MOLECULAR FARMING: PRODUCTION OF PHARMACEUTICALS IN TRANSGENIC TOBACCO

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Kafeel Ahmad
Department of Biology
University of Leicester

November 2010

Abstract

Title: Molecular Farming: Production of Pharmaceuticals in Transgenic Tobacco

Kafeel Ahmad

Molecular farming is an experimental application of biotechnology to modify crops in order to produce proteins and chemicals for medicinal and commercial interests. The vast majority in the developing world cannot afford the high cost of therapeutics produced by existing methods. We not only need to produce new therapeutics but also need to produce cheaper versions of the existing ones. Molecular farming could offer a viable option for this growing need for biopharmaceuticals.

Part of the thesis deals with investigating ways to produce DesB30 form of human insulin in transgenic tobacco. The human insulin was synthesized *in vitro* as strep-tag II-mini-insulin fusion protein. Expression of mini-insulin by transgenic tobacco was confirmed by RT-PCR, western blotting and ELISA. However, sufficient levels of purified insulin could not be obtained to carry out further functional assays. Strategies for increasing the yield of insulin by transgenic tobacco are discussed and further increases in yield would need to be developed for this to become a viable and cost effective source of this important pharmaceutical.

The second part of the thesis describes the production of a recombinant microbial polysaccharide in tobacco. Seven type 2 pneumococcal polysaccharide biosynthetic genes were expressed in a single tobacco plant, utilizing the plant Kex2 (Kexin protease 2) like protease system for multiple gene expression. Expression of these genes in transgenic tobacco was confirmed by RT-PCR and western blotting. Correct processing of the expressed proteins by the Kex2 protease system was confirmed. However, *In planta* production of type 2 polysaccharide could not be confirmed mainly as a result of high background from the wild type plant polysaccharide extracts. Strategies to overcome these issues are described. The usefulness of Kex2 protease system for multiple gene expression and metabolic pathways engineering is also discussed.

Dedication

Dedicated to my loving parents

Acknowledgements

All praise is for Allah Who is the most merciful, Knowledgeable and Worthy of all praises. He Knowth what before or after or behind us. And may His blessings, peace and favours descend in perpetuity on our beloved Prophet Muhammad *al-Mustafa*, *Sallallahu 'alaihi wa Sallam* who is mercy for all the worlds.

I am very grateful to many people helping me during my studies. This thesis would not have been possible without their support.

A special thank to my late supervisor Prof. Garry Whitelam for giving me a precious opportunity to join his lab and for his support, time and guidance during the initial stages of my PhD studies.

A big thank to my supervisor Prof. David Twell for taking the responsibilities of my supervision after the death of my supervisor. His sincere support, approachability and valuable suggestions had been of great help to complete the studies.

I also really appreciate all the help and supervision provided by Dr. Ben Maddison (ADAS). I find it difficult to describe in words his commitment and devotion for helping me. I learned all the lab work from him. His friendly nature, approachability and consistence guidance has been the main stimulant to accomplish my work.

Special thanks are extended to Dr. Kevin Gough for his guidance and valuable suggestions.

I also wish to thank Prof. Peter Andrew (Department of Infection, Immunity & Inflammation) for his guidance and kindly providing *Streptococcus pneumoniae* genomic DNA. Further thanks go to Dr. Sinead Drea for her encouragement.

I would like to express my humble gratitude to all the people in the Department of Biology who helped me at some point during my work and whose company has been a great source of happiness, amusement and enlightenment. I am indebted to my lab fellows (research staff, Department of Biology) specially Keith Bishop, Shila Patel, Dr. Helen Rees, Ceinwen Tilley, Dr. Claire Smith (Department of Infection, Immunity &

Inflammation), Dr. Daval Patel, Jonathon Owen and Claire Baker to help me during my stay at the department.

It is a pleasure to thank my friends from Pakistan whose company and support made my life at Leicester full of enjoyment and fulfilment.

Words are lacking to express my gratefulness to my parents, brothers and sister for their love, support and prayers which has been and will be a source of happiness and accomplishment. They provided me with the opportunity to be where I am. Without them, none of this would even be possible.

I am also thankful to the University of Peshawar and Higher Education Commission of Pakistan for providing financial support to perceive my PhD studies.

Contents

List of Figures

List of Tables

Abbreviations

%	Percentage
°C	degrees centigrade
A	adenine
BAP	6-benzylaminopurine
bp	base pair
BSA	bovine serum albumin
CaMV	cauliflower mosaic virus
cDNA	complementary DNA
cm	centimetre
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
ELISA	enzyme-linked immunosorbant assay
g	gram
g	gravity
G	guanine
H	hour
kb	kilobase pair
kDa	kilodalton
l	litre
LB media	Luria Bertani media
M	molar
mg	milligram
Min	minutes
ml	millilitre
mM	millimolar
mRNA	messenger RNA
MS	Murashige and Skoog
NAA	1-naphthaleneacetic acid
ng	nanogram
OD	optical density
p	plasmid
P	significance level
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pg	picogram
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcriptase-PCR
S	second
T	thymine
TBS	Tris Buffered Saline
TEMED	N,N,N’,N’-tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
Triton X-100	t-Octylphenoxypolyethoxyethanol

V	volts
WT	wild type
µg	microgram
µl	microlitre
µM	micromolar
$\chi 2$	Chi-square statistic

1 INTRODUCTION

1.1 Molecular farming

Molecular farming can be described as an experimental application of biotechnology to genetically engineer crops in order to produce proteins and chemicals for pharmaceutical and other commercial interests (Franken *et al*., 1997). Evidence of using plants for medicinal purposes can be traced back as early as the Neanderthal period (about 130,000 years ago), when plants were used to aid healing of wounds (Kleiner, 1995). Romans used willow bark for the treatment of fever which was later identified to contain aspirin (Mewett *et al.*, 2007). During the 16th century medicinal plants were grown for teaching medicine and as a source of treatment for various diseases (Akerele, 1993). The use of plants and plant extracts for medicinal purposes flourished until 17th century when more scientific pharmacological treatments were recommended (Trevelyan, 1993; Winslow & Kroll, 1998). One fourth of currently used medicines still have a plant origin (Winslow and Kroll, 1998).

Genetic engineering has recently opened up new opportunities for using plants as production factories for biopharmaceuticals. Human growth hormone was the first pharmaceutically important protein that was expressed in transgenic tobacco (Barta *et al.*, 1986). Since then transgenic plants expressing vaccines, therapeutics, industrial enzymes, antibodies, nutraceuticals, and other pharmaceutical proteins have been produced (Krebbers *et al*., 1992, Whitelam *et al*., 1993; Ma *et al.*, 1994; Ma *et al.*, 1995; Herbers and Sonnewald, 1999; De Jaeger *et al.*, 2002; Ma *et al.*, 2005; Nuttall *et al.*, 2005; Sexton *et al.*, 2009).

Historically both prokaryotic (Georgiou and Bowden, 1991) and eukaryotic (Harashima, 1994; Lubon *et al*., 1996) systems have been utilized to express recombinant proteins. Prokaryotic production systems are comparatively inexpensive and convenient compared to mammalian systems in terms of technology/equipments required. However, many mammalian proteins require post-translational modifications such as protein glycosylation for their biological activity which cannot be performed by prokaryotic production systems. The use of prokaryotic expression systems is therefore perhaps limited. The cost of producing proteins in mammalian cells is on the other hand very high in terms of

maintaining cell cultures and scale up. In contrast, pathways of protein synthesis and modification in plants are very similar to those in the protein synthesis pathways in animals facilitating protein modification analogous to that seen in animal cells (Cabanes-Macheteau *et al*., 1999). Mass production of heterologous proteins and biopharmaceuticals in plants can be achieved at much lower costs and contamination or co-purification of human or animal pathogens is also avoided (Evangelista *et al*., 1988; Verwoerd *et al*., 1995; Whitelam, 2000; Ziegler *et al*., 2000; Giddings *et al*., 2000). Additionally, plants possess natural protein storage organs and their seeds are easily distributed, allowing local production (Whitelam *et al*., 1993; Whitelam, 2000). A comparison of recombinant protein production in plants, yeast and mammalian systems is given in table-1 (modified from Fischer and Emans, 2000).

Table 1-1 Comparison of heterologous protein production in plants, yeast and mammalian systems (modified from Fischer and Emans, 2000)

	Transgenic plants	**Plant viruses**	**Yeast**	**Bacteria**	**Mammalian cell cultures**	**Transgenic animals**
Storage temperature	Cheap/RT	Cheap/−20°C	Cheap/−20°C	Cheap/−20° C	Liquid nitrogen	Liquid nitrogen
Distribution	Easy	Easy	Feasible	Feasible	Difficult	Difficult
Gene size	Not limited	Limited	Unknown	Unknown	Limited	Limited
Glycosylation	‘Correct'?	‘Correct'?	Incorrect	Absent	‘Correct’	‘Correct’
Multimeric protein assembly	Yes	No	No	No	No	Yes
Production cost	Low	Low	Medium	Medium	High	High
Production scale	Worldwide	Worldwide	Limited	Limited	Limited	Limited
Production vehicle	Yes	Yes	Yes	Yes	Yes	Yes
Propagation	Easy	Feasible	Easy	Easy	Hard	Feasible
Protein folding accuracy	High?	High?	Medium	Low	High	High
Protein homogeneity	High?	Medium	Medium	Low	Medium	High
Protein yield	High	Very high	High	Medium	Medium-High	High
Public perception of risk	High	High	Medium	Low	Medium	High
Scale-up costs	Low	Low	High**	High**	High**	High
Therapeutic risk*	Unknown	Unknown	Unknown	Yes	Yes	Yes
Production time required	Medium	Low	Medium	Low	High	High

*-Residual viral sequences, oncogenes, endotoxins; **-Large, expensive fermenters etc; ?- unclear (Fischer and Emans, 2000)

1.2 Selection of plant type for transformation

Theoretically, any type of plant can be transformed; however, it is useful to use plants which are well studied and characterized allowing effective risk assessment and tracking of transgene of interest. The appropriate type of plant selection is important for efficient protein production. Proper consideration should be given to use plants species with the least chance(s) of dissemination of trangene/gene product to other plants/environment. Use of self-pollinating plants can reduce chances of gene transfer to other plants. The choice of host plant also depends on the recombinant protein to be expressed, the host plant life cycle, biomass yield and productions costs (Sharma & Sharma, 2009). If the protein is to be expressed in green tissues then leafy plants like tobacco may serve as ideal plant with a huge biomass production capacity (Fischer and Emans, 2000). Tobacco has been used as the system of choice for a number of plant-derived recombinant proteins (Sexton *et al.*, 2009). Tobacco has the advantage of producing huge quantities of green leaf material per acre. The agrobacterium-mediated transformation is highly efficient in tobacco. The plant has a prolific seed production which could facilitate biomass scale-up. Tobacco plants mainly self-pollinate so there is little risk of transfer of genetic material to other plants. Tobacco is also a non food crop, so there is little risk of food chain contamination (Ma *et al.*, 2003). Other leafy crops that have been used for recombinant protein production include spinach (Yusibov *et al.*, 2002), lettuce (Ruhlman *et al.*, 2007) and alfalfa (Khoudi *et al.*, 1999). A disadvantage associated with leafy crops is that the expressed proteins may be unstable in the leaf environment which could interfere with yield and quality of the protein in question. Also the phenolic compounds released during extraction process could be detrimental for downstream processing (Ma *et al.*, 2003). Cereals and legumes can also be used as alternative production systems to overcome some of these problems (Rademacher *et al.*, 2008; Tsuboi *et al.*, 2008). Seeds have specialised storage compartments which help in reducing protein degradation, the exposure of the recombinant protein to phenolic compounds is also avoided thus improving downstream processing (Ma *et al.*, 2003). The overall yield of the recombinant protein expressed in seeds might be lower compared to the yield of the same protein expressed in leaf because of the high biomass yield in case of leaf (Stoger *et al.*, 2002). A number of cereal crops like maize (Rademacher *et al.*, 2008), rice (Qian *et al.*, 2008) and wheat (Tsuboi *et al.*, 2008) have been utilized for recombinant protein production. Legumes like pea (Perrin *et al.*, 2000)

and soybean (Moravec *et al.*, 2007) have also been used to express foreign proteins. Oil crops offer another inexpensive platform for expression of recombinant proteins. With Oleosin fusion technology, developed by SemBioSys Genetics Inc. (http://www.sembiosys.com/), the recombinant protein gene sequence is fused to the sequence of an oil-body-specific endogenous protein oleosin in rapeseed and safflower, after purification the protein is separated by an endoprotease digestion (Schillberga *et al*., 2005). Potato has also been exploited in the production of vaccines (Ma *et al.*, 2003). Other plants that have been used for biopharming include tomato, carrot, banana and papayas.

1.3 Sub-cellular targeting

Consideration should be given to where the protein of interest should be produced in plants. Targeting to sub-cellular compartments can help increase the production and recovery of the target proteins (Fischer and Emans, 2000). The endoplasmic reticulum is an important site for the processing, disulphide bond formation, assembly and glycosylation of proteins (Helenius and Aebi, 2001). Targeting proteins to apoplast can result in high level of expression and better downstream processing however endoplasmic reticulum retention has resulted in 10 to 100 fold more yields (Conrad and Fiedler, 1998). Another important sub-cellular compartment with capability of several hundredfold accumulation of recombinant proteins is the chloroplast which is capable of correct folding and disulfide bonds formation, but is not capable of glycosylation (Ruf *et al*., 2001). Recombinant proteins have also been targeted to protein storage vacuoles (Stoger *et al*., 2000). Rhizosecretion or the secretion of recombinant protein in the hydroponic medium in roots is another strategy that can help in simplifying the downstream processing and increasing protein yield (Drake *et al.*, 2009).

1.4 Plant transformation types

1.4.1 Nuclear transformation

Stable transformation involves the integration of foreign gene/s in the genome of plant. This can be achieved either by agrobacterium mediated transformation of dicotyledonous plants or through biolistic delivery (gene-gun) methods in monocots. These transformations result in heritable expression of recombinant protein, which is stable from generation to generation, so it is good for long term production of recombinant proteins. However, it is

time consuming as to develop stability transformed plants takes months or years depending on plant type used for recombinant protein expression. The strategy is also costly compared to transient expression. With nuclear transformation it is possible to target the protein of interest to various subcellular locations such as the nucleus, cytoplasm, endoplasmic reticulum, plastids, vacuole and apoplast. Correct subcellular targeting makes it possible for the correct posttranslational modifications to be carried out on the expressed protein in question. Another advantage of nuclear transformation is the high level of scalability that is possible as the stably transformed plant can be theoretically grown anywhere in world on huge acreages. However, there is the inherent chance of transfer of genetic material to other crops, so special regulatory measures (for details see section 1.10) must be taken to prevent transgene escape to non-target crops.

1.4.2 Transient expression

During transient expression the foreign genetic material does not integrate into the genome of plant. This can be achieved through agro-infiltration (using agro-bacteria), viral vectors or through biolistics. Transient expression offers some advantages, as data on whether a particular gene is being expressed or not can be obtained in days (Kapila *et al*., 1996). Transient expression can be used in pilot experiments before proceeding to the time consuming and costly stable transformation. Problems in protein expression can be identified and corrected so that the chances of producing the desired protein through stable transformation are made more likely. Through agroinfiltration a number of genes can be expressed at the same time which can help in studying the combined effect of multiple transgenes which are expressed (Johansen & Carrington, 2001). However, the yield of expressed protein is normally less and the plant material requires processing immediately as due to the perishable nature of leaves degradation during storage of plant tissue will result in further loss of the protein.

1.4.3 Chloroplast transformation

The gene of interest can also be incorporated into the chloroplast genome. The commonly used method to transform chloroplasts involves using a gene gun to incorporate the transgene into the chloroplast genome (Daniell, 2006). A mature leaf cell contains up to 100 chloroplasts and each chloroplast can contain up to 100 copies of

chloroplast genome. It is therefore likely that a higher yield of foreign protein could be produced from chloroplast transformation than the nuclear transformation (Daniell, 2006). Since the chloroplast is prokaryotic in nature, gene silencing is not observed unlike nuclear transformation. As the expressed protein is confined to the chloroplast therefore it has no toxic effects on the host plant (Daniell, 2006). Chloroplasts are maternally inherited; therefore, transfer to other plants through pollen is avoided. The technique is most commonly used in tobacco. However, the chloroplast does not provide post translational modifications such as glycosylation and hence the technology is not adopted widely (Boehm, 2007). As with nuclear transformation, chloroplast transformation is also time consuming as it requires the generation of stably transformed plants.

1.5 Stability of recombinant DNA inside plant

One possible outcome of insertion of recombinant DNA into the plant genome is that the recombinant DNA can sometimes undergo inactivation, preventing its expression inside the plant cell. Recombinant DNA inactivation has been attributed to multiple copy integration, different base composition between recombinant DNA and the integration site, overexpression effects and detrimental effects of sequences adjacent to the recombinant DNA integration site (Finnegan and McElroy, 1994). The presence of repeated homologous sequences, recombinant DNA methylation and co-suppression can also lead to recombinant DNA inactivation (Meyer and Saedler, 1996). Recombinant DNA inactivation can be avoided by selecting lines with a single insertion of the transgene, not using repetitive homologous sequences, selecting stable recombinant lines and creating site-specific recombination systems (Finnegan and McElroy, 1994).

1.6 Optimizing foreign gene sequences for expressing in plants

Plants have a different codon usage bias than animals, however foreign DNA can be optimized for expression in plants to increase translation and therefore obtain higher protein yields (Kusnadi *et al*., 1997). Expression can be increased further by use of tissue specific promoters, improving transcript stability and the use of viral sequences for translational enhancement (Gallie, 1998). The 35S promoter is suitable for dicotyledonous plants while the maize ubiquitin-1 promoter is normally used for monocot plants

(Christensen and Quail, 1996; Ma *et al*., 1995). Use of tissue specific promoters can also help in avoiding adverse effects on plant growth, development and environment (Stoger *et al*., 2000; Cramer *et al*., 1999).

Protein expression can be increased by the use of introns in the recombinant DNA molecule (Maas *et al*., 1991). However, the exact mechanism how introns enhance protein expression is not known. It has been found that transgene translation can be enhanced by the addition of un-translated leader sequence of alfalfa mosaic virus mRNA 4 (Datla *et al*., 1993). The stability of mRNA is also influenced by the polyadenylation sites in plant cells (Hunt, 1994; Ingelbrecht *et al*., 1989). These sites protect enzymatic degradation of mRNA. There also exist specific recognition sites that result in RNA degradation (Sullivan and Green, 1993). Some of these recognition sites have been discovered and are supposed to be involved in mRNA degradation as a result of their interaction with specific binding factors (Taylor and Green, 1995). It may be useful to screen for these sites and remove them by modifying the gene of interest to increase gene expression. Specific protein initiation sequences such as the Kozak sequence have been found to help in efficient translation in animals, however, this may be different in plants (Cavener and Ray, 1991; Lutcke *et al*., 1987). The level of amino acid can also be a limiting factor in the expression of particular proteins, it may therefore be necessary to alter amino acid synthesis pathways for the expression of some particular proteins (Matthews and Hughes, 1993, Singh and Matthews, 1994).

1.7 Expression of multiple genes

Traditionally, plants have been transformed with single genes to improve plant characteristics, to study plant gene expression and to express foreign proteins for industrial or pharmaceutical purposes. However, there are situations where transformation of plants with multiple genes is desirable. These include improving plants for multigenic traits, improving or altering metabolic pathways, expressing multimeric foreign proteins and expressing multiple enzymes involved in the synthesis of various compounds like bacterial polysaccharides (further described in this thesis) that have a pharmaceutical/commercial interest. Multiple gene engineering can be achieved by different ways.

1) IRESs or internal ribosome entry sites have been used primarily for the expression of two genes in the form of a bicistronic message (Hellen & Sarnow, 2001; Allera-Moreau *et al.*, 2006; Sasaki *et al.*, 2008). IRESs are nucleotide sequences that recruit eukaryotic ribosomes to mRNA to initiate protein translation in the middle of mRNA molecule without the requirement for a 5′ cap that is normally needed for translation initiation (Pelletier & Sonenberg, 1988). The main drawback of using IRES for multiple protein expression is that the expression of the IRES regulated gene is lower than the cap-dependant gene upstream of IRES sequence (Kaufman *et al.*, 1991; Zhou *et al.*, 1998; Houdebine & Attal, 1999).

2) Independent transgenic lines expressing one gene can be developed and then all the transgenic lines can be crossed together to combine all the genes responsible for the trait under study in a single plant (Ma *et al.*, 1995). However, this strategy is time consuming and laborious and cannot be used for many genes as the occurrence of transgenes on different loci makes the process of obtaining and maintaining homozygous plants complicated. Furthermore, the strategy cannot be used for vegetatively propagated plants.

3) A plant can be sequentially transformed with the transgenes of interest one by one or in units consisting of more than one gene. However, again this is time consuming and a different selection marker is required for each transformation event, and the number of available markers is limited.

4) The multiple genes can be expressed in the form of multiple expression cassettes linked together, each expression cassette with its own promoter and terminator (Slater *et al.*, 1999; Goderis *et al.*, 2002). However, multiple copies of the same promoter used for transgenes of interest can lead to gene silencing (Van Den Elzen *et al.*, 1993; Matzke & Matzke, 1998) and there is only a limited choice of different promoters available to overcome this problem.

6) Co-transformation is another method for multiple gene expression that involves the simultaneous transformation of plant with the transgenes of interest through biolistics or through agrobacterium mediated transformation (Zhang & Fauquet, 1998; Li *et al.*, 2003). Again the problem with co-transformation is that in many cases the plant cannot

be efficiently transformed with all the genes of interest and variable integrations of the genes involved can occur which makes the subsequent characterization difficult.

5) Multiple genes can also be expressed in the chloroplast genome (Daniell & Dhingra, 2002), however, this method is not suitable for proteins that need to be targeted to cell compartments other than chloroplast.

6) Another strategy that was investigated in this thesis involves linking multiple genes by the Kex2 protease cleavage sites. The strategy is described in detail in the following section.

1.7.1 The Kex2 protease system

The *Saccharomyces cerevisiae* Kex2 protein is a type1 (with the amino terminus outside of the membrane) integral membrane protein that resides in the *trans*-Golgi network (TGN) and functions in the eukaryotic secretory pathway to process polypeptide prohormone at pairs of basic residues (Julius *et al.*, 1984; Fuller *et al.*, 1989). Kex2 protease was one of the first discovered proprotein processing enzymes (Rockwell *et al.*, 2002). Initially, it was thought that the Kex2 gene is involved in the expression of *killer* toxin (Wickner & Leibowitz, 1976). Later on, it was found that the gene is involved in the synthesis of the mating pheromone, α-factor in *Saccharomyces cerevisiae* (Leibowitz & Wickner, 1976). Subsequently, it was found that the *killer* toxin and α-factor were actually excised from larger precursors to mature forms by specific endoproteolysis to become functional and a mutation in the Kex2 gene prevented such processing (Fuller *et al.*, 1988). Later, it was found that the Kex2 gene codes for a Ca^{2+}-dependant serine protease (Fuller *et al.*, 1989). This protease has been referred to as kexin, Kex2p, Kex2 or Kex2 protease by various researchers (Rockwell *et al.*, 2002). Homologues of Kex2 protease have been found to be involved in polypeptide prohormones processing at pairs of basic residues in a diverse group of organisms ranging from yeast, *Drosophila*, *Xenopus* to mammals (Brennan & Peach, 1988; Thomas *et al.*, 1988; Thomas *et al.*, 1991; Germain *et al.*, 1990; Bourbonnais *et al.*, 1991; Korner *et al.*, 1991; Hayflick *et al.*, 1992). The protease is highly conserved among the diverse groups of organisms (Thomas *et al.*, 1988).

A Kex2 protease-like pathway has been shown to exist in tobacco (Kinal *et al*, 1995; Jiang and Rogers, 1999), however, the enzymes involved in such processing have not been isolated. Kinal *et al* (1995) expressed the KP6 killer preprotoxin in transgenic tobacco to study its processing in tobacco. The KP6 killer preprotoxin of *Ustilago maydis* is not active until it is processed into α and β subunits by Kex2 protease (Tao *et al.*, 1990). The protoxin was correctly processed into α and β subunits in transgenic tobacco and the mature toxin produced killer activity similar to the native toxin. Jiang and Rogers (1999) demonstrated the substrate specificity and Golgi-localization of the plant Kex2-like protease in tobacco. The authors designed a reporter protein by fusing a mutated proaleurain to a domain of a protein known as BP-80. BP-80 is a vacuolar sorting receptor for Asn-Pro-Ile-Arg containing determinants (Kirsch *et al.*, 1994; Neuhaus & Rogers, 1998). Proaleurain is precursor of aleurain which is a vacuolar glycoprotein present in aleurone cells (Holwerda *et al.*, 1990). The proaleurain has its own vacuolar targeting determinants and after post-Golgi targeting to vacuole, it is converted to mature aleurain (Holwerda *et al.*, 1990). However, the mutated proaleurain used in the study lacked its own vacuolar targeting determinants (Jiang & Rogers, 1999). The mutated proaleurain and BP-80 fragments were either linked by a single Kex2 cleavage site (IGKRG) (Jiang & Rogers, 1998) or three tandemly repeated Kex2 cleavage sites (IGKRG IGKRG IGKRG) (Jiang & Rogers, 1999). The constructs were expressed in tobacco suspension culture. Traffic of the expressed reporter protein and cleavage of the mutated proaleurain from BP-80 was assessed by pulse-chase and immunoprecipitation followed by SDS-PAGE and autoradiography. No cleavage was observed in the construct containing a single Kex2 cleavage site between the mutated proaleurain and BP-80 (Jiang & Rogers, 1998). However, efficient Kex2 processing was observed in constructs containing three Kex2 cleavage sites between the mutated proaleurain and BP-80, with a large portion of mutated proaleurain secreted in the culture medium (Jiang & Rogers, 1999). It was also found that the Kex2-like processing takes place in the Golgi. By mutagenesis of the Kex2 cleavage site, they further showed that the plant Kex2-like protease shows substrate specificity similar to that in yeast with a requirement for hydrophobic-X-dibasic amino acids (X refers to amino acid).

Part of the thesis deals with the utilization of the plant Kex2-like protease system for expression of seven *Streptococcus pneumoniae* type 2 capsular polysaccharide biosynthetic genes in a single tobacco plant for the purpose of expressing the pneumococcal type 2

capsular polysaccharide in transgenic tobacco. Two expression cassettes A and B were designed one carrying four genes and the other carrying the remaining three genes from the *Streptococcus pneumoniae* type 2 polysaccharide operon. Both the expression cassettes were controlled by a single CaMV35S promoter. Tobacco plants were sequentially transformed with the two constructs to assemble all the seven genes in a single tobacco. The strategy is described in detail in chapter 4.

1.8 Purification and downstream processing of recombinant proteins

Downstream processing refers to the recovery and purification of the recombinant protein from plants. Recovery usually involves processing/fractionating of the plant tissue, protein extraction, solid–liquid separation, and concentration, whereas purification consists of immunoprecipitation, liquid–liquid extraction, membrane filtration, chromatography, etc. Processing of leaves requires special attention. Leaves must be processed immediately after harvest or must be frozen to prevent degradation of proteins by proteases, while seeds can be stored for longer periods as there are less chances of degradation of recombinant proteins expressed in the seed. Use of cell secretion systems could also be beneficial as there is no need to disrupt plant cells during recovery so phenolic compound release could be avoided, however, in the culture medium the recombinant protein may not be stable (Fischer *et al*., 2004). Another way of facilitating protein recovery is the use of affinity tags. Protein tags should be removed after purification to restore the structure of the purified protein to its native state (Fischer *et al*., 2004). Oleosin fusion technology, developed by SemBioSys Genetics Inc. (http://www.sembiosys.com/), is another system in which the recombinant protein gene sequence is fused to the sequence of an oil-body-specific endogenous protein oleosin in rapeseed and safflower, after purification the protein is separated by an endoprotease digestion (Schillberga *et al*., 2005). Problems encountered during protein extraction primarily include proteolytic degradation and structural modification due to reaction with phenolic compounds. When devising a strategy for heterologous protein production in plants, proper consideration should be given to downstream processing feasibility of the recombinant protein to get optimum protein yield.

1.9 Economics of plant made pharmaceuticals

The economic prospects for plant made proteins are quite high in comparison to conventional systems. There is a huge demand for many pharmaceutical proteins. Thanks to the scalability of transgenic plants this demand can be fulfilled at fractional costs of the traditional systems. It is possible to obtain yields of kilogram quantities of recombinant protein from just one hectare of transgenic tobacco (Fischer & Emans, 2000). Even if the expression level of protein in corn is 1% of dry weight and its recovery is only 50%, then the cost of its production may still only be 2-10% of microbial systems and may be even less than that for mammalian systems (Twyman *et al.*, 2003; Chen *et al.*, 2005). It has been estimated that at 20% total soluble seed protein expression level, one bushel (25 kg) of maize can produce the same amount of avidin as one tonne of chicken eggs and the cost is only 0.5% of that of chicken eggs (Hood, 2004). Several companies at the moment are involved in the commercial production of plant made recombinant proteins. For example, Maltagen Forshung GmbH has been able to produce 3 g human albumin, 2 g lactoferrin and 1.5 g lysozyme per kg of barley seeds respectively (http://www.maltagen.de/PDF/ Products. pdf). Several plant molecular farming products are currently under the advanced stages of development and some have been produced commercially. Some of these products are given in Table 1-2. Plants also have the advantage that several types of protein like the recombinant subunit vaccines can be administered in the form of raw or partially processed fruits and vegetables (Mason *et al.*, 2002). However, it has been quite hard for the industry to grow because of very strict regulations (section 1.10). The consumers concerns about safety of plant made pharmaceuticals further add to the difficulties involved in commercializing molecular faming products. Once the regulatory hurdles, and issues over public confidence can be overcome, then plant produced, molecular pharmed products have the potential to bring about the cost effective and large scale production of some important products, many with high medical value.

