Characterisation of the Potyvirus VPg Interacting Protein (PVIP) gene from *Nicotiana benthamiana*

Nitish Anand

A thesis submitted to
Auckland University of Technology
in partial fulfilment of the requirements for the degree
of Master of Applied Science (MAppSc)

2010

School of Applied Sciences

Primary Supervisor: Dr. Colleen Higgins

Table of contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a University or other institution of higher learning.

Nitish Anand.

Acknowledgements

This work was undertaken during the year 2009-2010 at the Faculty of Health and Environmental Sciences of Auckland University of Technology (AUT). I would like to acknowledge my work to people who, with their utmost support and inspiration made this project a success.

I would like to express my deep and sincere gratitude to my supervisor, Dr. Colleen Higgins. Her wide knowledge and her logical way of thinking have been of great value for me. Her understanding and personal guidance have provided a good basis for the thesis.

I owe my sincere appreciation to the technicians present in the laboratories of AUT who have also done a great deed with aiding me in this project.

I am indebted to my family and friends and offer my deepest gratitude to them. This thesis would not have been possible without their encouragement, support and understanding.

I would like to conclude with a vote of thanks to all the other people whose names cannot be given in a very short space for their patronage.

Abstract

According to a study by Dunoyer et al (2004), a protein called Potyvirus VPg-Interacting Protein (PVIP) is known to interact with potyviruses upon infection of the host plant *Arabidopsis thaliana* and help in their replication and movement. Potyvirus are the largest group of plant pathogens causing severe damage to most crops. PVIP appears to be a plant specific protein that occurs in uninfected tissues, suggesting a role which is independent of viral infection. Potyviruses appear to highjack this protein for their own needs. What the function of PVIP in healthy cells is unclear. This study aimed to determine the function of PVIP in *Nicotiana benthamiana.*

Bioinformatic analysis was used to identify sequence homologues of the *N.benthamiana* PVIP protein sequence, from all the plant species. Phylogenetic analysis of these homologues showed this to be an ancient gene, being present in moss, ferns, gymnosperms and angiosperms, and that there are likely to be two genes that have arisen via gene duplication. Only one gene from *N.benthamiana* has been identified, thus, there is likely to be another gene to be found. The function of many of these homologues, together with the identification of a PHD finger domain within PVIP from *A.thaliana* suggested a role for the protein in transcriptional regulation. The 'PHD finger' domain being plant specific; proteins with this domain are likely to have a role in a plant specific function.

Analysis of the promoter of the PVIP gene may provide more insight into gene functionality and also the hidden transcriptional network of the gene. Bioinformatic analysis of the 5' flanking regions of the PVIP homologues from *A.thaliana* and *Oryza sativa* (rice), predicted that the PVIP gene would be responsive to changes in conditions of light and have a role in development. The closest homologues from *A.thaliana*, OBERON (OBE) 1 and 2 have been shown to functionally overlap with roles in apical and root meristem development/maintenance. The OBERON proteins also control the transcription of genes required for auxin responses through the action of their PHD finger domains. This suggests that the rice homologues on chromosomes 11 and 12, as well as the *N. benthamiana* proteins are functionally equivalent. Southern hybridisation and inverse polymerase chain reaction were used to isolate the 5' flanking region from the *N.benthamiana* PVIP gene. However, all attempts using either technique were unsuccessful.

Quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) was used to study the relative amounts of PVIP mRNA in healthy tissues from *N.benthamiana* as well as how it responded to dark/light treatment. PVIP mRNA was shown to be expressed at the mRNA level in leaf, stem, root and flower tissue to varying degrees. The mRNA was most abundant in stem and least abundant in roots.

The PVIP gene was predicted to be light responsive since potential binding sites for light responsive transcription factors were identified within the 5' flanking regions of the *A.thaliana* and rice homologous genes. Leaf tissue from *N.benthamiana* exposed to continuous dark for 48 hours was shown to have 0.3 to 3 fold change of PVIP mRNA compared with leaf exposed to continuous cycles of of 16 hours light / 8 hours dark. Leaf that had been exposed to the dark for 24 hours and then returned to 16 hours light/8 hours dark for one cycle showed lower levels of PVIP mRNA than leaf continuously exposed to the standard day/night cycles. These findings suggest that PVIP is indeed responsive to changes in light conditions.

PVIP has previously been identified as a protein that interacts with potyviruses and has been implicated as having a role in viral movement through an infected plant. It's role in healthy tissue in unknown. However, bioinformatic analysis suggested that it has a role in transcription and it is likely to be developmentally and/or light regulated. Responsiveness to light was confirmed by RT-qPCR as was its expression in several tissues. This study suggests it has a role in flowering and shoot and root induction. It may also prevent pathogen attack and bring about DNA repair.

Chapter 1
Introduction

1.1 Gene structure

Deoxyribonucleic acid (DNA) is the basic unit of heredity and is the genetic material of both, eukaryotes and prokaryotes (Russell, 2001). It is also the genetic material of certain viruses which infects these eukaryotes and prokaryotes (Lodish et al., 2000). DNA consists of two strands, coiled in the shape of a double helix. Each strand is made up of a deoxyribose sugar, a phosphate group and a base. These three subunits together are known as nucleotides. It is the arrangement of these nucleotides which makes up the DNA double helix (Hartl & Jones, 2005). The four bases present in DNA are adenine, guanine, cytosine and thymine. The sequence or the order of these bases within the DNA determines the genetic information. The DNA strands are extremely long; the largest human chromosome, for example, is about 247 million base pairs in length (Hartwell et al., 2000).

Like DNA, RNA is also the genetic material of a few viruses. It differs from the DNA by being single stranded, the sugar being ribose instead of deoxyribose and by replacing the nucleotide thymine with uracil (Russell, 2001). The different types of RNA observed in eukaryotes are the messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) and the small nuclear RNA (snRNA) (Hartwell et al., 2000). RNA is essential in prokaryotes and eukaryotes for cell functions such as gene regulation and the formation of structures called ribosomes. It is to these ribosomes that the mRNA carries the genetic information from the DNA and it is the mRNA which specifies the amino acid sequences for proteins (Russell, 2001).

Within a cell, the genetic material is organised into structures known as chromosomes. A number of linear chromosomes are contained within the nucleus of eukaryotes and each of these chromosomes consists of a single DNA molecule complexed with proteins (Hartl & Jones, 2005). The numbers of chromosomes differ from one organism to another and the total amount of genetic material in an organism's nuclear chromosomes is called the

genome. Along with the nucleus, DNA is present within the mitochondria (for plants and animals) and in the chloroplasts (for plants) as well (Russell, 2001).

A gene consists of three basic components that is; the promoter, coding region and the terminator (Figure 1.1). Promoters are the DNA sequences present at the very beginning of the gene and which signal the enzyme RNA polymerase to the initiation site for transcription (Hartwell et al., 2000). The coding region is the region which encodes for the gene and, for genes transcribed by RNA polymerase II, is translated into protein. The sequence which stops the process of transcription is known as the terminator. The promoter and terminator along with the entire gene coding region are known as the transcription unit (Lodish et al., 2000).

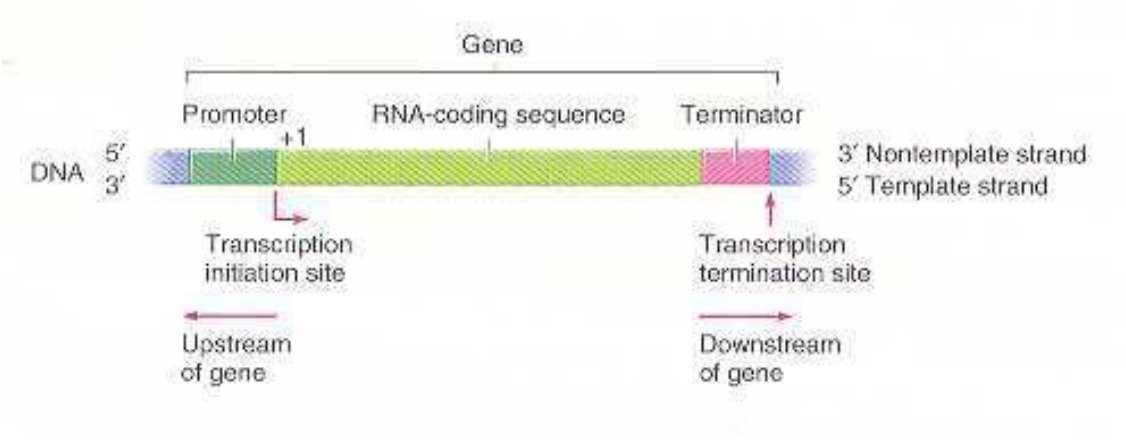


Figure 1.1. Generalised structure of a gene

From *iGenetics* p. 113, by P.J.Russel, 2001, San Francisco: Pearson Education, Inc.

It was Beadle and Tatum who, in the year 1941, showed that it is the genes which provide the necessary instructions for the synthesis of specific proteins. The process is known as the Central Dogma and involves two main steps, that of transcription and translation (Crick, 1970). Transcription involves the separation of the two DNA strands using the enzyme DNA helicase and the conversion of the DNA into RNA using the enzyme RNA polymerase. This process will be discussed further below. Translation involves the conversion of the mRNA into an amino acid sequence in proteins. The nucleotide information contained within the mRNA, specifying the amino acid sequence for protein synthesis is known as the genetic code and each amino acid is indicated by a three nucleotide sequence on the mRNA, known as the codon (Lodish et al., 2000).

Two types of genes are observed in any cell or tissue. The first being housekeeping or constitutive genes, which are always expressed within the cell and the second is facultative genes which are transcribed by the organism only when the need arises (Hartl & Jones, 2005). Transcription of the genes is further regulated by sequences known as enhancers and silencers. These sequences may be present anywhere within the genome. The former speeds up the process of transcription while the latter slows it down or might even stop the process. Enhancers operate by binding with certain protein factors called activators and form the enhanceosome, whereas the silencer binds with proteins called repressor; this keeps the DNA tightly packaged and unavailable for transcription (Lodish et al., 2000).

1.2 Promoter and transcription factors

Transcription of any gene is tightly controlled. The signal initiating transcription is embedded within the DNA itself and is known as the 'promoter' (Lodish et al., 2000). The promoter is a *cis*-acting element, very close to the coding region's transcription initiation site (TSS). The promoter also specifies the number of copies of a gene to be transcribed; that is, different promoters have different "strengths", some being strong promoters and causing efficient transcription, while others are weaker (Hartwell et al., 2000).

Interactions between proteins and DNA are essential for the genetic activity of any cell. This genetic activity may include transcription, DNA packaging, replication and also DNA repair (Rushton, Reinstadler, Lipka, Lippok & Somssich, 2002). Proteins that directly or indirectly interact with the *cis*-regulatory elements within a promoter are known as *trans*-acting elements and are encoded by their own genes. These *trans*-acting elements known as transcription factors are capable of activating or repressing the expression of the target gene. It is these transcription factors which influence the rate of transcription, that is, how "strong" a promoter is (Hartwell et al., 2000). *cis*-acting elements can be defined as DNA sequences which bind to transcription factors or DNA binding proteins thereby regulating the process of transcription. Transcription regulation depends upon the physical interaction between the transcription factors and the DNA binding sequences present within the promoters. These binding sites are, however, often poorly defined (Moyroud, Reymond, Hames, Parcy & Scutt, 2009). In eukaryotes, the regulation of genes by the transcription

factors is influenced by the degree of chromatin condensation, as also is promoter accessibility (Sekinger, Moqtaderi & Struhl, 2005). To access the promoter, the transcription factors have to alter the chromatin structure through a variety of mechanisms. Firstly, the newly synthesised DNA helices initially inherit just the parental histones and hence temporarily reveal half of their DNA to the proteins. The nucleosome-remodelling complex then displaces the protein octamers using ATP as energy, thereby increasing the association between the transcription factors and DNA (Narlikar, Fan & Kingston1, 2002). The acetylation of the N-terminal histone tail further influences the accessibility of DNA within the chromatin structure (Vettese-Dadey et al., 1996). The activator proteins then dissociate the histone proteins from the promoter and transcriptional elongation appears to disrupt nucleosomes throughout the protein coding region (Schwabish & Struhl, 2004). This greatly reduces the concentration of inappropriate binding sites, thereby providing an important mechanism to restrict gene regulation or transcription initiation only within the promoters (Sekinger, Moqtaderi & Struhl, 2005).

1.3 The core promoter

Knowledge about promoter architecture is essential in order to understand the regulation and also the evolution of promoters (Yamamoto et al., 2007). However, our current knowledge of promoters with respect to their core regions is relatively insignificant. In animals, the core promoter of protein coding genes is made up of functional elements that are common to most genes transcribed by RNA polymerase II. These elements include the TATA box, Initiator element (INR), Downstream Promoter Element (DPE), TFIIB-Recognition Element (BRE) and also the downstream core elements (DCE) (Malecova, Gross, Boyer-Guittaut, Yavuz & Oelgeschlager, 2007; Shi & Zhou, 2006). The transcriptional regulatory elements such as the DPE, DCE and BRE tend to show a random distribution within the promoters. These transcriptional regulatory elements along with the TATA box are the features which have been utilized for promoter identification (Yamamoto et al., 2007). Temporal and spatial gene expression are controlled by enhancers and silencer sequences that tend to be specific for each gene but have common sequence motifs for genes expressed in a similar manner (Venter, 2007). For genes encoded by RNA

polymerase II in plants, the promoter generally consists of the TATA box, INR, BRE and the DCE (Kornberg, 1996).

Promoter function can be analysed in two ways. One way is to alter the base sequences that is, frame-shift or point mutations (add or delete a base sequence) and to check whether these mutations affect transcription. If it affects transcription significantly, the affected nucleotides define important promoter elements. The other way may be to compare the DNA sequences present upstream of a number of protein coding genes and find regions with similar sequences. These results will show that the promoters of these genes contain the basal promoter elements and also the promoter proximal elements (Russell, 2001).

For genes transcribed by the RNA polymerase II, the basal promoter region is made up of the TATA box or the Goldberg-Hogness box (named after its discoverers) and a pyrimidine-rich sequence present near the transcription start site (TSS). In eukaryotes, this pyrimidine rich site is known as the initiator element (INR) (Russell, 2001). The TATA box indicates which strand holds the genetic information (Lodish et al., 2000). The TATA box consensus sequence for human was determined to be TATAWAWN and is located within a distance of 20 to 35 base pairs from the transcription start site (TSS). The letter W stands for either A or T and the letter N stands any of the four nucleotides present in DNA. The TATA box binding protein is known as the TBP (Promoter, 2009). It is also the first of all the core promoter elements to be identified in a eukaryotic gene encoding for different proteins (Shi & Zhou, 2006). The start codon in eukaryotes is generally AUG and this codon is present within the initiator, which is also known as the Kozak sequence (Russell, 2001). In certain TATA less promoters the initiator element acts as the initiator of transcription (Zhu, Dabi & Lamb, 1995). An initiator sequence (INR) is located around the TSS and has a sequence of $PyPyA_{+1}N(T/A)PyPy$, where Py represents pyrimidine, N represents any nucleotide and T/A means either a thymine (T) or an adenine (A) nucleotide (Lo & Smale, 1996). Interestingly, bioinformatic analysis indicates that only a minority of human genes, that is less than 20%, contain the TATA box whereas almost half of the genes contain the INR element (Malecova et al., 2007).

The sequence contained within 50-200 nucleotides upstream of the TSS is known as the promoter proximal region and contains sequences such as the CAAT box and the GC box. The CAAT binding protein is known as the CBP (Promoter, 2009). The consensus

sequence for these are CCAAT (Kusnetsov, Landsberger, Meurer & Oelmuller, 1999) and GGGCGG, respectively (Kim, Choi, Costa & An, 1992). In plants, the CAAT box is frequently referred to as the AGGA box and is located 80 bp upstream of TSS (Kusnetsov et al., 1999) whereas the GC box is located 90 bp upstream of the promoter. The CAAT and the GC boxes are known to bind to several transcription factors and hence play crucial roles within the promoter (Kim et al., 1992). Therefore, a number of these promoter elements can contribute to its function and it is not imperative for all the elements of the core promoter to be present within the promoters (Hartwell et al., 2000).

1.4 Transcription with the 'basal apparatus'

The 'basal apparatus' is comprised of a universal set of proteins which initiates the process of transcription by recognizing the core promoter. The basal apparatus constitutes RNA polymerase II, transcription factor II B (TFIIB), TFIIE, TFIIF, TFIIH, other general transcription factors and the TATA-binding protein. RNA polymerase II binds to the promoter and initiates the process of transcription at a nucleotide in or near the promoter called the TSS, usually present approximately 30 nucleotides downstream of the core promoter (Figure 1.2) (Hartl & Jones, 2005). Promoter sequences are different for different genes but are gene specific. Also, there are certain regions within the promoter to which certain transcription factors, namely enhancers or repressors, bind and these regions are located at a great distance upstream or downstream from the core promoter. Initially, the TFIID binds to the TATA box and hence forms the 'initial committed complex'. Other basal transcription complexes then follow and bind to this initial complex to form the 'minimal transcription initiation complex' and finally the 'complete transcription or the pre-initiation complex', because it is at this stage that the complex is ready to start the process of transcription (Russell, 2001). Thus, transcription involves the binding of the transcription factors to a specific *cis*-regulatory element within the promoter itself (Rushton et al., 2002).

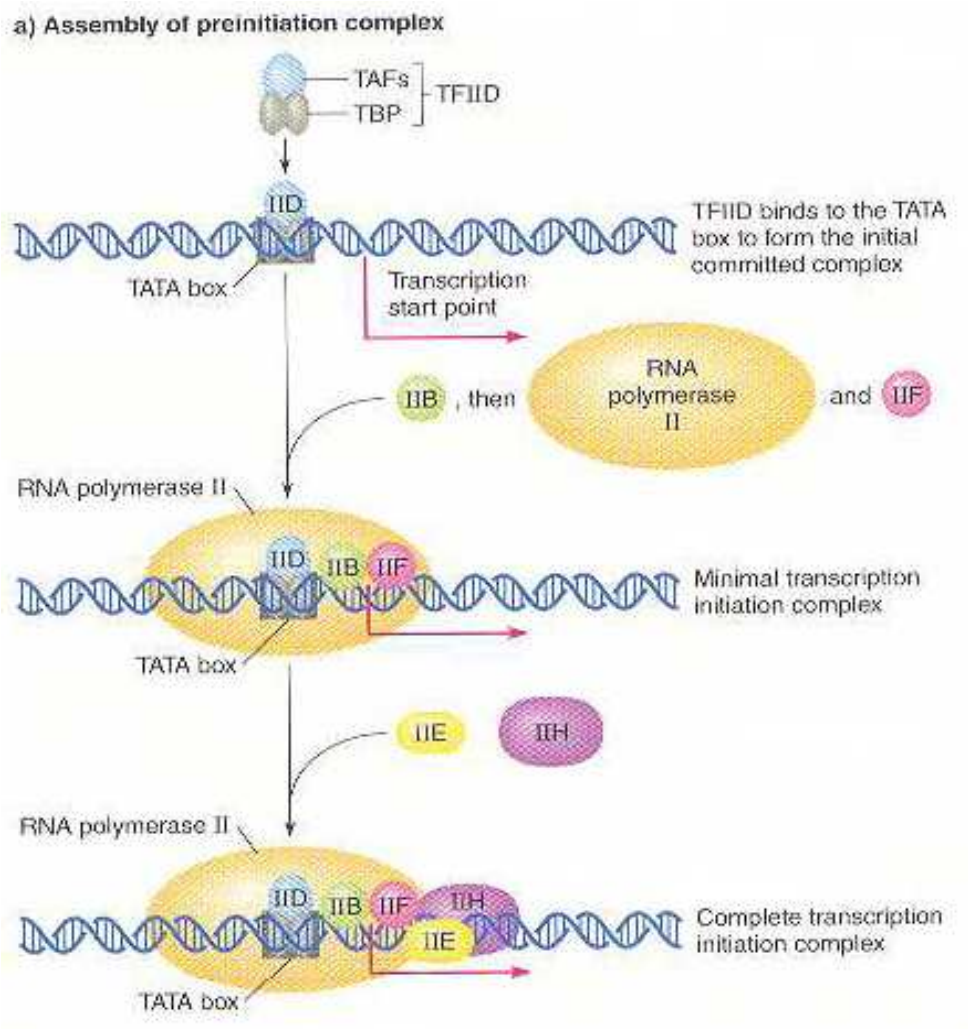


Figure 1.2. Initiation of RNA polymerase II directed transcription of DNA.

From *iGenetics* p. 118, by P.J.Russel, 2001, San Francisco: Pearson Education, Inc.

1.5 Classification of Transcription Factors (TF's)

Genes, as mentioned, can be expressed constitutively or else have a facultative expression that is, respond only to certain stimuli (Baerson & Lamppa, 1993). One aspect gene expression depends upon is the interaction of the transcription factors with the *cis*-regulatory elements, which in turn are important in the regulation of cellular activities (Guilfoyle, 1997). Transcription factors can be sub-divided or classified on the basis of their structural features, with families sometimes sub-divided according to the number and spacing of the conserved residues in the most similar domain (Guilfoyle, Ulmasov & Hagen, 1998). The DNA binding domain of the transcription factors within the promoter

regions are basic in nature, containing amino acid residues which determine the specificity of the protein (Aukerman, Schmidt, Burr & Burr, 1991). Other residues improve the binding between the two, that is, the transcription factor and the promoter by non-specific interaction with either the phosphate or the deoxyribose moieties of the promoter (Huang et al., 1996). Generally, in any plant, a transcription factor has only a single type of DNA-binding domain, occurring in either single or multiple copies (Liu, White & MacRae, 1999). Transcription factors have the tendency to function as activators or repressors. This depends upon the nature of transcription factors, whether to inhibit or stimulate the transcription of the particular gene. Transcription factors may act as repressors by competitive binding with the activators for the same domain in the *cis*-regulator region. Other mechanisms through which the transcription factors may act as repressors include the dimerization of transcription factors to prevent the regulation of the *cis*-elementary domain (Chern, Bobb & Bustos, 1996).

It has been observed that the transcription factors from the same family but from different eukaryotic organisms show quite a lot of structural and functional similarity, thereby indicating that these factors have evolved from a common ancestor. The gene distribution of the transcription factors might be redistributed through translocation after duplication and hence, the related family members may then be distributed throughout the genome or may be present within a single chromosome (Kerstetter et al., 1994). Alignment of transcription factor sequences showed that nucleotide substitution has played an important role in the evolution of the conserved sequences, whereas substitutions and insertions/deletions contributed to the variable regions (Purgganan & Wessler, 1994). In addition, exon capture through recombination may have led to the formation of totally new transcription factors (Schena & Davis, 1994). Entry of the transcription factors into the nucleus, although selective, is a prerequisite for their function. The position of these transcription factors also depends upon the environmental stimuli (Harter et al., 1994). The cloning and sequencing of transcription factor genes are important as is the examination of their expression for the functional domains, either during development or upon exposure to external stresses. X-ray crystallography and NMR spectroscopy, for example, are techniques that have been used to determine the structure of the transcription factors and their interaction with DNA (Liu et al., 1999).

1.6 Use of plant promoters

Knowledge of promoter function is growing but is still limited. The function of only a few promoters is well understood as is the minimum amount of sequence required for many promoters to be functional. Thus, there are only a few promoters available for use in transgenic plant systems (Rushton, Reinstadler, Lipka, Lippok & Somssich, 2002). The most commonly used promoter in transgenic plants is the *Caulifower mosaic virus* (CaMV) 35S promoter but this is only useful in dicot plants (Roychoudhary & Sengupta, 2009). Promoters that have been used in monocot plants include (maize) ubiquitin promoter (Wang et et al., 2004) and the (*Musa* species) actin promoter (Hermann, Harding & Dale, 2001).

Understanding promoter function is necessary from an academic point of view to understand gene function, but also from a commercial point of view in order to express transgenes in plants in a highly predictable and exquisite manner. The use of these promoters is that they can increase the expression of important traits in commercial species; that is, can be used to manipulate selected phenotypes/traits in commercial species. For example, researchers have identified several resistance genes which can be used to evade pathogens, thereby increasing the crop yield. These resistant genes are placed under the control of strong promoters which enables the constitutive expression of these genes that is, expressed in all plant tissues at all times. However, this can also be detrimental to plant growth. Finding promoters with the requisite expression pattern, directing controlled gene expression specifically at the locations of the probable pathogen invasion would be ideal. This gene expression could be responsive to either biotic or abiotic stress and would be a major scientific advance. It would also help engineer plants with an increased disease and stress resistance (Rushton et al., 2002).

Synthetic plant promoters could be useful to control or regulate gene expression in plants (Rushton et al., 2002). A synthetic plant promoter includes a stretch of DNA which comprises of a core promoter region along with multiple repeats of heterologous upstream regulatory elements. These regulatory elements might include enhancers, activators and/or repressors that bind to TFs expressed in response to specific conditions. The inability of the

conventional wild-type plant promoters to overcome these hurdles has renewed interest in the development of synthetic plant promoters. Hence, the use of synthetic promoters that allow for targeted inducibility is of considerable interest for plant engineering strategies (Venter, 2007). The challenge is if these synthetic plant promoters produced *in vitro* are able to perform their functions *in vivo* and bring about the desired expression pattern in the plant. The production of synthetic plant promoters can be used in engineered crops with an increased insect resistance or higher growth rate (Rushton et al., 2002).

1.7 Promoter evolution

It has been found that evolution has modified promoters to alter their length, depending upon the gene function. A strong relationship has been found between the length of the promoter and its responsiveness to a variety of stress conditions (Vinogradov, 2004). For example, long promoters appear to control stress induced genes. The simplest explanation of this is that long promoters allow for the integration of a larger number of regulatory inputs. Also, the more complex the function, the longer the promoter length in order to deal with the complex gene regulation. On the other hand, it is thought that short promoters may have been modified in order to fit into short genomes and bring about faster replication, thereby allowing faster proliferation (Kristiansson et al., 2009). Tissue specific genes are generally much longer than constitutive or housekeeping genes simply because of the more complex role and protein architecture of the specific facultative genes. Also, facultative genes tend to have more functional domains (Urrutia & Hurst, 2003). Evolution also plays its role in the modification of promoter length simply because the process of transcription and translation is an energetically costly process because it involves a lot of usage of ATP from plants and hence the expression of a long constitutive genes would be a rather costly affair as the gene would have to be transcribed throughout the life cycle of the plant and use up a lot of the plant's ATP (Vinogradov, 2004). This proves that promoters are an invaluable part of gene regulation and hence, are modified through evolution to suit the needs of the organism.

1.8 Responsiveness to biotic and abiotic stress

It is estimated that about 50% of crop losses occur by factors such as cold, drought, heat etc. while 10% of the losses are brought about by pathogen infections. Hence, it is essential to have an intensive study of the stress-induced changes in the plant gene expression. Many studies have been carried out on stress induction of plant gene expression (Kreps et al., 2002).

Transcriptional activation and repression of genes occur in response to several biotic and abiotic stresses. A number of genes that have been induced by any of these challenges, including those that encode for transcription factors, have been identified and have been shown to be important for stress tolerance (Chen et al., 2002). Under natural conditions, a plant may be challenged by several stresses simultaneously and hence a variety of signalling pathways which go hand in hand have evolved. These pathways can interact with each other either synergistically or antagonistically (Heil & Bostock, 2002). This indicates that a common set of transcription factors might be binding to the promoters to bring about the necessary stress response (Killian et al., 2007). During biotic stresses different sets of hormones are stimulated which results in the activation of different sets of defence-related genes (Vandenborre, Damme & Smagghe, 2009).

This shows that promoters and also the transcription factors are modified such that they suit the lifestyle of the plant in response to the different stressful conditions indicating the necessity of these elements for plant survival.

1.9 Potyviruses

1.9.1 Potyvirus genome structure

Viruses are intracellular pathogens having a very small genomic size and hence a limited protein coding capacity. As a result, several host factors are required for their replication. These host factors act either as components of the viral replication complex or directly bind to the viral genome (Browne, Li, Chong & Littman, 2005). Viruses have evolved in such a way so as to make use of almost all the resources of the host it infects. Viruses bring about

replication and encapsidation inside the host cell it infects and hence interacts with and manipulates the host's pathways so as to transform them into viral factories. This host-virus interaction includes a complex sequence of events which may be between the proteins encoded within the nucleic acids of the virus with the proteins of the host (Grainger, 2008).

The *Potyviridiae* are one of the largest and economically important families of the plant virus kingdom, most of them being transmitted by aphids (J. C. K. Ng & Falk, 2006). *Potyiviridiae* consists of about 200 species and causes serious diseases and damage to crops, both monocots and dicots (X. Wang, Ullah & Grumet, 2000). Hence, it is important to know more about how potyviruses infect plants.

The genus *Potyvirus* is one of the several genera within the *Potyviridiae*. Potyviruses are flexuous, non-enveloped and rod shaped having a length of about 680-900 nm and a diameter of 11-15nm. Their genetic material is a positive stranded RNA and is about 10 kb long and surrounded by around 2000 copies of the coat protein units. The genome consists of one long open reading frame (ORF), which is translated into a large 340-370 kDa polyprotein (Urcuqui-Inchima, Haenni & Bernardi, 2001). This polyprotein is then either co-translationally or post-translationally broken down into individual proteins by the cutting action of three viral encoded proteinases that is, the P1, HC-PRO and NIa (Plochocka, Welnicki, Zielenkiewicz & Ostoja-Zagorski, 1996). The proteins encoded by the potyviral genome are as shown in Figure 1.3.

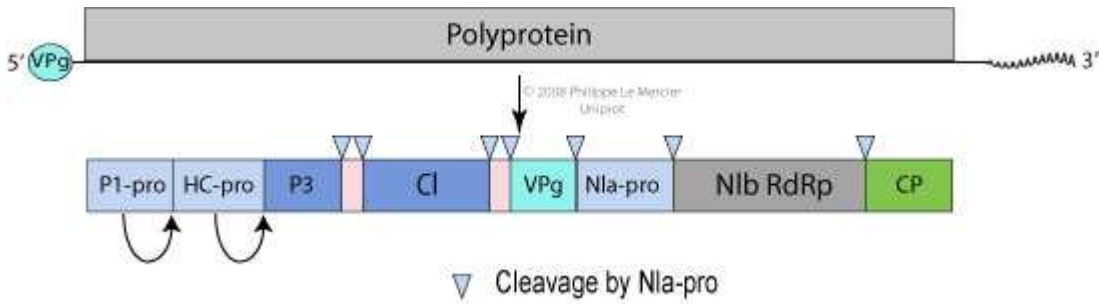


Figure 1.3.Genome structure and protein order within *Potyviridiae*.

From "Potyviridae" Retrieved Feb 13, 2010, from expasy.org/viralzone/all_by_species/48.html

The Potyviral genome is polyadenylated at the 3' end while the 5' end is bound by the viral protein VPg which is encoded by a portion of the NIa gene. The 5'end of the viral RNA is covalently attached to the VPg via a tyrosine residue which acts as a cap structure seen in eukaryotic mRNAs. The cap structures are responsible for the attachment of mRNAs to the small ribosomal subunits (Leonard et al., 2000).

Most of the viral proteins appear to be multifunctional and are often combined in homodimers or heterodimers to carry out their functions. The VPg has been shown to have several functions. The NIa protein, which is the precursor of the VPg protein has been found within the nucleus of *Tobacco etch virus* (TEV) infected plant cells (Carrington, Freed & Leinicke, 1991). This implies that the protein has certain important functions within the nucleus. However, mutations within the nuclear localisation sequence (NLS) resulted in high concentrations of the NIa protein in the cytoplasm which appeared to be disadvantageous for potyviral amplification (Leonard et al., 2000). The VPg is a protein of immense importance owing to the many roles it plays during potyviral infection. As stated above, it acts as a cap structure for translation of the viral RNA; the function of the VPg protein is to replace the 5' cap structure normally found on cellular mRNA. The position of VPg at the 5' end of the viral RNA genome instead of the mRNA cap-structure indicates that VPg plays an important role in the initiation of protein synthesis of the RNA (Leonard et al., 2000). VPg may also act as a primer for RNA replicase during virus replication, possibly through interaction with the viral RNA polymerase. VPg has been associated indirectly with the viral movement from one cell to another through the plasmodesmata and directly through the mutagenesis in the long distance translocation of the virus. VPg interacts with its adjacent protein which is NIa and has been known to show "intracellular targeting to the nucleus" (Dunoyer, Thomas, Harrison, Revers & Maule, 2004). Thus VPg is an important protein playing many roles such as a cap protein, primer for replication, viral movement, and role in the nucleus.

Roudet-Tavert et al. (2007) have elucidated the three dimensional structure of VPg. This study suggested that the amino acids 105-107 and 276-280 of VPg are involved in RNA binding within the host. The 3D structure of VPg had been predicted from the *Potato virus* Y. The prediction was based on an unrelated protein showing a similar distribution of hydrophobic/hydrophilic residues. VPg was observed to appear as an egg shaped protein,

with the domain involved in the interaction exposed at one pole. Although obtained by unconventional methods, this model is consistent with an involvement in protein interactions (Roudet-Tavert et al., 2007).

1.9.2 Potyvirus replication

Being positive stranded, the viral genome can behave as an mRNA as soon as the particles uncoat. Following the entry of the virus into the host cell, its capsid is disassembled and its genome translated, replicated and moved from one cell to another. Viral replication is a two step procedure. The first step involves the synthesis of a negative stranded replication intermediate. During this synthesis, the viral replicase, a RNA dependent RNA Polymerase (RdRp) makes a negative sense copy of the viral genome. The negative sense strand is then used by the viral RdRp as a template for the synthesis of the positive sense genome (Nagy, 2008).

1.9.3 Potyvirus translation

Potyviruses rely on host proteins to translate their genomes. In particular, the translation of potyviruses depends upon the host eukaryotic translation initiation factors namely, eIF4E and eIF(iso)4E (Robaglia & Caranta, 2006). Translation of plant mRNA is initiated by the binding of the protein complex eIF4F. In particular, eIF4E binds to the 5' end of mRNA, including potyviral RNA genomes (Gingras, Raught & Sonenberg, 1999). Another form of eIF4F is also present which is known as eIF(iso) 4F which contains eIF(iso)4E. eIF(iso)4F differs from eIF4F in being less efficient in promoting the internal initiation, cap dependent translation and translation of the structural RNA.

eIF4E in particular is important because in many plant- potyviral complexes, physical interaction between eIF4E and the VPg protein of potyviruses has been demonstrated (Nicaise, 2007). The eIF4E isoforms bind to the virus with different affinities. This interaction is however directly proportional that is, the greater the binding, the higher the rate of viral infection (Nicaise, 2007). The VPg of *Turnip mosaic virus* (TuMV) was

observed to interact with eIF(iso)4E of *Arabidopsis thaliana* (Schaad, Anderberg & Carrington, 2000). This interaction suggested that the VPg of the virus may be responsible for the initiation of translation of the viral RNA. Also, in TuMV any changes within the central domain of VPg can impair its interaction with eIF4E, thereby suggesting a direct interaction between the two. Also, VPg plays a role in overcoming resistance because the eIF factors have been shown to provide recessive resistance which can be overcome by variants of VPg. Some potyviruses require eIF(iso)4E for infection while some require eIF4E. Since both the isoforms are needed by the potyvirus, it is concluded that a common factor between the isoforms may be involved (Duprat, Caranta, Revers, Menand, Browning & Robaglia, 2002).

1.10 PVIP Protein

Certain host proteins have been shown to interact with potyviruses during the process of infection. In particular a protein identified in *Pisum sativum, Arabidopsis thaliana* and *Nicotiana benthamiana* called potyvirus interacting protein (PVIP) was found to bind to the VPg of the potyvirus TuMV. PVIP appears to support movement of potyvirus in plants (Dunoyer et al., 2004).

PVIP was initially identified in *P.sativum* (PVIPp) and found to be homologous to a small gene family in *Arabidopsis thaliana* (Dunoyer et al, 2004). Using the yeast two hybrid system it was shown that VPg from TuMV binds with the PVIP of *A.thaliana* and not just with PVIPp and that this binding depended upon the first 66 amino acids of the VPg. The sequence comparison between TuMV (turnip mosaic virus) and three other potyviruses, namely *Lettuce mosaic virus* (LMV), *Tobacco etch virus* (TEV) and *Pea seed-borne mosaic virus* (PSbMV) showed that the amino acid sequence likely to be involved in binding was determined within two regions of VPg that is, amino acids 1 to 16 and 42 to 66. Site directed mutagenesis determined that position 12 was important for binding between VPg and PVIP. The three dimensional structure of VPg showed that the N-terminal 16 amino acids is on the surface of VPg which is folded. On this folded domain, just the position 12 would be available for binding with the PVIPs (Dunoyer et al., 2004).

In contrast to the other viruses, the VPg of TEV did not bind to PVIP at position 12. The TEV VPg was found to have an aromatic amino acid at position 12 instead of methionine. Dunoyer et al. (2004) proposed that TEV binds with the PVIP protein using an alternative mode of interaction. The roles of the VPg-PVIP interaction in viral replication and viral movement were investigated and it was observed that the interaction is crucial but may not be imperative for potyviral multiplication considering the exception of TEV. This interaction also appeared to be essential for movement of the virus from one cell to another (Dunoyer et al., 2004).

Analysis of the PVIP amino acid sequence showed a "PHD finger like" domain, which was identified by a series of histidine and cysteine residues. These residues are found to frequently associate with proteins involved in protein synthesis regulation through chromosome remodelling, indicating that VPg might have a role to play in the nucleus. VPg has a nuclear targeting sequence and accumulates in the nucleus in the form of the precursor molecule which is NIa. VPg as part of the NIa may therefore play a role in regulating host protein synthesis during infection (Dunoyer et al., 2004).

1.11 Conclusion

There have been few studies of PVIP and its gene function. We know that it is present in a cell prior to viral infection and that it interacts with potyviruses via binding to the viral VPg protein. However, what is its normal role in an uninfected cell? Is the protein function the same in all plant species? How does it respond to environmental cues? Why would a potyvirus highjack this particular host protein for its own purposes? This study has attempted to understand the function of the PVIP gene in healthy *Nicotiana benthamiana* cells by analysing its promoter structure, and assessing how its expression is altered by environmental stimuli. Understanding PVIP's normal function may provide insights into why it interacts with the VPg and how it might be utilised to prevent viral infection.

