cell suspension culture occurs along the secretory pathway outside the ER (Sharp and Doran, 2001). Localization of CysP6 to the ER provides greater possibility of its involvement in IL-10 degradation, since IL-10 accumulates in the ER. The degradation process might be occurring along the secretory route when IL-10 is being processed or accumulated. However, more study should be conducted regarding the fate of IL-10, to prove the speculation that IL-10 degradation occurs along secretory pathway, thus affecting its overall yield in tobacco plant.

4.7 *CysP6* silenced lines show reduced *CysP6* expression in T_1 generation

CysP6 expression was measured in four *CysP6* silenced T_1 lines (2, 8, 9 and 15) by qRT-PCR and it was found that the *CysP6* transcript level was lower in all four lines in comparison to the IL-10 control plant (Figure 3.14). The normalized fold expression of *CysP6* varied among the lines, suggesting that the degree of *CysP6* silencing was different in each line. The differential level of silencing is a common character of RNAi-mediated gene silencing in plants (Kerschen et al., 2004). Multiple genes were silenced in *Arabidopsis* using different RNAi constructs, which showed significant transcript variability between the independent RNAi lines of same target gene (Kerschen et al., 2004). In tobacco, the degree of gene silencing was analyzed in three independent lines carrying 35S- β-*Glucorinidase* (*GUS*) transgenes, and it was found that the amount of antisense *GUS* RNA co-related with the extent of post-transcriptional silencing in each line (Hamilton and Baulcombe, 1999). Differential silencing was also observed in the expression of rice *CysP* (*OSEP3A*, encodes REP1), as small interfering (si) RNA was targeted to silence *OSEP3A* in transgenic rice cell lines expressing recombinant hGM-CSF (Kim et al., 2008).

The expression of siRNA and the silencing of a target gene depend on several factors, collectively referred to as the "position effect" (Matzke and Matzke, 1998). Such an effect is caused by the telomeric (stable expression) or centromeric (mosaic expression) integration of transgenes in the chromosome and transfer of the binary vector sequences that reduces transgene expression (Matzke and Matzke, 1998). Differential silencing of *CysP6* seen in the T_1 tobacco plants, which were obtained from the T_0 independent transgenic lines, might also be a result of the position effects.

A comparison of *CysP6* expression and IL-10 accumulations in the *CysP6* silenced T_1 lines showed no correlation. In the plants with lower *CysP6* expression (line 8 and 9), IL-10 accumulation was higher in comparison to the controls. Two of the *CysP6* silenced lines showed higher error in expression levels of *CysP6* (line 9 and 15), thus their relation to corresponding levels of IL-10 is not conclusive. A higher number of stable lines and biological replicates should be included to draw a conclusion regarding a

correlation, which might exist between the degree of *CysP6* silencing and resulting IL-10 accumulation in tobacco.

Chapter 5: Perspectives and Future work

Plants have emerged as major choice in producing biopharmaceutical proteins. High production costs hamper other production systems such as microbial fermentation or animal-based systems. Using plants, higher yields and economic production of recombinant proteins can be achieved by creating stable transgenic lines that can simply be propagated using agriculture.

Identification of the members of the tobacco *CysP* family is the first step in exploring the potential roles of many *CysPs* in plant growth and cellular responses to environmental stimuli, including foreign protein antigens. With at least 60 putative *CysPs*, the tobacco *CysP* gene family is possibly one of the largest plant *CysP* families. As plant CysPs are reported to be involved in the degradation of many heterologous proteins, it is necessary to understand the role of tobacco CysPs in IL-10 proteolysis. Identifying a CysP involved in the degradative process is a major step towards alleviating protease effects and improving protein yields in tobacco. Out of 10 candidate CysPs chosen for study, CysP6 was found to be potentially involved in IL-10 proteolysis. Stable silencing of *CysP6* resulted in an increase in IL-10 accumulations over two generations of tobacco. The potential link of *CysP6* to IL-10 degradation was also established by transiently silencing *CysP6*. Further characterization of CysP6 is necessary to confirm a direct role in IL-10 degradation, which can be done by using activity-based protein profiling (ABPP) techniques (van der Hoorn et al., 2004; Cravatt et al., 2008). More evidences can be generated by doing CysP activity assay and it can help to verify that functional protease activity is lower in *CysP6* silenced lines in comparison to the IL-10 control plants. Confirmation of *CysP6* roles could open opportunities such as creating *CysP6* depleted tobacco lines that can be utilized to study the accumulation of other high value therapeutic proteins.