Table 1-2 Molecular farming products that are in advanced stages of development or have already been commercialized (http://www.molecularfarming.com/PMPs-and-PMIPs.html), table last updated March 2010)

Company / Institution	Plant used	Product	Indication/Use	Clinical Stage
Meristem Therapeutics	Maize	Gastric lipase	Cystic fibrosis	Phase 2
	Maize	Lactoferrin	Gastrointestinal disorders	Phase 1
Planet Biotechnology	Tobacco	sIgA "CaroRx"	Prevention of tooth decay	Phase 2 but already granted an EU licence
	Tobacco	ICAM1	Receptor for common cold	Phase 1 ready, early 2008
Prodigene	Maize	Lt-B vaccine	Traveller's diarrhoea	Phase 1 complete
	Maize	TGE vaccine	Piglet gastroenteritis	Phase 1 complete
	Maize	Avidin	Diagnostic use	Available in Sigma catalogue
	Maize	Trypsin	Wound care / insulin manufacture	Available in Sigma catalogue
	Maize	Aprotinin	Non clinical use	No longer in Sigma catalogue
	Maize	GUS	Non clinical use	No longer in Sigma catalogue
LSBC	Tobacco	Vaccine	Non-Hodgkin's lymphoma	Phase 1 successful in 2002
	Tobacco	Vaccine	Feline Parvovirus	"very advanced" - Status confidential
	Tobacco	Aprotinin	Non-clinical use	Was in Sigma catalogue in 2005
Arizona State University	Potato	Vaccine	*E. coli*	Phase 1
	Potato	Vaccine	Hepatitis "B"	Phase 1
	Potato	Vaccine	Norwark virus	Phase 1
	Tobacco	Vaccine	Norwark virus	Phase 1 / 2
Ventria Biosciences	Rice	Lactoferrin	Infant formula enhancer	On sale to infant formula makers
	Rice	Lactoferrin	cell culture media	Available from company

Molecular farming products that are in advanced stages of development or have already been commercialized (http://www.molecularfarming.com/PMPs-and-PMIPs.html), Continued

Company / Institution	Plant used	Product	Indication/ Use	Clinical Stage
Ventria Biosciences	Rice	Lysozyme	For research purposes	Available from company
	Lemna	Alpha Interferon	Hepatitis "B" & "C" and Cancer	Phase 2b
Biolex	Lemna	"BLX-155"	Fibrinolytic "Clot-buster"	Phase 1 ready
Cobento AS (formerly Cobento Biotech AS)	Arabidopsis	Human Intrinsic Factor	Vitamin B12 deficiency	Approved Coban product launched
	Arabidopsis	Transcobalamin	Diagnostic / research	Available from company
Dow Agrisciences	Plant cell non-nicotine tobacco	Vaccine	Newcastle disease in poultry	USDA Approved in Feb.06
Protalix	Plant cell - tobacco	Vaccines	"diseases of horses, dogs and birds"	Undisclosed
	Plant cell	Glucocerebrosidase	Gaucher's disease	On Sale - named patient basis
Guardian Biosciences	Plant cell	Acetylcholinesterase	Biodefense	Phase 1
	Canola	Edible vaccine	Coccidiosis in poultry	CFIA phase 2
Nexgen Biotech	Oriental Melon	Thyroid stimulating hormone receptor	Diagnosis of Graves disease	Available from company
	Oriental Melon	Viral antigens	Rapid detection of Hantaan and Puumala	Available from company
CIGB (Cuba)	Tobacco	Recombinant Monoclonal (Mab)	Purification re-agent in Hep."B" vaccine	Approved mid 2006 in Cuba
D.Yusibov and others	Spinach	Vaccine	Rabies	Phase 1 successful in 2002
Farmacule	Tobacco	Virtonectin	Research use	Available from company mid 2007
NeoRx / Monsanto	Maize	Antibody	Cancer	Phase 2 (failed)
Applied Biotech. Institute & partner	Maize	Undisclosed	Undisclosed	Animal trials underway
Sembiosys	Safflower	Insulin	Diabetes	Phase 3
	Safflower	Apolipoprotein AI	Cardiovascular	Phase 1
Chlorogen & partner	Tobacco	TGF-Beta protein	Ovarian Cancer	Advanced animal trials
Biolex	Lemna	Alpha Interferon	Hepatitis "B" & "C" and Cancer	Phase 2b
	Lemna	"BLX-155"	Fibrinolytic "Clot-buster"	Phase 1 ready

1.10 Bio-safety and monitoring

Public concern about the introduction of genetically modified crops represents one of the most challenging issues. Lack of communication among the authorities dealing with the research, bio-safety and trade is an important issue that has hindered developments in molecular farming (Ramessar *et al.*, 2008). The long-term impact of molecular farming products on the environment is difficult to assess. An important concern is the contamination of food chain with plant made pharmaceuticals. This could happen as a result of transfer of genetic material from transgenic plants to food crops, using the same equipment for harvesting and processing of transgenic and food crop without proper decontamination, and growing food crops in the same field where a transgenic crop was grown previously and no decontamination was carried out (Rigano & Walmsley, 2005). To avoid food chain contamination, strict regulation needs to be put in place such as geographically isolating the transgenic crop, growing in greenhouses instead of open fields, and harvesting and processing transgenic plants using separate equipment or properly decontaminating the equipment if the same equipment is also applied to food crops (Rigano & Walmsley, 2005). Containment can also be achieved by using male- sterile traits and using chloroplast transformation of plants. As the chloroplast genome is maternally inherited the chances of transgene spread through pollen is controlled. It is also important to label genetically modified products so that the consumer has the choice to select according to his/her own preferences. Regulatory agencies are facing a number of challenges regarding regulation of transgenic crops. Each molecular farming product and each host system is unique so each case needs to be handled separately. A 0.5% presence of transgenic material in non-transgenic food or feed has been allowed by the European Parliament and the Council of the European Union in cases where the presence of transgenic material is unavoidable and its negative effects are dominated by its benefits (European Parliament, 2003). However, a 0.9% presence of molecular farming products other than pharmaceuticals is allowed in non-transgenic material (Spök, 2007). Efforts to confine transgenics and reduce environmental exposure have been made recently (Chen *et al.*, 2005; Julian *et al.*, 2005; Sparrow & Twyman, 2009), however, these regulations are still in infancy and a lot more effort is needed to overcome problems regarding regulations of molecular farming products.

1.11 Expression of human insulin in transgenic tobacco

1.11.1 Insulin

An important pharmaceutical compound that can be produced by molecular farming is the polypeptide hormone insulin. Insulin is derived from the Latin word insula which means “island” and it is because the hormone is produced in the Islets of Langerhans. Insulin was discovered by Banting and Best in 1921-22 at the University of Toronto. Insulin helps in transporting blood glucose into the body cells where the glucose is metabolized to produce energy. It maintains glucose concentration in the blood. When glucose concentration in the blood is increased, insulin lowers it by increasing glucose uptake by muscle, liver and fat cells. Excess glucose is converted to glycogen by these tissues. When glucose concentration is reduced in the blood, glycogen is converted back to glucose and released in the blood. It is involved in regulating amino acid uptake by increasing DNA replication and protein synthesis. Insulin helps in fatty acid synthesis through the uptake of lipid from blood by fat cells. Insulin also decreases proteinolysis, lipolysis and gluconeogenesis.

Insulin is synthesized by β-cells of the pancreas in the form of a single chain of three peptides B, C and A in the order: B chain-C peptide-A chain (Schwartz, 1990; Docherty and Steiner, 1982). A diagram of human proinsulin is shown in Figure 1-1 (Nolte and Karam, 2004). This proinsulin is converted to mature insulin after the removal of the central C-peptide by the action of proteolytic enzymes known as prohormone convertases PCl/PC3 and PC2 (Orci *et al*., 1987; Seidah and Chretien, 1992). The mature insulin consists of B-chain (30 amino acids) and A- chain (21 amino acids) linked by two inter-chain and one intra-chain disulphide bridge. Insulin structure has been highly conserved among vertebrates (Chance *et al.*, 1968; Cutfield *et al.*, 1979; BAJAJ *et al.*, 1983; Pollock *et al.*, 1987). There is only three amino acids difference between bovine and human insulin and only one amino acid difference between pig and human insulin.

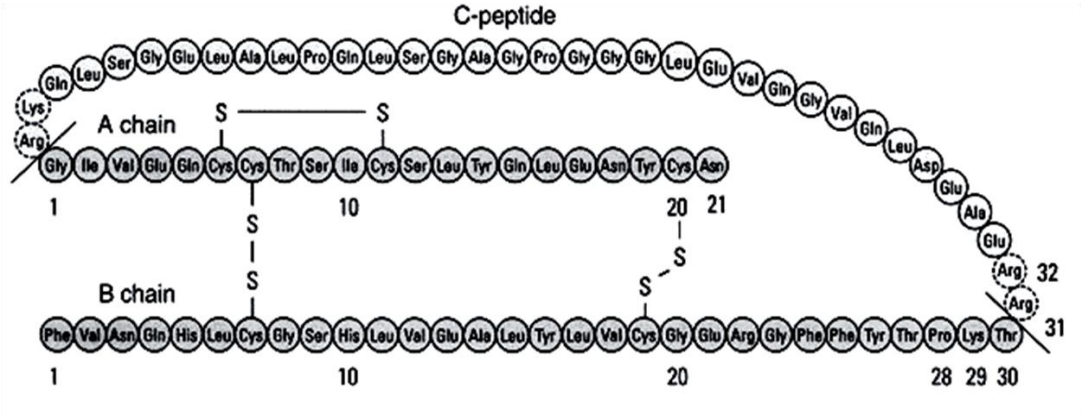


Figure 1-1 Diagram of human proinsulin (from Nolte and Karam, 2004): Proinsulin consists of B and A chains linked by the central C-peptide. B chain consists of 30 amino acids and A chain consists of 21 amino acids. Proinsulin is converted to mature insulin by prohormone convertases PC1/PC3 and PC2. The mature insulin consists of B and A chain. The two chains are linked by two inter chain disulphide bonds (one between cysteine residue 7 of B-chain and cysteine residue 7 of A-chain and another bond between cysteine residue 19 of B-chain and cysteine residue 20 of A-chain). An intra chain disulphide bond in the A chain also exists between cysteine residues 6 and 11.

1.11.2 Diabetes mellitus

Diabetes mellitus refers to a metabolic disorder characterized by chronic hyperglycaemia in which the patient experiences disturbances in metabolism of carbohydrate, fat and protein as a result of deficiency in insulin secretion, insulin action or both (Alberti *et al.*, 1998). Diabetes is the fifth leading cause of mortality in most developed countries and there are thoughts to be about 246 million people suffering from diabetes (diabetes atlas, IDF, 2007). It has been estimated that by 2025 there will be nearly 380 million people suffering from diabetes if not controlled (diabetes atlas, IDF, 2007). Diabetes leads to severe damage and failure of various organs. Symptoms include abnormal thirst, recurrent urination, blurring of vision, consistent hunger, dry mouth, weight loss, slow healing of wounds, coma, frequent infections and tiredness. Prolonged effects may lead to blindness, kidney failure, limb amputation, foot ulcers, heart attack and cerebrovascular disease.

1.11.2.1 Types of Diabetes

1.11.2.1.1 Type 1 diabetes

Type 1 diabetes is caused by the destruction of β-cells of pancreas that are involved in insulin synthesis. The destruction is caused by an auto-immune reaction where these cells are destroyed by the body's own defence system. As a result, very little or no

insulin is produced and hence carbohydrate, especially glucose, metabolism is disturbed leading to the characteristic symptoms. Type 1 diabetes is also known as insulin dependent, juvenile or immune- mediated diabetes. The causes of type 1 diabetes are still largely unknown; however, both genetic and environmental factors have been implicated. Children of parents with type 1 diabetes have an increased risk of developing the disease. Important environmental factors include obesity, diet, physical inactivity and viral infection. Type 1 diabetes mostly affects children and young adults. Patients suffering from type 1 diabetes consistently require insulin treatment (via injection or oral administration) for their continued survival.

1.11.2.1.2 Type 2 diabetes

Type 2 diabetes occurs as a result of insulin resistance and relative insulin deficiency. Insulin resistance refers to the condition where the insulin is not very effective in reducing the blood glucose levels. Type 2 diabetes can appear at any stage of life; however, it is mostly diagnosed in adults at the age of 40 or above. Symptoms of type 2 diabetes involve frequent urination, constant thirst, increased hunger, weight loss and tiredness. Like type 1 diabetes both genetic and environmental factors contribute to the development of type 2 diabetes. Genetic factors leading to type 2 diabetes are not well understood, while potential environmental factors include physical inactivity, obesity, diet and aging. Patients with type 2 diabetes generally do not need to be administered exogenous insulin, however, in extreme cases of hyperglycemia administration of exogenous insulin might be required. Treatment generally involves controlling blood glucose, blood pressure, lipids, changing lifestyle through exercise and medication.

1.11.3 Treatment of diabetes

Type 1 diabetes is mainly treated by administration of insulin, while type 2 diabetes patients only require exogenous insulin in the advance stages of the disease. Treatment involves regular monitoring of glucose levels in the blood so as to provide insulin when required to prevent hyperglycaemia. Necessary changes in lifestyle and diet are also required to keep diabetes in control. Insulin is primarily given in the form of injections, however recent advances in drug delivery have included pulmonary, oral and nasal administration routes which have been developed to overcome the inconveniences

associated with regular insulin injections (Modi *et al.*, 2002; Cefalu, 2004). Since the 1920s, insulin has been used for treatment of diabetes. The initial sources of insulin were from bovine, porcine, and equine. However, these sources of insulin have been of concern as some patients showed allergic reactions to them. Human insulin (humulin) was one of the first pharmaceutical proteins that were manufactured through recombinant DNA technology in the late 1970s, which has been marketed since 1982. Recombinant human insulin from *E. coli* and yeast has since been used for the routine treatment of diabetes.

1.11.4 Types of insulin according to mode of action

Healthy humans produce insulin at a basal level. The secretion of insulin increases to a peak level 1 hour after eating and dropping back to normal level after a further 2 hours. To achieve a normal 24 hour insulin profile and avoid nocturnal hypoglycaemia in diabetic patients, a single insulin formulation with a specific onset of action, peak effect time and duration of action cannot be used. Normally, when insulin is injected into the body, the insulin molecules form hexamers. For diffusion through interstitial fluid and penetration into the capillary walls to enter the bloodstream these hexamers need to dissociate into dimers and monomers (Guerci & Sauvanet, 2005). Therefore different formulations of insulin with different onset of action, peak effect and duration of action have been developed to meet specific needs of patients. These formulations change the rate of dissociation of hexamers into dimers and monomers and the resulting movement of free insulin molecules in the blood stream. These different types of insulin formulation are described in the following sections.

1.11.4.1.1 Rapid acting insulin

These insulin analogs have a more rapid onset of action (15-30 min) and shorter activity duration (4 to 5 hours). Their peak action ranges from 30-90 min post injection. By a single or two amino acid alterations in the insulin molecule, the ability to associate into hexamers is reduced such that they are readily absorbed, however, these modifications do not change the biological properties of these analogs (Howey *et al.*, 1994; Mudaliar *et al.*, 1999). Examples of rapid acting insulin include Lispro and Aspart. In Insulin Lispro (LysB28, ProB29), the positions of proline at position B28 and lysine at position

B29 in the B chain have been reversed (Howey *et al.*, 1994). In insulin Aspart, the proline at position 28 has been replaced by aspartic acid (Mudaliar *et al.*, 1999). These rapid acting analogs can be used at mealtime to achieve optimum level of insulin for utilization of glucose released after eating.

1.11.4.1.2 Short acting insulin

Short acting insulin analogs have an onset of action of around 0.5-1 hour, peak action of 2-4 hours and activity duration of 6-8 hours. Examples of these preparations include Actrapid, Humulin, Hypurin and Neutral. These insulin analogs should be injected into the body 20-30 minutes before meal so as to get optimum insulin activity for carbohydrate metabolism.

1.11.4.1.3 Intermediate acting insulin

Intermediate acting insulin analogs have an onset of action around 1-2 hours, peak action of 6-10 hours and activity duration of 10-16 hours. Examples of intermediate acting insulin include NPH (Neutral Protamine Hagedorn) and LENTE (from the Latin "lentus," meaning slow, or sluggish) insulin. The absorption rate of NPH insulin is reduced by the addition of protamine to the insulin preparation. In insulin LENTE, the same is achieved by the addition of zinc to the insulin preparation.

1.11.4.1.4 Long acting insulin

These insulin analogs have an onset of action around 2 hours, peak action (sometimes no peak action) of 6-20 hours and activity duration of up to 36 hours. One way to prolong insulin activity is designing analogs with more positively charged amino acids so as to raise the isoelectric point of insulin to near neutral pH (Rosskamp & Park, 1999). This helps in reducing the solubility of insulin at neutral pH after injection into the body and the absorption into the blood stream will be delayed. Some of the long acting insulin preparations also have protamine or zinc added to them to increase absorption time. Insulin detemir, also called desB30 insulin, is an example of long acting insulin. In insulin detemir, the threonine at position B30 in the B chain is removed and a 14-C fatty acid i.e. myristic acid is attached to the lysine at position B29 in the B chain (Markussen *et al.*, 1996). Attachment of myristic acid helps in insulin

hexamer formation and increases the binding of insulin to plasma albumin which delays the free insulin release and which prolongs the activity of insulin (Whittingham *et al.*, 1997).

1.11.5 Production systems for insulin: plants as alternative for mass scale production of insulin

As previously indicated, the initial insulin medicinal products were mainly from bovine and porcine origin. Methods to convert porcine inulin into the human insulin equivalent were developed in the 1970s and early 1980s (Morihara *et al.*, 1979; Markussen, 1980). Due to the production limit for an increasing demand for insulin and cases of allergic reaction in some patients to these insulin formulations these sources could not be relied upon for continued production. With the development in genetic engineering and modern biotechnology, efforts were shifted towards the production of insulin through recombinant DNA technology (Ullrich *et al*., 1977). Since 1980s, recombinant insulin from bacteria (Chan *et al.*, 1981) and yeast (Markussen *et al*., 1986) has been the main source of commercial insulin. These systems have been helping to meet the demand of this important pharmaceutical since this time. However, the incidence of diabetes is increasing and this demand for insulin cannot be provided by these sources in the near future due to the scale and costs involved. A cheaper, high productive system is required to meet the future demand for insulin. It has been estimated that the demand for insulin will double in the next 10 years (Boothe *et al.*, 2010). Developing countries are facing further challenges as patients in these countries are not able to afford the cost of this expensive drug. Furthermore, alternate delivery systems such as oral, inhalable and buccal forms which aim to provide relief to patients from the painful insulin injections require a higher amount of insulin to be effective as a treatment as these methods are not very efficient (Boothe *et al.*, 2010).

In view of the increasing demand for insulin, a plant based factory for the production of insulin would offer the potential for a high yielding and cost effective method of production for this important pharmaceutical (Arakawa *et al.*, 1998; Nykiforuk *et al.*, 2006; Ruhlman *et al.*, 2007). Arakawa *et al.* (1998) produced a cholera toxin B subunit-insulin fusion protein in transgenic potato tubers as an autoantigen against insulin-dependent diabetes mellitus. The fusion protein accumulated to 0.1% of total soluble

protein. The transgenic tobacco tuber tissues were fed to diabetic mice which helped in the reduction of insulitis and delayed the onset of diabetes. Nykiforuk *et al.* (2006) expressed the desB30 form of human insulin (lacking threonine at position B30) as an oleosin-insulin fusion protein in *Arabidopsis thaliana*. The fusion protein could be subsequently purified by the oleosin fusion technology to reduce the cost of production. They replaced the central c-peptide of human insulin with a trypsin cleavable mini c-peptide (AAK). The purified desB30 insulin could be enzymatically matured through trypsin treatment to remove the central mini c-peptide. Mature insulin accumulated upto 0.13% of total seed protein. Biological activity of the isolated insulin was demonstrated by an insulin tolerance test in mice and phosphorylation assay in a mammalian cell culture system. Boothe *et al.* (2010) utilized a similar strategy to produce the oleosin-insulin fusion protein in safflower seeds for industrial exploitation of the system. The safflower produced insulin is currently under clinical trials with promising initial results (Boothe *et al.*, 2010). Ruhlman *et al.* (2007) produced cholera toxin B subunit–human proinsulin (CTB-Pins) fusion protein in chloroplasts of transgenic tobacco and lettuce with a view of developing an oral delivery system for insulin (to target insulitis). Eight milligrams of leaf powder from transgenic tobacco expressing the fusion protein were fed to non-obese diabetic mice for seven weeks. After the study, normal blood and urine glucose level were observed in CTB-Pins treated mice with reduced insulitis and reduced damage of β-cells of pancreas.

In view of the increasing demand for insulin in the near future, part of my PhD research was focused on investigating tobacco leaves as a system for expression of the DesB30 form of insulin precursor (Kjeldsen *et al.*, 2001; Nykiforuk *et al.*, 2006) that could potentially provide a cheaper and less expensive platform for the production of this important pharmaceutical product.

1.12 Expression of type 2 pneumococcal polysaccharide in transgenic tobacco

1.12.1 *Streptococcus pneumoniae*

Streptococcus pneumoniae is a Gram-positive bacterium that can cause a variety of pneumococcal infections such as meningitis, septicaemia and pneumonia. Pneumococcal diseases are one of the most important causes of childhood morbidity and mortality, especially in developing countries (Shan, 1986; Mulholland, 1999). Each

year 1.6 million people including more than 800,000 children die due to pneumococcal diseases (World Health Organization, 2003). In healthy people the bacterium inhabits the upper part of the respiratory tract as a commensal organism. However, under certain conditions such as during viral infection, malnutrition or mucosal damage the bacterium can become pathogenic (Paton *et al.*, 1993). The bacterium can cause severe respiratory infections in the elderly in the developed countries while in developing countries it is mainly children who are affected. *Streptococcus pneumoniae* has the ability to produce a polysaccharide capsule which is a major virulence factor for this organism (Austrian, 1981). There are about 90 different polysaccharide types (serotypes) which are identified by unique antigenic determinants on the polysaccharide capsule (Jedrzejas, 2004).

1.12.1.1 Virulence of Streptococcus pneumoniae

The virulence factors of *Streptococcus pneumoniae* are made up of a number of structures found mostly on the surface of the bacterium. These include the polysaccharide capsule, the cell wall, pneumococcal surface protein A and some secreted virulence proteins such as pneumolysin, and autolysin (AlonsoDeVelasco *et al.*, 1995).

1.12.1.1.1 Polysaccharide capsule

The polysaccharide capsule is found outside to the cell wall and has been considered as one of the most important virulent factors (AlonsoDeVelasco *et al.*, 1995). It provides protection to the bacterium against phagocytosis by the host cells (Kamerling, 1999). Each of the 90 serotypes has a unique capsular polysaccharide. The polysaccharide is made of oligosaccharide units which are joined by glycosidic linkages. The capsule is permeable to nutrient molecules, ions and even proteins, but protects the bacterium from dehydration and infection by bacteriophages (Kamerling, 1999). It is also thought to help in adherence of the bacterium to target surfaces (Tuomanen & Masure, 1997). The role of the pneumococcal capsule in infection has also been demonstrated by the fact that anti-capsular antibodies have been proved effective against the infection (Snippe *et al.*, 1983).

1.12.1.1.2 Cell wall and cell wall polysaccharides

Purified cell wall polysaccharides and peptidoglycan have been found to induce inflammation in a similar way to the whole bacterium (AlonsoDeVelasco *et al.*, 1995). Antibodies targeted against cell wall polysaccharides have been shown to protect animals from pneumococcal infection (Briles *et al.*, 1989). It has been shown that the cell wall helps in the attachment of un-encapsulated pneumococci to human endothelial cells (Geelen *et al.*, 1993).

1.12.1.1.3 Pneumolysin

Pneumolysin is a 53 kDa protein and belongs to the thiol-activated toxin family (Boulnois, 1992). The protein is actually not found on surface of the bacterium but is a cytoplasmic enzyme and is released in response to another virulent factor, the autolysin. Pneumolysin causes transmembrane pores in mammalian cells that result in lysis of the cells (Boulnois, 1992). Human monocytes produce inflammatory cytokines in response to low concentration of pneumolysin (Houldsworth *et al.*, 1994). Pneumolysin under low concentrations can also inhibit beating of cilia on respiratory epithelial cells, decrease bactericidal activity, inhibit human neutrophils migration, and decrease production of lymphokines and immunoglobulins (Rubins & Janoff, 1998). Furthermore, pneumococcal mutants lacking pneumolysin have reduced virulence compared to the wild type strains (Berry *et al.*, 1989b).

1.12.1.1.4 Pneumococcal surface protein A

Pneumococcal surface protein A (PspA) is another virulence factor that is found on all strains of pneumococci (Swiatlo *et al.*, 2003). The protein is capable of inducing an antibody response in animal models (Briles *et al.*, 1996). PspA molecules are highly cross reactive meaning that immunization with one type of PspA family can provide protection against pneumococci that harbor a PspA of a different family (Briles *et al.*, 1996). However, the immunogenicity properties of PspA have been studied *in vitro* so far and for an antigen to be utilized as vaccine candidate *in vivo* studies are important (Swiatlo *et al.*, 2003).

1.12.1.1.5 Autolysin

Autolysins belong to a group of peptidoglycan degrading enzymes (Jedrzejas, 2001). The degradation by autolysins finally results in cell lysis (Tomasz, 1984). Autolysin has been thought to contribute to pneumococcal pathogenicity. The pathogenicity of autolysins is taught to be due to the release of cell wall components and pneumolysin by their action (Lock *et al.*, 1992). Reduction in virulence of autolysin negative pneumococcus mutants has been demonstrated (Berry *et al.*, 1989a). Furthermore, autolysin immunized mice showed some resistance against pneumococcal infection (Berry *et al.*, 1989a).

1.12.1.2 Vaccines against Streptococcus pneumoniae

The first pneumococcal vaccines consisted of the whole cells and as early as 1911, Sir Almroth E. Wright proposed the use of killed pneumococci for immunization against pneumococcal infections (Watson and Musher, 1999). However, the vaccine did not produce immunity because it contained only one serotype and the vaccine dosage was also suboptimal. Since this time extensive studies have been carried out to develop the following types of pneumococcal vaccines to assist in protection against infection by this organism.

1.12.1.2.1 Pneumococcal polysaccharide vaccines

The idea of developing a polysaccharide vaccine originated from the work of Avery and others who identified that a polysaccharide capsule existed on the outer surface of pneumococcus (Briles *et al.*, 1998). Anti capsular polysaccharide antibodies provided effective resistance against pneumococcal infection (Avery and Goebel, 1933). However, designing an effective vaccine was hindered by the fact that there are 90 different serotypes of pneumococcus and the capsular polysaccharide elicited immunity is type specific (Henrichsen, 1995). This problem was addressed by introduction of a 14-valent pneumococcal polysaccharide vaccine in 1977 (Austrian and Gold, 1964). In 1983, a 23-valent polysaccharide vaccine was developed (Robbins *et al.*, 1983). The vaccine effectively protected young adults (Austrian *et al.*, 1976), but was less effective in small children (Cowan *et al.*, 1978). Currently, a 23-valent polysaccharide vaccine

Pneumovax 23 is still marketed. Generally, polysaccharide vaccines are not very effective in children under the age of 2 years and some of the capsular polysaccharides are not highly immunogenic. New vaccines have been developed to address these problems; however, the polysaccharide vaccines have not been replaced completely (Bogaert *et al.*, 2004) as they are still 61-75% effective in adults.

1.12.1.2.2 Polysaccharide-protein conjugate vaccines

Avery and co-workers showed in 1929 that covalently binding capsular polysaccharides to proteins increases immunogenicity of polysaccharides (Avery and Goebel, 1929; Poland, 1999). Capsular polysaccharides can be coupled to proteins through reactive groups (Poland, 1999; Klein, 2000). Polysaccharide conjugate vaccines have been developed by coupling capsular polysaccharides to carrier molecules like tetanus toxoid, diphtheria toxoid, pneumolysin and meningococcal membrane proteins (Shelly *et al.*, 1997). For a conjugate vaccine to be effective it must have as many capsular polysaccharide types as possible (Bogaert *et al.*, 2004). The immunogenicity of polysaccharide-protein conjugate vaccines depends on the structure and length of the polysaccharide, the nature of the conjugated protein and the ratio of the polysaccharide to the protein coupled (Anderson *et al.*, 1989; Seppala & Makela, 1989; Van de Wijgert *et al.*, 1991).

1.12.1.2.3 Pneumococcal protein vaccines

Due to the high cost and limited serotype coverage of polysaccharide-protein conjugate vaccines, research on the suitability of pneumococcal proteins as potential vaccines is also being conducted (Bogaert *et al.*, 2004). To achieve this, various pneumococcal toxins and surface proteins have been used to immunize animals to check whether they provide protection against pneumococcal infection (Briles *et al.*, 1998). These include pneumolysin, pneumococcal surface protein A, pneumococcal surface adhesion A, choline binding protein A, neuraminidase and autolysin (Bogaert *et al.*, 2004). Of these pneumolysin and autolysin have been found to be most effective (Briles *et al.*, 1998). Monoclonal antibodies to pneumococcal proteins can also be utilized to identify immunogenic proteins (Briles *et al.*, 1998). This can be achieved by producing monoclonal antibodies by immunization with non-encapsulated pneumococci or

pneumococcal extracts and then utilizing the antibodies for the identification of candidate proteins.

1.12.2 The type 2 pneumococcal polysaccharide

As mentioned in section 1.12.1.1.1, there are about 90 different pneumococcal polysaccharide types (Jedrzejas, 2004). These polysaccharides are important constituents of pneumococcol vaccines (Robbins *et al.*, 1983). Among these an important polysaccharide is the type 2 capsular polysaccharide. The type 2 capsular polysaccharide is carried by a number of pathogenic strains (Iannelli *et al.*, 1999) which consists of singly branched hexasaccharide repeating units, each composed of one D-glucuronic acid, two D-glucose, and three L-rhamnose residues (Figure 1-2, Jansson *et al*., 1975).

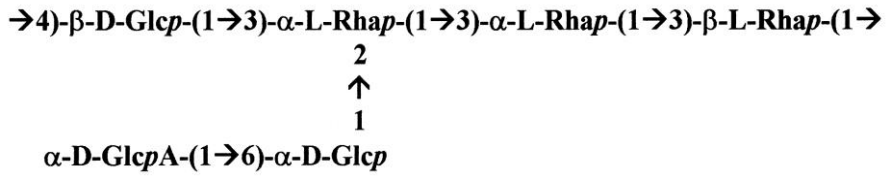


Figure 1-2 Structure of Type2 capsular polysaccharide: Glc, glucose; Rha, rhamnose; GlcA, glucuronic acid. Number indicates the linkages between sugar monomers (from Janson *et al*., 1975)

Gene organization of type 2 polysaccharide locus was characterized by Iannelli (1999) in D39 strain of *Streptococcus pneumoniae* (gene bank accession no. AF029368). 17 genes are responsible for the synthesis of type 2 polysaccharide in the pneumococcus strain D39 which are arranged in a single transcriptional unit (Figure 1-3, Iannelli *et al.*, 1999). Of these, ten genes *cps2*A, *cps2*B, *cps2*C, *cps2*D, *cps2*K, *cps2*P, *cps2*L, *cps2*M, *cps2*N and *cps2*O are involved in biosynthesis of simple sugar monomers (Garcia and Lopez, 1997; Guidolin *et al.*, 1994; Kolkman *et al.*, 1997; Morona *et al.*, 1997; Munoz *et al.*, 1997; Nassau *et al.*, 1996). The remaining 7 central genes *cps2*E, *cps2*T, *cps2*F, *cps2*G, *cps2*H, *cps2*I and *cps2*J encode five putative glycosyltransferases, a polysaccharide polymerase, and a repeat unit transporter (Table 1-3, Iannelli *et al.*, 1999).

Organization of genes involved in Synthesis of Type 2 polysaccharide in *S. pneumoniae* strain D39

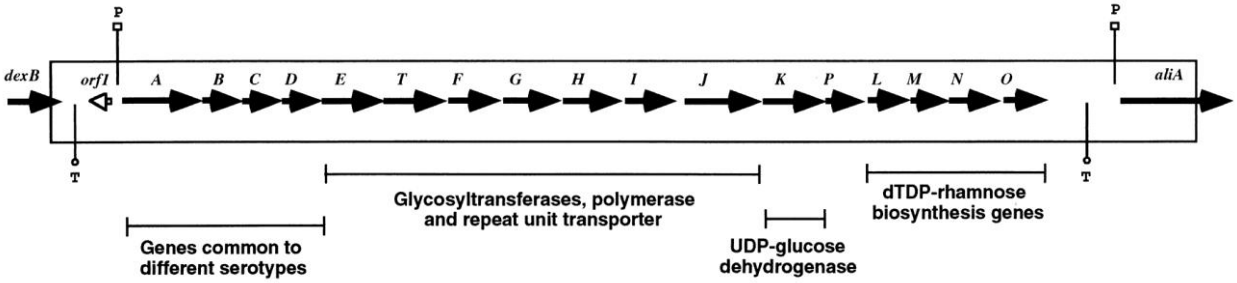


Figure 1-3 Structure of the type 2 capsular polysaccharide locus in *S. pneumoniae* strain D39: *cps*2 genes are indicated by their letters (from Iannelli *et al.*, 1999).