Chapter 2

Bioinformatic analysis of PVIP 5' flanking region

2.1 Introduction

The sea of data generated by science and technology is increasing at an incredible rate and hence there arises an increasing demand for tools and methods for data management, visualisation, prediction etc. Hence, bioinformatics plays a very important role in today's biological science (Rhee, Dickerson & Xu, 2006). Bioinformatics deals with the study of biological information using computer science, statistics and engineering. It can be divided into two categories that is, biological information management and computational biology. Biological information management deals with the research or application of computer technology to handle and integrate increasing biological data. This can even include the representation of data which may be visual or analytical. Computational biology involves the development and application of data-analysis and theoretical methods including mathematical modelling and computer simulation techniques to study the biological, social and behavioural systems (BISTI, 2000). Sequence analysis is one of the most commonly used features of bioinformatics which is used to compare the sequence from one organism to another (The_Arabidopsis_Genome_Initiative, 2000). Sequence analysis may involve techniques such as genome sequencing, genome annotation and sequence comparison. The sequences of DNA, RNA and proteins have been sequenced in model plants such as *Arabidopsis thaliana* and *Oryza sativa* (rice). Also, sequencing efforts are underway for other plants of commercial importance such as tomato, maize, sorghum etc. Genome sequences of model plants for which there is much genetic information will help us determine functional importance of sequences of other plants for which there is little genetic information. These sequences include regions that function as promoters, protein encoding regions, or the transcription factors (Rhee et al., 2006).

2.1.1 Genome sequencing

Developments in sequencing technologies have brought about an improvement in the management, processing and analysis of genomes. Until recently, the shotgun sequencing technique was the most commonly used method of genome analysis. It involves the random shearing of DNA into small pieces followed by cloning of the generated fragments and sequencing each one randomly (Myers, 1995). Different software such as Arachne, Phred/Phrap/Consed and GAP4, for example, have been developed which are used to piece together the random overlapping segments into a coherent and accurate contiguous sequence (Pop, Phillippy, Delcher & Salzberg, 2004). However, limitations of this procedure include the high cost involved as well as the assembly of highly repetitive sequences (Hinds et al., 2005). Alternative methods of sequencing such as differential hybridisation of oligonucleotide probes (Hinds et al., 2005) and polymorphism ratio sequencing (Blazej, Paegel & Mathies, 2003) have been developed which are cheaper as compared with shotgun sequencing. More recently, the development of massively parallel sequencing technologies has allowed the generation of whole genome sequences for significantly reduced cost. Sequencing repeat regions is still problematic as is the handling of the enormous amount of data generated (Hinds et al., 2005).

2.1.2 Genome annotation

Sequence information is only useful if the functional regions are interpreted and that interpretation made available as annotations to the sequence. Such annotations tell us where the functional regions are. For example, it tells us where a gene is positioned within a length of nucleotide sequence and where within that region the promoter, coding region and introns and exons might be found. Without such annotations, the sequence is merely a collection of letters that are of little use.

Several software packages have been developed which identify protein coding genes from non-coding regions (Zhang, 2002). Popular packages include GeneScan, GRAIL, Genie and GeneMarkHMM. However, this *ab initio* gene prediction may pose a challenge for large sized eukaryotic genomes (Brendel & Zhu, 2002). Genome annotation techniques are widely used during the analysis of transposons and genes in a genomic sequence using the

above algorithms. The current limitation of this technique however may be the inaccurate prediction of the transcription start sites (TSS), alternative splicing sites, non-coding genes and also the small genes encoding for less than 100 amino acids (Rhee et al., 2006).

2.1.3 Sequence comparison

Sequence comparison provides the framework for the analysis of functions, structure and also the evolution of genes and genomes. Therefore, many types of software have been developed for putative homology identification. Sequence comparison is a very important part of bioinformatics. A nucleotide or protein sequence can be compared with other sequences held within a database so that homologues can be identified. Regions of homology may provide clues about the function of a query sequence if the function of an identified homologue is known (Wan & Xu, 2005). The methods in sequence comparison can be further divided into pair-wise alignment, sequence-profile alignment and profile-profile comparison (Rhee et al., 2006).

The suite of algorithms known as the Basic Local Alignment Search Tool (BLAST) (Altschul, W. Gish, W. Miller, Myers & Lipman, 1990) is commonly used to identify regions of local similarity between a query sequence and sequences held in a database. The statistical significance of the matches is also calculated. A sequence-profile alignment is more sensitive to determine the homologous sequences. The protein sequence profile can be generated using a multiple sequence alignment of a group of closely related proteins. Multiple sequence alignments allow several homologous sequences to be aligned together so that the structural relationships may become clearer. Profile-profile alignment is more sensitive than the sequence-profile alignment (Yona & Levitt, 2002).

Bioinformatics helps in the sequence alignment of both, nucleotide and protein sequences. Protein sequence alignment is the task of identifying evolutionarily or structurally related positions in a collection of amino acid sequences whereas the nucleotide sequence alignment involves the alignment of nucleotides alone and hence helps us in the determination of any insertion and deletions within the genome that is, interpret the point mutations (Do & Katoh, 2008). In sequence alignments of proteins, the degree of similarity between the amino acids occupying a particular position in the sequence can be interpreted

as a rough measure of how conserved a particular region or sequence motif is among lineages. Although DNA and RNA nucleotide bases are more similar to each other than are amino acids, the conservation of base pairs can indicate a similar functional or structural role (P. C. Ng & Henikoff, 2001).

The classification of a protein generally depends upon its structure, function and sequence. Accordingly, several sequence-based methods have been developed to classify proteins on the basis of its similarity. The evolutionary relationships within and between protein families can be calculated and displayed as a phylogenetic tree. This type of an analysis can be used in comparative genomics and gene function prediction (Doolittle, 1999) and starts with the alignment of related proteins using multiple sequence alignment tools such as ClustalW (Larkin et al., 2007). The relationships determined by the alignment can be displayed as a tree which may be a minimum distance, maximum parsimony or maximum likelihood tree (Densmore, 2001). Several methods have been developed for calculating phylogenetic trees some of which are computer intense. Therefore, different methods can produce significantly different phylogenetic trees making manual assessment essential (Rhee et al., 2006).

Promoter analysis is an important part of bioinformatics because it helps to identify possible regulatory networks controlling a gene's function. The prediction of transcription factor binding sites allows prediction of a gene's function. Different software packages have been developed for this analysis such as TRANSFAC (http://www.biobase-international.com/pages/index.php?id=transfac) and MatInspector (http://www.genomatix.de/online_help/help_matinspector/matinspector_help.html) to name a few. MatInspector identifies the transcription factor binding sites (TFBS) within a promoter sequence using a large library of weight matrices. The weight matrix system is more advantageous than the simple IUPAC consensus sequence because it represents the complete nucleotide distribution for each single position. MatInspector uses a library containing 634 matrices and represents the largest library available to the public (Cartharius et al., 2005). The TRANSFAC database is used by software such as SignalScan and Matrix Search which is based on TRANSFAC version 2.5 while the software Transcription Element Search Site (TESS) uses a comparatively modern version of TRANSFAC i.e.

version 4. Although TRANSFAC version 8.4 has around 741 matrices it is not freely available to academic users (Matys et al., 2003).

At present the function of PVIP in *N.benthamiana* is unknown. Bioinformatic analysis of this gene and its protein product may provide clues. Using bioinformatics we can identify genes that have sequence similarity that may indicate the function of this gene as well as its evolutionary history. Further, analysis of the flanking regions from homologous genes from plants for which the genomes have been sequenced may provide indications as to how the *N.benthamiana* gene may be regulated. This chapter describes bioinformatic analysis of the *N.benthamiana* PVIP gene and its protein product in an attempt to determine its function.

2.2 Methods and Materials

2.2.1 Identification of PVIP homologues

Dunoyer et al (2004) described the PVIP mRNA sequence from *Nicotiana benthamiana* (accession number: AY271742), pea (accession number: AY271743) and *Arabidopsis thaliana* (accession numbers: At5g48160, At3g07780, At1g14740, At3g63500*)*. The *Nicotiana benthamiana* sequence (PVIP_Nb) was used as the query sequence in a BLASTn search for homologous sequences in the nucleotide collection database of the National Centre for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov).

The predicted linked protein sequence i.e. the PVIP protein sequence in *N. benthamiana* (accession number: AAP22954) was similarly used as the query sequence in a BLASTp search of the non-redundant protein database of NCBI using the default parameters.

The degree of identity and similarity between the *N.benthamiana* PVIP sequence (PVIP_Nb) and homologues from *A.thaliana* and *Oryza sativa japonica* (rice) was calculated using pairwise alignment software EMBOSS from the European Bioinformatics Institute (EMBL-EBI; http://www.ebi.ac.uk/Tools/emboss/align/index.html). Percent similarity and identity were determined between the nucleotide coding sequence of PVIP_Nb and the *A.thaliana* homologues OBE1 and OBE2. This was also done between PVIP_Nb and the closest *O.sativa* homologues found on chromosome 11 and 12. Similar analysis was done for the protein sequence.

2.2.2 Analysis of PVIP protein homologues

The relationships between PVIP_Nb and all homologues identified by BLAST were analysed by multiple sequence alignment and phylogeny. Multiple sequence alignment of PVIP protein sequences was carried out using ClustalW (Jeanmougin, Thompson, Gouy, Higgins & Gibson, 1998) within the software package GeneiousPro (Drummond et al., 2009). The alignment was optimized manually and the distances between the different sequences calculated using Neighbour Joining. Phylogenetic trees were calculated from the

distance measurements and tested with 100 bootstraps. *Malus X Domestica* (accession number: CAJ44362) was used as the outgroup.

2.2.3 Prediction of transcription factor binding sites within the PVIP 5' flanking region

2.2.3.1 Retrieval of the 5' flanking regions from *Arabidopsis thaliana* (PVIP_At) and *Oryza sativa Japonica* (PVIP_Os)

Analysis of potential or proven transcription factor binding sites (TFBS) within the 5' flanking regions of PVIP homologues from organisms for which the genomes have been sequenced may provide clues as to the potential TFBS within the *N.benthamiana* promoter region. The sequences of the 5' flanking region for the two homologues from both *A. thaliana* and *O.sativa Japonica* were therefore retrieved from genome databases and analysed for the common TFBS. *A.thaliana* and *O.sativa* were used as the homologue sequences because their genome had been completely sequenced at that time and the tomato genome sequence had not yet been released. Also, comparison of monocots (*O.sativa*) and dicots (*A.thaliana*) would provide some evidence of conservation.

BLAST analysis of the NCBI database identified the closest homologue to PVIP_Nb from *A. thaliana* (PVIP_At) as being OBE1 (At3g07780) on chromosome 3 and OBE2 (AT5g48160) on chromosome 5. The chromosomal co-ordinates for the OBE1 gene, as well as its direction of transcription, were noted as were the co-ordinates and the relative position and direction of transcription of the upstream gene. The region between the 5' end of the OBE1 mRNA and the 3' end of the mRNA of the preceding gene was downloaded and used as the 5' flanking region of the OBE1 gene in subsequent analyses. The same process was carried out to extract the 5' flanking sequence for the OBE2 gene.

Blast analysis of the NCBI database identified the closest homologue to PVIP_Nb in rice (PVIP_Os) to be on chromosomes 11 and 12. Extraction of the 5' flanking region of the chromosome 12 gene was carried as described for OBE1 gene from *A.thaliana*.

Extraction of the 5' flanking region of the chromosome 11 gene proved more problematic as it could not be downloaded from the NCBI database and an alternative strategy had to be

used. The PVIP_Os protein sequence from chromosome 11 (ABA94920) was used as a query sequence in a BLASTp search against the non-redundant protein database in NCBI using the default parameters. The top hit, EAZ19016 encoding the hypothetical protein Osj_34548 provided a link to its TREMBL entry, Q2ROP5, which in turn led to the sequence entry in the Gramene database (Gramene, 2009), LOC_Os11g41880. The link 'Best Hits in Gene Models' provided the coordinates of the gene which in turn gave the coordinates and orientation of the upstream gene i.e. LOC_Os11g41890. Using the same strategy as for OBE1, the intervening region between the genes was downloaded.

2.2.3.2 Identification of the probable TATA boxes within 5' flanking regions of PVIP_At and PVIP_Os

The consensus sequence for the TATA box is TATAAA and is estimated to be approximately 25 base pairs away from the transcription start site (Shi & Zhou, 2006). The predicted TATA box within each of the four 5' flanking regions from PVIP_At and PVIP_Os were identified by visual inspection of each sequence.

2.2.3.3 Determining the transcription factor binding sites within the 5' flanking regions

The 5' flanking sequences of the PVIP homologues from *A.thaliana* and *O.sativa* were analyzed for potential TFBS using the Transcription Element Search System (TESS) (Schug, 2003) using 'plants' as the organism classification and selecting the default parameters. The predicted TFBS for both the strands were identified and their relative positions and orientations were determined. The TFBS that were common to the sequences from both *A.thaliana* and *O.sativa* were identified and their functions were further analysed and studied.

2.2.3.4 Prediction of intron and exon regions of the PVIP_Nb

Predicting the structure of the PVIP gene within the homologues may yield some information regarding the gene structure within *N.benthamiana*. Hence, the exon/intron structure of the PVIP_Nb gene was predicted by comparing the exon/intron structure of the *A.thaliana* and *O.sativa* with the PVIP_Nb mRNA. The exon and intron regions of *A.thaliana* and *O.sativa* PVIP homologues were identified using the Gramene database

(Gramene, 2009). Gene ID numbers were used to download each sequence in turn. The transcript for each gene was downloaded by selecting the 'Transcript' tab and the intron/exon structure downloaded by selecting the 'exon' link. The gene structure obtained for *A.thaliana* and *O.sativa* were then compared manually.

2.2.4 Conserved Domain Analysis

Conserved domains are sequence regions that show higher degrees of conservation between sequence homologues. The fact that these domains have been conserved between species suggests its functional or structural importance. Therefore, it was hoped that determining the conserved domains of the PVIP homologues would help identify a possible function and/or expression (temporal/spatial) of the PVIP protein in *N.benthamiana*.

The GenBank entries for each PVIP protein homologues indicated the presence of a conserved domain. A common conserved domain was observed for each of the homologues i.e. PVIP_Nb, OBE1, OBE2, PVIP_chr11 Os and PVIP_chr12 Os. The region being the same in each of the PVIP homologues indicated the importance of the conserved domain. Hence, by determining the function of the conserved domain, the probable function of PVIP_Nb protein could be evaluated.

2.3 Results

2.3.1 Sequence homology search

Comparison of the PVIP sequence with sequences in the public domain was expected to provide some clues as to the protein's function in a healthy uninfected cell as well as how the gene may be regulated. It was hoped that this analysis might provide clues about the role of this protein in potyviral infection.

BLASTn and BLASTp analyses were carried out using the PVIP_Nb gene and protein as the query sequences. A number of homologues were identified and used for further analysis (refer to the appendices). Those sequences that had an expectation (E) value less than or equal to E^{-43} were used for further analysis. The E value greater than E^{-43} indicated that there was a reduced homology with the query sequence. The next value after this was greater than zero indicating no homology.

BLAST searches identified homologous sequences from plants only, suggesting that PVIP is a plant specific protein. Homologues were identified in a range of angiosperms such as the dicot species *A. thaliana*, *Solanum lycopersicum* (tomato), *Arachis hypogea* (peanut) and *Vitis vinifera* (grape) as well as monocot species such as *O. sativa* (rice) and *Zea mays* (maize) (Appendix Table 1). Homologues were also identified in the gymnosperm *Picea sitchensis* (Sitka spruce) as well as the moss *Physcomitrella patens.*

Functions of the homologous sequences included a zinc iron binding protein which has been found to be tolerant to oxidative damage and pathogens (*A.thaliana*, Accession numbers: NM_124190 and NM_111657) (Deák et al., 1999), a plant homeo domain (PHD) family protein (*A.thaliana*, accession number: DQ059085) which is involved in transcriptional regulation of a pathogen defence-related gene (Korfhage, Trezzini, Meier, Hahlbrock & Somssich, 1994) and a Constans interacting protein (*Lycopersicon esculentum*, accession number: AY490248) which has been found to promote flowering (Robson et al., 2001) and encodes a protein with similarities to zinc finger transcription factors. These functions may indicate possible roles for the PVIP protein in *N. benthamiana* as well.

The OBERON 1 and 2 (OBE1 and 2) proteins from *A.thaliana* (accession numbers: NP_566320 and NP_199627) were also found to be homologous. The genes have been found to control the transcription of genes required for auxin responses through the action of their PHD finger domains (Thomas, Schmidt, Bayer, Dreos & Maule, 2009). Auxin plays an important role in plant development. Being similar to the OBE proteins from *A.thaliana*, the PVIP protein from *N.benthamiana* might also be involved in auxin response i.e. might be responsible for plant growth and development.

As expected, protein analysis using BLASTp was more informative than nucleotide analysis using BLASTn. Analysing the nucleotide sequence alignments showed that the majority of nucleotide changes occurred in the third codon position. This was supported by the fact that greater homology was found between protein sequences than between nucleotide sequences, suggesting a greater evolutionary pressure to maintain the protein sequence compared to the nucleotide sequence.

2.3.2 Phylogenetic analysis

The phylogenetic analysis of the PVIP homologues is shown in Figure 2.1 and Figure 2.2. The Neighbour Joining analysis was rooted using the PVIP homologue from *Malus x Domestica* as the out-group. This sequence was the least homologous to the *N.benthamiana* sequence that still had a negative E value, i.e. E^{-43}. The majority of branches within the tree were statistically significant since the bootstrap values were above 50. Thus, the branching pattern within the tree was reliable.

One significant clade contained the PVIP_Nb sequence (accession number: AAP22954) that was used as the query in the BLASTp search. This clade was separated from the others with a boostrap value of 85, indicating strong support for this grouping. This clade is likely to contain sequences that are functionally homologous to PVIP. Sequences within the other clades are related to PVIP and may have functional overlap with this sequence with respect to functioning in a healthy cell, but they may be less likely to interact with the potyvirus VPg protein and therefore less likely to be involved in the virus infection process. However, this would need to be tested experimentally.

The clade containing the predicted functional homologues is broken up into sub-clades specific to the plant group from which the sequences were derived. All angiosperm

sequences are together and are separated according to monocot or dicot sources of the sequence. This grouping is separate from the other plant groups, gymnosperm, moss and fern (Figure 2.1). This suggests that PVIP is an ancient gene since it is present in the ancient lineages of moss and fern and that it has evolved as plants have evolved. This would need to be tested further with many more sequences as they become available.

As expected, PVIP_Nb is part of the dicot clade and is closest to the *Solanum lycopersicum* (tomato) sequence. This is not surprising as both tomato and *N.benthamiana* are both from the *Solanaceae* family of plants. These are the only two plants from this family included in this tree as no other sequences were available from the NCBI database at the time of analysis. Inclusion of sequences from the *Solanaceae* sequence databases would provide further support for this observation.

It is likely that there are two genes encoding PVIP in gymnosperms and angiosperms. Two distinct sequences were identified in rice and *A.thaliana*, where the complete genomes have been sequenced. The nucleotide sequences of the two genes from each species are more similar to each other suggesting a common ancestor and hence gene duplication. There are either two distinct sequences (as from *Populus trichocarpa*) or sequences that form two distinct subclades (*A.thaliana, O.sativa* and *Z.mays*). For some species, there is only one sequence, likely due to there being insufficient sequence information available. In some instances, for example *A.thaliana*, there are multiple sequences available that may represent allelic variation within one of the genes. Based on the dataset, it is impossible to say if a gene duplication event occurred when gymnosperms evolved or if this occurred before hand. More sequence information from moss and fern is required before this can be stated with any certainty. Figure 2.2 for example shows that there are two copies of PVIP present within *A.thaliana* and *O.sativa* (chr12 and chr11) and that there are likely to be additional different alleles of the PVIP gene within these species. Since there are two genes in both monocots (*O.sativa*) and dicot (*A.thaliana*) there may be a possibility of the gene duplication prior to angiosperm evolution.

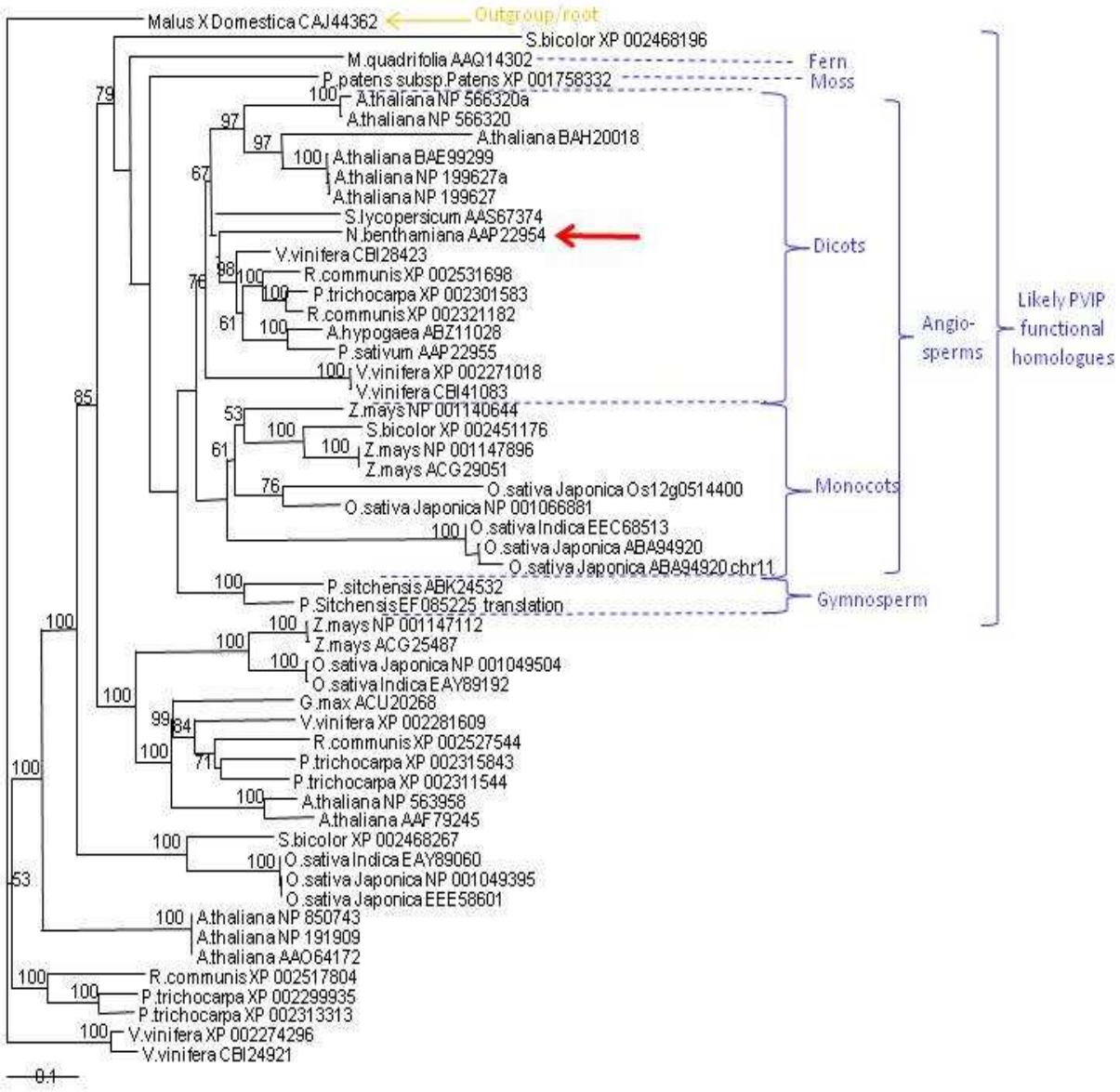


Figure 2.1. A Neighbour Joining tree of the PVIP protein homologues. The PVIP_Nb sequences is indicated with an arrow. The sequence form *Malus x Domestica* was used as the outgroup and is indicated in yellow. Likely PVIP functional homologues are indicated. PVIP homologues from each plant division are also indicated.

Also, as PVIP homologues have been identified in angiosperms, gymnosperms, moss and fern we could conclude that the PVIP sequence occurred very early in plant evolution and may have existed within plant ancestors as well. We could thereby conclude that the PVIP gene, being present since ancient times and having been maintained as plants evolved, has a very important function.

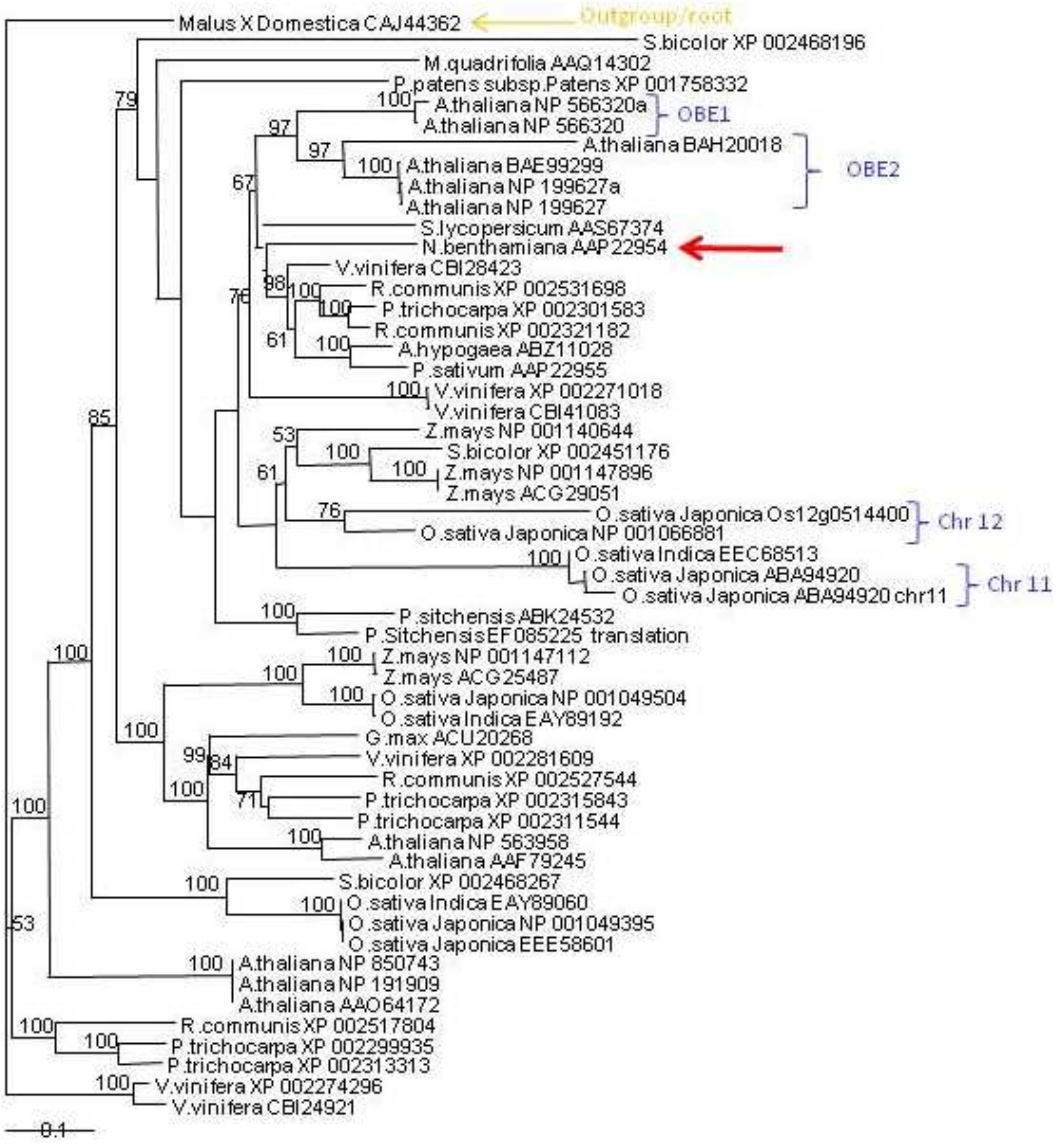


Figure 2.2. Neighbour Joining tree of the PVIP protein homologues with respect to *A.thaliana* and *O.sativa*. The PVIP_Nb sequences is indicated with an arrow. The sequence form *Malus x Domestica* was used as the outgroup and is indicated in yellow. The two alles of PVIP are shown in *A.thaliana* and *O.sativa* being marked in purple.

Comparison with sequences from the model species *A.thaliana* and *O.sativa* may provide further information about the function of PVIP_Nb. The genomes for both these species have been sequenced and there are many development/mutant studies available, particularly of *A.thaliana*, which may provide functional information. The phylogenetic tree (Figure 2.2) showed that the PVIP_Nb is more closely related to PVIP from *A.thaliana* than from *O.sativa*. Again, this is not surprising as *A.thaliana* is a dicot and rice is a monocot. It is

likely that in terms of gene structure, the PVIP_Nb gene may be more similar to that from *A. thaliana* than from *O.sativa*.

BLAST analysis of PVIP_Nb showed that the most homologous sequences from *A.thaliana* (PVIP_At) were OBERON1 (OBE1) and OBERON 2 (OBE2) and *O.sativa* sequences (PVIP_Os) from chromosomes 11 and 12. Percentage sequence similarity/identity values are given in Tables 2.1 and 2.2. The highest nucleotide sequence homology between PVIP_Nb and PVIP_At was with OBE1 (40.1 %,) while between PVIP_Nb and PVIP_Os it was with chromosome 12 (42.4 %). In contrast the highest protein homology was with PVIP_At_OBE2 (62.1 %,) and PVIP_Os_chr11 (56.4 %,).

Also, at the protein level, the *A.thaliana* proteins OBE1 and OBE2 were more similar to each other (86.3%) than they were to PVIP from rice encoded on either chromosome 11 (68.7%) or 12 (50.3%). In contrast, the rice sequences were not as similar to each as they were to the *A.thaliana* sequences. PVIP_Os from chromosome 11 was 44.5% similar to the chromosome 12 sequence, whereas it was 68.7% similar to OBE1 and 44.5% similar to OBE2 from *A.thaliana*. PVIP_Os from chromosome 12 was 50.3% similar to OBE1 and 69.4% similar to OBE2. This suggests that OBE1 and PVIP_Os from chromsome 11 are homologues and OBE2 and PVIP_Os from chromsome 12 may be homologues.

This suggests that there are likely to be two PVIP genes present within *N. benthamiana* as in *A. thaliana* and *O.sativa*, one showing homology with the PVIP_At_OBE1 and PVIP_Os_chr11 and the other having a greater similarity with PVIP_At_OBE2 and PVIP_Os_chr12 However, the available PVIP_Nb sequence shows greatest homology to OBE2 and the rice chromosome 11 sequence. Until the second PVIP gene is isolated and sequenced, the exact relationship between these sequences is unclear. However, it is expected that one gene will show higher identity to OBE1 and the other to OBE2 from *A.thaliana*.

Table 2.1. Percent identity between nucleotide sequences of PVIP_Nb and its homologues from *A. thaliana and O.sativa Japonica*.

	PVIP_Nb	PVIP_At_OBE1	PVIP_At_OBE2	PVIP_Os_chr11	PVIP_Os_chr12
PVIP_Nb		**40.1**	**36.7**	**34.2**	**42.4**
PVIP_At_OBE1			**45.5**	**43.5**	**43.9**
PVIP_At_OBE2				**44.1**	**44.7**
PVIP_Os_chr11					**47.9**
PVIP_Os_chr12					

Table 2.2. Percent similarity and identity between protein sequences of PVIP_Nb and its homologues from *A. thaliana* and *O.sativa Japonica*. The values in bold are the percent identity and the values in italics are the percent similarity.

	PVIP_Nb	PVIP_At_OBE1	PVIP_At_OBE2	PVIP_Os_chr11	PVIP_Os_chr12
PVIP_Nb	-	**57.7**	**62.1**	**56.4**	**32.5**
PVIP_At_OBE1	*73.3*	**-**	**76.2**	**53.7**	**31.4**
PVIP_At_OBE2	*76.0*	*86.3*	-	**33.0**	**53.9**
PVIP_Os_chr11	*70.1*	*68.7*	*48.8*	-	**32.8**
PVIP_Os_chr12	*51.1*	*50.3*	*69.4*	*44.5*	-

2.3.3 Identification of 5' flanking regions for PVIP_At and PVIP_Os genes

OBE1 was found on *A.thaliana* chromosome 3 between nucleotides 2484469 and 2487024 (Figure 2.3). It was found to be transcribed from the top strand with the preceding open reading frame (ORF) AT3G07770 lying between nucleotides 2479548 and 2484194. Hence, the 5' flanking region of OBE1 must lie between the 3'end of the previous gene AT3G07770 (position 2484194) and the 5' end of OBE1 ORF (position 2484469). This region is likely to contain the promoter region for OBE1 and is 275 bp long which is relatively small for a promoter. This sequence from nucleotide 2484194 to nucleotide 2484469 was retrieved from the *A.thaliana* chromosome 3 sequence in GenBank.

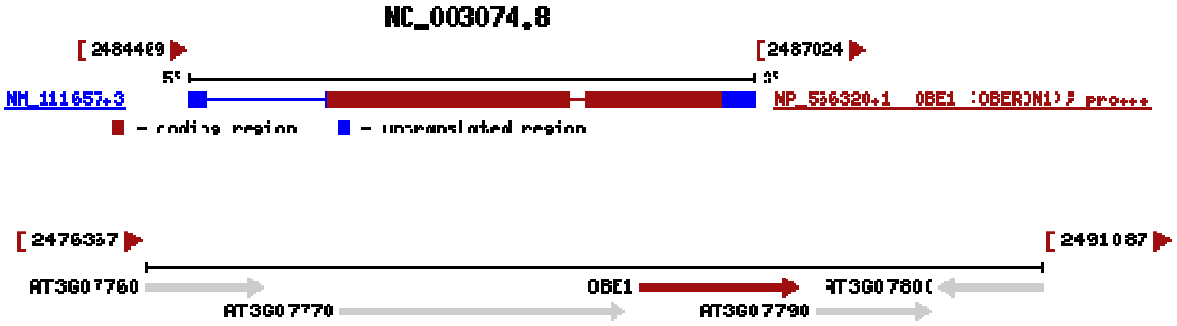


Figure 2.3. The structure of OBE1 mRNA and the position of the OBE1 gene on chromosome 3 of *A.thaliana.* The direction of transcription of OBE1 and surrounding genes is shown. (Source: NCBI).

The coding region for the PVIP homologue OBE2 present on chromosome 5 of *A.thaliana* was found to be transcribed from the bottom strand between the co-ordinates 19530316 at its 5' end and 19527833 at its 3' end (Figure 2.4). The ORF preceding OBE2 was found to be SLY2 having its co-ordinates between 19533265 at its 5' end and 19532537 at the 3' end. Hence, the promoter for OBE2 is likely to be between the 3' end of the ORF of SLY2 and the 5' end of the OBE2 ORF i.e. between nucleotides 19532537 and 19530316. This 5' flanking region for OBE2 was 2221 bp long and retrieved from the *A.thaliana* chromosome 5 sequence in GenBank.

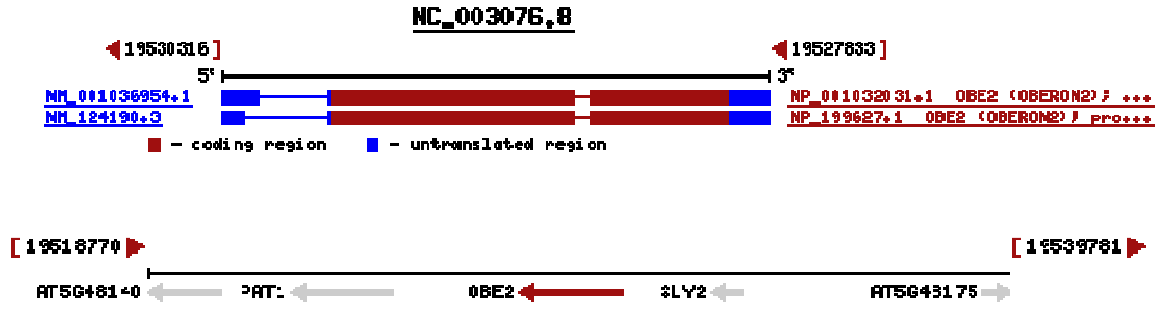


Figure 2.4.The structure of OBE2 mRNA and the position of the OBE2 gene on chromosome 5 of *A.thaliana* with respect to its co-ordinates, orientation and the adjacent genes. (Source: NCBI)

The closest rice homologues of PVIP (PVIP_Os) were found on chromosomes 11 and 12. The 5' flanking region for the homologue on chromosome 12 (Os12g0514400) was

retrieved in the same way as described in 2.2.3.1. The coding region for Os12g0514400 was found to be transcribed from the bottom strand between nucleotides 19958403 at the 3' end and 19955475 at its 5' end (Figure 2.5). The preceding ORF was found to be Os12g0514500 having its co-ordinates between 19968508 at the 5' end and 19962464 at its 3' end. The 5' flanking region for the PVIP homologue was in between the two ORFs i.e. between the 5' regions of the rice PVIP homologue Os12g0514400 (position 19958403) and 3' end of the Os12g0514500 ORF (position 19962464). This region was found to be around a 4061 bp long and was retrieved from the *O.sativa* chromosome 12 sequence in GenBank.

