Improving Expression Vectors for Recombinant Protein Production in Plants

by

Sun Hee Rosenthal

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2012 by the
Graduate Supervisory Committee:

Hugh Mason, Chair
Tsafrir Mor
Yung Chang
Charles Arntzen

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

Over the past decade, several high-value proteins have been produced using plant-based transient expression systems. However, these studies exposed some limitations that must be overcome to allow plant expression systems to reach their full potential. These limitations are the low level of recombinant protein accumulation achieved in some cases, and lack of efficient co-expression vectors for the production of multi-protein complexes. This study report that tobacco Extensin (Ext) gene 3' untranslated region (UTR) can be broadly used to enhance recombinant protein expression in plants. Extensin is the hydroxyproline-rich glycoprotein that constitutes the major protein component of cell walls. Using transient expression, it was found that the Ext 3' UTR increases recombinant protein expression up to 13.5- and 6-fold in non-replicating and replicating vector systems, respectively, compared to previously established terminators. Enhanced protein accumulation was correlated with increased mRNA levels associated with reduction in read-through transcription. Regions of Ext 3' UTR essential for maximum gene expression included a poly-purine sequence used as a major poly-adenylation site.

Furthermore, modified bean yellow dwarf virus (BeYDV)-based vectors designed to allow co-expression of multiple recombinant genes were constructed and tested for their performance in driving transient expression in plants. Robust co-expression and assembly of heavy and light chains of the anti-Ebola virus monoclonal antibody 6D8, as well as *E. coli* heat-labile toxin (LT) were achieved with the modified vectors. The simultaneous co-expression of three fluoroproteins

using the single replicon, triple cassette is demonstrated by confocal microscopy. In conclusion, this study provides an excellent tool for rapid, cost-effective, large-scale manufacturing of recombinant proteins for use in medicine and industry.

ACKNOWLEDGMENTS

I wish to thank my committee members who made this dissertation possible with their expertise and precious time. A special thanks to Dr. Hugh S. Mason, my committee chair for his vast knowledge, advice and patience. Without his support I could not have come this far. Thank you Dr. Tsafrir Mor, Dr. Yung Chang, and Dr. Charles Arntzen for your generous guidance and support throughout the entire process.

I would like to thank all members of Mason's lab for all their help and support over the years, especially to Sean, Jiman and Dalia. I also would like to acknowledge and thank former graduate students, Emel and Lolita for their friendship and helpful discussions. I must thank Dr. Douglas Daniel for his kind help with the confocal microscope in Chapter 4.

I cannot thank my family enough for their love and support. A special feeling of gratitude goes to my loving parents, Jin Young and Hae Geun, my dear sisters Eun Ju and Ju Yeun. I also owe my deepest gratitude to my brother Jun Sang. I must acknowledge my dearest husband Paul, without whose love, encouragement and sacrifices, I would not have done my Ph.D. I also give special thanks to my wonderful parents in law, Fredric and Nancy, and brother in law Jim for the love and care they have given throughout my doctoral dissertation journey.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

GENERAL INTRODUCTION

1.1 Plant expression system: approaches to enhance expression

Recombinant pharmaceutical proteins such as antibodies and vaccine antigens have many applications in diagnostics and treatment of a wide range of diseases (Paul and Ma, 2011; Obembe *et al*., 2011; Komarova *et al*., 2010). Historically, *Escherichia coli* was the first introduced and most commonly used expression system for the production of recombinant proteins. However, its inability to glycosylate recombinant proteins, required for proper function of many eukaryotic proteins, has encouraged the development of eukaryotic systems. Today, animal and insect cell culture based production systems have replaced *E. coli* for the expression of eukaryotic proteins and have yielded functional products such as monoclonal antibodies, vaccines and other therapeutics (Durocher and Butler, 2009). Nevertheless, the high production costs, in addition to the risk of mammalian pathogens/prions, are some of the drawbacks of the systems.

Plants have emerged as a promising alternative for the production of eukaryotic proteins due to their ability to produce complex and biologically-active proteins (Gomord *et al*., 2005; Boehm, 2007). Some of the advantages of plants over other expression systems include decreased cost, ease of large-scale production, and low risk of contamination with mammalian pathogens. The first generation of the plant-based expression systems relied on transgenic plants that involve stable integration of the gene of interest into the plant genome. Using

transgenic plants, a number of therapeutic proteins and vaccine antigens have been produced. For example, Norwalk virus capsid protein (NVCP) and hepatitis B virus S antigen (HBsAg), produced in transgenic potatos, were assessed in Phase I clinical trials (Tacket *et al*., 2000; Kong *et al*., 2001), while human glucocerebrosidase (prGCD) produced in transgenic carrot cell cultures to combat Gaucher's disease has progressed into Phase III clinical trials (Yusibov, 2011). Nevertheless, the system has major shortcomings; the levels of recombinant protein produced are too low and the time line for the generation of transgenic plants is overly long (>18 months) (Gomord *et al*., 2005).

One alternative strategy to overcome these limitations is *Agrobacterium*-mediated transient expression system. Following leaf infiltration of *Agrobacterium*, T-DNA is delivered into a plant cell by *Agrobacterium* and the DNA molecules remain episomal and can be transcribed and translated transiently. An obvious advantage of the transient expression system is speed, as a target protein can be expressed in a week from the delivery of *Agrobacterium* to plants. Since the first demonstration of β-glucuronidase (GUS) gene expression in phaseolus and tobacco leaves (Kapila *et al*, 1996), various recombinant therapeutic proteins have successfully been produced using this system, including antigens of human immune deficiency virus 1 (HIV-1) and human papillomavirus (HPV), as well as monoclonal antibodies against tumour cells (Komarova *et al*., 2010). As yield of recombinant products is a critical factor for the practical application of the plant expression system, further efforts have been made to boost protein accumulation (Obembe *et al*., 2011; Komarova *et al*., 2010). This chapter

discusses some of the strategies that are being considered for the optimization of plant expression vectors, which is the focus of this dissertation. These strategies include amplifying transgene, boosting mRNA accumulation using strong promoters and terminators, and controlling protein translation and stability.

1.2 Viral vector-based gene amplification

Most gene amplification strategies rely on plant viruses that rapidly replicate during their infection of host cells, resulting in a large number of genome copies (Scholthof *et al*., 1996). Plant virus-based vectors can be delivered into plant cells using the *Agrobacterium*-mediated transient expression method. Once delivered, the recombinant virus harboring the gene of interest rapidly replicates, consequently, generates high level of target mRNA and protein. Thus, this technology has become a fast-growing research area with significant application in biopharmaceutical production (Komarova *et al*., 2010). Some RNA and DNA plant virus based expression systems, including tobacco mosaic virus (TMV), potato virus X (PVX), cowpea mosaic virus (CPMV), and Bean Yellow Dwarf Virus (BeYDV), among others, are discussed in the next section.

1.2.1 Tobacco mosaic virus (TMV)

TMV, *Tobamovirus* genus, is one of the first rod-shaped viruses converted into an expression vector. It has a single stranded positive RNA genome and uses subgenomic promoters to express some of their ORFs. An early attempt was to replace the coat protein (CP) gene of TMV with the chloramphenicol acetyltransferase (CAT) gene (Brisson *et al*., 1984). This vector expressed CAT

but failed to spread efficiently through the plant because the coat protein of TMV is required for efficient viral movement. A duplicated copy of subgenomic coat protein promoter was applied to drive CAT expression as well as CP (Dawson *et al*., 1989). The vector was able to replicate, form subgenomic RNAs, assemble correctly, and produce reporter gene activity. However, the duplicated promoter regions caused instability leading to rapid loss of the CAT gene due to homologous recombination between the promoters. Subsequently, more stable TMV vectors have been developed using a CP promoter derived from a different virus belonging to the Tobamovirus genus. The reduced sequence similarity decreased the chance of homologous recombination and stabilized the vector. Using this system combined with agroinfiltration method, Dorokhov and coworkers (2007) demonstrated the expression of *Mycobacterium tuberculosis* antigen proteins. McCormick *et al* (2008) produced a single-chain Fv epitope vaccine candidate for the treatment of non-Hodgkin's lymphoma. Moreover, the TMV-based expression method further optimized to a 'deconstructed' TMV-based vector—a technology termed 'Magnifection' (Gleba *et al.*, 2005). This system uses various modules that are delivered into plant cells by a mixture of three or more agrobacteria carrying each module. The delivered DNA molecules assemble inside the cell with the help of a site-specific recombinase and transcribe generating a fully functional RNA replicon. Magnifection has provided up to 5 g of reporter protein per kilogram of fresh leaf weight (FLW), with the recombinant protein relative yield being up to 80% of the total soluble protein (Marillonnet *et al*., 2004).

1.2.2 Potato virus X (PVX)

PVX, *Potexviruses* genus, is another rod-shaped virus explored for its potential as an expression vector. Similar to TMV, replacement of the CP gene of PVX with the GUS gene provided high levels of GUS expression but failed to give a systemic virus spread, as the CP is essential for viral spread (Chapman *et al*., 1992). To overcome this, a PVX-based vector using the duplicated promoter of the CP gene was developed, pPVX-201(Chapman *et al*., 1992). This vector gave high levels of reporter gene expression throughout the plant. Nevertheless, the events of inserted gene loss were observed during prolonged infection periods with a positive correlation between the rate of gene loss and the inserted gene size (Avesani *et al*., 2007). Using the PVX vector system, similar to the pPVX-201, different proteins have been produced, including hepatitis B virus nucleocapsid protein, and *M. tuberculosis* ESAT6 antigen (Mechtcheriakova *et al*, 2006; Zelada *et al*., 2006).

1.2.3 Cowpea mosaic virus (CPMV)

CPMV, *Comoviruses genus,* contains a bipartite positive-stranded RNA genome, RNA1 and RNA2. RNA 1 is essential for viral replication, and RNA2 is required for cell-to-cell movement. The first CPMV-based vector was designed to create a fusion protein by inserting target sequences within the viral CP gene (between amino acids 98 and 99) of RNA2 (Porta *et al*., 1994). When the resulting chimaeric CPMV displaying epitope from human rhinovirus 14 (HRV-14) was purified and injected into rabbits, a strong immune response was generated. In another version, a 17-amino acid peptide from VP2 of canine parvovirus (CPV)

was inserted into the CPMV surface protein (between aa22 and 23) of RNA2 (Langeveld *et al*., 2001). Virus particles presenting VP2 epitope of CPV were efficiently expressed and administration of purified chiaemiric virus protected animals from a lethal challenge. Recently, deleted variants of RNA2, used to express whole antibody molecules, have achieved higher expression levels (Sainsbury *et al*., 2008).

1.2.4 Alfalfa mosaic virus (AMV)

AMV, *Alfamovirus* genus, is comprised of a tripartite RNA genome, RNA1, 2, and 3. RNA1 and RNA2 encode proteins required for viral replication (P1 and P2, respectively) while RNA3 encodes the capsid and movement proteins that are required for genome activation and replication. Early efforts have been made to develop an AMV-based expression system for epitope display. Antigenic epitopes (up to 50 aa) have been fused to N-terminus of the CP without interfering with viral assembly. For example, antigenic peptides from rabies virus and from HIV-1 fused to AMV CP were successfully displayed on viral particles; following immunizations of these purified recombinant particles a strong humoral immune response was observed (Yusibov *et al*., 1997). A recent version of the AMV-based expression system consists of three components: one for RNA1 and RNA2, a second for viral CP expression, and a third for RNA3 in which the CP gene is replaced with a foreign gene of interest. Using this system, therapeutic proteins have been expressed, including human growth hormone, hemagglutinin from influenza virus A (H3N2), and anthrax protective antigen fused to lichenase (Green *et al*., 2009).

1.2.5 Bean yellow dwarf virus (BeYDV)

BeYDV, *Mastrevirus* genus of *Geminiviridae*, consists of a circular single-stranded DNA (ssDNA) genome of 2.5 kb (Saunders et al., 1991; Palmer and Rybicki, 1998). The genomic ssDNA replicates following a rolling circle mechanism through a double-stranded (dsDNA) intermediate in the nucleus of host cells. The viral genome contains a long (LIR) and a short (SIR) intergenic region from which viral genes diverge in both the viral and complementary sense (V or C, Fig. 1.1) (Liu *et al.,* 1998). The LIR contains a bi-directional RNA polymerase II-type promoter and a stem-loop structure necessary for the replication of viral DNA. The SIR contains transcription termination and polyadenylation signals and serves as the origin of the complementary strand synthesis. The viral sense genome encodes the movement protein (MP) and the capsid protein (CP), while the complementary sense genome encodes replication associated proteins RepA, and after splicing Rep.

Rolling circle viral replication occurs in three stages (Liu *et al.*, 1998). In the first stage, viral ssDNA is converted into a dsDNA replicative form which serves as the template for viral transcription as well as a template for further replication. In the second stage, an additional dsDNA replicative form is generated. Rep, with nicking-closing activity, nicks the viral strand DNA at a specific sequence (TAATATT/AC) and then covalently binds to the 5’ terminus via a phosphotyrosine linkage. Nascent viral strand DNA is synthesized from the 3'-OH terminus, used as the primer, by host replication proteins. Rep nicks the completed nascent plus strand again and simultaneously ligates it by the closing

activity to create the circular form. In the process, Rep is transferred to the newly created 5' terminus. The last stage is for the accumulation of viral genomes for encapsidation. This stage is similar to the second stage but the priming is prevented, and thus ssDNA is the dominant product.

Several groups have demonstrated the feasibility of BeYDV as an expression vector in plants (Chen *et al*., 2011). The most widely used BeYDV-based vectors were created by replacing the CP and MP gene with an expression cassette driven by the cauliflower mosaic virus (CaMV) 35S promoter. The Rep/RepA is supplied from a separate expression vector to direct amplification of the recombinant replicon. Using this system, Mor *et al* (2003) showed up to 40-fold increase of GUS expression in tobacco cells. This study also showed that co-expression of Rep and RepA is essential for the efficient amplification of the replicon. Authors reasoned this may occur because BeYDV RepA has a similar function, as that of wheat dwarf virus (WDV, Mastrevirus genus), in maintaining the nucleus in S phase to facilitate DNA replication.

One of the first active proteins produced using this vector was a capsid protein of Norwalk virus (NVCP), a causative agent of gastroenteritis (Zhang and Mason, 2006). BeYDV replicon containing an expression cassette for NVCP was used to generate transgenic tobacco NT1 cells and whole potato plants. In this case, Rep/RepA gene under the control of an alcohol inducible promoter was delivered via a separate transgene chromosomal integration. Upon induction with alcohol, a substantial accumulation of replicon DNA was observed, while the

moderate increase of mRNA (~80-fold) and NVCP (~10-fold, up to 1.2% of total soluble protein, TSP) expression was obtained.

Further improvement of the BeYDV vectors employed co-delivery of two *Agrobacterium* strains to *Nicotiana benthamiana* plant leaves: one for a BeYDV-derived vector, containing the gene of interest, and another for the Rep/RepA-supplying vector (Huang *et al*., 2009). This approach resulted in very high levels of DNA amplification and expression of hepatitis B core antigen (HBcAg) at 0.8 mg per g LFW, and NVCP at 0.34 mg per g LFW within five days. Importantly, both of these transiently expressed antigens assembled virus like particles (VLPs) demonstrating the advantages of the BeYDV-derived DNA replicon system in VLP expression in plants. Recently, a single vector BeYDV-based system was developed by including Rep/RepA gene under the control of the relatively weak C-sense LIR promoter (Huang *et al*., 2010). Moreover, the same study demonstrated that the single vector system can contain two tandemly linked replicon cassettes, each containing a distinct gene of interet. The system was effective for the production of oligomeric protein complexes such as monoclonal antibodies (mAbs). Either using two separate single vectors or a dual replicon single vector, encoding both heavy and light chain subunits of an IgG mAb against Ebola virus GP1, a fully assembled and protective IgG mAb was produced at 0.5 mg per g LFW.

Regnard and his coworkers reported a DNA replicon system based on the phenotypically mild strain of the BeYDV, BeYDV-m (Accession number DQ458791, Regnard *et al*., 2010). Quantitative analysis on gene amplification

revealed that the BeYD-m vector was amplified *in planta* by up to two orders of magnitude. The system yielded 50 % more HPV-16 L1 and 3-7-fold more HIV-1 p24 expression compared to the non-replicating vector. The level of mRNA accumulation modulated by the system has not been discussed.

The high-level transgene expression based on the BeYDV relies on the hypothesis that an increase in gene copy number following replication of the viral replicon will in turn result in increased transcription and consequently translation of the recombinant gene. It is significant to note that while100-1000 fold increase of the in gene copy number (as a result of replication) is indeed observed with existing BeYDV-based vectors, these support only a very modest increase in accumulating mRNA and protein products. Thus, there is much more room to improve the system.

1.3 Boosting transcription by using strong promoters

Many efforts were directed at increasing the efficiency of transcription which is often the rate-limiting step of gene expression and ample evidence shows that increased level of transcription frequently yields higher level of recombinant proteins. In order to boost transcription efficiency, several natural promoters from plants and plant pathogens as well as synthetic promoters have been developed. Among others, Cauliflower mosaic virus (CaMV) 35S promoter is the best characterized and widely used promoter (Potenza *et al*., 2004). By stacking multiple copies of enhancer elements from the CaMV 35S, further boosting of expression was reported (Guerineau *et al*., 1992). Several promoters derived from

Agrobacterium tumefaciens are also commonly used such as the nopaline synthase (NOS) promoter (Shaw *et al*. 1984), the octopine synthase (OCS) promoter (Ellis *et al*. 1987) and mannopine synthetase (MAS) promoters (Comai *et al*., 1990). The NOS promoter is shown to be less effective compared to the CaMV 35S promoter (Sanders *et al*., 1987), while the MAS promoter was as efficient as the CaMV 35S promoter (Stefanov *et al*., 1991). Using a synthetic promoter that combines elements of the CaMV 35S promoter and the MAS promoter, several-fold higher expression has been reported (Comai *et al*., 1990). Besides promoters, further increased transcriptional activity has been achieved by including scaffold attachment region (SAR) or matrix attachment regions (MAR) on both sides of the expression cassette. Although no consensus sequences were found, all SARs or MARs are AT-rich DNA sequences. Inclusion of the yeast ARS-1 element that contains a SAR increased GUS expression by ~12-fold (Allen *et al*. 1993). Rb7, MAR from tobacco also showed ~4-fold increase of green fluorescent protein (GFP) expression (Halweg *et al*., 2005). Although molecular mechanisms for transcription activation are currently under investigation, their role in the structure of chromatin and suppression of gene silencing are suggested mechanisms (Abranches *et al*., 2005).

1.4 Selecting suitable terminator

To ensure high-level expression of proteins, the transcribed message should be properly processed. One of the most important steps is the pre-mRNA 3' end processing, whereby RNA transcription is terminated, the primary transcript

is cleaved and adenosine nucleotides are added to form the 3’ poly-adenosine [poly(A)] tail. This process is regulated by interaction between RNA *cis*-elements, located in pre-mRNA 3’ untranslated region (3’ UTR), and pre-mRNA processing complex. The proper 3’ end processing directly impacts mRNA stability, translatability, and nuclear-to-cytoplasmic export (Zhao *et al*., 1999; Moore and Proudfoot, 2009). Recent data also suggested that proper 3’ end processing stimulates transcription initiation by facilitating re-cycling of transcription factors (Mapendano *et al*., 2010). Therefore, it is critical to select an appropriate terminator in order to achieve high level of transgene expression. Several terminators from plant pathogens have been utilized to express recombinant proteins in plants such as the CaMV 35S terminator from CaMV (Pietrzak *et al*., 1986), and the nopaline synthease (NOS) terminator (MacFarlane *et al*., 1992), and the octopine synthase (OCS) teriminator (Ellis *et al*., 1987) from the Ti plasmid of A. tumefaciens. The glutelin GluB-1 terminator from rice (Yang *et al*., 2009) and the heat-shock protein 18.2 (HSP) terminator of *Arabidopsis thaliana* (Nagaya *et al*., 2010) are some of the recently developed terminators showing enhancing effects on protein accumulation compared to the NOS terminator.

1.4.1 Molecular mechanisms of eukaryotic pre-mRNA 3’ end processing

regulation are well reviewed by Millevoi and Vagner (2010). The pre-mRNA 3’end processing proceeds through the recognition of *cis*-acting elements, called the polyadenylation signals (PAS), by pre-mRNA 3’ end processing complex (Fig. 1.3). In mammals, four PAS elements were found: the highly conserved A(A/U)UAAA element located 10-30 nucleotides (nt) upstream of

cleavage/polyadenylation site [poly(A) site]; the U/GU-rich downstream element (DSE) located ~30 nt downstream of the poly(A) site; the U-rich upstream sequence elements (USE) found upstream of A(A/U)UAAA element; and the G-rich auxiliary downstream elements (auxDSE) which reside downstream of DSE (Danckwardt *et al*., 2008). In yeast, less conserved poly(A) signals were described: the A-rich positioning element (PE) that functions similar to the mammalian A(A/U)TAAA element; the UA-rich efficiency element (EE) situated upstream of PE; and the U-rich element located upstream or downstream of the poly(A) site (UUE, upstream U-rich element, and DUE, downstream U-rich element, respectively) (Licatalosi *et al*., 2002). In plants, three major PAS elements, less conserved than those found in mammals, were revealed: the A-rich near upstream elements [NUE, A(A/U)UAAA like element] located in 10-30 nt upstream of poly(A) site; the UG-rich far upstream elements (FUE) lies >50 nt upstream of the poly(A) site; and the cleavage element (CE), composed of cleavage site (CS) surrounded by two ~10 nt U-rich elements. The CS itself was found to reside within tracts of elevated U content (Hunt, 2008)

1.4.2 Pre-mRNA 3’ end processing complex

Extensive network of RNA-protein and protein-protein interactions is required for correct cleavage of the nascent mRNA 3’ end and its polyadenylation. In mammals, the pre-mRNA 3’ end processing complex contains several sub-complexes (Millevoi and Vagner, 2010). These include cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I and II (CF Im, and CF IIm), poly(A) polymerase (PAP),

poly(A)-binding proteins (PABPs), and RNA polymerase II large subunit (RNAP II). CPSF consisting of multi-subunit (CPSF160, 100, 73, 30, symplekin, hFip1, and Wdr33) recognizes the A(A/U)UAAA element and catalyzes cleavage through the interaction with other components of the 3' processing complex. CstF comprised of three subunits (CstF 77, 50 and 64) recognizes the DSE and promotes the efficiency of 3' end processing. Cooperative interactions between the CPSF and CstF are required to define the cleavage and polyadenylation site. Following cleavage, CPSF remains bound to the A(A/U)UAAA element and anchors PAP to the RNA for polyadenylation through direct interactions between CPSF160, hFip1, and PAP. Interactions between CF Im, consist of CF Im25 and CF Im 59 or 68, and RNA is also required for the PAP recruitment. Following cleavage, PAP initiates polyadenylation and the newly synthesized poly(A) tail is coated by PABP, and this interaction stimulates PAP activity. When the poly(A) tail reaches ~250 nt in length, polyadenylation is terminated by PABP-dependent mechanism (Kuhn *et al*., 2009).

In yeast, a similar set of proteins important for 3' end processing were identified: the cleavage and polyadenylation factor (CPF), the cleavage factor IA (CFIA) and the cleavage factor IB (CFIB) (Mandel *et al*., 2008). CPF is similar to mammalian CPSF and binds to sequences near or at the cleavage site (Chanfreau *et al*., 1996). CFIA is homologous to mammalian CF IIm and CstF, and binds to PE and the EE. *In vitro* studies showed that CFIA, CFIB and CFII are required for cleavage, while CPF, CFIA and CFIB are essential for polyadenylation.

Current understandings of mRNA 3' end formation in plants were reviewed by Hunt (2008). Analysis of *Arabidopsis* genome suggested probable plant genes that encode the complement proteins found in the mammalian and yeast polyadenylation apparatus. *Arabidopsis* CPSF complex (AtCPSF) includes AtCPSF30, 73-I, 73-II, 100, 160, AtFIPS5, and AtFY. AtCPSF100 serves as the core element of the AtCPSF complex and AtCPSF30 has RNA binding activity suggesting its role in 3' end processing (Zhao *et al*., 2009). AtFY is also an integral part of the core polyadenylation machinery and plays in alternative polyadenylation. AtCstF is composed of CstF 77 and 64. CstF64 is an RNA-binding protein and interacts with AtCstF77, which also interacts with AtFip1. The AtFip1 interacts with AtPAP and these interactions establish assumptions that AtCstF 64 and 77 are involved in mRNA 3' end formation. While RNA-protein interactions have been demonstrated for several *Arabidopsis* proteins participating in mRNA 3' end processing, the sequence specificity has not been determined. Thus, the association of the plant protein with a specific plant *cis*-element remains to be elucidated.

1.4.3 Aberrant transcription termination and gene silencing

Recent finding that improperly terminated and unpolyadenylated mRNA of sense transgene is targeted by RDR6 (RNA-DEPENDENT RNA POLYMERASE6)-mediated RNA silencing readdresses the importance of using a strong terminator for transgene expression (Luo *et al*., 2007). RNA silencing is an RNA-guided gene regulatory mechanism for genome protection against excessively expressed genes (Baulcombe, 1999). In plants, it also functions as an

antiviral response (Ding *et al*. 2004). Gene silencing is triggered when the type III endoribonuclease Dicer recognizes double-stranded RNA (dsRNA) (Hammond *et al*., 2000; Zamore *et al*., 2000) or imperfectly base-paired single-stranded RNA (ssRNA) (Hutvagner *et al*., 2001). Upon recognition, Dicer cleaves the RNA to 20-25nt, small interfering RNAs (siRNA). The resulting siRNA is incorporated into the RNA-induced silencing complex (RISC), and guides the RISC to target mRNA by siRNA/mRNA base-pairing for degradation (Hammond *et al*., 2001; Martinez *et al*., 2002).

During the expression of heterologous proteins in plants, gene silencing is a potential problem because silencing seems to be positively correlated with the level of expression, i.e. transgenes expressed from strong constitutive promoters are more likely to be silenced. This may because, as Luo suggested, highly transcribed transgenes may generate more aberrant RNAs as collateral products. Aberrantly transcribed RNA, such as improperly terminated and unpolyadenylated mRNA, may act as a template for the RDR6 and induce gene silencing (Luo *et al*., 2007). Thus, selecting appropriate terminator(s) mediating proper mRNA processing can reduce the silencing in some degree and ultimately increase the level of transgene expression. For example, placing tandemly linked 35S and NOS terminators in 3' of GUS gene, reduced GUS-specific small interfering RNA accumulation, and enhanced GUS gene expression. To prevent gene silencing viral suppressors of silencing such as P19 have been used quite effectively (Anandalakshmi *et al*., 1998).

1.5 Boosting protein translation and stability

Several approaches have been followed to increase efficiency of translation and stability of expressed foreign proteins in plants. Among others, these include using translational enhancer, optimizing codon, targeting proteins to sub-cellular compartments, and using fusion partners. Specific nucleotide sequences from single stranded positive-sense plant RNA viruses are often used to enhance the target protein expression. For example, the leader sequence of TMV and PVX enhanced recombinant protein levels by 2-100-fold, depending on target genes and expression systems used (Turnern and Foster, 1995). Recently, the enhancing effect of 5' and 3' UTRs from CPMVP RNA-2A (referred to as hyper translatable deleted RNA-2) was demonstrated (Sainsbury and Lomonossoff, 2008), while the mechanisms by which the enhancement occurs remains to be determined.

Codon-optimization of the target gene has been employed for foreign gene expression in plants. When target genes are originated from other organisms than expression hosts, there are high chances that the foreign genes contain codons that are rarely used in the expression hosts, and this may limit translational efficiency. Target gene codons can be optimized by replacing the rare codons to preferred codons of host plants. When optimizing codons it should be considered that codons within approximately 40 amino acids of the N terminus is especially important for increasing recombinant protein production (Batard *et al*., 2000). In addition, use of localized concentrations of a specific codon should be avoided, as it may cause the shortage of the corresponding tRNA. Extensive predicted mRNA

secondary structures that may hinder translation efficiency should also be excluded. A specific example can be found in the study of recombinant human acetylcholinesterase (AChE) accumulation in plants (Geyer *et al*., 2007). By substituting the rare codons present in the AChE gene to more frequently used codons in the most abundant plant protein, the small subunit of ribulose bisphosphate carboxylase (RuBisCO), a 5-10 fold increase in protein accumulation was achieved. It should be noted that authors also abolished the putative RNA polyadenylation signals and potential methylation signals. A detailed study demonstrated that the enhanced AChE expression was driven by the facilitated translation.

While the cytoplasm has been the most common place of expression, recombinant proteins can be targeted to a specific cellular compartment such as apoplast, plastid, mitochondrion, vacuolar compartment and nucleus by adding signal sequences. The subcellular location of a recombinant protein can greatly influence its accumulation (Chikwamba *et al*., 2003; Conrad and Fiedler, 1998). When selecting target organelles the posttranslational processing of recombinant proteins should be considered. For example, proteins targeted to the endoplasmic reticulum (ER) can be glycosylated. For several proteins, the addition of ER-retention signals (H/KDEL) to C-terminal of target proteins has demonstrated to increase their stability (Schouten *et al*., 1996).

Fusion of a foreign protein with a carrier protein that has been shown to be stably expressed in plants may also increase protein stability. For example, fusion of a foreign protein to the C-terminus of a single ubiquitin coding unit stabilized

expressed recombinant protein and resulted in a 10-fold increased expression level (Garbarino *et al*., 1995). The fusion of a tuberculosis antigen to the receptor binding subunit of the heat-labile toxin of *E. coli* follows this strategy (Rigano *et al*., 2004). In addition, fusion of a foreign protein to a recombinant protein with defined binding characteristics or to an affinity tag can facilitate purification (Airenne *et al*., 1999). However, later cleavage from the protein carrier may be necessary to release the desired product, adding to downstream processing costs.

1.6 Approaches to express heteromeric protein complexes

Expression of heteromeric complexes, consisting of more than two polypeptide chains (i.e. antibodies) in plants is a challenging task. In the transgenic method, two insertion strategies were used; the insertion of a single DNA construct encoding two genes of interest (i.e. heavy and light chain cassettes) or two independent insertions, each encoding one transgene. In the latter, two approaches are possible. First, the generation of independent transgenic plants, each expressing one gene of interest followed by cross-pollination of individuals to obtain plants expressing both genes (De Neve *et al*., 1993). Second, the concomitant co-transformation with both DNA constructs (De Neve *et al*., 1997). Although the formal approach is more time-consuming, better expression is achieved compared to the co-transformation method (Bouquin *et al*., 2002). Recently, expression and assembly of a noncovalently assembled mammalian B-cell receptor, the BCR, in transgenic *Nicotiana tabacum* was reported (Julian Ma, 2010). Plants initially expressing either Iga or Igb were generated, and these were

sexually crossed to produce plants expressing Iga／Igb heterodimer. Finally, these were crossed with plants expressing plasma membrane-retained antibody (mIg) to produce the BCR plants. However, as discussed above, the time necessary to generate the transgenic plants is very long (> 18 momonths) and yields are generally very low (1-40 μg/g LFW) (Giritch *et al*., 2006).

In addition to transgenic plants, viral vector-based transient expression systems were developed to co-express multiple target proteins. The first attempt was made by co-infecting *N. benthamiana* plants with *in vitro* synthesized transcripts of two recombinant tobacco mosaic viruses (TMV) encoding either the heavy or light chain of mAb against a colon cancer antigen (Verch *et al*., 1998). Upon co-infection, genes of heavy and light chains were expressed and assembled into a full-length antibody. However, when two *Agrobacterium* strains harboring TMV-based vectors were co-infiltrated to express two different proteins, an early segregation and spatial separation of the two distinct TMV populations were observed in the infiltrated tissues and, therefore, only 4-5% of infected leaf tissue coexpressed both genes (Giritch *et al*, 2006). This problem has been overcome by using two non-competing viral vectors, TMV and potato virus X (PVX) (Giritch *et al*, 2006). Synchronous coinfection of two virual vectors, each expressing a separate antibody chain effectively coexpressed the heavy and light chains in the same cell, resulting in a yield of up to 0.5g of assembled mAb per kg LFW. While it was successful, this method needed co-infiltration of six different *Agrobacterium* strains.

The CPMV based expression system was also tested (Sainsbury *et al*., 2008). In an early attempt, the RNA-1 vector encoding the replicase was co-infiltrated with two modified RNA-2 vectors encoding either the heavy or light chain. While the infiltrated tissues showed co-expression of both chains, the modified RNA-2 molecules segregated during the systemic spread of CPMV. A deleted version of RNA-2 (delRNA-2), in which the region encoding the movement protein and both coat proteins have been removed, was developed and a higher yield was obtained without segregation that was seen in the full-length versions of RNA-2 (Sainsbury *et al*., 2008).

Although successfully used in IgG expression, the RNA virus based expression systems seems impractical for the efficient co-expression and assembly of multi-protein complexes consisting of more than three subunit proteins that may require at least three compatible modules built in non-competing viruses. Recently, a study demonstrated the unique and advantageous features of the BeYDV-based expression system for the co-expression of multi proteins. When two *Agrobacterum* strains harboring two distinct BeYDV vectors were co-delivered in the same cell, the resulting two separate replicons replicated independently and non-competitively in the same cell. Using either co-delivery of separate vectors or delivery of a single vector that contained tandemly arranged replicons, the heavy and light chains of the anti-Ebola virus mAb 6D8 was expressed and fully assembled at ~0.5 mg of mAb per g LFW within 4 to 8 days post infiltration (DPI) of *N. benthamiana* leaves (Huang *et al*., 2010). This study suggests the BeYDV based expression system as a promising platform for the

production and assembly of multi-protein complexes such as secretory IgA and IgM.