Table 1-3 Putative functions of type 2 capsular polysaccharide genes (Iannelli *et al.*, 1999)

Type 2 gene	**Homologous gene**	**Origin species or *S. pneumoniae* type**	**Putative function**
cps2E	*cps23fE*	Type 23F	Undecaprenyl-phosphate glucose-1-phosphate transferase
	cps14E	Type 14	Glucosyl-1-phosphate transferase
	cps19fE	Type 19F	Uridine diphosphate glycosyltransferase
cps2T	*cps23fT*	Type 23F	Rhamnosyl transferase
cps2F	*cps23fU*	Type 23F	Galactosyltransferase
	cps14J	Type 14	ss-1,4-galactosyltransferase
	cps14I	Type 14	ss-1,3-*N*-acetylglucosaminyltransferase
cps2G	Sequence encoding protein RPN00103	Type 4	Undecaprenyl-phosphate galactosephosphotransferase
	icsA	*Neisseria meningitidis*	Lipopolysaccharide glycosyltransferase
cps2H	*cps23fI*	Type 23F	Polysaccharide polymerase
	cps14H	Type 14	Polysaccharide polymerase
	cps19fI	Type 19F	Polysaccharide polymerase
cps2I	*rfaK*	*N. meningitidis*	Alpha 1,2 *N*-acetylglucosamine transferase
	icsA	*N. meningitidis*	Lipopolysaccharide glycosyltransferase
cps2J	*cps23fJ*	Type 23F	Repeat unit transporter
	cps14L	Type 14	Repeat unit transporter

1.12.2.1 Expression of type 2 polysaccharide in plants

Traditional vaccines (section 1.12.1.2) used for prevention of pneumococcal infections are often too expensive to be distributed widely in developing countries. Plants offer a cheaper alternative for mass scale production of immunogenic pneumococcal molecules. There is the possibility of producing the capsular polysaccharides in plants if the plants can be engineered with genes responsible for production of these polysaccharides. Engineered metabolic pathways in plants could be directed to utilize the plant carbohydrate precursors (Fry, 1988) in the synthesis of pneumococcal polysaccharides. Previously, type 3 pneumococcal polysaccharide has been expressed in transgenic tobacco (Dr. Clair Smith; a PhD candidate in the department of Infection, Immunity and Inflammation, un-published). Type 3 polysaccharide is the simplest of the capsular polysaccharides. It consists of cellobiuronic acid repeating units (consisting of D-glucose and D-glucuronic acid) (Arrecubieta *et al.*, 1996). In *Streptococcus pneumoniae*, type 3 polysaccharide synthesis is thought to occur by the action of four enzymes; a glucose-1-phospate uridyltransferase (*cps*3U), a putative phosphoglucomutase (*cps*3M), a glucose dehydrogenase (*cps*3D) and a type 3 synthase (*cps*3S) (Dillard *et al.*, 1995; Forsee *et al.*, 2000). The first three enzymes are involved in the synthesis of the precursors UDP-glucose and UDP-glucuronic acid (Dillard *et al.*, 1995), while the fourth enzyme the type 3 synthase is responsible for the polymer formation (Forsee *et al.*, 2000). As the type 3 polysaccharide precursors are abundantly produced in plant leaves (Fry, 1988), it was reasoned that the expression of the type 3 synthase gene (*cps3*S) in transgenic tobacco would lead to the production of type 3 polysaccharide by the transgenic plants. For that purpose, the *cps3*S pneumococcal type 3 polysaccharide biosynthetic gene was expressed in tobacco. Production of type 3 polysaccharide by the transgenic tobacco was confirmed by Ouchterlony double-diffusion immune assay (Ouchterlony and Nilsson, 1973) and its immunogenicity was confirmed by administration to mice, and protection from infection by a type 3 *S. pneumoniae* challenge study (Clair Smith, unpublished observation).

As discussed in section 1.12.1.2.1, for a polysaccharide vaccine to be more effective, it needs to be made of a mixture of the different capsular polysaccharides. For that reason it is important to investigate the production of the remaining capsular polysaccharides,

especially the complex capsular polysaccharides *in planta*. Hence, the type 2 capsular polysaccharide, which is much more complex than type 3 polysaccharide, was selected for expression in transgenic tobacco. As described in section 1.12.2, of the 17 genes responsible for the synthesis of type 2 polysaccharide ten genes *cps2*A, *cps2*B, *cps2*C, *cps2*D, *cps2*K, *cps2*P, *cps2*L, *cps2*M, *cps2*N and *cps2*O are involved in biosynthesis of simple sugar monomers (Garcia and Lopez, 1997; Guidolin *et al.*, 1994; Kolkman *et al.*, 1997; Moron *et al.*, 1997; Munoz *et al.*, 1997; Nassau *et al.*, 1996). These sugar monomers are abundantly found in the apoplast of plant leaves (Fry, 1988; Voitsekhovskaya *et al.*, 2002; Büttner & Sauer, 2000; Yeo *et al.*, 1998; Velíšek & Cejpek, 2005). The remaining 7 central genes *cps2*E, *cps2*T, *cps2*F, *cps2*G, *cps2*H, *cps2*I and *cps2*J encode five putative glycosyltransferases, a polysaccharide polymerase, and a repeat unit transporter (Table 1-3, Iannelli *et al.*, 1999). It was expected that if these seven central genes (*cps2*E, *cps2*T, *cps2*F, *cps2*G, *cps2*H, *cps2*I and *cps2*J) could be expressed in a single tobacco plant, they would carry out the synthesis of type 2 polysaccharide *in planta*. The issue of expressing the 7 type 2 polysaccharide biosynthetic genes in a single tobacco plant was solved by utilizing the plant Kex2-like protease system (section 1.7.1). The strategy is described in detail in chapter 4. Briefly, the 7 genes were grouped into two sets. Set-A contained the four genes *cps2*E, *cps2*T, *cps2*F and *cps2*G while set-B contained the remaining three genes *cps2*H, *cps2*I and *cps2*J. Set-A genes were assembled in expression cassette-A and set-B genes were assembled in expression cassette-B. Both the expression cassettes were driven by a single 35S promoter. Three tandemly repeated Kex2 cleavage sites were introduced at the junction of two genes for Golgi localized Kex2 protease processing of the expressed proteins. PR1b endoplasmic reticulum targeting signal was attached to each polyprotein for targeting to ER and then transport to Golgi for processing and ultimate excretion to the apoplast of the processed proteins by the default pathway. Where the expressed proteins could carry out the expected synthesis of type 2 polysaccharide utilizing the available sugar precursors in the apoplast.

1.13 Aims and Objectives

There are two main objectives this thesis deals with. The first objective is to genetically engineer tobacco plants for the expression of $DesB_{30}$ (lacking threonine at position B30) form of human insulin precursor as method of producing low cost human insulin to meet

the future demands. The details of this part of the research are described in chapter 3. The second main objective of my PhD research is the cloning and expression of the 7 type 2 pneumococcal polysaccharide biosynthetic genes in a single transgenic tobacco plant, via utilization of the plant Kex2 protease like system. Chapter 4 details with this second part of my research. A discussion of the results obtained from projects, conclusions and future aspects are described in chapter 5.

2 MATERIAL AND METHODS

2.1 Bacterial manipulations

2.1.1 Strains used

E. coli strains used for plasmid production and DNA manipulation are described in Table 2-1. *Agrobacterium tumefaciens* strain GV3101was used for *Agrobacterium* mediated transformation of tobacco. GV3101 carries a rifampicin (*Rif*) chromosomal resistance gene. The disarmed T-helper plasmid (Ti-plasmid) of GV3101 has the *vir* genes responsible for T-DNA transfer, but has no functional T-DNA of its own. Ti-plasmid confers resistance to gentamycin (*Gen*).

Table 2-1 *E. coli* strains used for plasmid multiplication and isolation

Strain (supplier)	**Genotype and Description**
XL1-Blue heat shock Competent Cells (Stratagene, Cat. No. 200249)	**Genotype:** recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr) **Description:** XL1-Blue cells are endonuclease (*endA*) deficient and good for a higher plasmid DNA recovery. These cells are recombination deficient (*recA*) which improves insert stability. The *hsdR* mutation is helpful in overcoming cleavage of cloned DNA by the *Eco*K endonuclease system. *lacI*qZΔ*M15* gene on the F´ episome allows blue-white colour screening
ElectroMAX DH10B Electrocompetent Cells (Invitrogen, Cat. No. 18290-015)	**Genotype:** F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G **Description:** *mcr*A gene and the *mcr*BC (*mrr* deletion) carried by these cells are useful for cloning of DNA containing methylcytosine and methyladenine. These cells are useful for generating cDNA libraries. The φ80*lac*ZΔM15 gene allows blue/white screening
NEB 5-alpha Electrocompetent Cells (New England Biolabs, Cat. No. C2989K)	**Genotype:** *fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17* **Description:** Φ*80 Δ(lacZ)M15* gene and the (*argF-lacZ*) gene are useful for blue/white selection. These cells are recombination deficient (*recA*), endonuclease 1 deficient (*endA*), restriction deficient (*hsdR17*) and T1 phage resistant (*fhuA2*).

Table 2-1 *E. coli* strains used for plasmid multiplication and isolation (Continued)

Strain (supplier)	Genotype and Description
MegaX DH10B T1R Electrocompetent E. coli cells (Invitrogen, Cat. No C6400-03)	**Genotype:** F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZ ΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ - rpsL nupG tonA **Description:** These cells are resistant to T1 and T5 bacteriophages, mcrA gene and the mcrBC (mrr deletion) help in cloning of DNA containing methylcytosine and methyladenine. They are useful for generating cDNA libraries. The φ80lacZΔM15 gene allows blue/white screening

2.1.2 Media for bacterial culture

2.1.2.1 LB agar

LB (Luria Bertani) agar medium was made by mixing 10 g NaCl, 10 g Tryptone, 5 g Yeast Extract and 20 g agar in ddH2O to make final volume of 1 litre. The medium was autoclaved at 121°C temperature and 15 lb in^{-1} pressure for 20 min and then cooled to 55°C before adding appropriate antibiotic/s. The medium was then poured into plates under sterile conditions.

2.1.2.2 LB broth

LB (Luria Bertani) broth medium was made by mixing 10 g NaCl, 10 g Tryptone and 5 g Yeast Extract in ddH2O to make final volume of 1 litre. LB broth medium was autoclaved at 121°C temperature and 15 lb in^{-1} pressure for 20 min and then cooled to 55°C before adding appropriate antibiotic/s.

2.1.2.3 SOB medium (1 litre)

SOB (Super Optimal Broth) medium was made by mixing 20 g Tryptone, 5 g Yeast extract and 0.5 g NaCl in ddH2O to make final volume of 1 litre. SOB medium was autoclaved at 121°C temperature and 15 lb in^{-1} pressure for 20 min. 10 ml of filter sterilised 1 M $MgCl_2$ and 10 ml of filter sterilized 1M $MgSO_4$ were added to SOB medium before use.

2.1.2.4 SOC medium

SOC (as like SOB but the "C" in SOC supposedly refer to catabolite repression as a result of the added glucose) medium is a rich source of nutrients. The medium was used for recovery of competent bacterial cells after transformation. To make 100 ml SOC medium, 2 ml of filter sterilized 20% glucose was added to 98 ml of SOB medium. SOC medium was aliquated into 7 ml sterilin bijou tubes and stored at −20°C.

2.1.3 Bacterial growth conditions

Bacterial liquid cultures were grown in LB broth medium containing appropriate antibiotic/s at 37°C and with shaking at 220 rpm overnight. For selection of single colonies, bacterial cells were plated onto LB agar medium with appropriate antibiotic/s and incubated upside down at 37°C overnight. *Agrobacterium* liquid cultures were grown in LB broth medium containing appropriate antibiotics at 28-30°C for 36-48 h. For selection of single colonies, *Agrobacterium* cells were plated onto LB agar plates with appropriate antibiotics and incubated at 28-30°C for 36-48 h. Antibiotics used for selection of recombinant *E. coli* and *Agrobacterium* cells are given in Table 2-2. After preparation, antibiotic stock solutions were filter sterilized using 2-4 µM pore size sterile syringe filters. The stock solutions were aliquoted into micro-centrifuge tubes and stored at −20°C.

Table 2-2 Antibiotics used for selection of recombinant bacterial cells

Antibiotic	**Stock solution**	**Working solution**	**Solvent**
Kanamycin	10 mg/ml	25 µg/ml	ddH2O
Ampicillin	100 mg/ml	100 µg/ml	ddH2O
Hygromycin	10 mg/ml	25 µg/ml	ddH2O
Spectinomycin	100 mg/ml	100 µg/ml	ddH2O
Rifampicin	50 mg/ml	25-50 µg/ml	Methanol
Gentamycin	20 mg/ml	10 µg/ml	ddH2O

2.1.4 Storage of bacterial cultures

Bacterial cultures were stored in 20% glycerol at −80ºC. For making glycerol stocks, 200 µl of 100% glycerol was added to 800 µl of bacterial culture in LB broth and mixed on roller for 10 minutes. Glycerol stocks were then stored at −80°C. To recover a strain, a small amount of the frozen stock was streaked on LB agar plate. Single colonies from the plate were then grown in LB broth containing appropriate antibiotic/s.

2.1.5 Plasmid vectors used

2.1.5.1 pGEM-T Easy

Promega pGEM-T easy vector (Cat. No. A1380) was used for cloning of PCR products (containing 5′-A overhangs) and restricted DNA fragments as an easy and rapid DNA manipulation system. This vector is linearized with 3′-T overhangs at the insertion site which can be utilized for ligating PCR products generated by taq polymerase which introduces 5′-A overhangs into PCR products. The vector contains T7 and SP6 RNA polymerase promoters flanking a multiple cloning site within the α-peptide coding region of the enzyme β-galactosidase. The multiple cloning site can be used to clone a number of DNA fragments with different restriction enzyme cohesive ends in the circularized version of the vector. Size of the vector is 3015 bp. The vector carries an ampicillin resistance gene for selection.

2.1.5.2 SEAP-pPPZP212-6

SEAP-pPZP212-6 (Kevin C. Gough, University of Leicester, personal communication) is a derivative of the plant binary vector pPZP212 (Hajdukiewicz *et al.*, 1994). The vector was screened (section 3.3.2) for the isolation of CaMV35S promoter, tobacco PR1b endoplasmic reticulum targeting sequence (Cutt *et al.*, 1988) and the pea *rbc*S terminator sequence (Coruzzi *et al.*, 1984). However, the correct sequences of CaMV35S promoter, PR1b endoplasmic reticulum targeting signal peptide and the *rbc*S terminator could not be found in the vector SEAP-pPZP212-6. Therefore, three additional derivative vectors, i.e. TIR- SEKDEL, ESPA –SEKDEL and TIR- OX1 (Kevin C. Gough, personal communications) were screened by sequencing for the isolation of the desired sequences (for details see section 3.3.2). TIR-OX1was found to

contain the correct sequence of CaMV35S promoter, PR1b endoplasmic reticulum targeting sequence and *rbc*S terminator sequence. The desired sequences were then excised using restriction sites and cloned into relevant vectors.

2.1.5.3 pCAMBIA2301

Plant binary vector pCAMBIA2301 (Cambia) is a high copy number plasmid (Figure 2-1). The size of the vector is 11633 bp. It carries a kanamycin resistance gene for bacterial selection and an *npt*II kanamycin resistance gene for selection in plants. The vector has a *gus*A reporter gene for analysing presence of the transformed DNA in regenerated plants by GUS assay. The vector is useful for cloning genes with their own promoter and terminator sequences. The size of T-DNA is 5391 bp. It carries a pUC18 polylinker which offers a number of cloning sites for DNA fragments with different restriction enzyme overhangs.

The vector was used for cloning of the synthetic human insulin to generate pKAM1 and for cloning of type 2 *pneumococcal* polysaccharide set-A gene insert (carrying *cps*2T, *cps*2E, *cps*2F and *cps*2G type 2 polysaccharide genes) to generate pKAM3. The resultant vector pKAM1 was used in the *Agrobacterium* mediated transformation of tobacco for generation of transgenic tobacco expressing synthetic mini human insulin. Similarly, the resultant vector pKAM3 was used for generation of transgenic tobacco expressing type 2 pneumococcal polysaccharide set-A genes.

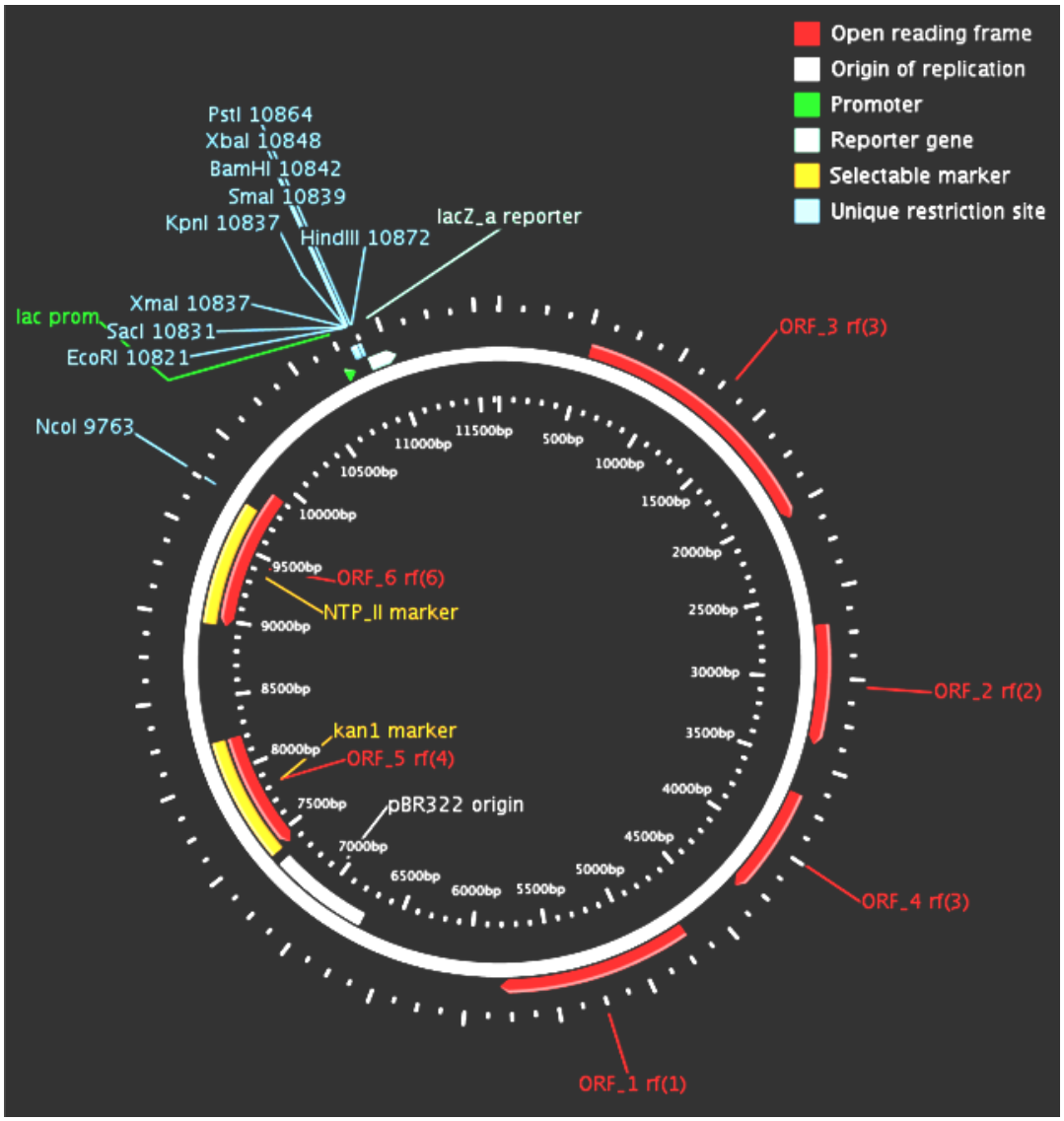


Figure 2-1 Map of pCAMBIA2301 created using PlasMapper (Dong *et al.*, 2004)

2.1.5.4 pCAMBIA1302

Plant binary vector pCAMBIA1302 (Cambia) is another high copy number plasmid that carries a hygromycin plant selection gene (Figure 2-2). Size of pCAMBIA1302 is 10549 bp. The vector carries a *gfp* reporter gene and has a T-DNA size of 4316 bp. It also carries a PUC18 polylinker which is helpful in cloning of inserts with a number of different restriction enzyme sites. Type 2 pneumococcal polysaccharide set-B genes (*cps*2I, *cps*2J and *cps2*H) were cloned into pCAMBIA1302 to generate vector pKAM8. pKAM8 was used in the double transformation of the transgenic tobacco plants expressing set-A type 2 pneumococcal polysaccharide genes (section 2.1.5.3) to

complete the assembly of all seven type 2 polysaccharide genes in a single tobacco plant. As pCAMBIA1302 carries a hygromycin plant selection gene, therefore, selection of the double transformed plants was possible.

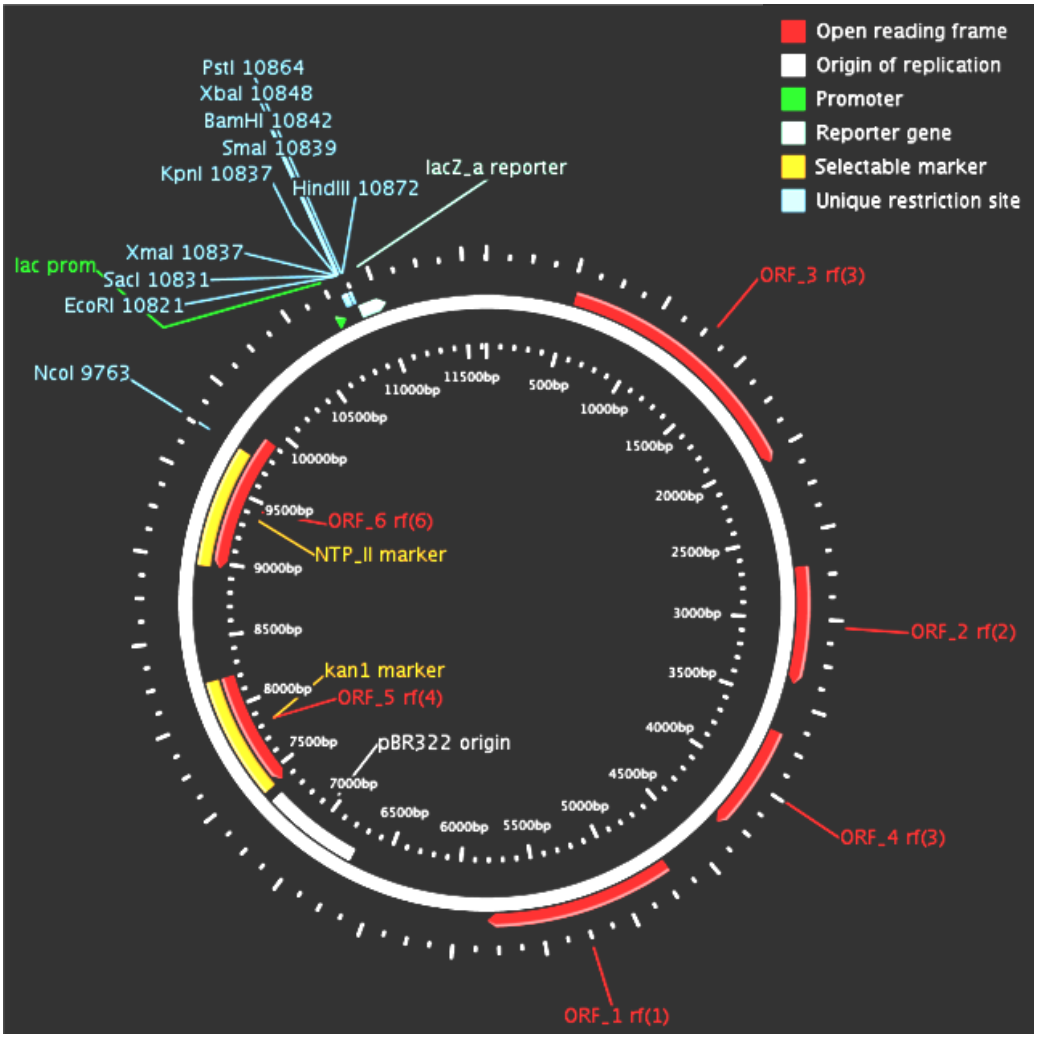


Figure 2-2 Map of pCAMBIA1302 created using PlasMapper (Dong *et al.*, 2004)

Various intermediate derivatives of the above vectors were developed for manipulation of various DNA segments and are described in detail in the respective results sections.

2.1.6 Preparation of electrocompetent XL1-Blue cells

LB broth medium (5 ml) was inoculated with a single colony of XL1-Blue cells and incubated at 37°C for 5 h to overnight with moderate shaking. The resultant culture (5 ml) was used to inoculate 500 ml of LB broth in a sterile flask. The culture was grown

at 37°C and 220 rpm to an OD_{600} value of 0.5 to 0.6. Cells were then chilled on ice for 10 to 15 minutes and transferred to two pre-chilled 250 ml centrifuge bottles. Cells were pelleted by centrifugation at 2°C and 3786 g for 20 minutes. The supernatant was poured off immediately and the cells re-suspended in 0.1 volume of ice-cold ddH20. A volume of ddH2O equal to the original culture volume was then added and mixed gently. Resuspended cells were centrifuged as before. The supernatant was immediately poured off and the cells re-suspended in the small amount of fluid remaining in the bottle. Another volume of ice-cold ddH2O equal to the original culture volume was added and mixed well. Cells were centrifuged as before. The supernatant was poured off and the cells resuspended in the remaining fluid. The cell suspension from the two centrifuge bottles was transferred to two pre-chilled 15 ml narrow bottom tubes. An amount of ice-cold 10 % glycerol equal to 0.08 of the original culture volume was added to the cells and mixed well. The suspension was centrifuged at 2 °C and 3786 g for 10 minutes. The supernatant was decanted and a volume of ice-cold 10 % glycerol equal to the pellet volume was added to resuspend the cells. The cell suspension was divided into pre-chilled microcentrifuge tubes (50 µl in each tube) and frozen by incubation in a dry ice/ethanol bath for 5 minutes. The cells were then stored at −80 °C.

2.1.7 *Escherichia coli* transformation by heat shock

Competent *E. coli* cells (Table 2-1) were thawed on ice and mixed gently. An aliquot of 100 µl was then added to a pre-chilled 15 ml tube. Then, 1.7 µl β-mercaptoethanol was added to the tube and the cells were incubated on ice for 10 minutes and mixed by swirling gently every 2 minutes. After this, 1-3 µl of ligation product (section 2.2.20) was added to the cells. The sample was swirled gently, incubated on ice for 30 minutes and heat-pulsed in a 42 °C water bath for 45 seconds. The sample was then incubated on ice for 2 minutes and 0.9 ml of preheated (42°C) SOC medium (section 2.1.2.4) was added to the sample. The cells were then incubated at 37 °C and shaken at 225-250 rpm for 1 h. The transformation mixture (50-200 µl) was then plated on LB agar plates containing appropriate antibiotic. Plates were incubated upside down at 37°C overnight. 16 to 24 h post transformation the plates were checked for transformed colonies.

2.1.8 *Escherichia coli* transformation by electroporation

For electroporation, 1-2 µl of DNA ligation mixture (section 2.2.19 and 2.2.20) was added to a pre-chilled microcentrifuge tube. Frozen electrocompetent *E.coli* cells (Table 2-1) were thawed on ice for 5 min. Thawed cells (20 µl) were added to the microcentrifuge tube containing DNA ligation mixture. The cells/DNA mixture was then transferred to an ice chilled cuvette (1 mm gap) and electroporated using the Bio-Rad GenePulser II electroporator set at 2.0 kV voltage, 200 Ω resistance and 25 µF capacitance. After electroporation, 1 ml of SOC recovery medium was immediately added to the cells in the cuvette. The cell suspension was then transferred to a 15 ml culture tube and incubated at 37°C and 225 rpm for 1 h. The cell suspension (50-200 µl) was then spread on pre-warmed (37°C) LB agar plates containing suitable antibiotics. The plates were incubated at 37°C overnight. 16 to 24 h after plating, the plates were checked for recombinant colonies growing on these plates.

2.1.9 Preparation of competent *Agrobacterium* cells for heat shock transformation

A single colony of *Agrobacterium* strain GV3101 was inoculated into 3 ml of LB broth medium (containing 10 µg/ml gentamycin and 30µg/ml rifampicin) and incubated at 30°C and 220 rpm overnight. The overnight culture (0.5 ml) was used to inoculate 50 ml of LB broth in a 250 ml flask. The culture was grown at 30°C and 220 rpm for 4-5 h. The culture was then chilled on ice for 5-10 minutes and centrifuged at 4°C and 2039 g for 5 min. The supernatant was removed and the pellet resuspended in 1 ml of ice cold 20 mM $CaCl_2$. For transformation, 0.1 ml of the bacterial suspension was used immediately.

2.1.10 Preparation of competent *Agrobacterium* cells for electroporation

A single colony of *Agrobacterium* strain GV3101 was inoculated into 3 ml of LB broth medium containing suitable antibiotics in a 15 ml snap-cap tube and incubated at 30 °C and 220 rpm overnight. Next day, 0.5 ml of the overnight culture was used to inoculate two 500 ml sterile flasks each containing 100 ml of LB broth medium with appropriate antibiotics. The cultures were grown at 30 °C with vigorous shaking to obtain an OD_{600}

value of 0.5-1.0. The cultures from both flasks were divided into six 50 ml falcon tubes and the tubes centrifuged at 4 °C and 3435 g for 5 min. The supernatants were poured off and cells in each tube were resuspended in 1/2 volume of ice-cold 10 % glycerol and centrifuged as before. The supernatants were discarded and the cells in each tube were resuspended in 4 ml of ice-cold 10 % glycerol. Cell suspensions from all the tubes were combined into two tubes and re-centrifuged. The supernatants were discarded. The cell pellets in each of the two tubes were re-suspended in 2 ml of ice-cold 10 % glycerol, combined into one tube and re-centrifuged. The final pellet was resuspended in 1.5 ml of ice-cold 10% glycerol. Cell suspension was then aliquoted into pre-chilled microcentrifuge tubes (100 µl per tube). The tubes were snap-frozen in liquid nitrogen and stored at −80 °C.

2.1.11 Transformation of *Agrobacterium* by heat shock

For heat shock transformation, 0.1 ml of competent *agrobacterium* suspension (section 2.1.9) was added to a pre-chilled 1.5 ml microcentrifuge tube on ice. To the cell suspension, 1 µg of plasmid DNA was added and mixed by tapping. The mixture was then frozen in liquid nitrogen and thawed at 37°C for 5 min. After this, 1 ml of LB broth medium was added to the tube and the contents were transferred to a 15 ml tube and incubated at 30 °C and 220 rpm for 2 h. The contents were then poured into a 1.5 ml microcentrifuge tube and spun at 1500 g for 5 min to pellet the cells. Supernatant was removed and the pellet resuspended in 100 µl of LB broth medium which was then plated on LB-agar medium containing suitable antibiotics. The plate was incubated at 30 °C for two days. Transformed colonies were visible on the second day of transformation.

2.1.12 Transformation of *Agrobacterium* by electroporation

Electrocompetent *Agrobacterium* cells (section 2.1.10) were thawed on ice for 5 min. An aliquot of thawed cells (50 µl) was transferred to a microcentrifuge tubes containing 50 ng of plasmid DNA and mixed by tapping the tube. The cell-DNA mixture was transferred to a pre-chilled electroproation cuvette (1 mm gap size). The Bio-Rad GenePulser II electroporator was set to 2.5 kV voltage, 25 µF capacitance and 400 Ω resistance. The cells in the cuvette were electroporated and 1ml of growth medium was

added to the cuvette immediately after the pulse. Contents of the cuvette were then transferred to a 15 ml tube and incubated at 30°C and 220 rpm for 2 h. The culture (50-200 µl) was plated on LB-agar plate containing suitable antibiotics and incubated at 30°C for 2 days. Recombinant colonies were visible on the second day after plating.