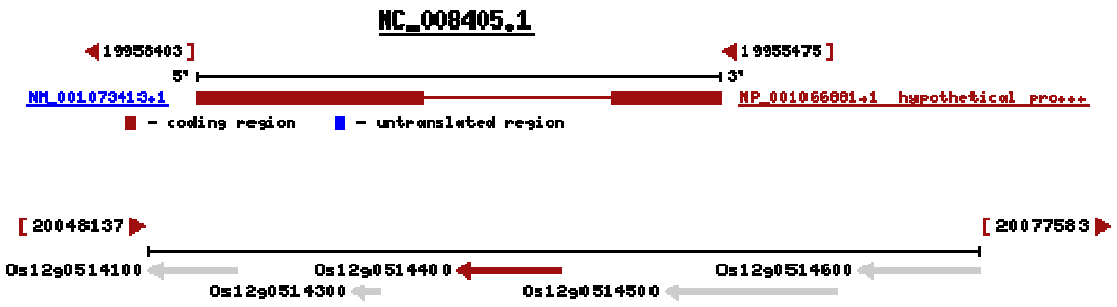


Figure 2.5.The relative position of the Os12g0514400 gene on chromosome 12 of *O.sativa* with respect to its co-ordinates, orientation and the adjacent genes. (Source: NCBI)

Retrieval of the 5' flanking region from chromosome 11 was more challenging. No direct links were available in the NCBI database to this region of the *O.sativa* genome, thus a different strategy had to be used. PVIP_Os from chromosome 11 (ABA94920) was used as a query sequence in a BLASTp search against the non-redundant protein database in NCBI using the default parameters. The top hit was found to be EAZ19016, the hypothetical protein Osj_34548. The protein had a link for 'Trembl Q2ROP5' which led to the Gramene database (Gramene, 2009) index page with LOC_Os11g41880 as the gene of interest. Clicking on 'Best Hits in Gene Models' the location and orientation of the gene was determined. The coding region was found to be transcribed from the bottom strand with 24,717,316 as its 5' end and 24,715,139 as its 3' end which gives a 5' flanking region of 2,177 bp long. The gene following LOC_Os11g41880 was LOC_Os11g41890 which was transcribed from the forward strand (Figure 2.6). The co-ordinates for this gene were

24,726,853 as its 5'end and 24,731,188 as its 3' end. Hence, the 5' flanking region of PVIP_Os lies between 24,717,316 bp to 24,726,853 bp which is 9,537 bp long.

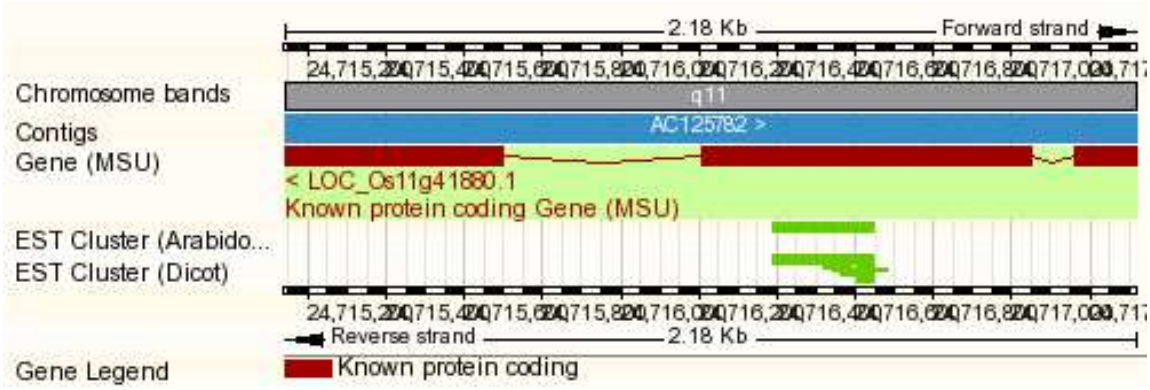


Figure 2.6.The 2.18 kb region of rice chromosome 11 containing the PVIP_Os gene (LOC_Os11g41880) from the Gramene database. The direction of transcription is indicated by the arrow at the top of the diagram, with the nucleotide coordinates below. Sequencing contigs and ESTS are indicated in grey and green while the exon/intorn structure is shown in red.

2.3.4 Prediction of TATA box

Several core-promoter elements or general Transcription Factor Binding Sites (TFBSs) have been previously identified in eukaryotes. They are characterized by a strong positional preference relative to the Transcriptional Start Site (TSS) as for instance the TATA-box in the [-30, -25] area. The TATA-box and TATA-variants are regulatory elements involved in the formation of a transcription initiation complex. Both have been conserved throughout evolution in a restricted region close to the Transcription Start Site (TSS) (Bernard, Brunaud & Lecharny, 2010). The TATA box is one of the best characterized transcription factor binding sites. The TATA box is recognized by TATA-binding protein (TBP), a component of the TFIID complex, which recruits RNA polymerase II and directs the assembly of the preinitiation complex (PIC). However, it is not a ubiquitous element of core promoters, and other sequence motifs such as Y Patches seem to play a major role in plants (Bernard et al., 2010). A group of similar T/C-rich motifs has been identified in *Arabidopsis* and rice which has been designated as the ''Y Patch'' (or pyrimidine patch). The Y Patch was shown to be a plant-specific element with strict direction sensitivity although its biochemical role is not known. The Y Patch as compared with the TATA box

is a less conserved sequence motif of rice core promoters. The Y Patch consensus sequence is CYTCYYCCYC (Civán & Svec, 2009). The position of binding sites of proteins belonging to the transcription complex is important for the functioning of promoters since it determines both the TSS location and the transcription direction. Thus, a strong positional conservation of a novel regulatory element would strongly indicate a functional role (Bernard et al., 2010).

Identification of TATA boxes and other regulatory elements signifies the DNA sequence has promoter activity. Manual inspection of the 5' flanking regions from rice and *A.thaliana* identified regions that may act as TATA boxes, although none matched the consensus exactly. The Y Patch found in plant promoter regions was less obvious. Regions rich in pyrimidines were found but none that matched the consensus CYTCYYCCYC (Civán & Svec, 2009) These sequences must have elements recognised by RNA polymerase II, what they are exactly is unclear and would need to be determined experimentally through mutation analysis.

a) GTTGGTTTAAAACTTTGTTTGGACAGATTCGACTCAAACGTTGAACGCGTTTAAAAAAGTCAACAAAAGTCGA
AGGTTACCGGCTGGTCAGGTTAACCGTTAGAG***ATTAAT***GAACTGGCAGCTGTATGTCTTTTTTATAGTTCTGTCT

b) CGAATCAAAGAGTTCCAATTTATAGTCAACTAGTAAGACCAAACCTCAAGGAATTTTTATTTTACAAAATCTG
GGAATCTTTTACCATACTTCTAATCTGACTTTTCG***ATAAA***GCGTATAGTACAGGTAAGAAACTTTCGACAATCCG

c) CTCCAGAGGCGGGCTGCCTACCCCAACCTTACCTAGGCCGAGCAGCCGCTCGCTCGCTCGCTCTAAGCACTAA
GCTACTCGCTAACTACGGCGGATGAAAACGTCC***TAAT***CCCAAACGAAACACAAACGGGCTGCGCTCCTCCTACCC

d) AAAATTTAGTTTAAAATTTAGGTAGCTCAGATACTTTTTTTTAAGAACTATAAAATTTCTCTTCT***T***ATAATAT
GACCTTTATGCAGCTGATGACTCATGGAG***TTAA***CTGTTCCTTGCAGGTTTAGTAGGGTTTTAAGTTTCGATGGCA

Figure 2.7. Potential TATA boxes within the first150 bp of the 5' flanking region upstream of the transcription start site (TSS) from a) OBE1, b) OBE2, c) Os12g0514400 and d) Os11g0637400, respectively. The sequences highlighted in yellow are those of the TSS and the sequence highlighted in green are the predicted TATA boxes.

2.3.5 Prediction of Transcription factor binding sites

The prediction of the transcription factor binding sites (TFBS) present within the promoters of *A. thaliana* and *O.sativa* is essential in order to predict the transcription factors binding within the *N. benthamiana* flanking region. This was done by comparing the TFBS within the flanking sequences of *A.thaliana* and *O.sativa* and identifying those that are common. These would be predicted to also be common to the PVIP_Nb promoter region. This will help the prediction of what cues the gene might be responsive to in *N. benthamiana*.

Table 2.3.Transcription factor binding sites (TFBS) that are common between the 5' flanking regions of PVIP_At OBE1 and OBE2 and PVIP_Os from Chromosomes 11 and 12.TFBS were identified using TESS software.

Transcription factor	**Function of transcription factor**
AP-3	Directs stamen development (Kaufmann, Nagasaki & Jáuregui, 2010).
ARR 10	A multifunctional domain responsible for nuclear localisation (Yokoyama et al., 2007).
Broad_Complex-2	Function includes ethylene biosynthesis, disease resistance, phototropism, root phototropism and leaf morphogenesis (Gingerich et al., 2005).
bZIP (Basic leucine zipper)	Control of gene expression for seed storage, photomorphogenesis and organ establishment (Foster, Takeshi & Chua, 1994).
Dof 2	It is a repressor which blocks the transactivation of Dof 1. Dof 1 is responsible for greening and etiolating the protoplasts (Yanagisawa & Schmidt, 1999).
Dof 3	In rice it is responsible for gibberelin related growth whereas in At it is responsible for seed germination (Yanagisawa & Schmidt, 1999).
Forkhead associated domains(FHA)	Involved in plant growth and pathogen responses. Involved in root, shoot, flower development. FHA defective plants also lack an early organ development (Morris, Chevalier & Walker, 2006).
GATA 1,2 and 3	Involved in light response transcription (Reyes, Muro-Pastor & Florencio, 2004).
GT-1a	Similar to GT in function (Lam, 1995).
GT-1	Involved in light induced expression (Zhou, 1999).
Homeobox	It regulates the different patterns of development. It switches on a cascade of other genes (Lincoln, J. Long, Yamaguchi, Serikawa & Hake, 1994).
HMG-1	Brings about DNA repair, recombination and replication in the nuclei and plastids of plant cells (Grasser, 1998).
Myb	Role in development and defense (Jin & Martin, 1999).
SEF 4	Serrated leaves and early flowering. Responsible for photoperiod independent early flowering. Required for normal vegetative and reproductive development. Under short day conditions, it is responsible for smaller leaves with serrated margins (March-Diaz, Garcia-Dominguez, Florencio & Reyes, 2007).
TGA1a	Expressed to a greater degree in roots. Enhances the transcription rate in response to xenobiotic stress (Yamazaki, Katagiri, Imaseki & Chua, 1990).
Ubx	Essential for developmental control. Its loss accelerated root development and the inflorescence shoots (Rancour, Park, Knight & Bednarek, 2004).

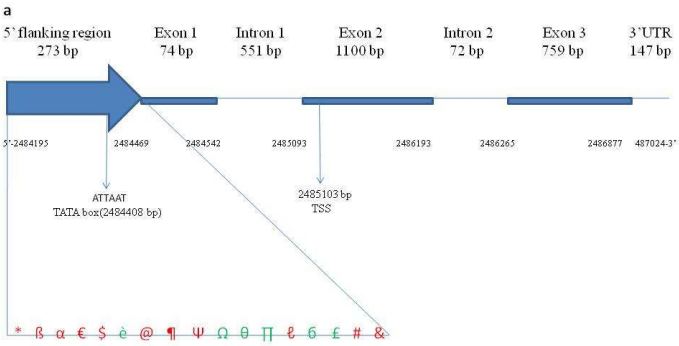
Analysis of possible TFBS for the 5' flanking regions of PVIP_At and PVIP_Os provides a means to predict how the PVIP_Nb gene might be regulated. The 5' flanking region for each gene was analysed for TFBS and sites common to the rice and Arabidopsis genes were identified. These common TFBS are presented in table 2.3. The PVIP genes are predicted to bind a variety of transcription factors ranging from environmental stimuli e.g. light response to pathogen response to plant repair and growth. It can be seen that many of these sites bind to transcription factors involved in plant development such as AP-3, bZIP, Dof3, SEF4 and UBx. This is not surprising given that OBE1 and OBE2 have been shown to have a role in plant development (Thomas et al., 2009). Thus, it is predicted that PVIP_Nb may be responsive to these or similar transcription factors.

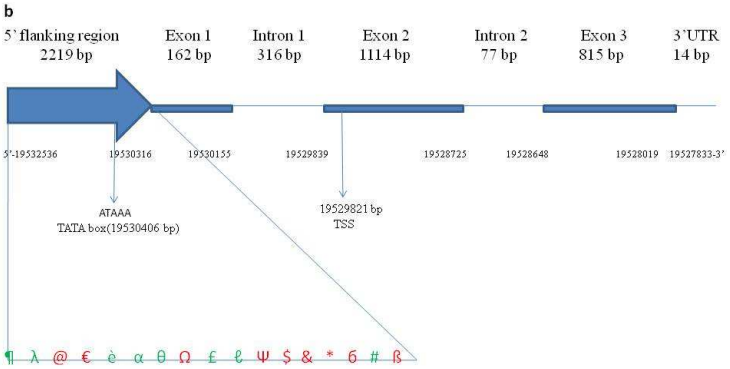
The PVIP gene may also be responsive to light. TFBS were predicted that have a role in photomorphogenesis (bZIP), light induced gene expression (GT-1), and phototropism (Broad_Complex-2). Thus, it is predicted that PVIP_Nb may be a light-responsive gene. Transcription factors that have been shown to have a role in pathogen infection (e.g. Forkhead associated domains) were also predicted to bind to the promoter region of the PVIP homologues. This suggests a potential role for the PVIP protein in response to pathogen attack which might be related to its role in potyvirus infection.

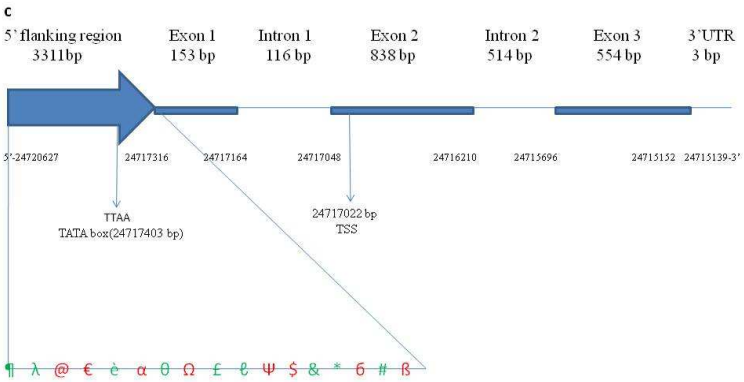
Hence, it is a possibility that the PVIP gene within *N. benthamiana* may respond to the same transcription factors as OBE1 and OBE2 and the rice homologues and may be responsive to developmental cues, pathogen attack and light. These responses would need to be proven experimentally.

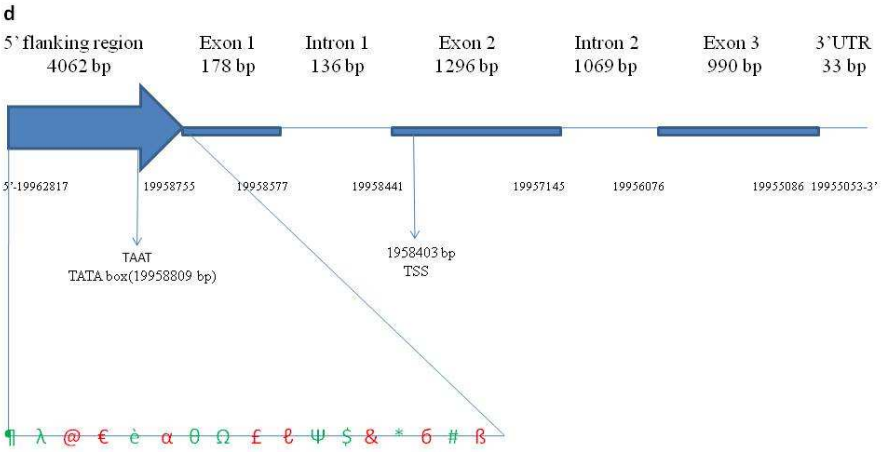
2.3.6 Predicted exon/intron gene structure

The exon-intron structures of the PVIP genes in *O. sativa* and the OBE genes in *A. thaliana* shows that each gene is made up of two introns and three exons, the sizes of which vary from one gene to another. The predicted transcription start site (TSS) for each of these genes was found in exon 2, indicating that exon 1 forms the 5' untranslated region of the mRNA. Since this gene structure is shared between a monocot and dicot species, it is likely that the PVIP_Nb gene will also have the same gene structure.


a
5`flanking region
273 bp
Exon 1
74 bp
Intron 1
551 bp
Exon 2
1100 bp
Intron 2
72 bp
Exon 3
759 bp
3`UTR
147 bp
5'-2484195
2484469
2484542
2485093
2486193
2486265
2486877
487024-3'
ATTAAT
TATA box(2484408 bp)
2485103 bp
TSS


b
5`flanking region
2219 bp
Exon 1
162 bp
Intron 1
316 bp
Exon 2
1114 bp
Intron 2
77 bp
Exon 3
815 bp
3`UTR
14 bp
5'-19532536
19530316
19530155
19529839
19528725
19528648
19528019
19527833-3'
ATAAA
TATA box(19530406 bp)
19529821 bp
TSS


c
5`flanking region
3311bp
Exon 1
153 bp
Intron 1
116 bp
Exon 2
838 bp
Intron 2
514 bp
Exon 3
554 bp
3`UTR
3 bp
5'-24720627
24717316
24717164
24717048
24716210
24715696
24715152
24715139-3'
TTAA
TATA box(24717403 bp)
24717022 bp
TSS


d
5`flanking region
4062 bp
Exon 1
178 bp
Intron 1
136 bp
Exon 2
1296 bp
Intron 2
1069 bp
Exon 3
990 bp
3`UTR
33 bp
5'-19962817
19958755
19958577
19958441
19957145
19956076
19955086
19955053-3'
TAAT
TATA box(19958809 bp)
1958403 bp
TSS

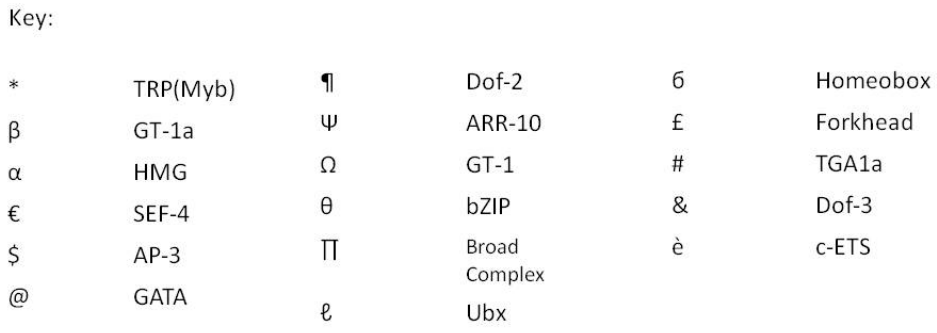


Figure 2.8. Gene structure for a) OBE1, b) OBE2, c) PVIP_Os chromosome 11, d) PVIP_Os chromosome 12. The transcription start site, predicted TATA box, the introns and exons are presented in this figure. Exons are shown as solid boxes with 5' and 3' untranslated regions and introns as lines. The coloured abbreviations are that of the predicted transcription factor binding sites, the key of which is present at the bottom of the figure.

Figure 2.8 shows that PVIP_At OBE1 and OBE2 like PVIP_Os chr11 and chr12 consists of 3 exons and 2 introns. Intron 1 of OBE1 and OBE2 was larger than intron 2 whereas intron 2 of PVIP_Os chr11 and chr12 was larger than intron 1. Exon 2 is the largest exon for both the species while exon 1 and a part of exon 2 form the 5' UTR. The gene structure has been maintained between monocots and dicots. Further, the relative sizes of the exons and introns between the two genes within a species provides further evidence of a gene duplication event giving rise to these two genes. Since *N. benthamiana* is a dicot, it is likely PVIP_Nb gene structure will be more similar to PVIP_At. Hence, it can be predicted that the 5'UTR will consist of exon 1 and some of exon 2 while intron 2 will be approximately 75 base pairs like intron 2 of PVIP_At_OBE1 and PVIP_At_OBE2 and not greater than 500 bp like the introns of PVIP_Os chr11 and chr12.

2.3.7 Identification of conserved domains

The PVIP_Nb protein sequence was assessed for the presence of conserved domains. Such domains can provide further clues as to the potential function of a protein. This analysis showed the presence of the DUF1423 (Domain of unknown function) (Bateman et al., 2004). It is of interest that the 5' flanking regions from the homologous genes from *A.thaliana and O.sativa* have a predicted binding site for HMG-1, a factor shown to have a role in DNA repair, recombination and replication. This may suggest that PVIP is part of a complex regulatory network involved in these nuclear processes.

The Plant HomeoDomain (PHD) zinc finger was observed within PVIP as well. PHD finger domains were first identified in plants and are now found to be present in nearly all the eukaryotes. The PHD finger domain consists of conserved domains which helps in the chromatin modification associated with transcriptional activity. Also, some of the PHD fingers might be responsible for the nuclear membrane localization of the protein(Safaee, Watkinson & Gillaspy, 2007). Hence, it might be this PHD finger domain which helps in the nuclear localization of PVIP, as suggested by Dunoyer et al (2004).

2.4 Discussion

Bioinformatic analysis can help to predict the putative functions of proteins. The identification of highly conserved, biologically characterised homologues from other species may also provide clues about the potential function of proteins. Sequence comparison may highlight the conservation of specific regions within the protein and biological evidence may provide clues about the specific functions of proteins.

There are many potential homologues of the *Nicotiana benthamiana* PVIP such as those from *Arabidopsis thaliana, Vitis vinifera, Oryza sativa* and *Pisum sativum*. Of all the species from which homologues were obtained, *A. thaliana* and *O.sativa japonica* have been well characterised having been used as model plants. The fact that their genomes have been sequenced provides important reference information for comparison with the PVIP sequence from *N.benthamiana*.

It appears that there are two PVIP genes within most plants, and in addition there may be allelic variation observed. Only the PVIP from pea, *N. benthamiana* and *A.thaliana* have been shown to have potyvirus VPg binding activity. Sequences within the same phylogenetic clade would be expected to have this activity also. There were several proteins which showed some homology to PVIP_Nb but as they were part of a different clade, it is predicted that these would not have VPg interacting activity. However, this would need to be determined experimentally.

On the basis of this homology one or more potential functions could be assigned to the PVIP protein in a healthy *N.benthamiana* plant. Several of the homologues have been

shown to have specific biological activities. For example, OBE1 and 2 from *A.thaliana* is a zinc iron binding protein. A sequence from tomato (accession number AY490248) has been shown to be a Constans interacting protein showing similarities to zinc finger transcription factors and known to promote flowering (Robson et al., 2001). Such homologies suggest that PVIP_Nb may act as a transcription factor. The fact that the OBE protein helps the transcription of genes required for auxin response indicates that the PVIP protein might have been involved in a similar function (Thomas et al., 2009). Auxin is responsible for plant growth and development. Hence, PVIP may indirectly be helping the growth and development of the plant. The binding of the transcription factor HMG-1 appears to be involved in DNA recombination, replication and repair (Grasser, 1998).

On the basis of pair-wise alignment analysis, PVIP from *N. benthamiana* was found to be closest to OBE2 and hence OBE2 may be the functional homologue in *A.thaliana* of PVIP_Nb. OBE1 is similar to OBE2, the two genes likely having arisen via gene duplication. Therefore, there is likely to be a second gene in *N.benthamiana* encoding a protein with greater homology to *A.thaliana* OBE1 (Figure 2.9). Further, PVIP genes are present in gymnosperms, fern and moss, indicating that this is an ancient gene and the extant genes have a common ancestor. This gene is also plant specific and therefore would be expected to encode a plant specific activity.

The closest homologues in *A.thaliana* have been studied. OBE1 and OBE2 appear to be plant homeodomain finger proteins that show functional overlap (Saiga et al, 2008) and are thought to control the transcription of genes required for auxin responses through the action of their PHD finger domains (Thomas et al, 2009). They appear to be necessary for the maintenance and/or establishment of both the shoot and root meristems (Saiga et al, 2008). Comparison of the 5' flanking regions of OBE1 and 2 with those from rice showed they had common predicted transcription factor binding sites relevant to plant development. This suggests the rice homologues have a similar function to OBE1 and 2. Further, since these are protein homologues of the PVIP_Nb sequence, this sequence too is likely to play a similar role in *N.benthamiana* (Thomas et al., 2009).

Analysis of the 5' flanking region from rice and *A.thaliana* genes predicts that PVIP should also be light responsive. Unpublished data suggested that the PVIP mRNA is induced in leaves when plants are subjected to the dark environment rather than light (C.Higgins, pers.

comm.). The relationship between light and dark responsiveness would need to be elucidated, however, it appears that the PVIP gene may be light responsive and that this activity may occur through the action of the transcription factors GT-1, GATA 1, 2 and 3 and also SEF4.

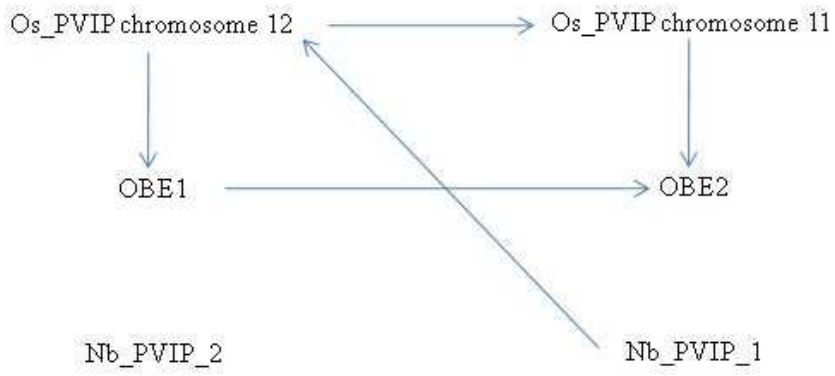


Figure 2.9.Relationship between the protein and nucleotide homologues. The nucleotide homologues show a greater identity with PVIP_Os_chr12 and OBE1 from *A.thaliana* whereas the protein homologues show a greater homology with PVIP_Os_chr11 and OBE2 of *A.thaliana*. The nucleotide similarity shows that the PVIP from *O.sativa* of chromosome 12 is similar to that of chromosome 11 and OBE1 is similar to OBE2 of *A.thaliana* which may be due to the allelic similarity. Also, PVIP_Nb is more similar to Os_PVIP chromosome 12 rather than chromosome 11. Hence, like PVIP_Nb_1 there may be another PVIP_Nb_2 gene which may be similar to chromosome11 of rice.

It is predicted that the PVIP_Nb gene would have the same structure as the genes from *A.thaliana* and rice. With the sequencing of an increasing number of *Solanaceae* genomes, it will be possible to compare the structure of these genes with those already identified. Isolating and sequencing the *N.benthamiana* gene and comparing it with the mRNA sequence would allow the location of the introns to be identified. Bioinformatic analysis of the structure of the 5'flanking region and possible promoter elements is also only predictive. Many plant genes do not have a TATA box, so that predicting the location of these allows the development of a hypothesis but functional tests must be carried out to determine the actual promoter elements.

The probable function of PVIP in *N.benthamiana* like *A.thaliana* might be to function as a zinc iron binding protein and hence protect the crop against oxidative stress and pathogens or as a plant homeo domain (PHD) family protein which has been found to be involved in transcriptional regulation of a pathogen defence-related gene (Korfhage et al., 1994). Like

tomato it may act as a Constans acting protein thereby facilitating flowering within the plant (Robson et al., 2001). The transcription factor Forkhead domain has been found to be involved in a similar function (Morris et al., 2006). Also, the transcription factors such as GT-1 and GATA were found to have an increased expression in light i.e. are light regulated (Chatterjee, Banerjee & Hannapel, 2007). The transcription factors belonging to the GATA family were found in photosynthetic tissue and showed an increased expression before dawn (Manfield, Devlin, Jen, Westhead & Gilmartin, 2007).

Bioinformatic analysis provides many predictions about gene function. However, these are just predictions and not a fool proof method for determining the function of a DNA sequence. Hence, it is essential to perform wet lab experiments to test the functional importance of specific regions within a promoter or gene. In order to do this, the region of interest must be isolated. The next chapter describes attempts to isolate the 5'flanking region from PVIP_Nb for functional testing.

Chapter 3

Promoter isolation

3.1 Introduction

The promoter is one of the most important controlling elements in gene expression; however, our current knowledge of promoters with respect to their functional regions is relatively insignificant (Ren et al., 2005). Hence, knowledge of promoter structure is essential in order to understand gene regulation (Yamamoto et al., 2007). Information regarding promoter function is essential for the use of promoters to express transgenes in transgenic plants in a predictable manner as well as for the creation of synthetic plant promoters (Rushton et al., 2002).

In order to understand the structure and function of a promoter, the region predicted to contain the promoter must be isolated and characterised. This characterisation would include sequencing and comparison of sequences with known promoters to identify regions of functional importance. This may provide clues about temporal and spatial gene expression. In order to confirm the predicted function of a promoter region, the predicted region must be isolated and its function tested. Joining a promoter to a reporter gene such as the green fluorescent protein (GFP) gene or the β-glucuronidase (GUS) gene has allowed the spatial and temporal functioning of a promoter to be determined (Sahooa et al., 2009).

3.1.1 Techniques for isolating promoters

Several techniques have been used to isolate promoters. These methods include PCR-based techniques and hybridisation-based techniques.

3.1.1.1 Polymerase chain reaction-based techniques for promoter isolation

The polymerase chain reaction (PCR) enables the enzymatic amplification of specific DNA regions without having to resort to the process of cloning (Ochman, Gerber & Hartl, 1988). Inverse polymerase chain reaction (IPCR) is an important and powerful technique for obtaining the flanking regions of the open reading frame (ORF) whose sequence information is known i.e. the region lying outside the boundary of a known gene (Huang, Zhang & Birch, 2000, Figure 3.1). IPCR is carried out like a standard PCR except that, rather than the primers facing inwards towards the gene of interest, the primers face outwards towards the flanking regions (Ochman, Gerber & Hartl, 1988). The approach of IPCR involves genomic DNA (gDNA) isolation and digestion with a restriction enzyme known to cut inside the coding region, downstream of the transcription start site (TSS). This enzyme would be expected to also cut somewhere upstream of the TSS and the region in between the two cut sites should include at least part of the promoter region. The digested gDNA is diluted and ligated so that the digested fragments will ligate only to themselves, thereby creating circular molecules. These monomeric circles are then used as substrates for the enzymatic amplification of the flanking regions with the oligonucleotide primers facing the opposite direction (Ochman et al., 1988). Primers designed to bind to sequence downstream of the TSS, but upstream of the original cut site and facing out from each other will allow amplification of the flanking region. This amplified product can then either be sequenced directly or cloned.

The unknown sequence flanking a small stretch of the known sequence of DNA cannot be amplified by the conventional PCR. Hence, the major advantage of inverse PCR (IPCR) is to amplify the flanking unknown sequence by using two gene-specific primers. IPCR, although very useful has certain drawbacks. One drawback may be that a specific product may not be amplified using just one pair of primers due to the complexity of the gDNA (Han, Ming, Wang & Guo, 2007) and the size of the flanking region must be around 2 to 3 kb or less or else the region won't amplify efficiently (Ochman, Gerber & Hartl, 1988).

There has been a growing interest in the isolation of gDNA for the purpose of cloning and sequencing (Beck, O'Keeffe, Coull & Koster, 1989). Also, for the purpose of plant research the isolation of long and pure quality DNA is essential. Until a few years back the major

setback for the isolation and detection of DNA had been the various contaminants which would interfere with the action of the enzyme restriction endonucleases during cloning leading to the isolation of DNA of insufficient length (Murray & Thompson, 1980).

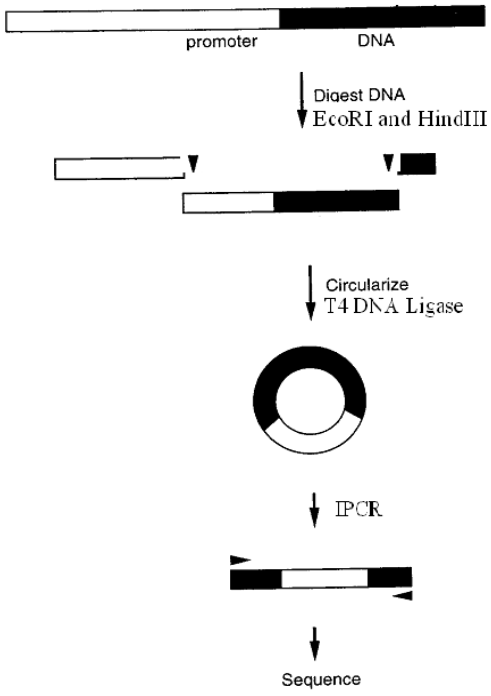


Figure 3.1. Diagrammatic representation of the IPCR.

Modified from "Inverse PCR (IPCR) for obtaining promoter sequence," by Tony Triglia, 2000, *Methods in Molecular Biology, 130, p.80.* Copyright © by Copyright Agency Limited.

3.1.1.2 Hybridisation-based techniques for promoter isolation

Hybridisation-based techniques for promoter isolation rely on the availability of a sequence to be used as a probe. gDNA is isolated, fragmented and the fragments cloned into an appropriate vector. The resulting fragment library is then screened using a labelled probe. The probe usually incorporates the 5' portion of the mRNA and the clones that hybridise to the probe would have this region together with some flanking region DNA. The advantages of hybridisation may include the precision i.e. the precise size of the target gene and also the sensitivity of the reaction. The disadvantages might be the possible nonspecific hybridisation of the probe to the a target gene (Lodish et al., 2000).

An adaptation of screening a genomic library is to create a sub-genomic library of fragments of a specific size. Southern hybridisation of digested gDNA will provide information about the sizes of 5' flanking fragments. Using a range of enzymes, a fragment of a suitable size can be identified and gDNA of this specific size can be gel purified and cloned into a vector. The specific sequence is then identified by colony hybridisation as for a complete genomic library. One advantage of this is that fewer clones need to be screened. A disadvantage may be that the particular region of interest may not be digestible by a useful restriction enzyme.

The technique of colony hybridisation (Figure 3.2) involves the growth or isolation of a culture in a suitable medium followed by the transfer of small amounts of the culture from the medium on to a nitrocellulose membrane where the cell is denatured and its DNA released. This is followed by the addition of a probe into the membrane and allowing the hybridisation to take place. Following hybridisation, the membrane is rinsed with water to let the non-hybridised probes drain off. For radioactive probes, the colony identification is brought about by a process called autoradiography in which the membrane is placed within an X-ray film. The hybridised colonies are observed as black spots within the membrane. It is these colonies which possess the necessary target sequence and by comparing the location of the target colony with its location in the master plate, the colony can be hybridised and sub-cultured if necessary (Sommerfelt, Grewal, Gaastra, Svennerholm & Bhan, 1992).

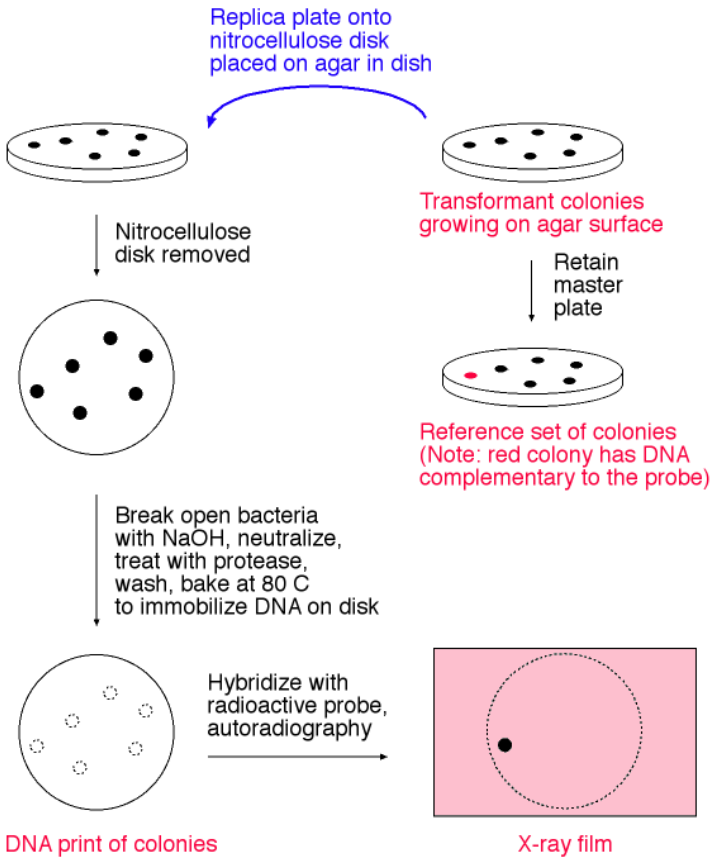


Figure 3.2.Diagrammatic representation of colony hybridisation.

From "Colony hybridisation" Retrieved May 28, 2010, from http://oregonstate.edu/instruct/bb492/figletters/F9.html

Colony hybridisation allows one to find a specific sequence within a complex mixture of similar sequences. Colony hybridisation, or any hybridisation for that matter, consists of reactions which occur in the presence of large quantities of molecules that are similar but not identical to the target site (Davis, Duerr & Jacobs, 1992). Hence, it is essential that a probe finds the target molecule in a mixture of millions of related but non-complementary molecules. These hybridization reactions are specific because the probes will bind only to sites that have complementary sequences (Born & Miyada, 1991). The advantages of this technique include the fact that the target gene can be selected from the millions of copies that are present. Also, unlike other hybridisation techniques, colony hybridisation does not require any nucleic acid isolation. However, there are certain disadvantages as well, such as the amount of time required. It is time consuming for a colony to grow and the hybridisation to show within the X-ray screen. Similarly, for the identification of the

microorganisms, the colonies will have to grow to a detectable size before the procedure can be efficiently used (Davis et al., 1992).

Southern hybridisation has many uses in the field of molecular biology and brings about the identification of DNA fragments which are complementary to a known DNA sequence (Lutz, 2003). The process of Southern blotting involves the transfer of DNA fragments by absorption, separated in electrophoretic gels to membrane filters for detection of specific base sequences by complementary probes. The chromosomal DNA is isolated from the organism of interest and digested completely with a restriction enzyme. These fragments are then transferred from the gel onto a nitrocellulose membrane usually performed by simple capillary action (LeBrasseur, 2010). One of the uses of Southern hybridisation is that it can provide a physical map of restriction sites within a gene in a chromosome and may also allow comparison between the genome of a particular organism and that of an available gene or gene fragment (Lutz, 2003). Other uses of Southern hybridisation include the detection of major gene rearrangements and deletions involved in disease, identifying structurally related genes within the same species or homologous genes in other species (LeBrasseur, 2010).