1.7 Dissertation outline

Recombinant protein expression in plants can potentially bring benefits for society. This research focuses on improving expression vectors to increase the efficiency of recombinant protein expression in plants. As a first step, a novel transcription terminator, tobacco Extensin 3' UTR was tested to optimize transient protein expression system in Chapter2. The Extensin 3' UTR was compared with commonly used plant terminators for the expression of GFP and NVCP using a non-replicating vector. Up to 13.5- fold increase in protein accumulation was achieved and the enhancing mechanism was uncovered. In Chapter 3, the BeYDV based vector was optimized by using the Ext 3' UTR. The elevated levels of GFP and NVCP expression were reported and the enhancing mechanism was elaborated. Chapter 4 discusses the creation of a new BeYDV-based vector system that allows a transient co-expression of three distinct proteins from a sigle vector, for the first time. The approach of this invention was not focused on the complex formation of the expressed proteins, but on the construction of the multi-protein expression vector and the successful co-expression of three separate proteins from the same cell.

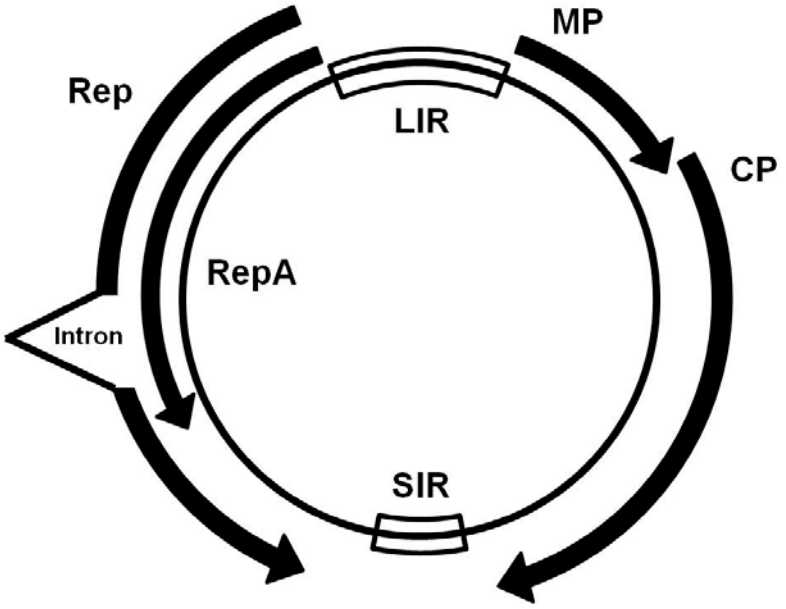


Figure 1.1 Genetic organization of bean yellow dwarf virus. LIR, long intergenic region; SIR, short intergenic region; MP, movement protein; CP, coat protein; Rep, replication initiation protein; RepA, replication associated protein.

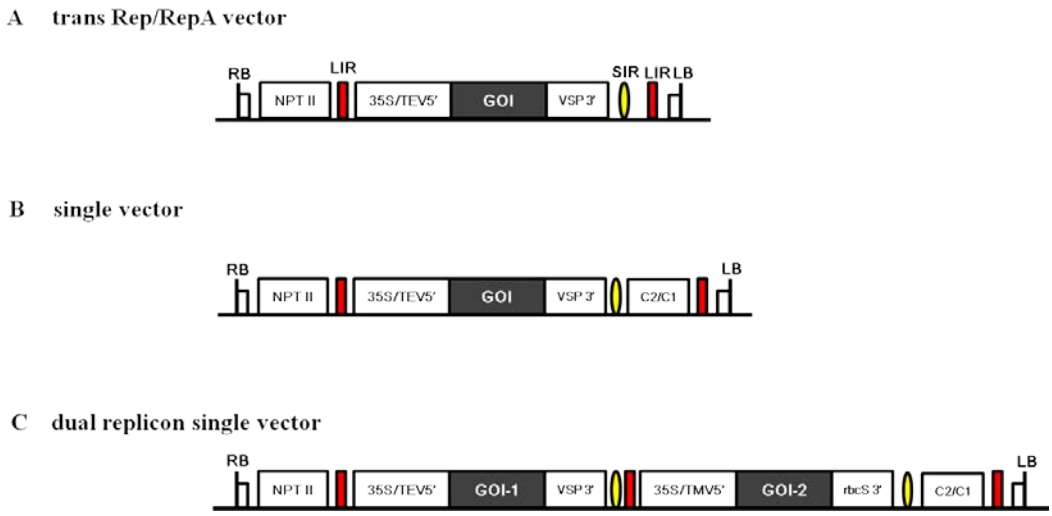


Figure 1.2 Structures of the T-DNA regions of BeYDV-based vectors.
35S/TEV 5', CaMV 35S promoter with tobacco etch virus 5' UTR; GOI, gene of interest; VSP 3', soybean vspB gene 3' element; rbcS 3', pea rbcS gene 3' element; NPT II, expression cassette encoding *nptII* gene for kanamycin resistance; LIR (red box), long intergenic region of BeYDV genome; SIR (yellow oval), short intergenic region of BeYDV genome; C2/C1, BeYDV C1 and C2, encoding replication initiation protein Rep and RepA; LB and RB, the left and right borders of the T-DNA region. In the dual replicon vector, the middle LIR functions in concert with the left-side LIR to release the left-side replicon, and it also functions in concert with the right-side LIR to release the right-side replicon. Adapted from reference, Chen *et al*., 2011.

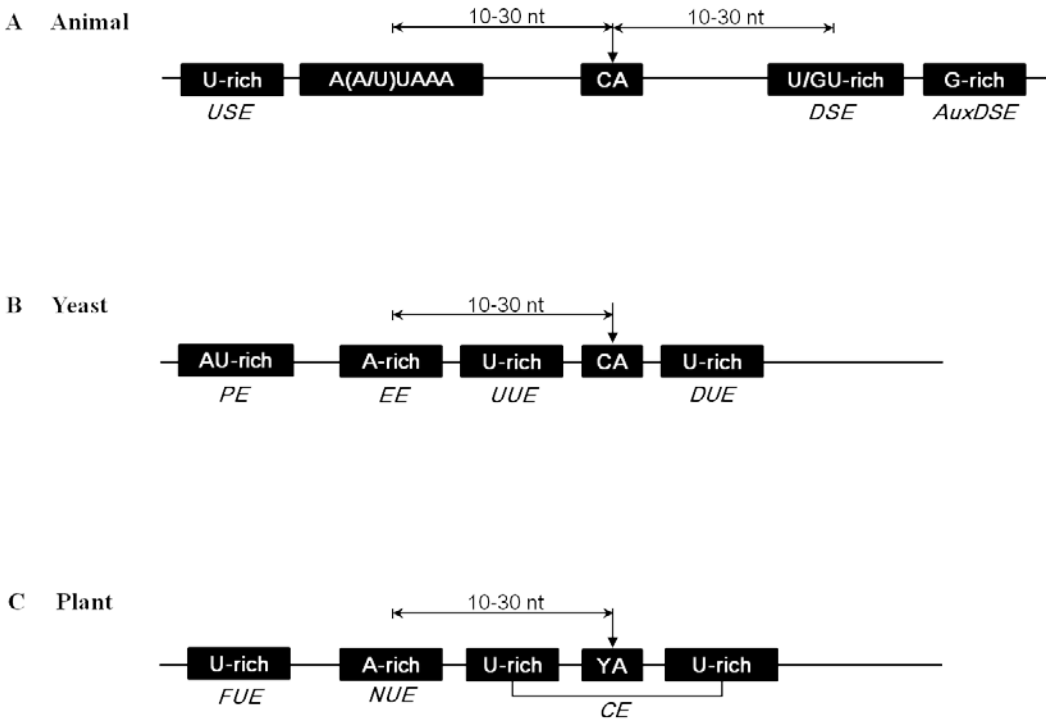


Figure 1.3 Schematic representation of poly(A) signals in animal, yeast, and plant. The sequence elements that comprise the poly(A) signals are indicated by rectangles. USE, upstream element; DSE, downstream element; AuxDSE, auxiliary downstream element; PE, positioning element; EE, efficiency element; UUE, upstream U-rich element; DUE, downstream U-rich element; FUE, far upstream element; NUE, near upstream element; CE, cleavage element. Cleavage sites are indicated with vertical arrows. Adapted from reference, Millevoi and Vanger (2010).

CHAPTER 2

TOBACCO EXTENSIN GENE TERMINATOR INCREASES TRANSGENE EXPRESSION BY IMPROVING TRANSCRIPT 3' END PROCESSING IN *NICOTIANA BENTHAMIANA*.

2.1 Abstract

To express high-value proteins in plants efficiently, the optimization of expression cassettes using appropriate regulatory sequences is critical. This chapter describes the tobacco Extensin (Ext) gene terminator, which increases transient transgene expression up to 13.5-fold compared to previously established terminators in *Nicotiana benthamiana.* Ext is a member of the hydroproline-rich glycoprotein (HRGP) superfamily and constitutes the major protein component of cell walls. The enhanced transgene expression was correlated with increased mRNA accumulation and reduced levels of read-through transcripts, which could impair gene expression. Moreover, this study reveales sequence elements responsible for the enhancing effects, including a polypurine sequence, which served as a dominant cleavage and polyadenylation site. Therefore, the use of Ext terminator may benefit the production of recombinant proteins in plants.

2.2 Introduction

Recombinant eukaryotic proteins produced from heterologous expression systems have been widely used for industrial and pharmaceutical purposes. Expression of eukaryotic proteins in plants has advantages over traditional expression systems, such as mammalian cells and insect cells, including low risk for contamination with animal pathogens, potentially lower costs, and high scalability (Thanavala *et al*., 2006; Koprowski and Yusibov, 2001; Arntzen, 2008). While stable transgenic plants are a well-established system, *Agrobacterium*-mediated transient expression is an attractive alternative because it significantly reduces development and production timelines. Because yield of the recombinant product is a critical factor for the practical application of expression systems, several strategies have been developed to augment expression levels in plants. Using strong promoters, elevating transcript stability, optimizing translation, and targeting the protein to a suitable sub-cellular location are some of the strategies that have been employed (Streatfield, 2007; Sharma and Sharma, 2009). However, another integral parameter is the transcription terminator element.

The terminator regulates gene expression by mediating RNA transcript termination and mRNA 3' end processing, whereby pre-mRNAs are cleaved and a tract of 100-200 adenosine residues is added to the 3' end of the cleaved RNA. Polyadenylation affects mRNA stability, translatability, and nuclear-to-cytoplasmic export (Zhao *et al.,* 1999; Moore and Proudfoot, 2009). In mammalian cells, core RNA sequence *cis* elements, called polyadenylation [poly(A)] signals, and protein factors involved in the 3' processing have been

identified. The poly(A) signals include the highly conserved sequence AAUAAA located 10-30 nt upstream of the RNA cleavage and poly(A) site, and the more variable but functionally important U-rich upstream element (USE), the U/GU-rich downstream element (DSE) found ~30 nt downstream of poly(A) site, and the auxiliary downstream elements (auxDSE) (Gilmartin, 2005; Millevoi and Vagner S, 2010). Mammalian 3' processing factors include the poly(A) polymerase (PAP), the poly(A)-binding proteins, the RNA polymerase II large subunit, and four multi-subunit protein complexes: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulating factor (CstF), and cleavage factors (CF Im and CFIIm) (Mandel *et al*., 2008). CPSF recognizes and binds to the AAUAAA element, recruits other components of the 3' processing complex, and catalyzes cleavage in association with other factors. Following cleavage, CPSF remains bound to the AAUAAA element and anchors PAP to the RNA for polyadenylation. Therefore, the efficient recognition of the AAUAAA element by the CPSF is the key step of 3' end processing.

Although less conserved, plant terminators contain similar poly(A) signals, including the far upstream element (FUE), near upstream element (NUE), and cleavage element (CE) (Loke *et al*., 2005; Ji *et al*., 2007; Hunt, 2008; Mathew *et al*., 2011). FUE is a combination of UG-rich motifs found approximately 50 or more nt upstream of the poly(A) site. NUE is an A-rich element located ~10 to 30 nt upstream of the poly(A) site (Ji et al., 2007). CE is composed of the poly(A) site and surrounding A/U-rich motifs. Despite the significant divergence in

sequence of the poly(A) signals, homologs of most mammalian 3' processing factors have been found in plants and yeast (Mandel *et al*., 2008).

Different terminators strongly influence the level of gene expression (Ingelbrecht *et al*., 1989). Specifically, in transgenic potato, the 3' untranslated region (UTR) of the potato *pin*II gene led to a 10-50 times greater accumulation of the hepatitis B virus surface antigen compared to the nopaline synthase (NOS) terminator derived from the Ti plasmid of *A. tumefaciens* (Richter *et al*., 2000). Another example is the heat shock protein 18.2 (HSP) terminator of *Arabidopsis thaliana* that increased the expression of β-glucuronidase (GUS) by 2- to 3-fold compared to the NOS terminator in both transient and stable transgenic *A. thaliana* (Nagaya *et al*., 2010). Similarly, in transgenic tobacco, the use of the octopine synthase (OCS) terminator from the Ti plasmid of *A. tumefaciens,* or the CaMV35S terminator from Cauliflower mosaic virus (CaMV) instead of the 3' UTR of a C4 photosynthesis gene of *Flaveria bidentis* increased GUS expression several fold (Ali and Taylor, 2001).

In this chapter, the tobacco Extensin (Ext) terminator, which enhances the levels of transgene expression up to 13.5-fold compared to NOS temrinator in *Nicothiana benthamiana* is described. Ext is a member of the hydroxyproline-rich glycoprotein (HRGP) superfamily of plant cell wall proteins that function in diverse aspects of plant growth and development (Showalter, 2010). Through a series of deletion analyses, sequence elements that are responsible for the enhancing effects were identified. Data indicated that the Ext 3' UTR is an efficient terminator in mediating transcription termination and mRNA 3' end

processing, resulting in higher levels of mRNA accumulation and protein expression.

2.3 Materials and Methods

2.3.1 Vector construction

For convenience, the sequence positions for the Ext terminator are given considering the U of the Ext stop codon as -3. The initial construct was the plasmid pBY027-EF, containing a fragment spanning the 1-731 nt from a D13951 *Nicotiana tabacum* genomic DNA cloned in pBY027 (Mor *et al*., 2003) using the *Sac*I and *EcoR*I sites. The fragment was amplified by PCR using primers Ext-1 and Ext-2. The sequences of all primers used in this study are listed in Table 2.1. The intron-less form of *Ext* terminator (spanning 252-731 nt) was amplified by PCR from pBY027-EF using primers EXT-3 and EXT-2 and cloned into the *Sac*I and *Eco*RI sites of pBY027 to produce pBY027-EU. A *Xho*I-*Eco*RI fragment of pBY027-EU was excised and replaced the corresponding fragment in pPS1 (Huang and Mason, 2004) to generate pgEU. The various forms of Ext terminator were similarly cloned using primer pairs of Ext-1 and Ext-2 (pgIEU), Ext-3 and Ext-4 (pgEU1), Ext-3 and Ext-6 (pgEU2), Ext-5 and Ext-2 (pgEU3), and Ext-1 and Ext-6 (pgIEU2). The control plasmids pgNOS, pg35S, and pgVSP were generated as follows. The NOS terminator was excised from construct pIBT211.1 (Richter *et al*., 2000) and cloned into pBY027 to yield pBY027-NOS using the SacI/EcoRI sites. The 35S terminator was amplified by PCR from pRTL2-GUS (Carrington *et al*., 1991) using the primers 35STm-1 and 35STm-2 and cloned

into pBY027 opened with *Sac*I/*EcoR*I to give pBY027-35S. The *Xho*I-*Eco*RI fragments of pBY027-NOS, pBY027-35S, and pBY027 were excised and replaced the corresponding fragment in pPS1 to produce pgNOS, pg35S, and pgVSP. The NVCP expression constructs pnEU, pn35S, pnVSP, pnNOS, pnEU1, pnEU2, pnEU3, pnIEU, and pnIEU2 were obtained by replacing GFP coding region of GFP expression constructs pgEU, pg35S, pgVSP, pgNOS, pgEU1, pgEU2, pgEU3, pgIEU, and pgIEU2 with the NVCP coding region excised from the plasmid pSNV210 (Zhang *et al*., 2006) with the *Xho*I and *Sac*I sites. A fragment that contains the intron of Ext terminator was amplified by PCR from pgIEU using primers Ext-1 and Ext-8 and cloned into pgNOS and pnNOS opened with *Sac*I to give pgINOS and pnINOS. The β-glucuronidase gene encoding GUS (with intron) was PCR-amplified from pGPTVK-GI (Collens *et al.*, 2007) using primers GUS-1 and GUS-2, cleaved with *Xho*I-*Sac*I, and cloned into pgEU, pg35S, pgVSP, and pgNOS by replacing the GFP gene to give puNOS, pu35S, puEU, and puVSP.

2.3.2 DNA constructs for PPS mutants

PPS mutant vectors were constructed using standard overlapping PCR and molecular cloning methods. For the construction of pEUd, primer sets of Ext-3 and Ed-2, and Ed-3 and Ext-6 were used for initial amplification in separate PCRs using pBY027.EF as a template. Resulting PCR fragments were mixed and amplified using primers Ext-3 and Ext-6, complementary to the ends of the two initial fragments. The resulting PCR product was inserted into pBY027 using *Sac*I/*EcoR*I sites and then subcloned into pPS1 via *Xho*I/*EcoR*I. For the

construction of pEUs, the similar steps as above were carried out amplifying two overlapping DNA fragments in separate PCRs. 5’ fragment was obtained by double round PCR: in the first round, primers Ext-1 and Et-2 were used with pBY027.EF as template and the resulting fragment was used for the second round amplification with primers Ext-1 and Vet-6. The 3’ fragment was also obtained by double round PCR: in the first round, primers Et-3 and Ext-2 were used with pBY027.EF as template and the resulting fragment was used for the second round amplification with primers Vet-5 and Ext-2. The resulting overlapping 5’ and 3’ fragments were mixed and amplified into full-length DNA fragment using primers Ext-3 and Ext-2. The final PCR product was inserted into pBY027 using *Sac*I/*EcoR*I sites and then subcloned via *Xho*I/*EcoR*I into pPS1. Construct pNgaa-1 was obtained as followings: to generate 5’ fragment, primer sets of gfp-3f and NE-2, and gfp-3f and VE-6 were used for the first and second round, respectively; to generate 3’ fragment primer sets of NE-3 and NOSREco, and VE-5 and NOSREco were used. The resulting overlapping 5’ and 3’ fragments were used to amplify full-length fragment with primers gfp-3f and NOSREco. The final PCR product was inserted to pgNOS using *Sac*I/*EcoR*I sites. Construct pNgat was obtained as the same way used to obtain pNgaa-1 but with different primers; primers NET-2, NET-3, VET-6 and VET-5 were used instead of NE-2, VE-6, NE-3, and VE-5. Constructs pNcaa and pNrad were obtained as the same way used to obtain pNgaa-1 but with different primers; primers CAA-2, CAA-4, CAA-3 and CAA-5 were used for construct pNcaa and primers RS-2, RS-4, RS-3, and RS-5 were used for the construct pNrad. Two consecutive sets of overlapping PCRs

were also required for construction of pNgaa-2: 5’ fragment was amplified with primer sets of gfp-3f and NE-4, and gfp-3f and VE-6, consecutively; 3’ fragment was amplified with primer sets of NE-5 and NOSREco, and VE-5 and NOSREco. The 5’ and 3’ fragments were mixed and amplified using primers gfp-3f and NOSREco to generate the full-length fragment. The subsequent cloning steps were as described for the construction of pNgaa-1. For the construction of pNgaa-3, to obtain 5’ overlapping fragment, a DNA fragment containing 40 nt of 5’ of NOS terminator and 26-100 nt of Ext 3’ UTR was synthesized by methods described by Withers-Martinez *et al*. (1999). Briefly, six oligos (Oli-1, -2, -3, -4, -5, and -6) were assembled using PCR and amplified with Oli-1 and Oli-6. 3’ overlapping fragment was amplified by PCR with NElink-3 and NOSREco. The 5’ and 3’ fragments were mixed and used to generate final full-length fragment with primers Oli-1 and NOSREco. The subsequent cloning steps were as described for the construction of pNgaa-1.

2.3.3 Agroinfiltration procedure

Binary vectors were separately introduced into *Agrobacterium tumefaciens* LBA4404 by electroporation. The resulting strains were verified by PCR, grown overnight at 30˚℃ and used to infiltrate leaves of 6- to 8-week-old greenhouse-grown *Nicotiana benthamiana* plants. Briefly, the bacteria were pelleted by centrifugation for 5 min at 5,000g and then resuspended in infiltration buffer (10 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH 5.5 and 10 mM $MgSO_4$) to OD600=0.2. The resulting bacterial suspensions were injected by using a syringe

without needle into fully expanded leaves through a small puncture (Huang and Mason, 2004).

2.3.4 GFP assay

Total protein was extracted from leaf samples harvested at 2 DPI with extraction buffer (25mM sodium phosphate buffer, pH 6.6, 100mM NaCl, 1mM EDTA, 0.05% Triton X-100, 50mM sodium ascorbate, and 10 µg/ml leupeptin) using a FastPrep machine (Bio101). Cleared supernatants were obtained by centrifugation at 13,000*g* for 10 min for protein assays. The protein concentration from the leaf samples were determined by Bradford methods using Bradford reagent (Bio-Rad) with bovine serum albumin (BSA) at known concentration as the standard. The GFP fluorescence intensity was examined on a microplate reader (Molecular Device Co, Spectra Max M2). GFP samples were prepared by serial 2-fold dilution with phosphate buffered saline (PBS, 137 mM NaCl, 2.6 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4, pH 7.4) and 50µl of each sample was added to black-wall 96-well plates (Corning), in duplicate. The excitation and emission wavelength were 485nm and 538nm, respectively. All measurements were performed at room temperature and the reading of negative control (extract of un-infiltrated plant leaf) was subtracted before graphing. *E. coli* expressed GFP was used to generate standard curve. GFP gene was cloned into pET28 expression vector (Invitrogen) and IPTG induced GFP was purified using TALON His-Tag purification resin (Clontech).

2.3.5 NVCP ELISA

For NVCP analysis, total leaf protein extracts prepared and normalized as described above were used in NVCP sandwich ELISA (Mason *et al*., 1996). Briefly, a rabbit polyclonal anti-NVCP and a guinea pig polyclonal anti-NVCP were used as capture and detection antibodies, respectively; insect cell-derived recombinant NVCP (Jiang *et al*., 1992) was used as the reference standard.

2.3.6 GUS activity assay

GUS activity was measured as described by Francis *et al*., (2005). Briefly, total leaf protein was extracted from 100 mg leaf tissue at 2 DPI with 1ml GUS Extraction Buffer [150mM sodium phosphate pH 7.0, 10mM EDTA, 10mM β-mercaptoethanol, 0.1% Triton X-100, 0.1% sarcosyl, 10/ml leupeptin] using a FastPrep machine (Bio101). 10µl of cleared extract were incubated with 130 µl Assay Buffer [GUS Extraction Buffer containing 1.2mM 4-methyl-umbelliferyl-β-D-glucuronide (MUG)(Sigma)] in a dark 37˚℃ incubator for 20 minutes. The reaction was stopped by transferring 10µl of the reaction to 190µl Stop Buffer [200mM sodium carbonate] in a black wall 96-well plate. Fluorescence was measured on a SpectraMax M2 microplate reader (Molecular Devices, USA) at 460nm when excited at 355nm. Standard curve was generated using 50, 25, 5, 2.5, 0.5, 0.25, and 0 µM 4-methylumbelliferone (MU) with every plate. Values from the fluorescence assay were converted to moles of MU/minute, and then standardized by protein concentration that determined by Bio-Rad *RC DC* protein assay kit according to the kit instructions. Known concentration of BSA was used as reference standard for protein concentration determination.

2.3.7 RNA extraction and Quantitative PCR and RT-PCR

Total RNA was purified from infiltrated tobacco leaves at 2 DPI using Plant RNA Reagent (Invitrogen), and residual DNA was removed by DNAfree system (Ambion). First-strand cDNA was synthesized from 1 μg of total RNA and oligo dT_{22} primer using the Superscript III First-Strand Synthesis System (Invitrogen) according to the manufacturer's protocol. Real-time PCR for each transcript of interest was performed on an IQ5 Real-Time PCR Detection System (Bio-Rad) using gene specific primers (gfp-f and gfp-r for GFP; sNV-f and sNV-r for NVCP) and custom-made Taqman FAM/MGB probes (GFP-p and sNV-p, Integrated DNA Technologies). Each sample was measured in triplicate for each transcript of interest and an internal reference gene. Elongation factor (EF1a) transcript served as an internal control (EFf, EFr and EFp, Integrated DNA Technologies). Transcript levels of GFP, NVCP and EF1a were quantified using separate standard curves prepared with plasmid pBY027, pSNV210, and pCR4-Topo-EF1a, respectively. The construct pCR4-Topo-EF1a was obtained as follows. RT-PCR was performed with cDNA from wild-type *N. benthamiana* RNA using EF1f and EF1r primers, which amplify 119bp (nt 167-285) fragment of EF1a gene (accession number AY206004). The amplified product was cloned to a PCR cloning vector, pCR4-Topo (Invitrogen). The relative quantification of GFP and NVCP transcripts was normalized against EF1a transcripts.

2.3.8 Read-through PCR

First-strand cDNA was synthesized from 1 μg of total RNA, described above, with a random primer. PCRs were performed to evaluate the transcription read-through of GFP and NVCP transcripts with four sets of primers. For GFP transcripts, a forward primer (gfp-3f), specific to GFP gene, was paired with four reverse primers RT-0, RT-1, RT-2, and RT-3, positioned at 42, 156, 291, and 389, respectively, downstream of the *EcoR*I site on the expression vectors. For NVCP transcripts, a NVCP specific sense primer sNV-3f was paired with the same four antisense primers used for GFP transcripts test.

2.3.9 Analysis of transcript 3' ends

First-strand cDNA was synthesized from the total RNA described above with oligo dT-anchor primer, DT-2. PCRs were performed by standard procedures with the cDNA obtained. For amplification of the GFP mRNA 3' end, a forward primmer gfp-3f was used with reverse primer, DT-4, that designed to bind to the anchor sequence of the oligo dT-anchor primer, DT-2. For NVCP transcripts, a forward primer sNV-3f was used with DT-4. The PCR products were cloned into a PCR cloning vector, pCR4-Topo (Invitrogen), and 20 positive clones were randomly selected and sequenced.

2.4 Results

2.4.1 Ext terminator increases transient transgene expression

Effects of the tobacco Ext terminator (Accession number D13951) on transient transgene expression were evaluated in comparison to other widely used

terminators, including NOS, CaMV 35S, and soybean vegetative storage protein (VSP). Ext terminator consists of 746 nt and contains an intron between nt 24 and 249, considering the U of termination codon as -3. For this comparison, Ext terminator (nt 252-731) and the other terminators were placed downstream of the green fluorescent protein gene (GFP), driven by the CaMV 35S promoter and tobacco etch virus (TEV) 5' UTR (Fig. 2.1A). The resulting constructs were introduced into *N. benthamiana* leaves by agro-infiltration, and at 2 days post infiltration (DPI) the level of GFP expression was analyzed. As shown in Figure 2.1B, Ext construct (pgEU) led to the highest GFP expression level. GFP expression levels with pgEU was 13.5-fold, 11.9-fold, and 2.8-fold higher than those with the NOS, VSP, and 35S constructs (pgNOS, pgVSP, and pg35S), respectively. To test whether the enhancing effect of the Ext terminator is gene specific, the GFP gene was replaced with the Norwalk virus capsid protein (NVCP) gene and the level of NVCP expression was compared. NVCP is a candidate vaccine antigen for the protection against Norwalk virus infections, which cause epidemic acute gastroenteritis in humans. Results using NVCP were similar to those with GFP, suggesting that the enhanced transgene expression by the Ext terminator is not gene specific (Fig. 2.1C). To determine whether the Ext terminator also functions without TEV 5' UTR, I directly fused the 35S promoter to a β-glucuronidase (GUS) reporter gene, followed by the various terminators (Fig. 2.1A). I found that the Ext terminator increased expression of the GUS gene without TEV 5' UTR (Fig. 2.1D). These data suggest that the ability of the Ext

terminator to increase expression level is independent of specific 5' UTR and transgene.

2.4.2 Deletion analysis of the Ext terminator

Because the quantitative analyses showed that the Ext terminator enhances gene expression, I asked whether specific sequence elements of Ext terminator control its activity. To address the question, I made a series of deletion constructs. Positions of four putative NUEs, intron, and polypurine sequence (PPS) of the Ext terminator are indicated in Figure 2.2A. Compared to the full-length construct EU (nt 252-731), the deletion construct pEU1 (nt 252-553) retains all four NUE, while pEU2 (nt 252-464) retains only the first three NUEs. pEU3 (nt 465-731) was designed to exclude the first three NUEs but to include the fourth NUE. The effects of these deletions were examined by measuring the transient expression of GFP and NVCP (Fig. 2.2B and 2.2C). Deletion construct pEU1 gave a similar level of GFP or NVCP expression compared to the full-length construct pEU, indicating that the sequence between 554-731 is not necessary for full-strength gene expression. In contrast, deletion construct pEU2 exhibited a substantially lower level of GFP or NVCP gene expression compared to construct EU (11-14% activity remained). These data indicate that the 89-bp sequence (nt 465-553), in which the fourth NUE is located, is essential to yield maximum gene expression. However, the presence of the 89-bp sequence without the sequence nt 252-464 was not sufficient to support strong gene expression as pEU3 showed no detectable level of GFP and NVCP expression. Collectively, these data indicate that sequence between nt 252-464 is absolutely required for optimal function of

the Ext terminator and that 89-bp sequence (465-553) is required in coordination with the sequence 252-464.

2.4.3 Influence of the Ext intron on gene expression

Based on a bioinformatic analysis of 32,955 annotated *Arabidopsis thaliana* protein-coding genes, Chung *et al.* (2006) reported that only 5.6% of the 3' UTR include introns, while 72% of protein-coding sequences and 19.9% of the 5' UTR contain introns. Interestingly, many cell wall protein coding genes have an intron located in the 3' UTR, including the HRGP homologues of teosinte, sorghum, and rice, and the tobacco Ext gene (Menossi *et al*., 2003). The presence of an intron in the tobacco Ext 3' UTR led us to study its effect on gene expression. For this, I generated intron-containing Ext terminator constructs: the Ext terminator ranging from nt 1-731 or nt 1-464, including the intron sequence (nt 24-249), was PCR amplified from tobacco genomic DNA and fused to the GFP gene (pIEU or pIEU2, respectively). GFP expression study revealed that the Ext intron has a deleterious effect on gene expression: the intron-containing constructs pIEU and pIEU2 showed significantly lower level of GFP expression compared to the intron-less construct pEU and pEU2 (~21-30% activity remained, Fig. 2.3A). Consistently, when the GFP gene was replaced with the NVCP gene, the NVCP expression of the intron-containing construct, pIEU or pIEU2, was also lower than that of the intronless constructs, pEU or pEU2 (~29-75% activity remained, Fig. 2.3B).

The effect of the Ext intron was also tested in context of the NOS terminator: PCR-amplified Ext intron was inserted to the 5' end of the NOS

terminator preceded by the GFP or NVCP gene. Unexpectedly, the addition of the Ext intron to NOS terminator caused 51% increase on GFP expression, and 34% increase on NVCP expression (Fig. 2.3, pNOS vs. pINOS). Taken together, these data indicate that the Ext intron has regulatory effects on transient transgene expression, in a context dependent manner.

2.4.4 Ext terminator increases mRNA accumulation

The upregulated transgene expression mediated by the Ext terminator could be caused by an increase in either mRNA level or translational efficiency. To investigate whether the Ext terminator affects mRNA accumulation, the level of accumulated transgene mRNA of Ext terminator constructs were compared with that of NOS terminator constructs. As shown in Figure 2.4, construct pgEU produced approximately 20-fold increase in GFP mRNA accumulation compared to construct pgNOS. Consistently, construct pnEU gave an approximately 10-fold higher NVCP mRNA accumulation compared to construct pnNOS. These results indicated that the enhanced transgene expression mediated by the Ext terminator is at least partially due to increased mRNA accumulation.

I also compared the level of mRNA accumulation between intron-less and intron-containing constructs (pEU vs. pIEU). The use of the Ext intron caused a 40-50% decrease in the mRNA accumulation (Fig. 2.4A and 2.4B), which was consistent with the protein expression data (Fig. 2.3A and 2.3B).

2.4.5 Ext terminator mediates efficient transcription termination.

The increased mRNA accumulation could be the result of improved mRNA stability and/or mRNA synthesis. The mRNA stability is greatly affected

by the efficiency of transcription termination and mRNA 3' end processing, because improperly terminated and unpolyadenylated mRNA is targeted by RNA-dependent RNA polymerase 6 (RDRP6)-mediated RNA silencing (Luo and Chen, 2007). To check if the Ext terminator increases the efficiency of transcription termination, I tested the presence of read-through transcripts from Ext and NOS terminator constructs using RT-PCR (Fig. 2.5). A forward primer specific to either the GFP or NVCP gene (gfp-3f or sNV-3f) was paired with one of four antisense primers (RT-0, RT-1, RT-2, and RT-3) specific to different downstream regions of the terminator. From the NOS terminator samples, specific RT-PCR bands of the same size as the positive controls were detected with the first three reverse primers, RT-0, RT-1, and RT-2, indicating that the NOS terminator was leaky for transcripts driven by the strong CaMV 35S promoter. In contrast, no detectable levels of specific bands were amplified in the Ext terminator samples with any primer set. These data demonstrate that the Ext terminator is more efficient in terminating the transcription of a transgene, and as a result, contributes to the higher mRNA and protein accumulation.