2.2 DNA techniques

2.2.1 Primer designing for the construction of a synthetic mini human insulin

The strep tag II -mini-insulin gene fusion molecule was constructed by designing four partially overlapping oligonucleotides (Table 2-3) with codon usage optimized for expression in tobacco, utilizing the GeneDesign software (Richardson *et al*, 2006). Oligonucleotides KA001and KA002 were designed to form the 5′ end of the gene fusion molecule and oligonucleotides KA003 and KA004 were designed to form the 3′ end of the gene fusion molecule. Synthesis of the mini-insulin fusion gene is described in detail in chapter 3. The complete sequence of the strep-tag II-mini-insulin fusion gene is given in Figure 2-3. Other primers used for PCR amplification, sequencing of the insulin expression cassette in different vectors and restriction enzyme sites attachment to the synthetic insulin gene are given in Table 2-1. Sequencing primers for screening of the various vectors used for isolation of CaMV35S promoter, PR1b endoplasmic reticulum targeting sequence and the *rbc*S terminator sequence (section 2.1.5.2) are also given in Table 2-3.

A. Strep tag II-mini insulin fusion gene DNA sequence (5′-3′):

TGGTCTCATCCTCAATTTGAAAAAGCTGCTAAGTTTGTTAATCAACATCTTTG
TGGATCTCATCTTGTTGAGGCTCTCTACCTTGTGTGTGGAGAAAGAGGATTTT
TCTACACTCCTAAGGCTGCTAAGGGAATTGTTGAACAATGTTGCACTTCTATT
TGCTCACTTTACCAATTGGAGAACTATTGCAACTGA

B. Strep tag II-mini insulin fusion gene amino acid sequence:

WSHPQFEKAAKFVNQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSI
CSLYQLENYCN*

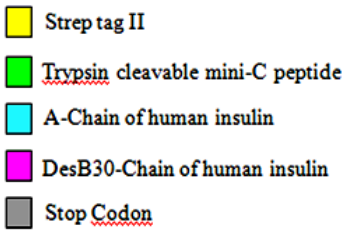


Figure 2-3 DNA and amino acid sequence of Strep-tag II-mini-insulin synthetic gene: Different parts of the gene are highlighted with different colours.

Table 2-3 List of primers/oligonucleotides used for the construction of strep-tag II-mini human insulin fusion protein expression cassette.

Name	Primer/ Oligonucleotide sequence (5′-3′) and description
KA001	TGGTCTCATCCTCAATTTGAAAAAGCTGCTAAGTTTGTTAATCAACATCT TTGTGGATCTCATCT**TGTTGAGGCTCTCTACCTTG** **Description:** Oligonucleotide-1 for synthetic insulin gene synthesis, letters in boldface show the complementary region of KA001 with KA002. Strep-tag II Trypsin cleavable mini-C peptide DesB30-Chain of human insulin
KA002	CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACA**CAAGGTAGAGAGCC TCAACA** **Description:** Oligonucleotide-2 for synthetic insulin gene synthesis, letters in boldface show the complementary region of KA002 with KA001, letters in green show *Bsu*361 restriction enzyme site.
KA003	CCTAAGG**CTGCTAAGGGAATTG** **Description:** Oligonucleotide-3 for synthetic insulin gene synthesis, letters in boldface show the complementary region of KA003 with KA004, letters in green show *Bsu*361 restriction enzyme site. A-Chain of human insulin
KA004	TCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGCAAC ATTGTTCAA**CAATTCCCTTAGCAGCCTT** **Description:** Oligonucleotide-4 for synthetic insulin gene synthesis letters in boldface show the complementary region of KA004 with KA003. Stop Codon
KA005	ACTGGTACCCTGGTCTCATCCTCAATTTG **Description:** Insulin fusion gene forward PCR primer with *kpn*1 site
KA006	GACCTGCAGTCAGTTGCAATAGTTCTCC **Description:** Insulin fusion gene reverse PCR primer with *pst*1 site

Table 2-3 List of primers/oligonucleotides used for the construction of strep-tag II-mini human insulin fusion protein expression cassette (continued).

Name	Primer/ Oligonucleotide sequence (5′-3′) and description
KA006	GACCTGCAGTCAGTTGCAATAGTTCTCC **Description:** Insulin fusion gene reverse PCR primer with *pst*1 site
KA007	ACTGGTACCTGGTCTCATCCTCAATTTG **Description:** The primer was not used due to change of strategy. This forward insulin gene primer was designed (as alternative to primer KA005) to make the insulin gene in frame with start codon of PR1b endoplasmic reticulum targeting sequence in vector SEAP-pPZP212-6 (section 2.1.5.2) by deleting deoxycytosine between *kpn*1 site and insulin gene in primer KA005. However, the correct sequences of CaMV35S promoter, PR1b endoplasmic reticulum targeting signal peptide and the *rbc*S terminator could not be found in the vector (section 2.1.5.2).
KA008	TGTGTGGAATTGTGAGCG **Description:** Forward sequencing primer for vectors SEAP-pPZP212-6, TIR- SEKDEL, ESPA –SEKDEL and TIR- OX1 for isolation of 35S promoter, PR1b endoplasmic reticulum targeting sequence and the *rbc*S terminator sequence. The primer binds 42 bp upstream the 35S promoter in all the four vectors (section 2.1.5.2).
KA009	TAAGTTGGGTAACGCCAG **Description:** Reverse sequencing primer for vectors SEAP-pPZP212-6, TIR- SEKDEL, ESPA –SEKDEL and TIR- OX1 for isolation of 35S promoter, PR1b endoplasmic reticulum targeting sequence and the *rbc*S terminator sequence. The primer binds 47 bp downstream the *rbc*S terminator in all the four vectors (section 2.1.5.2).
KA010	CTGACGGAATCGCATTTCC **Description:** Reverse sequencing primer for vectors TIR- SEKDEL and TIR- OX1 for isolation of 35S promoter, PR1b endoplasmic reticulum targeting sequence and the *rbc*S terminator sequence. The primer binds near the 3′ end of TIR gene to enable complete sequencing of 35S promoter and PR1b signal sequence (section 2.1.5.2).
KA011	GAGCTTTCGCATTCTTGTC **Description:** Reverse sequencing primer for vector ESPA –SEKDEL for isolation of 35S promoter, PR1b endoplasmic reticulum targeting sequence and the *rbc*S terminator sequence. The primer binds near the 3′ end of ESPA gene to enable complete sequencing of 35S promoter and PR1b signal sequence (section 2.1.5.2).

Table 2-3 List of primers/oligonucleotides used for the construction of strep-tag II-mini human insulin fusion protein expression cassette (continued).

Name	Primer/ Oligonucleotide sequence (5′-3′) and description
KA012	GCGACTAGTTGGTCTCATCCTCAATTTG **Description:** Forward primer for *spe*1 site incorporation into insulin insert instead of *Kpn*1 site for cloning into TIR- OX1 instead of cloning into SEAP-pPZP212-6 (containing *Kpn*1 site) as the desired sequences of 35S promoter, PR1b signal and the *rbc*S terminator sequences could not be found in the vector (section 2.1.5.2).
KA013	GATGCCTCTGCCGACAG **Description:** For sequencing the whole insulin expression cassette in the final plant binary vector pKAM1. However, the primer sequence was based in the 35S promoter region and there are two additional 35S promoters in pKAM1. Hence, the sequencing did not work. Subsequently, the sequencing was done with insulin insert forward primer KA012 and reverse primer KA006.

2.2.2 Primer designing for the construction of pneumococcal type 2 polysaccharide expression cassettes

Gene sequences of pneumococcal type 2 polysaccharide biosynthesis genes were obtained from NCBI gene bank (accession number: AF026471). The main steps in designing primers for the two type 2 polysaccharide expression cassettes, i.e. cassette A (carrying type 2 polysaccharide genes *cps2*E, *cps2*T, *cps2*F and *cps2*G) and cassette B (carrying type 2 polysaccharide genes *cps2*I, *cps2*J and *cps2*H) were: the attachment of different protein tags to the C-terminal end of each of these genes; the attachment of 3 tandemly repeated Kex2 plant-based cleavage sequences next to the tag sequence; the attachment of different restriction enzyme sites and the attachment of stop codon sequences at 3′ end of the last gene in each expression cassette. A few extra DNA bases (3 bases) were also included at the 5′ ends of the primers which contained a restriction enzyme site to increase cleavage efficiency. The DNA and amino acids sequences of the protein tags used and the Kex2 plant based processing sequences are given in Table 2-4. These sequences were optimized for expression in tobacco utilizing the GeneDesign software (Sarah *et al*, 2006). All of the seven type 2 polysaccharide genes (*cps2*E,

*cps2*T, *cps2*F, *cps2*G, *cps2*I, *cps2*J and *cps2*H) were PCR amplified from *Streptococcus pneumoniae* strain D39 genomic DNA. Genomic DNA of pneumococcal strain D39 was kindly provided by Professor Peter W Andrew (Department of infection immunity and inflammation, University of Leicester). Tag sequences were introduced by incorporation into PCR primers. Two versions of PCR reverse primers were designed for each type 2 polysaccharide gene: A shorter version of each reverse primer was designed to amplify each of the seven genes from *Streptococcus pneumoniae* genomic DNA and a second full length reverse primer was designed to incorporate the desirable sequences in the PCR product obtained with the shorter reverse primer. The list of primers used for amplification and sequencing of the seven pneumococcal type 2 polysaccharide genes is given in Table 2-5. The detailed description of these primers is given in the respective results sections on the cloning of these genes.

Table 2-4 DNA and amino acids sequences of the Kex2 plant-based processing peptide and the protein tags used in construction of pneumococcal type-2 polysaccharide biosynthetic genes expression cassettes.

Protein tag/ Processing peptide	**a. DNA sequence (5′-3′) and b. Amino acid sequence**	**Gene tagged/ Attachment site**
3×Kex2	a. ATTGGAAAGAGAGGAATTGGAAAGAGAGGAATTGGAAAGAGAGGA b. IGKRGIGKRGIGKRG	At the junction between two genes
Flag tag	a. GATTATAAGGATGATGATGATAAG b. DYKDDDDK	*cps2*T
HA tag	a. TATCCATATGATGTTCCAGATTATGCT b. YPYDVPDYA	*cps2*E
His tag	a. CATCATCATCATCATCAT b. HHHHHH	*cps2*F
E2 tag	a. TCTTCTACTTCTTCTGATTTTAGAGATAGA b. SSTSSDFRDR	*cps2*G
c-MYC tag	a. GAACAAAAGCTTATTTCTGAAGAAGATCTT b. EQKLISEEDL	*cps2*H
VSV-G tag	a. TATACTGATATTGAAATGAATAGACTTGGAAAG b. YTDIEMNRLGK	*cps2*I
Strep-II tag	a. TGGTCTCATCCACAATTTGAAAAG b. WSHPQFEK	*cps2*J

Table 2-5 List of primers used for amplification and sequencing of the seven pneumococcal type-2 polysaccharide genes: Different parts of the primers are highlighted in different colours.

Primer name	Primer sequence (5′-3′) and description
KA014	GCGACTAGTATGAAGAAGTCAGTTTATATC **Description:** Forward PCR primer for *cps2*T with *Spe*1 site
KA015	CGCGCGGCCGCTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATCTTATCATCATCATCCTTATAATCCTCACTTTTTCCCCCTTCAAAC **Description:** Reverse PCR primer for *cps2*T with Flag tag-3×Kex2- *Not*1 site
KA016	GCGGCGGCCGCGATGAATGGAAAAACAGTAAAG **Description:** Forward primer for *cps2*E with *Not*1 site
KA017	CGCCCATGGTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATAGCATAATCTGGAACATCATATGGATACTTCGCTCCATCTCTCATAAATAC **Description:** Reverse primer for *cps2*E with HA tag-3×Kex2-*Nco*1 site
KA018	CTCACTTTTTCCCCCTTCAAAC **Description:** Mini reverse primer for *cps2*T to optimize PCR protocol
KA019	CGCCTGCAGTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATCTTATCATCATCATCCTTATAATCCTCACTTTTTCCCCCTTCAAAC **Description:** Reverse PCR primer for *cps2*T with Flag tag-3×Kex2- *Pst*1 site (alternative to primer KA015 for revised strategy)
KA020	GCGCTGCAGATGAATGGAAAAACAGTAAAG **Description:** Forward primer for *cps2*E with *Pst*1 site (alternative to primer KA016 for revised strategy)
KA021	ACTTCGCTCCATCTCTCATAAATAC **Description:** Mini reverse primer for *cps2*E to optimize PCR protocol, letter highlighted in green is part of the attached HA tag.
KA022	TCGTTCGATCGGAGGCCG **Description:** 1st primer (forward) for sequencing *cps2*T in pKAM2. Total size of *cps2*T is 1263 bp, so three primers were used for its sequencing. The primer binds 356 bp downstream the start codon of *cps2*T.
KA023	ATATAGGTCGTCTTGGGC **Description:** 2nd primer (reverse) for sequencing *cps2*T in pKAM2. The primer binds 542 bp downstream the start codon of *cps2*T.

Table 2-5 List of primers used for amplification and sequencing of the seven pneumococcal type-2 polysaccharide genes (continued)

Primer name	Primer sequence (5′-3′) and description
KA024	GCTCAAAGATACAGGCTTC **Description:** 3rd primer (forward) for sequencing *cps2*T in pKAM2. The primer binds 791 bp downstream the start codon of *cps2*T.
KA025	GCGCCATGGATGGCAACTTATAATGGAGC **Description:** Forward primer for *cps2*F with *Nco*1 site
KA026	CGCGCATGCTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATATGATGATGATGATGATGAATAAACATTAACTCACCGATTAC **Description:** Reverse primer for *cps2*F with His tag-3×Kex2-*Sph*1 site
KA027	AATAAACATTAACTCACCGATTAC **Description:** Mini reverse primer for *cps2*F to optimize PCR protocol
KA028	GCGGCATGCATGACAGGTGGAATAAGAG **Description:** Forward primer for *cps2*G with *Sph*1 site
KA 029	CGCGGGCCCCTGCAGTTATCTATCTCTAAAATCAGAAGAAGTAGAAGATTTACCGTTTTCAATATATACC Description: Reverse primer for *cps2*G with E2 tag-stop codon-*Pst*1-*Apa*1 site
KA030	TTTACCGTTTTCAATATATACC **Description:** Mini reverse primer for *cps2*G to optimize PCR protocol
KA031	CTTATCCACCAATCATGC **Description:** Reverse sequencing primer for *cps2*F. The primer binds 545 bp downstream the start codon of *cps2*F.
KA032	ACGTTGTAAAACGACGGCCA **Description:** Reverse sequencing primer for *cps2*F and *cps2*G. The primer binds outside the T7 primer binding site in pGEM-T Easy vector.
KA033	GTGCAAGTAATCGGTGAG **Description:** Forward sequencing primer for *cps2*G. The primer binds 882 bp downstream the start codon of *cps2*F (adjacent to *cps2*G).

Table 2-5 List of primers used for amplification and sequencing of the seven pneumococcal type-2 polysaccharide genes

Primer name	Primer sequence (5′-3′) and description
KA034	CACAGCACTAAGTAGAT **Description:** Reverse sequencing primer for knowing the correct orientation of the *Pst*1-*cps*2E- *cps*2F- *cps*2G- *Pst*1 insert in pKAM2 (in frame with cps2T to give pKAM3). The primer binds 64 bp downstream the start codon of *cps2*E.
KA035	GCGACTAGTATGCTCTCTCTATATACAG **Description:** Forward primer for *cps2*H with *Spe*1 site
KA036	CGCCCATGGTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATAAGATCTTCTTCAGAAATAAGCTTTTGTTCTTTTTCTTGCTTAGTCAATCTC **Description:** Reverse primer for *cps2*H with c-Myc tag-3×Kex2-*Nco*1 site
KA037	TTTTTCTTGCTTAGTCAATCTC **Description:** Mini reverse primer for *cps2*H to optimize PCR protocol
KA038	GCGCCATGGATGACAAAAAGTATCTTAT **Description:** Forward primer for *cps2*I with *Nco*1 site
KA039	CGCGCATGCTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATCTTTCCAAGTCTATTCATTTCAATATCAGTATAATTTTCTAGTTCCTTATATAG **Description:** Reverse primer for *cps2*I with VSV-G tag-3×Kex2-*Sph*1 site
KA040	ATTTTCTAGTTCCTTATATAG **Description:** Mini reverse primer for *cps2* I to optimize PCR protocol
KA041	GCGGCATGCATGAGTAGAAGATATAATT **Description:** Forward primer for *cps2*J with *Sph*1 site
KA042	CGCGGGCCCCTGCAGTTACTTTTCAAATTGTGGATGAGACCATGTTAGAAACTTTTTTAATTCAC **Description:** Reverse primer for *cps2*J with strep II tag-stop codon-*Pst*1-*Apa*1 site
KA043	TGTTAGAAACTTTTTTAATTCAC **Description:** Mini reverse primer for *cps2*J to optimize PCR protocol

Table 2-5 List of primers used for amplification and sequencing of the seven pneumococcal type-2 polysaccharide genes

Primer name	Primer sequence (5′-3′) and description
KA044	CATTTATAGGAATGAGATTG **Description:** Forward sequencing primer for *cps2*I. The primer binds 1123 bp downstream the start codon of *cps2*H, adjacent to *cps2*I.
KA045	CTGTATGAACCAGGAGATTATATG **Description:** Forward sequencing primer for *cps2*J. The primer binds 981 bp downstream the start codon of *cps2*I, adjacent to *cps2*J.
KA046	GCGACTAGTATGACAAAAAGTATCTTAT **Description:** Revised forward primer for *cps2* I with *Spe*1 site
KA047	CGCCTGCAGTCCTCTCTTTCCAATTCCTCTCTTTCCAATTCCTCTCTTTCCAATCTTTTCAAATTGTGGATGAGACCATGTTAGAAACTTTTTTAATTCAC **Description:** Revised reverse primer for *cps2*J with strep II tag-3×Kex2-*Pst*1 site
KA048	GCGCTGCAGATGCTCTCTCTATATACAG **Description:** Revised forward primer for *cps2*H with *Pst*1 site
KA049	CGCGGGCCCCTGCAGTTAAAGATCTTCTTCAGAAATAAGCTTTTGTTCTTTTTCTTGCTTAGTCAATCTC **Description:** Revised reverse primer for *cps2*H with c-Myc tag- stop codon-*Pst*1 site-*Apa*1 site
KA050	CTGGTAAAATGATAATAAC **Description:** Reverse primer for sequencing *cps2*I. The primer binds 102 bp downstream the start codon of *cps2*I.
KA051	GAAGAGATGTATTTTCCTATG **Description:** Forward sequencing primer for sequencing the junction of *cps2*J and *cps2*H. The primer binds 1237 bp downstream the start codon of *cps2*J.

2.2.3 Primer stock preparation

A 100 µM stock solution of each primer was made in sterile ddH2O. The stock solution was incubated at room temperature for 15 min, vortexed and then stored at −20°C. For

PCR amplification, 1 µl of the primer stock solution was used in a 25 µl reaction volume.

2.2.4 Construction of strep-tag II-mini human insulin gene fusion molecule

The four oligonucleotides KA001, KA002, KA003, KA004 (Table 2-3) designed for synthetic mini-insulin gene were paired. Oligonucleotides KA001and KA002 were annealed and extended to form the 5′ end of the gene fusion, while oligonucleotides KA003 and KA004 were annealed and extended to form the 3′ end. Each 100 µl reaction mix consisted of 10 µl of 10x reaction buffer 2 (NEB buffer 2 was used as the buffer of choice for subsequent restriction of the two generated DNA segments with restriction enzyme *Bsu*361), 1 µl of 100x BSA, 2 µl of 25 µM oligonucleotide (KA001/ KA002 for reaction 1 and KA003/KA004 for reaction 2) and 74 µl of molecular biology grade water (Sigma) in a 0.5 ml microcentrifuge tube. The reaction tubes were transferred to a beaker of hot water (~ 95 °C) on a hotplate. The water was then cooled to room temperature to allow the oligonucleotides to anneal. After annealing, 1 µl of Klenow DNA polymerase I (NEB, Cat. No. M0210S) and 10 µl of dNTP mix (0.25 mM final concentration) were added to each reaction mix on ice and the tubes incubated at 37°C for 1 h. Klenow DNA polymerase was then heat inactivated at 75°C for 20 min. The reaction was then restriction digested by the addition of 1 µl of *Bsu*361 enzyme (NEB) to each tube followed by incubation at 37°C for 1 h. Digested DNA fragments were then purified using a Qiagen PCR purification kit (Cat. No. 28106). Ligation of the two purified products was then carried out using T4 DNA ligase (NEB, Cat. No. M0202S) in a 30 µl reaction volume to generate the full length strep-tag II-mini human insulin gene fusion product. The ligation reaction mix consisted of 10 µl of each of the two purified restricted DNA fragments, 3 µl of 10x T4 DNA ligase buffer, 1 µl T4 DNA ligase and 6 µl H2O (Sigma). The reaction mix was incubated at 16°C overnight. 1 µl of the ligation product was used for PCR amplification.

2.2.5 PCR with Taq DNA polymerase

Polymerase chain reaction (PCR) with Taq DNA polymerase (Bioline, Cat. No. M95801b) was carried out to amplify synthesized mini-insulin gene, confirmation of transgene integration in plant genome and amplification of cDNA. Amplification was

carried out in a 50 µl or 25 µl reaction volumes. A typical 50 µl reaction volume consisted of 5 µl of 10x Taq polymerase reaction buffer, 5 µl of 2.5 mM dNTP mix, 1.5 µl of 50mM MgCl2, 0.5 µl of 5U/µL Taq polymerase, 0.5 µl l of 100 µM forward primer and 0.5 µl of 100 µM reverse primer in molecular biology grade water. The quantity of template DNA amplified was generally 5-50 ng. Amplification was generally carried out for 30 cycles. Each cycle consisted of denaturation at 95°C for 30 seconds, annealing at 52°C for 30 seconds and extension at 72°C for 30-90 seconds depending on the length of template DNA (60 seconds extension for a DNA segment of 1kb). A long denaturation step at 95°C for 5 minutes was also included for efficient denaturation of the template before the first cycle. Similarly, the last cycle was followed by an extra extension step at 72°C for 7 minutes for completion of all the extension reactions.

2.2.6 PCR with Phusion DNA polymerase

Phusion DNA polymerase (NEB, Cat. No. F-530S) was used as the enzyme of choice for the amplification of the seven type 2 pneumococcal polysaccharide genes using *Streptococcus pneumoniae* strain D39 as template. This enzyme has increased fidelity compared to taq DNA polymerase and the error rates are much lower. The amplification conditions were optimized. PCR was carried out in a 50 µl reaction volume containing 10 µl of 5x phusion reaction buffer, 4 µl of 2.5 mM dNTP mix, 1 µl of 100 µM forward primer, 1 µl of 100 µM reverse primer, 5 µl (50ng) of template DNA, 0.5 µl of 2 U/µl phusion DNA polymerase, 0.5 µl (1.5 mM final concentration) of $MgCl_2$ and 28 µl of molecular biology grade water (sigma). Amplification temperature profile involved an initial denaturation step at 95°C for 5 minutes followed by 30-35 amplification cycles each consisting of a denaturation at 95°C for 30 seconds, annealing at 52°C for 30 seconds and extension at 72°C for 90 seconds. The final cycle was followed by an extension step at 72°C for 5 minutes.

2.2.7 PCR purification

PCR products were purified using a Qiagen PCR purification kit (Cat. No. 28106) as described by the manufacturer. Purified products were eluted in 30 µl ddH2O and used immediately or stored at −20°C.

2.2.8 Colony PCR

Colony PCR protocol was based on CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo) colony PCR protocol (CIMMYT, 2005). A single fresh colony was picked with a toothpick from a LB agar culture plate and suspended in 50 µl of TTE buffer (containing 1% triton X-100; 20 mM Tris HCl, pH 8.5; and 2 mM EDTA, pH 8.0) in 0.5 ml microcentrifuge tube. This was incubated at 95°C for 10 min to produce bacterial lysate. The lysed suspension was then spun down for 5 min and 2.5 µl of the supernatant was used for PCR amplification. The amplification was carried as described in section 2.2.5.

2.2.9 Plasmid DNA mini prep (Qiagen)

For the isolation of recombinant plasmids, 2 ml of LB broth contained in a 7 ml sterilin bijou tube was inoculated with a single resistant *E.coli* colony from a culture plate. The culture was grown overnight. The overnight culture (1.5 ml) was then transferred to a microcentrifuge tube and the tube centrifuged at 16100 g for 2 min. The supernatant was discarded and the bacterial pellet resuspended in 250 µl of buffer P1. Then, 250 µl of buffer P2 was added to the suspension and mixed well by inverting. After this, 350 µl of buffer N3 was added and mixed immediately by inverting the tube several times. The tube was then centrifuged for 10 min at 16100 g. The supernatant was transferred to QIAprep spin column and centrifuged for 30 seconds. The QIAprep spin column was washed with 0.5 ml of buffer PB and centrifuged for 30 seconds at 16100 g. It was followed by washing the spin column with 0.75 ml of buffer PE and centrifuging for 30 seconds at 16100 g. The flow-through was discarded and the column spun for an additional 1 min to remove traces of buffer PE. The QIAprep column was then placed in a clean 1.5 ml centrifuge tube and the plasmid DNA eluted by adding 50 µl molecular biology grade water to the centre of the spin column. The isolated plasmid DNA was stored at −20°C.

2.2.10 Diatomaceous earth plasmid DNA mini prep

Diatomaceous earth based plasmid DNA mini prep was used as a cheaper alternative to the commercially available kits for processing large number of samples. A single

antibiotic resistant colony was inoculated into 5 ml LB broth medium and incubated overnight at 37°C. 1.5 ml of the overnight culture was transferred to a microcentrifuge tube, the tube centrifuged for 2 min at 16100 g and the supernatant removed. The pellet was resuspended in 100 µl of solution I (50 mM Tris-HCl, pH 7.5; 10 mM EDTA; and 100 µg/ml RNAse A) and vortexed. Then, 150 µl of solution II (freshly made 0.2 M NaOH in 1% SDS) was added and mixed gently until clear. Next, 150 µl of solution III (60 ml 5 M Potassium Acetate, 11.5 ml glacial acetic acid and 28.5 ml distilled water) was added and mixed by inverting. The tube was then spun at 16100 g for 5 min. The supernatant was transferred to a fresh tube and 1 ml of Diatomaceous Earth resin (Sigma) was added. The tubes was inverted at room temperature for 1 min, spun for 1 min, left for 1 min and spun again for 1 min. The supernatant was discarded. Then, 1 ml of 80% isopropanol was added, the tube vortexed and spun for 1 min. The supernatant was discarded and 1ml of acetone was added. The sample was vortexed and spun again. The supernatant was discarded and the pellet air-dried for 10 min. The pellet was then dissolved in 100 µl ddH2O (the sample was incubated at 65°C for 5 min to completely dissolve DNA). The DNA/resin suspension was then centrifuged for 1 min and the supernatant containing DNA was transferred to a fresh tube. The plasmid DNA was stored at −20°C.

To make Diatomaceous Earth suspension, 25 grams of Diatomaceous Earth (Sigma d5384) was added to 500 ml distilled water and shaken. This was allowed to settle for 3 h and the supernatant was then aspirated. The matrix was resuspended in 100 ml of ddH2O and stored at room temperature. For making the resin, 1ml of Diatomaceous Earth suspension was added to a 25 ml solution of 4 M guanidine hydrochloride, 50 mM Tris-HCl (pH 7.0) and 10 mM EDTA. The resin was stored at room temperature in a dark bottle.

2.2.11 Plasmid DNA midi prep (Qiagen)

Single resistant *E. coli* colony was inoculated into 5 ml LB broth medium containing appropriate antibiotic. The bacterial culture was grown at 37°C and 250 rpm for 8 h. Then, 1ml of the culture was used to inoculate 100 ml of LB broth medium containing suitable antibiotic. The culture was grown at 37°C and 250 rpm overnight. Next day, the culture was divided into two 50 ml falcon tubes and the cells pelleted at 4°C and 6000 g

for 15 min. The pellets were resuspended in 4 ml of buffer P1 (50 mM Tris-Cl, pH 8.0; 10 mM EDTA; 100 µg/ml RNase A) and combined in one tube. To the cell suspension, 4 ml of buffer P2 (200 mM NaOH, 1% SDS) was then added. The sample was mixed well and incubated at room temperature for 5 min. After this, 4 ml of chilled buffer P3 (3.0 M potassium acetate, pH 5.5) was added and the tube incubated on ice for 15 min. It was followed by centrifugation at 4°C and 20,000 g for 30 min. The supernatant was transferred to a new tube and centrifuged again for 15 min at 4°C. The supernatant was then transferred to a new tube. A Qiagen column was equilibrated with 4 ml of buffer QBT (750 mM NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol; 0.15% Triton X-100). The supernatant was then transferred to the equilibrated column and allowed to enter the resin. After this, the column was washed 2 times with 10 ml of buffer QC (1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol). The plasmid DNA was then eluted in 5 ml of buffer QF (1.25 M NaCl; 50 mM Tris-HCl, pH 8.5; 15% isopropanol) in a 50 ml centrifuge tube. DNA was precipitated by adding 3.5 ml room-temperature isopropanol, mixing well and centrifuging for 30 min at 4°C. DNA pellet was washed with 70% ethanol and spun again at 16100 g. The supernatant was removed, pellet air dried and dissolved in 200 µl of water (Sigma).

2.2.12 Plant genomic DNA isolation and analysis by PCR

For PCR analysis, plant genomic DNA was isolated by the method of Edwards *et al* (1991). Fresh leaf sample weighing 100-150 mg was put in a 1.5 ml centrifuge tube and frozen in liquid nitrogen. The sample was then ground with eppendorf grinder and 400 µl of PCR extraction buffer (200 mM Tris-HCl, pH 7.5; 250 mM NaCl; 25 mM EDTA; 0.5% SDS) was added. The sample was vortexed, centrifuged at 16100 g for 10 min and the supernatant transferred to a fresh tube. Then, 400 µl of isopropanol was added to the tube and the sample incubated at room temperature for 2 min. The sample was then centrifuged at 16100 g for 10 min. The supernatant was decanted and the pellet air dried for 10 min. DNA pellet was resuspended in 50 µl of ddH2O and stored at −20°C. The resuspended DNA (2-5 µl) was then used for PCR analysis.

2.2.13 DNA quantification

DNA was quantified using a spectrophotometer or by gel electrophoresis using low DNA mass ladder (Invitrogen). For spectrophotometric determination, the DNA sample was diluted 1:50 in water to make a 200 µl solution. The sample was then transferred to a quartz cuvette. Water was used to blank the spectrophotometer. Absorbance was recorded at 260 nm and 280 nm with *Ultrospec 43000 pro* spectrophotometer (Amersham Biosciences). The ratio between the absorbance values at 260 nm and 280 nm was used to estimate DNA purity. Pure DNA generally has an A_{260}/A_{280} ratio between 1.7 and 2.0. The concentration of DNA was calculated from assuming an absorbance of 1 unit at 260 nm corresponds to 50 µg of DNA per ml. The following formula was used to calculate the concentration of DNA.

Concentration of DNA = 50 µg/ml × A_{260} × dilution factor

To estimate DNA concentrations by gel electrophoresis, a 1.5% agarose gel was used to run 4 µl of DNA sample along with 4 µl of low DNA mass ladder (invitrogen) at 100 V for 40-60 min. Low DNA mass ladder consists of six DNA fragments with predetermined amounts of DNA of each fragment. The amount of DNA in unknown DNA sample can be estimated by comparing by eye the band intensity of the unknown DNA sample with those of the low mass DNA ladder bands.

2.2.14 Agarose gel electrophoresis

Gel preparation and electrophoresis of the sample was carried out as described by Sambrook *et al*., (1989). DNA was separated using 0.8- 2.5 % agarose gels made with TAE buffer (4.84 g/l Tris base; 1.142 ml/l glacial acetic acid; 2 ml /l 0.5 M EDTA, pH 8.0) containing 0.1mg/ml ethidium bromide. Agarose was melted in TAE buffer and then cooled to about 55°C before addition of ethidium bromide inside a fume-hood. The gel solution was then slowly poured into the gel-casting tray with the comb positioned correctly. Air bubbles were pushed away to the side using a disposable tip. After the gel had solidified, the comb was removed carefully and the gel transferred into the gel tank. The gel was immersed in TAE buffer. DNA samples in 1x loading dye (6x loading dye stock contained 0.25% bromophenol blue, 0.25% xylene xyanol and 40% sucrose) and

DNA markers were added to the wells cast in the gel. Samples were electrophoresed for 1 h at 100 V. DNA bands were visualized under a UV transilluminator and recorded by photography.