The probe used for hybridisation can be radioactively labelled with ^{32}P (high energy β-particle emitter) or can be non-radiolabelled. If radiolabelled then the membrane is first washed to remove non-specifically bound probe and is then exposed to X-ray film (autoradiography). This allows the identification of the sizes and the number of fragments of chromosomal genes with strong similarity to the gene or gene fragment used as a probe. Southern blots can detect very small amounts of nucleic acids (Lutz, 2003). Radiolabels are often used because of their high energy and ease of incorporation into the phosphate groups of dNTPs (Lutz, 2003). However, the radiolabelled technique has disadvantages such as short half life (about 2 weeks) which means that the probes must be used immediately, and the labeling reagent cannot be stored for long. Researchers must have access to a dark room to set up and develop films and there are also contamination problems i.e. all materials and equipment must be dedicated to radioactive work only. Hence, regular lab-wide testing for contamination is required and the added expenditure for the disposal of radioactive waste. Therefore, detection using non-radiolabelled probes was developed and involves different detection systems such as colorimetric, fluorescent, and

chemiluminescent. Colorimetric detection generally involves the production of a coloured precipitate which can be seen with the naked eye. Usually the DNA probe itself will be labelled with an antigen such as digoxigenin; following hybridization to its target gene the antigen would be exposed to an anti-digoxigenin antibody conjugated to an enzyme capable of catalyzing a colorimetric reaction (one commonly used example is alkaline phosphatase which will act on substrates NBT & BCIP to produce a dark purple product). Fluorescent detection involves probes which are directly labelled with fluorophores, or more likely, probes which are coupled to fluorescent molecules indirectly e.g. if probe is labeled with biotin, it would be exposed to avidin conjugated to a fluorescent tag. Fluorophores emit light when excited by light of an appropriate wavelength. The third type which is the chemiluminescence technique is a combination of the colorimetric and fluorescent technique of detection i.e. an enzymatic reaction that triggers the release of ordinary visible light (Rogers, 2008).

As for PCR-based techniques, hybridisation techniques require the isolation of good quality gDNA. They also require the use of labelled probes. Hybridisation techniques have been developed using radioactively labelled probes with detection via exposure to X-ray film (Figure 3.3). Non-radioactive techniques have also been developed using chemiluminescent or colorimetric detection chemistries (Beck et al., 1989).

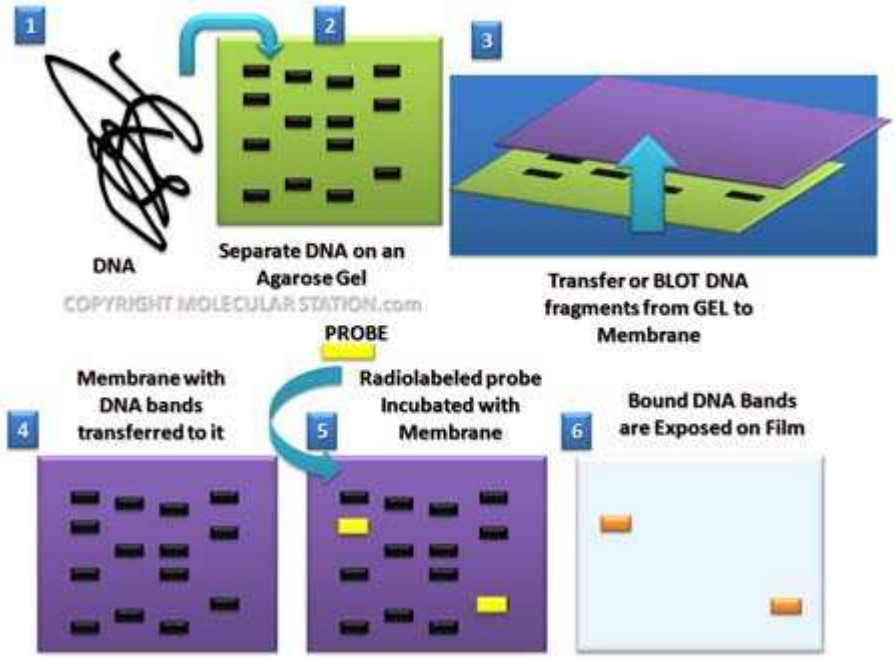


Figure 3.3.Diagrammatic representation of Southern hybridisation.

From "Southern blotting procedure (2007)". Retrieved May 28, 2010, from www.molecularstation.com/ro/dna/southern-blot/

The non-radioactive method of detection can be used instead of the radioactive method if it fulfils the requirements of sensitivity, reusability of the filters, stability of the probe used, low expense and also causes no health or environmental hazard. The non-radioactive technique has been successfully used for a variety of plants such as rice, potato, maize etc (Rogers, 2008). However, deciding which is preferred, the use of radioactive or non-radioactive techniques, is a question still being answered by the plant biologists and may be dependent on the particular experiment being carried out (Neuhaus-Url & Neuhaus, 1993). The use of digoxigenin itself is advantageous for the hybridisation technique because of the molecule's limited distribution within plants, resulting in no background reactions from non-specific antibody binding.

Isolation of the 5' flanking region from the *N.benthamiana* PVIP gene and assessment of its promoter activity has the potential to provide clues about the potential function of the PVIP protein. The aim of this part of the project was to isolate this flanking sequence. Two

techniques, IPCR and Southern hybridisation, were tried with a view to characterising the sequence of the PVIP 5' flanking region.

3.2 Methods and Materials

A general overview of the approaches attempted and planned for the isolation of the 5' flanking region of the *N.benthamiana* gene is given in Figure 3.4

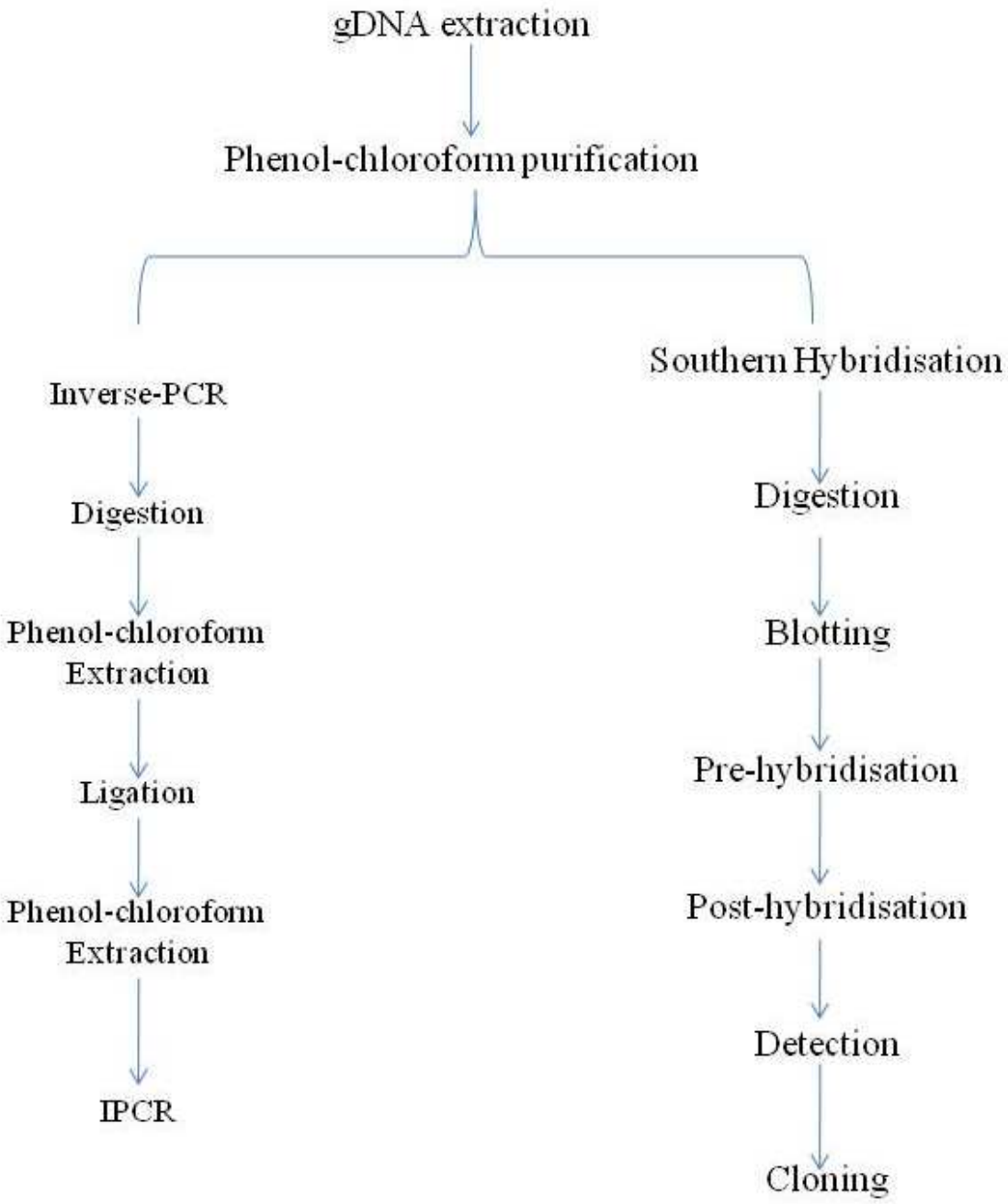


Figure 3.4. An overview of the approaches taken or planned to isolate the promoter region.

3.2.1 Sub-culturing *Nicotiana benthamiana*

The seeds of *Nicotiana benthamiana* were surface sterilized using Janola for 20 minutes followed by subsequent washing in sterile distilled water. These seeds were planted onto full strength Murashige and Skoog (MS) medium supplemented with 15g/L sucrose and incubated at 25^0C under 16 hours light and 8 hours dark conditions. These plants where grown until they reached the height of the culture vessel and were then sub-cultured by excising the young tissues (green leaves attached to the petiole) and planting it into fresh sterile medium. The MS medium (500ml) was made up as follows: 15g of sucrose, 5g of agar, and 2.2g of M.S agar made up to 500 ml of sterile distilled water.

3.2.2 Inverse PCR

3.2.2.1 gDNA extractions

Three techniques were tested for gDNA extraction. The first being the salt extraction method by Aljanabi and Martinez (1997), the second being the use of the InVitrogen PureLink DNA extraction kit and the third being the protocol by Dellaporta et al (1983).

3.2.1.1a Salt extraction by Aljanabi and Martinez

gDNA was extracted from *N.benthamiana* using a scaled up modification of the method of Aljanabi and Martinez (Aljanabi & Martinez, 1997). Approximately 0.4 g of leaf tissue was extracted rather than the 0.05 to 0.1 g reported by the authors. This was to ensure a high concentration and yield of the DNA for downstream processing. Leaf tissue was placed into 2 mL sterile salt homogenising buffer (0.4 M NaCl, 10mM Tris-HCl ph 8.0, 2 mM EDTA pH 8.0) and homogenised using a Polytron Tissue Homogeniser for 30-45 seconds. The homogenised samples were mixed with 200 µL of 10% SDS and 40 µL of 10mg/mL Proteinase K and then incubated at 55^0C overnight. The homogenised samples were then split (approximately 450 µL) into two 1.5 mL tubes. Salt solution (300 µL of 6M sodium chloride) was added into each of these tubes, vortexed for 30 seconds and then centrifuged at high speed i.e. 13,000 rpm for 30 minutes. The supernatant was transferred into a new 1.5 mL tube and an equal volume (750 µL) of iso-propanol was added, mixed well and

incubated at -20^{0}C for 1 hour. The samples were then centrifuged at 13,000 rpm for 20 minutes at 4^{0}C. The supernatant was discarded and the pellet was washed with 70% ethanol, dried and finally suspended in 300 µL of sterile distilled water. The gDNA was then stored at -20^{0}C.

3.2.1.1b InVitrogen PureLink Plant Extraction kit

The InVitrogen PureLink Plant Extraction kit was used to extract the gDNA from *N.benthamiana* samples as described by the manufacturer. Approximately 100 mg of the fresh leaf sample was ground in a sterile mortar and pestle using liquid nitrogen and transferred to a sterile 1.5 mL tubes. Resuspension buffer (250 µL) was added and 15 µL each of 20% SDS and 20 mg/ml of RNase A was added to the diluted plant material. The lysate was then incubated at 55^{0}C for 15 minutes to complete the lysis process after which it was centrifuged at 13,000 rpm for 5 minutes. The clear supernatant was then transferred into a new microcentrifuge tube and 100 µL of the 'precipitation buffer' was added, vortexed and incubated on ice for 5 minutes. The mixture was centrifuged again for 5 minutes at high speed and the clear supernatant was transferred into a fresh microcentrifuge tube into which 375 µL of the binding buffer with ethanol was added. This sample was then added into a PureLink spin cartridge which was attached to a 1.5 mL collection tube. The cartridge was then centrifuged at 10,000 g for 30 seconds at room temperature. After discarding the flow through, the spin cartridge was placed into a wash tube. Around 500 µL of the 'wash buffer' was added into the spin cartridge and centrifuged at 10,000 g for 30 seconds at room temperature. After discarding the supernatant, another 500 µL of the wash buffer was added into the spin column. The cartridge was centrifuged again at 10,000 g for 30 seconds at room temperature and the wash-through solution thrown away. The wash was performed again and the steps that were performed previously i.e. of centrifugation and decanting the supernatant were repeated. Any residual wash buffer from the tubes was discarded after centrifuging the tubes at maximum speed for 2 minutes. The spin column was then placed on a DNase-free 1.5 ml microcentrifuge tube and 100 µL of the elution buffer was added into the tube, incubated at room temperature for one minute and centrifuged at maximum speed at room temperature for a minute. This elution step was repeated. The DNA was eluted in a total of 200 µL and stored at – 20^{o}C.

3.2.1.1c Dellaporta et al (1983)

The *N.benthamiana* plants were subjected to dark treatment by covering them in foil for approximately two days and then extracted using the Dellaporta method (Dellaporta et al., 1983). One gram of plant tissue was ground to a fine powder in liquid nitrogen in a sterile mortar and pestle and transferred to a sterile 15 mL tube. Approximately 7.5 mL of extraction buffer (100mM Tris pH 8.0, 50 mM EDTA pH 8.0, 500 mM NaCl and 10 mM beta-mercaptoethanol) was added and mixed well before adding 0.5 mL of 10% SDS mixing gently and incubating at 65^{o}C for 15 minutes. Potassium acetate (3.5 mL of 5 M) was added to the solution, incubated on ice for 20 minutes and then centrifuged at 2800 rcf for 10 minutes. The supernatant was transferred to a new 15 mL Falcon tube and centrifuged again at 2800 rcf for 10 minutes. The clear supernatant was transferred into a fresh Falcon tube and 5 ml of iso-propanol was added to the mixture, incubated on ice for 20 minutes before centrifugation at 2800 rcf for 10 minutes. The supernatant was then discarded and the DNA pellet was air dried for 5 minutes before being resuspended in 200 µL of sterile TE buffer pH 8.0.

An equal volume i.e. 200 µL of phenol-chloroform- isoamyl alcohol (25:24:1) was added to the DNA, mixed gently by inverting and centrifuged at 13,000 rpm for 3 minutes. The aqueous phase was then dispensed into a fresh 1.5 mL tube and 200 µL of chloroform-isoamyl alcohol (24:1) was added and mixed gently by inverting. This tube was centrifuged at 13,000 rpm for 3 minutes and the aqueous phase was added into a fresh tube. The tubes were then subjected to ethanol precipitation i.e. 0.1 volume of 3 M Na-acetate pH 5.2 and two volumes of cold 100% ethanol were added and the mixture was incubated at -20^{o}C for a minimum of 2 hours or at -80^{o}C for a minimum of 30 minutes. The reaction mixture was then centrifuged at 13,000 rpm for 15 minutes. The ethanol was carefully removed and the pellet was washed with 200 µL of 70% ethanol. The pellet was then left to dry for 5 minutes before suspending the DNA in 50 µL of sterile distilled water.

In order to be sure that the DNA was free of any RNA, the extracted DNA sample was subjected to an RNase treatment. Approximately 5 µL of RNase A (100 µg /µL) was added to 50 µL of the DNA sample and incubated at 65 ^{o}C for 10 minutes. The treated samples were then purified using the phenol-chloroform extraction protocol in order to get rid of the remaining RNase enzyme.

3.2.2.2 Assessment of gDNA quality

3.2.2.2a Absorbance

The concentration and purity of the DNA was determined by measuring the absorbance at 260 nm of 2 µL of the sample using a Nanovue spectrophotometer. The DNA concentration was calculated by using the standard A_{260} value of 1 corresponding to a DNA concentration of 50µg/mL. The absorbance at 280nm was measured on the same sample. The purity of the DNA was determined by calculating the A_{260}:A_{280}.

3.2.2.2b Agarose gel electrophoresis:

The integrity of the extracted gDNA was confirmed by agarose gel electrophoresis as follows:

For DNA extracted using the Aljanabi and Martinez (1997) method a 1% agarose gel was run in1X TAE buffer (242g Tris base, 57.1mL glacial acetic acid and 18.6g EDTA in 1 litre of distilled water) for 3 hours at 45 volts. The molecular weight marker was Marker II i.e. λ *Hin*dIII (500 ng, Roche Applied Science). Approximately 250 ng of each of the samples was loaded onto the gel.

For DNA extracted using the InVitrogen extracted DNA a 1 % agarose gel was run in 1X TAE buffer for 2 hours at 65 volts. The molecular weight marker was Marker II i.e. λ *Hin*dIII (500 ng, Roche Applied Science) and around 144-272 ng (i.e. 10 µL) of each sample was run on the gel.

For DNA extracted using the Dellaporta (1983) method the gDNA was run on a 1% agarose gel in 0.5X TBE buffer (108g Tris Base, 55g Boric acid and 9.3g EDTA in 1 litre distilled water) at 65V for 90 minutes. The molecular weight markers used were marker XIV (750 ng, Roche Applied Science) and Marker II i.e. λ *Hin*dIII (750 ng, Roche Applied Science). Approximately 1.5 µg gDNA was run on the gel.

3.2.2.3 Restriction digestion

3.2.2.3a Digestion of gDNA extracted by the Aljanabi and Martinez method

Approximately 500 ng of gDNA was digested with 40 Units of 10 U/µL *Eco*RI and 10 U/µL *Hin*dIII (Fermentas) separately in 1X reaction buffer in a total volume of 60 µL. The reaction mixture was incubated at 37^{o}C overnight followed by the enzyme inactivation at 65^{o}C for 10 minutes and 20 minutes for *Eco*RI and *Hin*dIII, respectively.

3.2.2.3b Digestion of gDNA extracted by the InVitrogen method

Approximately 750 ng of gDNA was digested with 40 units of each enzyme as described in 3.2.2.3a.

3.2.2.3c Digestion of gDNA by the Dellaporta et al (1983) extraction method

Approximately 2.5 µg of gDNA was digested with 10 units of each of the two enzymes in 1X reaction buffer in a total volume of 50 µL. The reaction mixture was incubated at 37^{o}C overnight and the enzymes were then inactivated by heating the reaction mixture at 65^{o}C for 10 minutes for *Eco*RI and 20 minutes for *Hin*dIII respectively.

3.2.2.3d Post-digestion treatment

Digested DNA was extracted twice with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) with centrifugation at 13,000 rpm for 5 minutes in between. The DNA concentration was then determined using a Nanovue spectrophotometer as described3.2.2.2a.

3.2.2.4 Agarose gel electrophoresis of digested gDNA

The presence and the quality of the digested gDNA could be confirmed using an agarose gel electrophoresis.

3.2.2.4a Salt extracted gDNA by Aljanabi and Martinez method

Approximately 1.27 µg of the digested samples were subjected to electrophoresis through a 1% agarose gel in 0.5X TBE for 1 hour at 65V. Digested DNA was compared with an

equivalent amount of undigested gDNA, molecular weight marker XIV (500 ng, Roche Applied Science) and marker II i.e. λ HindIII (500 ng, Roche Applied Science).

3.2.2.4b InVitrogen extracted DNA

Approximately 145 ng of digested and undigested gDNA was subjected to electrophoresis as described in 3.2.2.4a. The gel was run for 90 minutes at 65V with molecular weight marker II (500 ng, Roche Applied Science).

3.2.2.4c gDNA extracted by the Dellaporta et al (1983) method

Approximately 1 µg of digested and undigested gDNA was run as described in 3.2.2.4a. Molecular weight marker II (750 ng, Roche Applied Science) was also run on the gel for comparison.

3.2.2.5 Ligation of digested gDNA

3.2.2.5a Test ligation reactions

The reaction conditions for ligating digested *N.bethamiana* DNA were optimised using λ *Hin*dIII (Roche Applied Science). The concentration of the λ *Hin*dIII varied from 50 ng to 750 ng while that of the enzyme T4 DNA ligase (Fermentas) varied from 1 µL (5U) to 5 µL (25 U). The total reaction volume was 50 µL, therefore, 5 µL of the 10X T4 DNA ligase buffer was included. The ligation reaction was then incubated at 16^{o}C overnight followed by the enzyme inactivation at 65^{o}C for 10 minutes.

3.2.2.4b Self Ligation of the digested Aljanabi and Martinez extracted DNA

Approximately 50 ng of the digested gDNA was self ligated with 15 units of T4 DNA ligase (Fermentas) with 1X ligase buffer in a total volume of 50 µL. The reactions were incubated at 16^{o}C overnight followed by the enzyme deactivation at 65^{o}C for 10 minutes. The ligated DNA was extracted twice with phenol: chloroform: isoamyl alcohol as described in 3.2.2.3d.

3.2.2.4c Self Ligation of the InVitrogen extracted DNA

Approximately 200 ng of the digested gDNA was self ligated with 20 units of the T4 DNA ligase in 1X ligase buffer in a total volume of 100 µL used for ligation. The reactions were treated as described in 3.2.2.4b.

3.2.2.4d Self Ligation of the digested Dellaporta DNA

Approximately 1 µg of the digested gDNA was self ligated with 75 units of the T4 DNA ligase in 1X ligase buffer in a total volume of 50 µL used for ligation. The reactions were treated as described in 3.2.2.4b.

3.2.2.6 Agarose gel electrophoresis of ligated DNA

3.2.2.6a Ligation test reaction

The test ligation reactions using λ *HindIII* were checked by running 10 µL of each sample on a 1% agarose gel with 0.5X TBE run at 65V for 60-90 minutes. Molecular weight marker II i.e. λ *Hin*dIII (750 ng, Roche Appleid Science) was run for comparison.

3.2.2.6b Aljanabi and Martinez (1997) extracted DNA

DNA ligation was confirmed by running approximately 300 ng and 120 ng of the *Eco*RI and *Hin*dII digested and ligated DNA through a 1% agarose/0.5X TBE gel. The gel was run at 65V for 60-90 minutes. For comparison, 150 ng of unligated DNA and molecular weight marker II (500 ng, Roche Applied Science) and marker XIV (500 ng, Roche Applied Science) were included.

3.2.2.6c InVitrogen extracted DNA

Ligation of the digested DNA was confirmed by running approximately 50 ng of the digested products as described in 3.2.2.6b.

3.2.2.6d Dellaporta et al (1983) extracted DNA

DNA ligation was confirmed by running approximately 650 ng for *Eco*RI digested DNA and 375 ng for *Hin*dIII digested DNA as described in 3.2.2.6b.

3.2.2.7 Inverse PCR (IPCR)

Four sets of primers were used for IPCR (Table 3.1). The table below gives the information regarding the primers with regards to their length, orientation and melting temperatures (Tm). The software OligoExplorer was used to design the primer sequences. Their relative positions and direction of amplification are shown in Figure 3.5.

Table 3.1. Primers used for IPCR.

Primer	**Length**	**Orientation**	**Sequence**	**GC%**	**Tm (^{o}C)**	**Nucleotide positions**
IPCR-1	20 bp	Reverse	GGCTCTCAAATTTTCTCCCC	50	60	59-39
IPCR-2	20 bp	Reverse	GGCAGCATTTTTCCGATACG	50	60	195-175
IPCR-3	18 bp	Forward	CAGCACAGGGAGGAGTTC	61.1	58	577-595
IPCR-4	21 bp	Forward	GTAGAGGTGATTTGACGGATG	47.6	62	620-641

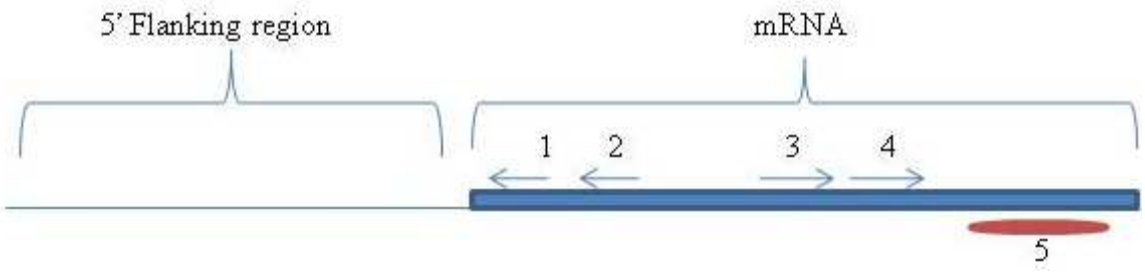


Figure 3.5.Relative positions of the IPCR primers and the probe used in Southern hybridisation with respect to *N.benthamiana* mRNA. The numbers 1 to 4 represent the corressponding primers and number 5 represents the location of the probe which is from 845 bp to 1291 bp.

The protocol followed for IPCR was very similar in all the three cases with respect to the kits and the conditions used. The kit used was the Expand High Fidelity PCR system from

Roche, Applied Science. The only factor that changed in each of the IPCR's was the amount of template added in the reaction.

Digested and ligated gDNA was amplified in a total volume of 25 µL containing 12.5 µM each of forward and reverse primer, 1X reaction buffer and 1 µL of the Expand High Fidelity enzyme mixture. For the Aljanabi and Martinez extracted gDNA 25 ng was used as the template, for the InVitrogen extracted gDNA 10 ng was used whereas for the Dellaporta extracted gDNA 500 ng of the gDNA was used as the template. The PCR conditions were as follows; Initial denaturation at 94^{o}C for 2 minutes followed by 10 cycles of denaturation, annealing and extension at 94^{o}C for 15 seconds, 50^{o}C for 30 seconds and 68^{o}C for 4 minutes respectively. This was followed by another 20 cycles of denaturation, annealing and extension at 94^{o}C for 15 seconds, 50^{o}C for 30 seconds and 68^{o}C for 4 minutes+5 seconds per cycle respectively. The final extension was at 72^{o}C for 7 minutes and the reactions were then held at 15^{o}C. Also, for the InVitrogen extracted DNA a gradient PCR was performed with the temperatures varying from 51.6 to 59.6. The reaction constituents and conditions were however, kept constant.

The amplified products were then run on a 1% agarose with 1X TBE gel at 65V for 90 minutes with 500 ng each of molecular weight markers II and XIV (Roche, Applied Science).

3.2.3 Southern Hybridisation

Southern hybridisation was performed to identify a suitable size fragment/s for isolation of the 5' flanking region from the PVIP gene of *N.benthamiana*. A portion of the *N.benthamiana* PVIP mRNA sequence was used as the probe.

3.2.3.1 Preparation of PVIP probe

3.2.3.1a Isolation of total RNA

The RNA used for the preparation of probe was extracted from the leaf tissue of *N.benthamiana*. RNA was extracted using the Spectrum Plant Total RNA kit from Sigma Aldrich, the protocol of which is discussed in Chapter 4, section 4.2.2. The quality of the

RNA was determined by running a 1% agarose gel of the extracted RNA with 0.5X TBE buffer at 65V for 90 minutes. A Nanovue spectrophotometer was used to determine the concentration of 2 uL RNA. The RNA concentration was calculated by using the standard A_{260} value of 1 corresponding to a RNA concentration of 40 µg/mL. The absorbance at 280nm was measured on the same sample. The purity of the DNA was determined by calculating the A_{260}:A_{280}.

3.2.3.1b DIG labelling probe

A 500bp region of the PVIP mRNA was DIG-labelled by two step reverse transcriptase PCR (RT-PCR). The PVIP mRNA was first converted to cDNA using the Quanta Biosciences Flexi cDNA synthesis kit following the manufacturer's instructions. Approximately 300 ng of the RNA was combined with 2 µL of the oligo dT and random primers from the kit and the reaction volume was made up to 15 µL using nuclease free water. The reaction mixture was then centrifuged and vortexed followed by the addition of 1 µL of the reverse transcriptase (RT) enzyme and 4 µL of the Flexi reaction mix (5X). Thus, the total volume was 20 µL. The conditions used for the cDNA synthesis were as follows; 25^{o}C for 10 minutes, 42^{o}C for 90 minutes and 85^{o}C for 5 minutes. The tubes were then kept at 4^{o}C.

The single stranded cDNA (2 µL) was used as a template for PCR using the Expand High Fidelity DIG-labelled PCR system from Roche, Applied Science. The other components of the reaction included 15.1 µL of sterile distilled water, 2.5 µL of the 10X buffer and 2.5 µL of DIG label. Forward and reverse primers used were PVIPT7FPRO and PVIPSP6R, respectively. The sequence of the two primers were as follows, PVIPF-T7Pro 5'-TAATACGACTCACTATAGGGAAATGGTTTCTGCAACCTTT-3'and that of PVIPR-SP6 was 5'-ATTTAGGTGACACTATAGGGATCCACATCAAGCTCTGG-3'. The primer concentration was 10 µM and 1.25 µL of each was added. Expand HiFi enzyme (0.4 µL) of was also added into the reaction mixture. The PCR conditions were as follows; 94^{o}C for 2 minutes, followed by 35 cycles of 94^{o}C for 15 seconds, 50^{o}C for 30 seconds and 68^{o}C for 1 minute. The tubes were finally treated at 68^{o}C for 5 minutes before holding them at 4^{o}C.

For comparison, control reactions were also set up. To ensure the amplification was specific to the mRNA, a cDNA synthesis reaction was set up without RT and then used as template in a PCR with and without the DIG label. To check the efficiency of the DIG labelling reaction, a parallel reaction was set up with cDNA template but no DIG label. Where no DIG was added, 2.5 µL of 2 mM dNTP was added rather than the DIG label.

10 µL of the PCR was then run on a 1% agarose gel with 1X TAE buffer at 65V for 90 minutes. The markers that were run were the 100 bp marker i.e. Marker XIV (750 ng) and also the DIG labelled marker II (λ *Hin*dIII, 400 ng).

3.2.3.2 gDNA extraction

DNA was extracted using the method of Dellaporta et al as described in section 3.2.2.1c.

3.2.3.3 Southern hybridisation setup

3.2.3.3a Restriction digestion of genomic DNA

Ten µg, 20 µg and 50 µg of gDNA were digested with 40 units of *Eco*RI or *Hin*dIII (Fermentas) in the presence of 1X restriction buffer in a total volume of 60 µL. Digestion was undertaken overnight at 37^{o}C after which the reactions were run on a 1% agarose gel with 0.5X TBE buffer. The gel was run for 1 hour and 45 minutes at 65V. An undigested sample of gDNA (750 ng) was run for comparison along with the three markers i.e. marker II (750 ng), marker XIV (750 ng) and DIG-labelled marker II (200 ng).

3.2.3.3b Southern blotting

Following electrophoresis, the gel was submerged in denaturing solution (0.5 M NaOH, 1.5 M NaCl). The gel was agitated for 10 minutes at room temperature and the denaturing solution replaced. The gel was agitated for a further 10 minutes at room temperature. The denaturing solution was then discarded and replaced with the neutralisation solution (3M NaCl and 0.5M Tris in distilled/deionized water). After 10 minutes agitation, it was

replaced with fresh solution and the gel treated for another 10 minutes at room temperature. The solution was then discarded after 10 minutes.

The gel was placed upside down on a wick of 3MM paper that had been soaked in 20X SSC and placed onto a support. The gel was surrounded with Glad wrap to ensure no short circuiting. A sheet of positively charged nylon membrane (Roche Applied Science) was placed over the gel and two sheets of dry 3MM paper placed on top of that. A series of dry towels followed by a heavy weight was placed above the 3MM paper and the blotting was allowed to proceed overnight. The blotting apparatus was dismantled and the membrane fixed under UV light for 3 minutes.

3.2.3.3c Southern hybridisation

The blot was pre-hybridised in a hybridisation bottle in 10 mL DIG Easy hybridisation solution (Roche Applied Science) at 65^{o}C. The DIG labelled probe (12 µL or 24 µL) was heated for 5 minutes at 95^{o}C and quenched on ice before being added into 6 mL of the hybridisation solution. This was then split between two bottles and the hybridisation allowed to proceed overnight at 65^{o}C.

After hybridisation, excess probe was removed by washing each blot with 100 ml of 2X SSC with 0.1% SDS at room temperature for 5 minutes. This liquid was discarded and replaced with another 100 ml of 2X SSC with 0.1% SDS. The bottles were placed on the rocking table again for 5 minutes at room temperature. The blots were than washed twice in 0.2X SSC/0.1% SDS at either room temperature (low stringency wash) or at 65^{o}C (high stringency wash) for 15 minutes.

The membranes were then washed with 100 mL of 1X washing buffer at room temperature for two minutes with constant shaking. The washing buffer was replaced with 100 mL of 1X blocking solution at room temperature for 30 minutes. The blot was then washed with the antibody solution made up of 20 mL of blocking solution and 4 µL of the antibody solution (Roche, Applied Science). The membranes were incubated at room temperature for 30 minutes before discarding the antibody solution and rinsing the membrane twice with 100 mL of the washing buffer at room temperature for 15 minutes with constant shaking. The membranes were then incubated at room temperature for 3 minutes with 20 mL of 1X detection buffer before being replaced by 10 mL of the colour substrate solution. The

colour solution is made up of 200 µL of the NBT/BCIP stock (Roche Applied Science) to 10 mL of the detection buffer. The membranes were then placed in dark and checked periodically for the development of any bands. The detection reaction was allowed to proceed for an hour after which the reaction was stopped by discarding the colour substrate and rinsing the membrane with 50 mL of TE buffer pH 8.0 for 5 minutes.

3.3 Results

3.3.1 IPCR approach

3.3.1.1gDNA extracted by Aljanabi and Martinez (1997)

3.3.1.1a Extraction of gDNA using the method of Aljanabi and Martinez (1997)

It can be seen from Figure 3.6 that the quantity of gDNA extracted from different samples using the method of Aljanabi and Martinez (1997) was consistent from sample to sample. There was a substantial amount of high molecular weight DNA, however some degradation was observed. This degradation is much more obvious in Figure 3.7. Also, the smear observed at the bottom of the gel was that of RNA and ribosomal RNA bands can be seen below 2322bp.

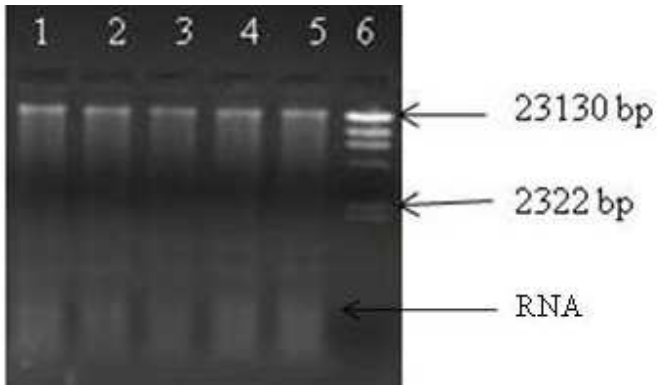


Figure 3.6. The DNA extracted by the Aljanabi and Martinez method. Lanes 1-5: the 5 samples extracted from *N.benthamiana*. Lane 6:1 ug of λ *Hin*dIII markers. Two size markers, 23130 and 2322 bp are indicated on the right.

3.3.1.1b Restriction digestion of gDNA

Figure 3.7 shows the gDNA following digestion with restriction enzymes *Eco*RI (lane 4) and *Hin*dIII (lane 5). It can be seen that compared to the undigested DNA in lane 3, the high molecular weight DNA has disappeared from lanes 4 and 5, suggesting that digestion with each enzyme has occurred.

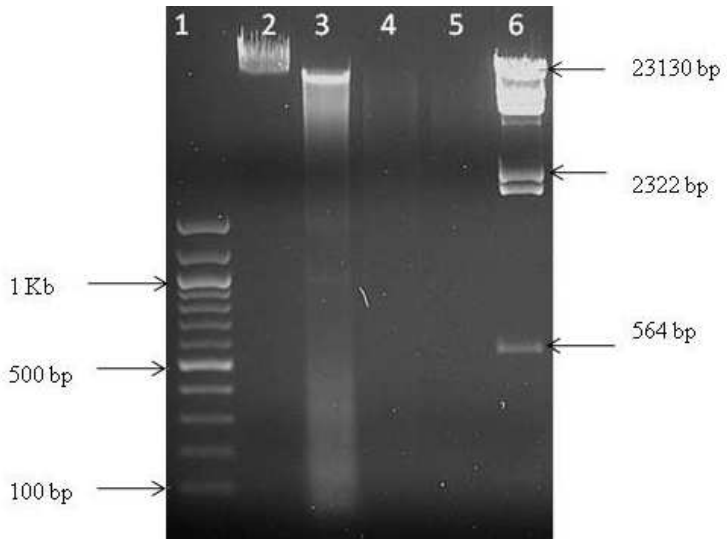


Figure 3.7. Restriction digestion of the Aljanabi and Martinez method of DNA extraction. Lane 1: Maker II; Lane 2: Uncut λ DNA (750ng) Lane 3: Undigested gDNA from *Nicotiana benthamiana* Lane 4: *Eco*RI digested DNA Lane 5: λ *Hin*dIII digested DNA (1 µg) Lane 6: Marker XIV (1 µg). Selected size markers are indicated on the right and left.