2.4.6 PPS serves as a dominant poly(A) site

The data presented thus far suggest that the Ext terminator enhances transgene expression through increased mRNA accumulation due to efficient transcription termination and/or 3' end processing. To characterize the sequence elements involved, the locations and features of the poly(A) sites on transgene transcripts were analyzed. The 3' ends of GFP and NVCP transcripts were RT-PCR amplified, cloned, and 20 randomly chosen clones were sequenced. From

construct pgEU, four poly(A)sites were found with 11, 2, 6, and 1 sequences starting the poly(A) tract at nt 62, 70, 82, and 180, respectively (Fig. 2.6A). Interestingly, three of the four sites representing 95% of the total sequenced transcripts (19 out of 20) were located in the PPS (nt 59-103). Similarly, six poly(A) sites were identified in transcripts from construct pnEU, with 9, 1, 4, 1, 2 and 3 sequences starting the poly(A) tract at nt 62, 70, 82, 85, 88, and 269, respectively. In this case, five of the six sites, 85% of the total sequenced transcripts (17 out of 20) were processed in the PPS.

Multiple poly(A) sites were also found in transcripts from constructs pgNOS and pnNOS (Fig. 2.6B). From pgNOS, eleven poly(A) sites were found, with three positions located at 14-63nt (14, 19, and 63nt) downstream from the first NUE (30% sequenced samples). The remaining 8 positions were located at 8-46nt (8, 11, 25, 28, 30, 40, 43, and 46nt) downstream from the second NUE (70% samples). Similarly, from the pnNOS construct, five positions were found, two located at 14-19 nt (14 and 19, 50% samples) downstream from the first NUE, and three located at 25-99 nt (25, 40, and 99nt, 50% samples) downstream from the second NUE.

2.4.7 PPS is an important element for the Ext terminator function

The analyses of poly(A) sites suggested that the PPS, the dominant poly(A) site, is critical for the Ext terminator function. To test this hypothesis, I deleted or substituted the PPS and tested its effects on GFP expression (Fig. 2.7). For the substitution, some adenines in the PPS were replaced with thymines, while maintaining the segment length (see Materials and Methods). Upon deletion or

substitution, the levels of GFP expression were substantially decreased (~36-40% activity remained), yet remained 4.7-5.2-fold higher than with the construct using the NOS terminator. These data indicate that the PPS is an important factor contributing to the enhancing effect of the Ext terminator. To map the locations of the poly(A) sites of the mutant constructs, transcript 3' end fragments were RT-PCR amplified, cloned and sequenced as described above. From the deletion mutant pgEUd, nine poly(A) sites were found among 20 sequenced clones with 1, 1, 2, 6, 1, 1, 5, 1, and 2 sequences starting the poly(A) tract at positions 95, 132, 133, 135, 149, 167, 224, 229, and 236, respectively. Regions nt 131-134 (9 of 20) and nt 223-235 (8 of 20) were the dominant poly(A) sites in the absence of PPS. Similarly, from the substitution mutant pgEUs, eight poly(A) sites were identified among 20 sequenced clones with position 140 nt (6 of 20) and region nt 178-185 (6 of 20) as the dominant sites.

In order to examine whether the PPS can enhance transgene expression in a context other than with Ext terminator, I inserted the PPS or variants either upstream or downstream of the first NUE of NOS terminator (Fig. 2.7A): PPS was placed 5 nt upstream (pNgaa-1) or 27 nt downstream (pNgaa-2) of the first NUE; PPS with an additional 30 nt derived from the Ext 3' UTR was linked to the first NUE (pNgaa-3); the thymine substituted sequence (pNgat), the cytosine substituted sequence (pNcaa), and random sequence (pNrad) were inserted 5 nt upstream of the first NUE. The modified sequences are shown in Figure 2.8. As shown in Figure 2.7C, insertion of the PPS either upstream (pNgaa-1) or downstream (pNgaa-2) of the first NUE increased GFP expression by 4.8- and

3.2-fold respectively. Conversely, insertion of the PPS with extra 30 nt derived from the Ext 3' UTR (pNgaa-3) decreased GFP expression by 44%. Unexpectedly, the thymine (pNgat) and cytosine (pNcaa) substitution mutants showed 2.3- and 3.3-fold higher expression than the unmodified NOS terminator. Random sequence insertion clone (pNrad) also gave 47% increased GFP expression. To identify poly(A) sites of the mutant constructs, transcript 3' end sequencing was conducted. This analysis revealed that the PPS also performs as a dominant poly(A) site in the context of the NOS terminator: in either pNgaa-1 or pNgaa-2, over 80% of poly(A) sites were clustered in the PPS. In contrast, no poly(A) site was located in either the thymine substituted sequence (pNgat) or random sequence (pNrad). Two poly(A) sites were found in the inserted sequence of construct pNcaa. Interestingly, the PPS in the pNgaa-3 also served as a poly(A) site (90%), although reduced expression was observed.

2.5 Discussion

Pre-mRNA 3' end processing is an essential step in the expression of genes and is greatly affected by transcription terminators. Hence, careful selection of an optimal terminator is important for high yield recombinant protein production. In this study, I evaluated the tobacco Ext terminator for its potential to enhance transgene expression. The Ext terminator produced higher levels (up to 13.5-fold) of transient transgene expression from three transgenes, GFP, NVCP, and GUS. Interestingly, the three transgenes showed similar expression level patterns in coordination with the tested terminators: i.e. all transgenes showed the

highest expression level with the Ext terminator, followed in order by the 35S, VSP, and NOS terminators (Fig. 2.1). These results suggest that none of these terminators are gene specific. I did not directly address the effect of the 5' UTR on the efficiency of the Ext terminator, but the enhancing effect of the Ext terminator was not obviously affected by 5' UTR, considering that the GFP and NVCP expressing vectors contained the TEV 5' UTR, while the GUS constructs did not. Whether the Ext 3' UTR will coordinate well with other promoters and/or 5' UTRs remains to be investigated.

In addition to terminators, introns can also significantly affect gene expression. For example, some introns boost expression by containing enhancer elements or through a poorly defined process termed intron-mediated enhancement (IME) (Mass *et al*., 1991; Rethmeier *et al*., 1997; Rose, 2004). Conversely, some introns are required for tissue specific or developmentally restricted gene expression: introns in the *Arabidopsis* agamous (AG) gene and the Seedstick (STK) gene restrict AG and STK expression to specific tissues within the flower (Sieburth and Meyerowitz, 1997; Kooiker *et al*., 2005); and an intron in the *Arabidopsis* floral repressor *Flowering Locus C* (*FLC*) gene reduces *FLC* expression in response to vernalization (Sheldon *et al*., 2002). It has been observed that the expression of Ext is regulated tissue specifically and developmentally, and is induced by various stress conditions (Showalter *et al*., 2010; Hirsinger *et al*., 1997). The presence of an intron in the 3' UTR of the Ext gene and its regulated expression might imply the involvement of the intron in the regulated Ext expression. I showed that the Ext intron has regulatory function on

transient transgene expression in a context dependent manner. Specifically, it showed a deleterious effect (up to 70% decrease) when linked with Ext terminator but showed a positive effect (up to 50% increase) when used in combination with NOS terminator (Fig. 2.3). Consistent with the data, Menossi *et al*. (2003) reported the tissue type dependent regulatory function of Maize HRGP gene 3' UTR: it increased GUS expression in fast-growing coleoptiles, but decreased GUS expression in maize suspension cells. Further studies of the Ext intron will provide valuable insight into the intron-mediated regulation of gene expression.

I showed that the Ext 3' UTR substantially reduces the level of readthrough transcripts compared to other terminators (Fig. 2.5). Although Xing *et al*. (2010) suggested that limited transctiption readthrough (~1% of transcripts) is a normal phenomenon of plant gene expression, there are strong data indicating that the readthrough transcripts have negative effects on gene expression. For example, readthrough transcripts can trigger RDRP6-mediated RNA silencing, which leads to a cascade of mRNA degradation (Luo and Chen, 2007). Best studied in this regard is the GUS transgene expression in *Arabidopsis thaliana*. A GUS transgene without a terminator produced readthrough mRNA and consistent RDRP6-dependent RNA silencing. However, this phenomenon was resolved when two 3' terminators were placed 3' of the GUS transgene: the level of readthrough transcripts and GUS-specific small interfering RNA were decreased, resulting in higher GUS expression.

Mapendano *et al*. (2010) demonstrated another possible mechanism whereby readthrough transcripts mediated down regulation of gene expression.

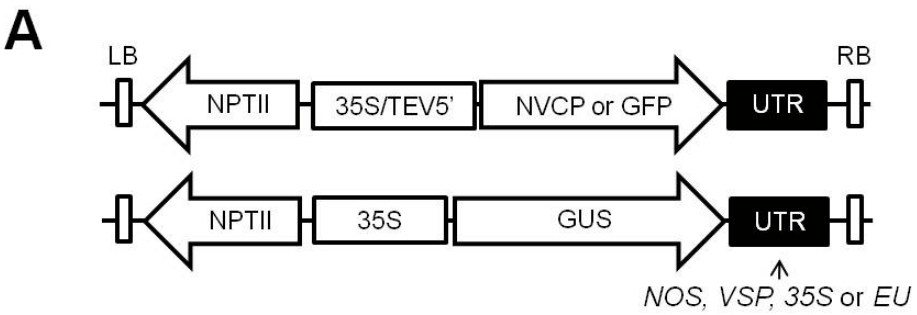
When the poly(A) signal was mutated, RNA polymerase II (RNAPII) complexes were found in the readthrough region of the mutated gene. This readthrough RNAPII, engaged on the mutated transcription units sequesters the transcription initiation/elongation factors, leading to their depletion at the promoter, thereby preventing sustained transcription initiation. Based on these references and the data showing that the Ext terminator decreases the level of readthrough transcripts and increases mRNA accumulation, I suggest that the Ext terminator increases transgene expression by preventing the activation of silencing and/or stimulating continued transcription.

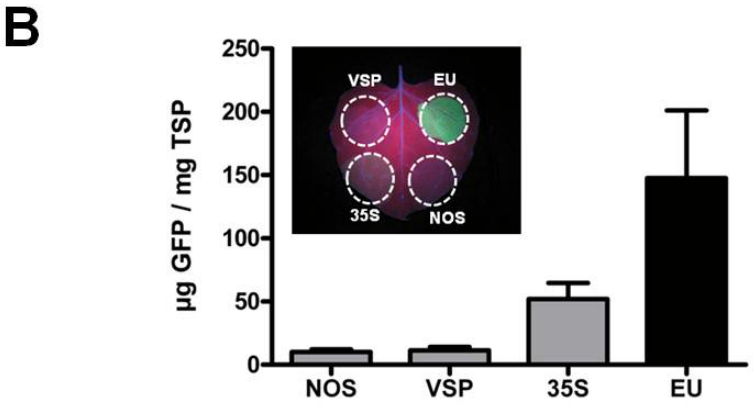
Ext 3' UTR contains four putative NUE (Fig. 2.2A). Sequence analyses of the 3' end of transgene transcripts revealed that most transcripts terminate shortly after the first and second NUEs (Fig. 2.6A). Although less than 15% of transcripts continued further downstream and terminated after the third or fourth NUE, the observation that the deletion of the fourth NUE substantially decreases transgene expression (Fig. 2.2B and 2.2C) suggests the importance of having the additional signals. Possibly, the additional signals may function to ensure termination of RNA transcription to minimize the generation of readthrough transcripts.

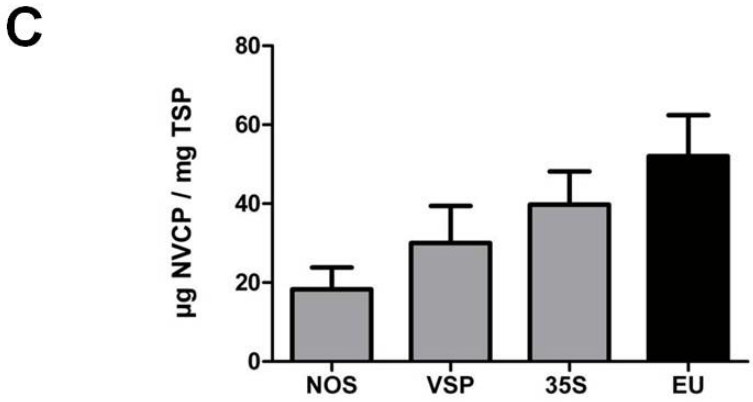
I showed that over 85 % of GFP and NVCP mRNAs were processed in the PPS of Ext terminator (Fig. 2.6A). Moreover, deletion of the PPS caused significant decrease in the level of expression (Fig. 2.7B). Based on the observations, I speculated that the PPS has an enhancing effect on mRNA 3' end processing. Then, what is the mechanism underlying the enhancing effect of the

PPS? In mammalian cells, it has been shown that the efficient recognition of the AAUAAA element by the CPSF is the key step for the stable 3' processing complex formation (Mandel *et al*., 2008). Although no conserved RNA structural elements have been found to be shared among poly(A) sites (Ji *et al*., 2007), studies indicated that an open or unstructured conformation facilitate the binding of CPSF at the poly(A) site and thus facilitate the subsequent assembly of the stable 3' processing complex (Gimmi *et al*., 1989; Graveley *et al*., 1996). RNA secondary structure analyses using *RNAfold* WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) revealed that PPS in the Ext terminator provides an extensively open structure (data not shown). Therefore, the data suggest that the polypurine sequence in the Ext terminator yields an enhancing effect on mRNA 3' end processing by providing a structurally open region encompassing the NUE, thereby promoting the initial interaction with CPSF.

In conclusion, I find that the production of recombinant proteins in plants is enhanced by use of the Ext terminator, with increases in transgene expression up to 13.5-fold. I also suggest that the additional poly(A) signals and PPS are important features of the Ext terminator that allow it to be more efficient in mRNA 3' end processing, thereby yielding enhanced transgene expression.


A
LB
RB
NPTII
35S/TEV5'
NVCP or GFP
UTR
NPTII
35S
GUS
UTR
NOS, VSP, 35S or EU


B
µg GFP / mg TSP
250
200
150
100
50
0
VSP
EU
35S
NOS
NOS
VSP
35S
EU


C
µg NVCP / mg TSP
80
60
40
20
0
NOS
VSP
35S
EU

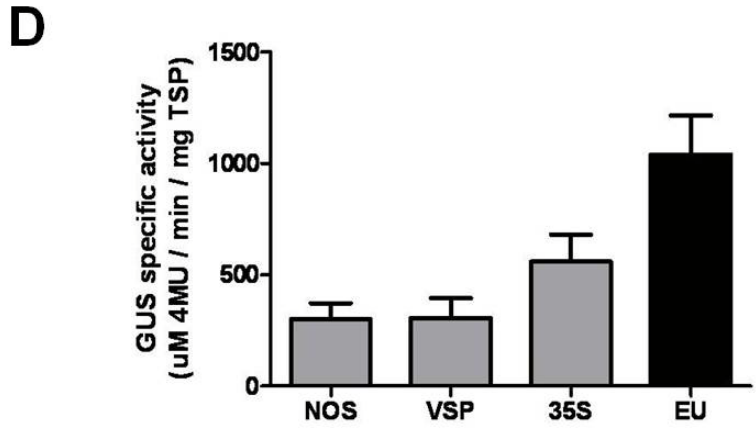
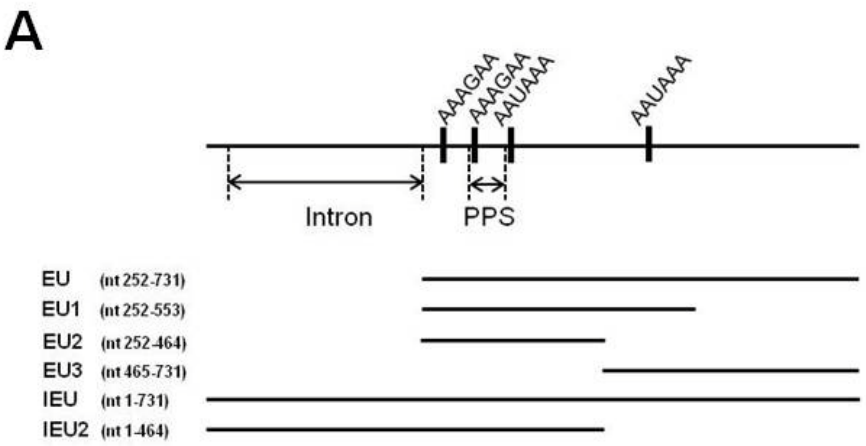
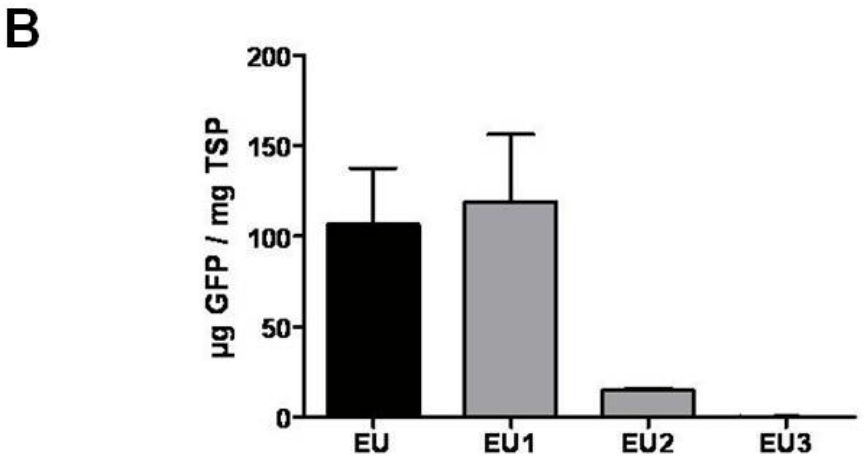

D
GUS specific activity
(uM 4MU / min / mg TSP)
1500
1000
500
0
NOS
VSP
35S
EU

Figure 2. 1 Ext terminator increases transient transgene expression in *N. benthamiana* leaves. (A) Diagram of terminator testing constructs. Different terminators are placed downstream of gene of interest. 35S, CaMV 35S promoter; 35S/TEV5', CaMV 35S promoter with tobacco etch virus 5' UTR; NOS, agro bacterium nopaline synthase gene 3' element; VSP, soybene vspB gene 3' element; 35S, CaMV 35S terminator; LB and RB, the left and right borders of the T-DNA region; NPTII, expression cassette encoding *nptII* gene for kanamycin resistance. **(B)** Fluorometric analysis of GFP expression. Infiltrated leaf extract were analyzed by spectrofluorimetry using excitation and emission wavelengths of 485nm and 538nm. Insert: Visualization of GFP expression. Infiltrated leaf was examined at 2 days post infiltration (DPI) under UV illumination (365 nm) generated by a B-100AP lamp (UVP, Upland, CA). **(C)** Expression of NVCP analyzed by ELISA. Insect-drived NVCP was used as the standard. **(D)** Fluorometric GUS activity assay measured by monitoring the cleavage of MUG (β-glucuronidase substrate, 4-methylumbelliferyl β-D-glucuronide) at 460nm when excited at 355nm. MU (4-methylumbelliferone) was used to generate standard curve. Columns represent means ± S.D. from six independently infiltrated samples (**B-D**).


A
AAAGAA
AAAGAA
AAUAAA
AAUAAA
Intron
PPS
EU (nt 252-731)
EU1 (nt 252-553)
EU2 (nt 252-464)
EU3 (nt 465-731)
IEU (nt 1-731)
IEU2 (nt 1-464)


B
µg GFP / mg TSP
200
150
100
50
0
EU
EU1
EU2
EU3

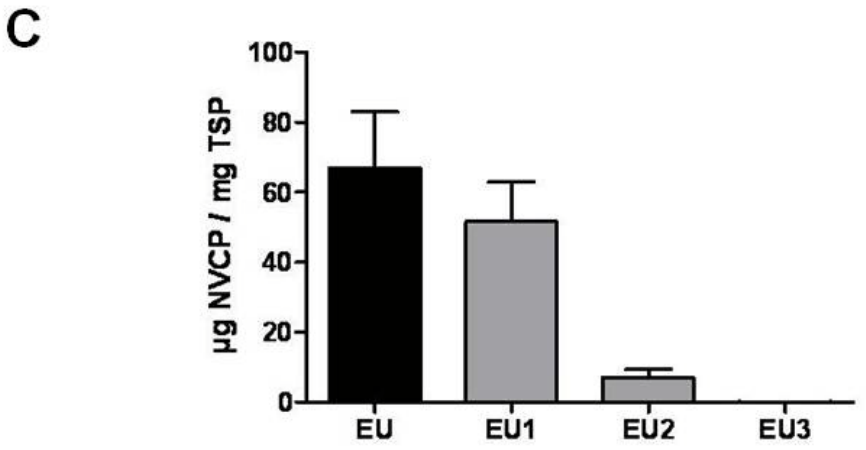

C
µg NVCP / mg TSP
100
80
60
40
20
0
EU
EU1
EU2
EU3

Figure 2.2 Influence of deletions in Ext terminator on gene expression. (A) Schematic representation of the deletion constructs. The vertical black bars indicate the putative near upstream elements (NUEs). The positions of intron and polypurine sequence (PPS) were indicated with arrows. **(B)** Fluorometric analysis of GFP expression. *N. benthamiana* leaves were infiltrated with the constructs depicted in **A** and the GFP concentrations were measured as described in Fig. 2.1B. **(C)** Expression of NVCP analyzed by ELISA as described in Fig. 2.1C. Data are means ± S.D. from three independently infiltrated samples (**B and C**).

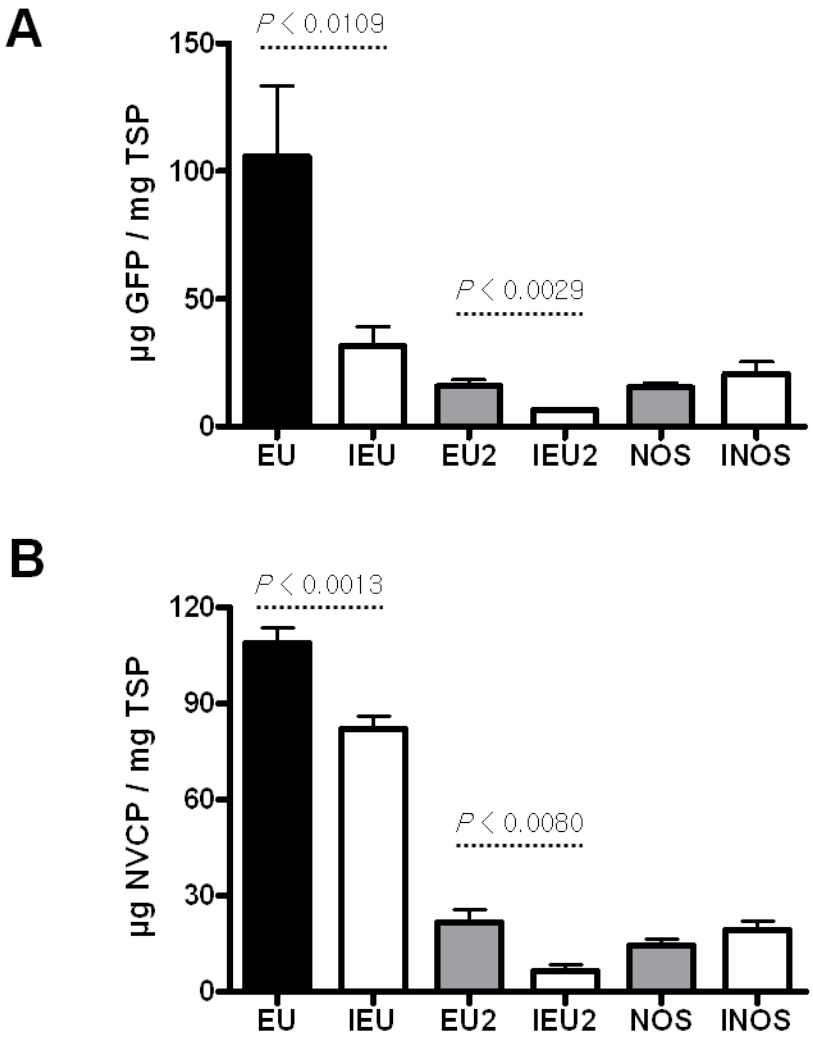


Figure 2.3 Influence of the Ext intron on gene expression. Constructs represented in Fig 2.2A were tested for GFP or NVCP expression. (A) GFP expression was determined by fluorometric analysis as described in Fig. 2.1B from three independently infiltrated samples. (B) NVCP expression was measured by ELISA from six independent samples. White or colored bars represent constructs with or without intron, respectively. Data are means ± S.D. from three independently infiltrated samples. P, *t* test.

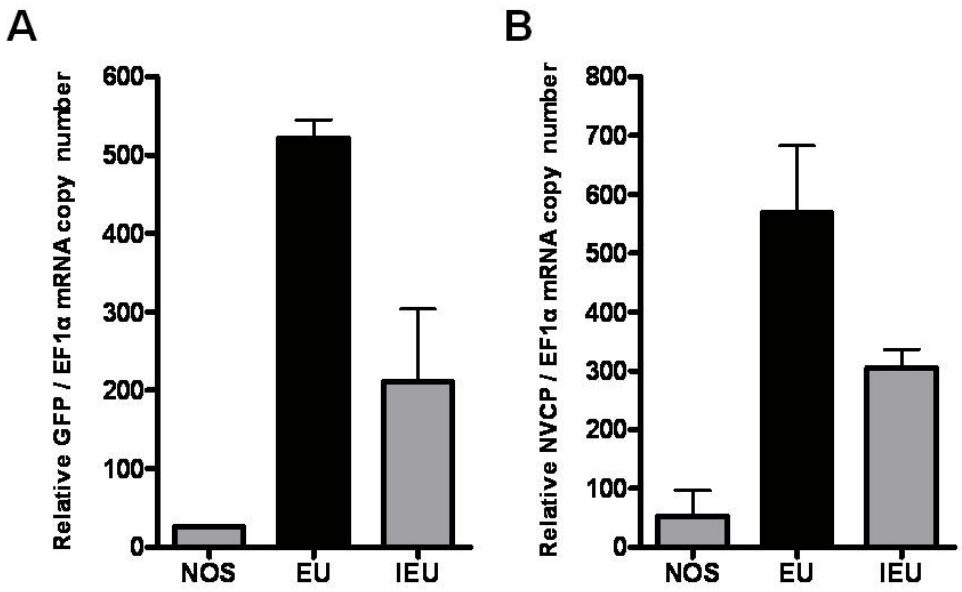
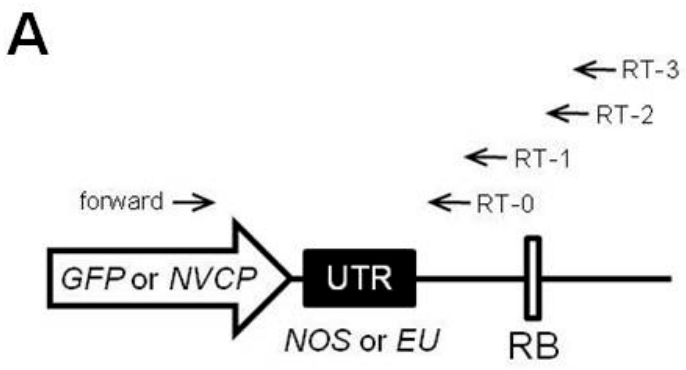
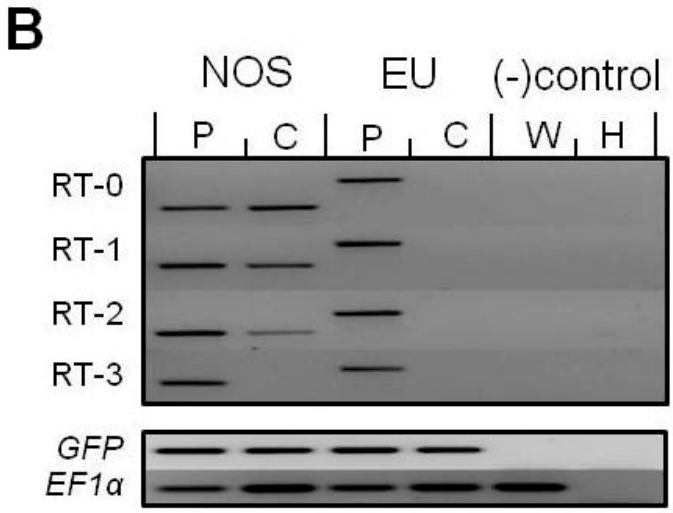


Figure 2.4 Ext terminator increases mRNA accumulation. Real-time RT-PCR analysis was performed with purified RNA from agro-infiltrated leaf samples after 2 DPI. Relative mRNA accumulations of GFP **(A)** and NVCP **(B)** are shown. Real-time PCR for each transcript of interest was performed on an IQ5 Real-Time PCR Detection System (Bio-Rad) using gene specific primers and custom made Taqman FAM/MGB probes. Each sample was measured in triplicate for each transcript of interest and an internal reference gene. Elongation factor (EF1a) transcript served as an internal control. Transcript levels of GFP, NVCP and EF1a were quantified using separate standard curves and then the mRNA copy numbers of GFP **(A)** and NVCP **(B)** are normalized against the mRNA copy numbers of EF1a. Data are means ± S.D. from three independently infiltrated samples.


A
RT-3
RT-2
RT-1
RT-0
forward
GFP or NVCP
UTR
NOS or EU
RB


B
NOS
EU
(-)control
P
C
P
C
W
H
RT-0
RT-1
RT-2
RT-3
GFP
EF1α

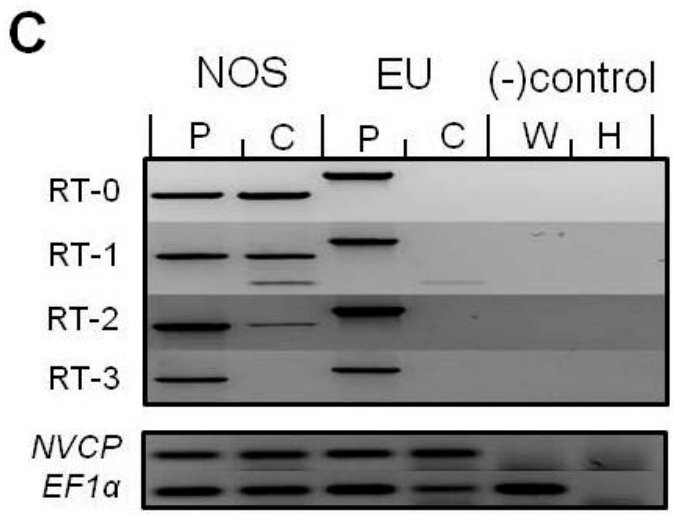
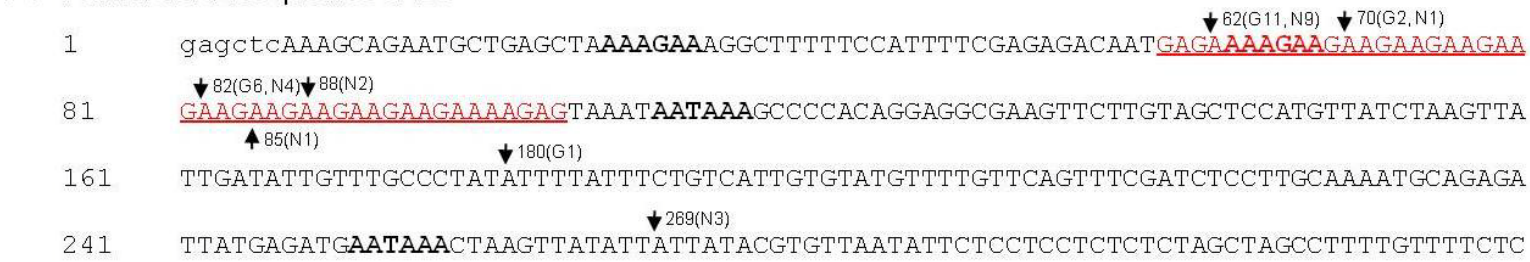

C
NOS
EU
(-)control
P
C
P
C
W
H
RT-0
RT-1
RT-2
RT-3
NVCP
EF1α

Figure 2.5 Ext terminator is efficient on transcription termination mediation. **(A)** Diagram of part of the T-DNA region of the constructs used for this study. NOS or Ext terminator is placed downstream of *GFP* or *NVCP* gene. RB, the right border of the T-DNA region. Horizontal arrows indicate primer regions used for the transcription read through detection by RT-PCR. **(B)** Agarose gel electrophoresis of RT-PCR product to check transcription read through. NOS, pgNOS; EU, pgEU. Template types are indicated: P, plasmid DNA; C, cDNA; W, cDNA of wild-type leaf sample, H, H2O. A gene specific forward primer was used with the reverse sense primers, binds to different regions downstream of the terminator, indicated on the left of upper panel. **(C)** Agarose gel electrophoresis of RT-PCR product to check transcription read through. NOS, pnNOS; EU, pnEU. To check RNA quality, RT-PCR was performed on the same RNA samples with *GFP* (B) or *NVCP* (C) and *EF1a* (B, C) primer sets.