2.2.15 DNA restriction digestion

DNA restriction digestion was conducted according to the manufacturer's recommendations. Restriction digestion mix consisted of 1x reaction buffer, DNA substrate, restriction enzyme and water to make the final volume. For some enzymes 1% BSA was also included in the reaction mix. Typical reaction volumes involved a 20 µl for a mini prep plasmid DNA restriction digestion (for an analytical gel) and 100 µl for a PCR product and a midi prep plasmid DNA restriction digestion (for a preparatory gel). Digestion was performed at 37°C for 1-2 h. Restriction digestion products were analysed on a gel (analytical gel) or processed further for gel elution (preparatory gel). For restriction digestion involving two enzymes, either a sequential digestion was carried out or the digestion was performed in a reaction buffer suitable for both the enzymes.

2.2.16 DNA dephosphorylation

DNA dephosphorylation was carried out to avoid self-ligation of digested DNA fragments. Dephosphorylation was carried out either directly in the restriction enzyme reaction buffer containing the digested DNA fragments or after purification of the digestion products. In the latter case, dephosphorylation was performed in shrimp alkaline phosphatase buffer (Promega). A concentration of 1 unit/µg DNA of shrimp alkaline phosphatase (Promega) was used to dephosphorylate DNA at 37°C for 20 min. After dephosphorylation, the enzyme was inactivated at 65°C for 15 min. The dephosphorylated products were purified using a Qiagen gel elution/ PCR purification kit.

2.2.17 DNA blunt end creation

For TA-cloning, restricted DNA fragments with 3′ overhangs were treated with Klenow DNA polymerase I large fragment (NEB, Cat. No. M0210S) to remove the 3′ overhangs before adding A-tail. A 50 µl reaction consisted of 20 µl DNA with 3′ overhangs, 5 µl

10x buffer2 (NEB), 2.5 µl of 2.5 mM dNTP mix, 1µl Klenow DNA polymerase I (NEB) and 21.5 µl H2O (Sigma). The reaction mix was incubated at 25°C for 15 min. The blunt-ended products were purified with PCR purification kit (Qiagen).

2.2.18 DNA A-tailing

For TA-cloning, A-tailing in blunt-ended DNA fragments was carried out with Biotaq DNA polymerase (Bioline, Cat. No. M95801b). The reaction was carried out in a 50 µl volume containing 5 µl of 10x reaction buffer, 5 µl of 2.5 mM dNTP mix, 1.5 µl of 50 mM MgCl2, 1.5 µl of Biotaq polymerase and 37 µl molecular biology grade water. The reaction mix was incubated at 70°C for 30 min. The products were then purified with PCR purification kit (Qiagen).

2.2.19 DNA ligation

DNA ligation reaction involved a 20 µl reaction volume for ligation of vector and insert DNA for transformation into *E.coli* and a 100 µl reaction volume for ligation of two PCR products for *in vitro* gene fusion. For vector/insert ligation, 1:1 to 1:5 ratio of vector: insert DNA was used. Vector/insert ligation mix (20 µl) consisted of 50 ng vector DNA, respective amount of insert DNA (amount determined by the following formula), 1 µl of 400 U/µl T4 DNA ligase (NEB, Cat. No. M0202S), 2 µl ligase buffer and water (Sigma) to make the final volume.

Formula to calculate amount of insert:

Insert amount (ng) = (molar ratio of insert / vector) × insert size / vector size × amount of vector (ng)

The 100 µl ligation mixture (*in vitro* gene fusion) consisted of 0.5-1.0 µg of each PCR product, 5 µl of 400 U/µl T4 DNA ligase, 10 µl ligase buffer and water (Sigma) to make the final volume. The ligation mixture was either incubated at 16°C overnight or at room temperature for 20 min. Ligated products were cleaned up (section 2.2.20) and used immediately or stored at −20°C.

2.2.20 DNA ligation mixture clean up

Ligation products were cleaned up by dialysis or by ethanol precipitation. For dialysis, the ligation mixture was transferred to a membrane filter (Millipore, pore size 0.15 µm, Cat. No. VMWP01300) floating on the surface of ddH20 in a Petri dish. Dialysis was carried for 30-45 min. The dialysed ligation products were then transferred to a microcentrifuge tube and 1-2 µl was used for transformation of competent *E.coli* cells.

For ethanol precipitation, ddH20 was added to the ligation mixture to bring the volume up to 100 µl, 10 µl (0.1 vol) of 3 M sodium acetate was added and mixed. This was followed by addition of 250 µl of 100% ethanol (2.5 vol) and mixed by vortexing. The sample was then incubated on ice for 20 min and centrifuged at 4°C for 15 min at 16100 g. The supernatant was removed, 250 µl of 70% ethanol added and the tube vortexed. The mixture was again centrifuged as before and the supernatant removed. The pellet was air-dried and resuspended in 10 µl H2O. For transformation of competent *E. coli* cells, 1-2 µl of the resuspended ligation product was used.

2.2.21 Gel purification

DNA fragments from restriction digestion or PCR were separated on agarose gel (1.5-3.0%) and the fragment of interest was excised from the gel with a sterile blade under UV light on a transilluminator. Efforts were made to expose the gel to as little UV light as possible to avoid degradation of DNA. The excised gel slice was transferred to a microcentrifuge tube. The DNA fragment was then purified with a Qiagen MinElute gel extraction kit (Cat. No. 28604) as per kit protocol.

2.2.22 DNA Sequencing

DNA sequencing work was done at Protein Nucleic Acid Chemistry Laboratory **(**PNACL), University of Leicester, utilizing an Applied Biosystems 3730 sequencer. The sequences obtained were analysed by Chromas DNA sequence analysing software (www.technelysium.com.au/chromas.html) and aligned with the original DNA sequences using Blastn online software package (www.blast.ncbi.nlm.nih.gov/) and EMBOSS Pairwise Alignment Algorithms (www.ebi.ac.uk/Tools/emboss/align/index.html) for pairwise alignment. ClustalW2

software package (www.ebi.ac.uk/Tools/clustalw2/index.html) was used for multiple sequence alignment.

2.3 Plant transformation

Plant transformation work was carried out under sterile conditions inside a laminar-flow workstation. Tobacco (*Nicotiana tabacum cv* SAMSUN) leaf disks were transformed with the *Agrobacterium* strain GV3101 (carrying the gene/s of interest) by method of Draper *et al*., (1988). A 5 ml overnight culture of transformed GV3101 strain was grown in LB broth medium containing 0.01 mg/ml Gentamycin, 0.03 mg/ml rifampicin, and 0.025 mg/ml kanamycin. The overnight culture was centrifuged at 4566 g for 15 min and the supernatant discarded. The pellet was resuspended in LB broth at a 1:20 dilution. The diluted culture was transferred to a sterile Petri dish. Young tobacco leaves were then harvested and dipped in 10 % bleach in a sterile beaker for 15 min. Air bubbles were removed from the leaf surface by gentle agitation with sterile spatula. Leaves were then thoroughly rinsed four times with sterile water. On a sterile white tile damaged areas and midribs were removed with a sterile scalpel blade (dissecting instruments were regularly sterilized by dipping in 100% alcohol, flaming and cooling before use). The leaves were cut into 0.5-1 cm squares with a sterile scalpel or punched with a sterile cork borer. Leaf discs were transferred to the Petri dish containing the diluted *Agrobacterium* culture and incubated for 5 min. The leaf discs were then transferred to sterile blotting paper to remove excess liquid. They were placed upside down on MS agar (4.32 g/l MS salt mixture, Sigma; 30 g/l sucrose and 8 g/l plant agar) in sterile Petri dishes. The plates were sealed with parafilm, wrapped in aluminium foil and incubated at room temperature for 2 days. Leaf discs were then transferred to selective MS agar shoot medium (containing 250 µg/ml cefotaxime to inhibit *Agrobacterium* growth, 100 µg/ml kanamycin or 30µg/ml hygromycin B depending on the type of plant antibiotic selection gene carried by the binary vector used for transformation, 100 µg/ml naphthalene acetic acid (NAA) and 1 mg/ml 6-benzylaminopurine) and incubated in growth room at 24-26°C in continuous light. Leaf discs were moved to new plates after every ten days. Shoot regeneration started 1-2 weeks post infection.

After 6-8 weeks, regenerated shoots were excised from leaf discs and transferred to MS agar rooting medium containing 3% (w/v) sucrose, 250 µg/ml cefotaxime and 100 µg/ml kanamycin and no added hormones. In case of double transformed plants (a plant sequentially transformed first with a plant binary vector carrying a kanamycin plant selection gene i.e. pKAM3 and the resultant transgenic plant retransformed with a second plant binary vector carrying a hygromycin plant selection gene i.e. pKAM8. 100 µg/ml kanamycin and 30µg/ml hygromycin B was included in the rooting medium. Root development was observed after 7-12 days.

After root development, the plantlets were transferred to soil in pots. The pots were then transferred to green house and covered with transparent plastic lids for 3-4 days to maintain humidity. The conditions in the green house were maintained at 60-65% humidity, 16 h photoperiod and 24-26°C temperature. The plants were allowed to grow until seed production. To avoid cross-fertilization, flower buds were covered with paper bags. After maturity, seeds were harvested and stored in bijou vials. Leaf material was harvested occasionally for isolation of DNA, RNA, proteins and polysaccharides.

For continuous supply of plant leaf material, transgenic plants were propagated using axillary buds. Stem from each transgenic seedling (growing on selective rooting medium) was cut into nodal sections (containing an axillary bud). Leaves were removed and each section was placed in selective MS agar medium in a sterile 50 ml vial and incubated in growth room at 24-26°C. After 1-2 weeks, stem sections developed new shoots and roots. Shoots were sub-cultured after every month.

2.4 Transgenic plants analysis

2.4.1 Seed screening

T1 seed from transgenic lines was sterilized with 70 % ethanol for 30 s and then washed 6 times with ddH2O. Sterilized seeds from each line were then placed on selective MS agar medium (containing 3% sucrose and 150 µg/ml kanamycin) in a 15 cm Petri plate. Plates were then sealed with parafilm and incubated at 24-26°C in continuous light. Seed germination started one week after incubation. Resistant plants were scored for segregation ratios after 3-4 weeks of germination. Resistant T1 seedlings (10-12) from

each line were transferred to soil to obtain T2 seed for the production of homozygous T2 plants and to obtain leaf material for transgene/s characterization.

2.4.2 Gene expression analysis at RNA level

2.4.2.1 Total plant RNA isolation

Total plant RNA was isolated by Qiagen RNeasy mini kit (Cat. No 74104). Leaf sample (100 mg) was thoroughly ground in 1.5 ml centrifuge tube and 450 µl of buffer RLT (containing 10 µl/ml β-Mercaptoethanol) was added immediately to the ground sample. The sample was vortexed vigorously and the lysate transferred to a QIAshredder spin column placed in a 2 ml collection tube. The column was then centrifuged at 16100 g for 2 min. The supernatant of the flow-through was transferred to a new microcentrifuge tube. It was followed by adding 0.5 volume of 100 % ethanol to the sample and mixing by pipetting. The sample was then transferred to an RNeasy spin column placed in a 2 ml collection tube and centrifuged at 16100 g for 15. The flow-through discarded and 700 µl of buffer RW1 was added to the column. The column was centrifuged as before and the flow-through discarded. The column was then washed twice with 500 µl of buffer RPE. After this, the column was placed in a new 2 ml collection tube and spun at 16100 g for 1 min to remove residual buffer RPE. Then, the column was placed in a new 1.5 ml collection tube and 40 µl of RNase-free water was added directly to the spin column membrane. RNA was eluted by centrifuging the column at 16100 g for 1 min and stored at −80°C.

2.4.2.2 RNA quantification

To determine RNA concentration, RNA samples were diluted 1:50 in RNase-free water. Diluted RNA sample (200 µl) was then transferred to a quartz cuvette. Prior to use, the cuvettes were washed with 0.1 M NaOH and 1 mM EDTA to remove RNAse contamination. Pure RNase-free water was used to blank the spectrophotometer. Absorbance was recorded at 260 nm and 280 nm with *Ultrospec 43000 pro* spectrophotometer (Amersham Biosciences). The ratio between the absorbance values at 260 nm and 280 nm was used to estimate RNA purity. Pure RNA has an A_{260}/A_{280} ratio of 1.9-2.1 in 10 mM Tris-HCl, pH 7.5. The ratio is influenced by change in pH. As pure water is not buffered, the ratio is subject to great variations. Therefore, the

absorbance reading should be taken using 10 mM Tris-HCl buffer (pH 7.5). The concentration of RNA was calculated by using the fact that an absorbance of 1 unit at 260 nm corresponds to 44 µg of RNA per ml. The following formula was used to work out RNA concentration.

Concentration of RNA sample= 44 µg/ml × A_{260} × dilution factor

2.4.2.3 RNA clean-up (rDNase 1 treatment)

Ambion DNA-*free* DNase treatment kit (Cat. No. AM1906) was used for the removal of DNA contamination from plant RNA (section 2.4.2.1). A 50 µl reaction consisted of 10 µg RNA, 0.1 volume 10x DNase 1 buffer, 1 µl rDNase 1 and RNase-free water to make the final volume. The reaction mixture was incubated at 37°C for 30 min. It was followed by adding 5 µl DNase inactivation reagent to the sample and incubating at room temperature for 2 min, with occasional mixing. The sample was then centrifuged at 16100 g for 2 min and the RNA supernatant was transferred to a new tube.

2.4.2.4 RT PCR

Gene expression at the RNA level was confirmed by RT-PCR. For the synthesis of first-strand cDNA, 2 µgs of plant RNA was mixed with 1 µg of gene specific reverse primer in a total volume of 11 µl in a sterile, nuclease-free microcentrifuge tube. The mixture was incubated at 70 °C for 5 min and subsequently chilled on ice for 5 min. The tube containing RNA/primer mix was then centrifuged briefly to collect the solution at the bottom of the tube. Then, 3 µl of AMV Reverse Transcriptase (Promega, Cat. No. M5101), 5 µl of AMV Reverse Transcriptase 5 x buffer, 2.5 µl of 10 mM dNTP mix, 1µl of RNasin Ribonuclease Inhibitor (Promega, Cat. No. N2111) and 2.5 µl of nuclease-free water was added to the annealed RNA/primer mix to make 25 µl final volume. The reaction mix was incubated at 42 °C for 60 min to make the first strand of cDNA. For PCR amplification, 2-3 µl of the cDNA solution was used (section 2.2.5).

2.5 Gene expression analysis at protein level

2.5.1 Protein extraction

2.5.1.1 Total plant protein extraction

Fresh leaf material (~0.15 g) was ground in liquid nitrogen in a 1.5 ml microcentrifuge tube with an eppendorf grinder. The ground tissue was homogenized in 200 µl of extraction buffer (100 mM Tris, pH 6.8; 10% glycerol; 0.5% SDS; 0.1% Triton X-100; 5 mM EDTA; freshly made 10 mM DTT; and one protease inhibitor cocktail tablet (Roche) for 10 ml buffer) and centrifuged at 4°C and 16100 g for 10 min. The supernatant was transferred to a fresh tube and used immediately or stored at −20°C.

2.5.1.2 Apoplast protein extraction

Freshly harvested tobacco leaves were cut into 1 cm square sections avoiding midribs. The cut sections were placed into a small beaker containing 50 ml of 50 mM $CaCl_2$. A conical flask which fitted well to the side of the beaker was then placed on the top of the liquid to ensure full immersion of the tissue pieces. The beaker was then enclosed in vacuum desiccators and the leaf sections were infiltrated by attaching the desiccators to a vacuum supply for a period of 30 minutes (by which time the tissue had taken on a transparent appearance). The leaf samples were then removed and dried gently on a paper towel before being transferred to the barrel of a 20 ml syringe with the plunger removed. The syringe containing the leaf sample was placed in a 50 ml tube and spun at 10°C and 1416 g for 10 minutes. The extra cellular wash accumulated at the bottom of the tube. The sample was spun again to ensure maximum recovery of the apoplastic wash. The fluid was used fresh or stored at −80°C.

2.5.1.3 Apoplast protein precipitation

To concentrate apoplast protein extract, 4 volumes of methanol were added to 1 volume of apoplastic fluid and the mixture incubated at −20 °C for 1-2 h. The sample was then centrifuged at 4 °C and 16100 g for 30 min. The supernatant was discarded and the protein pellet air-dried for 10 min. The dried pellet was resuspended in 100 µl of 2X protein loading buffer (1ml 1M Tris-HCl, PH 6.8; 2 ml 10% SDS; 4 ml 50% Glycerol; 1

ml 0.2% bromophenol blue and 2 ml ddH2O to make 10 ml final volume) and boiled for 10 minutes before being used for SDS-PAGE and western blotting analysis.

2.5.1.4 Protein quantification

Protein concentration was determined by Bradford assay (Bradford, 1976) using Sigma Bradford Reagent (Cat. No. B6916). Bovine serum albumen (BSA) was used as standard. Serial dilutions of BSA were made containing 0 mg/ml, 0.25 mg/ml, 0.5 mg/ml, 1 mg/ml and 1.4 mg/ml BSA respectively. A 200 µl working solution of each BSA dilution in ddH2O also contained 20 µl of protein extraction buffer (to account for any differences caused by extraction buffer in the plant protein extract). Similarly, 1/10 dilution of each plant protein sample was made in ddH2O. Next, BSA serial dilutions and diluted plant samples (50 µl each) were added into separate disposable cuvettes. 1.5 ml of Bradford reagent was added to each cuvette. The samples were mixed gently and incubated at room temperature for 5 min. Absorbance readings were then recorded at 595 nm with an *Ultrospec 43000 pro* spectrophotometer (Amersham Biosciences). A standard curve was plotted by using A_{595} readings of BSA serial dilutions. Protein concentration in the unknown sample was then determined by comparison with the standard curve.

2.5.2 Polyacrylamide gel electrophoresis (SDS-PAGE)

BIORAD mini-protein II gel system was used for separation of proteins through polyacrymamide gel electrophoresis (SDS-PAGE). The apparatus was cleaned and the glass plates, spacers and combs were rinsed with 100% ethanol and dried before use.

2.5.2.1 Tris-Tricine SDS-PAGE

Tris-Tricine SDS-PAGE was used for separation of the synthetic mini human insulin (size of mini-insulin is ~ 6 kDa) from plant proteins. The recipe for making Tris-Tricine gel is given in Table 2-6. Resolving gel solution was poured into the casting system with a 5 ml disposable pipette. At the top of the gel, 400 µl of water saturated isoamyl alcohol (50:50 v/v) was added to isolate polymerization from oxygen. After 30 min polymerisation, the water saturated isoamyl alcohol was discarded and top of the gel

was washed with ddH2O. Gel combs were inserted into the casting system and the stacking gel solution was added with 1000 µl pipette onto the top of the resolving gel.

Table 2-6 Recipe for making 16% Tris-Tricine gel

Stock Solution	Resolving Gel	Stacking Gel
29:1 acrylamide/bisacrylamide	10.86 ml	2.42 ml
Tris-Cl/SDS	10.00 ml	6.2 ml
(3M Tris-Cl, 0.3% SDS, pH8.45)		
water	5.97 ml	16.38 ml
Glycerol	3.17 ml	---------
25%(w/v) ammonium persulfate	50 µl (freshly made)	50 µl (freshly made)
TEMED	15 µl	30 µl

After polymerisation the gel combs were removed and the wells were washed with ddH2O. Cast gels were then transferred to the gel running tank. The upper gel tank was filled with 1x cathode buffer (0.1 M Tris, 0.1 M Tricine and 0.1% SDS) and the bottom tank was filled with 1x anode buffer. Before loading onto the gel, the protein samples were denatured in 2x protein loading buffer (1ml 1M Tris-HCl, pH 6.8; 2 ml 10% SDS; 4 ml 50% Glycerol; 1 ml 0.2% bromophenol blue and 2 ml ddH2O to make 10 ml final volume) by boiling for 10 min. Protein samples (10-30 µg) and suitable protein marker (Invitrogen SeeBlue Plus2 protein standard, Cat. No. LC5925 or Sigma Ultra Low Range colour marker, Cat. No. C6210) was loaded into the wells and the proteins were electrophoretically separated at 120 volts for 45-60 min. After SDS-PAGE, separated proteins were analysed by coomassie staining or by western blotting.

2.5.2.2 Tris-Glycine SDS- PAGE

Tris-Glycine SDS-PAGE was performed for separation of the larger pneumococcal type 2 polysaccharide proteins. Recipe for making Tris-Glycine gel is given in Table 2-7. Procedure for gel casting and electrophoresis was the same as for Tris-Tricine SDS-PAGE (section 2.5.2.1) except that electrophoresis was performed with a single running buffer (3.03 g/l Tris, 1 g/l SDS and 14.4 g/l glycine) both in the upper gel tank and the bottom gel tank.

Table 2-7 Recipe for making 12% Tris-Glycine gel

Stock Solution	Separating Gel	Stacking Gel
29:1 acrylamide/bisacrylamide	8.0 ml	2.42 ml
Tris 1.5 M , pH 8.8	5.0 ml	------
Tris 1 M, pH 6.8	-------	1.25 ml
10% SDS	0.2 ml	0.1 ml
TEMED	8.0 µl	10 µl
25% ammonium persulfate	200 µl (freshly made)	100 µl (freshly made)
H2O	6.6 ml	6.8 ml

2.5.3 Coomassie Staining

For Coomassie blue staining, the protein gel containing the separated proteins was transferred to a staining solution (100 ml/l acetic acid, 450 ml/l ddH2O, 450 ml/l methanol and 2.5 g/l Coomassie blue G-250) and incubated at room temperature overnight on orbital shaker. The gel was then destained in destaining solution (100 ml/l acetic acid, 450 ml/l ddH2O and 450 ml/l methanol) for 4-6 h changing the solution several times. The protein gel was then photographed with a digital camera or a transilluminator using a white light converter plate.

2.5.4 Western blotting

For western blot detection, SDS-PAGE separated proteins were transferred from the gel to a PVDF (Polyvinylidene fluoride) membrane or nitrocellulose membrane using a BIORAD mini-protein II tank blot system. PVDF membranes were briefly wetted in methanol. The membrane and protein gel were then equilibrated in transfer buffer (3.03 g/l and 14.4 g/l Glycine) for 2-3 min. The blotting apparatus was assembled according to the manufacturer instructions. Protein transfer was carried out at 100 V for 60- 90 min. After blotting, the membrane was washed in PBS (8 g/l NaCl, 0.2 g/l KCl, 1.44 g/l Na2HPO4 and 0.24 g/l KH2PO4, pH 7.4) and then blocked in PBST (PBS with 0.05 % Tween-20) containing 4 % skimmed milk for one 1 h at room temperature on orbital shaker or overnight at 4 °C in a refrigerator. The membrane was then incubated with a suitable dilution of primary antibody (Table 2-8) in PBST-0.5% Marvel for 1 h. The membrane was then washed 3 times in PBST-0.5% milk for 10 min each. The

membrane was incubated with 1/2000 dilution of polyclonal goat anti-mouse horse radish peroxidise (HRP) conjugate secondary antibody (DakoCytomation, Cat. No. P0447) in PBST-0.5% milk for 75 min. The membrane was then washed 3 times in PBST-0.5% milk for 10 min each. Proteins were then detected using the EZ-ECL detection kit (Amersham, Cat. No 20-500-120A & 20-500-120B). An equal volume of EZ-ECL solution A and B was mixed to get sufficient solution to cover the membrane. The mixture was equilibrated for 5 min and then added directly onto the blot membrane (protein side up). After 1-3 min incubation, the membrane was wrapped in a saran wrap. Proteins on the membrane were detected by exposing the membrane to an ECL Hyperfilm (Amersham, Cat. No. 28-9068-38) from 30 s to 15 min. The X-rays film was then developed using a Xograph Compact X4 imaging system. Alternatively, the proteins were detected with a Photek HRPCS photon counting system.

For a protein dot blot, 2-3 µl of protein solution was first spotted onto a nitrocellulose membrane and the membrane was allowed to dry. The rest of blotting methodology was essentially the same as for the western blotting above.

Table 2-8 Primary antibodies used for detection of the expressed proteins in transgenic tobacco

Primary antibody	**Protein detected**	**Dilution used**	**Source**
Mouse Monoclonal (E2E3) to Insulin	insulin	1/200	Abcam, Cat. No. ab9569
Mouse Anti-Flag M2 Monoclonal Antibody	cps2T	0.2 µg/ml	Sigma, Cat. No. F3165
Mouse Monoclonal Anti-HA Clone HA-7	cps2E	1/10000, 1/5000	Sigma, Cat. No. H9658
Mouse Monoclonal (5E11) to E2 tag	Cps2G	1/500	Abcam, Cat. No. ab977
Mouse Anti c-MYC (9E10) (Culture supernatant)	cps2H	1/200	ADAS, University of Leicester, UK
Mouse Strep-tag II Monoclonal Antibody	insulin and cps2J	1/1000	Novagen, Cat. No. 71590-3
Mouse Tetra-His Antibody	cps2F	1/5000	Qiagen, Cat. No. 000000000000034670

2.5.5 Strep-tag II purification of synthetic mini human insulin

Fresh tobacco leaves (weighing 2-4 gram) were ground in liquid nitrogen using a mortar and pestle. The ground tissue was homogenized in protein extraction buffer (section 2.5.1.1) or strep tag extraction buffer (100 mM Tris, pH 8.0; 5 mM EGTA; 5 mM EDTA; 150 mM NaCl; 10 mM DTT; 0.5% Triton X-100 and 1 protease inhibitor cocktail tablet for 10 ml buffer). The slurry was transferred to a 15 ml centrifuge tube and centrifuged at 4566 g for 20 min. The supernatant was transferred to a fresh tube and 150 µl of Strep-Tactin Sepharose 50% suspension (IBA, Cat. No. 2-1201-002) was added to it. The mixture was incubated at 4 °C for 1-2 h on a rotating disk. The sample was then centrifuged and the supernatant removed. The resin was washed 5 times with 4 ml of Buffer W (100 mM Tris-HCl, pH 8.0; 150 mM NaCl and 1 mM EDTA). After washing, 300 µl of 2x protein loading buffer was added and the sample boiled for 5-10 min before being used in SDS-PAGE. Alternatively, 200 µl of elution buffer (100 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA and 2.5 mM desthiobiotin) was added to the washed resin, mixed well and centrifuged at 16100 g for 2 min. The supernatant was transferred to a fresh tube and used immediately or stored at −80 °C.

2.5.6 Trypsin maturation of the synthetic mini human insulin

Synthetic mini human insulin was matured by trypsin treatment to remove the attached strep-tag II and the central mini-C pro-peptide from the recombinant insulin following the method of Nykiforuk *et al* (2006). A 1/5 dilution of plant protein extract (containing no added protease inhibitors) was made in 50 mM Tris-HCl (pH 6.8) and treated with trypsin (from bovine pancreas; Sigma-Aldrich, Cat. No. T1426) at a trypsin: total protein ratio of 1: 300 in 50 mM Tris-HCl (pH 8.0) on ice for 90 min. The reaction was stopped by adding protease inhibitor cocktail (Roche). Trypsin treated protein was then analysed using the Mercodia Ultrasensitive Insulin ELISA kit (Cat. No. 10-1132-01).

2.6 Insulin ELISA

2.6.1 Plate trapped antigen (PTA) ELISA for insulin

An optimized plate trapped insulin ELISA assay was developed. Microplate wells (Thermo Fisher Scientific, Cat. No. 439454) were coated with serial dilutions of bovine

insulin (Sigma, Cat. No. 15500-50MG) and plant protein extracts (diluted in PBS with 2% SDS and boiled for 5 min). The plate was incubated overnight at room temperature. Next day, the plate was washed 3 times with PBS, tapping the plate on tissues between washes. The plate was then blocked by filling each well with 400 µl PBS-3 % Marvel milk and incubating for 1 h at room temperature. After blocking, the plate was washed 3 times with PBS and 100 µl of a 1/1000 dilution of mouse monoclonal (E2E3) insulin primary antibody (Table 2-8) in PBS-3% milk was added to each well. The plate was incubated at room temperature for 1 h followed by washing the plate 3 times in PBS. Then, 100 µl of a 1/2000 dilution of AP-Goat anti mouse secondary antibody (Zymed, Cat. No. 62-6622) in PBST-3% milk was added to each well and the plate incubated at room temperature for 1 h. Unbound secondary antibody was then removed by washing the microplate wells 3 times in PBST and 3 times in PBS sequentially. Substrate was prepared by dissolving one SIGMA*FAST* p-Nitrophenyl phosphate tablet and one Tris buffer tablet (Cat. No. N2770-5SET) in 20 ml ddH2O. 100 µl of the substrate solution was added to each well and the plate incubated at room temperature for development of yellow colour. After sufficient yellow colour had developed, absorbance was read at 405 nm (reference 630 nm) with a Tecan Genios plus plate reader. A standard curve was developed from the OD_{405} readings of the standard bovine insulin serial dilutions, which was then used to determine the amount of insulin in the plant extract. In instances of very weak signals or no signal, the plate with substrate was wrapped in cling film and left in dark at room temperature overnight and OD_{405} readings were recorded again in the morning.

2.6.2 Mercodia Ultrasensitive Insulin ELISA

Mercodia Ultrasensitive Insulin ELISA (Cat. No. 10-1132-01) is based on a direct sandwich ELISA technique, which utilizes two monoclonal antibodies directed against separate epitopes on insulin molecule. One monoclonal antibody is bound to microplate well to which the insulin molecule binds. Another peroxidise-conjugated monoclonal antibody then binds to the bound insulin molecule. Washing after incubation removes the unbound labelled antibody. The bound labelled antibody is then detected by reaction with 3,3´,5,5´-tetramethylbenzidine (TMB).

The ELISA assay was conducted as suggested by the manufacturer. 25 µl of each Calibrator (containing 0, 0.13, 0.865, 2.5875, 8.62 and 17.24 pg/µl recombinant human insulin respectively) and plant protein sample was added into microplate wells in duplicate. Then, 100 µl of peroxidase conjugated mouse monoclonal ant-insulin antibody solution was added into each well and the plate incubated on a plate shaker at room temperature and 700 rpm for 1 h. The plate was then washed 6 times with wash buffer followed by adding 200 µl substrate TMB solution into each well. The plate was incubated for 30 min at room temperature. The reaction was then stopped by adding 50 µl of stop solution (0.5 M H_2SO_4) into each well. Optical density readings were recorded at 450 nm with a Tecan Genios plus plate reader. A standard curve was developed from the OD_{450} readings of calibrators, which was then used to work out the amount of insulin in the plant extract.

2.7 Type 2 polysaccharide analysis

2.7.1 Polysaccharide extraction

2.7.1.1 Crude extraction in water

Total plant polysaccharide was extracted in water with (and without) sonication. Midribs from fresh tobacco leaves were removed and 2 grams of leaf tissue were ground in liquid nitrogen using a mortar and pestle. The ground tissue was homogenized in 1 ml ddH2O. The slurry was transferred to a 50 ml tube and sonicated on ice using the small sonicatior probe (Sonitech) at an amplitude of 50 microns with 30 seconds sonication and 1 minute cooling, repeated for 6-8 bursts, or until the colour of the suspension appeared more green compared to the original slurry before sonication. The lysate was then centrifuged at 4°C and 16100 g for 5 min. The supernatant was transferred to a fresh tube and stored at 4°C.

2.7.1.2 Squeezed plant extract (leaf juice)

For isolation of leaf juice, fresh leaf material was squeezed in an eppendorf tube with the help of an eppendorf grinder. The sample was then centrifuged at 16100 g for 10 min and the supernatant was transferred to a fresh tube and stored at 4°C.