3.3.1.1c Testing ligation reaction conditions for IPCR

To ensure the conditions for ligation of the digested gDNA were optimal, test ligations were carried out on λ *Hin*dIII DNA. Figure 3.8 shows the ligation products following ligation of different amounts of λ DNA with 5, 15, or 25 Weiss units of T4 DNA ligase. Not surprisingly, ligation was more likely to occur as the amount of DNA decreased and the amount of ligase increased. Reactions containing 250 or 50 ng of DNA were ligated to a greater extent in the presence of 15 (Figure 3.8 B) and 25 Weiss units of ligase (Figure 3.8 C) than in the presence of 1 unit (Figure 3.8 A). Bands of higher molecular weight can be observed for these samples, whereas for reactions containing higher amounts of DNA, these bands were not observed. From these results, it was decided that ligation of the digested *N.benthamiana* DNA would be best done using no more than 750 ug of gDNA with 15-25 Weiss units of T4 DNA ligase.

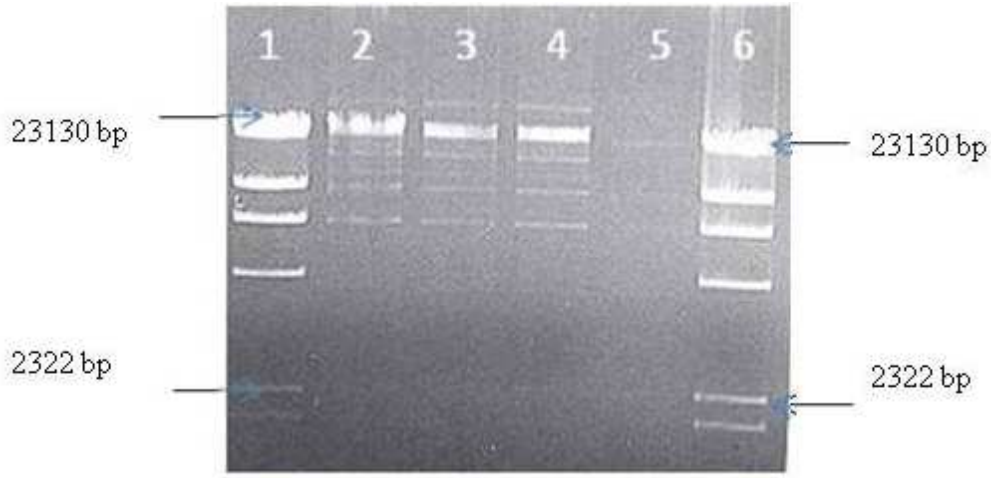


Figure 3.8. A.Ligation of λ *Hin*dIII with 1 µL (5 Weiss units) T4 DNA ligase. Lane 1: 750 ng marker XIV; Lane 2: 750 ng λ *Hin*dIII DNA; Lane 3: 500 ng λ *Hin*dIII DNA; Lane 4: 250 ng λ *Hin*dIII DNA; Lane 5: 50 ng λ *Hin*dIII DNA; Lane 6: 750 ng Marker II. Selected size markers are indicated on the right and left.

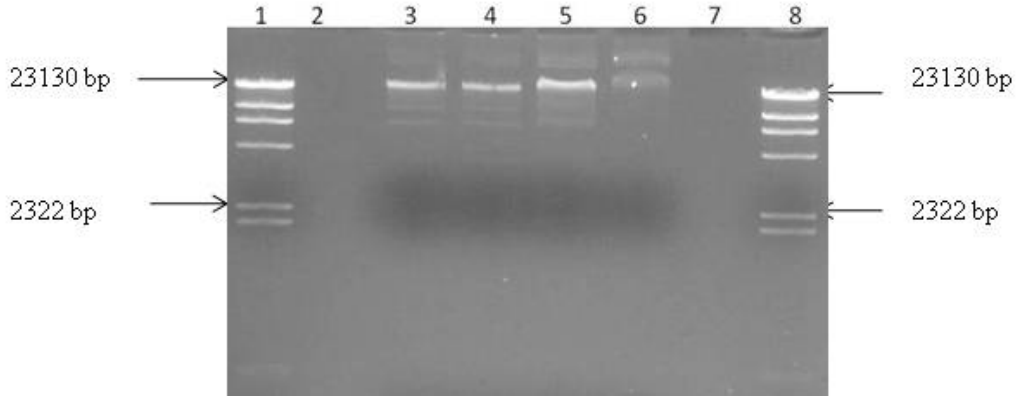


Figure 3.8B.Ligation of λ *Hin*dIII with 3 µL (15 Weiss units) T4 DNA ligase. Lane 1: 750ng marker XIV; Lane 2: no sample; Lane 3: 750 ng λ *Hin*dIII DNA; Lane 4: 500 ng λ *Hin*dIII DNA; Lane 5: 250 ng λ *Hin*dIII DNA; Lane 6: 50 ng λ *Hin*dIII DNA; Lane 7: no sample; Lane 8: 750ng Marker II. Selected size markers are indicated on the right and left.

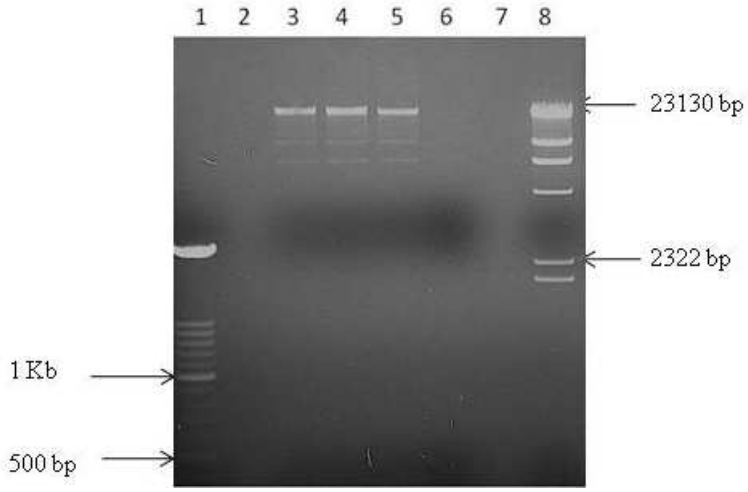


Figure 3.8C. Ligation of λ *Hin*dIII with 5 µL (25 units) of the T4 DNA ligase. Lane 1: 1 µg λ *Hin*dIII marker; lane 2: no sample; Lane 3 is that of 750 ng Lambda Hind III DNA followed by lanes 4, 5 and 6 which have 500 ng, 250 ng and 50ng of the Lambda Hind III DNA respectively; lane 7: no sample; lane 8: 1 µg λ *Hin*dIII marker. Selected size markers are indicated on the right and left.

3.3.1.1d Ligation of digested gDNA

No ligation products were observed following ligation of gDNA extracted using the method of Aljanabi and Martinez. The agarose gel in Figure 3.9 contained 10 uL from a total of 50 uL of ligation reaction. Given the DNA was diluted prior to ligation, the lack of reaction products was likely due to the concentration of the gDNA in the reaction being too low to see. Since the ligation contained less DNA than the test ligations of λ *Hind*III DNA with 15 Weiss unit of ligase, it was assumed that ligation had occurred.

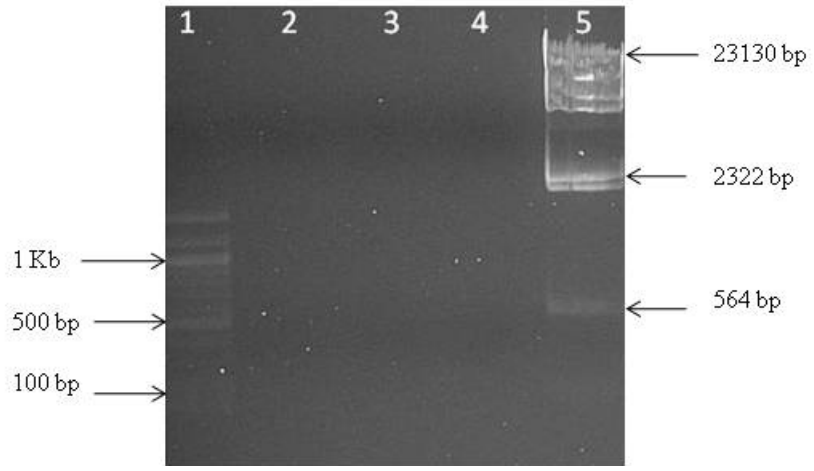


Figure 3.9.Ligation of the DNA products following digestion. Lane1: 750 ng of marker II; Lane 5: marker XIV; Lane 2: 10 µL of *Eco*RI digested and ligated DNA; lane3:10 µL λ *Hin*dIII digested and ligated DNA; Lane4: no sample. Selected size markers are indicated on the right and left.

3.3.1.1e Inverse PCR

Inverse PCR was carried out in a total volume of 25 uL from 25 ng of the ligation reaction using the primers IPCR 2 and IPCR4. Figure 3.10 shows that no PCR product was observed for either DNA digested with *Eco*RI or *Hin*dIII and subsequently ligated. Thus, the 5' flanking region was not amplified using gDNA extracted using the Aljanabi and Martinez method.

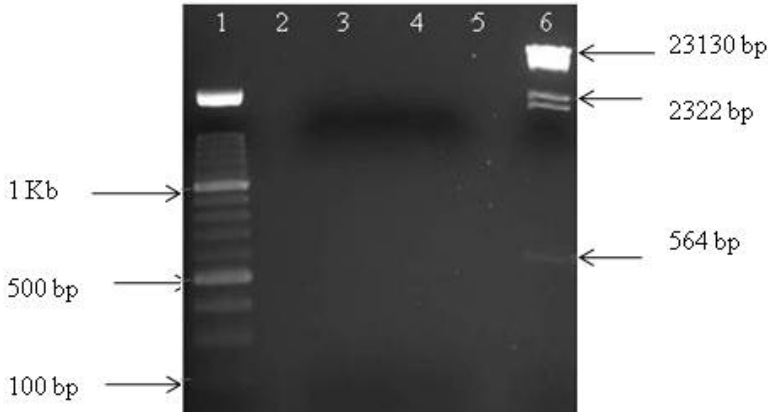


Figure 3.10.Gel results of IPCR on the Aljanabi and Martinez extracted DNA. IPCR products run on the gel against a no template control (NTC) with the two markers i.e. Lanes 1 and 6 : 750 ng of Marker XIV and Marker II respectively. Lane 2 and 5: no sample; lane 3: EcoRI digested and ligated DNA; 4: λ HindIII digested and ligated DNA.

3.3.1.2 InVitrogen extracted DNA

3.3.1.2a gDNA extraction

gDNA was extracted from *N.benthamiana* using the InVitrogen Purelink genomic DNA extraction kit. The extraction procedure was quite easy to carry out; however, the quality of the DNA was poor. Figure 3.11 shows the gDNA extracted from two leaf samples appeared as smears on an agarose gel, suggesting they were somewhat degraded. This gDNA did not appear to contain RNA as there was no smaller molecular weight nucleic acid present on the gel.

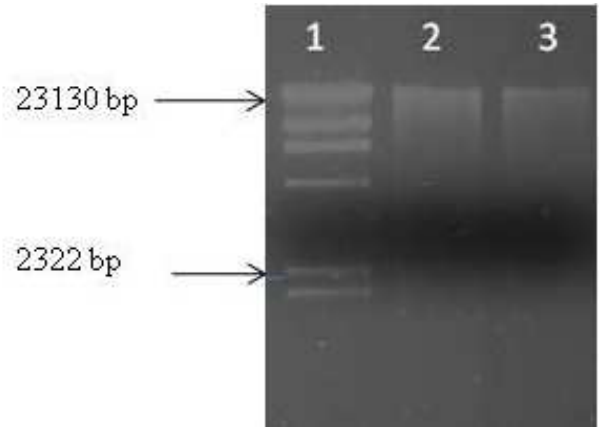


Figure 3.11.DNA extracted by the InVitrogen Plant Extraction DNA kit. Lane 1: Marker II Lane 2 and Lane 3: *N.benthamiana* gDNA. Selected size markers are indicated on the left.

3.3.1.2b Restriction digestion

The undigested sample in lane 1 of Figure 3.12 had a high molecular weight band of approximately 23,130 bp which is not visible in lanes 2 and 3. This indicates that the process of digestion has indeed taken place. However, that starting material was degraded prior to digestion.

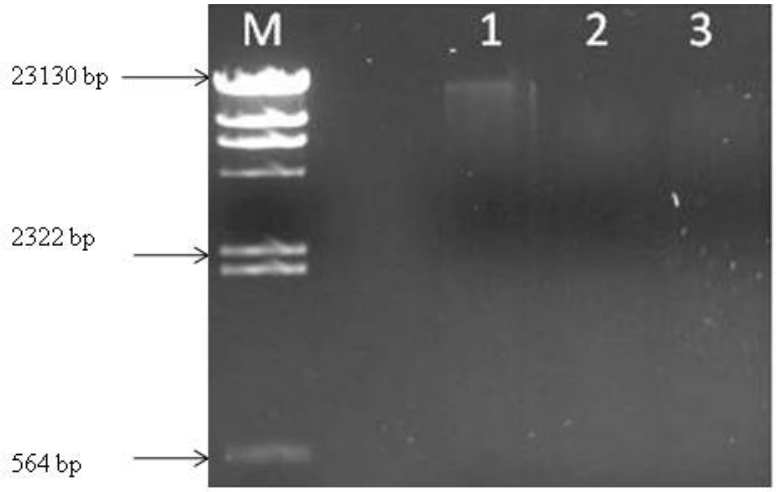


Figure 3.12. Digestion of the InVitrogen extracted gDNA from *Nicotiana benthamiana*. Lanes 2 and 3 when compared against the undigested DNA on lane 1 indicate the efficiency of the restriction digestion. Selected size markers are indicated on the left.

3.3.1.2c Ligation

As for the ligation reaction described above, the amount of DNA on the gel was very low as the DNA was diluted to allow self-ligation to occur. This made it difficult to see the ligation products (data not shown). Again, given that the amount of DNA was within the acceptable range as determined before by the test ligations of λ *Hin*dIII, it was assumed that ligation had occurred.

3.3.1.2d Inverse PCR

IPCR using the InVitrogen extracted gDNA as the template with primers IPCR2 and IPCR4 resulted in a smear (Figure 3.13). A smear as a product from an IPCR may indicate a high concentration of the template. However, this smear was observed irrespective of the amount of template included. The amounts of template used varied from 3 µL to 5 µL i.e. 10 ng to 50 ng.

Nested PCR has been suggested as an alternative for amplification using an IPCR approach. Given that a smear was observed with IPCR primers 2 and 4, it was felt that a PCR with primers IPCR 2 and 3, followed by nested PCR with primers IPCR 1 and IPCR 4 might result in a discrete band. However, this was not the case (data not shown). Amplification

with IPCR 2 and IPCR 3 did not result in any visible products at all, nor did nested amplification with IPCR 1 and IPCR 4.

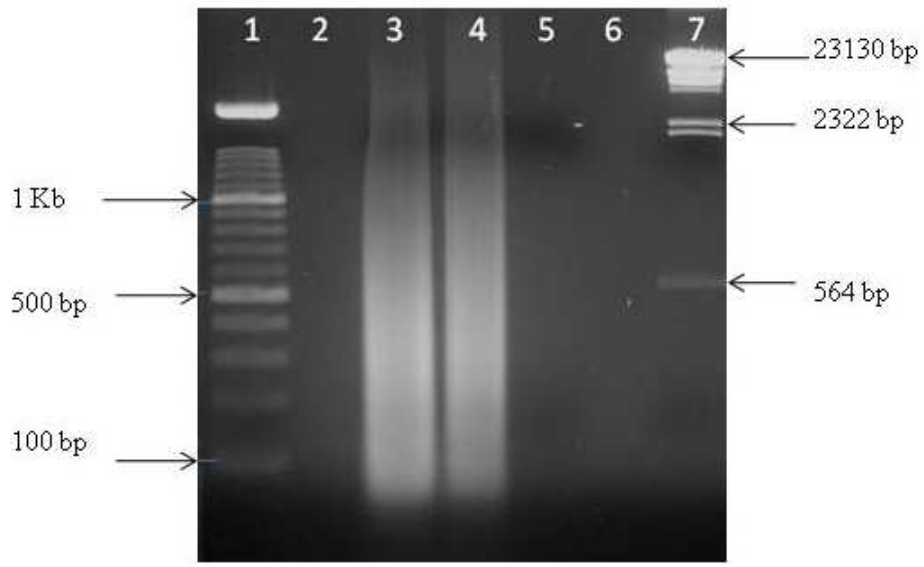


Figure 3.13. Gel result of an IPCR with the DNA extracted using the InVitrogen Purelink DNA extraction kit. Lane 1: 1µg Marker XIV; Lane 2: no sample; Lane3: *Eco*RI digested and ligated PCR product; Lane4: λ *Hin*dIII digested and ligated PCR product; Lane 5: no template control; Lane 6: no sample; Lane 7: 1 µg Marker II. Selected size markers are indicated on the right and left.

3.3.1.3 gDNA extracted by the Dellaporta et al (1983) method

3.3.1.3a gDNA extraction

Of all the three extractions, this technique gave the best quality gDNA (Figure 3.14). A discrete high molecular weight band could be seen with no DNA smear however, a lower molecular weight smear was observed which was likely to be RNA.

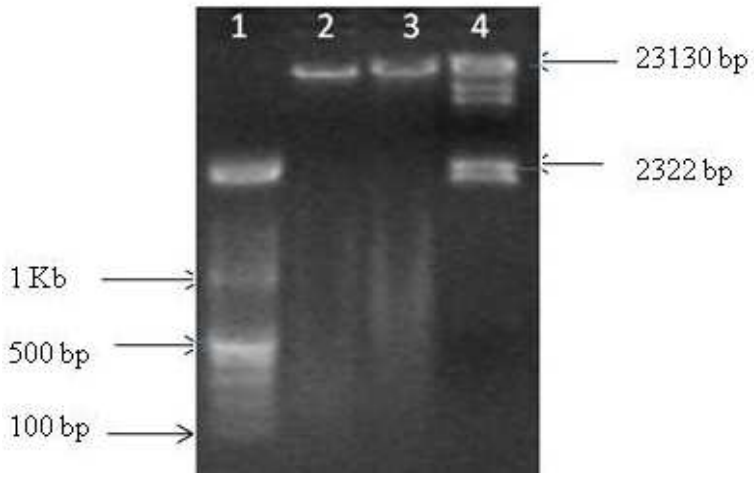


Figure 3.14. The gDNA extracted by the Dellaporta et al (1983). The DNA appears to be of a high molecular weight and of a good quality with low RNA concentration (the smears present in each lane). Lane1: 750 ng of Marker XIV; Lane 2 and 3; extracted gDNA; lane5: 1 µg of Marker II. Selected markers are indicated on the left and right.

The DNA samples treated with the RNase enzyme were run on a gel as well. However, the DNA degraded on treatment. Hence, the RNase enzyme must have been contaminated with DNase as well. This was believed because inspite of the enzyme being treated at 92 ^{o}C the second time to deactivate the DNase, the DNA was degraded once again indicating the DNase activity.

3.3.1.3b Restriction digestion

The high molecular weight band observed in the uncut gDNA sample seems to have disappeared from the digested samples (Figure 3.15, lanes 2 and 3 compared with lane 1). Also, a faint smear of lower molecular weight was observed indicating that the gDNA may have been digested with *Eco*RI and *Hin*dIII.

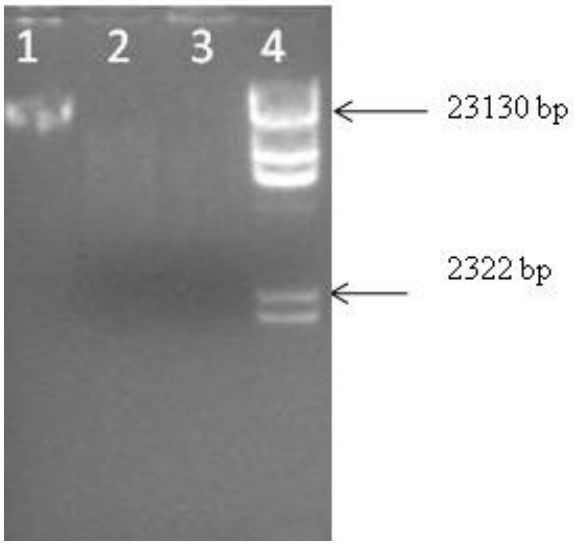


Figure 3.15. Digestion of the Dellaprta et al extracted DNA. The digested gDNA on lanes 2 and 3 when compared with the undigested and intact DNA on lane 1 indicates the efficiency of restriction digestion by *Eco*RI and *Hin*dIII. Lane 1 is a high molecular weight DNA fragment while lanes 2 and 3 are light, smeary and have no intact band. Lane4: 1 µg Marker II. Selected markers are indicated on the right.

3.3.1.3c Ligation

As for the other ligation reactions using the other gDNA material, no band or smear was observed (data not shown). This was again assumed to be due to a low concentration of ligated products; however, the parameters of the ligation were such that ligation was expected to have occurred.

3.3.1.3d Inverse PCR

IPCR with the Dellaporta et al (1983) extracted gDNA and primers IPCR 2 and IPCR 4 did not result in a PCR product (Figure 3.16).

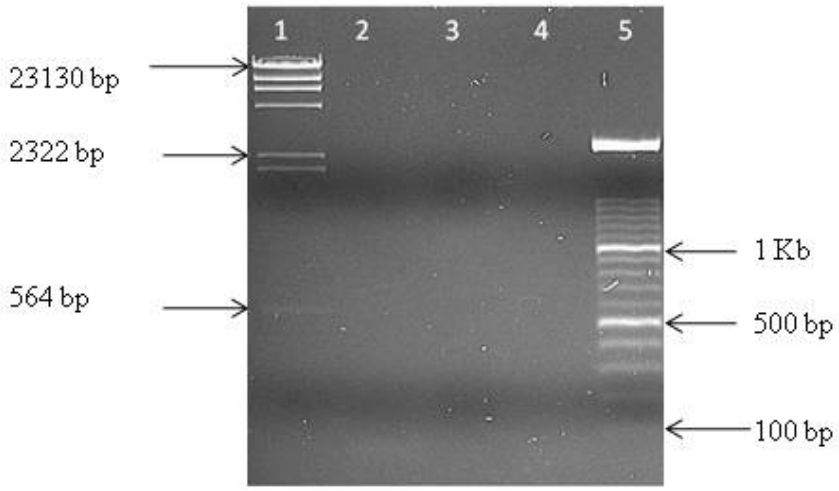


Figure 3.16. IPCR result for the DNA extratcted by Dellaporta et al (1983). Gel with the IPCR amplified products loaded on the gel with the markers II and XIV on lanes 1 and 5 respectively (750 ng). However no products were seen to be amplified. Lane3: *Eco*RI digested and ligated PCR product; Lane4: λ *Hin*dIII digested and ligated PCR product; Lane5: no template control. Selected markers are indicated on the left and right.

3.3.2 Southern Hybridisation

3.3.2.1 Labelling of probes

Figure 3.17 lane 2 shows a PCR product of just greater than 500bp which corresponded to the DIG-labelled PVIP sequence. This band was slightly greater in size than the unlabelled 500bp product in lane 3. This size difference indicated that the probe had labelled well and the probe was suitable to use in a hybridisation experiment. Control reactions showed that where no reverse transcriptase was added (Figure 3.17, lane 5).

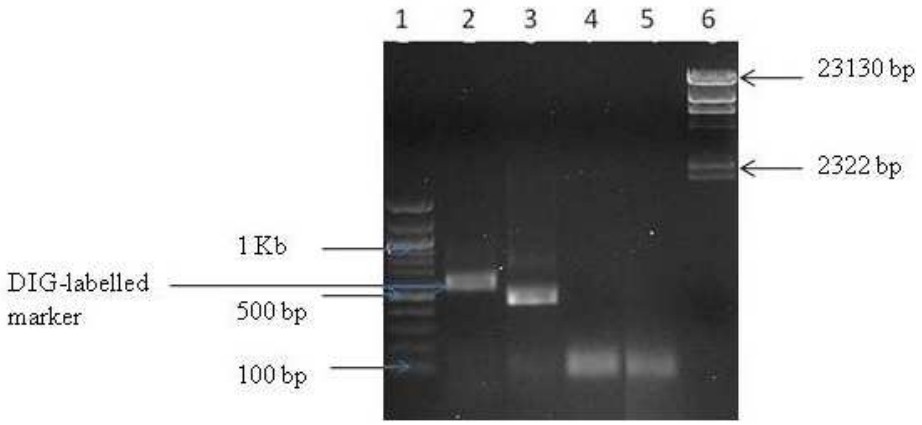
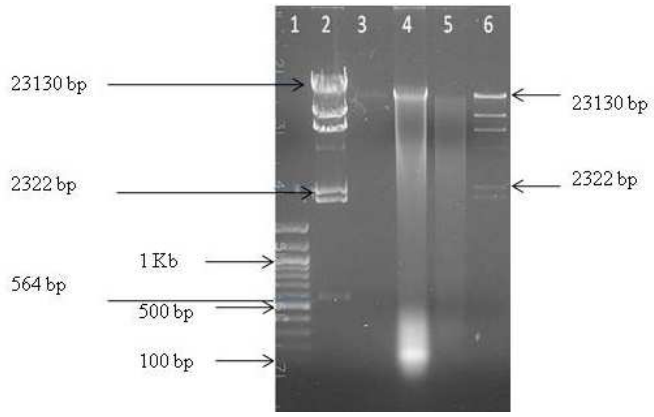


Figure 3.17. PVIP probe synthesis gel result. Lane 1 and 6 are that of the markers while lane 2 is the PVIP DIG-labelled probe. Lane 3 is the PVIP DNA sequence without the DIG label attached to it; Lane 4 is the cDNA without the RT enzyme whereas lanes 5 is the control reactions with the DIG label. Selected markers are indicated on the left and right.

3.3.3.2 Southern blotting and hybridisation to Aljanabi and Martinez extracted DNA

Approximately 10 µg of digested *N.benthamiana* gDNA was electrophoresed (Figure 3.18) and transferred to a nylon membrane. The gDNA appeared to be digested to completion with *Eco*RI and *Hin*dIII (Figure 3.18 A and B, lanes 4 and 5) compared with the undigested DNA (lane 3). However, it is likely there is a background of degraded DNA as the gDNA extracted using this method had some degradation (see Figure 3.5 and 3.6). Thus, the DNA appeared to have been digested and the gels were blotted prior to hybridisation.

A



B

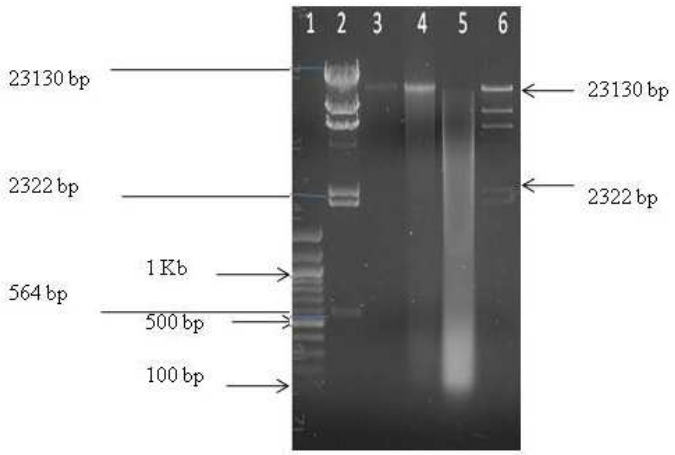


Figure 3.18. Digestion of gDNA as extracted by the Aljanabi and Martinez method prior to Southern blotting. A) The gel prepared for low stringency and (B) high stringency for Southern hybridisation. Lane 1: marker XIV (750 ng) lane2 : marker II (750 ng) Lane 3: Undigested gDNA (1µg) lane 4: *EcoRI* digested DNA lane 5: *HindIII* digested DNA lane 6: DIG labelled marker II (1µg). Selected markers are indicated on the left and right.

Southern hybridisation was carried out to determine the sizes of restriction fragments carrying a portion of the PVIP mRNA attached to its 5' flanking region. Of the two gels that were blotted and hybridised, one was treated with a low stringency post-hybridisation wash to allow detection of somewhat similar DNA fragments, while the other was treated with a high stringency post-hybridisation wash to detect highly similar sequences. Detection of the hybridised probe following the post-hybridisation washes was the same for

both blots. Figure 3.19 shows that neither low nor high stringency post-hybridisation washes gave rise to any detectable fragments.

The detection of the DIG label clearly worked as the DIG-labelled markers could be easily seen. Thus, it could have been a problem with the hybridisation itself or with the DNA transfer to the membrane. It was unlikely to be the latter as the gel was stained with ethidium bromide after blotting to ensure the DNA had moved from the gel. Thus, it was likely to be a problem with the hybridisation. This could have been due to there being too little gDNA on the gel or there may not have been enough probe added to the hybridisation. Non-radioactive detection of DIG using colorimeteric detection is known to be insensitive (Beck, O'Keeffe, Coull & Koster, 1989); either of these possibilities could have been the reason why no detection was observed. Therefore, it was decided that future Southern hybridisation experiments should use increased amounts of high quality gDNA and also test using an increased concentration of the probe.

A B

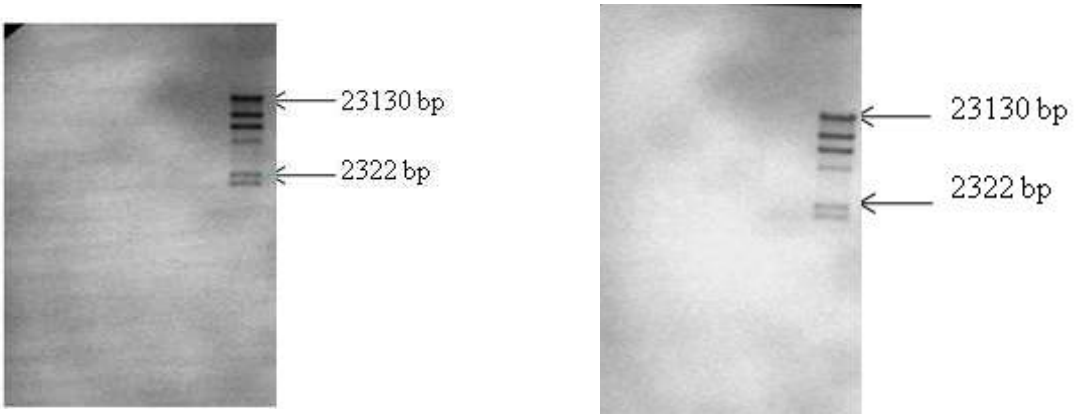
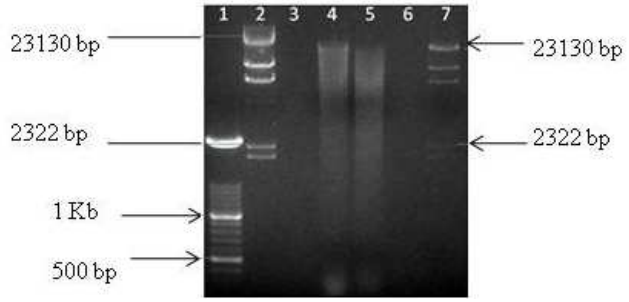


Figure 3.19.The positively charged Nylon membranes after the first Southern hybridisation experiment using the Aljanabi and Martinez extracted DNA. Nothing but the DIG-labelled marker II can be observed in the membrane. The membrane on the left is the result of gel1 and the membrane on the right is the result of gel2 i.e. the gels obtained after digestion. Selected markers are indicated on the right.

3.3.3.3 Southern blotting and hybridisation to Dellaporta extracted DNA

Based on the result using DNA extracted by the Aljanabi and Martinez method, it was decided that the experiment should be repeated with more, better quality gDNA. Thus, Southern blots were prepared using 20µg (Figure 3.20) or 50µg (data not shown) of digested gDNA that had been extracted using the Dellaporta method. The gDNA appeared to be digested to completion with both *Eco*RI and *Hin*dIII (Figure 3.21 A and B, lanes 4 and 5). The quality of the starting gDNA was superior to the previous method as the majority of the DNA was of high molecular weight. This meant that any fragment detected by Southern hybridisation would have been generated by digestion and not by DNA degradation. Too little undigested DNA was loaded onto the gels for comparison, however based on Figure 3.14, it was expected the DNA was of high molecular weight.

A)



B)

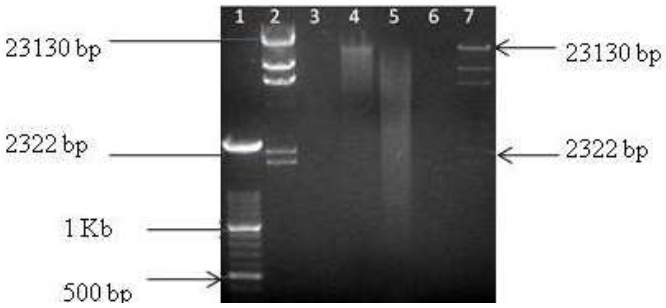


Figure 3.20. The digested product of 20µg of the Dellaporta extracted DNA prior to Southern hybridisation. Lanes 4 and 5 are the digested DNA compared against the 3 markers i.e. marker XIV (lane1), II (lane 2) both of which were 750 ng, and DIG-labelled marker II (200ng). Selected markers are indicated on the left and right.

A) B)

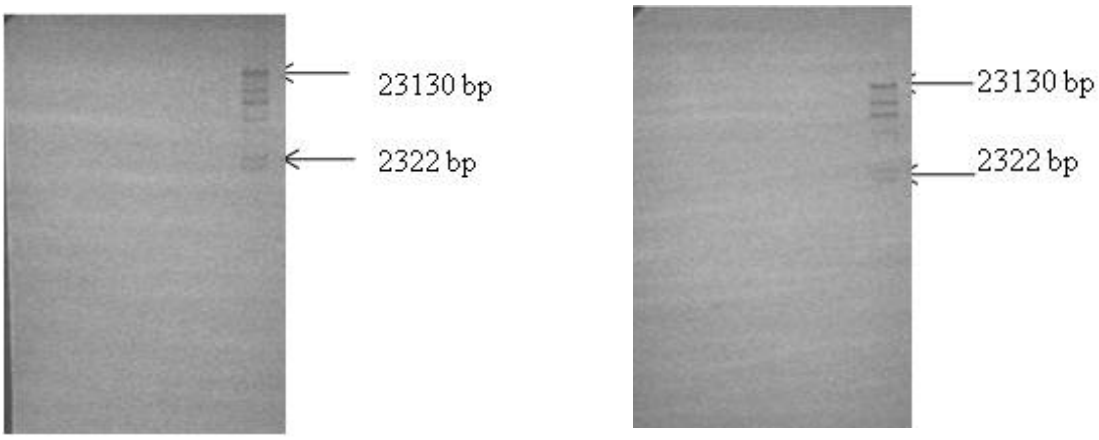


Figure 3.21. The positively charged Nylon membranes after the second Southern hybridisation experiment using the Dellaporta extraction technique. Nothing but the DIG-labelled marker II can be observed in the membrane. The membrane A is the result of gel1 and the membrane B is the result of gel2 i.e. the gels obtained after digestion. Selected markers are indicated on the right.

Hybridisation was carried out on 20 µg gDNA using an increased amount of probe as compared to that of Aljanabi and Martinez extracted DNA i.e. increase from 2 µL per mL to 4 µL per mL of the hybridisation solution. The blots were washed at either low or high stringency following hybridisation. No hybridising fragments were observed for either (Figure 3.21 A and B). Hence, increasing the probe concentration to 4 uL per mL of the hybridisation solution made no difference, as no fragments could be detected. Increasing the amount of gDNA to 50ug in each lane also did not allow detection of any fragments. Thus, the PVIP flanking region could not be detected using these conditions for Southern hybridisation.

3.4 Discussion

Knowledge of promoter structure is essential in order to understand gene regulation and potential promoter function. Hence, isolation of the 5' flanking region from the *N.benthamiana* PVIP gene and assessment of its promoter activity would have had the potential to provide clues of the function of the PVIP protein.

In this study, IPCR was attempted to isolate the 5' flaking region from the PVIP gene of *N.benthamiana*. Isolation using this technique would have allowed the fragment to be cloned and functional analyses to be carried out. Whether or not this flanking region fragment has promoter activity could be tested by joining it to a reporter gene and testing the expression of the reporter in transgenic plant systems. The sequence could be mutagenised to identify important functional regions contained within the fragment.

All attempts to isolate the 5' flanking region by IPCR were unsuccessful. It is important for this technique to start with high molecular weight DNA. Degradation of the DNA can result in fragments that cannot be amplified by PCR. The concentration of the gDNA can also be an issue, less is better. However, it was difficult to accurately measure the DNA concentration as RNA was co-extracted in two of the methods- the Aljanabi & Martinez method and the Dellaporta method. Even though no RNA appeared to be co-extracted with the InVitrogen extracted DNA, the DNA was degraded. Attempts to remove the RNA resulted in the gDNA being degraded further, suggesting the RNAse, while sold as not having DNAse contamination, appeared to have this activity as heating at 95^{o}C in order to deactivate the DNase degraded the DNA.

For the PCR, the template concentration was varied, from 10 ng to 500 ng. However, it would have been ideal to start with a lower template concentration and gradually increase until the IPCR resulted in a product. The primer concentration appeared appropriate as no primer-dimers were observed in any of the IPCR reactions. Also, adding 1.25µL of the 10µM primers should have been adequate based on the fact that most of the PCRs are carried out at this or similar concentrations. Anything lower may reduce the reaction efficiency. However, one could test a range of primer concentrations as well.

The PCR cycling was tested using a gradient PCR for a wide range of annealing temperatures from 51.6^{0}C to 59.6^{0}C. However, there may be a possibility that the PCR

conditions might not have been optimal for the primer sequences. The primer sequences were designed using OligoExplorer based on the *N. benthamiana* PVIP mRNA sequence. However, we could not be absolutely sure about the relative success of the primers as we could not test the primers due to their inverse nature and the lack of a positive control. A series of primer combinations were used such as IPCR 1 with IPCR primers 3 and 4 and also IPCR primer 2 with IPCR primers 3 and 4. However, none of them yielded any product. The technique of IPCR involves a number of steps i.e. that of digestion, ligation and purification and hence IPCR is a technically challenging process. Any one of these steps could be sub-optimal and interfere with the success of the process.