As shown in Figure 3.14, IL-10 accumulation in the T_1 generation of *CysP6* silenced lines and a differential *CysP6* expression in their respective biological replicates indicated the inconclusive correlation between degree of *CysP6* silencing and resulting IL-10 accumulation. A higher number of stable transgenic lines and biological replicates

should be analyzed to study correlation between degree of *CysP6* silencing and IL-10 accumulation. Moreover, even though the silencing construct was targeted to the unique region in *CysP6*, it is reasonable to study if silencing is specific to *CysP6* only. The study of silencing specificity can be done by a transcript level analysis of all other candidate *CysPs* in the *CysP6* silenced lines. It will be helpful to know if *CysPs* other than the *CysP6* have a potential impact in IL-10 accumulation.

Subcellular localization of CysP6 in the ER, even without presence of an ER retention signal, could be due to presence of a signal peptide which directs proteins to the secretory pathway and is functionally required to perform roles similar to the ER resident proteases. ER resident proteins are involved in degradation of proteins and some of them lack an ER retention signal (Gu et al., 2012). Based on CysP6 localization study, it is plausible that ER could potentially be the site of IL-10 proteolysis as IL-10 is also targeted to accumulate in the ER. To confirm that IL-10 is degraded in the ER, subcellular fractioning can be done followed by immunoblotting with anti-IL-10 antibodies.

References

Aquea, F., Gutierrez, F., Medina, C., and Arce-Johnson, P. (2008). A novel otubain-like *cysteine protease* gene is preferentially expressed during somatic embryogenesis in *Pinus radiata*. Mol Biol Rep **35,** 567-573.

Asadullah, K., Sterry, W., and Volk, H.D. (2003). Interleukin-10 therapy-review of a new approach. Pharmacol Rev **55,** 241-269.

Avesani, L., Marconi, G., Morandini, F., Albertini, E., Bruschetta, M., Bortesi, L., Pezzotti, M., and Porceddu, A. (2007). Stability of Potato virus X expression vectors is related to insert size: implications for replication models and risk assessment. Transgenic Res **16,** 587-597.

Bakker, H., Bardor, M., Molthoff, J.W., Gomord, V., Elbers, I., Stevens, L.H., Jordi, W., Lommen, A., Faye, L., Lerouge, P., and Bosch, D. (2001). Galactose-extended glycans of antibodies produced by transgenic plants. Proc Natl Acad Sci U S A **98,** 2899-2904.

Balakirev, M.Y., Tcherniuk, S.O., Jaquinod, M., and Chroboczek, J. (2003). Otubains: a new family of cysteine proteases in the ubiquitin pathway. EMBO Rep **4,** 517-522.

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M.A., and Matzke, A.J. (1986). The expression of a *nopaline synthase* - human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. Plant Mol Biol **6,** 347-357.

Beers, E.P., Jones, A.M., and Dickerman, A.W. (2004). The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. Phytochemistry **65,** 43-58.

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

Bernoux, M., Timmers, T., Jauneau, A., Briere, C., de Wit, P.J., Marco, Y., and Deslandes, L. (2008). RD19, an *Arabidopsis* cysteine protease required for RRS1-R-mediated resistance, is relocalized to the nucleus by the *Ralstonia solanacearum* PopP2 effector. Plant Cell **20,** 2252-2264.

Bertazzon, N., Raiola, A., Castiglioni, C., Gardiman, M., Angelini, E., Borgo, M., and Ferrari, S. (2012). Transient silencing of the grapevine gene *VvPGIP1* by agroinfiltration with a construct for RNA interference. Plant Cell Rep **31,** 133-143.

Beyene, G., Foyer, C.H., and Kunert, K.J. (2006). Two new cysteine proteinases with specific expression patterns in mature and senescent tobacco (*Nicotiana tabacum L.*) leaves. J Exp Bot **57,** 1431-1443.

Bock, R., and Warzecha, H. (2010). Solar-powered factories for new vaccines and antibiotics. Trends Biotechnol **28,** 246-252.

Bouquin, T., Thomsen, M., Nielsen, L.K., Green, T.H., Mundy, J., and Dziegiel, M.H. (2002). Human anti-rhesus D IgG1 antibody produced in transgenic plants. Transgenic Res **11,** 115-122.

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 **72,** 248-254.