A Partial cDNA sequence of EU



58

B Partial cDNA sequence of NOS

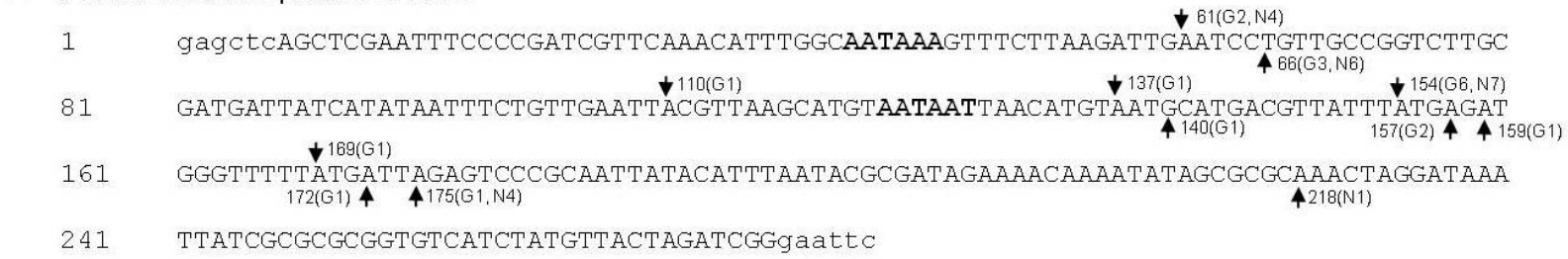


Figure 2.6 Identification of poly(A) sites of terminator constructs. The RNA samples described in Fig. 2.2 were reverse transcribed with oligo dT-anchor primer DT-2 and used for subsequent PCR amplification. A forward primer specific to GFP or NVCP gene, gfp-3f or sNV-3f, and a reverse primer DT-4, binding to the anchor of the DT-2 primer, were used for the PCR and resulting products were cloned. 20 positive clones were sequenced. (A) The partial cDNA sequence of Ext 3' UTR is shown. The conserved NUE AATAAA and AATAAA like sequence AAAGAA are in bold. The poly-purine sequence (PPS) in the Ext terminator is underlined. The arrows pointing the cDNA sequence indicate poly(A) sites of the terminator constructs. Numbers and letters next to the arrow, such as 62 (G11, N9), indicate that 11 clones of gEU and 9 clones of nEU constructs have the same poly(A) site starting at position 62. (B) The cDNA sequence of NOS termiantor is shown. The conserved NUE AATAAA and AATAAA like sequence AATAAT are in bold. Restriction enzymes sites used for cloning are shown in lowercase letter.

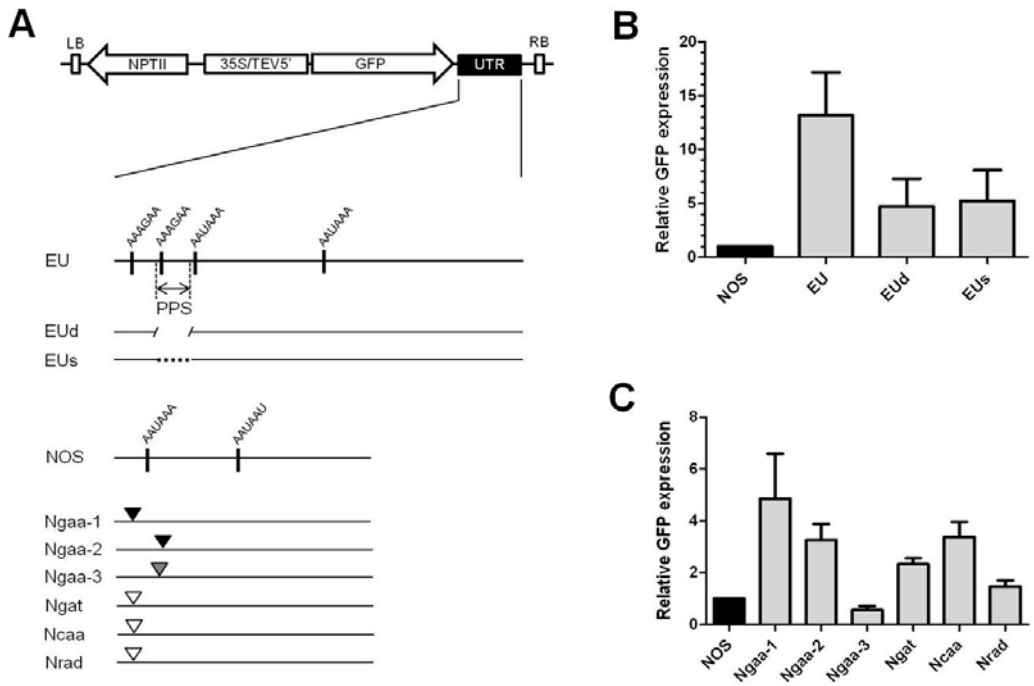


Figure 2.7 PPS is an important element for the Ext terminator function. (A) Diagram of PPS testing constructs. Modified terminators are placed downstream of GFP gene under the control 35S/TEV5', CaMV 35S promoter with tobacco etch virus 5' UTR; LB and RB, the left and right borders of the T-DNA region; NPTII, expression cassette encoding *nptII* gene for kanamycin resistance. The positions of the putative NUEs are shown on the constructs with Ext (EU) and NOS (NOS) terminators. PPS is indicated by vertical dotted line and arrow. The insertion position of the PPS (black triangle), PPS with extra sequence (gray triangle), and variant sequences (white triangle) are shown. (**B** and **C**) Fluorometric analysis of GFP expression. Leaf extract from four to five independently infiltrated samples were analyzed by spectrofluorimetry using excitation and emission wavelengths of 485nm and 538nm. The columns show the relative GFP expression of each construct: values form the fluorometric analyses were converted to µg GFP/mg TSP, and then normalized against the GFP expression of gNOS (indicated by black column). Data are means ± S.D. from four independently infiltrated samples

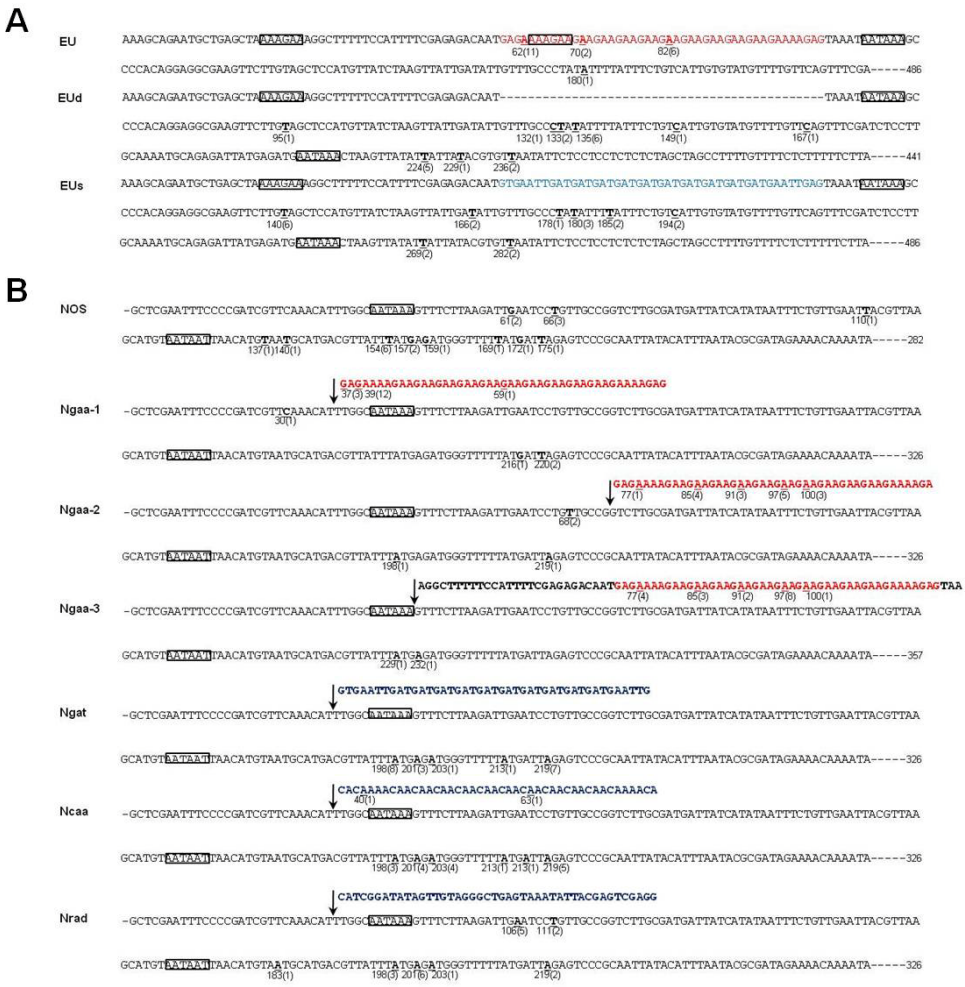


Figure 2.8 Identification of poly(A) sites of PPS testing constructs. The partial cDNA sequence of Ext 3' UTR and mutants (A), and NOS terminator and mutants (B). The RNA samples used in Fig. 2.7 were RT-PCR amplified, cloned and 20 positive clones were sequenced as described in Fig. 2.6. The partial cDNA sequence of each construct depicted on the left is shown. The putative NUEs are in boxes. The poly-purine sequence (PPS) in the Ext terminator is indicated in red. The thymine or cytosine substituted sequence is indicated in blue. The arrows above the cDNA sequence indicate the insertion position of the sequence next to the arrows. The poly(A) sites are indicated by the underlined bold letters. The numbers below the poly(A) sites, such as 62 (11), indicate that 11 clones EU constructs have the same poly(A) site starting at position 62. The total length of each terminator is indicated at the end of the terminator sequence.

Table 2.1 Sequences of oligonucleotides used in this study

Name	Sequence (5' to 3')
CAA-2	GTTGTTGTTGTTGTTTTGTGATGTTTGAACGATCGGGG
CAA-3	CAACAACAACAACAACAAAACATTGGCAATAAAGTTTCTTAAGATTG
CAA-4	TGTTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTTGTG
CAA-5	CACAAAACAACAACAACAACAACAACAACAACAACAAAACA
35STm-1	GTGAGCTCGTCCGCAAAAATCACCAG
35STm-2	CAGAATTCGTCACTGGATTTTGGTTTTAGG
DT-2	CGCTGTGCTGCCGTGCCTTTTTTTTTTTTTTTTTT
DT-4	CGCTGTGCTGCCGTGCCTT
EFf	CTGGTGGTTTTGAAGCTGGTA
EFp	FAM-CACGCATTGCTTGCTTTCACCC-TAMRA
EFr	GGTGGTAGCATCCATCTTGTT
Et-2	ATCATCATCATCATCAATTCACATTGTCTCTCGAAAATGGAAAAAG
Et-3	GATGATGATGATGATGAATTGTGTAAATAATAAAGCCCCACAGG
Ext-1	GTGAGCTCGAAGTGACATCACAAAGTTGAAG
Ext-2	CAGAATTCGTCATAACTGTAGAAATGATTCC
Ext-3	GTGAGCTCAAAGCAGAATGCTGAGCTA
Ext-4	CAGAATTCGGCTAGCTAGAGAGAGGAGGAGAAT
Ext-6	CAGAATTCGAAACTGAACAAAACATACACAATGACAG
Ext-8	TGCGAGCTCCTGCATAAGAATATACATTGTGTG
gfp-3f	ATGGTCCTGCTGGAGTTCGTGACC
gfp-f	GTCCAGGAGCGCACCATCTTCT
gfp-p	FAM-CGGGTCTTGTAGTTGCCGTCGTCCTTG-TAMRA
gfp-r	GATGCCCTTCAGCTCGATGCGGTT
GUS-1	GTCTCGAGAACAATGTTACGTCCTGTAGAAACC
GUS-2	ACGAGCTCTCATTGTTTGCCTCCCTGC
NE-2	TTCTTCTTCTTCTTCTTTTCTCATGTTTGAACGATCGGGG
NE-2m	TTCTTCTTCTTCTTCTTTTCTCCGGCAACAGGATTCAATC
NE-3	GAAGAAGAAGAAGAAGAAAAGATTGGCAATAAAGTTTCTTAAGATTG
NE-3m	GAAGAAGAAGAAGAAGAAAAGAGTCTTGCGATGATTATCATATAATTTCTG
NE-4	TTCTTCTTCTTCTTCTTTTCTCAGGATTCAATCTTAAGAAAC
NE-5	GAAGAAGAAGAAGAAGAAAAGAGGTTGCCGGTCTTGCGATG
NElink-3	AGAAGAAAAGAGTAAGTTTCTTAAGATTGAATCCTGTTGC
Oli-1	GTGAGCTCGAATTTCCCCGATCGTTCAAACATTTG
Oli-2	GAAAAAGCCTTTTATTGCCAAATGTTTGAACGATCG
Oli-3	GCAATAAAAGGCTTTTTCCATTTTCGAGAGACAATG
Oli-4	CTTCTTCTTCTTCTTTTCTCATTGTCTCTCGAAAATG
Oli-5	AGAAAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA
Oli-6	CTTAAGAAACTTACTCTTTTCTTCTTCTTCTTCTTCTT
NET-2	ATCATCATCATCATCAATTCACATGTTTGAACGATCGGGG
NET-2m	ATCATCATCATCATCAATTCACCGGCAACAGGATTCAATC
NET-3	GATGATGATGATGATGAATTGTTTGGCAATAAAGTTTCTTAAGATTG

Name	Sequence (5' to 3')
NET-3m	GATGATGATGATGATGAATTGTGTCTTGCGATGATTATCATATAATTTCTG
NOSREco	CTCGAATTCCCGATCTAGTAAC
RS-2	CAGCCCTACAACTATATCCGATGATGTTTGAACGATCGGGG
RS-3	GAGTAAATATTACGAGTCGAGGTTGGCAATAAAGTTTCTTAAGATTG
RS-4	CCTCGACTCGTAATATTTACTCAGCCCTACAACTATATCCGATG
RS-5	CATCGGATATAGTTGTAGGGCTGAGTAAATATTACGAGTCGAGG
RT-0	AACGAT TGAAGGAGCCACTCA
RT-1	ACTGAAGGCGGGAAACGAC
RT-2	GTGGTTGGCATGCACATACAA
RT-3	ATTTGCCTGCGCACCTGT
sNV-3f	AATGGTGCTAGCAGCGGTCCAC
sNV-f	CAATGTCCAAAGTGTGCAATTCC
sNV-p	FAM-TGATGGAAGACTTGTTGGCACC-TAMRA
sNV-r	CCTCTTATCTTGGCAACATGTGAC
VE-5	GAGAAAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAAGA
VE-6	TCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTTTCTC
Vet-5	GTGAATTGATGATGATGATGATGATGATGATGATGAATTGT
Vet-6	ACAATTCATCATCATCATCATCATCATCATCATCAATTCAC

CHAPTER 3

IMPROVED BEAN YELLOW DWARF VIRUS BASED VECTOR USING TOBACCO EXTENSIN TERMINATOR FOR HIGH LEVEL PROTEIN EXPRESSION IN PLANTS.

3.1 Abstract

Bean yellow dwarf virus (BeYDV) is a single-stranded circular DNA virus that belongs to the *Geminiviridae*, *Mastrevirus* genus. The previously developed BeYDV-based vector system has a great potential in rapid production of important pharmaceutical proteins such as antigens and antibodies in plants. This chapter describes further development of the BeYDV-based vector using tobacco Extensin (Ext) terminator. Extensin is a member of the hydroxyproline-rich glycoprotein (HRGP) superfamily and constitutes the major protein component of cell walls. Incorporation of the Ext terminator into the expression cassette of the BeYDV system increased the level of the Norwalk virus capsid protein (NVCP) expression up to 6-fold. Deletion analysis reveled that the 5' region of the Ext terminator containing three putative polyadenylation near upstream elements plays an important role in the enhancing effect. Moreover, the increased NVCP expression with Ext terminator was correlated with more efficient mRNA 3' end processing, as demonstrated by the decreased production of read-through transcripts, compared to the soybean *vspB* terminator. The optimized BeYDV vector will be an important tool for the rapid and robust production of pharmaceutical proteins in plants.

3.2 Introduction

Over the past years, bean yellow dwarf virus (BeYDV) has been developed as a recombinant protein expression system for the efficient expression of foreign proteins in plants (Huang *et al*., 2009, 2010; Chen *et al*., 2011). BeYDV is a member of the *Mastrevirus* genus of *Geminiviridae* and infects dicotyledonous plants. BeYDV contains a single-stranded circular DNA genome that replicates to a very high-copy number via a rolling circle mechanism in the nuclei of infected cell (Saunders *et al*., 1991; Palmer and Rybicki, 1998). The viral genome encodes the replication initiator proteins (Rep/RepA) and two structural proteins, the movement and coat proteins flanked by viral *cis*-acting elements, the long intergenic region (LIR) and the short intergenic region (SIR). The LIR contains the origin of rolling-circle replication and promoter elements, and the SIR contains transcription terminator and polyadenylation signals. BeYDV based expression systems involve the replacement of the viral structural genes encoding movement and coat proteins with an expression cassette for a gene of interest (Lico *et al*., 2008). The recombinant BeYDV vectors delivered to plant cells via *Agrobacterium*-mediated method results in high-level amplification of the DNA replicon that contains an expression cassette for a gene of interest, and thus increases levels of mRNA and protein of interest.

A number of vaccine antigens have been produced using the BeYDV expression system, such as Norwalk virus capsid protein (NVCP) and hepatitis B core antigen (HBcAg), which self-assembled into virus like particles (Huang *et al*., 2009). Recently, the production of multi-subunit proteins, for instance, protective

monoclonal antibody (mAb) against Ebola virus has been also achieved using the BeYDV system (Huang *et al*. 2010). While BeYDV vectors are successful for the expression of single or multi-component protein complexes, it has been noted that the expression level was relatively low compared to the level of the accumulated replicon DNA. For example, using the inducible transgenic BeYDV system that composed of the BeYDV replicon vector, and Rep providing vector under the control of alcohol-inducible promoter, Zhang and Mason showed the substantially increased BeYDV replicon, and up to 80-fold increased target gene mRNA, but moderately increased target protein (up to 10-fold) (Zhang and Mason, 2006). Similarly, with BeYDV vector based on the mild strain (BeYDV-m), Regnard and colleagues found 100-1000-fold increase of replicon accumulation versus 1.5-7-fold increase of target protein expression (Regnard *et al*., 2010). These observations suggested that the BeYDV expression system can be further improved by finding ways to increase the mRNA accumulation and/or the efficiency of translation. I hypothesized that it should be possible to achieve this by employing an efficient transcription terminator.

Transcription terminators regulate gene expression by modulating pre-mRNA 3' end processing, whereby RNA transcription is terminated and poly-adenosine [poly(A)] tail is added (Moore and Proudfoot, 2009). This process is triggered by the recognition of poly(A) signals, located in pre-mRNA 3' untranslated region by pre-mRNA processing complex (Millevoi and Vagner, 2010). Mammalian poly(A) signals are composed of four elements: the most conserved AAUAAA hexamer located 10-30 nucleotides upstream of the cleavage

and poly(A) site; U/GU-rich downstream element (DSE) found ~30 downstream of the poly(A) site; G-rich auxiliary downstream element (AuxDSE) located near downstream of the DSE; and U-rich upstream elements situated at variable distances upstream of the poly(A) site (Danckwardt *et al*., 2008). These signals are recognized by pre-mRNA 3' end processing complex, composed of the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulation factor (CstF), cleavage factors (CFIm and CFIIm), poly(A) polymerase (PAP), poly(A) binding proteins, and RNA polymerase II large subunit (Millevoi and Vagner, 2010). CPSF recognizes the AAUAAA element in coordination with CstF bound to DSE, and catalyzes cleavage. Following cleavage, CPSF interacts with CF Im for the recruitment of PAP which adds ~250 nt poly(A) tail to the cleaved mRNA. In plants, less conserved poly(A) signals are found, including a UG-rich far upstream element (FUE), an A-rich near upstream element (NUE), and a U-rich cleavage element (CE) (Hunt, 2008). The NUE is located ~10-30 nucleotides upstream of the poly(A) site and functions similar to the mammalian AAUAAA element. The FUE is found ~50 upstream of the poly(A) site, and the cleavage elements is composed of the cleavage site and surrounding two U-rich elements. In spite of the significant divergence in sequence of the poly(A) signals, homologues of most of the mammalian 3' processing factors have been found in plants (Mandel *et al*., 2008).

The efficiency of the pre-mRNA 3' end processing subsequently affects stability and the translation efficiency of the mRNA (Ingelbrecht *et al*., 1989). Thus, in plants, many transcription terminator sequences have been characterized

and used to optimize the transgene expression. Among others, nopaline synthase (NOS) terminator derived from the Ti plasmid of *Agrobacterium* tumefaciens, CaMV 35S terminator from cauliflower mosaic virus (CaMV), and soybean vegetative storage protein (VSP) 3' untranslated region (UTR) are those have been widely used in plant expression vectors. Potato proteinase inhibitor II (Pin II) terminator and Rubisco small subunit (rbcS) terminator are also known as strong terminators (Richter *et al*., 2000; Outchkourov et al., 2003).

The purpose of this study was to investigate an efficient terminator in coordination with BeYDV replication elements to boost the level of transgene expression. For this, I evaluated the tobacco extensin (Ext) terminator in comparison with other terminators. Ext is a member of the hydroxyproline-rich glycoprotein (HRGP) superfamily of plant cell wall proteins that function in diverse aspects of plant growth and development (Showalter, 2010). I reported in Chapter 2 that the Ext terminator enhances transgene expression in non-replicating vector by increasing mRNA accumulation via efficient regulation of the transcription termination and mRNA 3' end processing. In this chapter, I tested the effects of the Ext terminator in the BeYDV vector on the expression of green fluorescence protein (GFP) and NVCP. The protein analyses revealed that the BeYDV vector with Ext terminator increases transgene expression up to 6-fold. The enhanced expression was correlated with the increased efficiency of the mRNA 3’ end processing, as demonstrated by reduced levels of read-through transcripts. I propose that the improved BeYDV vector system containing the Ext

terminator will facilitate the economical production of valuable recombinant proteins in plants.

3.3 Materials and Methods

3.3.1 Vector construction

pBYGFP.REF was obtained by replacing VSP terminator in pBYGFP.R (Huang et al., 2009) with Ext terminator. A SacI/NdeI fragment from the pBY027.EF (described in Chapter 2) was ligated with SacI/FseI and FseI/NdeI fragments of pBYGFP.R (Huang *et al.,* 2009). In the construct pBYGFP.REF, the Ext 3' UTR was placed downstream of the GFP gene, which was under the control of the CaMV35S promoter and TEV 5' UTR. A diagram of the T-DNA of the vectors is shown in Figure 3.1. Constructs pBYNVCP.R and pBYNVCP.REF were generated by replacing the GFP coding sequence of the pBYGFP.R and pBYGFP.REF with NVCP gene coding sequence from psNV210 (Zhang *et al*., 2006) using XhoI and SacI sites. pBYGFP.ET2 was generated as follows. A PCR fragment spanning 1-464 nt (considering the T of the Ext gene stop codon as -3) of Ext 3' UTR was amplified from pBY027.EF using primers Ext-1 and Ext-6, and cloned into pBY027 via SacI/EcoRI sites to give pBY027.ET2. A SacI/NdeI fragment from pBY027.ET2 was replaced with a SacI/NdeI fragment from pBYGFP.R to yield pBYGFP.RET2. To obtain intron-less constructs pBYGFP.REU and pBYGFP.REU2, fragments of Ext 3' UTR spanning 257-731 nt and 252-464 nt were PCR amplified using primer sets Ext-3/Ext-2 and Ext-3/Ext-6, respectively. The amplified fragments were inserted into pBY027 using

SacI/EcoRI sites to give pBY027.EU and pBY027.EU2. SacI/NdeI fragments from the resulting clones were used to replace the corresponding fragments of the pBYGFP.REF and pBYGFP.RET2 to give pBYGFP.REU and pBYGFP.REU2. To create pBYGFP.REI, which contains the first 23 nt and the intron of Ext 3'UTR at the 5’ of VSP terminator, a fragment spanning 1-251 nt of Ext terminator was amplified from pBY027.EF using primers Ext-1 and Ext-8 and inserted to pBYGFP.R opened with SacI. Replacement of GFP coding sequence of constructs pBYGFP.REI, pBYGFP.REU, pBYGFP.RET2, and pBYGFP.REU2 with NVCP coding sequence using XhoI and SacI sites resulted in constructs pBYNVCP.REI, pBYNVCP.REU, pBYNVCP.RET2, and pBYNVCP.REU2. Constructs pBYGFP.RNOS and pBYGFP.R35S were cloned as follows. The NOS terminator digested from pBI101 (Jefferson *et al*., 1987) or the CaMV35S terminator amplified from pRTL2-GUS (Carrington & Freed, 1991) using primers 35STm-1 and 35STm-2 were inserted into pBY027 opened with SacI and EcoRI, resulting pBY027.NOS or pBY027.35S. A SacI/NdeI fragment from pBY027.NOS or the pBY027.35S was used to replace VSP terminator of pBYGFP.R to yield pBYGFP.RNOS or pBYGFP.R35S. pBYGFP.RSIR, which does not contain terminator sequence other than the SIR of BeYDV, was made by replacing SacI/NdeI fragment of pBYGFP.R with a SacI/NdeI fragment from pBY027.SD12. pBY027.SD12 was obtained by inserting an adaptor (annealed primers SD12-Sm and SD12-Rm) into pBY027 opened with SacI and EcoRI. Constructs pBYNVCP.RNOS, pBYNVCP.R35S, and pBYNVCP.RSIR were obtained by replacing the GFP gene from constructs pBYGFP.RNOS,

pBYGFP.R35S and pBYGFP.RSIR with NVCP gene from psNV210 using XhoI and SacI sites.

3.3.2 Agroinfiltration procedure

Greenhouse-grown *Nicotiana benthamiana* plants (6 to 8 week-old) were infiltrated with *Agrobacterium tumefaciens* strain LBA4404 containing the experimental constructs. The leaf infiltration procedure is described previously (Huang and Mason, 2004). Briefly, the bacteria grown overnight at 30˚C were pelleted by centrifugation for 5 min at 5,000*g* and then resuspended in infiltration buffer [10 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH5.5 and 10 mM MgSO4] to OD600 = 0.2. The resulting bacterial suspensions were infiltrated into fully expanded leaves by using a syringe without needle through a small puncture. Infiltrated plants were incubated in a growth chamber for indicated time period.

3.3.3 GFP analysis

Total leaf protein was prepared from leaf samples harvested at specified days of post infiltration (DPI) with extraction buffer (25mM sodium phosphate buffer, pH 6.6, 100mM NaCl, 1mM EDTA, 0.05% Triton X-100, 50mM sodium ascorbate, and 10 μg/ml leupeptin) using a FastPrep machine (Bio101). Homogenates were centrifuged at 13,000*g* for 10 min, and cleared supernatants were used for protein assays. The protein concentrations from the leaf samples were determined using Bradford reagent (Bio-Rad) with bovine serum albumin (BSA) as standard. A serial 2-fold dilution of GFP samples were made using phosphate buffered saline (PBS, 137 mM NaCl, 2.6 mM KCl, 10 mM Na_2HPO4, and 1.8 mM KH_2PO4, pH 7.4) and 50 μL of each sample was added to black-wall

96-well plates (Corning), in duplicate. The GFP Fluorescence intensity was examined on a microplate reader (Molecular Device Co, Spectra Max M2) using the excitation and emission wavelength of 485nm and 538nm, respectively, at room temperature. The known concentration of *E. coli* expressed and purified GFP ranging from 5 to 100 ng were used to generate the standard curve. The reading of negative control, extract of uninfiltrated plant leaf, was subtracted before graphing.

For GFP SDS-PAGE analysis, 15µg of total soluble protein (TSP) was separated on 4-15% polyacrylamide gels in the presence of 0.1% SDS with or without 0.5 M DTT as a reducing reagent. The GFP on the non-reducing gel was visualized on the UV light (365 nm). The reducing gel was stained with PageBlue protein staining solution (Fermentas) to show the normalized sample loading.

3.3.4 ELISA

Total leaf proteins were extracted from leaf materials harvested at 4 DPI and normalized using a Bradford assay with BSA as standard. NVCP concentration was analyzed by sandwich ELISA as described by Mason *et al.* (1996). Briefly, a rabbit polyclonal anti-NVCP antibody was bound to 96-well plates, and the plates were blocked with 5 % nonfat dry milk (DM) in PBS. After washing the wells with PBST (PBS with 0.05 % Tween 20), the plant extracts were added and incubated. The bound NVCP were detected in succession with guinea pig polyclonal anti-NVCP antibody and goat anti-guinea pig IgG-horseradish peroxidase conjugate. The plate was developed with TMB substrate

(Pierce) and the absorbance was read at 450 nm. Insect cell-derived recombinant NVCP (i-rNV) (Jiang *et al*., 1992) was used as the reference standard.

3.3.5 Western blotting

15µg of TSP was separated on 4-15% polyacrylamide gels and either transferred to PVDF membrane or stained with PageBlue protein staining solution (Fermentas) to show the normalized sample loading. The protein transferred membranes were blocked with 5% DM/PBST for 1 h at 37˚C and probed in succession with guinea pig anti-(i-rNV) and goat anti-guinea pig IgG-horseradish peroxidase conjugated (Sigma) each diluted 1:10,000 in 1% DM/PBST. Bound antibody was detected with ECL reagent (Amersham).

3.3.6 Quantification of GFP gene

To quantify the GFP gene copy number, Real Time PCR was performed on 25ng samples of purified genomic DNA. The genomic DNA was purified from infiltrated leaf samples harvested at a specified DPI using the Spin Plant Mini Kit (Qiagen) as recommended by manufacturer. Real-Time PCR was performed on an IQ5 Real-Time PCR Detection System (Bio-Rad) using *Taq* DNA polymerase with ThermoPol Buffer (NEB). GFP gene specific primers (GFP-f and GFP-r) and custom made Taqman probe GFP-pro (Integrated DNA Technologies) were used. Gene copy number was deduced based on the standard curve generated using the plasmid pBY027 with concentration of 10^8-10^1 copies/reaction.

3.3.7 Quantification of mRNA

Total RNA was prepared using Plant RNA Reagent (Invitrogen) according to the manufacturer's protocol and residual DNA was removed by DNAfree system (Ambion). Aliquots of 1 μg of total RNA were subjected to first-strand cDNA synthesis with oligo dT_{20} primer using a Superscript III First-Strand Synthesis System (Invitrogen) according to the manufacturer's protocol. Real-Time PCR was performed as described previously. For GFP transcripts, gene specific primers (GFP-f and GFP-r) and custom made Taqman probe (GFP-pro, Integrated DNA Technologies) were used. As an internal control, *N. benthamiana* translation elongation factor 1 alpha (*EF1a*, Acession number AY206004) was used (primers EF1f, EF1r and EF1p, Integrated DNA Technologies). Each sample was measured in triplicate for GFP transcripts and an internal reference gene. Transcript levels of GFP and EF1*a* were quantified using separate standard curves prepared with plasmid pBY027 and pCR4-Topo-EF1*a* (see below), respectively. The relative quantification of GFP transcripts was normalized against *EF1a* transcripts. The construct pCR4-Topo-EF1*a* was constructed as follows. RT-PCR was performed with cDNA from wild-type *N. benthamiana* RNA using EF1f and EF1r primers, which amplify 119bp (167nt-285nt) fragment of EF1*a* gene. The amplified product was cloned to a PCR cloning vector, pCR4-Topo (Invitrogen) and used to generate standard curve.

3.3.8 RT-PCR

First-strand cDNA was synthesized from 1 μg of total RNA with either random primers or the gene specific primer (rRep3-4 or SNV-3f). To detect read-

through transcripts PCR was performed with three sets of primers on the random primed cDNA. A forward primer, SNV-3f, specific to the NVCP gene (positioned at 138 nt upstream of stop codon) was paired with four reverse primers SIR5R, rRep3-4, and rRep3-2, positioned at 38, 207, and 544, respectively, downstream of the EcoRI site on the expression vectors. Primers SNV-3f and rRep3-4 were used for the PCR on the cDNA synthesized with gene specific primers. Resulting PCR products were separated on agarose gel with SYBR Safe DNA gel stain.

3.4 Results

3.4.1 The effect of replicon amplification on mRNA and protein accumulation

I have previously shown that BeYDV vector led a substantial increase of DNA replicon but moderate mRNA and relatively low target protein accumulation, when alcohol inducible transgenic system was used (Zhang and Mason, 2006). In this study, I tested the correlation of BeYDV replicon amplification with mRNA and protein accumulation using *Agrobacterium*-mediated transient expression system. A single-replicon BeYDV vector containing a built-in Rep/RepA cassette, pBYGFP.R (Zhong *et al*., 2009), was compared with non-replicating vector, pPSGFP.VSP (Fig. 3.1). For both constructs, expression of target protein, is driven by CaMV 35S promoter with the tobacco etch virus (TEV) 5’ UTR, and terminated by VSP 3’ UTR. Four independent plants were infiltrated with *Agrobacterium* harboring either pBYGFP.R or pPSGFP.VSP, and leaf samples were harvested over a 5-day period for DNA, mRNA, and protein analyses.

I first measured the copy number of GFP DNA in total genomic DNA samples using real-time PCR with gene specific primers (GFP-f and GFP-r) and custom made Taqman probe (GFP-pro, Integrated DNA Technologies). Plasmid DNA containing the GFP gene was used to plot the standard curve (see Materials and Methods). As expected, I observed efficient amplification of the BeYDV replicon (Fig. 3.2A). GFP gene copy number was increased from 10^6 to 10^9 per 25 ng genomic DNA by 3 DPI, and the increased DNA level was maintained throughout the remainder of the time trial. By contrast, the copy number of the GFP gene in the non-replicating vector pPSGFP.VSP remained constant at the input level, 10^6 copy/25ng genomic DNA.