Knowing the size of the region containing promoter activity would allow gDNA of this size to be gel purified, cloned into a vector to create a sub-genomic library which could be screened for the desired fragment. This fragment could then be isolated and used as described above to test for promoter activity. Southern hybridisation was used to try to identify a fragment/s that contained the 5'flanking region, however no fragments could be detected irrespective of the amount of gDNA analysed or the amount of probe used. The digestion conditions appeared appropriate as the agarose gels run after each digestion showed the expected smear of DNA. The quality of the starting DNA was best using the Dellaporta extraction method; therefore any fragments that would have been detected using this method would have been likely to be restriction fragments and not degradation products.

Increasing the amount of gDNA assayed and the probe concentration, did not improve the detection. DIG-labelled probes are not as sensitive as radio-actively labelled probes and colorimetric detection is not as sensitive as chemiluminescent detection. It appears that the detection of the PVIP sequence is beyond the limit of colorimetric detection; however, neither radioactive nor chemiluminescence detection could be carried out at AUT. If this work is to continue, then this limitation would need to be overcome.

Since the promoter region could not be isolated using the techniques employed, we could not compare the flanking sequences of *Nicotiana benthamiana* with those from *Arabaidopsis thaliana* and *Oryza sativa*. Hence, we could not distinguish the size and structure of the PVIP promoter and also could not predict the transcription factors that

might bind to the promoter of PVIP within *N.benthamiana* based on the flanking regions actual sequence.

Promoter analysis i.e. comparing the sequence information of the 5' flanking region from *N.benthamiana* with the promoters of *A.thaliana* and *O.sativa* could have determined the probable information about the function of the PVIP gene and also the TF's the promoter binds to which would in turn help us understand the different stresses the promoter responds to. The mRNA quantification by real-time PCR will help us obtain information about the genes environmental responsiveness and also check some of the predictions made in chapter 2.

Chapter 4

Relative quantification of the PVIP gene.

4.1 Introduction

Gene expression is used by all life forms to generate the macromolecular machinery of life which may be proteins or even non-protein coding genes such as rRNA or tRNA. Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. Gene expression quantifies the level at which a particular gene is expressed within a cell, tissue or an organism. Gene expression may depend upon the stresses the eukaryotes are going through. Sensing the stresses and making amendments in order to adjust with these stresses activates a cascade of transcription factors which in turn regulate the gene expression, thereby helping in the protection and repair of the plant. If successful in overcoming stresses, these transcription factors may result in the adaptation and also tolerance of the organism, thereby creating a new genetic code for the organism. It is this genetic code of a gene which defines the level of gene expression, and the properties of the expression products gives rise to the organism's phenotype (Rensink et al., 2005).

4.1.1 Methods for measuring mRNA levels

4.1.1.1 Northern hybridisation

Gene expression can be measured at the RNA and/or protein level (if protein is the final product of the gene). Methods used for the quantification of gene expression at the mRNA level include real-time PCR (qPCR), northern hybridization and microarrays. Northern analysis remains a standard method for detection and quantification of mRNA levels despite the advent of powerful techniques, such as RT-PCR, gene array analysis and nuclease protection assays. Northern hybridisation requires the extraction of RNA from selected tissue which is then separated on a denaturing gel. The gel is blotted onto a membrane and hybridised with a labelled probe as described for Southern blotting in

chapter 3. Northern analysis can provide a direct relative comparison of message abundance between samples on a single membrane if measured against the amount of RNA loaded. It is the preferred method for determining transcript size and for detecting alternatively spliced transcripts. However, like other techniques there are limitations associated with northern analysis as well. The first limitation being that, northern analysis is not a quantitative method of RNA determination and only gives an indication of the relative amounts of RNA. Secondly, if the RNA samples are even slightly degraded, the quality of the data and the ability to quantitate the expression may be severely compromised. Northern hybridization is a far less sensitive technique than nuclease protection assays and also RT-PCR. However, improvements have been made in this field. An increased sensitivity can be achieved by using highly specific activity antisense RNA probes, optimized hybridization buffers and positively charged nylon membranes. Although, even with each of these optimised, it is only considered semi-quantitative. The third limitation of northern blotting has been the difficulty associated with multiple probe analysis. To detect more than one probe, it is imperitive to unhybridize (removal of probes from blots) the initial probe before hybridizing with a second probe. This process can be time consuming and problematic, since harsh treatment is required to unhybridize the probe ("The Basics: Northern Analysis," 2010).

4.1.1.2 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

Polymerase chain reaction (PCR) involves the production of a large number of copies of a specific DNA sequence without having to resort to the process of cloning. The starting material of PCR is a double stranded DNA and a pair of oligonucleotide primers flanking the DNA sequence. An RNA template is amplified using reverse transcriptase polymerase chain reaction (RT-PCR). The RNA must first be reverse transcribed into DNA by reverse transcriptase and then the specific sequence amplified by PCR. The PCR steps include the denaturation of the DNA at 94-95^{0}C followed by annealing of the primers to the template DNA. The last step involves the extension of the DNA strand using the enzyme *Taq* polymerase at 70-75^{0}C. At each of these cycles, a two fold increase in the amount of DNA is observed. A typical PCR reaction involves usually around 25-35 cycles and the number of copies at the end of the cycle can estimated by the formula 2^n, where n represents the number of cycles (Russell, 2001). The amplification products are examined at the end of

the reaction on an agarose gel. This is known as end point RT-PCR. The limitation of RT-PCR may be that the quantitation might not be accurate since ethidium bromide reliance is insensitive and also when the band is detected, the exponential stage of amplification is over. Hence, RT-PCR is sometimes known as “semi-quantitative” PCR.

A more modern (quantitative and sensitive) approach for measuring mRNA abundance is reverse transcription coupled with quantitative analysis i.e. RT-qPCR.

The documentation of real time PCR (qPCR) had been first recorded in the year 1993, but it is just recently i.e. post 2000 that the technology has gained popularity (Ginzinger, 2002) and has hence changed the way the analysis of gene expression is made (Wong & Medrano, 2005). Real-time PCR functions by collecting the data throughout the PCR process as it occurs, thereby combining both, amplification and detection within a single step (Wong & Medrano, 2005). Real-time PCR has a number of advantages over certain conventional techniques. Sensitivity, large dynamic range and the potential for accurate quantification are just to name a few (Huggett, Dheda, Bustin & Zumla, 2005). However, there are certain disadvantages as well such as the high cost of the machines involved (Ginzinger, 2002) or the inherent variability of the RNA being used or the different reverse transcription and PCR efficiencies (Bustin & Nolan, 2004). Q-PCR can be used for a number of applications such as quantifying mRNA i.e. the mRNA expression studies, measurement of the DNA copy number, allelic discrimination assays and also the confirmation of microarray data (Ginzinger, 2002).

Real time PCR is a dependable detection and measurement technique of the products generated during each step of the PCR reaction. These products are directly proportional to the amount of template used initially i.e. at the start of the PCR reaction. Real time PCR involves four main steps or phases i.e. linear ground phase, exponential phase, log-linear phase and the plateau phase. The linear phase is just the beginning within which the fluorescence emission has not reached the peak i.e. above the background. This usually takes place within the first 10-15 cycles. Within the exponential phase the fluorescence reaches a threshold value i.e. the value greater than the background values. The cycle at which the fluorescence is greater than the background level is known as the threshold cycle (Ct) and the fluorescence level indicates the copy number in the original template and is hence used to calculate the amount of template. The optimal level of the PCR is reached

within the log-linear phase in which the PCR product doubles with every reaction cycle under ideal conditions. Within the plateau stage the reaction components become limited and hence the fluorescence intensity is of not much use for data analysis (Wong & Medrano, 2005). The most commonly used chemistries employed for real time PCR includes the use of TaqMan probes, SYBR Green I intercalating dyes and molecular beacons amongst many more (Ginzinger, 2002).

For mRNA quantification, qPCR can be used either as a one-step or a two-step reaction. A one step RT-qPCR involves both the enzymatic reactions within a single tube and hence involves the direct use of RNA, which is prone to degradation. However, the chances of contamination are reduced as well. A two-step process involves the generation of cDNA and then the DNA product. The cDNA can be stored relatively easily and can be subjected to several different real time PCR assays on dilutions of a single cDNA (Wong & Medrano, 2005).

mRNA can be quantified by RT-qPCR either relatively or by absolute quantification. During relative quantification, an external reference or a house keeping gene is required against which the variation of the target gene expression can be measured. This reference gene is known as the calibrator and the results are expressed as a target/reference ratio. For absolute quantification, serially diluted standards of a known concentration are used to generate a standard curve. The standard curve thus generated must produce a linear relationship between the Ct value and the initial amounts of template used (either cDNA/RNA) (Wong & Medrano, 2005).

Absolute quantification when compared with relative quantification is a lot more labour intensive. This is because, in the case of absolute quantification, it is essential to develop a reliable standard and thereby include this standard in every PCR henceforth. In case of relative quantification however, the standard values we develop are only arbitrary and are hence applicable only to the samples run within the same PCR reaction (Wong & Medrano, 2005).

The efficiency and sensitivity of PCR largely depends upon the efficiency of the primers, quality of the DNA/RNA and also the accuracy of pipetting. The primers which are unique for the target sequence to be amplified should fulfill certain criteria such as primer length,

percent GC, annealing and melting temperature, 5 ‘end stability, and 3' end specificity (Singh, Govindarajan, Naik & Kumar, 2000).

The PCR efficiencies of all the genes present in the reaction must be similar and preferably at or above 90% for relative quantification. This can be analysed by plotting a standard-curve of the diluted samples. The cDNA concentration may be diluted in a 10 fold series and the $\log_{[10]}$ concentration is plotted against Ct. The resulting slope (M) which is derived from the straight line obtained through the standard curve must be -3.32 for a 100% efficiency reaction. The standard curve method is best applicable for the relative quantification of a small number of genes within a few or many samples (Ginzinger, 2002). The relative quantification of mRNA can be calculated using the ΔΔCt method. This technique involves the use of a single positive sample termed as the ‘calibrator’ against which the unknown genes are compared. The formula that is used is

$$\text{Fold induction} = 2^{-[\Delta\Delta Ct]}$$

Where ΔΔCt= [Ct (gene of interest)-Ct (reference gene)] – [Ct (gene of interest of calibrator sample)-(reference gene of calibrator)].

However, this formula is based on the assumption that the rate of Ct change and the target copy number is the same for the gene of interest and reference gene and the copy number increases by a factor of two per cycle. These assumptions may hardly ever be exact but increasing the number of replicates analysed may increase the quantitative accuracy by the standard curve method (Winer, Jung, Shackel & Williams, 1999). The reaction efficiency can be calculated using the formula $10^{(-1/slope)}$-1, where the slope is the value obtained after the standard curve experiments, that is M.

One of the problems of qPCR is also that of normalisation and several theories have been proposed for the normalisation of the RT-qPCR data namely from using a similar sample size to using an internal housekeeping or a reference gene. Normalising the real time data using a reference gene is a simple and a popular technique for internally controlling any error within the RT-qPCR reaction. This is because both, the target gene and the reference gene will be amplified within the same reaction and hence under the same reaction conditions (Ståhlberg, Kubista & Pfaffl, 2004). The ideal reference genes must essentially be transcribed in all the cells and tissues and not vary with the treatment being used. Also,

the reference genes must be subjected to all the steps of the PCR reaction in an identical manner to that of the target gene (Radonic et al., 2004) The commonly used reference genes include the 18S rRNA, actin, ubiquitin, glyceraldehyde-3-phosphate dehydrogenase (GADPH) and hypoxanthine-guanine phosphoribosyl transferase (HPRT) (Lovdal & Lillo, 2009). It is desirable to perform an experiment with more than one reference gene in order to obtain a reliable result in the gene transcription analysis (Radonic et al., 2004). However, some of the well known reference genes such as glyceraldehyde-3-phosphate-dehydrogenase (GADPH) and β-actin have been found to be affected by different treatments, biological processes and even different tissues and cell types. Also, within the family of rRNA i.e. 18S and 28S, 28S is considered to be more representative of the mRNA integrity because 18S may remain intact even in samples with degraded mRNA. The rRNAs are transcribed with a different polymerase enzyme than the mRNAs, so any change in either of the polymerase activity may not affect both types of RNA expression equally (Wong & Medrano, 2005) The use of multiple reference genes may be impractical for high throughput applications because of the increase in sample number, cost and setup (Phillips, D'Auria, Luck & Gershenzon, 2009).

A similar experiment with RT-PCR and end-point PCR can be performed as well which may provide information about the product being of the desired size or not after running it on an agarose gel. However, it will not be able to distinguish the different amounts of mRNA in the treated plants and the different tissues of the same plant. This is where real-time PCR becomes a valuable technique.

4.1.1.3 Micorarrays

Northern blots and RT-qPCR are good for detecting whether a single gene is being expressed, but these quickly become impractical if many genes within the sample are being studied. Using microarrays, the expression profile for many genes can be measured at once. Recent advances in microarray technology allow for the quantification, on a single array, of transcript levels for every known gene in several organisms' genomes (Wheelan, Murillo & Boeke, 2008).

Using the relative quantification method of RT-qPCR we can determine the relative abundance of the PVIP mRNA as well. By quantifying the gene of interest in different

tissues and under different environmental conditions we can determine in which tissue and under which environmental conditions it is expressed to a greater degree, which in turn may provide some information about the possible function of PVIP and how it might be involved in potyviral infection.

4.2 Methods and Materials

4.2.1 Sub-culturing *Nicotiana benthamiana*

Refer section 3.2.1

4.2.2 RNA extraction

Different tissues from *N. benthamiana* i.e. stem, leaf, roots and flowers were subjected to RNA extraction using the Spectrum Plant Total RNA extraction kit from Sigma Aldrich according to the manufactuer's instructions. Approximately 100 mg of each of the plant tissues were used for extraction. These tissues were ground in a sterile mortar and pestle using liquid nitrogen. Approximately 500 µL of the lysis buffer with beta-mercaptoethanol was added to the 100 mg of ground tissue and vortexed vigorously for 30 seconds and then incubated at 56^{o}C for 5 minutes. The reaction mixture was then centrifuged at high speed i.e. 13,000 rpm for 3 minutes. The supernatant of the lysate was added into a filtration column i.e. with a blue retainer ring which was attached to a 2 mL collection tube and centrifuged at 13,000 rpm for 1 minute and the clarified flow-through lysate was preserved. Five hundred µL of the binding buffer was added to this clarified lysate and mixed thoroughly. Seven hundred µL of this mixture was added into the binding column with the red retainer ring which in turn was attached to the 2 ml collection tube. The tube was then centrifuged at 13,000 rpm for 1 minute. The flow through was discarded and the remaining mixture of the lysate and the binding buffer was added into the retainer ring and centrifuged once again for 1 minute at 13,000 rpm.

Into the binding column 300 µL of the wash solution 1 was added and centrifuged at 13,000 rpm for 1 minute. The wash buffer was then decanted. This RNA extraction procedure included a DNase treatment as well and for every sample 10 µL of the DNase was mixed with 70 µL of the DNase digestion buffer and the entire 80 µL was loaded onto the binding column. The tube was then incubated at room temperature for 15 minutes before adding 500 µL of the wash buffer 1 again and centrifuging at the maximum speed for 1 minute. The flow through was discarded once again and 500 µL of the wash buffer 2 with ethanol was added into the binding column. The microcentrifuge tube was then centrifuged at

13,000 rpm for 30 seconds before discarding the flow through and drying the tubes on a clean absorbent paper. The tube was then centrifuged once again, at the same speed for the same amount of time to remove any of the remaining wash buffers. The binding column was then transferred to a new 2 ml collection tube and 50 µL of the elution buffer was added into the binding column, incubated at room temperature for a minute and then centrifuged at maximum speed for a minute. The RNA was present in the flow through elution buffer and was stored at -80^{o}C. The concentrations of the RNA samples were then determined using the Nanovue as described in chapter 3, section 3.2.2.2a.

The presence and the size of the extracted RNA preparations were confirmed by agarose gel electrophoresis as follows:

A 1% agarose gel was run in 1X TAE buffer for 90 minutes at 65 volts. The molecular weight marker was Marker XIV, that is 100 bp (750 ng, Roche Applied Science) and approximately 800 ng of each of the samples were loaded onto the gel.

4.2.3 Reverse Transcriptase PCR (RT-PCR) for primer and product analysis

To confirm whether the desired fragments were amplified and of the expected size an end point RT-PCR experiment was performed on the RNA extracted. The extracted *N. benthamiana* RNA were subjected to a two-step RT-PCR using the qScript Flex cDNA Synthesis kit from Quanta Biosciences. The first step of the RT-PCR was the synthesis of the cDNA using 10 µL of the template RNA (approximately 400ng), 1 µL of the nuclease free water, 2 µL of the random primers and 2 µL of the oligo dT primers. The reaction mixture was vortexed and centrifuged for 10 seconds before keeping it for incubation at 65^{o}C for 5 minutes. Following incubation, 4 µL of the qScript Flex Reaction Mix (5X) was added along with 1 µL of the qScript Reverse Transcriptase. The tubes were vortexed, centrifuged for another 10 seconds and then incubated under the following cycling conditions: 90 minutes at 42^{o}C, 85^{o}C for 5 minutes and the final hold at 4^{o}C. The cDNA was then stored at -20^{0}C.

PCR amplification was carried out with PVIP specific primers, as well as primers for the 18S RNA, actin and ubiquitin mRNAs. The sequences of the primers for these genes are given in Table 4.1. The primer sequences were designed using the software OligoExplorer.

The actin and ubiquitin primers were obtained from the tomato nucleotide sequences from the Gene Index database (http://compbio.dfci.harvard.edu/tgi/cgi-in/tgi/gimain.pl?gudb=tomato) by searching for tentative consensus sequences (TCs) previously analyzed in silico for their potential as reference genes. The accession numbers for actin and ubiquitin primers being TC194780 and TC193502 respectively (Lovdal & Lillo, 2009). The expected size of the actin, ubiquitin, 18S rRNA and the PVIP (PVIPF and PVIPR) fragments was approximately 100 bp, whereas for the PVIPT7FPRO and PVIPSP6 it was around 500bp.

Table 4.1.Primer sequences used for RT-PCR and qRT-PCR.

Gene	Primer name	Orientation	Sequence
Actin	ACTF	Forward	5'-TGGTAGAACCACCACTGAGC-3'
Actin	ACTR	Reverse	5'-GGAATCCACGAGACCACATA-3'
Ubiquitin	UBIF	Forward	5'-CCTCAGCAAACAAAGCAC-3'
Ubiquitin	UBIR	Reverse	5'-GTCTCCGTGGTGGTTTCTA-3'
18S rRNA	18SF	Forward	5'-GTAACCCGTTGAACCCCATT-3'
18S rRNA	18SR	Reverse	5'-CCATCCAATCGGTAGTAGCG-3′
PVIP-100bp	PVIPF	Forward	5'-GCACCAAAAGAGTTCGTGTC-3'
PVIP-100bp	PVIPR	Reverse	5'-TCCCAGATGATGTACAGCAA-3'
PVIP-500bp	PVIPT7FPRO	Forward	5'-AAATGGTTTCTGCAACCTTT-3'
PVIP-500bp	PVIPRSP6	Reverse	5'-GGATCCACATCAAGCTCTGG-3'

PCR was carried out as follows. First stand cDNA (2 µL) was combined with 10µL of the 2 X Promega Go Green PCR master mixes, 5.5 µL of sterile distilled water and 1.25 µL of 10 µM forward and reverse primers. Initial denaturation was carried out at 94^0C for 2 minutes

followed by 40 cycles of denaturation at 94^0C for 2 minutes, annealing at 50^0C for 30 seconds and extension at 68^0C for 1 minute. The final extension was at 68^0C for 5 minutes followed by hold at 15^0C.

In order to confirm the results obtained through end-point PCR, 20µL of the reaction was run on a 1% agarose gel in 0.5X TBE buffer for 90 minutes at 65 V. The marker run was marker XIV i.e. 100 bp marker (1µg).

4.2.4 Validation of the internal control genes using real time RT-PCR (RT-qPCR)

Validation of the reference genes was done by one step real time PCR analysis. mRNA levels in each of the *N.benthamiana* tissues was analysed along with each of the three reference genes. The reactions were carried out in a 96 well Eppendorf twin.tec real-time PCR plates covered with Applied Biosystem Optical Adhesive Film.

Each reaction included 5 µL of the One Step SYBR Green Master Mix (Quanta BioSciences), 0.2 µL of 10 µM forward and reverse primers, 0.2 µL of the qScript One-Step RT enzyme, 3.4 µL of nuclease free water and 1 µL of the RNA (100ng) template. The total reaction volume was 10 µL which was scaled down from the volume recommended in the qScript One-Step SYBR Green qRT-PCR kit from Quanta Biosciences. Each PCR was carried out in triplicate, except for the negative controls which were done in duplicate. The reaction conditions were that of the 3-step PCR cycling i.e. 49^0C for 10 minutes, 95^0C for 5 minutes followed by 40 cycles of denaturation at 95^0C for 10 seconds, annealing at 50^0C for 20 seconds and extension at 68^0C for 30 seconds. An additional melting curve cycle was also included in order to confirm that all the products obtained are the same products and to distinguish the products from primer-dimers and other types of contamination. Reactions were cycled in an Eppendorf Realplex gradient PCR machine.

4.2.5 Determining the PCR efficiency for each mRNA

RNA from different tissues were diluted in a ten fold dilution series using sterile distilled water. The PCR efficiency obtained for the leaf sample from this experiment was assumed to be applicable for the leaf samples subjected to the different light treatment experiment as well. Each reaction consisted of 5 µL of the One Step SYBR Green Master Mix (Quanta BioSciences) , 0.2 µL of 10 µM forward and reverse primer, 0.2 µL of the qScript One-Step RT enzyme, 3.4 µL of the nuclease free water and 1 µL of the RNA (100-150 ng). Each of the RNA samples was tested in triplicate, except for the negative controls which were done in duplicates.

The reaction conditions were that of the 3-step PCR cycling i.e. 49^0C for 10 minutes, 95^0C for 5 minutes followed by 40 cycles of denaturation at 95^0C for 10 seconds, annealing at 50^0C for 20 seconds and extension at 68^0C for 30 seconds. An additional melting curve cycle was also included in order to confirm that all the products obtained are the same products and to distinguish the products from primer-dimers and other types of contamination.

The Ct values once obtained were plotted on the graph against the $\log_{10}$ concentration of the RNA used. The graph was then used to determine the slope and the efficiency of the real-time reaction.

4.2.6 Determining the variability of the PVIP gene in different tissues and under different environmental cues:

4.2.6.1 Relative quantification of the PVIP mRNA in different tissues of *N.benthamiana*

In order to determine whether the PVIP gene is amplified in all the tissues and to what extent the concentrations differ from one tissue to another, a relative quantification of the PVIP mRNA from different tissues was carried out, namely leaf, stem, flower and root tissue. The reaction conditions and constituents were the same as in section 4.2.4. Amplification of the actin gene product was used as the reference gene. The calibrator used

in this case was the leaf sample; therefore the relative amounts of PVIP mRNA in other tissues were measured with respect to leaf.

4.2.6.2 Relative quantification of the PVIP mRNA in response to light/dark

N.benthamiana plants were grown under continous cycles of 16 hours light/ 8 hours dark. Three plants were transferred to the dark by covering with foil for 48 hours before total RNA was extracted. Three other plants were transferred to the dark 24 hours before being exposed to 1 cycle of 16 hours light/ 8 hours dark prior to RNA extraction. RNA was extracted at the same time from three plants that had not been covered in foil but rather treated with continous 16 hour light/ 8 hour dark cycles for the entire period. RNA was extracted from the leaves using the method described in section 4.2.2.

In order to determine whether the PVIP gene was amplified under the different conditions of light, relative quantification of the mRNA from the leaf sample of different plants was carried out. RNA was extracted from the three plants for each treatment i.e. three plants for light treatment, three for dark and three for light and dark treatment. The relative amounts of mRNA varying with respect to light were tested in triplicates, except for the negative control which was done in duplicate. Leaf tissues from light grown plants were used as the calibrator samples. Actin was used as the reference gene against which the gene of interest (PVIP) was normalised and any errors in pipetting determined. The reaction conditions and constituents were the same as on section 4.2.4.

A melting curve analysis was included in each of the real-time experiments to confirm that all the products obtained are the same products i.e. the same product will have a single melting curve and also to distinguish the products from primer-dimers and other types of contamination.

During the reaction analysis, the cycle threshold (Ct) was set manually at a point where the fluorescence value begins to increase considerably above that of the background. The ΔΔCt value was provided by the Realplex machine itself when the option for relative

quantification was selected. This value was confirmed manually as well. The relative amounts of the RNAs were also determined manually using the formula

$$R = 2^{-\Delta\Delta Ct} = 2^{-(\Delta Ct\ sample - \Delta Ct\ calibrator)} = 2^{(Ctsample-CtHKG) - (Ctcalibrator-CtHKG)]}$$

Where R represents expression levels and HKG is the house keeping gene (Winer, Jung, Shackel & Williams, 1999).

The reaction efficiency was then determined by the formula $10^{(-1/slope)}$-1. The slope is obtained from the graph plotted while determining the standard curve of the reaction (Ginzinger, 2002).

4.2.6.3 Agarose gel electrophoresis

In order to confirm the RT-qPCR results, i.e. formation of one product alone and the product formation in the NTC if any, is contamination or a primer-dimer, agarose gel electrophoresis was performed. This provided confidence in the melting curve analyses in identifying any contamination and primer dimer.

Hence, a 1% agarose gel in 0.5X TBE buffer was used. The gel was run for 90 minutes at 65 V. The marker run was marker XIV i.e. 100 bp marker (1µg) and the amount of sample loaded into the gel was 25µL.

4.3 Results

4.3.1 End point RT-PCR of mRNAs in *N.benthamiana* tissues

4.3.1.1 End point RT-PCR of PVIP mRNA

The RNA extracted from *N.benthamiana* leaves, roots, petioles, stem and flowers appeared to be intact. Figure 4.1 shows the presence of the 18S and 28S rRNA bands from the different tissues. The background smear would be the mRNA present in the sample. Absorbance readings indicated the RNA yield from leaf tissue was the greatest. This was confirmed by agarose gel electrophoresis; the samples in Figure 4.1 were loaded as equal volumes. This is not surprising as, of all the tissues sampled, leaf is probably the most metabolically active.

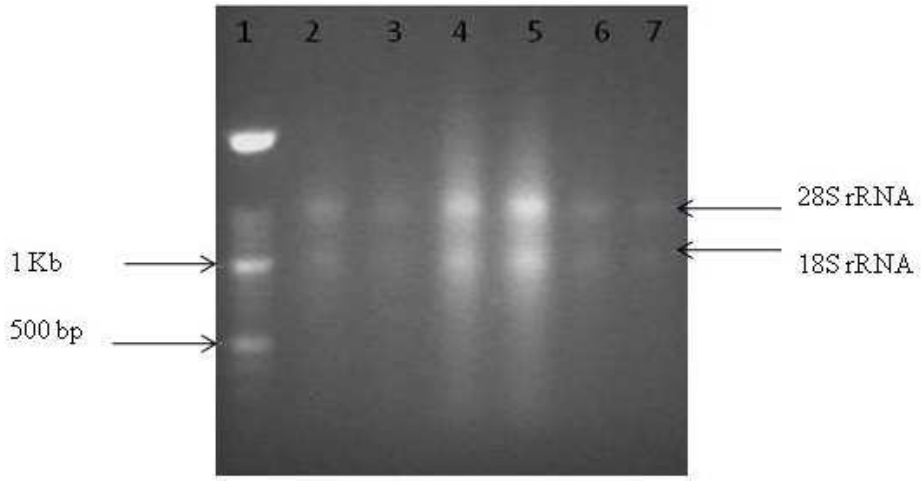


Figure 4.1. Agarose gel electrophoresis of total RNA extracted from *N.benthamiana* tissue. Lane1 is that of marker XIV i.e. 100bp marker (750ng) while lanes 2 to 7 are the RNA from petiole, stem, leaf1, leaf2, roots and flowers, respectively. The 28S and 18S rRNA bands are indicated on the right while molecular size markers are indicated on the left.

RT-PCR amplification of PVIP was carried out using two different primer sets. The primers PVIPFT7PRO and PVIPRSP6 were expected to give rise to a product of approximately 500bp (Figure 4.2, lanes 2 to 6). PVIP was amplified from all the tissues except that of flowers (lane 6). Also, there was no formation of a primer-dimer in lane 6 wheras we can observe the primer-dimers in the remaining lanes including the no template control (NTC) which is lane 7. Therefore, PVIP may not be present within flowers but is

present within the remaining tissues; however this would need to be repeated to confirm this result.

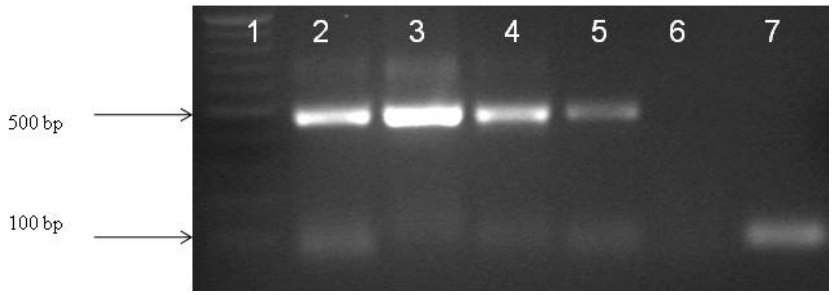


Figure 4.2. RT-PCR of *N.benthamiana* tissue using the PVIP primers PVIPFT7PRO and PVIPRSP6. Lane 1: 500 ng marker XIV i.e.100 bp marker; Lanes 2-6 RT-PCR of petiole, stem, leaf, roots and flowers, respectively; lane 7: no template control. Selected marker shown on the left.

The primers PVIPF and PVIPR were expected to produce an approximately 100bp product which can be seen in Figure 4.3, lanes 2-6. The expected products were amplified from all tissues, indicating that PVIP is expressed in all the tissues analysed. Amplification of the 100bp product with the primers PVIPF and PVIPR also indicated that these primers would be suitable for use in a quantitative RT-PCR. Figure 4.3 shows the presence of PVIP within flowers but figure 4.2 does not. Hence, PVIP is present within flowers but the high molecular weight PVIP (500 bp) could not be amplified from the flower sample whereas the low molecular weight PVIP could.

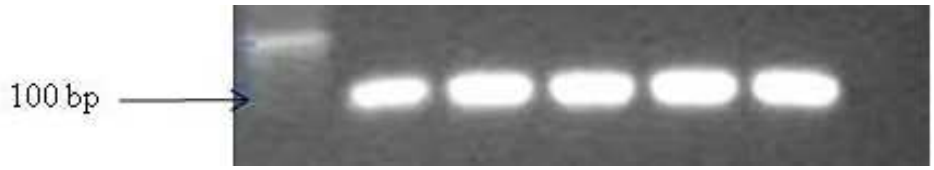


Figure 4.3. RT-PCR of *N.benthamiana* tissue using the PVIP primers PVIPF and PVIPR. Lane 1: 750 ng marker XIV i.e.100 bp marker; Lanes 2-6 RT-PCR of petiole, stem, leaf, roots and flowers, respectively. Selected marker shown on the left.

4.3.1.2 End point RT-PCR of reference gene mRNAs

The mRNAs of 18S, actin and ubiquitin were considered as reference genes for RT-qPCR. End point RT-PCR was carried out in order to confirm the primers designed to amplify these fragments were suitable i.e. specific products could be amplified and they were of the expected size which was around a 100 bp.

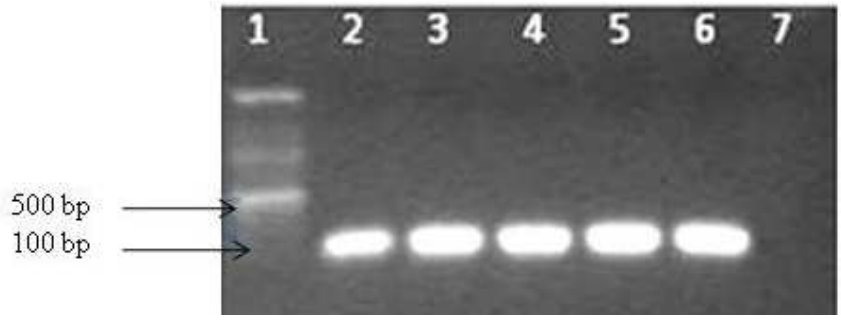


Figure 4.4.RT-PCR with the 18S primers. Lane 1: marker XIV i.e. 100 bp markers (750 ng); lanes 2 to 6: 18S gene from petiole, stem, leaf, roots and flowers, respectively. Selected markers are indicated on the left.

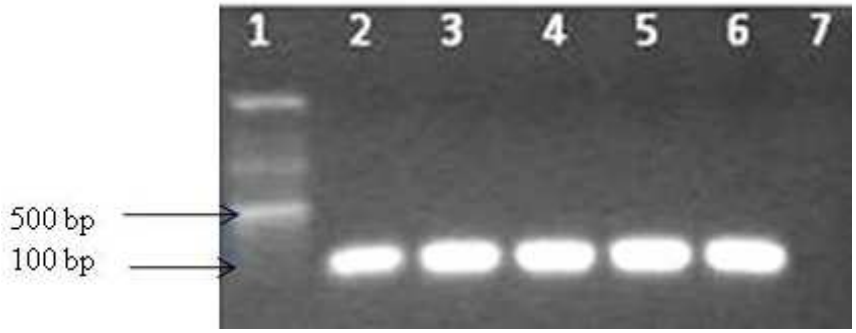


Figure 4.5. RNA amplified with the actin primers ACTF and ACTR. Lane 1: marker XIV i.e. 100 bp markers (750 ng); lanes 2 to 6: petiole, stem, leaf, roots and flowers, respectively. Lane 7 is a no template control (NTC). Selected markers are indicated on the left.

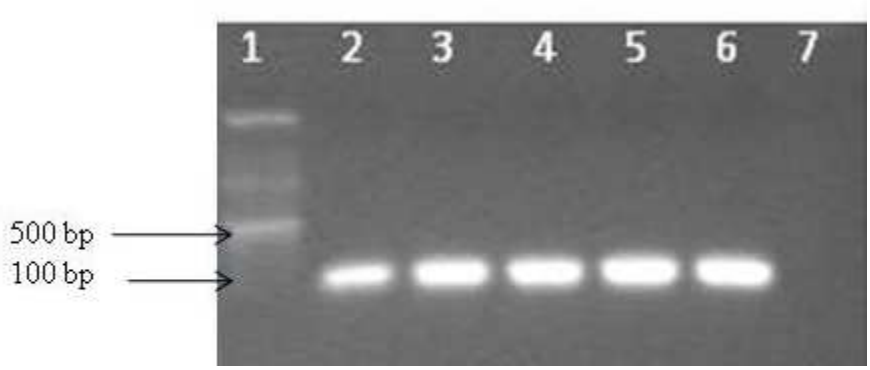


Figure 4.6. RNA amplified with the ubiquitin primers UbiF and UbiR. Lane 1: marker XIV i.e. 100 bp marker (750 ng); lanes 2 to 6: petiole, stem, leaf, roots and flowers, respectively. Lane 7 is a no template control (NTC). Selected markers are indicated on the left.

Figures 4.4, 4.5 and 4.6 shows that the expected 100bp products for the 18S, actin and ubiquitin genes were amplified in all the four tissues. Also, these products could be amplified from the leaf tissues that had been placed in the dark as well as the tissues placed in dark/light environment (Figures 4.7 and 4.8). Thus, these primers would be suitable for use in a RT-qPCR reaction.

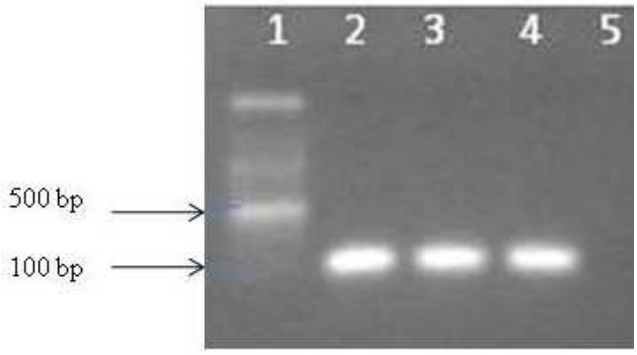


Figure 4.7. Dark treated RNA amplified with the 18sRNA, actin and ubiquitin primers respectively. Lane 1: marker XIV i.e. 100 bp marker (750 ng); lanes 2 to 4: 18S RNA, actin and ubiquitin products respectively. Lane 5: no template control (NTC). Selected markers on the left.

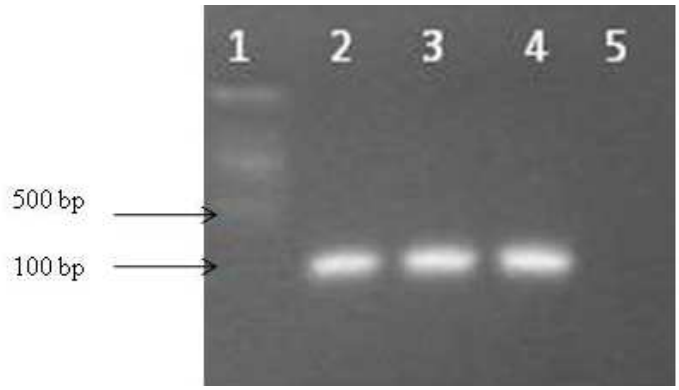


Figure 4.8. Dark and light treated RNA amplified with the 18sRNA, actin and ubiquitin primers respectively. Lane 1: marker XIV i.e. 100 bp marker (750 ng); lanes 2 to 4: 18S RNA, actin and ubiquitin products respectively. Lane 5: no template control (NTC). Selected markers on the left.

4.3.2 Validation of reference genes

4.3.2.1 Abundance of each reference mRNA in *N.benthamiana* tissues

In order for a reference gene to be regarded useful, it must not vary between samples or treatments. The candidate reference genes, i.e. 18S, actin and ubiquitin, had to be validated to ensure that their expression levels did not vary between the different *N.benthamiana* tissues. Validation of the reference genes was essential in order to decide which of the three genes to use for future experiments i.e. 18S, actin or ubiquitin.