2.7.2 Purification of polysaccharides

Polysaccharides were purified by the method of Gilbert *et al* (2000). Proteins from the plant polysaccharide extract were precipitated by adding 25% (v/v) TCA and incubating on ice for 15 min. The sample was then centrifuged at 4 °C and 16100 g for 5 min. The supernatant was transferred to a fresh tube and an equal volume of ice-cold 100 % acetone was added to precipitate polysaccharides. The sample was centrifuged again, supernatant discarded and the pellet washed with acetone. The pellet was then re-dissolved in 5% (w/v) sodium acetate. The pellet was extracted twice with 1/5 volume of chloroform:butanol (5:1), precipitated again in an equal volume of acetone and finally dissolved in ddH2O (to completely dissolve the pellet, the solution was warmed at 70°C for 5 min).

2.7.3 Polysaccharide Quantification

Polysaccharides were quantified by the method of Dubois *et al* (1956). Polysaccharides, polysaccharide derivatives, oligosaccharides and simple sugars with free or potentially free reducing groups react with phenol and sulphuric acid to give a stable orange-yellow colour (Dubois *et al.*, 1956). The colour is proportional to the relative concentration of the reacting sugar/s and hence can be used for quantification of polysaccharides.

Glucose was used as standard for generation of standard curve. Serial dilutions of 1 mg/ml glucose solution were made in 0.5 ml ddH20 each. Similarly, 1-20 µl of plant polysaccharide sample was mixed with 0.5 ml ddH2O. After this, 12.5 µl of an 80% (v/v) phenol solution was added to the glucose serial dilutions and plant samples inside a fume hood. To this, 1.25 ml of concentrated sulphuric acid was added quickly (to prevent evaporation due to the heat produced during the reaction). The samples were incubated at room temperature for 10 min and 1 ml of each sample was then transferred to a disposable cuvette. OD_{490} readings were recorded with *Ultrospec 43000 pro* spectrophotometer (Amersham Biosciences). A standard curve was made from the OD_{490} readings of glucose serial dilutions for determination of polysaccharides in the plant samples.

2.7.4 Type 2 polysaccharide detection

2.7.4.1 Ouchterlony assay

Ouchterlony double-diffusion immuno assay (Ouchterlony and Nilsson, 1973) was used for detection of type 2 pneumococcal polysaccharide in plant extracts. The assay is based on diffusion of both the antigen and antibody molecules through agar and formation of a precipitate (antibody-antigen complex) at the site where the antibody and antigen meet in the agar. The precipitate is visible as a white arc.

Microscope slides (76 mm × 26 mm × 1 mm) were first coated with 0.2% (w/v) Ouchterlony agarose (Difco) in Barbitone buffer (1.84 g/l Diethylbarbituric acid, 10.3 g/l sodium diethylbarbiturate, pH 8.6 with NaOH). The slides were left to dry and then overlaid slowly with 4.5 ml of 1% (w/v) agarose in Baribitone buffer. After the slides had set, equally spaced holes (one hole in centre and the remaining surrounding the central hole) of 4 mm in diameter were made in the solidified agarose using a cork borer. Excised portions were removed with a needle. To the outer holes, 20 µl of each transgenic plant extract, wild type plant extract and 20 µl (0.5 µg/ µl) of type 2 pneumococcal polysaccharide (LGC Standards, Cat. No. ATCC-165-X) was added. To the central hole, 20 µl of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) was added. The slide was then incubated at 4°C for 1-2 weeks in a humidity box to prevent drying of samples. The wells were refilled after every 1-2 days to provide a continuous supply of antigen and antibody for diffusion into the agarose. After enough antigen-antibody complex had formed, a white arc was visible at the antigen-antibody meeting point (only in the positive control sample). The slide was then photographed to record the results.

2.7.4.2 Polysaccharide ELISA

An indirect type 2 polysaccharide ELISA assay was developed. Microplate wells (Thermo Fisher Scientific, Cat. No. 439454) were coated with serial dilutions (in PBS) of type 2 pneumococcal polysaccharide (LGC Standards, Cat. No. ATCC-165-X) and plant polysaccharide extracts (diluted in PBS). The plate was wrapped in cling film and incubated at room temperature overnight. The plate was then washed 3 times with PBS, tapping the plate on tissues between washes. Afterwards, the plate was blocked by

filling each well with 400 µl PBS containing 3% Marvel milk and incubating at room temperature for 1 h. After blocking, the plate was washed 3 times with PBS and 100 µl of a 1/1000 dilution of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) in PBS containing 3% milk was then added to each well. The plate was incubated at room temperature for 1 h and then washed 3 times in PBS. Then, 100 µl of a 1/2000 dilution of Goat Anti-Rabbit IgG (H+L) alkaline phosphatase conjugate secondary antibody (Zymax, Cat. No. 81-6122) in PBST-3% milk was added to each well and the plate incubated at room temperature for 1 h. Rest of ELISA procedure was the same as for Insulin ELISA (section 2.6.1).

2.7.4.3 Tissue printing

Tissue printing method was based on a modified version of Pont-Lezica and Pont-Lezica (2009). Tobacco leaf petiole sections of ~1 mm thickness were made with a sharp razor and pressed firmly on nitrocellulose membrane. The membrane was air-dried and then blocked in TBST (0.05% Tween 20, 0.8 g/l NaCl, 0.02 g/l KCl, 0.3 g/l Tris base, pH 7.5) containing 3% milk for 1 h (or overnight). After blocking, the membrane was incubated in 1/1000 to 1/15000 dilution of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) in TBS-3% milk for 1-2 h at room temperature. The membrane was then washed 3 times in TBST for 10 min each. It was followed by incubating the membrane with 1/2000 dilution of Goat Anti-Rabbit IgG (H+L) alkaline phosphatase conjugate secondary antibody (Zymax, Cat. No. 81-6122) in TBS-3% milk for 1-2 h at room temperature. Then, the membrane was washed 3 times in TBST for 10 min each. After washing, the membrane was covered in SIGMA*FAST* BCIP/NBT alkaline phosphatase substrate (Cat. No. B-5655) for 5-15 min for signal development. The membrane was then washed in ddH2O and photographed.

2.7.4.4 Polysaccharide dot blot

An easy and rapid dot blot assay for detecting nano gram quantities of type 2 pneumococcal polysaccharide was developed. Serial dilutions of type 2 pneumococcal polysaccharide (LGC Standards, Cat. No. ATCC-165-X) were made in ddH2O and in wild type plant extract. Next, the standard serial dilutions and 3-4 µl of each plant leaf

juice sample (section 2.7.1.2) were spotted on nitrocellulose membrane. The membrane was air dried and blocked in PBST containing 4% milk for one 1h at room temperature on orbital shaker or overnight at 4 °C in a refrigerator. The membrane was then incubated with 1/2000 to 1/8000 dilution of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) in PBS containing 1% milk for 1-2 h at room temperature. After this, the membrane was washed 3 times in PBST containing 0.5% milk for 10 min each. Rest of the detection procedure was the same as for tissue printing (section 2.7.4.3).

In some cases the interfering chlorophyll in the leaf extract was removed by incubating the spotted membrane in 1-2% NaOCl solution for 2-5 min before blocking the membrane.

2.7.4.5 Polysaccharide gel electrophoresis and immunoblotting

Type 2 polysaccharide SDS-PAGE and immunoblotting analysis was based on the method of Bender *et al* (2003). Plant polysaccharides samples were treated with 1 µg/ µl Proteinase K (Sigma, Cat. No. P-6556) for 1 h at 37 °C. The plant polysaccharide samples and serial dilutions of type 2 pneumococcal polysaccharide (LGC Standards, Cat. No. ATCC-165-X) were then boiled in 2x loading buffer (1ml 1M Tris-HCl, pH 6.8; 2 ml 10% SDS; 4 ml 50% Glycerol; 1 ml 0.2% bromophenol blue and 2 ml ddH2O to make 10 ml final volume) for 5-10 min. The boiled samples were then loaded onto a 12 % Tris-Glycine gel. Polysaccharides were electrophoretically separated at 100 V until the dye front reached the bottom of the gel. Polysaccharides from the gel were then transferred to a nitrocellulose membrane using a BIORAD mini-protein II tank blot system at 100 V for 60- 90 min. After blotting, the membrane was washed in PBS and then blocked in PBST containing 4 % skimmed milk for one 1 h at room temperature. Next, the membrane was incubated with 1:1000 dilution of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code 16745) in PBS containing 0.5 % milk for 2 h at room temperature. The membrane was then washed 3 times in PBST-0.5% milk for 10 min each. Afterwards, the membrane was incubated with 1/2000 dilution of Goat Anti-Rabbit IgG (H+L) alkaline phosphatase conjugate secondary antibody (Zymax, Cat. No. 81-6122) in PBS with 3 % milk for 1 h at room temperature. The membrane was then washed 3 times in PBST-0.5 % Marvel for 10 min each. After

washing, the membrane was incubated in SIGMA*FAST* BCIP/NBT alkaline phosphatase substrate (Cat. No. B-5655) for 5-15 min for signal development. After sufficient colour had developed, the membrane was washed in ddH2O and photographed.

2.7.4.6 Polysaccharide immunolocalization

Immunolocalization protocol was based on the method of Willats *et al* (2001). Tobacco leaf petioles were cut into 100-300 µm sections with a sharp blade. Sections were transferred quickly to fixation solution containing 4% freshly made paraformaldehyde in PIPES buffer (50 mM PIPES, 5 mM MgSO4 and 5 mM EGTA). Fixation was carried out for 30 min. Sections were then washed in PIPES buffer and incubated in 1/30 to 1/200 dilution of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) in PBS-5% milk for 1 h. Afterwards, sections were washed 3 times in PBS and incubated in 1/150 dilution of Goat Anti-Rabbit IgG FITC conjugate secondary antibody (Sigma, Cat. No. F-0511) for 1 h. Sections were then washed 3 times in PBS and mounted in anti-fade agent Vectashield (Vector Laboratories, Cat. No. H-100) on microscope slides. The slides were examined with a Nikon C1Si confocal microscope and data was recorded by photography.

3 EXPRESSION OF SYNTHETIC HUMAN INSULIN IN TRANSGENIC TOBACCO

3.1 Chapter overview

This chapter describes the expression of human insulin in transgenic tobacco as an alternative platform for mass production of this important pharmaceutical compound to serve as a cheaper method of production to meet future demands. This was carried out by the synthesis of mini human insulin gene (it owes the name to the fact that the central full length C- peptide of native human insulin is replaced by a trypsin cleavable mini C-peptide that is only 3 amino acids in length), cloning of the mini-insulin into various vectors for expression in tobacco. The construct produced was used for *Agrobacterium* mediated transformation of tobacco, for *in planta* expression of synthetic mini-insulin. This chapter describes the analysis of the transgene integration in the plant genome, analysis of insulin gene expression at RNA level, and the analysis of insulin protein expression.

3.2 Construction of strep-tag II-mini human insulin gene fusion molecule

A diagram depicting various parts of the strep-tag II-mini-insulin gene fusion molecule and an outline of its isolation from plant is given in Figure 3-1. The synthetic mini-insulin gene was constructed *in vitro* as a strep tag II-mini-insulin gene fusion molecule by attaching a strep tag II sequence to the 5′ end of mini-insulin gene. The synthetic mini-insulin is encoded by B-chain and A-chain of DesB30 form of human insulin interposed by a trypsin cleavable mini C peptide (AAK). The DesB30 form of insulin precursor lacks the amino acid threonine at position B30 in the B-chain. Hence, B-chain of DesB30 insulin carries 29 amino acids instead of 30 amino acids in the native insulin molecule. This deletion helps in improving the pharmacokinetic properties of insulin molecule. Attachment of the strep tag II in principle should allow subsequent purification of insulin from tobacco plant material. Strep-tag II is an 8-amino acids (WSHPQFEK) protein tag which binds to Strep Tactin matrix. The bound protein attached to the Strep-tag II can be eluted from the StrepTactin matrix by washing with biotin or desthiobiotin containing buffers (Skerra and Schmidt, 2000). A trypsin-cleavable mini peptide (AAK) was also introduced in between the strep tag II and mini-

insulin. Tobacco PR1b endoplasmic reticulum targeting sequence (Cutt *et al.*, 1988) was attached to the 5′ end of mini-insulin gene fusion molecule. The endoplasmic reticulum is the site for disulphide bond formation. After processing, the protein would be exported to the apoplast, which should result in higher accumulation of the protein and make the purification process easier and cheaper. After purification, the recombinant desB30 insulin can be matured *in vitro* by trypsin treatment that cleaves the molecule at the two mini C-peptide sites (AAK) and releases the B and A chains as a single insulin molecule (the two chains are still attached by the two inter chain disulphide bonds), see Figure 3-1.

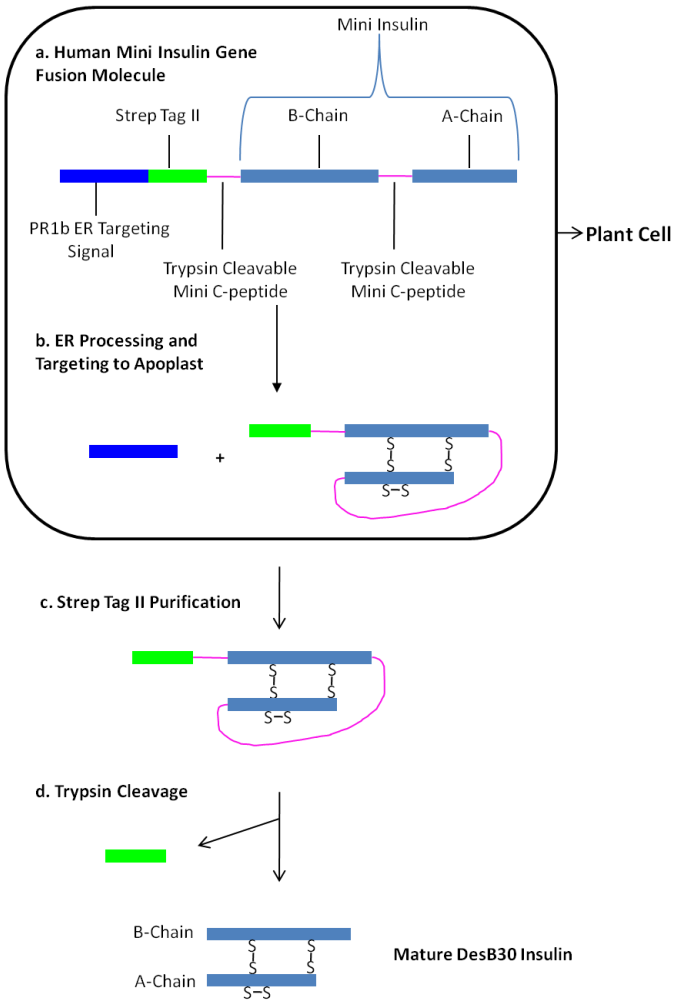


Figure 3-1 Diagram of the recombinant human mini-insulin gene fusion molecule and outline of mature recombinant insulin isolation from plant: (a) The synthetic mini-insulin gene fusion molecule is encoded by the B and A chain of desB30 form (lacking amino acid threonine at position B30 in the B-chain) of human insulin precursor. The two chains are separated by a trypsin cleavable mini C-peptide (AAK). A strep tag II protein tag is attached to the 5′ end of the mini-insulin and is separated by another trypsin cleavable mini C-peptide (AAK) from the mini-insulin. The tobacco PR1b endoplasmic reticulum targeting sequence is attached next to the strep tag II for ER targeting. **(b)** After translation in the

cytoplasm, the mini-insulin fusion protein is targeted to the endoplasm reticulum by the PR1b signal. Here in the endoplasm reticulum, the two inter chain disulphide bonds (one bond between cysteine residue 7 of B-chain and cysteine residue 7 of A-chain and another bond between cysteine residue 19 of B-chain and cysteine residue 20 of A-chain) and one intra chain disulphide bond (between cysteine residues 6 and 11 of A-chain) are formed. The PR1b signal peptide is removed and the protein is exported to the apoplast. **(c)** The expressed mini-insulin fusion protein can in principle be separated from the plant proteins by the attached strep-tag II using Strep Tactin purification columns. **(d)** The purified insulin can be matured *in vitro* by trypsin treatment. Trypsin cleaves the fusion protein molecule at the two AAK trypsin cleavable sites and releases the B and A chains of insulin molecule which are still attached to each other by the two disulphide bonds.

The strep-tag II-mini-insulin fusion insert was synthesized *in vitro* from four partially overlapping oligonucleotides KA001, KA002, KA003 and KA004 (see section 2.2.1 and 2.2.4) optimized for expression in tobacco. Oligonucleotides KA001and KA002 were annealed and extended to form the 5′ end of the gene fusion and oligonucleotides KA003 and KA004 were annealed and extended to form the 3′ end of the gene fusion. The two halves were then digested with *Bsu*361 restriction enzyme. The purified digested products were subsequently ligated to yield the complete coding sequence for strep-tag II-mini-insulin fusion molecule. A schematic diagram for synthesis of the strep-tag II-mini-insulin gene fusion is given in Figure 3-2.

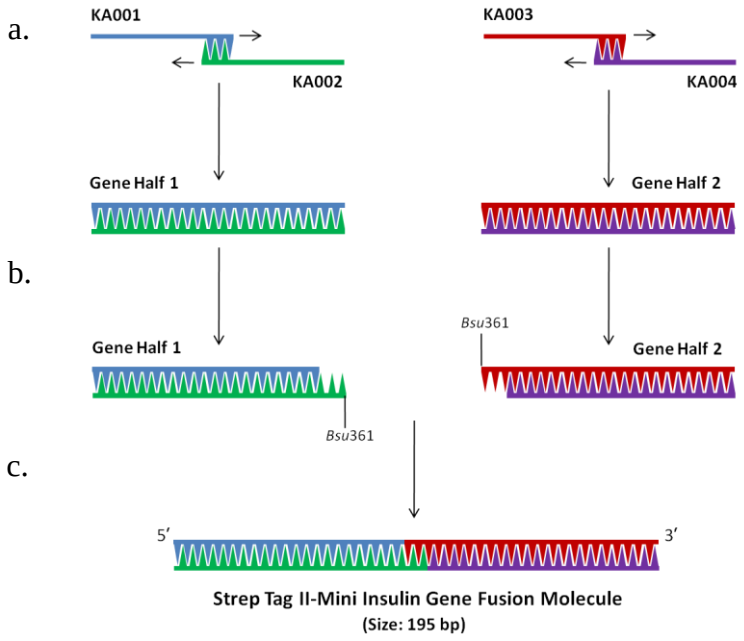


Figure 3-2 Strategy for synthesis of strep-tag II-mini-insulin gene fusion molecule. (a) The four partially overlapping oligonucleotides KA001, KA002, KA003 and KA004 were paired. Oligonucleotides KA001and KA002 were annealed and extended to form the 5′ end of the gene fusion molecule while oligonucleotides KA003 and KA004 were annealed and extended to form the 3′ prime end, **(b)** The

resulting two halves were then restricted with *Bsu*361 enzyme to generate sticky ends, **(c)** The two digested gene halves were ligated to obtain the full length strep-tag II- mini-insulin gene fusion insert.

The synthesized recombinant DNA molecule was amplified by PCR with Taq DNA polymerase using insulin forward primer KA005 (carrying a *Kpn*1 restriction site) and reverse primer KA006 (carrying a *Pst*1 restriction site). PCR reactions used 1µl and 0.5 µl of the ligated gene product as template DNA respectively. Negative controls reactions used 1 µl of gene half 1, 1 µl of gene half 2 and no DNA as template respectively. PCR products (10 µl) were analysed on 2 % agarose gel. A band of expected size (213 bp) was observed in both the PCR samples containing 0.5 µl and 1 µl of the ligated gene product as template DNA respectively while no amplification was observed in the negative controls (Figure 3-3). Taq DNA polymerase also incorporated 5′-A overhangs in the PCR product which were subsequently used in cloning of the strep-tag II-mini-insulin gene fusion into the pGEM-T easy vector.

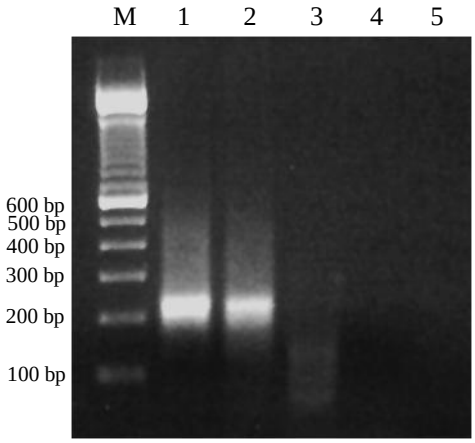


Figure 3-3 PCR amplification and confirmation of the newly synthesized strep-tag II-mini-insulin fusion gene. Amplification of correct size fragment (213 bp) can be seen in lane 1 and 2 containing 1µl and 0.5 µl of the ligated gene product as template DNA respectively. No amplification was observed in the negative controls (lane 3 -5). Lane 3 contained gene half 1 DNA as template, Lane 4 contained gene halve 2 DNA as template and lane 5 contained no DNA. Invitrogen 100 bp ladder (Lane M) was used for size comparison.

3.3 Cloning of strep-tag II-mini-insulin fusion gene

3.3.1 Cloning into pGEM-T Easy vector

The PCR amplified strep-tag II-mini-insulin fusion gene with 5′-A overhangs (section 3.2) was cloned into the complementary 3′-T overhangs of pGEM-T easy vector to generate vector pGEMKA1 (Figure 3-4).

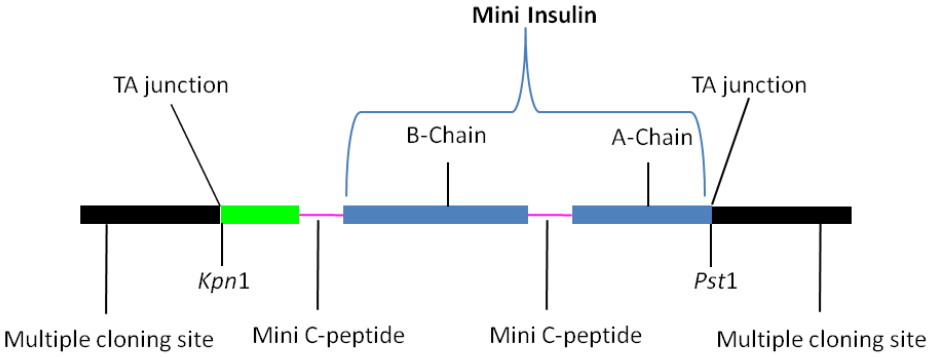


Figure 3-4 pGEMKA1. The strep-tag II-mini-insulin fusion gene with 5′-A overhangs was cloned into 3′-T overhangs site of pGEM-T easy vector to generate pGEMKA1. The vector was used for sequencing and multiplication of strep-tag II-mini-insulin fusion gene. The *Kpn*1 and *Pst*1 sites flanking the fusion gene can be used in restriction digestion for confirmation of the transgene integration.

pGEMKA1 was transformed into Stratagene XL1-Blue cells (Table 2-1) by heat shock transformation . The transformed cells were grown overnight on LB agar plates containing 100 µg/ml ampicilin. Twenty-four randomly selected resistant colonies were used for mini prep plasmid DNA extraction. Integration of the insert was confirmed by restriction digestion of the 24 plasmid DNA mini prep samples with restriction enzymes *Kpn*1 and *Pst*1 (flanking the fusion gene). The digested products were analysed on 2% gel (Figure 3-5). A restriction fragment of correct size (207 bp) was observed in most of the samples. Mini prep samples 5, 19, 20 and 24 were selected for sequencing. Sequencing was performed with the commercially available SP6 primer that is located outside of the multiple cloning site of pGEM-T easy vector. Sample 5 was found to contain the correct sequence of the integrated strep-tag II-mini-insulin fusion gene and hence was used in further cloning experiments.

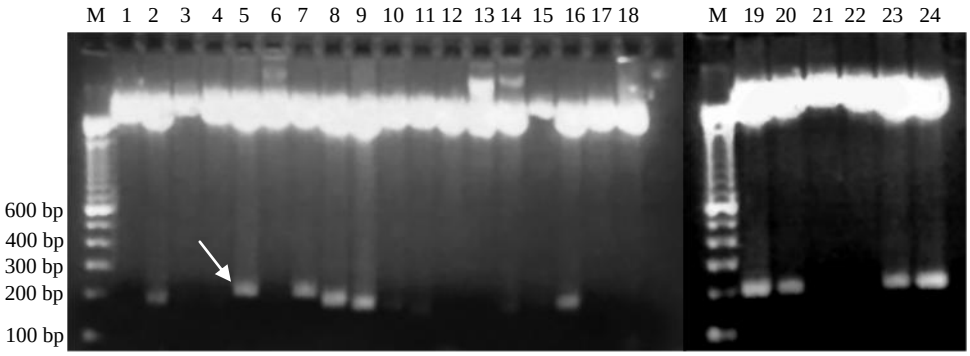


Figure 3-5 Restriction digestion of pGEMKA1 with *Kpn*1 and *Pst*1 for confirmation of strep-tag II-mini-insulin fusion gene integration. (M) Invitrogen 100 bp ladder, the size of each band is shown in base pairs; Lanes 1-24 contain pGEMKA1 mini prep samples, A DNA restriction fragment of 207 bp (shown by an arrow) was observed in most of the samples.

3.3.2 Cloning into SEAP-pPZP212-6 and TIR-OX1

SEAP-pPZP212-6, a derivative of the plant binary vector pPZP212 (Dr Kevin C. Gough, University of Leicester, personal communication), was supposed to contain the CaMV35S promoter sequence, tobacco PR1b endoplasmic reticulum targeting sequence (Cutt *et al.*, 1988) and the pea *rbc*S terminator sequence (Coruzzi *et al.*, 1984). For the attachment of these sequences to the strep-tag II-mini-insulin gene, the mini-insulin fusion gene segment was excised from pGEMKA1 (section 3.3.1) by cutting the vector with *Kpn*1 and *Pst*1 restriction enzymes. Vector SEAP-pPZP212-6 was also cut at the *Kpn*1 and *Pst*1 sites and dephosphorylated. The excised mini-insulin fusion gene segment was then cloned into the *Kpn*1 and *Pst*1 sites of the vector SEAP-pPZP212-6. The resultant vector was used to transform Invitrogen ElectroMAX DH10B Electrocompetent Cells (Table 2-1) by electroporation. Transformed cells were cultured on LB agar plates containing 100 µg/ml spectinomycin. Small-scale plasmid DNA was isolated from 20 randomly selected resistant colonies. The plasmid DNA samples were digested with *Kpn*1/*Pst*1enzymes and analysed on 2% agarose gel for confirmation of the transgene integration (Figure 3-6). A fragment of correct size (207 bp) was observed in many of the samples.

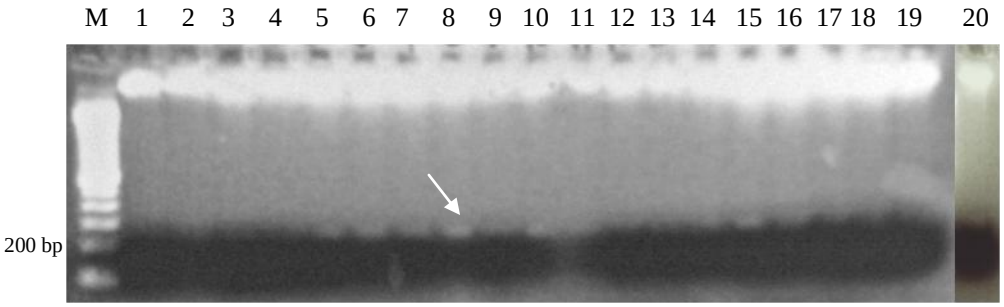


Figure 3-6 Restriction digestion analysis of strep-tag II-mini-insulin fusion gene integration in the vector SEAP-pPZP212-6. Twenty mini prep plasmid DNA samples (lane 1-20) were digested with *Kpn*1/*Pst*1enzymes**.** A fragment of 207 bp (indicated by an arrow) can be seen in many of the samples. Invitrogen 100 bp ladder **(**lane M) was used for size comparison.

Clone 6 and 8 were selected for sequencing. Sequencing was carried out using the insulin gene forward primer KA005 and reverse primer KA006 (Table 2-3). The sequencing results showed that the correct sequences of CaMV35S promoter, tobacco PR1b endoplasmic reticulum targeting signal and the pea *rbc*S terminator were not present in the vector SEAP-pPZP212-6. Three additional derivative vectors, i.e. TIR-SEKDEL, ESPA –SEKDEL and TIR- OX1 (Dr Kevin C. Gough, personal communications) were screened by sequencing for the presence of these desired sequences. Sequencing was performed with primers KA008, KA009, KA010 and KA011 (Table 2-3). Forward primer KA008 binds 42 bp upstream the 35S promoter in all the three vectors. Reverse primer KA009 binds 47 bp downstream the *rbc*S terminator in all the three vectors. Reverse primer KA010 binds near the 3′ end of TIR gene in the vector TIR-SEKDEL and TIR-OX1 to enable the complete sequencing of 35S promoter and PR1b signal sequence in these vectors. The reverse primer KA011 binds near the 3′ end of ESPA gene in vector ESPA-SEKDEL and enabled the complete sequencing of 35S promoter and PR1b signal sequence in this vector. Among the three vectors screened, TIR-OX1was found to contain the correct sequences of CaMV35S promoter, PR1b endoplasmic reticulum targeting sequence and *rbc*S terminator. To clone the mini-insulin fusion gene in the vector TIR-OX1in frame with the PR1b signal sequence, the restriction enzyme site at the 5′ end of the insulin fusion gene had to be changed from *Kpn*1 to *Spe*1 site. This was achieved by designing forward primer KA012 and re-amplifying the strep-tag II-mini-insulin gene with the forward primer KA012 and reverse primer KA006. The amplified PCR product was restricted with *Spe*1/*Pst*1 restriction enzymes. Similarly, vector TIR-OX1 was also restricted with

*Spe*1/*Pst*1 restriction enzymes and dephosphorylated. The strep-tag II-mini-insulin insert was then cloned into *Spe*1/*Pst*1 sites of TIR-OX1 in frame with the PR1b signal. The resultant vector was named pPZPKA1 (Figure 3-7). pPZPKA1 was used to transform XL1-Blue cells (Table 2-1) by heat shock transformation. Plasmid DNA was miniprepped from 10 randomly selected resistant colonies. The DNA samples were digested with *Spe*1/*Pst*1 restriction enzymes and analysed on 2% agarose gel for confirmation of the transgene integration. A fragment of correct size (207 bp) was observed in most of the samples (figure not shown). Plasmid DNA was prepared from two pPZPKA1 clones and restricted with *Eco*R1 and *Hin*d III enzymes flanking the strep-tag II-mini-insulin expression cassette (Figure 3-7) to release the expression cassette for further cloning into plant binary vector pCAMBIA2301(Cambia, Canada).

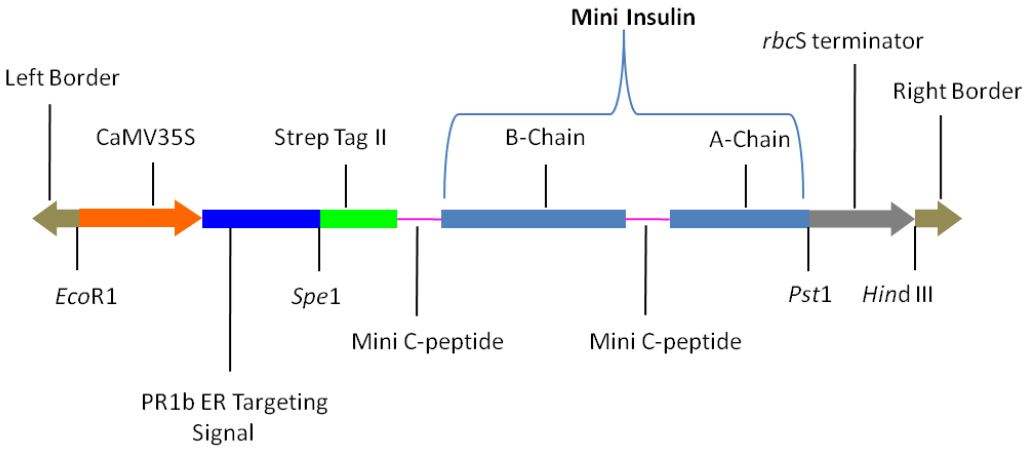


Figure 3-7 pPZPKA1. The construct was generated by cloning the strep-tag II-mini-insulin insert in the *Spe*1/*Pst*1 site of vector TIR-OX1 in frame with the PR1b signal.