Next I measured the relative GFP mRNA level using real-time RT-PCR. Total RNA samples were purified from the harvested leaf materials, and used for cDNA synthesis using oligo-dT_{20} primer. Real-time PCR was performed and GFP mRNA copy numbers were deduced. To normalize the total mRNA, *N. benthamiana* translation elongation factor 1 alpha (EF1a, accession number AY206004) was used as an internal control. Figure 3.2B shows that the level of GFP mRNA increases in response to the BeYDV replicon amplification, consistent with previous data (Zhang and Mason, 2006). For the construct pBYGFP.R, the mRNA level reached the maximum value at 4 DPI, and substantially decreased after that. For the non-replicating vector pPSGFP.VSP, the maximum mRNA copy was reached at 3 DPI. The maximum GFP mRNA level of the pBYGFP.R was 12.5-fold higher than that of the pPSGFP.VSP.

I then analyzed the level of GFP expression. Total protein was extracted from the leaf samples and the concentration of GFP was measured using a fluorometer. *E.coli* expressed and purified GFP was used to generate the standard curve. The time course profile of GFP expression was consistent with that of the BeYDV replicon amplification and mRNA accumulation (Fig. 3.2C). Construct pBYGFP.R showed its maximum GFP level at 4 DPI, whereas pPSGFP.VSP reached to the maximum GFP level at 3 DPI. The maximum GFP level of the pBYGFP.R construct was approximately 15.5-fold higher than that from the pPSGFP.VSP. As seen in previous studies, the level of target protein expression was substantially lower than replicon accumulation (850-fold increase of replicon vs. 15.5-fold increase of GFP expression).

3.4.2 Enhanced BeYDV vector using Ext terminator

I hypothesized that the low GFP protein accumulation relative to replicon copy number (Fig. 3.2) is due in part to either inefficient transcription or mRNA instability. I investigated the potential of tobacco extension (Ext) terminator (accession number D13951) to increase transient transgene expression in the context of BeYDV vector. For this, the VSP terminator of pBYGFP.R was replaced with the Ext terminator resulting pBYGFP.REF (Fig. 3.1). *Agrobacterium* strains containing either pBYGFP.R or pBYGFP.REF were infiltrated into leaves of *N. benthamiana* plants, and leaf samples were harvested at 3, 4, and 5 DPI. The measurement of GFP fluorescence revealed that the amount of GFP produced from pBYGFP.REF was approximately 2-fold higher than that from the pBYGFP.R at all time points (Fig. 3.3A). To test whether the

enhancing effect of the Ext terminator was gene specific, I replaced the GFP gene of pBYGFP.R and pBYGFP.REF with the NVCP coding sequence, resulting pBYNVCP.R and pBYNVCP.REF (Fig. 3.1). Quantification of the transiently expressed NVCP 4 DPI samples showed that construct pBYNVCP.REF yields 6-fold higher NVCP expression compared to pBYNVCP.R (Fig. 3.3B). From these data, I conclude that the Ext 3' UTR is more efficient in supporting transgene expression compared to the VSP terminator in a manner that is independent of the transgene used. Moreover, these data suggest possible further enhancement of the BeYDV vector through careful optimization of the expression cassette.

3.4.3 Ext terminator increases mRNA accumulation

The greater accumulation of transgene products mediated by the Ext 3' UTR could be caused by an increase in either mRNA level or translational efficiency. To determine whether the Ext 3' UTR affects the mRNA level, total RNA was isolated from leaf samples infiltrated with either pBYGFP.R or pBYGFP.REF and subjected to real-time RT-PCR analysis as described previously. The level of GFP mRNA observed was approximately 3.0-fold greater from the Ext 3' UTR construct (pBYGFP.REF) than from the VSP terminator construct (pBYGFP.R) (Fig. 3.4), suggesting that the *Ext* 3' UTR is involved in maintaining the stability of and/or increasing the synthesis of transgene mRNA.

3.4.4 Deletion analysis of the Ext terminator

An interesting feature of the Ext terminator is the presence of an intron sequence at its 5' end (24-249 nt). To assess whether the intron plays a role in the enhancing effect of the Ext terminator, I made a deletion construct, pBYGFP.REU,

lacking the intron (Fig. 3.5). This mutant was constructed by replacing the Ext terminator of pBYGFP.REF with the fragment of Ext 3' UTR spanning 252-731 nt. The mutant construct was introduced into *N. benthamiana* using *Agrobacterium*. Three days after agroinfiltration, the infiltrated leaf was collected for protein extraction. Fluorometric GFP analysis of the mutant construct, pBYGFP.REU, showed approximately 25% decreased expression compared to the intron-containing construct, pBYGFP.REF (Fig. 3.5B). Although decreased, the expression level of pBYGFP.REU was still 1.5-fold higher than that of pBYGFP.R, the construct with VSP terminator. The influence of the Ext intron was also assessed on NVCP expression, by replacing the GFP gene of pBYGFP.REU with NVCP, resulting pBYNVCP.REU. Deletion of the intron had no significant effect on the NVCP expression (Fig. 3.5C). Taken together, these data suggest that the enhancing effect of the Ext terminator is mainly derived from the 3' UTR sequence, and not from the intron.

Ext terminator contains four putative NUEs. In order to examine the contribution of the NUEs to protein expression, I constructed truncation mutants in which one of the four NUEs was deleted (Fig. 3.5A). These mutants were constructed by replacing the Ext terminator of pBYGFP.REF or pBYNVCP.REF with the fragment of Ext 3' UTR spanning 252-464 nt, yielding constructs pBYGFP.RET2 and pBYNVCPP.RET2. As shown in Figure 3.5, the deletion of the fourth NUE caused approximately 20-25% decrease in expression of either GPF or NVCP. The remaining 75-80 % of expression levels represent approximately 1.5- to 5-fold higher GFP and NVCP expression, respectively,

compared to those of VSP terminator constructs. These data suggest that Ext terminator sequence between 252-464 nt provides most of the enhancing effect, while sequence between 465-731 nt further supports optimal terminator function.

The effect of the Ext intron was further investigated using the truncated Ext terminator. For this, intron sequences present in pBYGFP.RET2 and pBYNVCP.RET2 were deleted to generate pBYGFP.REU2 and pBYNVCP.REU2. Comparison of GFP and NVCP expression between intron-containing and intron-less constructs showed no significant differences on protein expression levels (Fig. 3.5B and 3.5C). These data are consistent with the conclusion that the Ext sequence between 252-464 nt, which contains three putative NUEs, contains the main elements that enhance transgene expression.

To determine if the Ext intron functions in combination with other terminators, we fused the Ext intron to the 5' of VSP terminator (pBYGFP.REI and pBYNVCP.REI). The addition of the intron did not significantly change the level of GFP or NVCP expression (Fig. 3.5B and 3.5C), confirming the conclusion that the intron contributes little to the enhancing effect.

3.4.5 Comparison of Ext 3' UTR with other terminators

Next, I compared the efficiency of Ext terminator with NOS and CaMV 35S terminator. For this, VSP terminator from pBYGFP.R and pBYNVCP.R was replaced with a NOS or CaMV35S terminator (Fig. 3.6A). As a control, VSP terminators from pBYGFP.R and pBYNVCP.R were removed to give pBYGFP.SIR and pBYNVCP.SIR. It is important to note that in the BeYDV vectors, the bidirectional terminator of BeYDV, SIR, is located at downstream of

the expression cassette. The constructs with various terminators were infiltrated into *N. benthamiana* leaves and leaf samples were harvested at 3 DPI for GFP and NVCP expression tests. Ext terminator yielded the highest GFP expression level, as showed either by SDS-PAGE followed by UV illumination (365nm) or fluorometric quantification (Fig. 3.6A and 3.6B). Similarly, the Ext terminator resulted in the highest NVCP expression, as demonstrated by Western blotting and ELISA (Fig. 3.7A and 3.7B). These results indicate that the Ext 3' UTR is more effective than previously established terminators in facilitation of transgene expression.

3.4.6 Ext 3' UTR mediates efficient transcription termination.

I hypothesized that the increased protein expression could be related to efficient transcription termination, because improperly terminated mRNA is targeted by RNA-dependent RNA polymerase RDR6-mediated RNA silencing (Luo and Chen, 2007). To test the hypothesis, I analyzed the NVCP transcripts by RT-PCR with three sets of primers. One sense primer (SNV-3f) specific to the NVCP coding region was paired with one of three antisense primers (SIR5R, rREP3-4, or rRep3-2) corresponding to the sequence downstream of the terminator sequences in the expression vectors (Fig. 3.8A). SIR-5R binds to the SIR, while rRep3-4 and rRep3-2 bind to the BeYDV C1/C2 coding sequence (Fig. 3.8A). RT-PCR using the primer sets would produce DNA fragments of expected size only when read-through RNA transcripts are present.

First, I analyzed transcription termination using cDNA, synthesized with random primers (Fig. 3.8B). When the primer set SNV-3f/SIR5R or SNV-

3f/rRep3-4 was used, all constructs except the pBYNVCP.REF showed RNA transcription read-through bands. When SNV-3f/rRep3-2 was used, construct SIR produced an intense read-through band and constructs pBYNVCP.RNOS and pBYNVCP.R35S produced faint bands of expected size in addition to smaller faint bands. In contrast, the construct pBYNVCP.REF did not produce any detectable read-through band with any primer set, indicating efficient transcription termination.

Second, to exclude the possibility that the read-through PCR products were amplified from the Rep/RepA transcripsts driven by the LIR promoter, I performed RT-PCR using cDNA synthesized with gene specific primer rRep3-4 or SNV-3f (Fig. 3.8C and 3.8D). For amplification of read-through transcripts, primer set SNV-3f/rRep3-4 was used. When cDNA primed with rRep3-4 was used, all constructs except the pBYNVCP.REF produced the read-through band of expected size, similar to the random primer data. In contrast, when cDNA synthesized with SNV-3f was used, only construct pBYNVCP.RSIR showed a faint band. These data indicate that the read-through transcripts are mainly derived from the CaMV35S/TEV5'-NVCP-terminator cassette not from the LIR-Rep-SIR cassette. Together, these data suggest that the Ext 3' UTR directs more efficient mRNA 3' end formation in the BeYDV vector than other terminators tested.

3.5 Discussion

In this study, I report an optimized BeYDV-based vector system, using

tobacco Ext terminator for the expression cassette. The delivery of the BeYDV vector to leaf cells via agroinfiltration results in a very high level accumulation of the DNA replicon, and thereby high yields of mRNA and protein of interest (Chen *et al*., 2010). Using the BeYDV vector, a number of vaccine antigens have been produced, including NVCP, HBV core, and HIV-1 p24 (Huang *et al*, 2009; Regnard *et al*., 2010). Although the BeYDV vector markedly improved target protein expression, it has been noted that the DNA replicon was amplified to a greater extent than the mRNA (Zhang and Mason, 2006; Regnard *et al*, 2010). In agreement with the previous findings, in this study, I showed that BeYDV DNA replicon accumulated to a level 850-fold higher than that of the non-replicating vector. Nevertheless, the mRNA and protein of interest were increased by only ~12.5- and 15.5-folds, respectively (Fig. 3.2). As suggested elsewhere, this may be a result of saturation of the cellular transcription and translation capacity when the transcription template (replicon) is extremely abundant (Zhang and Mason, 2006; Regnard *et al*, 2010). However, considering that the concomitant use of the silencing suppressor P19 improved HBc mRNA accumulation 2- to 4-fold and enhanced HBc protein yield ~4-fold (Huang *et al*., 2009), I hypothesized that it could be possible to further increase the target protein expression by increasing mRNA accumulation and/or translatability.

My approach in this study was to optimize the terminator, which regulates gene expression by mediating RNA transcript termination and mRNA 3' end processing, thereby playing an important role in mRNA stability and translatability (Zhao *et al*., 1999 Moore and Proudfoot, 2009). Previously, I

explored various terminators in combination with the BeYDV vector system: I have used VSP terminator for the expression of GUS protein (Mor *et al*., 2003); rbcS 3' UTR for GFP and capsid proteins of hepatitis B and Norwalk viruses (Huang *et al*., 2009); and pin II terminator for monoclonal antibody against Ebola virus (Huang *et al*., 2010). Here, I provide data on the direct comparison of the various terminators to evaluate their efficiency on transgene expression in combination with the BeYDV vector system. I found that tobacco Ext terminator is the most efficient among tested terminators, enhancing GFP and NVCP expression by up to 6.2-fold. This led us to test the correlation of mRNA accumulation and transgene expression when Ext terminator was used. The level of GFP mRNA was approximately 3.0-fold greater from the Ext 3' UTR construct than from VSP terminator construct (Fig. 3.4), indicating that the Ext 3' UTR may be involved in the maintaining the stability and/or increasing the synthesis of transgene mRNA.

BeYDV vector has received considerable attention for the development as a co-expression vector. The non-competitive replication of the BeYDV replicons in the same cells enabled the efficient co-expression of two or more genes (Huang *et al*., 2010). For example, Huang *et al* showed robust expression of the heavy and light chains of the anti-Ebola virus mAb 6D8, using either co-delivery of separate vectors or delivery of a single vector that contained two tandem arranged replicons. Following the Ext terminator, the CaMV 35S terminator also enhanced GFP and NVCP expression by 1.9-4.5-fold (Fig. 3.6 and 3.7). The direct comparison on the efficiency of the various terminators will provide better

options for the choice of terminators for the economic use of limited cellular factors. Often times, co-expression of multiple proteins requires regulation of abundance of each protein (eg., influenza hemagglutinin, neuraminidase and matrix M1 proteins), which might be achieved by the choice of terminators.

Ext terminator is unusual in that it contains an intron at its 5' end. Based on a bioinfomatic analysis of 32,955 annotated *Arabidopsis thaliana* protein-coding genes, only 5.6% of the 3' UTRs include an intron, while 72% of protein-coding sequences and 19.9% of the 5' UTRs contain introns (Chung *et al*., 2006). In Chapter 2, I reported that the Ext intron has a regulatory function on transient transgene expression in a context dependent manner. Specifically, it showed a deleterious effect (up to 70% decrease) when linked with Ext terminator but showed a positive effect (up to 50% increase) when used in combination with NOS terminator. Similar to the data, Menossi *et al*. reported the tissue type dependent regulatory function of Maize HRGP gene 3' UTR: it increased GUS expression in fast-growing coleoptiles, but decreased GUS expression in maize suspension cells (Menossi *et al*., 2003). In order to continue the study on the Ext intron, I tested GFP and NVCP expression controlled by Ext terminator with or without Ext intron in the BeYDV vector system. Unexpectedly, Ext intron did not show a deleterious effect, rather showed a positive effect for GFP expression (Fig 3.5). At present, I do not understand the mechanism causing this difference between non-replicating and replication vector systems. Further studies of the Ext intron will give us valuable insight into the intron-mediated regulation of gene expression.

I provided detailed studies on the sequence elements of Ext terminator. Ext terminator contains four NUEs and a 45-base long polypurine sequence (PPS). In our previous report, I showed that the 4th NUE plays an important role in the function of the Ext terminator in coordination with the other NUEs: the deletion of the 4th NUE reduced the protein expression level by 90%, but the 4th NUE alone without the other NUEs was not sufficient to support gene expression (Chapter 2). Interestingly, in BeYDV vector, the deletion of the 4th NUE caused only 20~25% decrease of GFP and NVCP expression. I speculate that this is because of the presence of SIR in the BeYDV vector system: the Ext terminator is fused to the SIR which functions as a bidirectional terminator (Jeske, 2009). I hypothesisize that when the 1st terminator provides weak activity (e.g., the deletion mutants), the second terminator supports the function of the first terminator. This hypothesis is supported by Luo and Chen's study showing the double terminator effects: when a tandemly linked NOS and 35S terminator was used, higher GUS expression was achieved compared to the individual terminators (Luo and Chen, 2007).

Terminator competency can be determined by testing for transcription readthrough products. When a terminator is leaky, the RNA polymerase complex passes beyond the termination signals resulting in improperly terminated and unpolyadenylated transcripts. Such transcripts can trigger RDRP6-mediated RNA silencing, which leads to a cascade of mRNA degradation (Luo and Chen, 2007). Consistent with Chapter 2, I found that the Ext 3' UTR tightly controls the transcription termination and/or 3' end processing (Fig. 3.8). I conclude that the

decreased level of read-through transcripts supports mRNA stability and higher accumulation, thereby increasing transgene expression by preventing activation of gene silencing.

While incorporation of the Ext terminator significantly improved target protein expression with the BeYDV replicon vector, further optimization might be possible. I observed mild necrosis in leaves infiltrated with BeYDV vectors upon extended incubation (~10 DPI, data not shown). As suggested elsewhere, an excessive accumulation of either Rep/RepA or replicons could be the cause of the necrosis (Zhang and Mason, 2006). Thus, optimization of Rep/RepA expression and replicon amplification may further increase the expression level by preventing unnecessary stress on plant cells. Moreover, the general cell death suppressor, AvrPtoB, may be co-expressed to prevent the necrotic response (Abramovitch *et al*., 2006). I also observed sudden decrease of GFP mRNA on 5 DPI (Fig. 3.2B). This may indicate the possible onset of gene silencing, which can be easily overcome by the use of anti-silencing element P19. As suggested by others, co-expression of molecular chaperones can be also beneficial to assist correct folding of the expressed proteins (Nuttall *et al*., 2002).

In conclusion, this paper describes an enhanced BeYDV expression system for a rapid and robust transgene expression in plants. This enhanced system provides a valuable tool for the production of pharmaceutical proteins in plants, including antibodies and vaccine antigens.

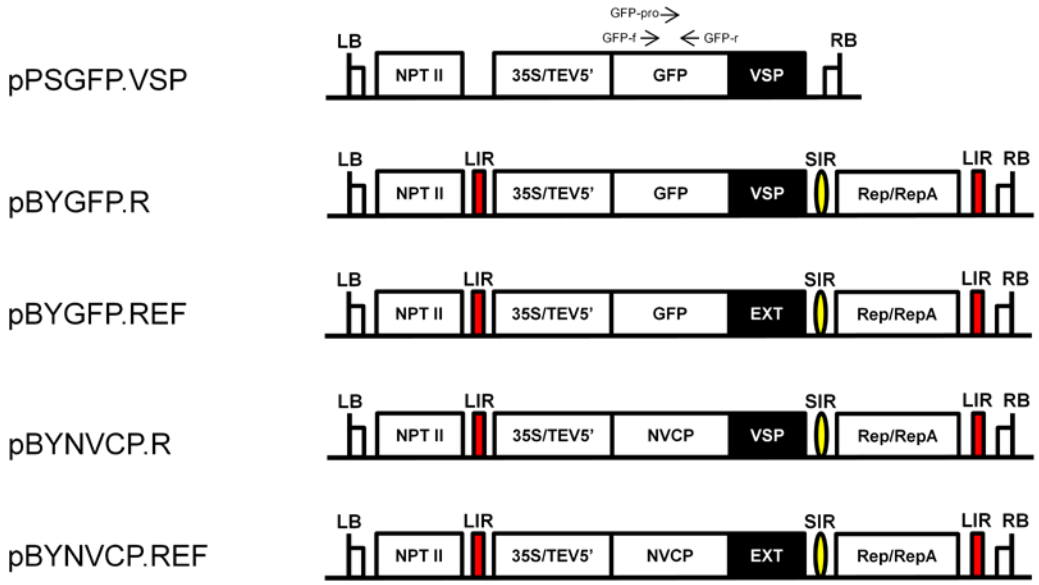


Figure 3.1 Schematic representation of the T-DNA regions of the vectors used in this study. pPSGFP.VSP, non-replicating vector with VSP terminator. pBYGFP.R, BeYDV vector with VSP terminator. pBYGFP.REF, BeYDV vector with Ext terminator. VSP, soybean vspB gene 3' element; Ext, tobacco extensin gene 3' element; GFP, green fluorescent protein gene; NVCP, Norwalk virus capsid protein, 35S/TEV 5', CaMV 35S promoter with tobacco etch virus 5' UTR; NPTII, expression cassette encoding nptII gene for kanamycin resistance; LIR, long intergenic region of BeYDV genome; SIR, short intergenic region of BeYDV genome; Rep/RepA, replication initiation protein gene of BeYDV; LB and RB, the left and right borders of the T-DNA region. Primers and probe used in the Real-time PCR are indicated with horizontal arrows.

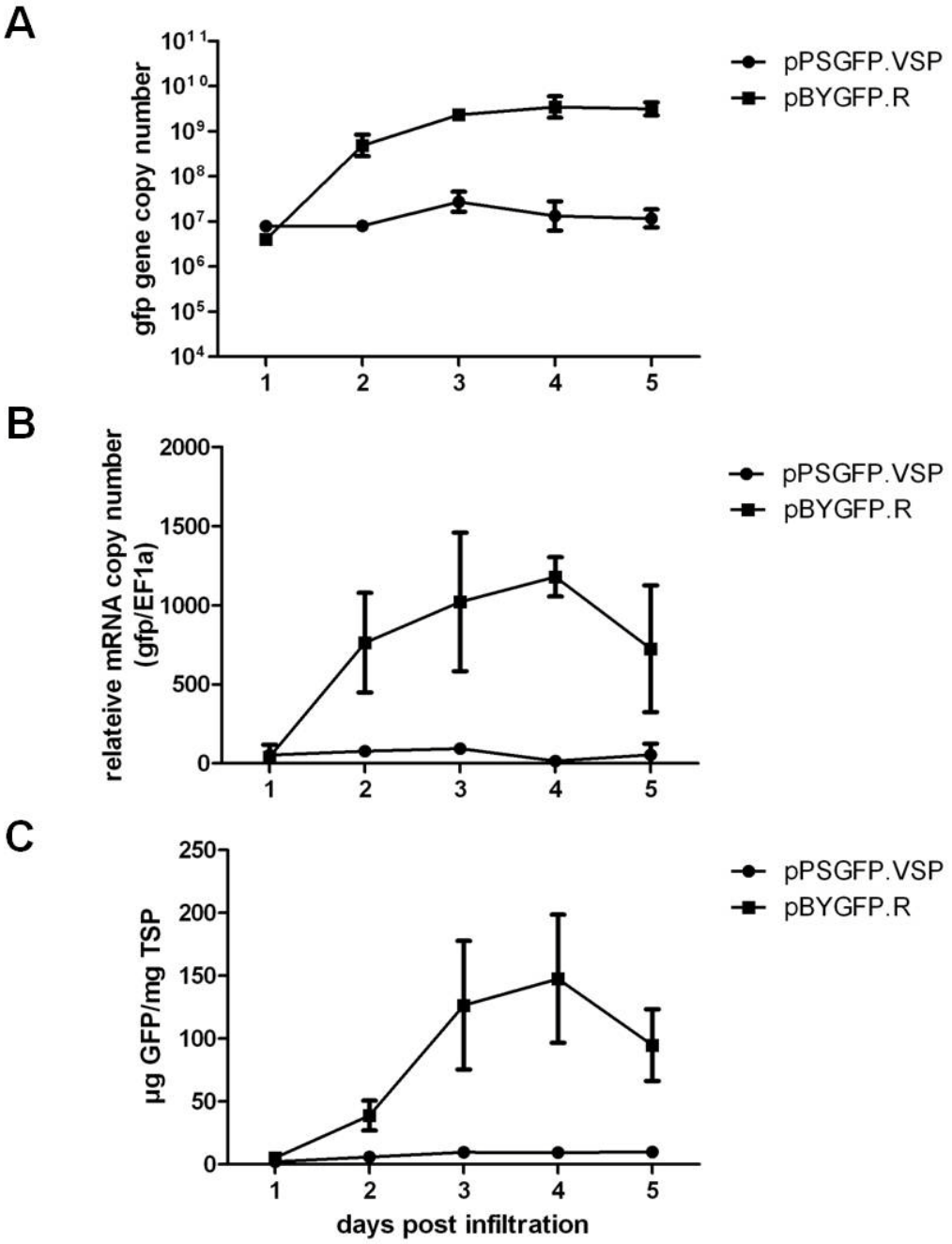

A
gfp gene copy number
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
1
2
3
4
5
pPSGFP.VSP
pBYGFP.R
B
relateive mRNA copy number (gfp/EF1a)
2000
1500
1000
500
0
1
2
3
4
5
pPSGFP.VSP
pBYGFP.R
C
µg GFP/mg TSP
250
200
150
100
50
0
1
2
3
4
5
days post infiltration
pPSGFP.VSP
pBYGFP.R

Figure 3.2 Effect of BeYDV replicon amplification on mRNA and protein accumulation. (A) GFP gene copy number per 25ng total genomic DNA measured by real-time PCR. (B) Relative GFP mRNA copy number measured by real-time PCR. (C) Fluorometric analysis of GFP expression. *Agrobacterium* strains harboring either the replicating PBYGFP.R vector or non-replicating pPSGFP.VSP vector were infiltrated and leaf samples were harvested over 5 days post infiltration. Data are means ± S.D. of samples from four independent infiltration experiments.

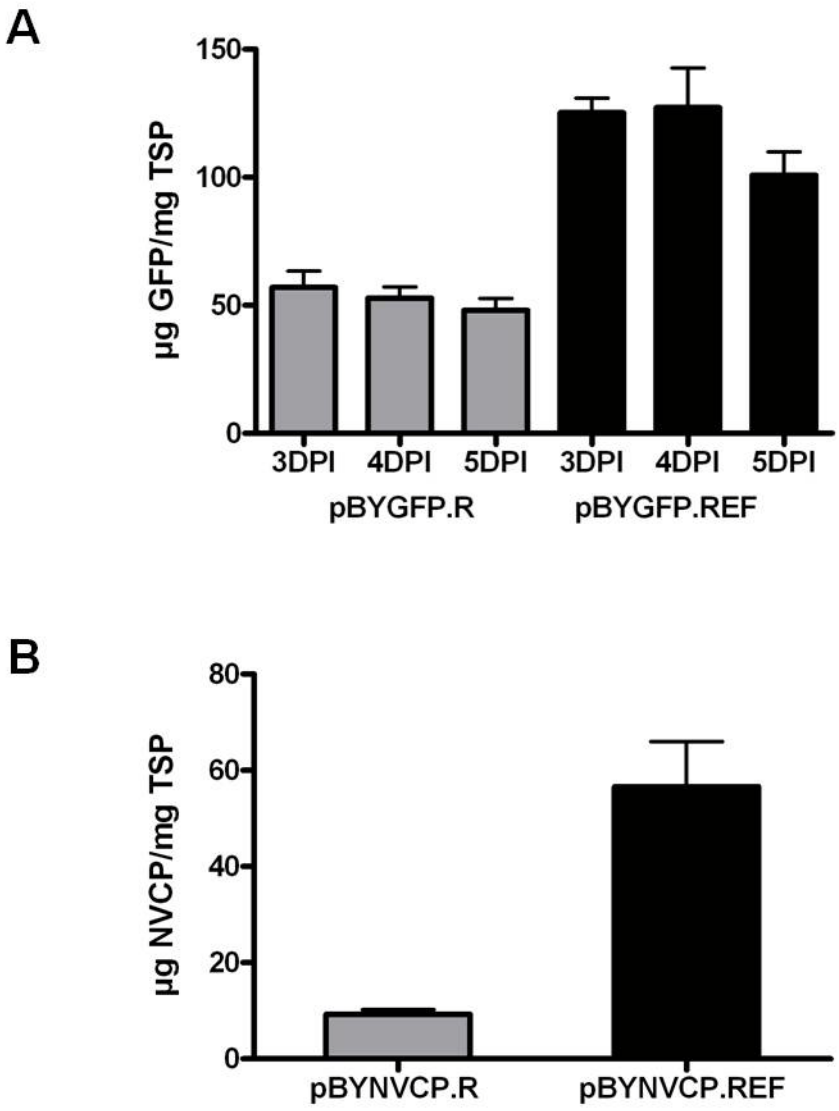


Figure 3.3 Ext terminator increases the level of transgene expression. (A) Fluorometric analysis of GFP expression. Extracts from *N. benthamiana* leaves infiltrated with constructs pByGFP.R and pByGFP.REF were analyzed for GFP expression by spectrofluorimetry using excitation and emission wavelengths of 485nm and 538nm. Columns represent means ± S.D. from six independently infiltrated samples. (B) Expression of NVCP. Extracts from *N. benthamiana* leaves infiltrated with constructs pBYNVCP.R and pBYNVCP.REF were analyzed for NVCP expression by ELISA. Columns represent means ± S.D. from nine independently infiltrated samples.

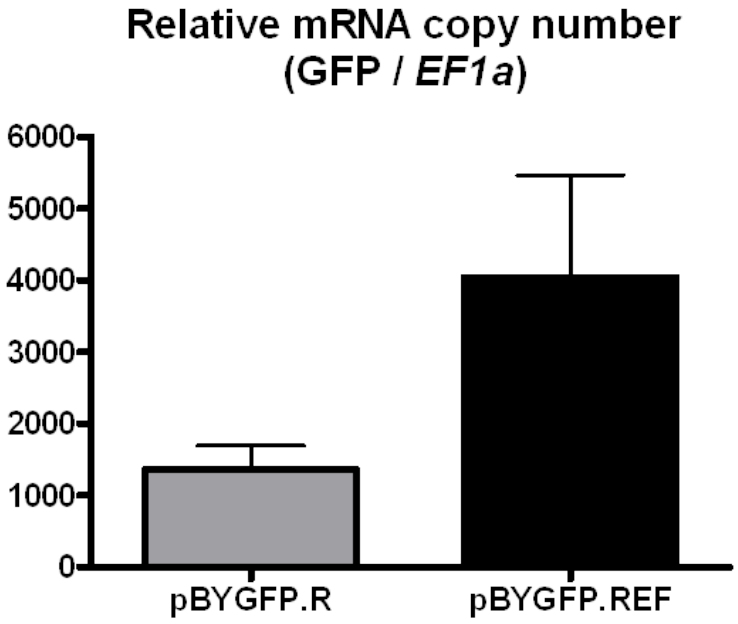


Figure 3.4 Ext terminator increases mRNA accumulation. Total RNA from *N. benthamiana* leaves infiltrates with constructs pBYGFP.R or pBYGFP.REF were analyzed for GFP mRNA accumulation by real-time RT-PCR. Translation elongation factor 1 alpha (*EF1a*) was used as an internal control. Relative mRNA copy numbers are shown. Columns represent means ± S.D. from four independently infiltrated samples.

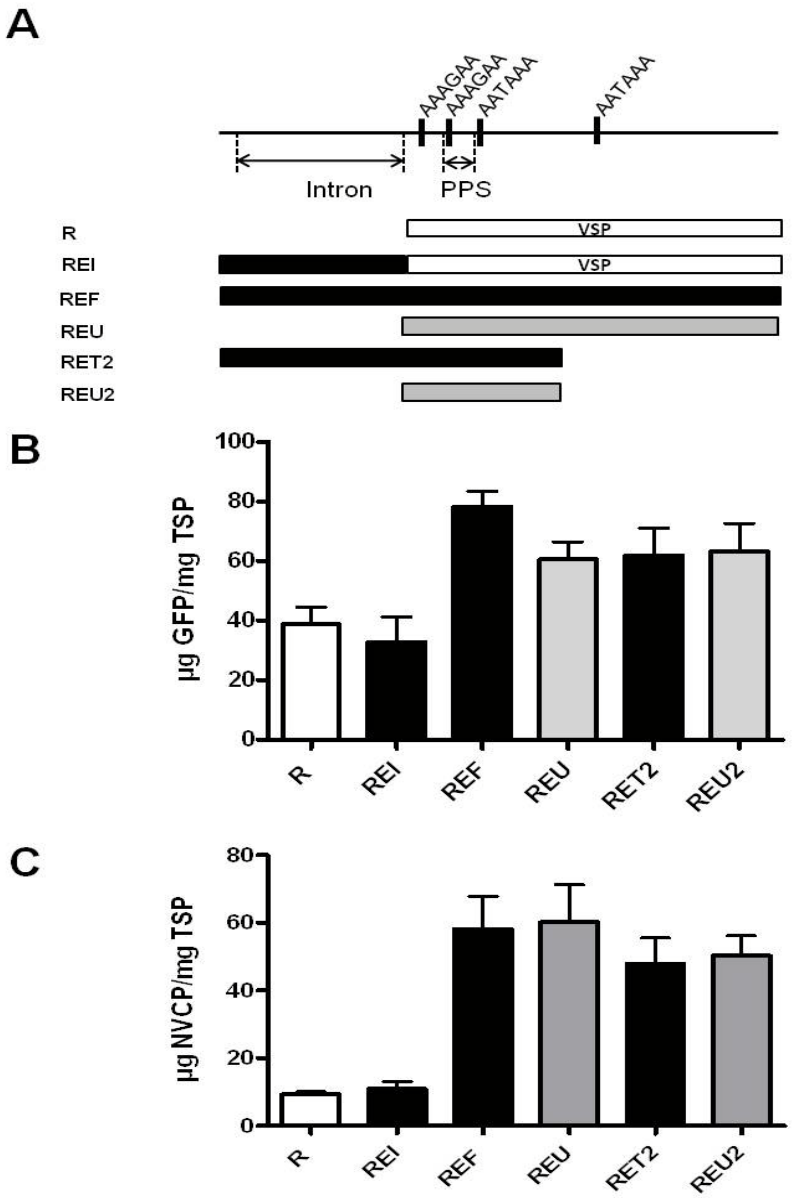


Figure 3.5 Deletion analysis of Ext terminator. (A) Schematic representation of *Ext* 3' UTR and deletion variants. Putative polyadenylation signals and positions of intron and poly-purine sequence are shown. R, pBYGFP.R or pBYNVCP.R; REI, pBYGFP.REI or pBYNVCP.REI; REF, pBYGFP.REF or pBYNVCP.REF; REU, pBYGFP.REU or pBYNVCP.REU; RET2, pBYGFP.RET2 or pBYNVCP.RET2; REU2, pBYGFP.REU2 or pBYNVCP.REU2. Extracts from *N. benthamiana* leaves infiltrated with indicated constructs were analyzed for GFP expression by spectrofluorometry (**B**) and for NVCP expression by ELISA (**C**), as described in materials and methods. Columns represent means ± S.D. from seven to nine independently infiltrated samples.