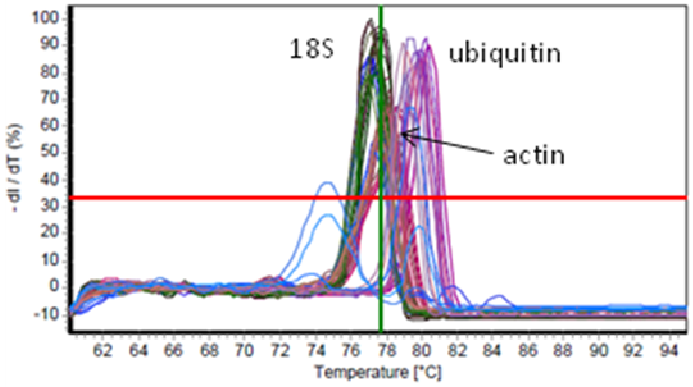


Figure 4.9. The melting curve of the products for the reference genes and the negative controls. Temperature is plotted against –dl/dt in a melting curve analysis.The fluorescence profiles are labelled and coloured for clarity: green curves are that of 18S RNA, pink are ubiqutin and red are actin. The blue coloured fluorescence is that of the negative control.

Melting curve analysis showed that each amplicon was the only PCR product in each reaction and, as expected, each had a different melting temperature (Figure 4.9). The melting curve analysis showed three main peaks which indicated that three main products were formed which are most likely to be 18sRNA, ubiquitin and actin. The peaks in blue were most likely primer-dimers since their melting temperatures did not match with the temperatures of the products formed.

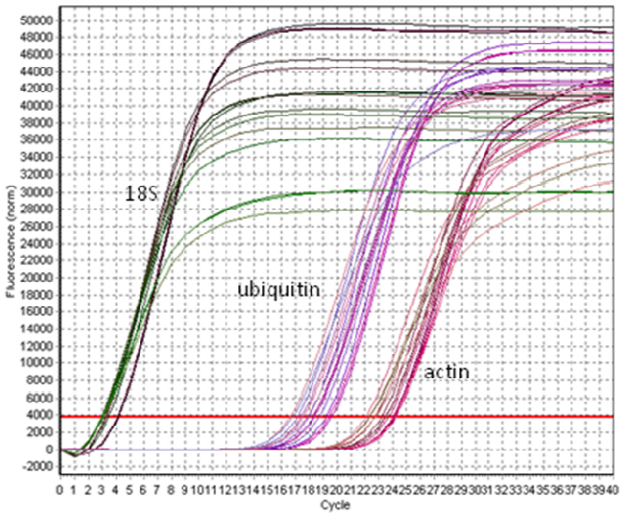


Figure 4.10. Validation of 18S, ubiquitin and actin mRNAs by RT-qPCR in all *N.benthamiana* tissues. The fluorescence profiles are labelled and coloured for clarity: green curves are that of 18S RNA, pink are ubiqutin and red are actin. The graph is plotted with the cycle number (Ct) on the X axis and the fluorescence on the Y axis.

Figure 4.10 shows the RT-qPCR results comparing expression levels of all three genes in the different tissues. It can be seen that for each gene, the expression level is within a specific range that differs from the other genes. 18S showed the least variation while ubiquitin showed the greatest. Further, the expression level of 18S was much greater than that of either ubiquitin or actin, with a very low Ct value of 3-4. This was not surprising as 18S rRNA is known to be very abundant. Actin showed the lowest expression in all tissues.

An ideal reference gene would be one which did not vary in abundance between the different tissues analysed. Of all the three candidate reference genes that varied, 18S varied the least. The amount of 18S present in all the tissues was very high since the Ct value was very low i.e. 3 to 4. Hence, it was felt that this would cause difficulty when normalising the data and was considered to be the least suitable to use. It was known from a trial experiment that the PVIP expression level in leaves was around the same as actin. Also, this gene shows less variation between tissues than ubiquitin. Therefore this gene was chosen as the reference gene for the relative quantification of PVIP expression.

4.3.2.2 Comparison of PCR amplification efficiency

Reaction efficiency is essential to determine whether the products increase at the rate of two-fold per cycle. The reaction efficiency can be determined by performing a serial dilution of the cDNA produced from the original RNA template, determining the Ct for each template, plotting a graph of Ct against $\log_{10}$ of the RNA concentration and calculating the slope of the graph. A slope of -3.1 to -3.32 is indicative of an efficient reaction.

It is considered that for a reference gene to be considered valid, it must amplify with the same efficiency as the gene of interest. Hence, for a reaction to be considered optimally efficient, the slope of the values generated must be within the range of -3.1 to -3.32. A slope of this value indicates a doubling of the amount of PCR product with each cycle. This efficiency can vary depending upon the pipetting accuracy and the contamination within the

RNA sample. So, while the efficiencies of the amplification for different products may not be the same, they might still be determined to accurately measure relative quantification.

The expression level of the PVIP mRNA in different *N.benthamiana* tissues was determined. The cDNA synthesised from leaf, stem, root and flower tissues were subjected to a ten fold serial dilution (1, 1:10, 1:100, 1:1000). Actin, ubiquitin and PVIP mRNAs were then amplified by qPCR. Amplification for each gene was carried out in triplicate. Following the qPCR reaction, the Ct values were determined.

Table 4.2 shows that the reaction efficiencies varied from tissue to tissue for each gene showing a range of efficiencies from 44% to 227%. The REST software by Qiagen was used to determine the reaction efficiency. Also, care was taken to manually remove any of the outliers that were present within the triplicates. The amplification efficiency for each of the candidate reference genes did not match exactly that observed for PVIP; however, the variation between tissues observed for PVIP was generally matched by actin i.e. the differences between the efficiencies of PVIP amplification and those of reference genes were generally less for actin than ubiquitin.

The reaction efficiencies were generally not within the desired range of 90% to 100% (Dorak, 2006). This was most likely due to the limited number of data points that is, limited number of dilutions to test the efficiency. Had the number of dilutions been increased, the data obtained would have been more robust. However, due to the limited number of RT-qPCR assays available within the budget of the project the data size could not be increased. Also, pipetting errors may have caused the efficiency of the reaction to reduce considerably. Other factors such as poor quality of the extracted RNA and the contaminants or inhibitors present may deteriorate the efficiency even more. Irrespective of the low efficiencies, the experiments were continued in order to get an indication of the expression level of PVIP in different tissues and under different conditions of light. Reaction efficiency greater than 100% usually indicates the formation of primer dimers because a greater amount of double stranded products are now available for the SYBR Green to bind to and give out fluorescence. This, however, was not observed when the samples were run on an agarose gel (figure 4.13) but seen in figure 4.12 in the form of small peaks. The concentration of these primer-dimers might have been too low to be seen in the agarose gel.

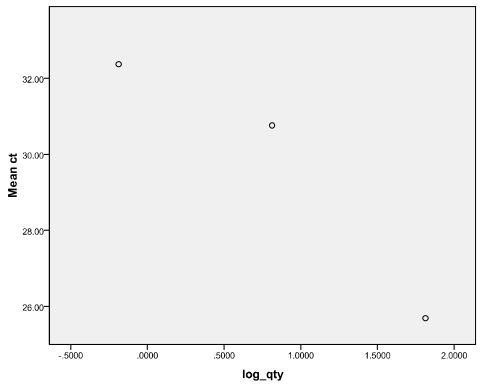
Hence, the increase in reaction efficiency might be because of the formation of primer-dimers.

Figure 4.11 shows the graph of the Ct values plotted against the $\log_{10}$ to the concentration of RNA. The value of the slopes was determined from these graphs.

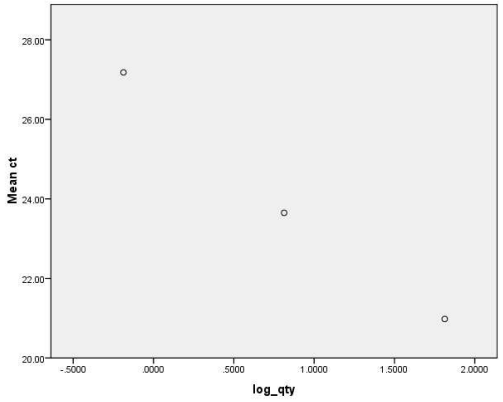
Table 4.2. The slopes and the reaction efficiencies of the PVIP, actin and ubiquitin mRNAs following amplification from *N.benthamiana* tissues.

Tissue	**Value**	**PVIP**	**Actin**	**Ubiquitin**
Leaf	Slope	-3.34	-3.2	-2.6
	Reaction efficiency	98%	110%	142%
Flower	Slope	-6.2	-5.3	-1.9
	Reaction efficiency	44%	53%	227%
Root	Slope	-6.41	-4.1	-2.32
	Reaction efficiency	43%	73%	172%
Stem	Slope	-2.9	-4.09	-3.2
	Reaction efficiency	120%	75%	102%

A)



B)



C)

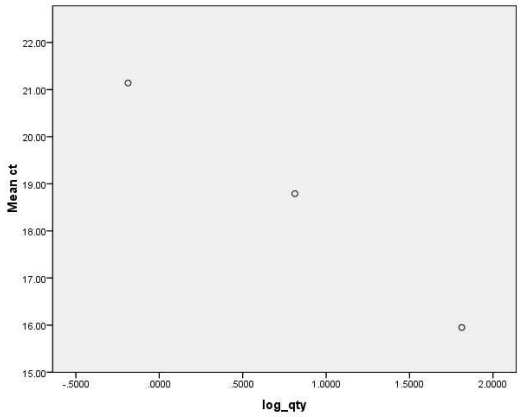


Figure 4.11. The scatterplots of the Ct values obtained against Log_{10} of the RNA concentration from leaf tissue. The points obtained should be in a straight line with a regression efficiency of -1.0. A) Represents that of PVIP, B) represents actin and C) represents ubiquitin.

The slopes obtained were between -2 and -5 for most of the reactions. Hence, the values obtained were not within the range of -3.1 and -3.4 indicating that the reaction efficiency needed to be improved.

4.3.3 Relative quantification of PVIP in different *N.benthamiana* tissues

End point RT-PCR showed that PVIP mRNA is expressed in leaves, stems, petioles, roots and flowers (Figure 4.3). RT-qPCR was used to determine the relative abundance of PVIP mRNA in these tissues. For these experiments, due to the limited number of assays available, petiole tissue was excluded. Amplification of the PVIP mRNA by RT-qPCR using the PVIP primers i.e. PVIPF and PVIPR resulted in a single product as determined by the the melting curve analysis (Figure 4.12) and agarose gel electrophoresis (figure 4.13) for each of the tissues tested. Agarose gel electrophoresis (Figure 4.13) confirmed the size of the PCR product to be the expected 100bp product as observed previously.

Amplification of the actin mRNA by RT-qPCR using the primers ACTF and ACTR resulted in a single product which was determined by the melting curve analysis in each of the tissues tested (Figure 4.12 respectively). This was also confirmed by agarose gel electrophoresis (figure 4.13) which showed the product to be the expected 100 bp product as observed previously.

The amplified products in figures 4.14 and 4.15 with a Ct value of 30+ are that of the no template control (NTC). The fact that no 100 bp fragment was observed by agarose gel in the NTC implies that the amplified products seen during RT-qPCR were primer-dimers or even potential contamination. Also, the melting curve experiment showed that the melting temperature of these products is at a temperature different from that of the desired products, adding further evidence of the NTC being a primer-dimer rather than a contamination.

Amplification of both, PVIP and actin mRNA was carried out for each of the three plants in triplicate. Ct values obtained for each plant i.e. the Ct values of all the triplicates within a single plant were averaged since the difference between the outliers were marginal. The Ct values for each tissue from each plant were normalised against actin to give a ΔCt value for

each tissue. All expression levels were calibrated against leaf tissue, meaning that the normalised expression level of PVIP in each tissue was compared against that seen in leaf. This difference was calculated as the ΔΔCt for each tissue. The amount of PVIP mRNA in a tissue, normalised against the reference gene actin and relative to the leaf samples was calculated as $2^{-\Delta\Delta Ct}$.

A) B)

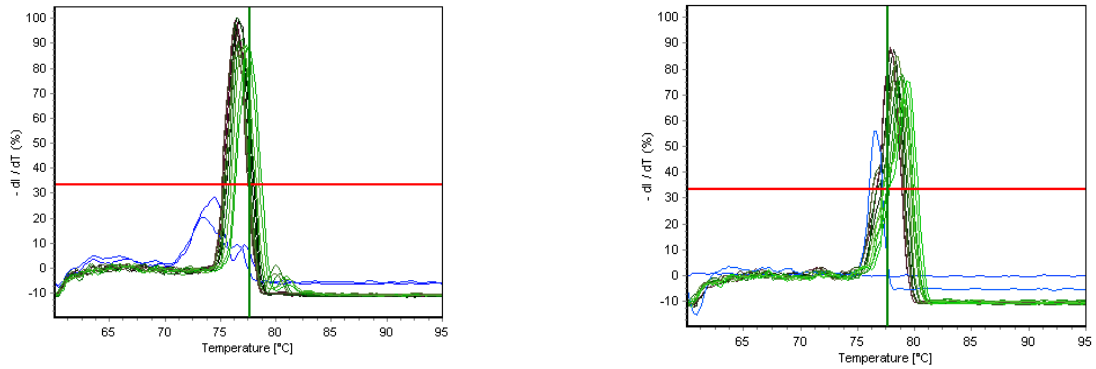


Figure 4.12. The melting curve of the PVIP product (A) and the actin product (B). The blue fluorescence alongside the tall peaks is the NTC. The melting curves of the NTC being much less compared with the tall peaks indicate that the products are different from the desired products which could be primer dimers or a potential source of contamination.

. A) B)

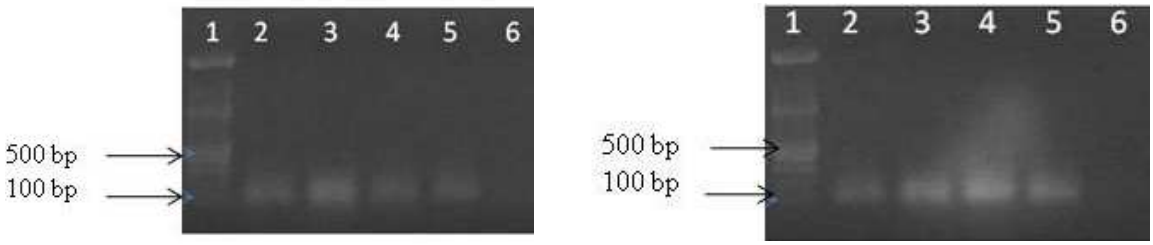


Figure 4.13. Agrose gel electrophoresis of A) PVIP and B) actin RT-qPCR products amplified from stem (lane 2), leaf (lane 3), flower (lane 4) and roots (lane 5). Lane 1 shows 1µg of the 100 bp marker and lane 6 shows the no template control (NTC)

A) B) C)

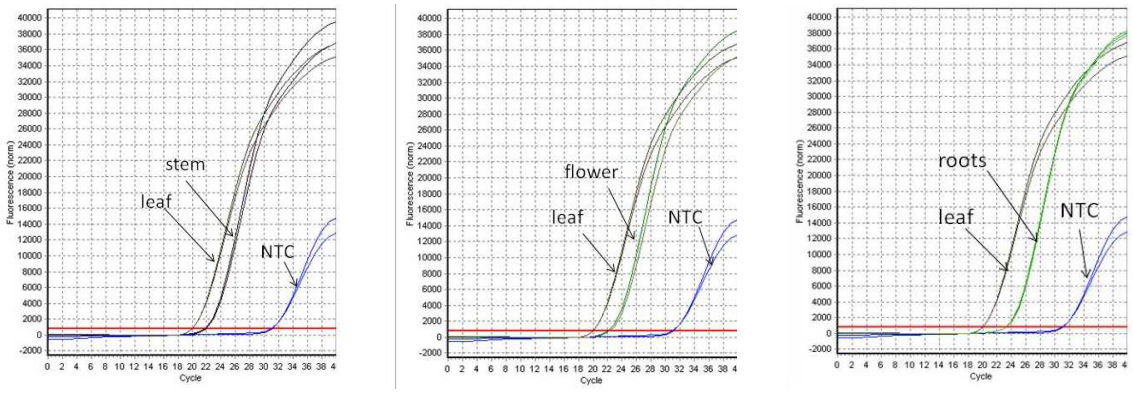


Figure 4.14. Amplification plots of PVIP RT-qPCR from leaf compared with that from (A) stems (B) flowers and (C) roots. The amplification plot with the lowest Ct value in each of the three figures is that of leaf followed by the respective tissues. The blue coloured amplification observed at a cycle close to 32 is the no template control (NTC).

A) B) C)

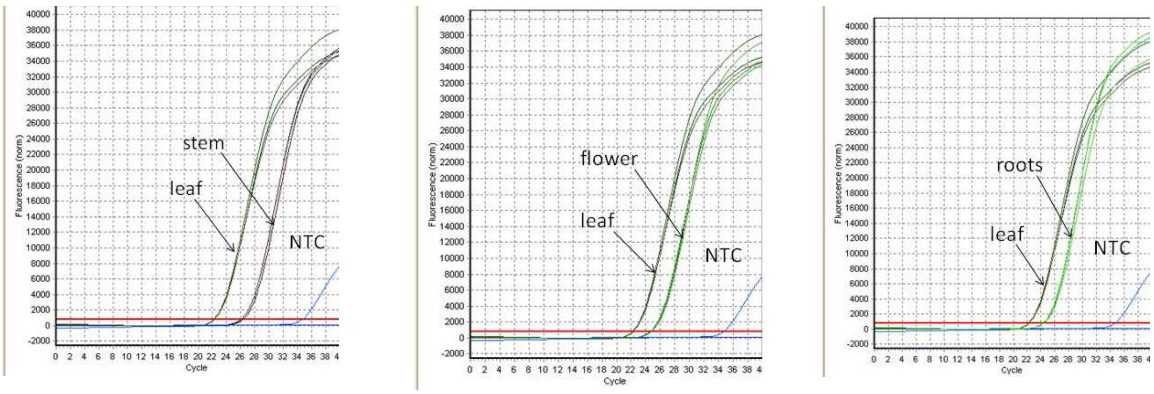


Figure 4.15. Amplification plots of Actin RT-qPCR from leaf compared with that from (A) stems (B) flowers and (C) roots. The amplification plot with the lowest Ct value in each of the three figures is that of leaf followed by the respective tissues. The blue coloured amplification observed at a cycle close to 34 is the no template control (NTC).

Table 4.3 shows that the expression levels of PVIP in each tissue tested relative to that seen in leaf. Stem appeared to have a five fold higher level of PVIP mRNA while root tissue had

half the amount compared with leaf tissue. The level in flower was slightly higher than that seen in leaf. Hence, the order of expression is stem>flower>leaf>roots.

Table 4.3. The ΔCt, ΔΔCt, and the fold difference of the PVIP gene in different tissues with respect to the calibrator i.e. leaf.

Tissue	ΔCt	ΔΔCt	Fold Difference
Stem	-4.37	-2.31	4.95
Leaf	-2.06	0	1
Flower	-2.52	-0.46	1.37
Roots	-1.19	0.87	0.55

4.3.4 Relative quantification of PVIP in light/dark treated *N.benthamiana* leaves

Bioinformatic analysis of the predicted transcription factor binding sites within the 5'flanking regions of homologous genes from *A.thaliana* and rice suggested that the PVIP gene may be responsive to light (see chapter 2). This was tested by comparing the PVIP mRNA levels in plants that had been treated with different light regimes. Nine plants were grown continously under standard cycles of 16 hours light and 8 hours dark. Three were held under this regime (light treated) while six plants were shifted to continuous dark, three of which were held for 48 hours (dark treated) while the remaining three were returned after 24 hours to one cycle of 16 hours light/ 8 hours dark (dark/light treated). These latter samples were treated with dark followed by light to determine if any affect of the dark treatment would be reversible on being subjected to light.

Amplification of the PVIP mRNA by RT-qPCR (Figure 4.18) using the primers PVIPF and PVIPR resulted in a single product as determined by melting curve analysis in each of the tissues tested (Figure 4.16). This was confirmed by agarose gel electrophoresis (figure

4.17) showing the product to be the expected 100bp product as observed previously. Similarly, amplification of the actin mRNA by RT-qPCR using the primers ACTF and ACTR primers (Figures 4.18) resulted in a single product formation which was determined by melting curve analysis as well (Figure 4.16). This was also confirmed by agarose gel electrophoresis (figure 4.17) showing the product to be the expected 100bp product as observed previously and confirming that the product if formed in the no template contral is a primer-dimer rather than a contamination.

The calibrators used in this case were the light treated samples (i.e. those grown continuously under 16hrs day/8hrs dark conditions). These samples were taken as one group by the Realplex real time PCR machine and the two treatments compared against this calibrator. All samples were normalised against actin.

A) B)

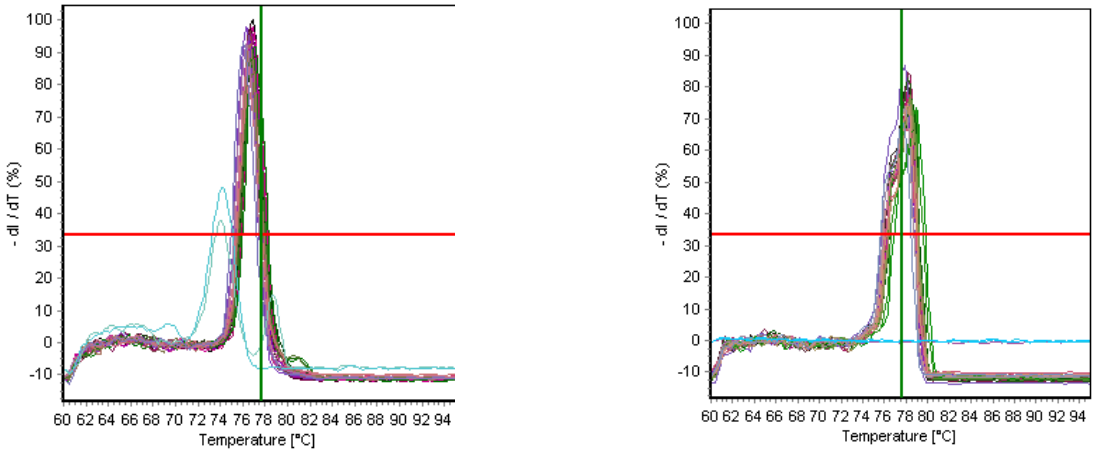


Figure 4.16.The melting curves of A) PVIP and B) actin from all samples. The blue curve is the no template control.

A) B)

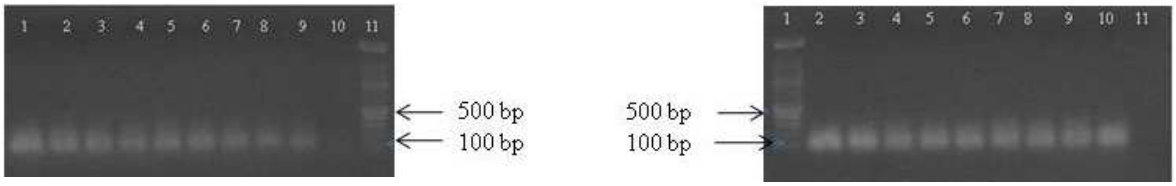


Figure 4.17. Agarose gel electrophoresis of the RT-qPCR products for A) PVIP and B) actin in leaves treated with normal day/night conditions (lanes 1 to 3 for A and 2 to 4 for B), 48 hrs continuous dark treated (lanes 4 to 6 in A and 5 to 7 in B) and leaves treated with 24 hrs dark followed by 24hrs standard day/night conditions (lanes 7 to 9 in A and 8 to 10 in B). The no template controls are in lane 10 in A and lane 11 in B.

A) B)

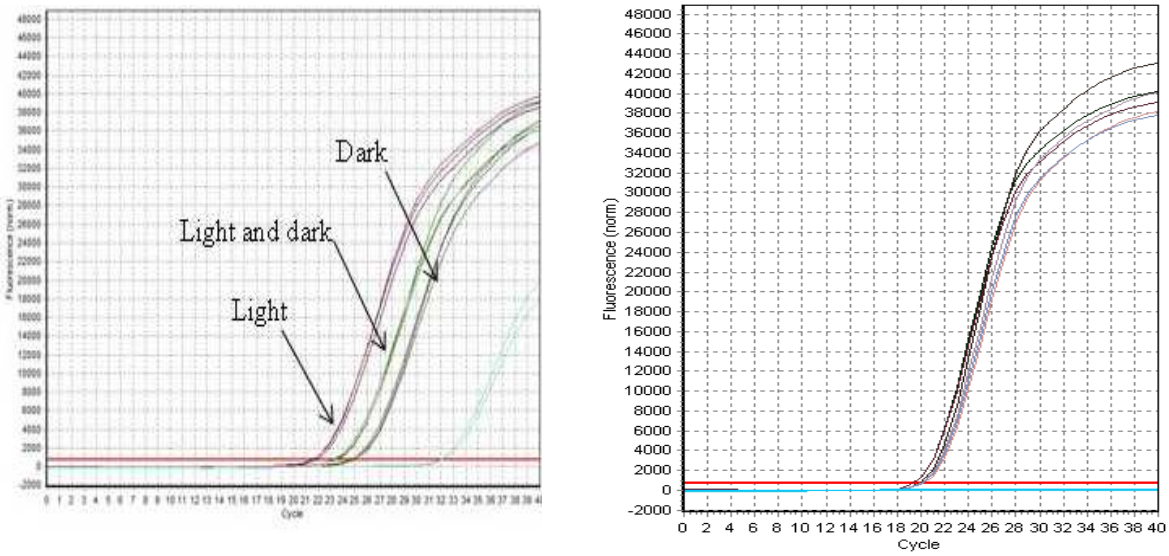


Figure 4.18. Example amplification plots of (A) PVIP and (B) Actin RT-qPCR from leaf under different conditions of light i.e. light treatment, dark treatment and dark/light treatment. The amplification plot with the lowest Ct value in each of the three figures is that of leaf under different environmental conditions. The blue coloured amplification observed at a cycle close to 32 in (A) is the no template control (NTC). The amplification of Actin in (B) are very close to each other making it difficult to segregate the amplification from one condition to another.

Table 4.4. The ΔCt, ΔΔCt, and the fold difference of the PIVP gene leaf from the different light treatment tissues with respect to the calibrator that is light treated leaf. The data for each replicate plant is the average triplicate samples.

Condition	**Replicate plant**	**ΔCt**	**ΔΔCt**	**Fold Difference**
Dark treated	1	4.06	1.79	0.30
	2	1.64	-0.94	1.92
	3	0.80	-1.62	3.07
Dark/light treated	1	2.84	0.57	0.67
	2	3.06	0.48	0.72
	3	2.68	0.26	0.84

Table 4.4 gives the expression levels of the treated tissues tested relative to that observed in leaves grown under dark as well as normal day/night conditions of 16 hrs light/ 8hrs dark. The levels of PVIP mRNA appeared to vary between treatments, particularly for those samples reated with continuous dark for 48 hrs. It appeared that the PVIP mRNA levels subjected to dark treatment for 24 hours ranged from 0.3-3.0 times the level of PVIP mRNA compared with those exposed to standard day/night conditions. Further, leaves exposed to 24 hrs of continuous dark followed by standard day/night (dark/light treated) appeared to have lower levels of PVIP mRNA (approximately 0.7 times) than those leaves exposed only to normal day/night.

The variation observed within the different treatments was however thought to have occurred because no separate validation experiment was performed for the actin genes against the different environmental treatments. The reaction efficiency was determined just for the light treated PVIP sample in leaf and ideally should have been determined for both PVIP and actin under different environmental conditions. However, the number of assays

available within the given budget restricted the number of experiments that could be performed.

The large variation observed in the dark treatments may be due to pipetting errors when dispensing the template. However, given the limited variation in the other samples tested, the pipetting error can't account entirely for the variation observed. Also, the NTC appeared to be contaminated with low levels of template showing amplification at cycle number 31. Hence, it is possible that the dark samples showing a high level of PVIP mRNA might also be contaminated. There might also be a possibility that while dark treatment does induce changes in the PVIP mRNA levels, however, these levels have not stabilised by 48 hours. However, if this is the case then why would only one plant show a decrease in the mRNA levels while the other plants show an increase in the mRNA levels? The data, when all treatments are considered together, do suggest that PVIP mRNA levels are responsive to changes in light levels. Given that the mRNA levels in the dark/light treated leaf tissue is lower than in the light treated tissue, it is surprising that dark treatment would induce and increased level of expression. It might be expected that the dark/light treatment might be intermediate between the light and the dark treated tissues. Therefore, it is more likely that the mRNA levels are reduced in the dark treatments and the order of expression of the PVIP mRNA subjected to light treatment might be

Light treatment > light and dark treatment > dark treatment.

4.4 Discussion

Real-time PCR was observed to be a good method for quantifying differential gene expression because it involves the use of only small amounts of cDNA and also the relative quantification by the ΔΔCt method although the least reliable method with respect to quantification is well suited for this kind of an experimental system.

The study involved the analysis and evaluation of three endogenous reference genes namely 18S RNA, actin and ubiquitin because these genes are commonly used as reference genes. These genes were chosen because they were found to be least affected by light stress, in fact actin and ubiquitin were ranked as the most stable reference genes under different light conditions in tomato (Lovdal & Lillo, 2009) and the experiment involved the determination of relative amounts of the PVIP gene with respect to light treatment. Also, actin was found to be one of the stable reference genes when experimented for light treatment in *A.thaliana* (Hu, Fan, Li, Zhang & Fu, 2009). Some of the well known reference genes like GADPH have been found to be affected by different treatments, biological processes and even different tissues and cell types and hence the validation of these reference genes is essential. The assessment of the transcription levels of the three reference genes that is, 18S, actin and ubiquitin showed hardly any difference in regulation from one tissue to another, thereby fulfilling the general requirement for a reference gene for comparison of expression in different tissues that is of constant expression in all the tissues.

The relative quantification of the PVIP gene in *N.benthamiana* using RT-qPCR indicated that the PVIP mRNA is expressed at a greater level within stems and flowers compared with leaves and roots. The order of expression being stem>flower>leaves>roots. PVIP_Nb protein was shown to have high homology with OBE1 and OBE2 from *A.thaliana*. These proteins have been shown to have a role in shoot and root induction. However, the amount of OBE1 and OBE2 in *A.thaliana* was found to be uniform that is, the concentration of OBE1 or OBE2 proteins are found to be constant throughout the plant (Saiga et al., 2008). Having strongest sequence homology at the protein level with the OBE2 protein and also being present in shoots (which includes stem) and roots, the PVIP protein might be having similar functions.

OBE1 and OBE2 have been found to regulate the response of auxins in plants. Auxin plays an important role in plant development. However, the mechanism between light regulation and the amounts of auxin produced have not been determined as yet. Indole-3-actic acid (IAA), the most important member of the auxin family has been seen to vary with regards to the different light treatments. IAA has been found to generate the majority of auxin effects in healthy plants (Laxmi, Pan, Morsy & Chen, 2007). The levels of IAA were experimented with different spectra of light treatments and it was observed that the level of IAA produced in the range 700-800nm light, that is, not within the visible range, was approximately 1.5 to 2.0 fold the amounts of auxin produced under normal conditions. This suggests that light does have an effect on the amounts of IAA produced and that it increases as the spectrum shifts from the visible to the invisible spectrum. Hence, this might indicate that the levels of IAA might have a similar pattern of expression in a dark environment as well (Kurepin, Emery, Pharis & Reid, 2007). This might imply that the concentration of OBE proteins might increase in the dark which in turn increases the concentration of IAA or rather auxin. However, rather contradictory results have been found in *N.benthamiana* because of a potential decrease in the PVIP mRNA when the plant is exposed in dark.

The presence of PVIP mRNA in leaves and flowers might indicate some functional role within these parts of the plant as well, which has yet to be defined. However, some of the transcription factors that were identified in chapter 2 have been shown to be involved in early flowering and flower development that is, SEF4 and FHA, respectively. The mRNA being present in leaves and flowers might indicate that the PVIP protein could be involved in the photosynthetic machinery and also the development of flower such as stamens development etc. However, to date little has been studied about the function of PVIP and its distribution within the different tissues, but the fact that we found the PVIP gene to be expressed in these different tissues suggests it is necessary throughout the plant.

PVIP_Nb mRNA appears to be responsive to changes in light levels. This confirms an earlier observation based on end point RT-PCR (C.Higgins, pers.comm.). Nothing has been published about the light response of the OBE1 and OBE2 mRNA in *A.thaliana*. The PVIP_Nb mRNA in leaf tissue treated with different light regimes showed that when plants are placed in continuous dark, the level of PVIP mRNA appars to be reduced. While the dark treated tissue showed variation in the levels of PVIP mRNA, it is thought that the dark

treated tissue is more likely to have a lower PVIP mRNA leve than dark/light treated, which would suggest that light induces the expresson of PVIP mRNA.

General Discussion

PVIP was initially identified by Dunoyer et al (2004) as being a protein that interacts with the VPg protein of potyviruses. An earlier study showed that the PVIP mRNA in *N.benthamiana* was present in several different plant tissues (C.Higgins, pers. Comm.). This showed that PVIP is expressed in healthy, unifected tissues and therefore must have a role that is independent of viral infection. The overall aim of this project was to determine this role.

BLAST analysis identified a series of close homologues of the PVIP sequence from *Nicotiana benthamiana* such as those from *Arabidopsis thaliana, Vitis vinifera, Oryza sativa, Pisum sativum* etc. The transcription factor HMG-1 has been shown to support the theory of PVIP being involved in DNA recombination and repair. PVIP may also act as a Constans acting protein thereby bringing about flowering just like the transcription factor FHA which might bring about early flowering as well. Consisting of the PHD finger domain and the possibility of PVIP acting like its homologues involved in zinc finger binding protein, PVIP might be responsible to respond to pathogen response. Thus it is quite possible for PVIP to be a multi-functional protein involved with the development and the survival of plant under different conditions of stress.

Real-time PCR experiments showed that the PVIP mRNA is present throughout the plant but in different amounts. Through the homology analysis it was observed that the PVIP protein is similar to OBE2 from *A.thaliana*. OBE2 is required for the maintenance and establishment of the shoot and root apical meristems and is located within the nucleus. Thus, it is quite possible that the PVIP from *N.benthamiana* performs similar functions and is specific for the nucleus as well.

The probable responses and the environmental stimuli the PVIP gene responds to were predicted by analysing the predicted transcription factor binding sites within the 5'flanking regions of homologues from *A.thaliana* and rice. On the basis of the results obtained by this TESS analysis, PVIP_Nb was predicted to be responsive to light. Factors such as GT-1, GATA 1, 2 and 3 and SEF4 may bind to the prmoter region of PVIP_Nb. This however needs to be proven. Relative quantification of the PVIP mRNA showed that it is expressed

at different levels in various tissues and that the mRNA levels in leaf tissues is indeed responsive to light. The data suggests that the PVP mRNA is induced by light, but this remains inconclusive. This would need to be confirmed.

Two PVIP genes were observed within *A.thaliana* and *O.sativa*. The fact that *A.thaliana* is a dicot whereas *O.sativa* is a monocot proves that gene duplication probably occurred very early i.e. before the plants evolved into monocots and dicots. Hence it is assumed that there are likely to be two genes of PVIP present within *N.benthamiana* as well, with one gene being similar to OBE1 and the other gene being similar to OBE2. Phylogenetic analysis suggested that the PVIP is an ancient gene since it is present in all plant groups, that is moss and ferns as well as gymnosperms and angiosperms.

Given that OBE1 and OBE2 are important for apical meristem establishment and maintence, it seems likely that the PVIP_Nb gene is quite important for the development of the plant and is not just involved in viral multiplication and movement. Indeed, the PVIP_Nb gene is expressed to varying degrees in many healthy tissues and is responsive to changes in light levels in uninfected plants. Why potyviruses highjack this particular protein for its own use is unclear. It has been suggested that the virus uses PVIP to assist its movement (Dunoyer et al, 2004). The viral VPg protein that interacts with PVIP is nuclear targetted. Predictions of PVIP's function suggest it is also present in the nucleus. The location of the interaction between PVIP and VPg is unknown and may take place in the nucleus. Given the virus replicates in the cytoplasm and moves through the plasmodesmata makes it unclear how the virus can be using PVIP for movement. There may be an as yet unidentified nuclear located process important to the virus.

References

The Basics: Northern Analysis. (2010). Retrieved 3 June, 2010, from http://ambion.com/techlib/basics/northerns/index.html

Gene Expression Measurement. (2010). Retrieved 1 June, 2010, from http://www.news-medical.net/health/Gene-Expression-Measurement.aspx

Aljanabi, S. M. & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques *Nucleic Acids Research 25* (22), 4692-4693.

Altschul, S. F., W. Gish, W., W. Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol, 215*, 403-410.

Aukerman, M. J., Schmidt, R. J., Burr, B. & Burr, F. A. (1991). An arginine to lysine substitution in the bZIP domain of an opaque-2 mutant in maize abolishes specific DNA binding. *Genes and Development, 5*(2), 310-320.

Baerson, S. R. & Lamppa, G. K. (1993). Developmental regulation of an acyl carrier protein gene promoter in vegetative and reproductive tissues. *Plant Molecular Biology, 22* (2), 255-267.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. et al. (2004). The Pfam protein families database. *Nucleic Acids Research, 32*, 138-141.

Beck, S., O'Keeffe, T., Coull, J. M. & Koster, H. (1989). Chemiluminescent detection of DNA: application for DNA sequencing and hybridization. *Nucleic Acids Research, 17*(13), 5115-5123.

Bernard, V., Brunaud, V. & Lecharny, A. (2010). TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics, 12*(1).

BISTI. (2000). *NIH WORKING DEFINITION OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY*. Retrieved 3 June, 2010, from http://www.bisti.nih.gov/docs/CompuBioDef.pdf

Blazej, R. G., Paegel, B. M. & Mathies, R. A. (2003). Polymorphism Ratio Sequencing: A New Approach for Single Nucleotide Polymorphism Discovery and Genotyping. *Genome Res, 13*(2), 287–293.