3.3.3 Cloning into plant binary vector pCAMBIA2301

Plant binary vector pCAMBIA2301 (Cambia) was cut at the *Eco*R1/*Hin*d III sites located in the multiple cloning region of the plasmid. The strep-tag II-mini-insulin expression cassette (section 3.3.2) with *Eco*R1/*Hin*d III sites was cloned into the *Eco*R1/*Hin*d III sites of pCAMBIA2301 to generate the final plant expression vector pKAM1 (Figure 3-8).

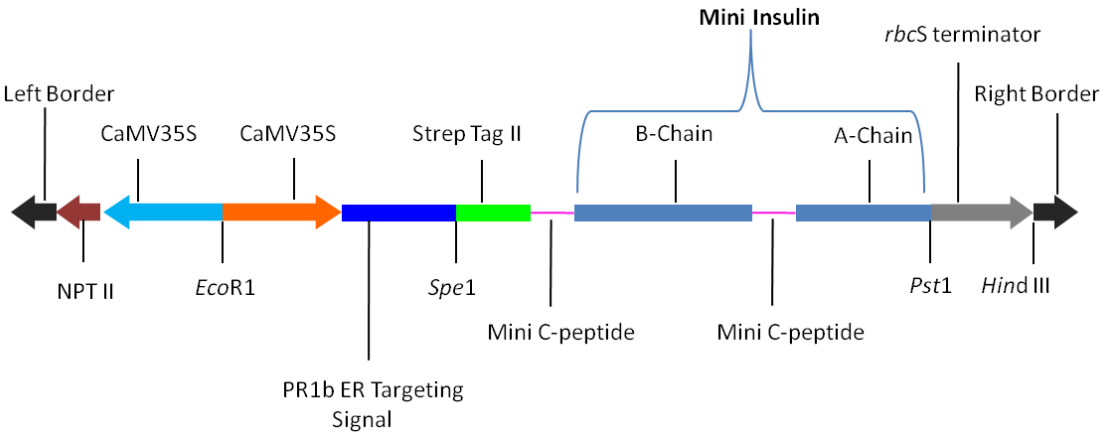


Figure 3-8 pKAM1: Plant expression vector carrying strep-tag II-mini-insulin expression cassette. The construct was generated by cloning strep-tag II-mini-insulin expression cassette into the *Eco*R1/*Hin*d III sites of pCAMBIA2301. The expression cassette is driven by CaMV35S promoter. The PR1b signal targets the fusion protein to ER. The *rbc*S terminator sequence directs RNA transcription termination.

pKAM1 was transformed into Invitrogen ElectroMAX DH10B Electrocompetent Cells (Table 2-1). Transformed cells were cultured on LB agar plates containing 25 µg/ml kanamycin. Small-scale plasmid DNA was isolated from 20 randomly selected resistant colonies. The plasmid DNA samples were digested with *Eco*R1/*Hin*d III enzymes and analysed on 1.5% agarose gel for confirmation of the expression cassette integration into the vector (Figure 3-9). A DNA fragment (1803 bp) consisting of 35S promoter, PR1b signal, strep-tag II-mini-insulin fusion gene and *rbc*S terminator sequence was detected in most of the samples. Clone 3 and 18 were selected for sequencing of the complete expression cassette. Sequencing primers used included KA006, KA012, KA008, KA009 (Table 2-3). The sequencing primers enabled the complete sequencing of the insulin expression cassette. A midi scale pKAM1 DNA was isolated for subsequent transformation of *Agrobacterium* strain GV3101.

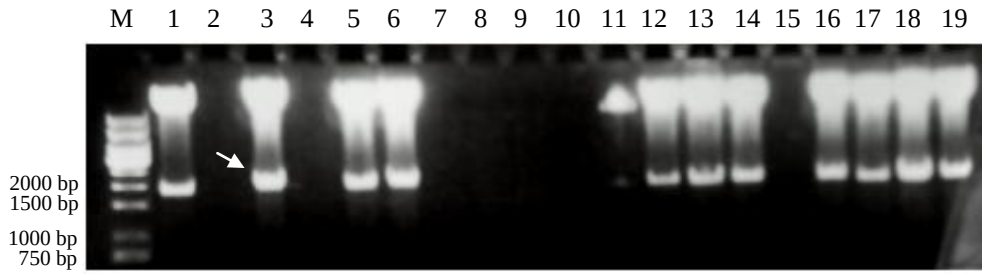


Figure 3-9 Restriction digestion of pKAM1 with *Eco*R1/*Hin*d III to confirm integration of insulin expression cassette. Twenty mini prep plasmid DNA samples (lane 1-19, lane 20 not shown) were

digested with *EcoR1/hind* III restriction enzymes. A DNA fragment of 1803 bp (indicated by an arrow) corresponding to the complete expression cassette was detected in most of the samples. 1kb ladder (Fermentas GeneRuler) (lane M) was used for size comparison.

3.4 Tobacco transformation with pKAM1

pKAM1 was introduced into *Agrobacterium tumefaciens* strain GV3101 by heat shock transformation (section 2.1.11). The resulting *Agrobacterium* strain was used to transform tobacco (*Nicotiana tabacum cv* SAMSUN) leaf disks (section 2.3). Regeneration of shoots (on selective MS agar medium containing 3% sucrose, 250 µg/ml cefotaxime and 100 µg/ml kanamycin) from the wounded areas of infected leaf disks was evident 1-2 weeks post infection (Figure 3-10 A). Non-infected wild type leaf disks developed no callus or shoots and dried out (Figure 3-10 B). After 6-8 weeks, regenerated shoots were excised from leaf discs and transferred to MS agar rooting medium containing 3% (w/v) sucrose, 250 µg/ml cefotaxime and 100 µg/ml kanamycin with no additional hormones. Root development started after 7-12 days (Figure 3-10 C). Some escape shoots were also observed that failed to develop any roots on selective rooting medium (Figure 3-10 D). After roots had developed, 25 healthy transgenic plantlets (designated as INS1 to INS25) were transferred to pots containing soil and were allowed to grow in green house conditions (Figure 3-10 E). The transgenic line INS1 died in the green house during early growth stage. The remaining 24 plants were allowed to grow until maturity to obtain T1 seeds. To avoid cross-fertilization, flower buds were covered with paper bags. Transgenic lines INS7, INS21 and INS23 did not produce any seed. After maturation, seeds from the remaining transgenic lines were harvested and stored in bijou vials.

Figure 3-10 Development of transgenic tobacco plants expressing human mini-insulin. A. Four week old regenerated shoots from transformed leaf disks on kanamycin selective shoot medium; **B.** Non-transformed (negative control) leaf disks failed to regenerate any shoots/callus on selective shoot medium and dried; **C.** Regenerated transformed shoots developed roots on selective medium containing 100 µg/ml kanamycin; **D.** Escapes (non-transformed regenerated shoots) failed to develop any roots and became chlorotic on kanamycin selective rooting medium; **E.** Transgenic tobacco lines growing in the green house.

3.5 Analysis of transgenic plants

3.5.1 PCR analysis of the mini-insulin gene integration in the plant genome

The integration of mini-insulin fusion gene in the plant genome was demonstrated by PCR. DNA was isolated from 20 transgenic lines and 1 wild type plant, while pKAM1 DNA (section 3.3.3) was used as positive control. PCR was carried out with insulin forward primer KA012 and reverse primer KA006 (Table 2-3). PCR products were analysed on 2.5% agarose gel. The results are shown in Figure 3-11.

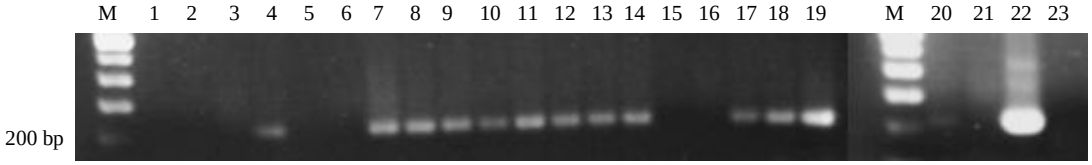


Figure 3-11 PCR analysis of the mini-insulin gene integration in the plant genome: Lane M, Invitrogen 100 bp ladder; lane 1, INS2; lane 2, INS3; lane 3, INS4; lane 4, INS5; lane 5, INS6; lane 6, INS7; lane 7, INS8; lane 8, INS9; lane 9, INS10; lane10, INS11; lane 11, INS12; lane 12, INS13; lane 13, INS14; lane 14, INS15; lane 15, INS16; lane 16, INS17; lane 17, INS18; lane 18, INS19; lane 19, INS20; lane20, INS21; Lane 21, Wild type plant DNA as negative control; lane 22, pKAM1 DNA as positive control; lane 23, no template DNA as negative control. PCR was carried out with insulin forward primer KA012 and reverse primer KA006. A band of correct size of 207 bp (shown by arrow) corresponding to the mini-insulin gene can be seen in most of the transgenic tobacco lines (lane 1-20) and positive control pKAM1 (lane 22). No amplification was observed in the wild type plant (lane21) and negative control containing no template DNA (lane 23).

An amplification of correct size (207 bp) corresponding to the mini-insulin gene was observed in most of the transgenic lines (lane 1-21) and the positive control pKAM1 (lane 22). However, no amplification was observed in the wild type plant sample (lane 21) and the sample containing no template DNA (lane 23). No amplification was observed in tobacco lines INS2, INS3, INS4, INS6, INS7, INS16 and INS17 (lane 1, lane 2, lane 3, lane 5, lane 6, lane 15 and lane 16). Among these PCR negative lines, T1 progeny of INS3 line was able to survive on kanamycin selection medium and segregated in the expected Mendelian segregation ratio of 3:1 for a single copy insertion (section 3.5.2). Furthermore, the RT-PCR and protein western blot analysis confirmed that INS3 is indeed a transgenic line. The nature of INS7 line could not be determined as the plant was not able to produce any seeds for further testing. INS7 could be an escape plant or the PCR might not have worked for this sample as the quality of the plant DNA used for PCR was not very good. T1 seedlings from PCR negative tobacco

lines INS2, INS4, INS6, INS16 and INS17 failed to survive on kanamycin MS agar medium during T1 seed screening indicating that these five lines were possibly not transformants but escapes that managed to survive the initial selection conditions during shoot regeneration and root development.

3.5.2 Seed screening

T1 seeds from 21 transgenic tobacco lines were screened on selective MS agar medium (containing 3% sucrose and 150 µg/ml kanamycin) for segregation analysis. Susceptible seedlings became chlorotic and died a few days after germination while resistant lines remained healthy and green. All germinated seedlings from tobacco lines INS2, INS4, INS6, INS16, INS17, INS19 and INS22 died on selective medium. Among these susceptible lines INS2, INS4, INS6, INS16 and INS17 were also PCR negative (section 3.5.1) suggesting that these lines could be escapes that somehow survived the harsh selection conditions during shoot regeneration and root development in the early stages of their life cycle. On the other hand, susceptible tobacco lines INS19 and INS22 were PCR positive; hence, the transgenes in these plants might have been silenced not allowing T1 seedlings to survive on selective medium. After 3-4 weeks of germination, T1 progenies from the remaining 14 resistant lines were scored as resistant and susceptible for segregation analysis. Segregation ratios were calculated and compared with the expected Mendelian segregation ratios of 3:1(resistant: susceptible) for a single copy insertion or 15:1(resistant: susceptible) for a double copy insertion using the chi-square test (Table 3-1). Based on the observed segregation ratio, 9 transgenic lines INS3, INS5, INS8, INS9, INS11, INS13, INS14, INS15 and INS18 were tested against an expected segregation ratio of 3:1 (resistant: susceptible) while 5 transgenic lines i.e. INS10, INS12, INS20, INS24 and INS25 were tested against an expected segregation ratio of 15:1 (resistant: susceptible). The null-hypothesis that the observed segregation ratios do not differ significantly from Mendelian segregation ratio of either a 3:1 or 15:1 was not rejected at 10% level ($p>0.100$) (Pearson and Hartley, 1966) for each of the lines tested. Among the transgenic lines tested for a 3:1 segregation ratio, all lines showed the standard Mendelian segregation ratio except INS11. INS11 showed an observed segregation ratio of 1.8:1 (resistant: susceptible), which is closer to a 2:1 ratio of a single copy insertion resulting in embryo lethal phenotype. In such a situation the

homozygous T1 seeds carrying two copies of the transgene will not be able to germinate while the remaining heterozygous seeds and seeds with no copy of the transgene will show a 2:1 ratio (resistant: susceptible). The hypothesis of an embryo lethal phenotype is further supported by the fact that 18% of T1 seed from INS11 did not germinate. Among the transgenic lines tested for a 15:1 segregation ratio, all the lines followed standard Mendelian segregation ratio except INS12 that showed a 100% resistance for all the T1 seeds tested. Lack of any segregation in T1 seed of INS12 could be due to multiple copy insertions of the transgene in the plant genome. To obtain homozygous T2 transgenic lines, 8-10 resistant seedlings were selected from each of the 14 transgenic lines and grown in the green house until maturity. All the plants were self-pollinated as before. T2 seed was grown on selective MS agar medium (containing 3% sucrose and 150 µg/ml kanamycin) for segregation analysis. Homozygous lines were identified as those T2 lines that did not segregate.

Table 3-1 Segregation analysis of T1 transgenic tobacco lines engineered for expressing human mini-insulin

Line	Number of seedlings screened	Resistant seedlings	Susceptible Seedlings	Observed segregation ratio	Expected segregation ratio	Chi square Value	P value
INS3	96	70	26	2.7:1	3:1	0.21	0.64
INS5	95	71	24	3:1	3:1	0	1
INS8	94	73	21	3.5:1	3:1	0.48	0.49
INS9	97	74	23	3.2:1	3:1	0.05	0.82
INS10	96	92	4	23:1	15:1	0.71	0.4
INS11	85	55	30	1.8:1	3:1	5.33	0.02
INS12	96	96	0	96:0	15:1	6.38	0.01
INS13	83	59	24	2.5:1	3:1	0.85	0.36
INS14	101	76	25	3:1	3:1	0	1
INS15	71	58	13	4.5:1	3:1	2.61	0.11
INS18	95	73	22	3.3:1	3:1	0.21	0.64
INS20	94	91	3	30.3:1	15:1	1.6	0.21
INS24	92	84	8	10.5:1	15:1	1.6	0.21
INS25	100	95	5	19:1	15:1	0.18	0.67

3.5.3 RT- PCR to analyse expression of mini-insulin by transgenic tobacco

Total RNA from three T1 transgenic tobacco lines (INS3, INS5 and INS25) and one wild type plant was subjected to RT-PCR. Insulin reverse primer KA006 was used for synthesis of first strand cDNA using reverse transcriptase enzyme. The resultant cDNA was then amplified by PCR using insulin forward primer KA012 and reverse primer KA006. To eliminate chances of DNA contamination, pure RNA extracts (without RT reaction) from each transgenic line and wild type plant were also included in the PCR. The results of RT-PCR are shown in Figure 3-12. A fragment of correct size (207 bp) corresponding to the mini-insulin was amplified in all the three transgenic lines samples (lane3, lane5 and lane7) that were treated with reverse transcriptase enzyme. No amplification was observed in the reverse transcriptase treated wild type plant (lane1). Similarly, no amplification was observed in any of the samples that were not treated with reverse transcriptase prior to PCR (lane2, lane4, lane6 and lane8) thus eliminating the possibility that the observed fragment was amplified from residual DNA in the RNA samples. This confirms that the mini human insulin gene is being expressed by the transgenic tobacco lines.

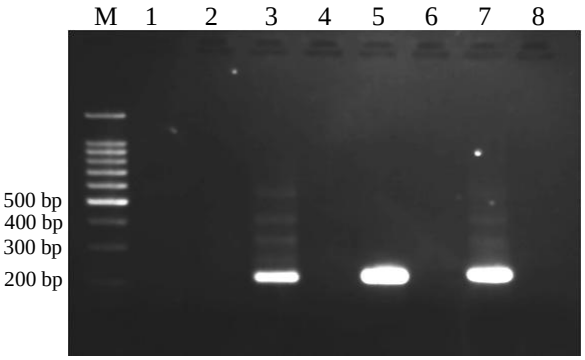


Figure 3-12 RT-PCR to detect human min insulin expression by transgenic tobacco plants: First strand cDNA was made from RNA samples of three transgenic lines INS3 (lane 3), INS5 (lane 5) and INS25 (lane 7) and one wild type plant (lane 1). Insulin reverse primer KA006 was used to make insulin cDNA in a reverse transcription reaction. The resultant cDNA was amplified by PCR using insulin forward primer KA012 and reverse primer KA006. To eliminate the chances of genomic DNA contamination, RNA extracts (without any reverse transcription reaction) from each transgenic line (INS3; lane 4, INS5; lane 6 and INS25; lane 8) and wild type plant (lane 2) were also included in PCR. An amplified fragment of correct size (207 bp) corresponding to the mini-insulin can be seen in all the three transgenic lines samples (lane3, lane5 and lane7) that were treated with reverse transcriptase. This band is not present in the wild type plant sample that was treated with reverse transcriptase (lane1). The fragment was also not observed in any of the samples that were not treated with reverse transcriptase prior

to PCR (INS3; lane 4, INS5; lane 6, INS25; lane 8 and wild type plant; lane 2) thus eliminating the possibility that the observed fragment was amplified from residual DNA in the RNA samples. This confirms the RNA level expression of mini human insulin by the transgenic tobacco lines. Promega 100 bp DNA ladder **(**lane M) was used for size comparison.

3.5.4 Western blot and coomassie staining to detect expression of strep-tag II-mini-insulin fusion protein by transgenic tobacco

Plant proteins (total crude protein and apoplastic fluid protein) were extracted and separated by Tris-Tricine SDS-PAGE. The protein gel was either subjected to coomassie staining for protein visualization, or the proteins were transferred to a PVDF membrane for western blot detection. Coomassie staining did not show any unique bands related to strep-tag II-mini-insulin (data not shown). Western blot detection was carried out with two different antibodies; one mouse monoclonal insulin E2E3 antibody (Abcam, Table 2-8) targeted against the mini-insulin and another mouse strep-tag II monoclonal antibody (Novagen, Table 2-8) targeted against the strep-tag II attached to the mini-insulin. In both cases, goat anti-mouse horseradish peroxidise (HRP) conjugate secondary antibody (DakoCytomation) was used at 1/2000 dilution to detect the bound primary antibody with EZ-ECL based detection system. Protein extraction and western blotting protocols were optimized for each of the two primary antibodies used.

3.5.4.1 Western blot detection of mini-insulin in total plant protein extract

Figure 3-13 shows result of western blot analysis of total plant protein extracts using mouse monoclonal insulin E2E3 antibody (Table 2-8) targeted against mini-insulin. Crude protein extracts from transgenic tobacco lines (INS3; lane 3, INS10; lane 5, INS12; lane 6 and INS25; lane 7) and wild type plant (lane 1) along with 200 ng bovine insulin (Sigma) as positive control were analysed. A positive signal of approximate size (~ 6 kDa) corresponding to mini-insulin was detected in transgenic tobacco lines (INS3; lane 4, INS12; lane 6 and INS25; lane 7) and positive control (lane 3). Strongest signal was detected in INS3 (lane 4). No signal was detected in wild type plant. Signal was also not detected in transgenic line INS10 (lane5). Similar results were obtained with the remaining transgenic lines (data not shown). Variation in expression could be due to clonal variation among the transformants.

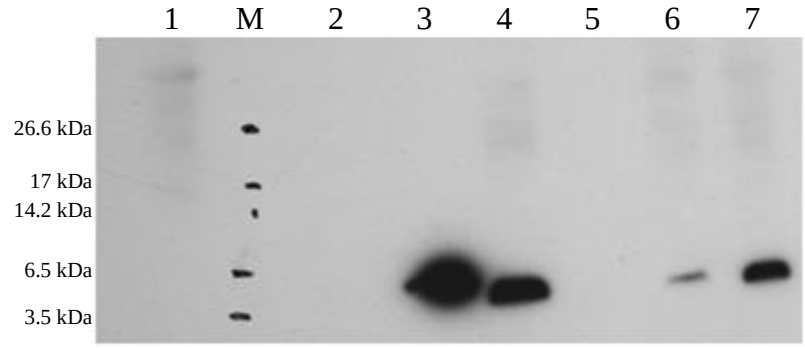


Figure 3-13 Western blot of crude plant protein extracts to detect expression of strep-tag II-mini-insulin in transgenic tobacco with mouse monoclonal insulin antibody (E2E3): Total plant proteins were separated through Tris-Tricine SDS-PAGE and subsequently transferred to a PVDF membrane. The strep-tag II-mini-insulin fusion protein was detected with 1/200 dilution of mouse monoclonal insulin antibody (E2E3, Abcam). A band of expected size (~ 6 kDa) was observed in the transgenic tobacco lines INS 3 (lane 4, the strongest signal), INS 12 (lane 6) and INS 25 (lane 7). Bovine insulin (lane 3) was used as positive control. No signal was detected in the wild type plant (lane 1). Signal was also not detected in the transgenic line INS 10 (lane 5). No sample was loaded in lane 2. Sigma ultra low range protein marker (lane M) was used for size comparison.

Western blot detection of strep-tag II- mini-insulin in crude protein extracts was also carried out with mouse strep-tag II monoclonal antibody (Table 2-8) targeted against the strep-tag II attached to the mini-insulin. The results are shown in Figure 3-14. A positive band of approximate size (~ 6 kDa) corresponding to strep-tag II-insulin fusion protein was detected in transgenic tobacco lines INS3 (lane 4), INS12 (lane 6) and INS25 (lane 7). The signal was stronger in INS3 as seen in western blot detection with mouse monoclonal insulin antibody (E2E3). No signal was detected in wild type plant. Signal was also not detected in transgenic line INS10 (lane5). A high molecular weight non-specific band (shown by asterisks) was observed in the transgenic lines as well as in the wild type plant.

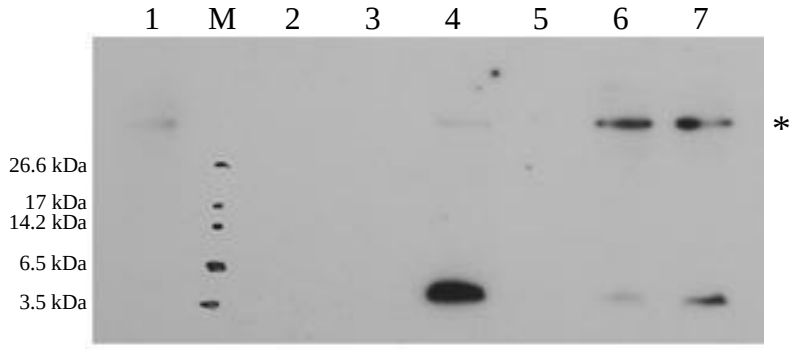


Figure 3-14 Western blot of crude plant protein extract to detect expression of strep-tag II-mini-insulin in transgenic tobacco with mouse strep-tag II monoclonal antibody: Total plant proteins were separated through Tris-Tricine SDS-PAGE and subsequently transferred to a PVDF membrane. The strep-tag II-mini-insulin fusion protein was detected with 1/1000 dilution of mouse strep-tag II monoclonal antibody (Novagen). A band of expected size (~ 6 kDa) was observed in the transgenic tobacco lines INS 3 (lane 4, the strongest signal), INS 12 (lane 6) and INS 25 (lane 7) but not in the wild type plant (lane 1). No signal was detected in sigma bovine insulin (lane 3) as unlike the recombinant mini-insulin it is not tagged with strep-tag II. Signal was also not detected in the transgenic line INS 10 (lane 5). Lane 2 contained no sample. A high molecular non-specific band (shown by asterisks) was observed in the transgenic lines as well as in the wild type plant. Sigma ultra low range protein marker (lane M) was used for size comparison.

3.5.4.2 Western blot detection of insulin in apoplastic fluid

Apoplast proteins from transgenic plants and wild type plants were extracted and analysed by western blotting. However, the strep-tag II-mini-insulin protein could not be detected in the apoplastic fluid with any of the two primary antibodies used. Apoplastic fluid was precipitated with methanol with the view that concentrating the extract might help in detecting the mini-insulin; however, no signal was detected in the concentrated extract. Presence of the fusion protein in the crude extract and absence in the apoplastic fluid suggest that some sequences in the fusion protein might be responsible for the ER retention of the fusion protein thus not allowing the secretion of fusion protein into the apoplast. Lack of secretion of a recombinant protein (single-chain antibody fragments specific for beet necrotic yellow vein virus coat protein) into the culture fluids of *Nicotiana benthamiana* carrying a *Phaseolus vulgaris* phytohemagglutinin (PHA) secretory signal has been reported (Fecker *et al.*, 1996).

3.5.4.3 Western blot detection of Strep-Tactin purified mini-insulin fusion protein

The strep-tag II attached to the mini-insulin has affinity to Strep-Tactin matrix and hence the mini-insulin can be purified from crude protein extract using Strep-Tactin columns. The bound protein attached to the Strep-tag II can be eluted from the StrepTactin matrix by washing with biotin or desthiobiotin (a competitor of strep-tag II) containing buffers (Skerra and Schmidt, 2000). Purification of the strep-tag II-mini-insulin fusion protein from three transgenic tobacco lines (INS3, INS5 and INS25) was carried out by applying crude plant extract to a 50% suspension of Strep-Tactin Sepharose (section 2.5.5). A wild type plant was included as a negative control. The purified resin-mini-insulin complex was boiled in 2X protein loading buffer for 5 min before running on 16% Tris-Tricine gel and analysing by western blotting. Bovine insulin (400 ng) was used as positive control. Figure 3-15 shows result of western blot detection of the purified insulin with mouse monoclonal E2E3 insulin antibody (Table 2-8). A band of approximate size (~ 6 kDa) related to the mini-insulin was observed in the three transgenic lines INS3, INS5 and INS25 but not in the wild type plant. The intensity of the band was much higher in the transgenic line INS3 as observed in western blot detection of insulin in crude plant extract. However, the purification did not result in high enrichment of the strep-tag II-mini-insulin. A strong non-specific band given by the StrepTactin beads at about 10 kDa was also observed (not shown). In another experiment, attempts were made to elute the mini-insulin fusion protein from the washed StrepTactin matrix. The elution buffer contained 100 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA and 2.5 mM desthiobiotin. The eluted sample was boiled in protein loading buffer and run on 16% Tris-Tricine gel. Western blot detection was conducted as before. However, no signal was detected in any of the samples. It might be possible that the concentration of desthiobiotin (which competes with strep-tag II for binding with StrepTactin matrix) in elution buffer was not enough to allow the release of strep-tagged mini-insulin. Witte *et al.* (2004) purified a strep II tagged membrane-anchored protein kinase (NtCDPK2) from *Nicotiana benthamiana* leaf extract using a 10 mM concentration of desthiobiotin in the elution buffer. Aggregation of insulin in dimers and hexamers might be another factor that hinders the purification of insulin. Another factor that could be responsible for no recovery of this protein might be the partial degradation of the attached strep-tag II. (Van Esse *et al.*, 2006) failed to purify

the C and N terminally Strep II-tagged *Cladosporium fulvum* effector proteins targeted to tomato leaf apoplast (also to tobacco leaf apoplast). Based on their findings they suggested that *in planta* instability (apoplast contains many proteases) of the expressed fusion protein causes the removal of affinity-tags independent of the nature of the tag and its position in the fusion protein; resulting in no purification of the fusion protein.

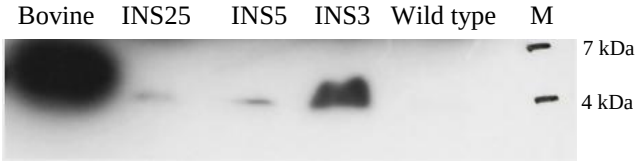


Figure 3-15 Western blot detection of Strep-Tactin purified mini-insulin fusion protein with mouse monoclonal insulin antibody (E2E3): The strep-tag II-mini-insulin fusion protein was purified from crude plant extract by affinity purification using Strep-Tactin Sepharose resin. The strep-tag II attached to the mini-insulin has affinity to the Strep-Tactin matrix. The bound protein with the matrix was boiled in 2x protein loading buffer for 5-10 min before being used in SDS-PAGE. Detection of the fusion protein was done with mouse monoclonal insulin antibody (E2E3, Abcam). A band of expected size (~ 6 kDa) related to the mini-insulin was detected in the three transgenic lines INS3, INS5 and INS25 but not in the wild type plant. The intensity of the band was much higher in the transgenic line INS3 than the other two lines. Bovine insulin (Sigma) was used as positive control. Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for size comparison.

3.5.5 Insulin ELISA

3.5.5.1 Plate trapped antigen (PTA) ELISA for insulin

A plate trapped antigen ELISA assay was developed for detection of insulin in transgenic tobacco protein extract (section 2.6.1). Bovine insulin (Sigma) was used as a standard for determination of optimum conditions to generate a reliable standard curve. Initially, the ELISA was performed with PBS as coating buffer for the native bovine insulin serial dilutions. However, no measurable signal was obtained. Therefore, the assay was optimized by adopting several strategies. One strategy involved using samples denatured in various buffers before coating instead of the native bovine insulin serial dilutions. To achieve this, bovine insulin serial dilutions were first denatured by; (1) boiling for 5 min in PBS, (2) boiling for 5 min in PBS containing 2% SDS and (3) incubating in 8 M urea at 37 °C for 30 min before coating the ELISA plate wells. The ELISA was conducted as before. Only samples denatured in PBS containing 2% SDS produced measurable signals and generated a linear standard curve. In the next

optimization step, bovine insulin serial dilutions were mixed with three different volumes (30 µl, 20 µl and 5 µl) of wild type plant protein extract to analyse the effect of plant background on the detection of insulin. It was found that amount of plant extract used had significant inhibitory effect on signal development with 30 µl of plant extract having the worst effect. Not only the signal was very weak but it was variable as well among the different serial dilutions irrespective of the amount of insulin in each dilution. On the other hand, 5 µl of plant extract mixed with each serial dilution had the least matrix effect. The signal was still weak but it resulted in the generation of a standard curve with an R^2 value (Pearson co-efficient of determination) of 0.986 (mean of three experiments; an example of standard curve can be seen in Figure 3-16). The optimised ELISA assay was used with total plant protein from three transgenic lines (INS3, INS5 and INS25) and three wild type plants, along with bovine insulin serial dilutions (denatured in PBS containing 2% SDS). Plant protein samples (15 µg) and bovine serial dilutions (mixed with 5 µl plant extract to account for any differences caused by plant background) were then analysed by ELISA using the same coating buffer (PBS containing 2% SDS). Absorbance readings were taken at 405 nm and are shown in Table 3-2. A standard curve was generated using OD_{405} readings of the bovine insulin serial dilutions after subtracting zero insulin background values (Figure 3-16). The curve was used to determine amount of insulin in transgenic tobacco lines. Figure 3-17 shows detectable amount of insulin in the three transgenic tobacco lines tested. Highest amount of insulin was detected in transgenic line INS3 (14.32 µg/ 100 µg total plant protein) followed by INS5 (8.06 µg/ 100 µg total plant protein) and INS25 (4.52 µg/ 100 µg total plant protein) respectively. The results further support the western blotting results (section 3.5.4.1 and 3.5.4.3) in which strongest signals were detected in transgenic line INS3 protein samples.