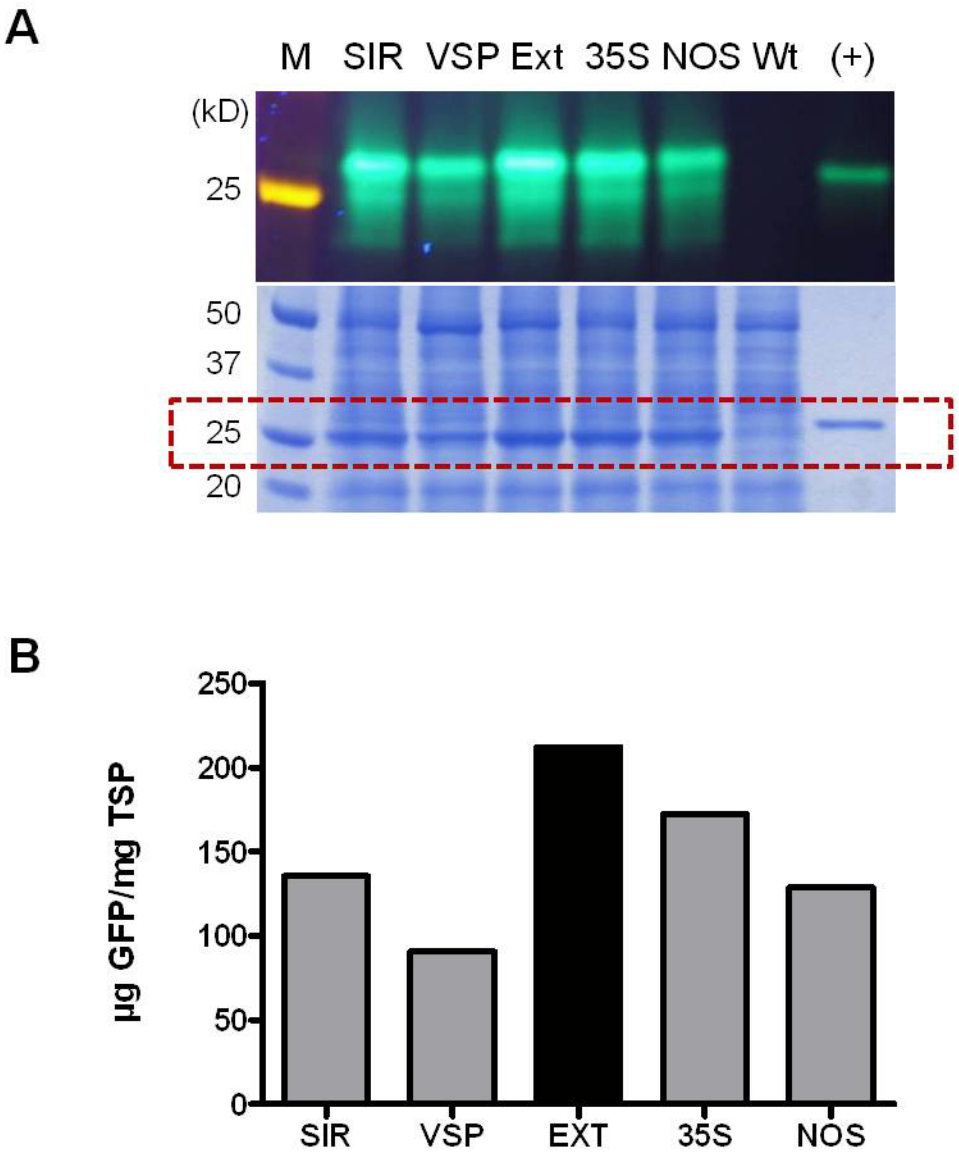


Figure 3.6 Comparison of terminataors for GFP expression. (**A)** Extracts from *N. benthamiana* leaves infiltrated with indicated constructs were analyzed by SDS-PAGE followed by either observation under UV light (upper) or Coomassie stain (lower). SIR, pBYGFP.RSIR; VSP, pBYGFP.R; EXT, pBYGFP.REF; 35S, pBYGFP.R35S; NOS, pBYGFP.RNOS; Wt, wild-type *N. benthamiana*; (+), positive control, purified *E. coli* expressed GFP. (**B)** GFP expression measured by spectrofluorimetry using excitation and emission wavelengths of 485nm and 538nm. Columns represent value from the single sample.

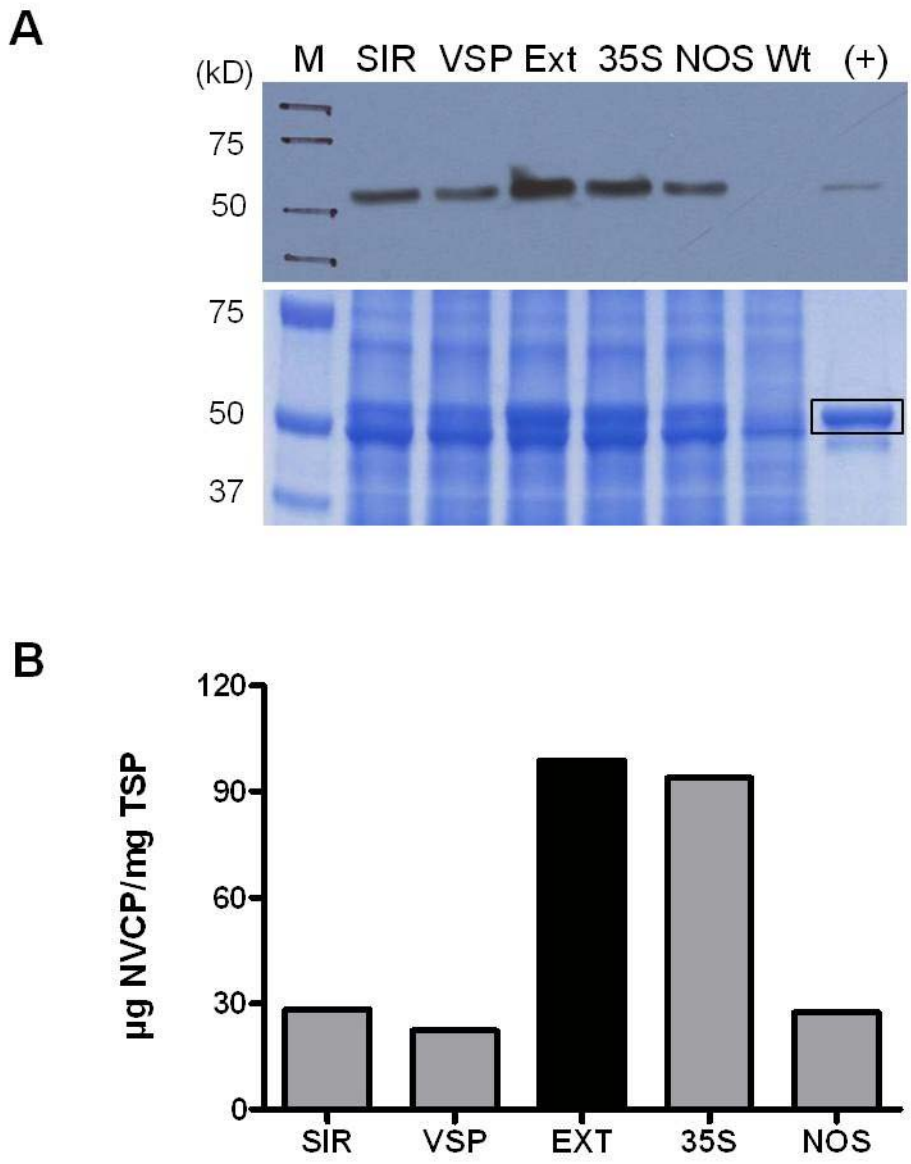


Figure 3.7 Comparison of terminators for NVCP expression. (A) Extracts from *N. benthamiana* leaves infiltrated with indicated constructs were analyzed by SDS-PAGE followed by either immuno-detection (upper) or Coomassie stain (lower). SIR, pBYNVCP.RSIR; VSP, pBYNVCP.R; EXT, pBYNVCP.REF; 35S, pBYNVCP.R35S; NOS, pBYNVCP.RNOS; Wt, wild-type *N. benthamiana*; (+), positive control, purified insect cell expressed NVCP. (B) NVCP expression measured by ELISA. Columns represent value from the single sample.

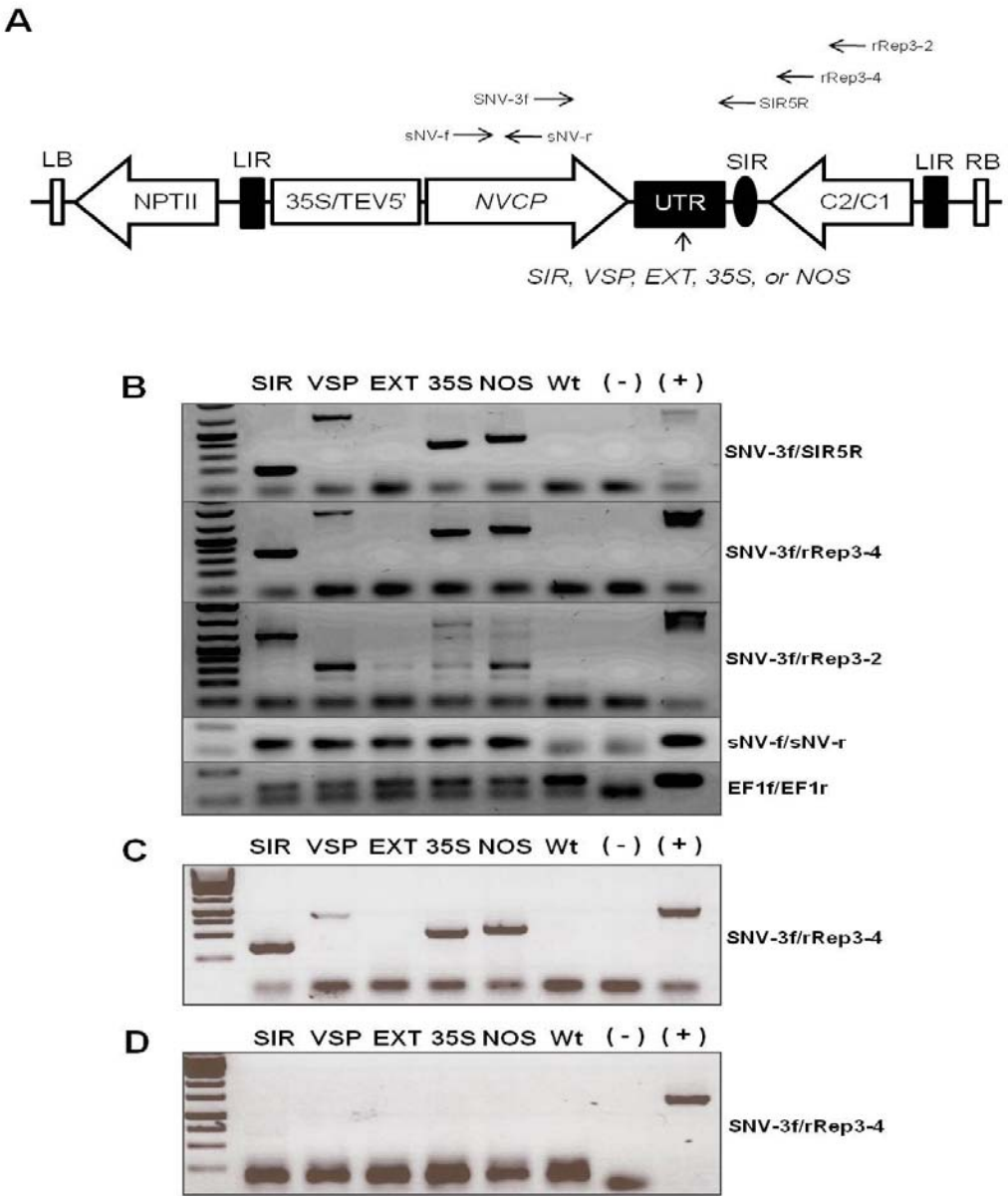


Figure 3.8 Ext terminator is efficient on transcription termination mediation. (A) Diagram of the T-DNA region of the vectors used in this study. SIR, pBYNVCP.RSIR; VSP, pBYNVCP.R; EXT, pBYNVCP.REF; 35S, pBYNVCP.R35S; NOS, pBYNVCP.RNOS; C1/C2, Rep/RepA cds. Primer binding sites are indicated by horizontal arrows. (B-D) Agarose gel electrophoresis of RT-PCR products. Total RNA was purified from leaf samples infiltrated with indicated constructs. cDNA was synthesized with random primer (B), gene specific primers rRep3-4 (C) or sNV-3f (D). Wt, wild-type *N. benthamiana*. (-), no template sample, (+), plasmid DNA of pBYNVCP,R was used as a template. Primers used for PCR are indicated on the right panel.

Table 3.1. Sequences of oligonucleotides used in this study

Name	**Sequence (5' to 3')**
35STm-1	GTGAGCTCGTCCGCAAAAATCACCAG
35STm-2	CAGAATTCGTCACTGGATTTTGGTTTTAGG
EFf	CTGGTGGTTTTGAAGCTGGTA
EFp	FAM-CACGCATTGCTTGCTTTCACCC-TAMRA
EFr	GGTGGTAGCATCCATCTTGTT
Ext-1	GTGAGCTCGAAGTGACATCACAAAGTTGAAG
Ext-2	CAGAATTCGTCATAACTGTAGAAATGATTCC
Ext-3	GTGAGCTCAAAGCAGAATGCTGAGCTA
Ext-6	CAGAATTCGAAACTGAACAAAACATACACAATGACAG
Ext-8	TGCGAGCTCCTGCATAAGAATATACATTGTGTG
GFP-f	GTCCAGGAGCGCACCATCTTCT
GFP-pro	FAM-CGGGTCTTGTAGTTGCCGTCGTCCTTG-TAMRA
GFP-r	GATGCCCTTCAGCTCGATGCGGTT
SD12-Sm	/5Phos/CTAAATGCCCCACAGGAGG
SD-12-Rm	/5Phos/AATTCCTCCTGTGGGGCATTTAGAGCT
SNV-3f	AATGGTGCTAGCAGCGGTCCAC
sNV-f	CAATGTCCAAAGTGTGCAATTCC
sNV-r	CCTCTTATCTTGGCAACATGTGAC
SIR5R	GATTTCCAACTGACTTGAAGTACACTC
rRep3-2	TCGTACCGGAAAGACCACCTGG
rRep3-4	GCTCGGGAATCGTCGAGTCACTGA

CHAPTER 4

SINGLE REPLICON BEAN YELLOW DWARF VIRUS VECTOR SYSTEM FOR HETERO-OLIGOMERIC PROTEIN COMPLEX EXPRESSION IN PLANTS

4.1 Abstract

The development of effective and simple plant viral vectors is essential to express important hetero-oligomeric pharmaceutical proteins in plants. Previously, a bean yellow dwarf virus (BeYDV)-based dual-replicon single vector system to express two distinct proteins has been reported. This chapter describes construction and testing of a simplified BeYDV single replicon system for co-expression of up to three recombinant proteins. This system relies on replacement of movement protein and coat protein gene of BeYDV by two or three transgene expression cassettes. The vector yielded simultaneous expression of three different fluorescent reporter proteins in the same cell. Moreover, the single replicon multi-cassette vectors allowed efficient expression and assembly of tetrameric monoclonal antibody against Ebola virus, and hexameric *E. coli* heat labile enterotoxin. Once optimized, this technology will be a versatile tool for the production of a wide spectrum of pharmaceutical multi-protein complexes in a fast, powerful, and cost-effective way.

4.2 Introduction

Plants offer numerous advantages over bacteria and mammalian cells for expression of pharmaceutical and industrial proteins. The outstanding benefits of the plant expression system include limited risk of contamination with animal pathogens, potentially lower cost, and easy scale-up (Daniell *et al.* 2009; Arntzen, 2008). While various immunologically active proteins have been successfully produced using a traditional plant expression method (stable transgenic plants), the long generation time and low yield of target proteins are major limitations of the system (Gomord *et al*., 2005). On the other hand, plant virus-based transient expression systems provide a rapid and high-yield alternative to the stable transgenic method: the gene of interest inserted among viral replicating elements is delivered into plant cells by *Agrobacterium*, amplified to a very high copy number without genomic integration, and translated into desired protein in a short period of time (Komarova *et al*., 2010). For these reasons, many DNA- and RNA-plant virus based expression systems have been developed and used to produce a variety of classes of therapeutic polypeptides and homooligomers (Lico *et al.* 2008).

Because many biologically and commercially important proteins, including antibodies, are composed of multiple heterologous subunits, recent efforts have been made to develop efficient co-expression viral vectors (Giritch *et al*., 2006; Roy *et al*., 2011). However, the competing nature of viruses made the task challenging. When viral vectors that share the same virus backbone are co-delivered into the same cell, the result is early segregation and subsequent

preferential amplification of one of the vectors in each cell. Therefore, the amount of cells co-expressing all of the desired proteins becomes extremely small (Dietrich and Maiss, Diveki *et al*., 2002). This problem has been overcome by utilizing two sets of vectors derived from non-competing viruses, tobacco mosaic virus (TMV) and potato virus X (PXV), as demonstrated by expression of fully assembled IgG at a level of 0.5 mg/g fresh leaf weight (FLW) (Giritch *et al*. 2006). Following this, co-infiltration of two full-length or deleted versions of cowpea mosaic virus (CPMV) RNA-2 constructs in the presence of RNA1 has found to be effective in expressing full-size IgG (Sainsbury *et al*., 2008). Recently, simultaneous expression of two or three proteins using a disarmed tobacco etch virus (TEV) was reported (Bedoya *et al*., 2010). The TEV RNA polymerase (NIb) cistron was replaced by a cassette containing two or three target genes flanked by protease cleavage motifs, and the deleted NIb was supplied *in trans* either from transgenic plant or by co-infiltration strategies. However, attempts to express two or more different proteins from one viral vector have failed, probably due to the decreased viral replication efficiency in relation to increase in inserted gene length, resulting in drastic reduction of expression level. Therefore, there is a need to develop a single viral vector to ensure co-expression of all target proteins in the same cell, which in turn will facilitate accurate assembly of multi-protein complexes.

Bean yellow dwarf virus (BeYDV, *Geminiviridae*, *Mastrevirus* genus) has a single-stranded circular DNA genome (~ 2.5 kb), and replicates via a rolling circle mechanism (Liu *et al*., 1998; Palmer and Rybicki, 1998). The genome is

separated by long and short intergenic regions (LIR and SIR, respectively) and encodes four viral proteins: movement protein (MP), coat protein (CP) and two replication-associated proteins, Rep and RepA. For its replication, unlike other viral expression systems that require near-full-length viral genome, the BeYDV requires only the two cis elements, LIR and SIR, and the proteins Rep/RepA. Thus, the coding sequence of MP and CP can be replaced with an expression cassette containing a gene of interest (BeYDV replicon vector), without compromising viral replication (Mor *et al*., 2003). Moreover, such BeYDV vectors are non-competing as co-delivery of two replicon vectors allows efficient simultaneous expression of two different proteins in the same cells (Huang *et al*., 2010). These features made BeYDV suitable to be developed as a single viral vector for multi-protein expression.

Previously, Huang et al reported BeYDV-based dual replicon single vector system that contains two tandemly linked replicons (Huang *et al*., 2010). The dual replicon vector permitted high-level amplification of two distinct DNA replicons and efficient expression of both proteins in the same cells. Using the system, the protective IgG mAb 6D8 against Ebola virus GP1 was produced at a level of 0.5 mg/g FLW, within 4 days post infiltration. In this study, I show that the BeYDV-based dual replcon system can be simplified to have a single replicon that contains two or three expression cassettes. Using the single replicon vector I demonstrate 1) co-expression of green fluorescent protein (GFP) and *Discosoma* sp. red fluorescent protein (DsRed); 2) co-expression and assembly of the heavy and light chain molecules of anti-Ebola virus mAb 6D8; and 3) co-expression and

assembly of A and B subunits of *E.coil* heat labile enterotoxin. I further demonstrate that the BeYDV single replicon vector allows efficient simultaneous expression of three fluorescent report proteins in the same cell. This system represents a significant advance in transient expression technology that may allow a rapid and robust expression of multi protein complexes in plants.

4.3 Materials and Methods

4.3.1 Vector construction

BeYDV-based tandem dual replicon constructs pBYGFPDsRed.R and pBY-HL(6D8).R were obtained from H. Mason (Huang *et al*., 2010) and renamed as pBY-G(SL)R and pBY-H(SL)L in this study (Fig. 4.1). Construct pBY-GR and pBY-HL were designed to contain two expression cassettes in tandem and to be flanked by LIR and SIR. For the construction of pBY-GR (Fig. 4.1), standard overlapping PCR and molecular cloning methods were used. Primer sets GR5-1 /GR5-2 and GR3-1/GR3-2 were used for initial amplification in separate PCR reactions using pBY-G(SL)R as a template. The resulting PCR fragments were mixed and amplified using primers GR5-1 and GR3-2, complementary to the ends of the two initial fragments. The resulting PCR product was digested with SacI/KpnI and ligated with pBY-G(SL)R digested likewise, to yield pBY-GR. For the construction of pBY-HL (Fig. 4.1), similar steps were carried out with the primer sets HL5-1/HL5-2, and HL3-1/ HL3-2 for initial amplifications using pBY-H(SL)L as template, and the primer set HL5-1/HL3-2 for overlapping PCR.

The final PCR product was digested with SacI/KpnI and ligated with pBY-H(SL)L digested likewise.

Vectors expressing *E. coli* heat labile enterotoxin subunits A and B (LTA, LTB) were obtained from H. Mason: pBYRsLTAR72 contains an expression cassette encoding LTA (substitution mutant A72R), pBYRsLTB contains an LTB cassette, and pBYRsLTR72 contains tandemly linked LTA and LTB cassettes. In this study, constructs pBYRsLTAR72, pBYRsLTB, and pBYRsLTR72 were renamed as pBY-LTA, pBY-LTB, and pBY-LTAB, respectively.

The construction of plasmids pBYGFP and pBYDsRed has been previously described (Huang *et al*., 2010). For the construction of pBYCFP, CFP gene (Accession number EU530627) was PCR amplified from plasmid pIBT-PR7:eCFP (gift from Dr. Z. Huang, Arizona State University) with primers GFP-BsaF and GFP-PacI, digested with BsaI and PacI, and ligated with pBYR7, a derivative of pBYR2 (Chen *et al*., 2011). The CFP cassette was obtained by PCR with primers U35S-SpeF and Ext6-SalR using pBYCFP as a template, digested with SpeI-SalI, and ligated with pBY-GR digested likewise, to yield the single-replicon three-expression cassette vector pBY-GCR.

4.3.2 Agroinfiltration

The binary T-DNA plasmid constructs were separately introduced into *Agrobacterium tumefaciens* LBA4404 by electroporation. The agroinfiltration procedure was performed as previously described (Huang *et al*., 2009). Briefly, *Agrobacterium* strains were pelleted by centrifugation at 4,000 *g* for 5 min and

resuspended in the infiltration buffer (10 mM MES, pH 5.5 and 10 mM $MgSO_4$,) to reach OD_{600}=0.2. The *Agrobacterium* suspensions were infiltrated, either alone or as a mixture of several strains, into 6-8 weeks old *N. benthamiana* leaves. The infiltrated plants were incubated in a plant growth chamber until harvested on 2 or 3 DPI.

4.3.3 Plant DNA extraction and DNA replicon analysis

Total plant DNA was extracted using DNeasy Plant Mini kit (Qiagen) according to the manufacturer's instructions. Purified DNA (1 µg) was digested with indicated restriction enzymes (Fig. 4.3 and 4.7) and run on a 1% agarose gel with ethidium bromide for DNA staining.

4.3.4 Protein extraction

Total soluble protein (TSP) was extracted from infiltrated leaf tissues by homogenizing in the extraction buffer (25mM sodium phosphate, pH 6.6, 100mM NaCl, 1mM EDTA, 0.05 % TritonX-100, 50mM sodium ascorbate and 10µg/ml leupeptin) using a FastPrep machine (Bio101, Vista, CA). The crude plant extract was clarified by centrifugation at 13,000*g* for 10 min at 4°C and TSP concentration was measured using Bradford reagent (Bio-Rad) with known concentration of bovine serum albumin (BSA) as the standard.

4.3.5 Analysis of fluorescent proteins

GFP and DsRed fluorescence intensity was examined on a microplate reader (Spectra Max M2, Molecular Device) at room temperature. TSP samples (25µg) were loaded to black-wall 96-well plates (Corning) in duplicate and read with excitation and emission wavelength of 485nm and 538nm, respectively, for

GFP, and 544 nm and 590 nm for DsRed. The fluorescence value of the negative control (extract of un-infiltrated plant leaf) was subtracted before graphing. Expression levels are reported as fluorescence units (FU) per 25 μg TSP (Fig. 4.2).

Fluorescence microscope images were taken using a Zeiss LSM 5 DUO (Carl Zeiss) laser scanning confocal microscope. Infiltrated leaf tissue sections were mounted with water and imaged with a Zeiss EC Plan-Neofluar 40X/1.3 oil immersion lens. Fluorescence signals for GFP, CFP, and DsRed were sequentially scanned with excitation lasers of 488, 458 and 543nm, respectively, and detection windows of 550-560, 470-500 and 614-646 nm, respectively. For plant chlorophyll autofluorescence detection, the excitation laser of 633 nm with detection window of 630-700 nm was used. All images were taken at 512 x 512 pixel resolution covering an area of 318 x 318 μm^2. An 8-line average was applied to all scans with the scan speed set to 6.39 μs/pixel.

4.3.6 SDS-PAGE and western blot

SDS-PAGE and western blotting for mAb 6D8 were performed as previously described (Phoolcharoen *et al*., 2011). Briefly, plant extracts were mixed with SDS sample buffer either under reducing [3.5% β- mercaptoethanol (V/V)] or non-reducing condition and then subjected to 4-20% SDS-PAGE. PAGE gels were either stained with PageBlue protein staining solution (Fermentas) or used to transfer proteins onto PVDF membranes (Amersham Pharmacia). To detect heavy chain or assembled heavy and light chains, the membrane was probed with goat anti-human IgG-HRP conjugate (Southern Biotech) diluted at 1:5000 in 1% skim milk in PBST. To detect light chain, goat

anti-human kappa-HRP conjugate (Southern Biotech) diluted at 1:10000 in 1% skim milk in PBST was used. Bound antibodies were detected by using ECL plus detection reagent (Amersham, NJ).

To analyze the expression of LTB and holotoxin the clarified total protein extracts were subjected to 4-20% SDS-PAGE either reducing [3.5% β-mercaptoethanol (V/V)] or non-reducing condition. Proteins were transferred onto PVDF membrane and probed either with goat anti LTB serum (1:10000) or goat anti LTA (1:5000) serum followed by rabbit anti-goat IgG-HRP conjugate (Sigma, 1:5000). The LTA and LTB antisera were produced by immunization of a goat with recombinant His-tagged LTA or LTB cloned into pQE60 (Qiagen, Valencia, CA) and expressed in *E. coli*. The membranes were developed by chemiluminescence using ECL plus detection reagent (Amersham, NJ). Positive control LT was expressed in *E. coli BL-21* transformed with pCS96 (Bowman & Clements, 2001) and purified by galactose affinity chromatography as described (Bowman & Clements, 2001).

4.3.7 Ganglioside binding-dependent ELISA

The ELISA for ganglioside binding-dependent detection of assembled LT holotoxin was done as described (Cardenas & Clements, 1993) with modifications. Briefly, serial diluted plant crude extracts ranging from 500 to 18 ng TSP were loaded to the ganglioside (Calbiochem)-bound micro plates and LT holotoxin was detected in succession with goat anti-LTA serum (1:500 dilution) and rabbit anti-goat IgG-HRP conjugate (Sigma, 1:5000 dilution).

4.4 Results

4.4.1 BeYDV single replicon vector allows co-expression of two proteins.

Previous study showed (Huang *et al*., 2010) that a BeYDV-derived single-vector multi-replicon system containing two tandemly linked replicons [e.g., pBYR-G(SL)R in Fig. 4.1] leads to high-level amplification of two distinct DNA replicons and efficient expression of both proteins. In this study, I tested whether the single-vector multi-replicon system can be simplified to a single replicon system for co-expression of heterologous proteins. For this, the LIR and SIR elements located between the two expression cassettes in pBYR-G(SL)R were removed to give pBYR-GR (Fig. 4.1). The resulting construct contains two tandemly arranged expression cassettes encoding GFP and DsRed flanked by LIR and SIR. To test the co-expression of two reporter proteins, I infiltrated *N. benthamiana* plants with *Agrobacterium* strains harboring either pBYR-G(SL)R or pBYR-GR. Three days post infiltration (DPI), the single replicon vector showed co-expression of GFP and DSRed proteins as efficiently as the double replicon (Fig. 4.2). Flurorescence of GFP and DsRed from the single replicon vector (pBYR-GR) was similar to that from the double replicon vector [pBYR-G(SL)R].

The replicon size of the pBYR-GR is 215% larger than that of the native BeYDV replcon. To determine whether the pBYR-GR permits efficient replicon amplification, total plant DNA samples were purified and subjected to restriction enzyme digestion followed by agarose gel electrophoresis. As shown in Figure 4.3, gel showed the expected ~5.4kb band from pBY-GR sample, indicating

efficient formation of the DNA replicon. Double digested pBY-GR DNA sample showed two bands at ~ 2.0 kb and ~ 3.4 kb, which corresponding to the fragments derived from the DsRed and GFP genes, respectively. In agreement with the previous study (Huang el al., 2010), the double replicon construct pBYR-G(SL)R DNA sample produced a GFP replicon at ~2.6 kb and a DsRed replicon (containing the Rep/RepA coding sequence) at ~3.5 kb. Comparison of the replicon band intensity showed that the single replicon amplified slightly less than the dual replicons, probably due to the larger replicon size. However, the levels of GFP and DsRed expression were nearly equal between the single and dual replicon systems (Fig. 4.2). Together, these results demonstrate that the BeYDV replicon can contain multiple expression cassettes for expression of distinct proteins.

4.4.2 Expression and assembly of IgG mAb 6D8 against Ebola virus GP1

To demonstrate the utility of the single replicon double cassettes expression system, I tested expression of mAb 6D8 against Ebola virus glycoprotein GP1. For this, two expression cassettes encoding 6D8 heavy chain or light chain were linked in tandem and placed between LIR and SIR. The Rep/RepA coding sequence remained in its normal position in the viral genome (Fig. 4.1). The resulting construct pBYR-HL was infiltrated into *N. benthamiana* leaves, and at 3 DPI leaf samples were harvested for protein assay. Western blot with SDS-PAGE under reducing condition confirmed the presence of both heavy and light chains with the expected molecular weights of 50 kDa and 25 kDa,

respectively (Fig. 4.4A and 4.4B). When the same samples were analyzed under non-reducing condition, a ~150 kDa band was observed, which represents the fully assembled hetero-tetrameric form of 6D8 mAb (Fig. 4.4C). The western blot analysis also showed that the expression levels of the 6D8 heavy and light chains and the assembled mAb were comparable with those from the dual replicon construct pBYR-H(SL)L. These results indicate that the single replicon vector is as efficient as the dual replicon vector, not only for co-expression but also for correct assembly of two subunits, while providing simplicity.

4.4.3 Expression and assembly of A and B subunits of *E.coil* heat labile enterotoxin

The BeYDV single replicon system was applied for expression of *E. coli* heat labile enterotoxin (LT), which assembles into a hexamer composed of five identical B subunits and one A subunit (AB5, holotoxin). LT is a potent mucosal immunogen, and have a strong mucosal adjuvant quality (Clements *et al*., 1988; Liang and Hajishengallis, 2010). Pentameric LTB5, with or without LTA, binds to the ganglioside G_{M1} on the surface of gut mucosal cells (Willams *et al*., 1999). Two expression cassettes encoding the LTA (A72R mutant) and LTB subunits were cloned into the BeYDV single repicon system (Fig. 4.1). The resulting construct, pBYR-LTAB was agroinfiltrated into *N. benthamiana* leaves, and infiltrated leaf areas were collected for protein assay. Expression of the LTA and LTB was analyzed by SDS-PAGE under reducing conditions followed by Western blotting probed with LTA- and LTB-specific antibodies. The expected

molecular weights of 27.0 kDa and 11.6 kDa corresponding to the LTA and LTB, respectively, were observed. To test whether the LTA and LTB assembled into AB5 holotoxin, the crude protein extract was separated on SDS-PAGE under nonreducing condition. Western blot with LTB specific antibody showed two distinct bands as seen in the *E. coli* expressed and purified positive control. Two bands of ~40 kDa and ~60 kDa most likely correspond to pentameric B subunit and AB5 holotoxin, respectively.