Born, T. L. & Miyada, C. G. (1991). Stained colonies facilitate alignment in a nonradioactive colony hybridization. *BioTechniques, 10*(4), 480-481.

Brendel, V. & Zhu, W. (2002). Computational modeling of gene structure in *Arabidopsis thaliana. Plant Molecular Biology, 48*, 49-58.

Browne, E. P., Li, J., Chong, M. & Littman, D. R. (2005). Virus-host interactions: new insights from the small RNA world. *Genome Biology, 6*(238), 1-4.

Bustin, S. A. & Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of biomolecular techniques, 15* (3), 155-166.

Carrington, J. C., Freed, D. D. & Leinicke, A. J. (1991). Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein. *Plant Cell, 3*(9), 953-962.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A. et al. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics, 21*(13), 2933-2942.

Chatterjee, M., Banerjee, A. K. & Hannapel, D. J. (2007). A BELL1-like gene of potato is light activated and wound inducible. *Plant Physiology, 145*, 1435-1443.

Chern, M.-S., Bobb, A. J. & Bustos, M. M. (1996). The regulator of MAT2 (ROM2) protein binds to early maturation promoters and represses PvALF-activated transcription. *Plant Cell, 8*(2), 305-321.

Civán, P. & Svec, M. (2009). Genome-wide analysis of rice (Oryza sativa L. subsp. japonica) TATA box and Y Patch promoter elements. *Genome, 52*(3), 294-297.

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature, 227*, 561-563.

Deák, M., Horváth, G. V., Davletova, S., Török, K., Sass, L., Vass, I. et al. (1999). Plants ectopically expressing the iron-binding protein, ferritin, are tolerant to oxidative damage and pathogens. *Nat Biotechnol, 17*(2), 192-196.

Dellaporta, S. L., Wood, J. D. & Hicks, J. B. (1983). A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter, 1* (4), 19-21.

Densmore, L. D. (2001). Phylogenetic inference and parsimony analysis. *Methods in molecular biology, 176*, 23-36.

Do, C. B. & Katoh, K. (2008). Protein Multiple Sequence Alignment. *Methods in Molecular Biology, 484*, 379-413.

Doolittle, W. F. (1999). Phylogenetic Classification and the Universal Tree. *Science, 284*(5423), 2124-2128.

Dorak, M. T. (2006). *Real-time PCR*. Retrieved 22 June, 2010, from http://www.dorak.info/genetics/realtime.html

Drummond, A., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R. et al. (2009). *Geneious v4.7*, from http://www.geneious.com/

Dunoyer, P., Thomas, C., Harrison, S., Revers, F. & Maule, A. (2004). A cysteine-rich plant protein potentiates *Potyvius* movement through an interaction with the virus genome linked protein Vpg. *Journal of Virology, 78*(5), 2301-2309.

Duprat, A. M., Caranta, C., Revers, F. E., Menand, B., Browning, K. S. & Robaglia, C. (2002). The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant Journal, 32*(6), 927-934.

EMBL-EBI. *EMBOSS Pairwise Alignment Algorithms*. Retrieved 21 Sept, 2009, from http://www.ebi.ac.uk/Tools/emboss/align/index.html

Foster, R., Takeshi, I. & Chua, N.-H. (1994). Plant bZIP proteins gather at ACGT elements. *FASEB Journal, 8*(2), 192-200.

Gingerich, D. J., Gagne, J. M., Salter, D. W., Hellman, H., Estelle, M., Ma, L. et al. (2005). Cullins 3a and 3b assemble with members of the Broad Complex/Tramback/Bric-a-Brac (BTB) protein family to form essential Ubiquitin-protein ligases (E3s) in Arabidopsis. *The Journal of Biological Chemistry, 280*(19), 18810-18821.

Gingras, A.-C., Raught, B. & Sonenberg, N. (1999). eIF4 initiation factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. *Annual Review of Biochemistry, 68*, 913-963.

Ginzinger, D. G. (2002). Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream *Experimental Hematology 30*(6), 503-512.

Grainger, J. (2008). *Virus Diseases of Plants*: Read Books.

Gramene. (2009). *A Resource for Comparative Grass Genomics*. Retrieved 3 Nov, 2009, from http://www.gramene.org/

Grasser, K. D. (1998). HMG1 and HU proteins: architectural elements in plant chromatin. *Trends in Plant Science, 3*(7), 260-263.

Guilfoyle, T. J. (1997). The structure of plant gene promoters. *Genet. Engin, 19*, 15-47.

Guilfoyle, T. J., Ulmasov, T. & Hagen, G. (1998). The ARF family of transcription factors and their role in plant hormone-responsive transcription. *Cellular and Molecular Life Sciences, 54*(7), 619-627.

Han, Y.-Y., Ming, F., Wang, J.-W. & Guo, B. (2007). Nested Polymerase Chain Reactions: An effective method for cloning of ful-length sequences of Chalcone synthase. *Plant Mol Biol, 25*, 63-69.

Harter, K., Kircher, S., Frohnmeyer, H., Krenz, M., Nagy, F. & Schafer, E. (1994). Light-regulated modification and nuclear translocation of cytosolic G-box binding factors in parsley. *Plant Cell, 6*(4), 545-559.

Hartl, D. L. & Jones, E. W. (2005). *Genetics: analysis of genes and genome* (6 ed.): Rober W. Holland, Jr.

Hartwell, L. H., Hood, L., Goldberg, M. L., Reynolds, A. E., Silver, L. M. & Veres, R. C. (2000). *Genetics: from genes to genomes* (1 ed.): James M. Smith.

Heil, M. & Bostock, R. M. (2002). Induced systemic resistance (ISR) against pathogens in the context of induced plant defences. *Annals of Botany, 89*, 503-512.

Hermann, S. R., Harding, R. M. & Dale, J. L. (2001). The banana actin 1 promoter drives near-constitutive transgene expression in vegetative tissues of banana (Musa spp.). *Plant Cell Rep, 20*, 525-530.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G. et al. (2005). Whole-Genome Patterns of Common DNA Variation in Three Human Populations. *Science, 307*(5712), 1072 - 1079.

Hu, R., Fan, C., Li, H., Zhang, Q. & Fu, Y.-F. (2009). Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Molecular Biology, 10*(93), 1-12.

Huang, G., Zhang, L. & Birch, R. G. (2000). Rapid amplification and cloning of Tn5 flanking fragments by Inverse PCR. *Letters in Applied Microbiology, 31*, 149-153.

Huggett, J., Dheda, K., Bustin, S. & Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity, 6* (4), 279-284.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences, 23*(10), 403-405.

Jin, H. & Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. . *Plant Molecular Biology, 41*, 577-585.

Kadare, G. & Haenni, A.-L. (1997). Virus-Encoded RNA Helicases. *American Society for Microbiology, 71*(4), 2583-2590.

Kerstetter, R., Vollbrecht, E., Lowe, B., Veit, B., Yamaguchi, J. & Hake, S. (1994). Sequence analysis and expression patterns divide the maize knotted1-like homeobox genes into two classes. *Plant Cell, 6*(12), 1877-1887.

Killian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O. et al. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *The Plant Journal, 50*, 347-363.

Kim, S.-R., Choi, J.-L., Costa, M. A. & An, G. (1992). Identification of G-box sequence as an essential element for methyl jasmonate response of potato proteinase inhibitor II promoter. *Plant Physiol, 99*, 627-631.

Korfhage, U., Trezzini, G. F., Meier, I., Hahlbrock, K. & Somssich, I. E. (1994). Plant homeodomain protein involved in transcriptional regulation of a pathogen defense-related gene. *Plant Cell 6*(5), 695-708.

Kornberg, R. D. (1996). RNA Polymerase II transcription control. *TIBS, 21*, 325-326.

Kreps, J. A., Wu, Y., Chang, H.-S., Zhu, T., Wang, X. & Harper, J. F. (2002). Transcriptome changes for Arabidopsis in response to salt, osmotic and cold stress. *Plant Physiology, 130*, 2129-2141.

Kristiansson, E., Thorsen, M., Tamas, M. J. & Nerman, O. (2009). Evolutionary forces act on promoter length: Identification of enriched cis-regulatory elements. *Mol Biol Evol, 26*(6), 1299-1307.

Kurepin, L. V., Emery, R. J. N., Pharis, R. P. & Reid, D. M. (2007). The interaction of light quality and irradiance with gibberellins, cytokinins and auxin in regulating growth of *Helianthus annuus* hypocotyls. *Plant, Cell and Environment, 30*, 147-155.

Kusnetsov, V., Landsberger, M., Meurer, J. & Oelmuller, R. (1999). The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids. *The Journal of Biological Chemistry, 274*(50), 36009-36014.

Lam, E. (1995). Domain analysis of the plant DNA -binding protein GT1a: Requirement of four putative alpha-helices for DNA binding and identification of a novel oligomerization region. *Molecular and Cellular Biology, 15*(2), 1014-1020.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*, 2947-2948.

Laxmi, A., Pan, J., Morsy, M. & Chen, R. (2007). Light plays an essential role in intracellular distribution of Auxin efflux carrier PIN2 in *Arabidopsis thaliana. PLoS ONE, 1*, 1-11.

LeBrasseur, N. D. (2010). *Blotting Analysis*. Retrieved 28 May, 2010, from http://science.jrank.org/pages/971/Blotting-Analysis.html

Leonard, S., Plante, D., Whittmann, S., Daigneault, N., Fortin, M. G. & Laliberte, J.-F. (2000). Complex formation between Potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *Journal of Virology, 74*(17), 7730-7737.

Lincoln, C., J. Long, J., Yamaguchi, J., Serikawa, K. & Hake, S. (1994). A Knotted1-like Homeobox gene in Arabidopsis is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *The Plant Cell, 6*, 1859-1876.

Liu, L., White, M. J. & MacRae, T. H. (1999). Transcription factors and their genes in higher plants. *Eur. J. Biochem, 262*, 247-257.

Lo, K. & Smale, S. T. (1996). Generality of a functional initiator consensus sequence. *Gene, 182*, 13-22.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. & Darnell, J. (2000). *Molecular Cell Biology* (4 ed.): W. H. Freeman.

Lovdal, T. & Lillo, C. (2009). Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Analytical Biochemistry, 387*, 238-242.

Lozsa, R., Csorba, T., Lakatos, L. & Burgyan, J. (2008). Inhibition of 3' modification of small RNAs in virus-infected plants require spatial and temporal co-expression of small RNAs and viral silencing-suppressor proteins. *Nucleic Acids Research, 36*(12), 4099-4107.

Lutz, D. E. (2003). *Southern/Northern Analysis*. Retrieved 28 May, 2010, from http://homepages.strath.ac.uk/~dfs99109/BB211/RecombDNAtechlect2.html

Malecova, B., Gross, P., Boyer-Guittaut, M., Yavuz, S. & Oelgeschlager, T. (2007). The initiator core promoter element antagonizes repression of TATA-directed transcription by negative cofactor NC2. *The Journal of Biological Chemistry, 282*(34), 14767-24776.

Manfield, I. W., Devlin, P. F., Jen, C.-H., Westhead, D. R. & Gilmartin, P. M. (2007). Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family. *Plant Physiology, 143*, 941-958.

March-Diaz, R., Garcia-Dominguez, M., Florencio, F. J. & Reyes, J. C. (2007). SEF, a new protein required for flowering repression in Arabidopsis interacts with PIE1 and ARP6. *Plant Physiology, 143*, 893-901.

Matys, V., Fricke, E., Geffers, R., Gobling, E., Haubrock, M., Hehl, R. et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Research, 31*(1), 374-378.

Matzke, M., Daxinger, W. A. K., Papp, I., Mette, M. F. & Matzke, A. J. M. (2004). Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochimica et Biophysica Acta, 1677*, 129–141.

Morris, E. R., Chevalier, D. & Walker, J. C. (2006). DAWDLE, a Forkhead-associated domain gene, regulates multiple aspects of plant development. *Plant Physiology, 141*, 932-941.

Moyroud, E., Reymond, M. C. A., Hames, C., Parcy, F. & Scutt, C. P. (2009). The analysis of entire gene promoters by surface plasmon resonance. *The Plant Journal, 59*, 851-858.

Murray, M. G. & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research, 8*(19), 4321-4325.

Myers, E. W. (1995). Toward simplifying and accurately formulating fragment assembly. *Journal of computational biology: a journal of computational molecular cell biology, 2*(2), 275-290.

Nagy, P. D. (2008). Yeast as a model host to explore plant virus-host interactions. *The Annual Review of Phytopathology, 46*, 217-242.

Narlikar, G. J., Fan, H.-Y. & Kingston1, R. E. (2002). Cooperation between complexes review that regulate chromatin structure and transcription. *Cell, 108*, 475-487.

NCBI. *Basic Loal Alignment Sarch Tool*. Retrieved 17 Aug, 2009, from http://blast.ncbi.nlm.nih.gov/Blast.cgi

Neuhaus-Url, G. & Neuhaus, G. (1993). The use of nonradioactive digoxigenin chemiluminescent technology for plant genomic Southern blot hybridization: a comparison with radioactivity. *Transgenic Research, 2*, 115-120.

Ng, J. C. K. & Falk, B. W. (2006). Virus-Vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annual Review of Phytopathology, 44*, 183-212.

Ng, P. C. & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res, 11*(5), 863-874.

Nicaise, V. r., Gallois, J.-L., Chafiai, F., Allen, L. M., Schurdi-Levrau, V. r., Browning, K. S. et al. (2007). Coordinated and selective recruitment of eIF4E and eIF4G factors for potyvirus infection in Arabidopsis thaliana. *Federation of European Biochemical Societies, 581*, 1041-1046.

Ochman, H., Gerber, A. S. & Hartl, D. L. (1988). Genetic applications of an Inverse Polymerase Chain Reaction. *Genetics, 120*, 621-623.

Phillips, M. A., D'Auria, J. C., Luck, K. & Gershenzon, J. (2009). Evaluation of candidate reference genes for real-time quantitative PCR of plant samples using purified cDNA as template. *Plant Mol Biol Rep, 27*, 407-416.

Plochocka, D., Welnicki, M., Zielenkiewicz, P. & Ostoja-Zagorski, W. (1996). Three-dimensional model of the potyviral genome-linked protein. *Proc. Natl. Acad. Sci.USA, 93*, 12150-12154.

Pop, M., Phillippy, A., Delcher, A. L. & Salzberg, S. L. (2004). Comparative genome assembly. *Briefings in Bioinformatics, 5*(3), 237-248.

Promoter. (2009). *Molecular Biology Web Book*. Retrieved September 28, 2009, from http://www.web-books.com/MoBio/Free/Ch4C1.htm

Purgganan, M. D. & Wessler, S. R. (1994). Molecular evolution of the plant R regulatory gene family. *Genetics, 138*(3), 849-854.

Radonic, A., Thulke, S., Mackay, I. M., Landtd, O., Siegert, W. & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications, 313*(4), 856-862.

Rancour, D. M., Park, S., Knight, S. D. & Bednarek, S. Y. (2004). Plant UBX domain-containing protein 1, PUX1, regulates the oligomeric structure and activity of Arabidopsis CDC48. *The Journal of Biological Chemistry, 279*(52), 54264-54274.

Reyes, J. C., Muro-Pastor, M. I. & Florencio, F. J. (2004). The GATA family of transcription factors in Arabidopsis and Rice. *Plant physiology, 134*, 1718-1732.

Rhee, S. Y., Dickerson, J. & Xu, D. (2006). Bioinformatics and its application in plant biology. *Annaul Reviews of Plant Biology, 57*, 335-360.

Riechmann, J. L., Lain, S. & García, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology, 73*(1), 1-16.

Robaglia, C. & Caranta, C. (2006). Translation initiation factors: a weak link in plant RNA virus infection. *TRENDS in Plant Science, 11*(1), 40-45.

Robson, F., Costa, M. M. R., Hepworth, S. R., Vizir, I., eiro, M. P., Reeves, P. H. et al. (2001). Functional importance of conserved domains in the fowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *The Plant Journal, 28*(6), 619-631.

Rogers, D. A. (2008). *Southern blotting: Probe labeling & Detection*. Retrieved 28 May, 2010, from http://www.csus.edu/indiv/r/rogersa/Bio181/probelabeling.pdf

Roudet-Tavert, G., Michon, T., Walter, J., Delaunay, T., Redondot, E. & Gall, O. L. (2007). Central domain of a potyvirus VPg is involved in the interaction with the host translation initiation factor eIF4E and the viral protein HcPRO. *Journal of General Virology, 88*, 1029-1033.

Roychoudhary, A. & Sengupta, D. N. (2009). The promoter-elements of some abiotic stress-inducible genes from cereals interact with a nuclear protein from tobacco. *Biologia Plantarum, 53*(3), 583-587.

Rushton, P. J., Reinstadler, A., Lipka, V., Lippok, B. & Somssich, I. E. (2002). Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen and wound induced signaling. *The Plant Cell, 14*, 749-762.

Russell, P. J. (2001). *iGenetics*. (1 ed.): Benjamin Cummings.

Safaee, N. M., Watkinson, J. & Gillaspy, G. E. (2007). An analysis of the PHD-finger domain in arabidopsis and rice using bioinformatics, functional genomics and biochemistry. *The FASEB Journal., 21*(781.11), 1-8.

Sahooa, D. K., Ranjana, R., Kumara, D., Kumara, A., Sahooa, B. S., Rahab, S. et al. (2009). An alternative method of promoter assessment by confocal laser scanning microscopy. *Journal of Virological Methods, 16*, 114-121.

Saiga, S., Furumizu, C., Yokoyama, R., Kurata, T., Sato, S., Kato, T. et al. (2008). The Arabidopsis OBERON1 and OBERON2 genes encode plant homeodomain finger proteins and are required for apical meristem maintenance. *Development, 135*, 1751-1759.

Sanfacon, H. (2005). Replication of positive strand RNA viruses in plants: contact points between plant and virus components. *Canadian Journal of Botany, 83*, 1529-1549.

Schaad, M. C., Anderberg, R. J. & Carrington, J. C. (2000). Strain-Specific interaction of the Tobacco Etch Virus NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. *Virology, 273*, 300-306.

Schena, M. & Davis, R. W. (1994). Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. *Proceedings of the National Academy of Sciences of the United States of America, 91*(18), 8393-8397.

Schug, J. (2003). *Current Protocols in Bioinformatics*: J. Wiley and Sons.

Schwabish, M. A. & Struhl, K. (2004). Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA Polymerase II. *Molecular and Cellular Biology, 24*(23), 10111-10117.

Sekinger, E. A., Moqtaderi, Z. & Struhl, K. (2005). Intrinsic Histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Molecular Cell, 18*, 735-748.

Shi, W. & Zhou, W. (2006). Frequency distribution of TATA Box and extension sequences on human promoters. *BMC Bioinformatics, 7*, 1-12.

Singh, V. K., Govindarajan, R., Naik, S. & Kumar, A. (2000). The effect of hairpin structure on PCR amplification efficiency. *Molecular Biology Today, 1*(3), 67-69.

Sommerfelt, H., Grewal, H. M. S., Gaastra, W., Svennerholm, A.-M. & Bhan, M. K. (1992). Use of Nonradioactive DNA Hybridization for Identification of Enterotoxigenic Escherichia coli Harboring Genes for Colonization Factor Antigen I, Coli Surface Antigen 4, or Putative Colonization Factor 0166. *Journal of Clinical Microbiology, 30*(7), 1823-1828.

Ståhlberg, A., Kubista, M. & Pfaffl, M. W. (2004). Comparison of reverse transcriptases in gene expression analysis. *Clinical Chemistry 50* (9), 1678-1680.

The_Arabidopsis_Genome_Initiative. (2000). Analysis of the genome sequence of the fowering plant *Arabidopsis thaliana. Nature, 408*, 796-815.

Thomas, C. L., Schmidt, D., Bayer, E. M., Dreos, R. & Maule, A. J. (2009). Arabidopsis plant homeodomain finger proteins operate downstream of auxin accumulation in specifying the vasculature and primary root meristem. *The Plant Journal, 59*, 426-436.

Triglia, T. (2000). Inverse PCR (IPCR) for obtaining promoter sequence. *Methods in Molecular Biology, 130*, 79-83.

Urcuqui-Inchima, S., Haenni, A.-L. & Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Research, 74*, 157-175.

Urrutia, A. O. & Hurst, L. D. (2003). The signature of selection mediated by expression on human genes. *Genome Research, 13*, 2260-2264.

Vandenborre, G., Damme, E. J. M. V. & Smagghe, G. (2009). *Nicotiana tabacum* agglutinin expression in response to different biotic challengers. *Anthropod-Plant Interactions, 3*, 193-202.

Venter, M. (2007). Synthetic promoters: genetic control through cis engineering. *TRENDS in Plant Science, 12*(3), 118-124.

Vettese-Dadey, M., A.Grant, P., Hebbes, T. R., Crane-Robinson, C., Allis, C. D. & L.Workman, J. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *The EMBO Journal, 15*(10), 2508-2518.

Vinogradov, A. E. (2004). Compactness of human housekeeping genes: selection for economy or genomic design? *TRENDS in Genetics, 20*, 248-253.

Wan, X.-F. & Xu, D. (2005). Computational Methods for Remote Homolog Identification. *Current Protein and Peptide Science, 6*(6), 527-546.

Wang, X., Ullah, Z. & Grumet, R. (2000). Interaction between Zucchini Yellow Mosaic Potyvirus RNA-Dependent RNA Polymerase and Host Poly-(A) Binding Protein. *Virology, 275*, 433-443.

Wang, Z., Chen, C., Jiang, Y. X., Han, Y., Xu, Z. & Chong, K. (2004). A practical vector for efficient knockdown of gene expression in rice (Oryza sativa L.). *Plant Molecular Biology Reporter, 22*, 409-417.

Wheelan, S. J., Murillo, F. M. & Boeke, J. D. (2008). The incredible shrinking world of DNA microarrays. *Mol Biosyst, 4*(7), 726-732.

Winer, J., Jung, C. K. S., Shackel, I. & Williams, P. M. (1999). Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro. *Analytical Biochemistry 270* (1), 41-49.

Wong, M. L. & Medrano, J. F. (2005). Real-time PCR for mRNA quantification. *BioTechniques, 39*, 75-85.

Yamamoto, Y. Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M. et al. (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics*.

Yamazaki, K.-I., Katagiri, F., Imaseki, H. & Chua, N.-H. (1990). TGA1a, a tobacco DNA-binding protein, increases the rate of initiation in a plant in vitro transcription system. *Proc.National Acad, 87*, 7035-7039.

Yanagisawa, S. & Schmidt, R. J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. . *The Plant Journal, 17*(2), 209-214.

Yona, G. & Levitt, M. (2002). Within the Twilight Zone: A Sensitive Profile-Profile Comparison Tool Based on Information Theory. *J. Mol. Biol, 315*, 1257-1275.

Zhang, M. Q. (2002). Computational prediction of eukaryotic protein-coding genes. *Nature, 3*, 698-709.

Zhou, D.-X. (1999). Regulatory mechanism of plant gene transcription by GT-elements and GT-factors. . *Trends in plant science, 4*, 1360-1385.

Zhu, Q., Dabi, T. & Lamb, C. (1995). TATA Box and initiator functions in the accurate transcription of a plant minimal promoter in vitro. *The Plant Cell, 7*, 1681-1689.

Appendices

Table A1. The BLASTp result of PVIP protein from *Nicotiana benthamiana*. The table gives the information regarding the closest homologue of the PVIP_ *Nicotiana Benthamiana* with respect to its E-value, its protein homologue, the species the homologue is present in and the common name of the species (Source: NCBI)

Accession number	Gene name	Product	E-value	Common name
AAP22954.1	*Nicotiana benthamiana*	Potyvirus VPg interacting protein	0	
XP_002270767.1	*Vitis vinifera*	PREDICTED: hypothetical protein	0	Grapes
XP_002321182.1	*Populus trichocarpa*	predicted protein	0	Poplar
XP_002531698.1	*Ricinus communis*	protein binding protein, putative	0	Castor oil plant
XP_002301583.1	*Populus trichocarpa*	predicted protein	0	Poplar
ABZ11028.1	*Arachis hypogaea*	Potyvirus VPg interacting protein	0	Peanut
NP_199627.1	*Arabidopsis thaliana*	OBE2 (OBERON2); protein binding / zinc ion binding	0	Mouse ear cress
BAH20018.1	*Arabidopsis thaliana*	AT5G48160	0	Mouse ear cress
BAE99299.1	*Arabidopsis thaliana*	hypothetical protein	0	Mouse ear cress
NP_001066881.1	*Oryza sativa*	Potyvirus VPg interacting protein, putative	0	Rice
NP_001140644.1	*Zea mays*	hypothetical protein	0	Maize
ABK24532.1	*Picea sitchensis*	unknown	0	Sitka spruce
AAP22955.1	*Pisum sativum*	Potyvirus VPg interacting protein	0	Pea
NP_566320.1	*Arabidopsis thaliana*	OBE1 (OBERON1); protein binding / zinc ion binding	0	Mouse ear cress

NP_001147896.1	*Zea mays*	potyvirus VPg interacting protein	0	Maize
ACG29051.1	*Zea mays*	potyvirus VPg interacting protein	0	Maize
AAS67374.1	*Solanum lycopersicum*	CONSTANS interacting protein 6	0	Tomato
XP_002451176.1	*Sorghum bicolor*	hypothetical protein	0	Sorghum
CBI28423.1	*Vitis vinifera*	unnamed protein product	3.00E-176	Grapes
XP_002271018.1	*Vitis vinifera*	PREDICTED: hypothetical protein	6.00E-156	Grapes
CBI41083.1	*Vitis vinifera*	unnamed protein product	1.00E-155	Grapes
XP_001760086.1	*Physcomitrella patens subsp. Patens*	predicted protein	6.00E-144	Moss
XP_001758332.1	*Physcomitrella patens subsp. Patens*	predicted protein	3.00E-137	Moss
XP_002313313.1	*Populus trichocarpa*	predicted protein	2.00E-81	Poplar
ACU21522.1	*Glycine max*	unknown	2.00E-81	soy bean
EEC68513.1	*Oryza sativa Indica Group*	hypothetical protein	1.00E-80	Rice
ABA94920.1	*Oryza sativa japonica*	Potyvirus VPg interacting protein, putative [	1.00E-80	Rice
XP_002527544.1	*Ricinus communis*	protein binding protein, putative	4.00E-80	Castor oil plant
XP_002517804.1	*Ricinus communis*	protein binding protein, putative	1.00E-78	Castor oil plant
XP_002274296.1	*Vitis vinifera*	PREDICTED: hypothetical protein	3.00E-78	Grapes
XP_002299935.1	*Populus trichocarpa*	predicted protein	5.00E-77	Poplar
XP_002281609.1	*Vitis vinifera*	PREDICTED: hypothetical protein	1.00E-75	Grapes
XP_002468196.1	*Sorghum bicolor*	hypothetical protein	2.00E-74	Sorghum

NP_001049504.1	*Oryza sativa japonica*	tropomyosin, putative, expressed	4.00E-74	Rice
EAY89192.1	*Oryza sativa Indica*	hypothetical protein	4.00E-74	Rice
XP_002315843.1	*Populus trichocarpa*	predicted protein	7.00E-73	Poplar
NP_001147112.1	*Zea mays*	CONSTANS interacting protein 6	1.00E-72	Maize
ACG25487.1	*Zea mays*	CONSTANS interacting protein 6	2.00E-72	Maize
AAO64172.1	*Arabidopsis thaliana*	unknown protein	1.00E-70	Mouse ear cress
NP_850743.1	*Arabidopsis thaliana*	unknown protein	2.00E-70	Mouse ear cress
EEE58601.1	*Oryza sativa Japonica Group*	hypothetical protein	2.00E-70	Rice
NP_191909.1	*Arabidopsis thaliana*	unknown protein	2.00E-70	Mouse ear cress
NP_001049395.1	*Oryza sativa japonica*	expressed protein	3.00E-70	Rice
EAY89060.1	*Oryza sativa Indica*	hypothetical protein	3.00E-70	Rice
XP_002311544.1	*Populus trichocarpa*	predicted protein	1.00E-69	Poplar
NP_563958.1	*Arabidopsis thaliana*	unknown protein	6.00E-69	Mouse ear cress
AAF79245.1	*Arabidopsis thaliana*	F10B6.14	3.00E-68	Mouse ear cress
XP_002468267.1	*Sorghum bicolor*	hypothetical protein	5.00E-68	Sorghum
ACU20268.1	*Glycine max*	unknown	6.00E-63	soy bean
BAD94916.1	*Arabidopsis thaliana*	putative protein	3.00E-62	Mouse ear cress
BAF29900.2	*Oryza sativa Japonica Group*	Os12g0514400	1.00E-60	Rice
CBI24921.1	*Vitis vinifera*	unnamed protein product	2.00E-52	Grapes

CAJ44362.1	*Malus x domestica*	hypothetical protein	4.00E-43	Apple

Table A2. The BLASTn result of the accession number AY271742 i.e. PVIP nucleotide sequence from *N.benthamiana*. The table gives the information regarding the closest homologue of PVIP_ *Nicotiana Benthamiana* with respect to its E-value, its protein homologue, the species the homologue is present in and the common name of the species (Source: NCBI).

Accession number	*Gene name*	Product	E-value	Common name
AY271742.1	*Nicotiana benthamiana*	Potyvirus VPg interacting protein mRNA, complete cds	0	
AK319266.1	*Solanum lycopersicum*	cDNA, clone: LEFL1031CF10, HTC in leaf	0	Tomato
AC189287.2	*Solanum lycopersicum*	chromosome 5 clone C05HBa0166A02, complete sequence	0	Tomato
XM_002270731.1	*Vitis vinifera*	PREDICTED: hypothetical protein LOC100259651 (LOC100259651), mRNA	0	Grape
XM_002321146.1	*Populus trichocarpa*	predicted protein, mRNA	0	Poplar
XM_002531652.1	*Ricinus communis*	protein binding protein, putative, mRNA	0	Castor oil plant
XM_002301547.1	*Populus trichocarpa*	predicted protein, mRNA	0	Poplar
EU408476.1	*Arachis hypogaea*	Potyvirus VPg interacting protein (VPgB) mRNA, complete cds	0	Peanut
AM424745.2	*Vitis vinifera*	contig VV78X024126.8, whole genome shotgun sequence	6.00E-179	Grape
AY271743.1	*Pisum sativum*	Potyvirus VPg interacting protein mRNA, partial cds	7.00E-172	Pea

XM_002270982.1	*Vitis vinifera*	PREDICTED: hypothetical protein LOC100255772 (LOC100255772), mRNA	1.00E-168	Grape
AK320037.1	*Solanum lycopersicum*	cDNA, clone: LEFL1004BE08, HTC in leaf	3.00E-158	Tomato
EZ172572.1	*TSA: Artemisia annua strain*	Artemis Contig31978, mRNA sequence	2.00E-153	Absinth Wormwood
AY490248.1	*Lycopersicon esculentum*	CONSTANS interacting protein 6 (CIP6) mRNA, partial cds	7.00E-153	Tomato
AC215624.1	*Populus trichocarpa*	clone POP006-B08, complete sequence	8.00E-152	Poplar
AK317672.1	*Arabidopsis thaliana*	AT5G48160 mRNA, complete cds, clone: RAFL21-56-J04	2.00E-141	Mouse ear cress
NM_124190.3	*Arabidopsis thaliana*	OBE2 (OBERON2); protein binding / zinc ion binding (OBE2) mRNA, complete cds	2.00E-141	Mouse ear cress
NM_001036954.1	*Arabidopsis thaliana*	OBE2 (OBERON2); protein binding / zinc ion binding (OBE2) mRNA, complete cds	2.00E-141	Mouse ear cress
AK317346.1	*Arabidopsis thaliana*	AT5G48160 mRNA, complete cds, clone: RAFL24-15-I08	2.00E-140	Mouse ear cress
AK227275.1	*Arabidopsis thaliana*	mRNA for hypothetical protein, complete cds, clone: RAFL11-09-N19	2.00E-140	Mouse ear cress
NM_111657.3	*Arabidopsis thaliana*	OBE1 (OBERON1); protein binding / zinc ion binding (OBE1) mRNA, complete cds	9.00E-139	Mouse ear cress
AY113947.1	*Arabidopsis thaliana*	unknown protein (At3g07780) mRNA, complete cds	9.00E-139	Mouse ear cress
AY045899.1	*Arabidopsis thaliana*	unknown protein (At3g07780) mRNA, complete cds	9.00E-139	Mouse ear cress

DQ059085.1	*Arabidopsis thaliana*	PHD family protein (At3g07780) mRNA, complete cds	9.00E-139	Mouse ear cress
AM474144.1	*Vitis vinifera,*	whole genome shotgun sequence, contig VV78X029207.10, clone ENTAV 115	9.00E-120	Grape
AP010534.1	*Lotus japonicus*	genomic DNA, clone: LjT44E18, TM1828, complete sequence	6.00E-116	
AK241392.1	*Oryza sativa Japonica Group*	cDNA, clone: J065157H03, full insert sequence	1.00E-104	Rice
NM_001073413.1	*Oryza sativa japonica*	mRNA, complete cds	1.00E-104	Rice
EF085225.1	*Picea sitchensis*	clone WS0271_B11 unknown mRNA	6.00E-103	Sitka spruce
AC174287.13	*Medicago truncatula*	clone mth2-32a8, complete sequence	7.00E-102	Barrel clover
CT571263.1	*Medicago truncatula*	chromosome 5 clone mte1-7c20, COMPLETE SEQUENCE	7.00E-102	Barrel clover
BT000803.1	*Arabidopsis thaliana*	clone At3g07780 mRNA sequence	3.00E-101	Mouse ear cress
BT067495.1	*Zea mays*	full-length cDNA clone ZM_BFc0066J18 mRNA, complete cds	3.00E-100	Maize
NM_001147172.1	*Zea mays*	full-length cDNA clone ZM_BFc0002B06 mRNA, complete cds	3.00E-100	Maize
EZ345161.1	*TSA: Artemisia annua strain Uganda*	Contig12377, mRNA sequence	3.00E-94	Absinth Wormwood
EZ301413.1	*TSA: Artemisia annua strain Madagascar*	Contig7298, mRNA sequence	1.00E-93	Absinth Wormwood
BT118139.1	*Picea glauca*	clone GQ04002_N11 mRNA sequence	2.00E-91	White spruce

CU224999.1	*Populus*	EST from leave	7.00E-83	Poplar
AC009176.5	*Arabidopsis thaliana*	chromosome III P1 MLP3 genomic sequence, complete sequence	3.00E-81	Mouse ear cress
AC013483.7	*Arabidopsis thaliana*	chromosome III BAC F17A17 genomic sequence, complete sequence	3.00E-81	Mouse ear cress
XM_002451131.1	*Sorghum bicolor*	hypothetical protein, mRNA	2.00E-78	Sorghum
AL731888.5	*Oryza sativa japonica*	BAC OJ1118_C12 of library Monsanto from chromosome 12 of cultivar Nipponbare	2.00E-77	rice
AB023039.1	*Arabidopsis thaliana*	genomic DNA, chromosome 5, P1 clone:MIF21	8.00E-70	Mouse ear cress
BT097348.1	*Soybean clone*	JCVI-FLGm-16G17 unknown mRNA	3.00E-62	soya bean
NM_001154424.1	*Zea mays*	potyvirus VPg interacting protein (LOC100281506), mRNA	3.00E-62	Maize
EU957697.1	*Zea mays*	clone 1610438 potyvirus VPg interacting protein mRNA, complete cds	3.00E-62	Maize
EU956933.1	*Zea mays*	clone 1577747 potyvirus VPg interacting protein mRNA, complete cds	3.00E-62	Maize
EU946173.1	*Zea mays*	clone 299625 mRNA sequence	3.00E-62	Maize
EZ351004.1	*TSA: Artemisia annua strain Uganda*	Contig18220, mRNA sequence	5.00E-60	Absinth Wormwood
AC189305.1	*Brassica rapa subsp. pekinensis*	clone KBrB031O20, complete sequence	2.00E-59	
EZ383068.1	*TSA: Artemisia annua strain Uganda*	Contig16519, mRNA sequence	5.00E-53	Absinth Wormwood

EZ172424.1	*TSA: Artemisia annua strain Artemis*	Contig31830, mRNA sequence	7.00E-52	Absinth Wormwood
EZ298733.1	*TSA: Artemisia annua strain Madagascar*	Contig4618, mRNA sequence	4.00E-48	Absinth Wormwood

List of abbreviations

BLAST	Basic local alignment search tool
BRE	TFIIB recognition element
bZIP	Basic leucine zipper
CBP	CAAT-box binding protein
Ct	Threshold cycle
DCE	Downstream core elements
DNA	Deoxyribose nucleic acid
dNTP	Deoxyribonucleotide triphosphate
DPE	Downstream promoter element
DUF	DNA unwinding factor
EDTA	Ethylenediamine tetra acetic acid
FHA	Forkhead associated domains
GADPH	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	genomic DNA
GFP	Green fluorescent protein
GUS	B-glucoronidase
HPRT	Hypoxanthine-guanine phosphoribosyl transferase
IAA	Indole-3-Acteic acid
INR	Initiator
IPCR	Inverse PCR
LMV	Lettuce mosaic virus
M.S. medium	Murashige and Skoog medium

mRNA	messenger RNA
NLS	Nuclear localisation sequence
OBE	Oberon
ORF	Open reading frame
PCR	Polymerase chain reaction
PHD	Plant homeo domain
PIC	Pre-initiation complex
PVIP	Potyvirus VPg interacting protein
qPCR	Real-time PCR
RdRp	RNA dependant RNA polymerase
RNA	Ribose nucleic acid
rRNA	ribosomal RNA
RT-PCR	Reverse transcriptase PCR
SDS	Sodium dodecyl sulphate
SnRNA	Small nuclear RNA
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TBP	TATA-box binding protein
TEV	Tobacco etch virus
TF	Transcription factor
TFBS	Transcription factor binding site
tRNA	transfer RNA
TSS	Transcription start site
TuMV	Turnip mosaic virus