Cabanes-Macheteau, M., Fitchette-Laine, A.C., Loutelier-Bourhis, C., Lange, C., Vine, N.D., Ma, J.K., Lerouge, P., and Faye, L. (1999). N-Glycosylation of a mouse IgG expressed in transgenic tobacco plants. Glycobiology **9,** 365-372.

Chen, L., Dempsey, B.R., Gyenis, L., Menassa, R., Brandle, J.E., and Dhaubhadel, S. (2013). Identification of the factors that control synthesis and accumulation of a therapeutic protein, human immune-regulatory interleukin-10, in *Arabidopsis thaliana*. Plant Biotechnol J **11,** 546-554.

Christensen, A.H., and Quail, P.H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res **5,** 213-218.

Christou, P. (1995). Particle bombardment. Methods Cell Biol **50,** 375-382.

Cravatt, B.F., Wright, A.T., and Kozarich, J.W. (2008). Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. Annu Rev Biochem **77,** 383-414.

D'Aoust, M.A., Lavoie, P.O., Couture, M.M., Trepanier, S., Guay, J.M., Dargis, M., Mongrand, S., Landry, N., Ward, B.J., and Vezina, L.P. (2008). Influenza virus-like particles produced by transient expression in *Nicotiana benthamiana* induce a protective immune response against a lethal viral challenge in mice. Plant Biotechnol J **6,** 930-940.

Daniell, H., Streatfield, S.J., and Wycoff, K. (2001). Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. Trends Plant Sci **6,** 219-226.

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*. Transgenic Res **2,** 227-237.

Decker, E.L., and Reski, R. (2008). Current achievements in the production of complex biopharmaceuticals with moss bioreactors. Bioproc Biosyst Eng **31,** 3-9.

Dickinson, D.P. (2002). Cysteine peptidases of mammals: their biological roles and potential effects in the oral cavity and other tissues in health and disease. Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists **13,** 238-275.

Doran, P.M. (2006). Foreign protein degradation and instability in plants and plant tissue cultures. Trends in Biotechnol **24,** 426-432.

Durocher, Y., and Butler, M. (2009). Expression systems for therapeutic glycoprotein production. Curr Opin Biotechnol **20,** 700-707.

Evangelista, R.L., Kusnadi, A.R., Howard, J.A., and Nikolov, Z.L. (1998). Process and economic evaluation of the extraction and purification of recombinant beta-glucuronidase from transgenic corn. Biotechnol Prog **14,** 607-614.

Faye, L., Boulaflous, A., Benchabane, M., Gomord, V., and Michaud, D. (2005). Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. Vaccine **23,** 1770-1778.

Gan, S., and Amasino, R.M. (1995). Inhibition of leaf senescence by autoregulated production of cytokinin. Science **270,** 1986-1988.

Garcia-Lorenzo, M., Sjodin, A., Jansson, S., and Funk, C. (2006). Protease gene families in *Populus* and *Arabidopsis*. BMC Plant Biol **6,** 30.

Ge, X., Dietrich, C., Matsuno, M., Li, G., Berg, H., and Xia, Y. (2005). An *Arabidopsis* aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. EMBO Rep **6,** 282-288.

Gietl, C., and Schmid, M. (2001). Ricinosomes: an organelle for developmentally regulated programmed cell death in senescing plant tissues. Naturwissenschaften **88,** 49-58.

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

Grudkowska, M., and Zagdanska, B. (2004). Multifunctional role of plant cysteine proteinases. Acta Biochim Pol **51,** 609-624.

Grzonka, Z., Jankowska, E., Kasprzykowski, F., Kasprzykowska, R., Lankiewicz, L., Wiczk, W., Wieczerzak, E., Ciarkowski, J., Drabik, P., Janowski, R.,

Kozak, M., Jaskolski, M., and Grubb, A. (2001). Structural studies of cysteine proteases and their inhibitors. Acta Biochim Pol **48,** 1-20.

Gu, C., Shabab, M., Strasser, R., Wolters, P.J., Shindo, T., Niemer, M., Kaschani, F., Mach, L., and van der Hoorn, R.A. (2012). Post-translational regulation and trafficking of the granulin-containing protease RD21 of *Arabidopsis thaliana*. PLoS One **7,** e32422.

Gutierrez, S.P., Saberianfar, R., Kohalmi, S.E., and Menassa, R. (2013). Protein body formation in stable transgenic tobacco expressing elastin-like polypeptide and hydrophobin fusion proteins. BMC Biotechnol **13,** 40.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286,** 950-952.