Next, plant material was analyzed by ganglioside binding-dependent ELISA for estimation of holotoxin assembly (Fig. 4.6). Crude extracts used in the Western blot analysis were loaded into microtiter plates coated with ganglioside G_{M1} and probed with LTA-specific antibody. Because LTA does not bind to ganglioside unless assembled into holotoxin with pentameric LTB, the ganglioside-dependent ELISA for LTA verifies holotoxin assembly. The high absorbance values obtained from pBYR-LTAB samples confirmed the efficient assembly of LT holotoxins.

4.4.4 Simultaneous expression of three fluorescent proteins by a single replicon vector

So far, I showed that the BeYDV single replicon system provides the efficient co-expression of two target proteins. In this study, I first tested whether three distinct proteins can be co-expressed using the BeYDV single vector system. Three expression cassettes encoding three fluorescent proteins, GFP, CFP, and DsRed were inserted into the 3-cassette vector pBYR-GCR (Fig. 4.1). Each cassette was designed to have a different 5’ untranslated region and terminator to

avoid undesired recombination or RNAi-mediated mRNA degradation, although the same CaMV 35 promoters were used. Construct pBYR-GCR was delivered into *N. benthamiana* and the co-expression of the three reporter fluorescent proteins was analyzed at 2 DPI. Confocal microscopic examination revealed colocalized green, cyon, and red fluorescence, indicating efficient simultaneous expression of the three different proteins. In contrast, leaf samples individually infiltrated with pBYR-GPF, pBYR-CFP, or pBY-DsRed, which contain one expression cassette, showed only single fluorescent signal (green, cyon, or red, respectively) (Fig. 4.7).

Analysis of DNA from infiltrated leaf samples showed expected ~7.1 kb band of the pBYR-GCR replicon (Fig. 4.8). I observed that the amount of the replicon is reduced in correlation to the replicon size. Overall, these data indicate that the single replicon vector with three target protein expression cassettes allows replicon amplification and co-expression of three different target proteins.

4.5 Discussion

In many cases, co-expression of two or more proteins is required to produce biologically active pharmaceutical proteins, such as monoclonal antibodies and multi-component virus like particles. Therefore, the co-expression techniques have become the major focus of recombinant protein expression studies. Many systems are now available, including *E. coli*, eukaryotic cells and plants. Plants are one of the best systems to express the multimeric complex molecules as highlighted by the production of IgG and IgA molecules (Hiatt *et al*.,

1989; Ma *et al*., 1995). Traditionally, the transgenic method was used for multi-protein complex expression in plants. This involves two insertion strategies; the insertion of a single DNA construct encoding two genes of interest (i.e. heavy and light chain cassettes) or two independent insertions each encoding one transgene followed by cross-pollination (Bouquin *et al*. 2002). The transgenic approach, in recent years, has been replaced by transient expression strategies as the latter allows better yield of recombinant protein complexes in short period of time (Obembe *et al*., 2011). In the transient expression systems, two strategies have been applied: co-infiltration of multiple vectors based on non-competing viruses and infiltration of a single vector that contains multiple expression cassettes.

Previously, Huang et al reported that the single BeYDV vector containing two tandemly linked replicons yields ~95% co-expression efficiency of two target proteins within same cells, while co-infiltration of two vectors each expressing one target protein resulted in ~80% co-expression efficiency (Huang et al., 2010). A similar result was seen in the insect cell/baculovirus expression system (Vieira *et al*., 2005). When a single-infection method using a tricistronic multi-gene vector was compared with co-infection method using three monocistronic baculovirus vectors, the single infection strategy was more effective for multi-subunit rotavirus-like particle production. These results demonstrate the benefits of the single vector approach in facilitating the co-expression of all target proteins in the same cell, which is essential for the multimeic protein complex formation. Moreover, the method also enables a further co-expression of additional

components such as gene silencing suppressor or chaperones carried on supplemental vectors.

In this study, I focused on simplifying the BeYDV dual replicon system to improve its feasibility for multimeric protein complex expression. I created BeYDV single replicon vectors harboring two or three expression cassettes. GFP and DsRed expression studies showed that the single replicon system is as efficient as the previously reported double replicon vector (Fig. 4.2). Furthermore, the confocal study showed that the single replicon system is efficient in directing simultaneous expression of three different fluorescent proteins.

Next, I evaluated the utility and capacity of the single replicon system by producing a protective mAb against Ebola virus GP1. The single replicon vector pBY-HL, containing the heavy chain and light chain expression cassettes, resulted in the co-expression of both molecules, which correctly assembled to its tetrameric structure (Fig. 4.5C). The expression level was as high as that from the double replicon vector, pBYR-H(SL)L, which produced 6D8 mAb at a level of ~0.5 mg/g leaf fresh weight. This level is comparable to the highest level ever achieved with a plant-based expression system (Giritch *et al*., 2006). In addition, I have also successfully expressed *E. coli* heat labile enterotoxin. Western blot and ganglioside dependent ELISA demonstrated that the LTA and LTB subunits assembled into holotoxin. The success in producing the fully assembled tetrameric IgG and hexameric LT holotoxin strongly suggests that the single replicon vector is highly efficient for the multimetic protein complex expression. This finding may indicate the possible construction of double replicon double

cassettes vector for production of four heterologous proteins in a single BeYDV system.

BeYDV requires only the two viral cis elements LIR and SIR, and Rep/RepA proteins for its replication. This feature facilitates insertion of large amounts of heterologous genetic information into a BeYDV vector. A notable finding from the current study is that BeYDV replicon can be enlarged as much as 280% of the native viral replicon size. Transient infiltration of constructs pBYR-GR and pBYR-GCR resulted in efficient formation of ~5.4 kb and ~7.1 kb replicons, respectively (Fig. 4.8). I observed that the accumulation of the enlarged replicons was decreased by 38-64% compared to the smaller replicon (i.e. pBYR-GFP in Fig. 4.8). However, the pBYR-GR resulted in GFP and DsRed expression that was comparable with pBY-G(SL)R. This may because BeYDV vectors yield a greater degree of replicon amplification than that of target protein expression (Zhang and Mason, 2006; Regnard *et al*, 2010). Thus, I expect that the optimal expression level can be maintained although the level of replicon accumulation is decreased in some degree. The maximum amount of genetic information that can be placed into a BeYDV replicon vector without compromising the expression level remains to be studied.

Despite the efficient co-expression obtained using the single replicon vector, this approach has some limitations. Aside from the sequential steps needed for a construct generation, it requires multiple cloning sites for each cloning step. This disadvantage could be addressed by using the Gateway® cloning technology based on bacteriophage lambda site-specific recombination system to facilitate

cloning of target sequences. In addition, repetitive use of the same promoters, CaMV 35S could possibly result in undesired recombination during replicon amplification. If necessary, different promoters could be incorporated into the system, such as the nopaline synthase (NOS), octopine synthase (OCS), and mannopine synthetase (MAS) promoters from the Ti plasmid of *A. tumefaciens* (Shaw *et al*. 1984; Ellis *et al*. 1987; Comai *et al*., 1990). By controlling the strength of a promoter in combination with a translation enhancer in the 5'UTR, and/or particular terminator element, the single replicon system can be further optimized.

In conclusion, this chapter describes a simple and high-yield BeYDV-based single replicon vector for co-expression of two or more recombinant proteins. Once optimized, this technology will be a versatile platform for the production of a wide spectrum of pharmaceutical multi-protein complexes in a fast, efficient, and cost-effective way.

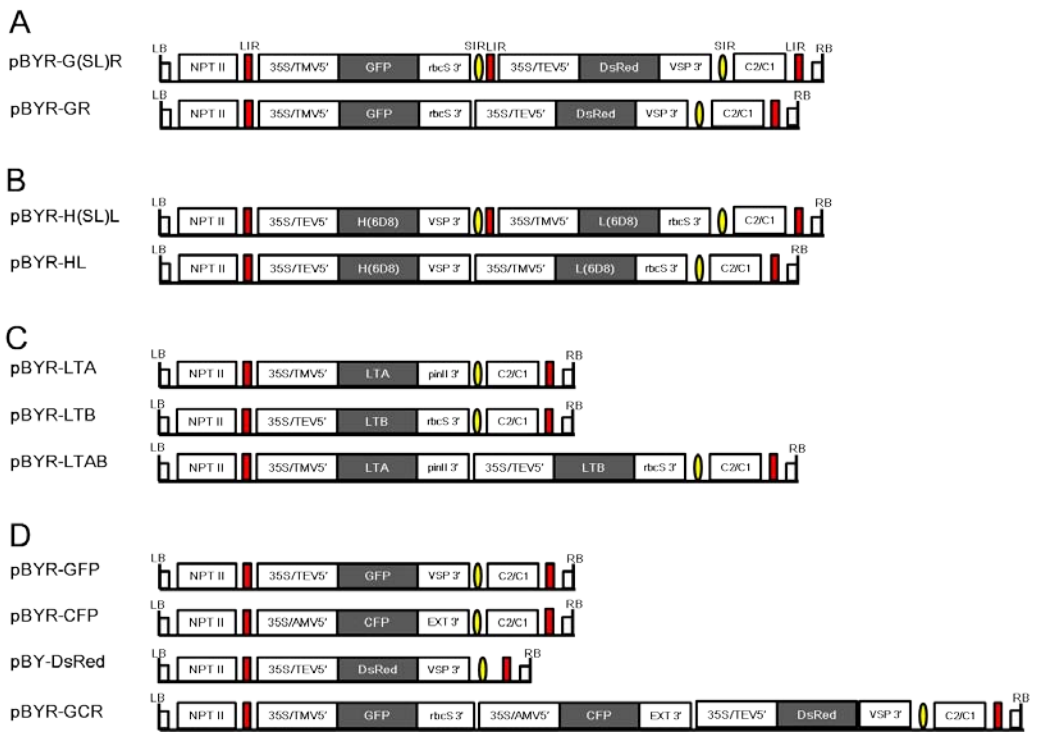


Figure 4.1 Plasmid constructs. (A) Double [pBYR-G(SL)R] and single replicon (pBYR-GR) vectors used for co-expression of GFP and DsRed. (B) Double [pBYR-H(SL)L] and single replicon (pBYR-HL) vectors for expression of mAb against Ebola virus. (C) Constructs for co-expression of LTA and LTB: either two separate vectors (pBYR-LTA and pBYR-LTB) or single vector (pBYR-LTAB) were used. (D) Vectors used for co-expression of three fluorescence proteins. 35S/TEV 5', CaMV 35S promoter with tobacco etch virus 5'UTR; 35S/TMV 5', CaMV 35S promoter with tobacco mosaic virus 5'UTR; 35S/AMV 5', CaMV 35S promoter with alfalfa mosaic virus 5'UTR; VSP 3', soybean vspB gene 3' element; rbcS 3', pea rbcS gene 3' element; pinII 3', potato potato proteinase inhibitor II terminator; EXT 3', tobacco Extensin 3' terminator; NPT II, expression cassette encoding *nptII* gene for kanamycin resistance; LIR (red box), long intergenic region of BeYDV genome; SIR (yellow oval), short intergenic region of BeYDV genome; C2/C1, BeYDV ORFs C1 and C2, encoding replication initiation protein Rep and RepA; LB and RB, the left and right borders of the T-DNA region; GFP, green fluorescent protein gene, CFP, cyan fluorescent protein gene, DsRed, DsRed fluorescent protein gene; H(6D8) and L(6D8), heavy and light chain gene of a protective IgG mAb 6D8 against Ebola virus GP1 (Wilson *et al*. 2000); LTA and LTB, genes of A and B subunits of *E. coli* heat-labile enterotoxin.

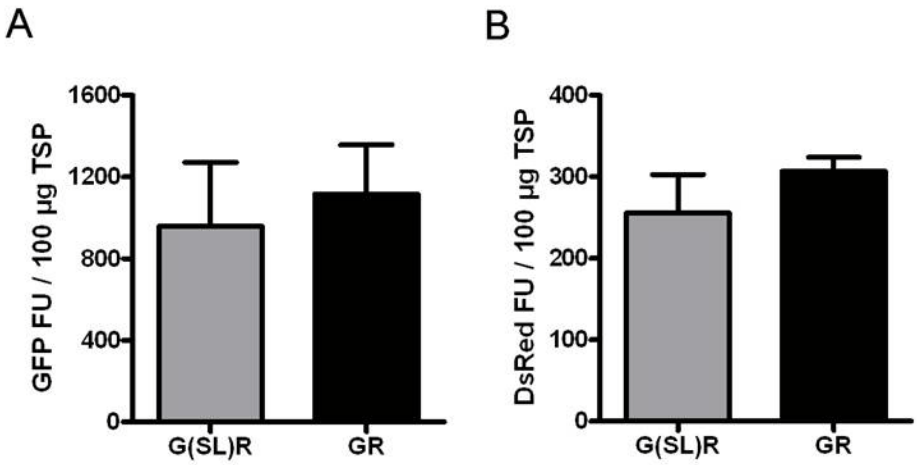


Figure 4.2 Expression of GFP and DsRed in plant leaf cells. Fluorimetric analysis of GFP (A) and DsRed (B) accumulation showing efficient co-expression of two fluorescent proteins from either single [pBYR-GR] or dual [pBYR-G(SL)R] replicon vectors. Dilutions of total soluble protein extracts were subjected to spectrofluorimetry with a Spectra Max M2 (Molecular Devices) using excitation and emission wavelengths of 485nm and 538nm for GFP, and 544nm and 590nm for DsRed. G(SL)R, pBYR-G(SL)R; GR, pBYR-G(SL)R. Data are means ± S.D. from three independently infiltrated samples.

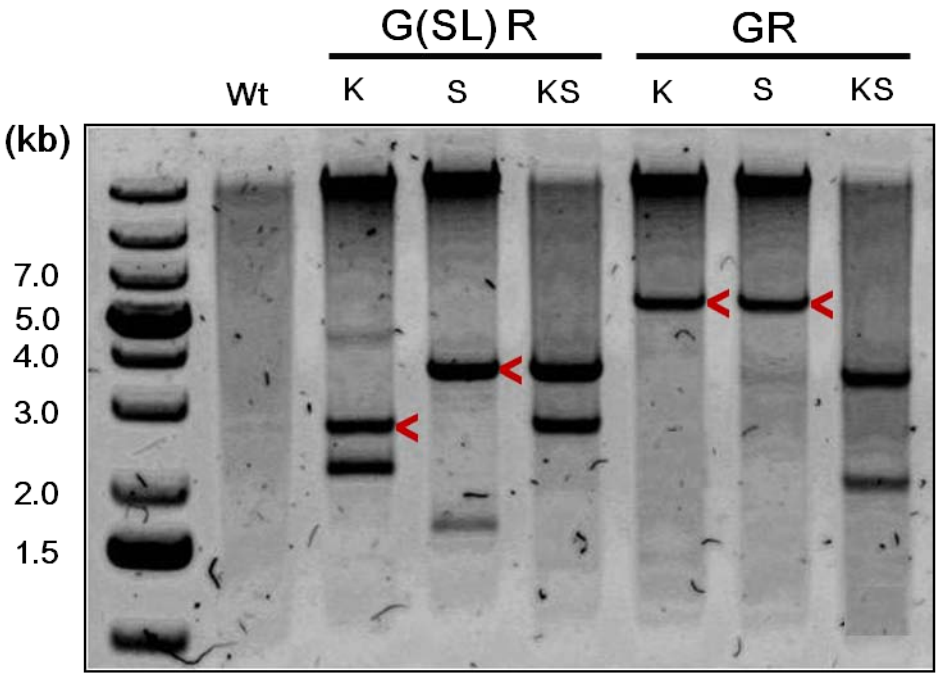


Figure 4.3 Agarose gel stained with ethidium bromide showing efficient replicon amplification. Leaf DNA was extracted from 3 DPI samples of uninfiltrated (Wt), or infiltrated with the indicated vectors, digested with indicated restriction enzymes (K, KpnI; S, SacI) and run on a 1% agarose gel. G(SL)R, pBYR-G(SL)R; GR, pBYR-G(SL)R. Replicon positions are indicated with arrow head.

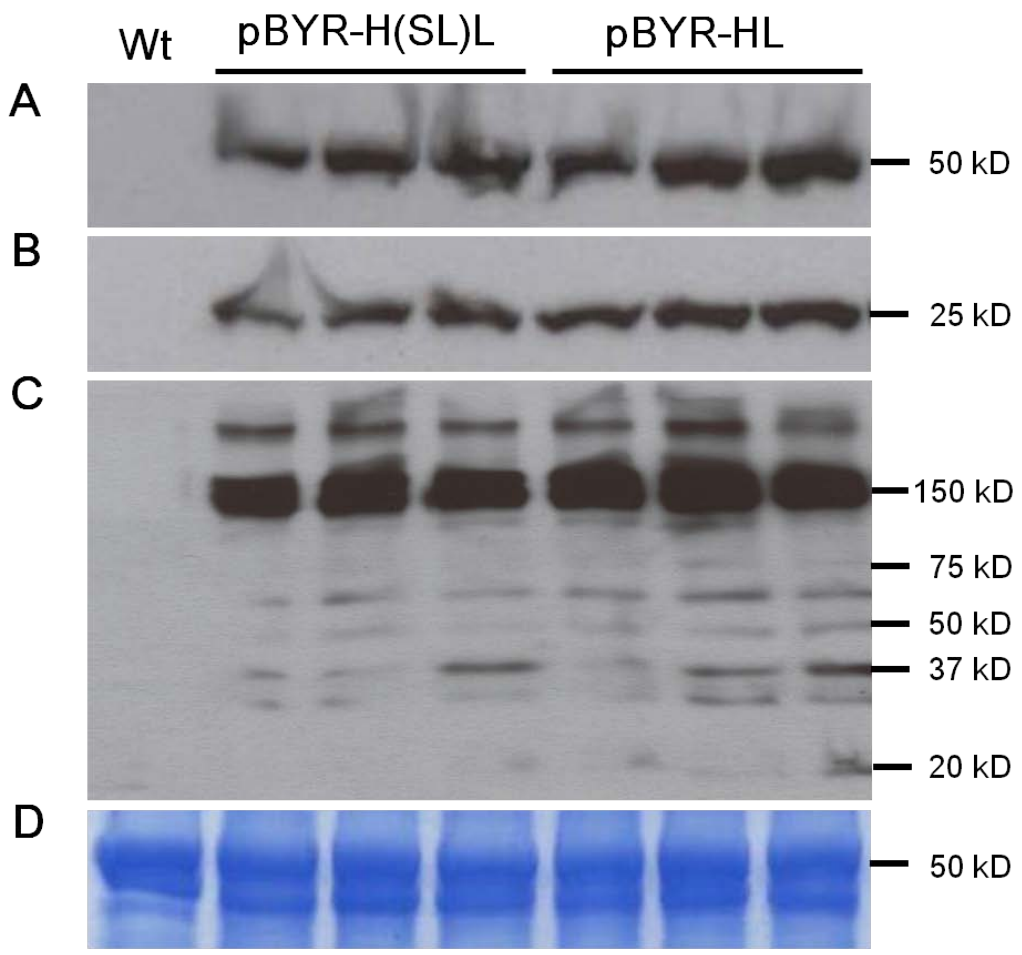


Figure 4.4 Western blot analysis of plant-derived 6D8. Protein samples were separated on a 4-20% SDS-PAGE gradient gel under denaturing and reducing condition (A and B) or under non-reducing condition (C) and blotted onto a PVDF membrane. The membrane was incubated with a goat anti-human gamma chain antibody or goat anti-human kappa chain antibody to detect heavy chain (A) or light chain (B and C). Wt: Protein samples extracted from uninfiltrated leaves; lanes marked pBYR-H(SL)L: protein samples extracted from the leaves infiltrated with dual replicon construct pBYR-H(SL)L; lanes marked pBYR-HL: protein extracted from the single replicon construct pBYR-HL. Commassie blue stained gel is shown for normalized total protein loading.

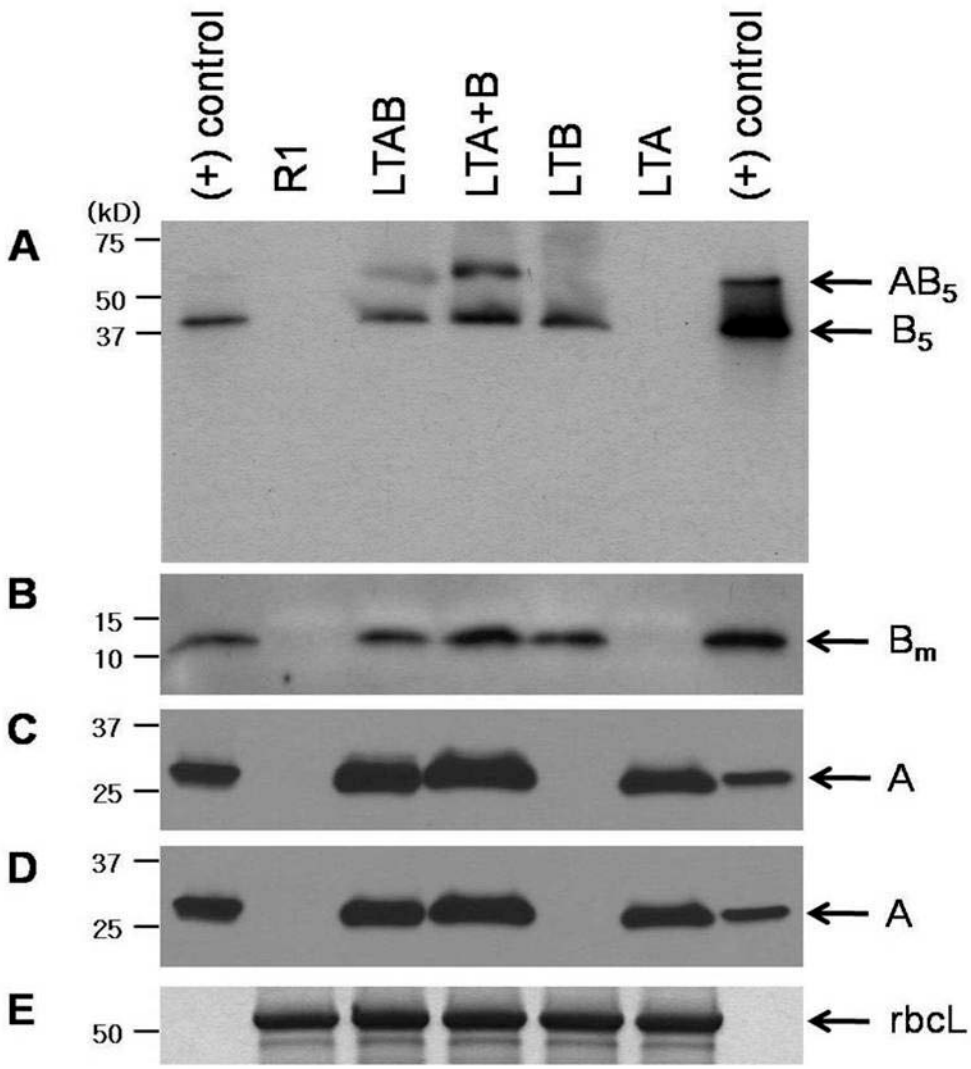


Figure 4.5 Western blot analyses of transiently expressed LTB and LTA. Protein samples were subjected to SDS-PAGE under denaturing and non-reducing condition (A and C) or denaturing and reducing condition (B, D and E). After electrophoresis, the gel was stained with Coomassie brilliant blue (E) or transferred on to PVDF membrane and then immunodecorated with either anti-LTB serum (A and B) or anti-LTA serum (C and D) followed by horseradish peroxidase (HRP)-conjugated anti-rabbit antibody and ECL detection. The sizes of the proteins are shown on the left. The positions of the holotoxin (AB_5), B-subunit pentamer (B_5) and monomer (B_m), the A-subunit (A) and the large subunit of Rubisco (rbcL) are indicated on the right. LTAB, pBYR-LTAB; LTA+LTB, co-infiltration of pBYR-LTA and pPBYR-LTB; LTA, pBYR-LTA; LTB, pBYR-LTB; R1, empty vector pBYR1 as a negative control.

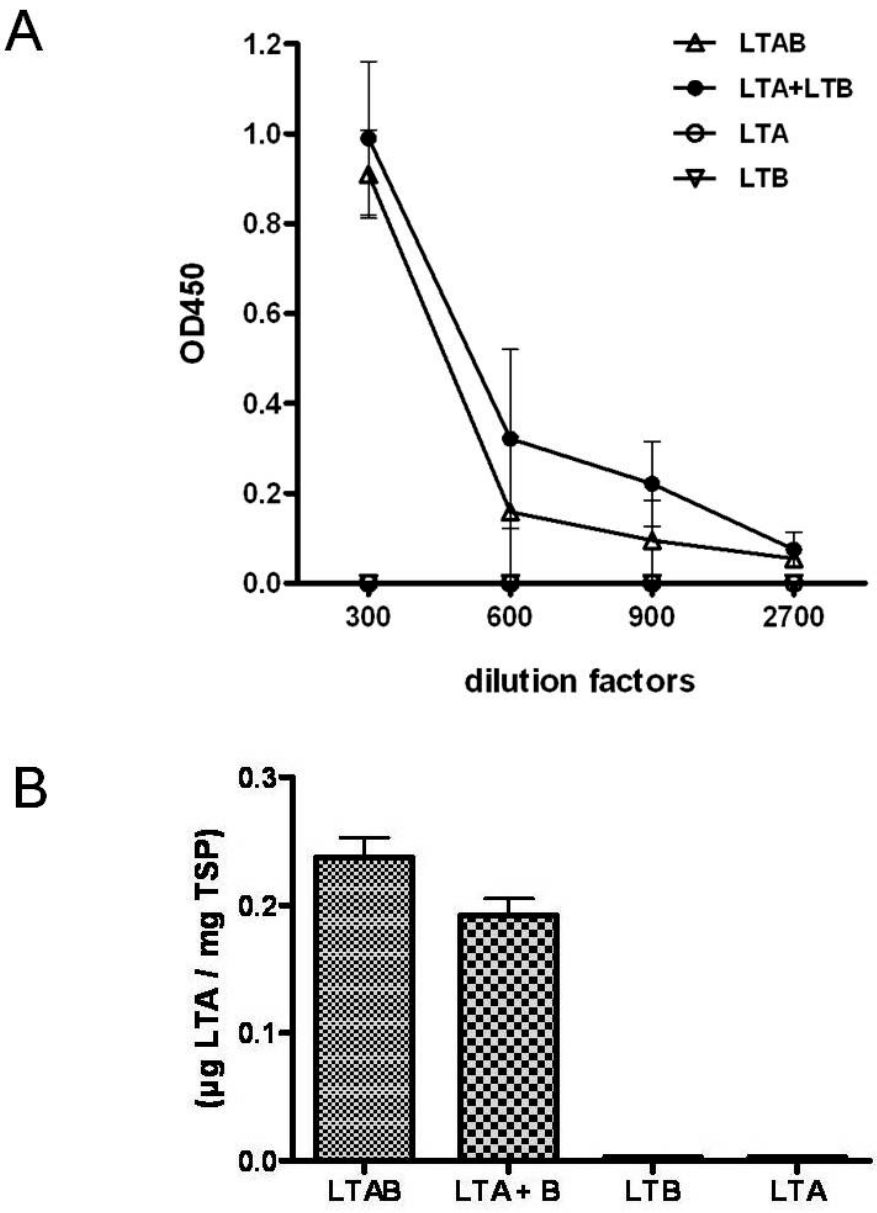


Figure 4.6 Formation of hexameric LT holotoxin. Crude extract from leaf samples infiltrated with indicated constructs were tested using ganglioside-dependent ELISA to analyze LT holotoxin assembly. (A) Serial 3-fold dilutions of crude extracts were loaded to ganglioside bound plates and probed with goat anti-LTA serum. (B) *E. coli* expressed and purified LT was used as the standard. LTAB, infiltrated with pBYR-LTAB; LTA+LTB, infiltrated with *Agrobacterium* mixture of pBYR-LTA and pBYR-LTB; LTA, infiltrated with pBYR-LTA; LTB, infiltrated with pBYR-LTB. All samples were harvested at 3 DPI.

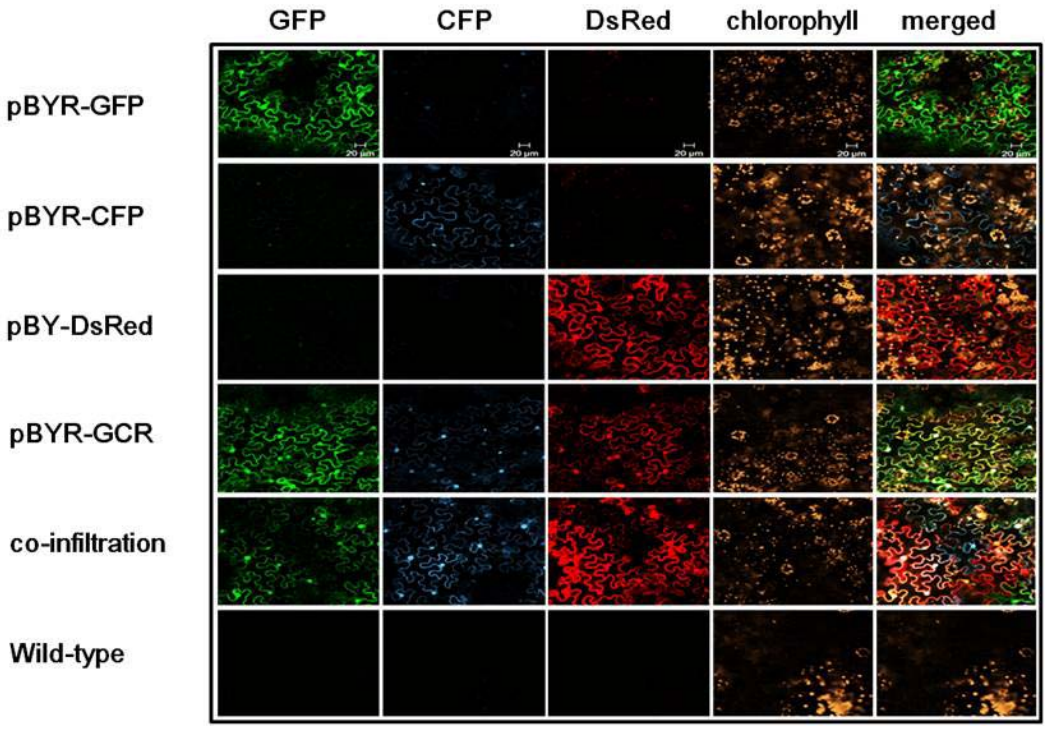


Figure 4.7 Simultaneous co-expression of three fluoresecent proteins. *N. benthamiana* leaves were infiltrated with *Agrobacterium* strains harboring expression vectors as indicated (on the left) and at 2 DPI the infiltrated leaf samples were examined with confocal laser scanning microscope. For co-infiltration, mixture of *Agrobacterium* strains harboring pBYR-GFP, pBYR-CFP, and pBY-DsRed were used. Excitation lasers of 488, 458 and 543 nm and detection windows of 550-560, 470-500 and 614-646 nm were employed to detect GFP, CFP and DsRed signals, respectively. For plant autofluorescence chlorophyll detection, the excitation laser of 633 nm with detection window of 630-700 nm was used.

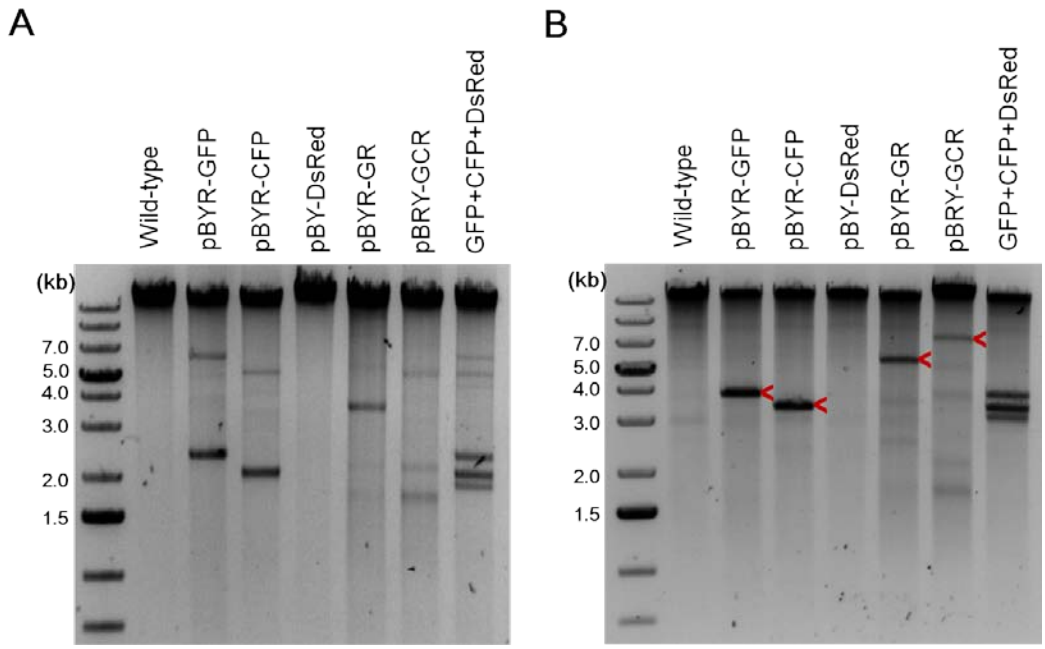


Figure 4.8 Agarose gel stained with ethidium bromide showing replicon amplification. DNA from leaves of uninfiltrated (Wt), or infiltrated with indicated vectors were test on a 0.8% agarose gel before (A) and after (B) restriction enzyme digestion. GFP+CFP+DsRed indicates co-infiltrated sample with *Agrobacterium* mixture of pBYR-GFP, pBYR-CFP and pBY-DsRed. Note that construct pBY-DsRed does not contain Rep/RepA gene. Restriction enzyme XhoI was used for pBYR-GFP, pBYR-CFP, pBY-DsRed, pBYR-GR, and GFP+CFP+DsRed. For pBYR-GCR, restriction enzyme SalI was used. Expected replicon positions are indicated with arrow heads.