Hamilton, S.R., Bobrowicz, P., Bobrowicz, B., Davidson, R.C., Li, H., Mitchell, T., Nett, J.H., Rausch, S., Stadheim, T.A., Wischnewski, H., Wildt, S., and Gerngross, T.U. (2003). Production of complex human glycoproteins in yeast. Science **301,** 1244-1246.

Hammond, C., and Helenius, A. (1995). Quality control in the secretory pathway. Curr Opin Cell Biol **7,** 523-529.

He, Y., Ning, T., Xie, T., Qiu, Q., Zhang, L., Sun, Y., Jiang, D., Fu, K., Yin, F., Zhang, W., Shen, L., Wang, H., Li, J., Lin, Q., Sun, Y., Li, H., Zhu, Y., and Yang, D. (2011). Large-scale production of functional human serum albumin from transgenic rice seeds. Proc Natl Acad Sci U S A **108,** 19078-19083.

Hehle, V.K., Paul, M.J., Drake, P.M., Ma, J.K., and van Dolleweerd, C.J. (2011). Antibody degradation in tobacco plants: a predominantly apoplastic process. BMC Biotechnol **11,** 128.

Ho, S.L., Tong, W.F., and Yu, S.M. (2000). Multiple mode regulation of a *cysteine proteinase* gene expression in rice. Plant Physiol **122,** 57-66.

Houdebine, L.M. (2009). Production of pharmaceutical proteins by transgenic animals. Comp Immunol Microbiol Infect Dis **32,** 107-121.

Jobling, S.A., Jarman, C., Teh, M.M., Holmberg, N., Blake, C., and Verhoeyen, M.E. (2003). Immunomodulation of enzyme function in plants by single-domain antibody fragments. Nat Biotechnol **21,** 77-80.

Kaldis, A., Ahmad, A., Reid, A., McGarvey, B., Brandle, J., Ma, S., Jevnikar, A., Kohalmi, S.E., and Menassa, R. (2013). High-level production of human interleukin-10 fusions in tobacco cell suspension cultures. Plant Biotechnol J **11**, 535-545.

Karrer, K.M., Peiffer, S.L., and DiTomas, M.E. (1993). Two distinct gene subfamilies within the family of *cysteine protease* genes. Proc Natl Acad Sci U S A **90,** 3063-3067.

Kato, H., and Minamikawa, T. (1996). Identification and characterization of a rice cysteine endopeptidase that digests glutelin. Eur J Biochem / FEBS **239,** 310-316.

Kerschen, A., Napoli, C.A., Jorgensen, R.A., and Muller, A.E. (2004). Effectiveness of RNA interference in transgenic plants. FEBS Lett **566,** 223-228.

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.

Kim, N.S., Kim, T.G., Kim, O.H., Ko, E.M., Jang, Y.S., Jung, E.S., Kwon, T.H., and Yang, M.S. (2008). Improvement of recombinant hGM-CSF production by suppression of *cysteine proteinase* gene expression using RNA interference in a transgenic rice culture. Plant Mol Biol **68,** 263-275.

Kohl, T., Hitzeroth, II, Stewart, D., Varsani, A., Govan, V.A., Christensen, N.D., Williamson, A.L., and Rybicki, E.P. (2006). Plant-produced cottontail rabbit

papillomavirus L1 protein protects against tumor challenge: a proof-of-concept study. Clinical and vaccine immunology : CVI **13,** 845-853.

Koizumi, M., Yamaguchi-Shinozaki, K., Tsuji, H., and Shinozaki, K. (1993). Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*. Gene **129,** 175-182.

Krichevsky, A., Zaltsman, A., Lacroix, B., and Citovsky, V. (2011). Involvement of KDM1C histone demethylase-OTLD1 otubain-like histone deubiquitinase complexes in plant gene repression. Proc Natl Acad Sci U S A **108,** 11157-11162.

Kusnadi, A.R., Nikolov, Z.L., and Howard, J.A. (1997). Production of recombinant proteins in transgenic plants: Practical considerations. Biotechnol Bioeng **56,** 473-484.

Lee, S., Jung, K.H., An, G., and Chung, Y.Y. (2004). Isolation and characterization of a rice cysteine protease gene, *OsCP1*, using T-DNA gene-trap system. Plant Mol Biol **54,** 755-765.

Lehfeldt, C., Shirley, A.M., Meyer, K., Ruegger, M.O., Cusumano, J.C., Viitanen, P.V., Strack, D., and Chapple, C. (2000). Cloning of the *SNG1* gene of *Arabidopsis* reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. Plant Cell **12,** 1295-1306.

Leitch, I.J., Hanson, L., Lim, K.Y., Kovarik, A., Chase, M.W., Clarkson, J.J., and Leitch, A.R. (2008). The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Ann Bot **101,** 805-814.

Lienard, D., Sourrouille, C., Gomord, V., and Faye, L. (2007). Pharming and transgenic plants. Biotechnol Ann Rev **13,** 115-147.

Ling, J.Q., Kojima, T., Shiraiwa, M., and Takahara, H. (2003). Cloning of two cysteine proteinase genes, *CysP1* and *CysP2*, from soybean cotyledons by cDNA representational difference analysis. Biochim Biophys Acta **1627,** 129-139.

Linthorst, H.J., van der Does, C., Brederode, F.T., and Bol, J.F. (1993). Circadian expression and induction by wounding of tobacco genes for cysteine proteinase. Plant Mol Biol **21,** 685-694.

Lombardi, R., Circelli, P., Villani, M.E., Buriani, G., Nardi, L., Coppola, V., Bianco, L., Benvenuto, E., Donini, M., and Marusic, C. (2009). High-level HIV-1 Nef transient expression in Nicotiana benthamiana using the P19 gene silencing suppressor protein of Artichoke Mottled Crinckle Virus. BMC Biotechnol **9,** 96.

Ma, J.K., Drake, P.M., and Christou, P. (2003). The production of recombinant pharmaceutical proteins in plants. Nat Rev Genet **4,** 794-805.

Matzke, A.J., and Matzke, M.A. (1998). Position effects and epigenetic silencing of plant transgenes. Curr Opin Plant Biol **1,** 142-148.

McCormick, S., Niedermeyer, J., Fry, J., Barnason, A., Horsch, R., and Fraley, R. (1986). Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*. Plant Cell Rep **5,** 81-84.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature **431,** 343-349.

Menassa, R., Nguyen, V., Jevnikar, A., and Brandle, J. (2001). A self-contained system for the field production of plant recombinant interleukin-10. Molecular Breeding **8,** 177-185.

Menassa, R., Du, C., Yin, Z.Q., Ma, S., Poussier, P., Brandle, J., and Jevnikar, A.M. (2007). Therapeutic effectiveness of orally administered transgenic low-alkaloid tobacco expressing human interleukin-10 in a mouse model of colitis. Plant Biotechnol J **5,** 50-59.

Nelson, B.K., Cai, X., and Nebenfuhr, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. Plant J **51,** 1126-1136.

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.

Paccalet, T., Bardor, M., Rihouey, C., Delmas, F., Chevalier, C., D'Aoust, M.A., Faye, L., Vezina, L., Gomord, V., and Lerouge, P. (2007). Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthesizing enzymes. Plant Biotechnol J **5,** 16-25.

Palomares, L.A., Estrada-Mondaca, S., and Ramirez, O.T. (2004). Production of recombinant proteins: challenges and solutions. Methods Mol Biol **267,** 15-52.

Pan, L., Zhang, Y., Wang, Y., Wang, B., Wang, W., Fang, Y., Jiang, S., Lv, J., Wang, W., Sun, Y., and Xie, Q. (2008). Foliar extracts from transgenic tomato plants expressing the structural polyprotein, P1-2A, and protease, 3C, from foot-and-mouth disease virus elicit a protective response in guinea pigs. Vet Immunol Immunopathol **121,** 83-90.

Russell, D.A. (1999). Feasibility of antibody production in plants for human therapeutic use. Curr Top Microbiol Immunol **240,** 119-138.

Russell, D.A., Spatola, L.A., Dian, T., Paradkar, V.M., Dufield, D.R., Carroll, J.A., and Schlittler, M.R. (2005). Host limits to accurate human growth hormone production in multiple plant systems. Biotechnol Bioeng **89,** 775-782.

Santi, L., Giritch, A., Roy, C.J., Marillonnet, S., Klimyuk, V., Gleba, Y., Webb, R., Arntzen, C.J., and Mason, H.S. (2006). Protection conferred by recombinant Yersinia pestis antigens produced by a rapid and highly scalable plant expression system. Proc Natl Acad Sci U S A **103,** 861-866.