Table 4.1 Sequence of oligonucleotides used in this study

Name	Sequence (5' to 3')
GR5-1	GCGGTACCCAATTCGCCCTATAGTGAGTCG
GR5-2	GTGTCGTGCTCCACCATGCCGTCGACGCACTAGTCGATAGCTTGATGCATGTTGTC
GR3-1	GACAACATGCATCAAGCTATCGACTAGTGCGTCGACGGCATGGTGGAGCACGACAC
GR3-2	GAGAGCTCCACCGCGGTGGC
HL5-1	GTGGTACCGAGCTCTCTCAACAATCTAGC
HL5-2	GTGTCGTGCTCCACCATGCCACCGGTACACTAGTCGCTTCAAGACGTGCTCAAATC
HL3-1	GATTTGAGCACGTCTTGAAGCGACTAGTGTACCGGTGGCATGGTGGAGCACGACAC
HL3-2	TGGGTACCTAACACTCTCCCCTGTTGAAGCTC

REFERENCES

Chapter 1

Abranches R, Shultz RW, Thompson WF, Allen GC. Matrix attachment regions and regulated transcription increase and stabilize transgene expression. Plant Biotechnol J. 2005;3(5):535-543.

Airenne KJ, Laitinen OH, Alenius H, Mikkola J, Kalkkinen N, Arif SA, Yeang HY, Palosuo T, Kulomaa MS. Avidin is a promising tag for fusion proteins produced in baculovirus-infected insect cells. Protein Expr Purif. 1999;17(1):139-145.

Allen GC, Hall GE Jr, Childs LC, Weissinger AK, Spiker S, Thompson WF. Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. Plant Cell. 1993;5(6):603-613.

Avesani L, Marconi G, Morandini F, Albertini E, Bruschetta M, Bortesi L, Pezzotti M, Porceddu A. Stability of Potato virus X expression vectors is related to insert size: implications for replication models and risk assessment. Transgenic Res. 2007;16(5):587-597.

Barbi T, Irons SL, Pepponi I, Hawes C, Ma JK, Drake PM. Expression and plasma membrane localization of the mammalian B-cell receptor complex in transgenic Nicotiana tabacum. Plant Biotechnol J. 2011;9(4):455-465.

Batard Y, Hehn A, Nedelkina S, Schalk M, Pallett K, Schaller H, Werck-Reichhart D. Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems. Arch Biochem Biophys. 2000;379(1):161-169.

Baulcombe DC. Gene silencing: RNA makes RNA makes no protein. Curr Biol. 1999;9(16):R599-601.

Gleave AP. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Mol Biol. 1992;20(6):1203-1207.

Boehm R. Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. Ann N Y Acad Sci. 2007;1102:121-134.

Bouquin T, Thomsen M, Nielsen LK, Green TH, Mundy J, Dziegiel MH. Human anti-rhesus D IgG1 antibody produced in transgenic plants. Transgenic Res. 2002;11(2):115-122.

Brisson N, Paszkowski J, Penswick R, Gronenbom B, Potrykus I, Hohn T. Expression of a bacterial gene in plants by using a viral vector. Nature. 1984;310:511-514.

Chanfreau G, Noble SM, Guthrie C. Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF) Science. 1996;274:1511–1514.

Chapman S, Kavanagh T, Baulcombe D. Potato virus X as a vector for gene expression in plants. Plant J. 1992;2(4):549-557.

Chen Q, He J, Phoolcharoen W, Mason HS. Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. Hum Vaccin. 2011;7(3):331-338.

Chikwamba RK, Scott MP, Mejía LB, Mason HS, Wang K. Localization of a bacterial protein in starch granules of transgenic maize kernels. Proc Natl Acad Sci U S A. 2003;100(19):11127-11132.

Comai L, Moran P, Maslyar D. Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. Plant Mol Biol. 1990;15(3):373-381.

Conrad U, Fiedler U. Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. Plant Mol Biol. 1998 ;38(1-2):101-109.

Danckwardt S, Hentze MW, Kulozik AE. 3' end mRNA processing: molecular mechanisms and implications for health and disease. EMBO J. 2008;27(3):482-498.

Dawson, W. O., P. Bubrick and G. L. Grantham. Modifications of the tobacco mosaic virus coat protein gene affecting replication, movement, and symptomatology. Phytopathol. 1998;78:783-789.

De Neve M, De Buck S, Jacobs A, Van Montagu M, Depicker A. T-DNA integration patterns in co-transformed plant cells suggest that T-DNA repeats originate from co-integration of separate T-DNAs. Plant J. 1997;11(1):15-29.

De Neve M, De Loose M, Jacobs A, Van Houdt H, Kaluza B, Weidle U, Van Montagu M, Depicker A. Assembly of an antibody and its derived antibody fragment in Nicotiana and Arabidopsis. Transgenic Res. 1993;2(4):227-237.

Ding SW, Li H, Lu R, Li F, Li WX. RNA silencing: a conserved antiviral immunity of plants and animals. Virus Res. 2004;102(1):109-115.

Dorokhov YL, Sheveleva AA, Frolova OY, Komarova TV, Zvereva AS, Ivanov PA, Atabekov JG. Superexpression of tuberculosis antigens in plant leaves. Tuberculosis (Edinb). 2007;87(3):218-224.

Durocher Y, Butler M. Expression systems for therapeutic glycoprotein production. Curr Opin Biotechnol. 2009;20(6):700-707.

Ellis JG, Llewellyn DJ, Walker JC, Dennis ES, Peacock WJ. The ocs element: a 16 base pair palindrome essential for activity of the octopine synthase enhancer. EMBO J. 1987;6(11):3203-3208.

Garbarino JE, Oosumi T, Belknap WR. Isolation of a polyubiquitin promoter and its expression in transgenic potato plants. Plant Physiol. 1995;109(4):1371-1378.

Geyer BC, Fletcher SP, Griffin TA, Lopker MJ, Soreq H, Mor TS. Translational control of recombinant human acetylcholinesterase accumulation in plants. BMC Biotechnol. 2007;7:27-40.

Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y. Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. Proc Natl Acad Sci U S A. 2006;103(40):14701-14706.

Gleba Y, Klimyuk V, Marillonnet S. Magnifection-a new platform for expressing recombinant vaccines in plants. Vaccine. 2005;23(17-18):2042-2048.

Gomord V, Chamberlain P, Jefferis R, Faye L. Biopharmaceutical production in plants: problems, solutions and opportunities. Trends Biotechnol. 2005;23(11):559-565.

Green BJ, Fujiki M, Mett V, Kaczmarczyk J, Shamloul M, Musiychuk K, Underkoffler S, Yusibov V, Mett V. Transient protein expression in three Pisum sativum (green pea) varieties. Biotechnol J. 2009;4(2):230-237.

Guerineau, F., Lucy, A. and Mullineaux, P. Effect of two consensus sequences preceding the translation initiator codon on gene expression in plant protoplasts. Plant Mol. Biol. 1992; 18:815–818.

Halweg C, Thompson WF, Spiker S. The rb7 matrix attachment region increases the likelihood and magnitude of transgene expression in tobacco cells: a flow cytometric study. Plant Cell. 2005;17(2):418-429.

Hammond SM, Bernstein E, Beach D, Hannon GJ. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature. 2000;404(6775):293-296.

Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. Argonaute2, a link between genetic and biochemical analyses of RNAi. Science. 2001;293(5532):1146-1150.

Huang Z, Chen Q, Hjelm B, Arntzen C, Mason H. A DNA replicon system for rapid high-level production of virus-like particles in plants. Biotechnol Bioeng. 2009;103(4):706-714.

Huang Z, Phoolcharoen W, Lai H, Piensook K, Cardineau G, Zeitlin L, Whaley KJ, Arntzen CJ, Mason HS, Chen Q. High-level rapid production of full-size monoclonal antibodies in plants by a single-vector DNA replicon system. Biotechnol Bioeng. 2010;106(1):9-17.

Hunt AG. Messenger RNA 3' end formation in plants. Curr Top Microbiol Immunol. 2008;326:151-177.

Hutvágner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 2001;293(5531):834-848.

Kapila J, De Rycke R, Van Montagu M, Angenon G. An Agrobacterium-mediated transient gene expression system for intact leaves.Plant Science. 1997; 122(1):101-108.

Komarova TV, Baschieri S, Donini M, Marusic C, Benvenuto E, Dorokhov YL. Transient expression systems for plant-derived biopharmaceuticals. Expert Rev Vaccines. 2010;9(8):859-876.

Kong Q, Richter L, Yang YF, Arntzen CJ, Mason HS, Thanavala Y. Oral immunization with hepatitis B surface antigen expressed in transgenic plants. Proc Natl Acad Sci USA 2001; 98:11539-11544.

Kühn U, Gündel M, Knoth A, Kerwitz Y, Rüdel S, Wahle E. Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. J Biol Chem. 2009;284(34):22803-22814.

Langeveld JP, Brennan FR, Martínez-Torrecuadrada JL, Jones TD, Boshuizen RS, Vela C, Casal JI, Kamstrup S, Dalsgaard K, Meloen RH, Bendig MM,

Hamilton WD. Inactivated recombinant plant virus protects dogs from a lethal challenge with canine parvovirus. Vaccine. 2001;19(27):3661-3670.

Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol Cell. 2002;9(5):1101-1111.

Liu L, Davies JW, Stanley J. Mutational analysis of bean yellow dwarf virus, a geminivirus of the genus Mastrevirus that is adapted to dicotyledonous plants. J Gen Virol 1998; 79:2265-2274.

Luo Z, Chen Z. Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. Plant Cell. 2007;19(3):943-958.

MacFarlane SA, Gilmer D, Davies JW. Efficient inoculation with CaMV 35 S promoter-driven DNA clones of the tobravirus PEBV. Virology. 1992;187(2):829-831.

Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci. 2008;65(7-8):1099-1122.

Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. Crosstalk between mRNA 3' end processing and transcription initiation. Mol Cell. 2010;40(3):410-422.

Marillonnet S, Giritch A, Gils M, Kandzia R, Klimyuk V, Gleba Y. In planta engineering of viral RNA replicons: efficient assembly by recombination of DNA modules delivered by Agrobacterium. Proc Natl Acad Sci U S A. 2004;101(18):6852-6857.

McCormick AA, Reddy S, Reinl SJ, Cameron TI, Czerwinkski DK, Vojdani F, Hanley KM, Garger SJ, White EL, Novak J, Barrett J, Holtz RB, Tusé D, Levy R. Plant-produced idiotype vaccines for the treatment of non-Hodgkin's lymphoma: safety and immunogenicity in a phase I clinical study. Proc Natl Acad Sci U S A. 2008;105(29):10131-10136.

Mechtcheriakova IA, Eldarov MA, Nicholson L, Shanks M, Skryabin KG, Lomonossoff GP. The use of viral vectors to produce hepatitis B virus core particles in plants. J Virol Methods. 2006;131(1):10-15.

Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res. 2010;38(9):2757-2774.

Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell. 2009 Feb 20;136(4):688-700.

Mor TS, Moon YS, Palmer KE, Mason HS. Geminivirus vectors for high-level expression of foreign proteins in plant cells. Biotechnol Bioeng. 2003;81(4):430-437.

Nagaya S, Kawamura K, Shinmyo A, Kato K. The HSP terminator of Arabidopsis thaliana increases gene expression in plant cells. Plant Cell Physiol. 2010;51(2):328-332.

Obembe OO, Popoola JO, Leelavathi S, Reddy SV. Advances in plant molecular farming. Biotechnol Adv. 2011;29(2):210-222

Palmer KE, Rybicki EP. The molecular biology of mastreviruses. Adv Virus Res 1998; 50:183-234.

Paul M, Ma JK. Plant-made pharmaceuticals: leading products and production platforms. Biotechnol Appl Biochem. 2011;58(1):58-67.

Pietrzak M, Shillito RD, Hohn T, Potrykus I. Expression in plants of two bacterial antibiotic resistance genes after protoplast transformation with a new plant expression vector. Nucleic Acids Res. 1986;14(14):5857-5868.

Porta C, Spall VE, Loveland J, Johnson JE, Barker PJ, Lomonossoff GP. Development of cowpea mosaic virus as a high-yielding system for the presentation of foreign peptides. Virology. 1994;202(2):949-955.

Potenza C, Aleman L, Sengupta-Gopalan C. Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. In Vitro Cell Dev Biol. 2004;40:1–22

Regnard GL, Halley-Stott RP, Tanzer FL, Hitzeroth II, Rybicki EP. High level protein expression in plants through the use of a novel autonomously replicating geminivirus shuttle vector. Plant Biotechnol J. 2010;8(1):38-46.

Rigano MM, Alvarez ML, Pinkhasov J, Jin Y, Sala F, Arntzen CJ, Walmsley AM. Production of a fusion protein consisting of the enterotoxigenic Escherichia coli heat-labile toxin B subunit and a tuberculosis antigen in Arabidopsis thaliana. Plant Cell Rep. 2004;22(7):502-508.

Sainsbury F, Lavoie PO, D'Aoust MA, Vézina LP, Lomonossoff GP. Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. Plant Biotechnol J. 2008;6(1):82-92.

Sainsbury F, Lomonossoff GP. Extremely high-level and rapid transient protein production in plants without the use of viral replication. Plant Physiol. 2008;148(3):1212-1218.

Sainsbury F, Lavoie PO, D'Aoust MA, Vézina LP, Lomonossoff GP. Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. Plant Biotechnol. J. 2008; 6:82–92.

Saunders K, Lucy A, Stanley J. DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. Nucleic Acids Res. 1991;19(9):2325-2330.

Scholthof HB, Scholthof KB, Jackson AO. Plant virus gene vectors for transient expression of foreign proteins in plants. Annu Rev Phytopathol. 1996;34:299-323.

Schouten A, Roosien J, van Engelen FA, de Jong GA, Borst-Vrenssen AW, Zilverentant JF, Bosch D, Stiekema WJ, Gommers FJ, Schots A, Bakker J. The C-terminal KDEL sequence increases the expression level of a single-chain antibody designed to be targeted to both the cytosol and the secretory pathway in transgenic tobacco. Plant Mol Biol. 1996;30(4):781-793.

Shaw CH, Carter GH, Watson MD, Shaw CH. A functional map of the nopaline synthase promoter. Nucleic Acids Res. 1984;12(20):7831-7846.

Stefanov I, Ilubaev S, Fehér A, Margóczi K, Dudits D. Promoter and genotype dependent transient expression of a reporter gene in plant protoplasts. Acta Biol Hung. 1991;42(4):323-330.

Tacket CO, Mason HS, Losonsky G, Estes MK, Levine MM, Arntzen CJ. Human immune responses to a novel norwalk virus vaccine delivered in transgenic potatoes. J Infect Dis 2000; 182:302-305.

Turner R, Foster GD. The potential exploitation of plant viral translational enhancers in biotechnology for increased gene expression. Mol Biotechnol. 1995;3(3):225-236.

Verch T, Yusibov V, Koprowski H. Expression and assembly of a full-length monoclonal antibody in plants using a plant virus vector. J Immunol Methods. 1998;220(1-2):69-75.

Yang L, Wakasa Y, Kawakatsu T, Takaiwa F. The 3'-untranslated region of rice glutelin GluB-1 affects accumulation of heterologous protein in transgenic rice. Biotechnol Lett. 2009;31(10):1625-1631.

Yusibov V, Modelska A, Steplewski K, Agadjanyan M, Weiner D, Hooper DC, Koprowski H. Antigens produced in plants by infection with chimeric plant viruses immunize against rabies virus and HIV-1. Proc Natl Acad Sci U S A. 1997;94(11):5784-5788.

Yusibov V, Streatfield SJ, Kushnir N. Clinical development of plant-produced recombinant pharmaceuticals: vaccines, antibodies and beyond. Hum Vaccin. 2011;7(3):313-321.

Zamore PD, Tuschl T, Sharp PA, Bartel DP. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell. 2000;101(1):25-33.

Zelada AM, Calamante G, de la Paz Santangelo M, Bigi F, Verna F, Mentaberry A, Cataldi A. Expression of tuberculosis antigen ESAT-6 in Nicotiana tabacum using a potato virus X-based vector. Tuberculosis (Edinb). 2006; 86:263–267.

Zhang X, Mason H. Bean Yellow Dwarf Virus replicons for high-level transgene expression in transgenic plants and cell cultures. Biotechnol Bioeng. 2006;93(2):271-279.

Zhao H, Xing D, Li QQ. Unique features of plant cleavage and polyadenylation specificity factor revealed by proteomic studies. Plant Physiol. 2009;151(3):1546-1556.

Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev. 1999;63(2):405-445.

Chapter 2

Ali S, Taylor WC. The 3' non-coding region of a C4 photosynthesis gene increases transgene expression when combined with heterologous promoters. Plant Mol Biol. 2001;46(3);325-333.

Arntzen CJ. Plant science. Using tobacco to treat cancer. Science. 2008;321(5892):1052-1053.

Carrington JC, Freed DD, Leinicke AJ. Bipartite signal sequence mediateds nuclear translocation of the plant potyviral NIa protein. Plant Cell. 1991;3(9):953-962.

Chung BY, Simons C, Firth AE, Brown CM, Hellens RP. Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. BMC Genomics. 2006;7:120.

Collens JI, Mason HS, Curtis WR. Agrobacterium-mediated viral vector-amplified transient gene expression in Nicotiana glutinosa plant tissue culture. Biotechnol Prog. 2007;23(3):570-576.

Gilmartin GM. Eukaryotic mRNA 3' processing: a common means to different ends. Genes Dev. 2005;19(21):2517-2521.

Gimmi ER, Reff ME, Deckman IC. Alterations in the pre-mRNA topology of the bovine growth hormone polyadenylation region decrease poly(A) site efficiency. Nucleic Acids Res. 1989;17(17):6983-6998.

Graveley BR, Fleming ES, Gilmartin GM. RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor. Mol Cell Biol. 1996;16(9):4942-4951.

Hirsinger C, Parmentier Y, Durr A, Fleck J, Jamet E. Characterization of a tobacco extensin gene and regulation of its gene family in healthy plants and under various stress conditions. Plant Mol Biol. 1997;33(2):279-289.

Huang Z, Mason HS. Conformational analysis of hepatitis B surface antigen fusions in an Agrobacterium-mediated transient expression system. Plant Biotechnol J. 2004;2(3):241-249.

Hunt AG. Messenger RNA 3' end formation in plants. Curr Top Microbiol Immunol. 2008;326:151-177.

Ingelbrecht IL, Herman LM, Dekeyser RA, Van Montagu MC, Depicker AG. Different 3' end regions strongly influence the level of gene expres-sion in plant cells. Plant Cell. 1989;1(7):671-680.

Ji G, Zheng J, Shen Y, Wu X, Jiang R, Lin Y, Loke JC, Davis KM, Reese GJ, Li QQ. Predictive modeling of plant messenger RNA polyadenylation sites. BMC Bioinformatics. 2007;8:43.

Jiang X, Wang M, Graham DY, Estes MK. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. J Virol. 1992;66(11):6527-6532.

Kooiker M, Airoldi CA, Losa A, Manzotti PS, Finzi L, Kater MM, Colombo L. BASIC PENTACYSTEINE1, a GA binding protein that induces confor-mational changes in the regulatory region of the homeotic Arabidopsis gene SEEDSTICK. Plant Cell. 2005;17(3):722-729.

Koprowski H, Yusibov V. The green revolution: plants as heterologous expression vectors. Vaccine. 2001;19(17-19):2735-3741.

Loke JC, Stahlberg EA, Strenski DG, Haas BJ, Wood PC, Li QQ. Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures. Plant Physiol. 2005;138(3):1457-1468.

Luo Z, Chen Z. Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. Plant Cell. 2007;19(3):943-958.

Mathew LG, Maloney B, Takeda N, Mason HS. Spurious polyadenylation of Norovirus Narita 104 capsid protein mRNA in transgenic plants. Plant Mol Biol. 2011;75(3):263-275.

Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci. 2008;65(7-8):1099-1122.

Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. Crosstalk between mRNA 3' end processing and transcription initia-tion. Mol Cell. 2010;40(3):410-422.

Menossi M, Rabaneda F, Puigdomènech P, Martínez-Izquierdo JA. Analysis of regulatory elements of the promoter and the 3' untranslated region of the maize Hrgp gene coding for a cell wall protein. Plant Cell Rep. 2003;21(9):916-923.

Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res. 2010;38(9):2757-2774.

Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell. 2009;136(4):688-700.

Mor TS, Moon YS, Palmer KE, Mason HS. Geminivirus vectors for high-level expression of foreign proteins in plant cells. Biotechnol Bioeng. 2003;81(4):430-437.

Nagaya S, Kawamura K, Shinmyo A, Kato K. The HSP terminator of Arabidopsis thaliana increases gene expression in plant cells. Plant Cell Physiol. 2010;51(2):328-332.

Rose AB. Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol. 2008;326:277-290.

Richter LJ, Thanavala Y, Arntzen CJ, Mason HS. Production of hepatitis B surface antigen in transgenic plants for oral immunization. Nat Biotechnol. 2000; 18(11):1167-1171.

Sharma AK, Sharma MK. Plants as bioreactors: Recent developments and emerging opportunities. Biotechnol Adv. 2009;27(6):811-832.

Sheldon CC, Conn AB, Dennis ES, Peacock WJ. Different regulatory regions are required for the vernalization-induced repression of FLOWERING LOCUS C and for the epigenetic maintenance of repression. Plant Cell. 2002;14(10):2527-2537.

Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR. A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiol. 2010;153(2):485-513.

Sieburth LE, Meyerowitz EM. Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. Plant Cell. 1997;9(3):355-365.

Streatfield SJ. Approaches to achieve high-level heterologous protein production in plants. Plant Biotechnol J. 2007;5(1):2-15.

Thanavala Y, Huang Z, Mason HS. Plant-derived vaccines: a look back at the highlights and a view to the challenges on the road ahead. Expert Rev Vaccines. 2006;5(2):249-260.

Withers-Martinez C, Carpenter EP, Hackett F, Ely B, Sajid M, Grainger M, Blackman MJ. PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. Protein Eng. 1999;12(12):1113-1120.

Xing A, Moon BP, Mills KM, Falco SC, Li Z. Revealing frequent alternative polyadenylation and widespread low-level transcription read-through of novel plant transcription terminators. Plant Biotechnol J. 2010;8(7):772-782.

Yang L, Wakasa Y, Kawakatsu T, Takaiwa F. The 3'-untranslated region of rice glutelin GluB-1 affects accumulation of heterologous protein in transgenic rice. Biotechnol Lett. 2009;31(10):1625-1631.

Zhang X, Mason H. Bean Yellow Dwarf virus replicons for high-level transgene expression in transgenic plants and cell cultures. Biotechnol Bioeng. 2006;93(2):271-279.

Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev. 1999;63(2):405-445.

Chapter 3

Abramovitch RB, Janjusevic R, Stebbins CE, Martin GB. Type III effector AvrPtoB requires intrinsic E3 ubiquitin ligase activity to suppress plant cell death and immunity. Proc Natl Acad Sci U S A. 2006;103(8):2851-2856.

Carrington JC, Freed DD, Leinicke AJ. Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein. Plant Cell. 1991;3(9):953-962.

Chen Q, He J, Phoolcharoen W, Mason HS. Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. Hum Vaccin. 2011;7(3):331-338.

Chung BY, Simons C, Firth AE, Brown CM, Hellens RP. Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. BMC Genomics. 2006;7:120.

Danckwardt S, Hentze MW, Kulozik AE. 3' end mRNA processing: molecular mechanisms and implications for health and disease. EMBO J. 2008;27(3):482-498.

Huang Z, Chen Q, Hjelm B, Arntzen C, Mason H. A DNA replicon system for rapid high-level production of virus-like particles in plants. Biotechnol Bioeng. 2009;103(4):706-714.

Huang Z, Elkin G, Maloney BJ, Beuhner N, Arntzen CJ, Thanavala Y, Mason HS. Virus-like particle expression and assembly in plants: hepatitis B and Norwalk viruses. Vaccine. 2005;23(15):1851-1858.

Huang Z, Mason HS. Conformational analysis of hepatitis B surface antigen fusions in an Agrobacterium-mediated transient expression system. Plant Biotechnol J. 2004;2(3):241-249.

Huang Z, Phoolcharoen W, Lai H, Piensook K, Cardineau G, Zeitlin L, Whaley KJ, Arntzen CJ, Mason HS, Chen Q. High-level rapid production of full-size monoclonal antibodies in plants by a single-vector DNA replicon system. Biotechnol Bioeng. 2010;106(1):9-17.

Hunt AG. Messenger RNA 3' end formation in plants. Curr Top Microbiol Immunol. 2008;326:151-177.

Ingelbrecht IL, Herman LM, Dekeyser RA, Van Montagu MC, Depicker AG. Different 3' end regions strongly influence the level of gene expression in plant cells. Plant Cell. 1989;1(7):671-680.

Jefferson RA, Kavanagh TA, Bevan MW. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 1987;6(13):3901-3907.

Jeske H. Geminiviruses. Curr Top Microbiol Immunol. 2009;331:185-226.

Jiang X, Wang M, Graham DY, Estes MK. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. J Virol. 1992;66(11):6527-6532.

Lico C, Chen Q, Santi L. Viral vectors for production of recombinant proteins in plants. J Cell Physiol. 2008;216(2):366-377.

Luo Z, Chen Z. Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. Plant Cell. 2007;19(3):943-958.

Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci. 2008;65(7-8):1099-1122.

Mason HS, Ball JM, Shi JJ, Jiang X, Estes MK, Arntzen CJ. Expression of Norwalk virus capsid protein in transgenic tobacco and potato and its oral immunogenicity in mice. Proc Natl Acad Sci U S A. 1996;93(11):5335-5340.

Menossi M, Rabaneda F, Puigdomènech P, Martínez-Izquierdo JA. Analysis of regulatory elements of the promoter and the 3' untranslated region of the maize Hrgp gene coding for a cell wall protein. Plant Cell Rep. 2003;21(9):916-923.

Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res. 2010;38(9):2757-2774.

Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell. 2009 Feb 20;136(4):688-700.

Mor TS, Moon YS, Palmer KE, Mason HS. Geminivirus vectors for high-level expression of foreign proteins in plant cells. Biotechnol Bioeng. 2003;81(4):430-437.

Outchkourov NS, Peters J, de Jong J, Rademakers W, Jongsma MA. The promoter-terminator of chrysanthemum rbcS1 directs very high expression levels in plants. Planta. 2003;216(6):1003-10012.

Palmer KE, Rybicki EP. The molecular biology of mastreviruses. Adv Virus Res. 1998;50:183-234.

Regnard GL, Halley-Stott RP, Tanzer FL, Hitzeroth II, Rybicki EP. High level protein expression in plants through the use of a novel autonomously replicating geminivirus shuttle vector. Plant Biotechnol J. 2010;8(1):38-46.

Richter LJ, Thanavala Y, Arntzen CJ, Mason HS. Production of hepatitis B surface antigen in transgenic plants for oral immunization. Nat Biotechnol. 2000; 18(11):1167-1171.

Saunders K, Lucy A, Stanley J. DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. Nucleic Acids Res. 1991;19(9):2325-2330.

Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR. A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiol. 2010;153(2):485-513.

Zhang X, Mason H. Bean Yellow Dwarf Virus replicons for high-level transgene expression in transgenic plants and cell cultures. Biotechnol Bioeng. 2006;93(2):271-279.

Chapter 4

Arntzen CJ. Plant science. Using tobacco to treat cancer. Science. 2008;321(5892):1052-1053.

Bedoya L, Martínez F, Rubio L, Daròs JA. Simultaneous equimolar expression of multiple proteins in plants from a disarmed potyvirus vector. J Biotechnol. 2010;150(2):268-275.

Bouquin T, Thomsen M, Nielsen LK, Green TH, Mundy J, Dziegiel MH. Human anti-rhesus D IgG1 antibody produced in transgenic plants. Transgenic Res. 2002; 11(2):115-122.

Bowman CC, Clements JD. Differential biological and adjuvant activities of cholera toxin and Escherichia coli heat-labile enterotoxin hybrids. Infect Immun. 2001; 69(3):1528-1535.

Cárdenas L, Clements JD. Development of mucosal protection against the heat-stable enterotoxin (ST) of Escherichia coli by oral immunization with a genetic fusion delivered by a bacterial vector. Infect Immun. 1993; 61(11):4629-4636.

Clements JD, Hartzog NM, Lyon FL. Adjuvant activity of Escherichia coli heat-labile enterotoxin and effect on the induction of oral tolerance in mice to unrelated protein antigens. Vaccine. 1988; 6(3):269-277.

Comai L, Moran P, Maslyar D. Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. Plant Mol Biol. 1990; 15(3):373-381.

Daniell H, Singh ND, Mason H, Streatfield SJ. Plant-made vaccine antigens and biopharmaceuticals. Trends Plant Sci. 2009;14(12):669-679.

Dietrich C, Maiss E. Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected Nicotiana benthamiana plants. J Gen Virol. 2003;84(Pt 10):2871-2876.

Divéki Z, Salánki K, Balázs E. Limited utility of blue fluorescent protein (BFP) in monitoring plant virus movement. Biochimie. 2002;84(10):997-1002.

Ellis JG, Llewellyn DJ, Walker JC, Dennis ES, Peacock WJ. The ocs element: a 16 base pair palindrome essential for activity of the octopine synthase enhancer. EMBO J. 1987; 6(11):3203-3208.

Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y. Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. Proc Natl Acad Sci U S A. 2006;103(40):14701-14706.

Gomord V, Chamberlain P, Jefferis R, Faye L. Biopharmaceutical production in plants: problems, solutions and opportunities. Trends Biotechnol. 2005;23(11):559-565.

Hiatt A, Cafferkey R, Bowdish K. Production of antibodies in transgenic plants. Nature. 1989; 342(6245):76-78.

Huang Z, Chen Q, Hjelm B, Arntzen C, Mason H. A DNA replicon system for rapid high-level production of virus-like particles in plants. Biotechnol Bioeng. 2009;103(4):706-714.

Huang Z, Phoolcharoen W, Lai H, Piensook K, Cardineau G, Zeitlin L, Whaley KJ, Arntzen CJ, Mason HS, Chen Q. High-level rapid production of full-size monoclonal antibodies in plants by a single-vector DNA replicon system. Biotechnol Bioeng. 2010; 106(1):9-17.

Komarova TV, Baschieri S, Donini M, Marusic C, Benvenuto E, Dorokhov YL. Transient expression systems for plant-derived biopharmaceuticals. Expert Rev Vaccines. 2010;9(8):859-876.

Liang S, Hajishengallis G. Heat-labile enterotoxins as adjuvants or anti-inflammatory agents. Immunol Invest. 2010; 39(4-5):449-467.

Lico C, Chen Q, Santi L. Viral vectors for production of recombinant proteins in plants. J Cell Physiol. 2008;216(2):366-377.

Liu L, Davies JW, Stanley J. Mutational analysis of bean yellow dwarf virus, a geminivirus of the genus Mastrevirus that is adapted to dicotyledonous plants. J Gen Virol 1998; 79:2265-2274.

Ma JK, Hiatt A, Hein M, Vine ND, Wang F, Stabila P, van Dolleweerd C, Mostov K, Lehner T. Generation and assembly of secretory antibodies in plants. Science. 1995; 268(5211):716-719.

Mor TS, Moon YS, Palmer KE, Mason HS. Geminivirus vectors for high-level expression of foreign proteins in plant cells. Biotechnol Bioeng. 2003; 81(4):430-437.

Obembe OO, Popoola JO, Leelavathi S, Reddy SV. Advances in plant molecular farming. Biotechnol Adv. 2011;29(2):210-222

Palmer KE, Rybicki EP. The molecular biology of mastreviruses. Adv Virus Res. 1998; 50:183-234.

Phoolcharoen W, Bhoo SH, Lai H, Ma J, Arntzen CJ, Chen Q, Mason HS. Expression of an immunogenic Ebola immune complex in Nicotiana benthamiana. Plant Biotechnol J. 2011; 9(7):807-816.

Regnard GL, Halley-Stott RP, Tanzer FL, Hitzeroth II, Rybicki EP. High level protein expression in plants through the use of a novel autonomously replicating geminivirus shuttle vector. Plant Biotechnol J. 2010; 8(1):38-46.

Roy G, Weisburg S, Foy K, Rabindran S, Mett V, Yusibov V. Co-expression of multiple target proteins in plants from a tobacco mosaic virus vector using a combination of homologous and heterologous subgenomic promoters. Arch Virol. 2011;156(11):2057-2061.

Sainsbury F, Lavoie PO, D'Aoust MA, Vézina LP, Lomonossoff GP. Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. Plant Biotechnol J. 2008; 6(1):82-92.

Shaw CH, Carter GH, Watson MD, Shaw CH. A functional map of the nopaline synthase promoter. Nucleic Acids Res. 1984;12(20):7831-7846.

Vieira HL, Estêvão C, Roldão A, Peixoto CC, Sousa MF, Cruz PE, Carrondo MJ, Alves PM. Triple layered rotavirus VLP production: kinetics of vector replication, mRNA stability and recombinant protein production. J Biotechnol. 2005;120(1):72-82.

Williams NA, Hirst TR, Nashar TO. Immune modulation by the cholera-like enterotoxins: from adjuvant to therapeutic. Immunol Today. 1999; 20(2):95-101.

Zhang X, Mason H. Bean Yellow Dwarf Virus replicons for high-level transgene expression in transgenic plants and cell cultures. Biotechnol Bioeng. 2006; 93(2):271-279.