Schaffer, M.A., and Fischer, R.L. (1988). Analysis of mRNAs that accumulate in response to low temperature identifies a *thiol protease* gene in tomato. Plant Physiol **87,** 431-436.

Schechter, I., and Berger, A. (2012). On the size of the active site in proteases. I. Papain. 1967. Biochem Biophys Res Commun **425,** 497-502.

Scheller, J., Guhrs, K.H., Grosse, F., and Conrad, U. (2001). Production of spider silk proteins in tobacco and potato. Nat Biotechnol **19,** 573-577.

Schiermeyer, A., Schinkel, H., Apel, S., Fischer, R., and Schillberg, S. (2005). Production of *Desmodus rotundus* salivary plasminogen activator alpha1 (DSPAalpha1) in tobacco is hampered by proteolysis. Biotechnol Bioeng **89,** 848-858.

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

Shenai, B.R., Sijwali, P.S., Singh, A., and Rosenthal, P.J. (2000). Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of Plasmodium falciparum. J Biol Chem **275,** 29000-29010.

Sourrouille, C., Marquet-Blouin, E., D'Aoust, M.A., Kiefer-Meyer, M.C., Seveno, M., Pagny-Salehabadi, S., Bardor, M., Durambur, G., Lerouge, P., Vezina, L., and Gomord, V. (2008). Down-regulated expression of plant-specific glycoepitopes in alfalfa. Plant Biotechnol J **6,** 702-721.

Staub, J.M., Garcia, B., Graves, J., Hajdukiewicz, P.T., Hunter, P., Nehra, N., Paradkar, V., Schlittler, M., Carroll, J.A., Spatola, L., Ward, D., Ye, G., and Russell, D.A. (2000). High-yield production of a human therapeutic protein in tobacco chloroplasts. Nat Biotechnol **18,** 333-338.

Stevens, L.H., Stoopen, G.M., Elbers, I.J., Molthoff, J.W., Bakker, H.A., 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.

Stewart, C.N., Jr., and Via, L.E. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. BioTechniques **14,** 748-750.

Strasser, R., Stadlmann, J., Schahs, M., Stiegler, G., Quendler, H., Mach, L., Glossl, J., Weterings, K., Pabst, M., and Steinkellner, H. (2008). Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. Plant Biotechnol J **6,** 392-402.

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.

Tanaka, H., Onouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., Machida, C., and Machida, Y. (2001). A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants. Development **128,** 4681-4689.

Tolkatchev, D., Xu, P., and Ni, F. (2001). A peptide derived from the C-terminal part of a plant cysteine protease folds into a stack of two beta-hairpins, a scaffold present in the emerging family of granulin-like growth factors. Peptide Res **57,** 227-233.

Turk, D., Guncar, G., Podobnik, M., and Turk, B. (1998). Revised definition of substrate binding sites of papain-like cysteine proteases. Biol Chem **379,** 137-147.

Twyman, R.M., Stoger, E., Schillberg, S., Christou, P., and Fischer, R. (2003). Molecular farming in plants: host systems and expression technology. Trends Biotechnol **21,** 570-578.

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.

Vain, P., Finer, K.R., Engler, D.E., Pratt, R.C., and Finer, J.J. (1996). Intron-mediated enhancement of gene expression in maize (*Zea mays L.*) and bluegrass (*Poa pratensis L.*). Plant Cell Rep **15,** 489-494.

van der Hoorn, R.A. (2008). Plant proteases: from phenotypes to molecular mechanisms. Annu Rev Plant Biol **59,** 191-223.

van der Hoorn, R.A., Leeuwenburgh, M.A., Bogyo, M., Joosten, M.H., and Peck, S.C. (2004). Activity profiling of papain-like cysteine proteases in plants. Plant Physiol **135,** 1170-1178.

van Engelen, F.A., Schouten, A., Molthoff, J.W., Roosien, J., Salinas, J., Dirkse, W.G., Schots, A., Bakker, J., Gommers, F.J., Jongsma, M.A., and et al. (1994). Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco. Plant Mol Biol **26,** 1701-1710.

Vaquero, C., Sack, M., Chandler, J., Drossard, J., Schuster, F., Monecke, M., Schillberg, S., and Fischer, R. (1999). Transient expression of a tumor-specific single-chain fragment and a chimeric antibody in tobacco leaves. Proc Natl Acad Sci U S A **96,** 11128-11133.

Vernet, T., Tessier, D.C., Chatellier, J., Plouffe, C., Lee, T.S., Thomas, D.Y., Storer, A.C., and Menard, R. (1995). Structural and functional roles of asparagine 175 in the cysteine protease papain. J Biol Chem **270,** 16645-16652.

Vierstra, R.D. (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci **8,** 135-142.

Von Groll, U., Berger, D., and Altmann, T. (2002). The subtilisin-like serine protease SDD1 mediates cell-to-cell signaling during *Arabidopsis* stomatal development. Plant Cell **14,** 1527-1539.

Wang, Y., Zhao, S., Bai, L., Fan, J., and Liu, E. (2013). Expression systems and species used for transgenic animal bioreactors. Biomed Res Int **2013,** 580463.

Werner, S., Breus, O., Symonenko, Y., Marillonnet, S., and Gleba, Y. (2011). High-level recombinant protein expression in transgenic plants by using a double-inducible viral vector. Proc Natl Acad Sci U S A **108,** 14061-14066.

Wiederanders, B., Kaulmann, G., and Schilling, K. (2003). Functions of propeptide parts in cysteine proteases. Current protein & peptide science **4,** 309-326.

Woltering, E.J. (2004). Death proteases come alive. Trends Plant Sci **9,** 469-472.

Xia, Y., Suzuki, H., Borevitz, J., Blount, J., Guo, Z., Patel, K., Dixon, R.A., and Lamb, C. (2004). An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. Embo J **23,** 980-988.

Zhang, X.M., Wang, Y., Lv, X.M., Li, H., Sun, P., Lu, H., and Li, F.L. (2009). NtCP56, a new cysteine protease in *Nicotiana tabacum L.*, involved in pollen grain development. J Exp Bot **60,** 1569-1577.

Zhao, P., Zhou, X.M., Zhang, L.Y., Wang, W., Ma, L.G., Yang, L.B., Peng, X.B., Bozhkov, P.V., and Sun, M.X. (2013). A bipartite molecular module controls cell death activation in the Basal cell lineage of plant embryos. PLoS Biol **11,** e1001655.

CURRICULUM VITAE

Kishor Duwadi

EDUCATION

2012 – Present	M.Sc. candidate (Biology)	The University of Western Ontario London, Ontario, Canada
2006 – 2010	B.Tech. in Biotechnology	Kathmandu University Dhulikhel, Kavre, Nepal

AWARDS AND SCHOLARSHIP

2012 – 2014	Western Graduate Research Scholarship
2013	Best Poster Award, Canadian Society of Plant Biologists, Eastern Regional Meeting, University of Toronto Mississauga
2013	Honorable Mention - Poster Award, Biology Graduates Research Forum, The University of Western Ontario
2010	Korean Government Scholarship

WORK EXPERIENCE

2012 – Present	Graduate Research Assistant, The University of Western Ontario
2012-2014	Graduate Teaching Assistant, The University of Western Ontario
2011-2012	Secretary, Research Institute for Bioscience and Biotechnology, Kathmandu, Nepal
2010	Volunteer, AstraZeneca Pharma India Limited, Bangalore, India

ORAL PRESENTATION

Duwadi, K., Chen L., Menassa, R., and Dhaubhadel, S. (2014). Improvement in recombinant interleukin-10 production by suppression of a *cysteine protease* gene in transgenic tobacco. 9th Canadian Plant Biotechnology Conference, 2014, Montreal, Quebec

POSTER PRESENTATIONS

Duwadi, K., Chen L., Menassa, R., and Dhaubhadel, S. (2013). Improvement in recombinant interleukin-10 production by RNAi silencing of a *cysteine protease* gene in tobacco. Canadian Society of Plant Biologists, Eastern Regional Meeting, Missisauga, Ontario

Duwadi, K., Chen L., Menassa, R., and Dhaubhadel, S. (2013). Identification and characterization of a c*ysteine Protease* gene in tobacco for use in recombinant protein production, Biology Graduates Research Forum (BGRF 2012), London, Ontario

Duwadi, K., Malla, K., Paudel, M., and Lamichhane, J. (2011) Phytochemical and biological screening of high-altitude medicinal plants, *Equisetum diffisum* and *Lagerstroemia parviflora*. International DNA Day, White House Institute of Science and Technology, Kathmandu