Table 3-2 Mean absorbance values of bovine insulin serial dilutions, three transgenic tobacco lines (INS3, INS5 AND INS25) and three wild type plants (Wild1, Wild2 and Wild3) obtained from insulin plate trapped antigen (PTA) ELISA

Sample	Mean OD_{405}
0µg Bovine insulin	0.0484
1µg Bovine insulin	0.0628
2.5µg Bovine insulin	0.0810
5µg Bovine insulin	0.1041
INS3	0.0743
INS5	0.0640
INS25	0.0582
Wild1	0.0441
Wild2	0.0480
Wild3	0.0410

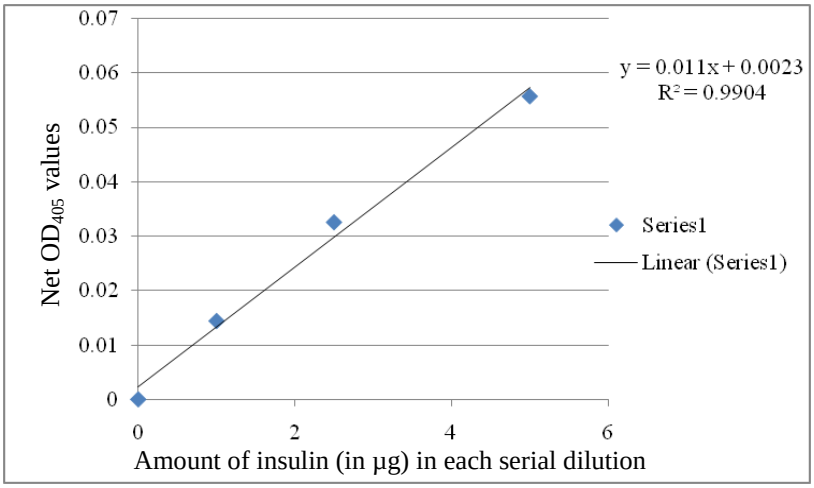


Figure 3-16 Bovine insulin standard ELISA curve: Serial dilutions (0 µg, 1 µg, 2.5µg and 5 µg respectively) of bovine insulin in wild type plant extract were boiled in PBS containing 2% SDS for 5 min. The boiled samples were used to coat ELISA plate wells overnight. A 1/1000 dilution of mouse monoclonal (E2E3) insulin antibody (Abcam) was used as the primary antibody and 1/2000 dilution of AP-Goat anti mouse antibody (Zymed) was used as the secondary antibody. Detection was done with SIGMA*FAST* p-Nitrophenyl phosphate substrate. Absorbance readings were recorded at 405 nm. The standard curve was generated by subtracting zero insulin background readings from each serial dilution reading and plotting the net absorbance values against amount of insulin in each serial dilution.

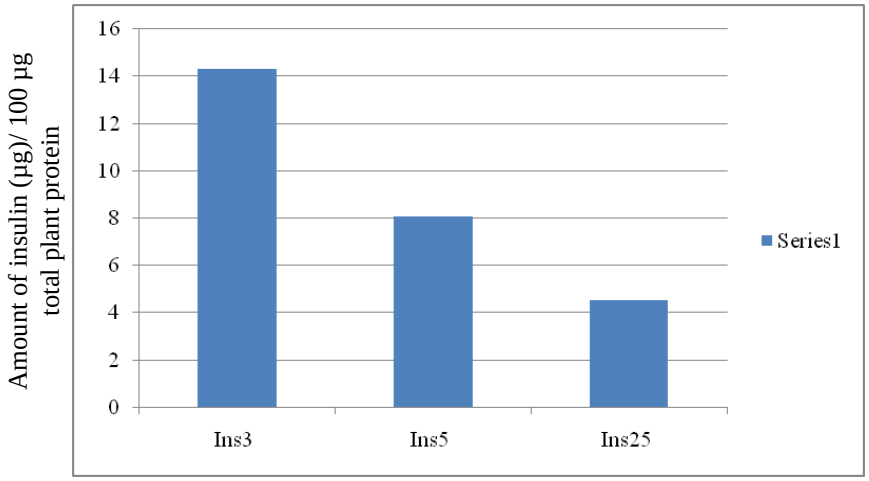


Figure 3-17 Detectable amounts of insulin in three transgenic tobacco lines (INS3, INS5 and INS25) protein extracts obtained by plate trapped antigen (PTA) ELISA: Highest amount of insulin was detected in transgenic line INS3 (14.32 µg/ 100 µg total plant protein) followed by INS5 (8.06 µg/ 100 µg total plant protein) . Detectable amount of insulin in transgenic line INS25 was 4.52 µg/ 100 µg total plant protein. The results support the western blotting results in which stronger signal was detected in transgenic line INS3.

3.5.5.2 Mercodia Ultrasensitive Insulin ELISA

Mercodia Ultrasensitive Insulin ELISA was carried to further analyse the presence of insulin in transgenic plant protein extracts. Human recombinant insulin calibrators (Table 3-3), provided with the kit, were used to generate a standard curve. To determine the matrix effect of plant background on insulin detection, different amounts of wild type plant protein extract (5 µl, 10 µl, 20 µl and 50 µl) were mixed with calibrator 4 and analysed. Mean OD_{450} values obtained are given in Table 3-3. The effect of plant extract matrix on the detection of insulin in calibrator 4 is given in Figure 3-18. It can be seen that the plant extract has a significant inhibitory effect on signal development. The highest inhibitory effect was shown by sample containing 50 µl of plant extract with a 56.14 % signal reduction over samples with no plant extract. Sample with least amount of plant extract (5 µl) had the least inhibitory effect. However, variation in signal reduction among Calibrator 4 samples containing various amounts of plant extract was less than variation between Calibrator 4 alone and Calibrator 4 containing any amount of plant extract. For example, signal was reduced by 38.86 % in sample containing 5 µl plant extract and there was a further reduction of 17.28 % in sample containing 50 µl plant extract. Based on these results, 15 µg of total plant protein (contained in ~ 4-5 µl of protein extract) was used for all subsequent experiments. The

Mecrodia ELISA assay is designed for detection of mature insulin in human serum or plasma and has < 0.01 % reactivity to proinsulin (proinsulin consists of the B-chain and A-chain of insulin joined by the central C-peptide). Therefore, plant protein extracts from transgenic plant INS3 and wild type plant were treated with trypsin (section 2.5.6) to release mature insulin by cleaving the two min-C peptides (one between the DesB30-chain and A-chain of mini-insulin and another between the strep-tag II and mini-insulin) with a view to minimize the effect of the attached strep-tag II and mini-C peptides on the detectibility of insulin. The trypsin-treated and non-treated plant extracts from transgenic line INS3 and wild type plant were analysed by the ELISA. Absorbance readings recorded at 450 nm are given in Table 3-4. As can be seen from the table, trypsin-treated transgenic line INS3 sample showed significantly high absorbance value than non-treated INS3 sample, however, no significant difference in absorbance was observed in trypsin-treated and non-treated wild type plant samples. A standard curve was generated for Mercodia insulin standards (Figure 3-19) to estimate detectable amounts of insulin in the transgenic line INS3 samples. Figure 3-20 shows amounts of insulin detected in the trypsin-treated and non-treated transgenic line INS3. The amount of insulin detected in trypsin treated extract was 4.0 pg/ 100 µg total plant protein while in the non-treated extract the amount of insulin detected was 0.3 pg/ 100 µg total plant protein. The amount of insulin detected in INS3 samples by Mercodia ELISA is much less than the amount of insulin detected by PTA ELISA (section 3.5.5.1). This could be due to different reactivities of the antibodies used in both the detection systems to the mini-insulin. Also, the reactivity of Mercodia ELISA to commercially available insulin detemir (DesB30 insulin is a precursor of insulin detemir) is <0.0000007 % (Mercodia); another factor contributing to such low detectable levels of insulin in the transgenic plant samples.

Table 3-3 Mercodia Ultrasensitive Insulin ELISA: Mean OD_{450} values for Mercodia insulin Calibrators (0, 1, 2, 3, 4, 5) and calibrator 4 mixed with different amounts (5 µl , 10µl, 20µl and 50µl) of wild type plant extract

Sample	Amount of insulin in each calibrator (pg)	OD_{450}
Calibrator 0	0	0.0389
Calibrator 1	0.13	0.0459
Calibrator 2	0.8625	0.1284
Calibrator 3	2.5875	0.3393
Calibrator 4	8.62	1.0278
Calibrator 5	17.24	1.8042
Calibrator 4+5µl plant extract	8.62	0.6284
Calibrator 4+10µl plant extract	8.62	0.5304
Calibrator 4+20µl plant extract	8.62	0.4992
Calibrator 4+50µl plant extract	8.62	0.4504

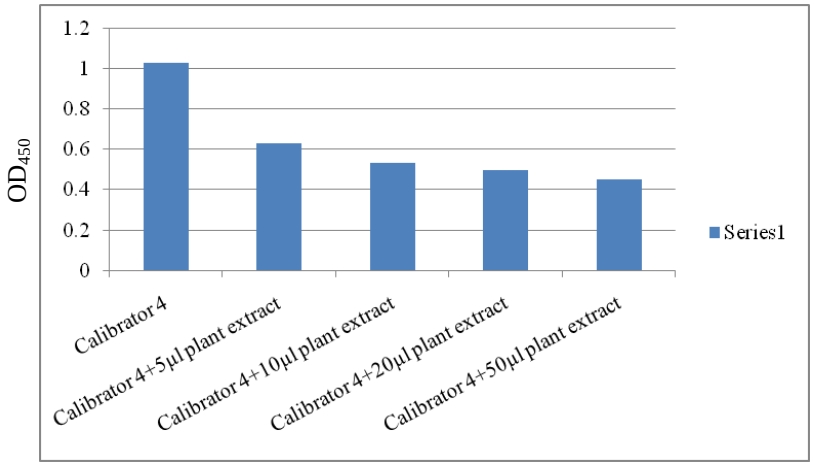


Figure 3-18 Mercodia insulin ELISA for analysing the effect of plant extract on insulin detection: To analyse the detection of insulin in plant background, different amounts (5 µl**,** 10 µl, 20 µl and 50 µl) of wild type plant extract were mixed with Mercodia Calibrator 4 (containing 8.62 pg insulin) and analysed by Mercodia ultrasensitive insulin ELISA. Mean absorbance readings were recorded for each sample (Table 3-3) and plotted. The plant extract was found to have severe inhibitory effect on signal development. Signal was reduced by 56.14% in sample containing 50 µl of plant extract. Variation in signal reduction was less pronounced among the samples with added plant extracts compared to the variation between Calibrator 4 alone and samples containing Calibrator 4 with plant extract added.

Table 3-4 Mercodia Ultrasensitive Insulin ELISA: Mean OD_{450} values for Mercodia insulin Calibrators (0, 1, 2, 3, 4 and 5) and trypsin-treated and non-treated plant extracts

Sample	Mean OD450
Calibrator 0	0.02935
Calibrator 1	0.03785
Calibrator 2	0.11905
Calibrator 3	0.26195
Calibrator 4	0.941
Calibrator 5	1.76085
Negative1 trypsin treated	0.0271
Negative1 NON treated	0.02895
INS3 trypsin treated	0.08855
INS3 NON treated	0.03185

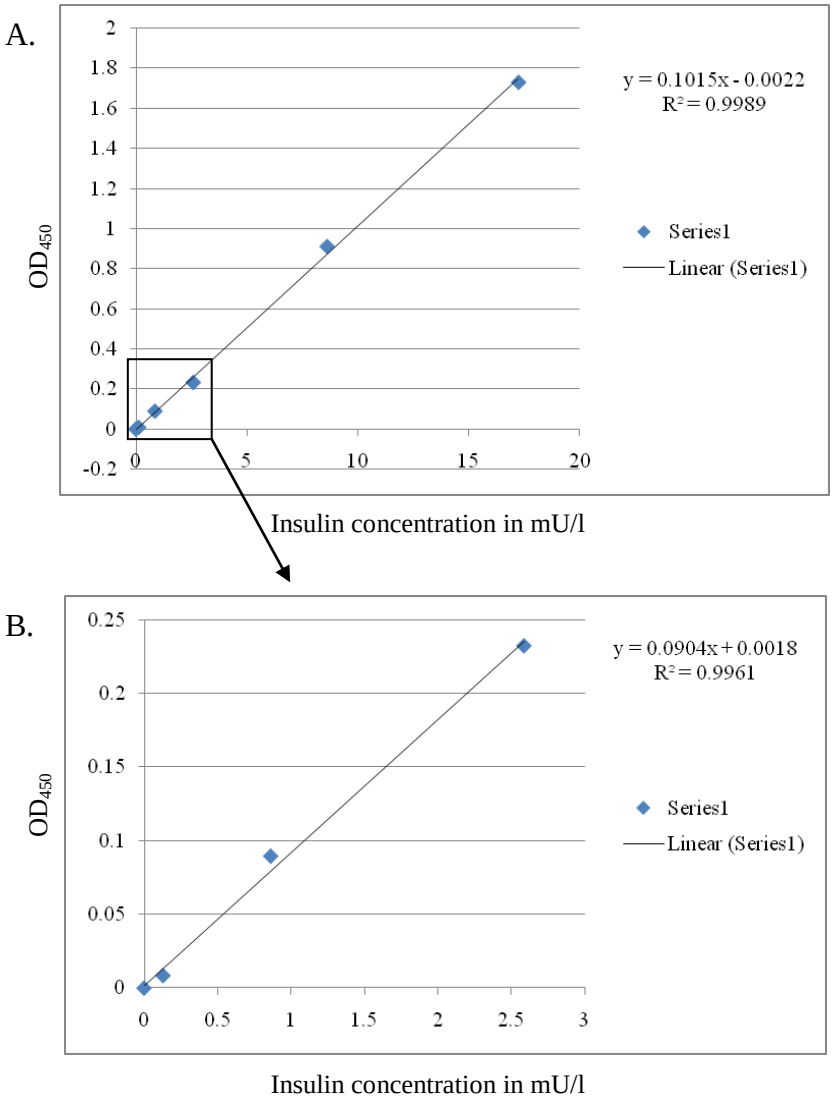


Figure 3-19 Mercodia Ultrasensitive Insulin ELISA standard curve: Mercodia Calibrators 0, 1, 2, 3, 4 and 5 were assayed and OD_{450} values were plotted against insulin concentration in each sample (after subtracting the Calibrator 0 value from all OD_{450} values). A. Standard curve for all the five calibrators (0, 1, 2, 3, 4, 5), B. Standard curve for Calibrators 0, 1, 2 and 3 (lower OD_{450} range)

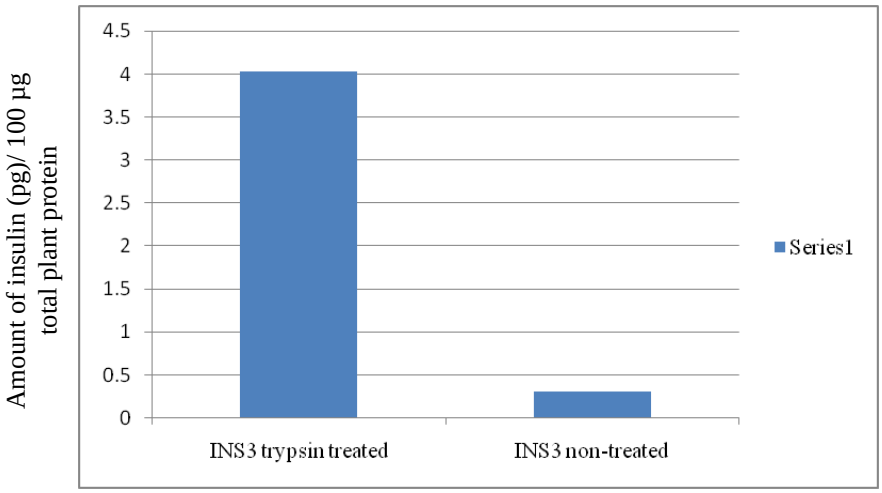


Figure 3-20 Amount of insulin in transgenic tobacco line INS3 as detected by Mercodia ultra sensitive insulin ELISA: Trypsin-treated and non-treated crude plant protein samples (in 2 replicates as suggested by the manufacture) were analysed by Mercodia ELISA. Detectability of insulin in plant extract was significantly increased in trypsin-treated transgenic plant INS3 protein extract (4.0328 pg/100 µg total plant protein) than non-treated INS3 plant extract (0.3087 pg/100 µg total plant protein). No significant differences in absorbance values were seen in the trypsin-treated and non-treated wild type plant extract (Table 3-4).

The plant-made insulin could not be correctly quantified by the above two ELISA assays. However, the western blot data (Figure 3-13) could be used to get a rough estimation of insulin production by the transgenic tobacco. In Figure 3-13, the band intensity of the bovine insulin (100 ng) was compared by eye with those of the transgenic tobacco protein samples (containing 20µg crude plant protein each). The highest amount of insulin was detected in transgenic line INS3 (0.15 % insulin per total plant protein).

Further tests to compare the properties of the plant made insulin with the native/commercial human insulin are desirable. These include *in vitro* bioassays for affinity, insulin receptor binding assays and tests for intrinsic activity. Similarly, *in vivo* activity studies could be conducted by administration of plant made insulin to diabetic mice and determing their response. However, due to time shortage and funds shortage these asssays could not be conducted.

4 CLONING OF *STREPTOCOCCUS PNEUMONIAE* TYPE 2 CAPSULAR POLYSACCHARIDE GENES AND EXPRESSION OF TYPE 2 POLYSACCHARIDE IN TRANSGENIC TOBACCO

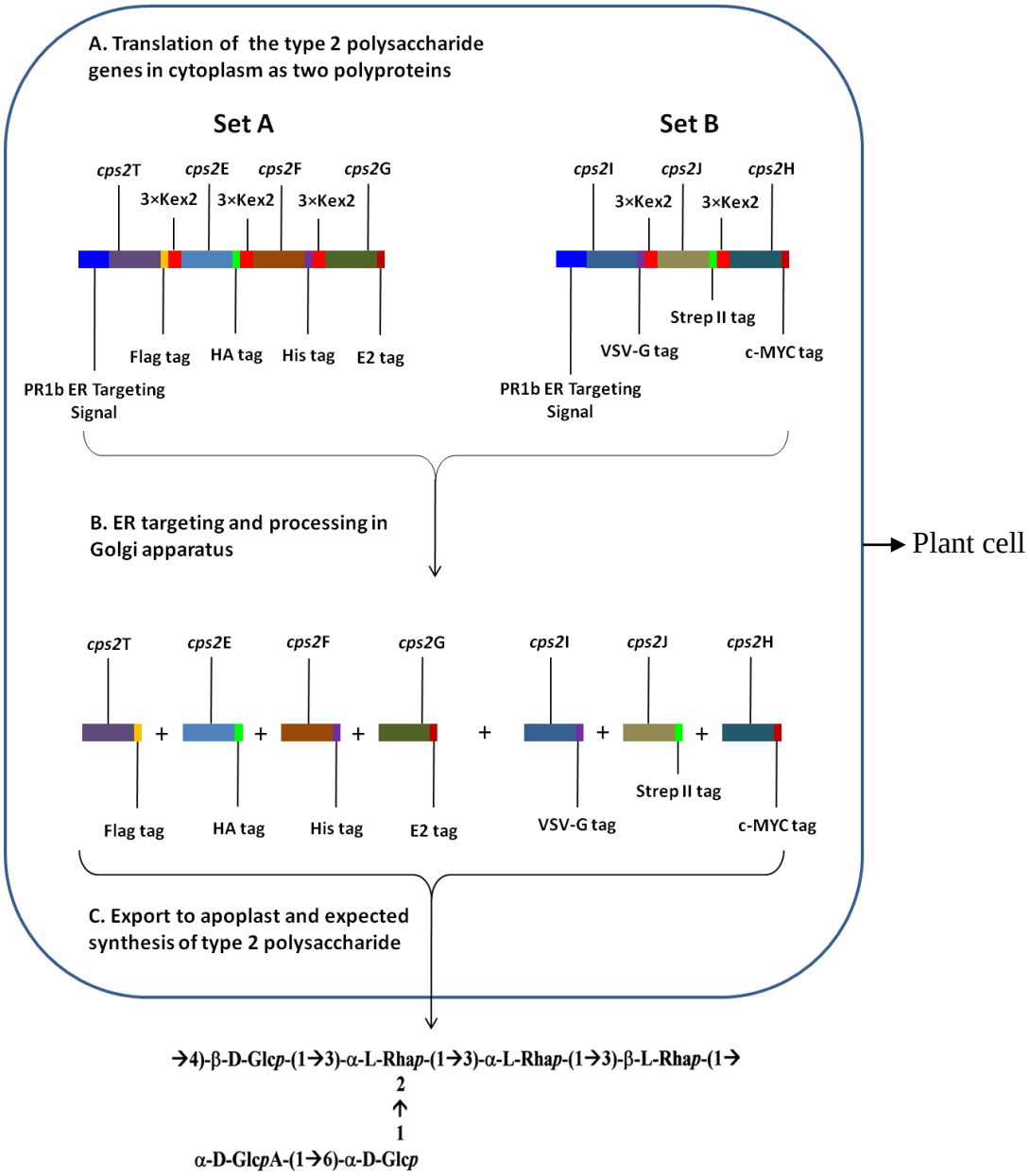
4.1 Chapter overview

This chapter describes the cloning of 7 *Streptococcus pneumoniae* type 2 capsular polysaccharide genes and their novel expression in tobacco, for the production of type 2 polysaccharide *in planta*. The project was aimed at investigating a possible way of producing type 2 polysaccharide vaccine in large quantities at low costs to meet the global demands of this important vaccine. The chapter discusses in detail: the construction of two expression cassettes carrying the 7 type 2 polysaccharide genes; *Agrobacterium* mediated transformation of tobacco with the two expression cassettes; analysis of the transgenes integration in the plant genome, analysis of gene expression at RNA and protein level, and detection of type 2 polysaccharide production in the plant.

4.2 Strategy for expression of the 7 type 2 polysaccharide genes in a single tobacco plant

Figure 4-1 outlines the strategy for *in planta* expression of pneumococcal type 2 polysaccharide. Two expression cassettes were designed for the expression of the 7 type 2 polysaccharide genes in a single tobacco plant. Cassette A carried the polysaccharide genes *cps2*E, *cps2*T, *cps2*F and *cps2*G (set-A genes) while cassette B carried the polysaccharide genes *cps2*I, *cps2*J and *cps2*H (set-B genes). Both expression cassettes were driven by the constitutive CaMV35S promoter for expression in tobacco leaves. Different protein tag sequences were attached to the 3′ end of each gene for subsequent immuno detection. For plant based processing of the individual proteins, 3 tandemly repeated Kex2 sequences were introduced at the junction of two genes (next to the tag sequence of the preceding gene and before the start of the following gene in the expression cassette). Tobacco PR1b endoplasmic reticulum targeting signal sequence (Cutt *et al.*, 1988) was attached to the 5′ end of the polygene in each expression cassette.

The 7 type 2 polysaccharide proteins after expression in the cytoplasm in the form of two polyproteins would be targeted to the endoplasmic reticulum by the respective attached PR1b endoplasmic reticulum targeting signals (Cutt *et al.*, 1988). From the endoplasmic reticulum, the two polyproteins would be transported to the Golgi bodies by the default pathway where the Kex2 protease processing would result in the generation of the 7 type 2 polysaccharide proteins which would then be secreted to the apoplast. Here in the apoplast, the 7 type 2 polysaccharide enzymes would carry out the expected synthesis of the pneumococcal type 2 polysaccharide utilizing the sugar precursors present in the apoplastic fluid (Voitsekhovskaya *et al.*, 2002; Büttner & Sauer, 2000; Yeo *et al.*, 1998; Velíšek & Cejpek, 2005).



Type 2 polysaccharide structure (Janson *et al.*, 1975)

Figure 4-1 Strategy for *Streptococcus pneumoniae* type 2 capsular polysaccharide expression in transgenic tobacco: The 7 type 2 polysaccharide biosynthetic genes were grouped into two sets. Set- A consisted of the polysaccharide genes *cps2*T, *cps2*E, *cps2*F and *cps2*G. Set-A genes were assembled in expression cassette A. Set-B consisted of polysaccharide genes *cps2*I, *cps2*J and *cps2*H. Set-B genes were assembled in expression cassette B. Both expression cassettes were driven by a CaMV35S promoter. The endoplasmic reticulum targeting signal was attached to the N-terminal of each polygene. Different protein tags were attached to the C-terminal end of each gene for subsequent detection. For plant based processing of the individual proteins, 3 tandemly repeated Kex2 sequences were introduced at the junction of two genes. After translation in the cytoplasm, the two polyproteins would be targeted to the endoplasmic reticulum by the respective attached PR1b signal peptides (the signal peptides are then cleaved off in ER). The polyproteins would then be transported into the Golgi apparatus where the 7 type 2 polysaccharide proteins would be released from the two polyproteins as a result of the Kex2 protease processing. These proteins would then be secreted to the apoplast to carry out the expected synthesis of type 2 polysaccharide.

4.3 Cloning of type 2 polysaccharide set-A genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G)

A strategy was designed which involved cloning the *cps2*T gene directly into the *Spe*1/*Pst*1 site of the plant binary vector pKAM1 (section 3.3.3) in frame with the PR1b signal replacing the insulin fusion gene. The resulting vector was named pKAM2 (section 4.3.1). The remaining three genes (*cps2*E, *cps2*F and *cps2*G) were first sequentially cloned into pGEMKA1 (section 3.3.1) multiple cloning site in the following order; *Pst*1-*cps2*E-HA tag-3×Kex2-*Nco*1-*cps2*F-His tag-3×Kex2-*Sph*1-*cps2*G-E2 tag-stop-*Pst*1-*Apa*1 to generate vector pGEMKA4 (section 4.3.2). The DNA segment containing the three genes was then excised from pGEMKA4 at the two *Pst*1 sites flanking these genes and cloned into the *Pst*1 site of the construct pKAM2 in frame with the *cps2*T gene to generate pKAM3 (section 4.3.2).

4.3.1 Cloning of *cps2*T

*cps2*T was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using phusion DNA polymerase. Amplification was carried out with forward primer KA014 (with *Spe*1 restriction site at 5′ end) and two versions of reverse PCR primer; the mini reverse primer KA018 and the full-length reverse primer KA019. KA018 did not contain additional DNA segments apart from the sequence complementary to cps2T while KA019 carried the flag tag sequence, 3×Kex2 cleavage sites and *Pst*1 restriction site in addition to the complementary sequence to *cps2*T. KA014 and KA018 were used to amplify *cps2*T from genomic DNA. The resultant PCR product was then re-amplified with forward primer KA014 and reverse primer KA019 incorporating the Flag tag, 3×Kex2 cleavage sites and *Pst*1 restriction site to the 3′ end of the gene (Figure 4-3A). PCR products were analysed on a 1.5 % agarose gel. The results are shown in Figure 4-2. Lane 1 contained the full-length gene product (1263 bp) while lane 2 contained the shorter version of the gene product (1188 bp).

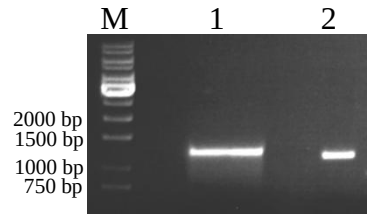


Figure 4-2 PCR amplification of *cps2*T: *cps2*T gene was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using phusion DNA polymerase. To facilitate efficient amplification, *cps2*T was first amplified with forward primer KA014 (with *Spe*1 restriction site at 5′ end) and mini reverse primer KA018 (lane 2, 1188 bp). The PCR product was then re-amplified with forward primer KA014 and full length reverse primer KA019 which incorporated the Flag tag, 3×Kex2 cleavage sites and *Pst*1 restriction site to the 3′ end (lane 1, 1263 bp). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

The *cps2*T gene PCR product and the construct pKAM1 (section 3.3.3) were digested with *Spe*1/*Pst*1 restriction enzymes. The digested *cps2*T fusion gene was then cloned into the *Spe*1/*Pst*1 sites of pKAM1 in frame with PR1b signal to generate pKAM2 (Figure 4-3B). pKAM2 was transformed into XL1-Blue cells and selected on LB agar plates containing 25 µg/ml kanamycin. DNA minipreps from this transformation were digested with *Spe*1/*Pst*1 restriction enzymes and analysed on 1.5% agarose gel for confirmation of *cps2*T fusion gene integration into the vector (Figure 4-4). A DNA fragment of expected size (1263 bp) was observed in all samples except sample1 (lane 1) and sample 7 (lane 7). Clone 2 and 3 were selected for sequencing. Sequencing was carried out with forward primer KA022, reverse prime KA023 and forward primer KA024. After confirmation of correct sequence, clone 3 was selected for further cloning.

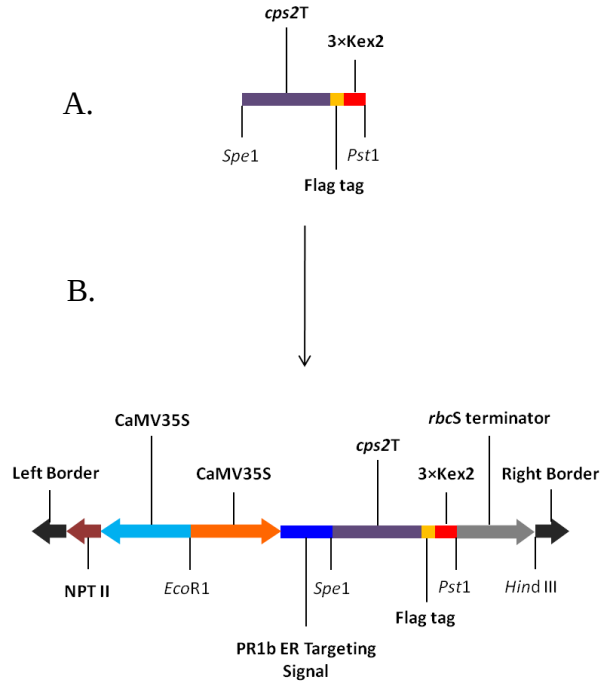


Figure 4-3 Cloning of *cps2*T to generate pKAM2: A. *cps2*T was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA. This amplification also incorporated flag tag, 3×Kex2 cleavage sites and restriction enzyme sites in the amplified product. **B.** The amplified *cps2*T fusion gene was cloned into the *Spe*1/*Pst*1 sites of pKAM1 (section 3.3.3) in frame with PR1b signal to replace insulin fusion gene. The resultant vector was named pKAM2.

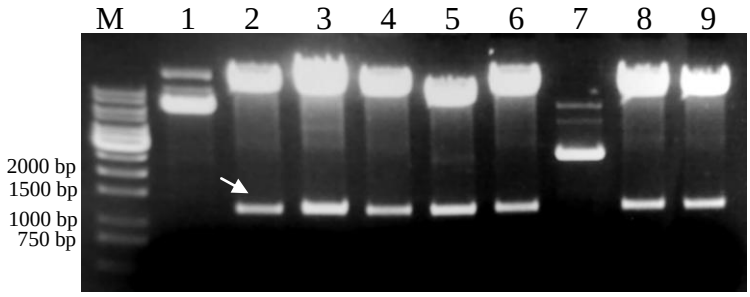


Figure 4-4 Restriction digestion of pKAM2 with *Spe*1/*Pst*1 restriction enzymes for confirmation of *cps2*T integration: Ten mini prep plasmid DNA samples (lane 1-9, sample 10 not shown) were restricted with *Spe*1/*Pst*1 restriction enzymes and analysed on 1.5 % agarose gel. A band of expected size (1263 bp, indicated by an arrow) corresponding to *cps2*T was observed in all samples except sample 1(lane 1) and sample 7 (lane7). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

4.3.2 Cloning of *cps2*E, *cps2*F and *cps2*G

Set-A type 2 polysaccharide remaining three genes *cps2*E, *cps2*F and *cps2*G were amplified from *Streptococcus pneumoniae* strain D39 genomic DNA. The required tag sequences, Kex2 sites and restriction enzyme sites were also introduced during PCR amplification. The amplified gene products were sequentially cloned into pGEMKA1

(section 3.3.1) to generate pGEMKA4 (Figure 4-5). The DNA segment consisting of the three genes was then excised from pGEMKA4 and cloned into the *Pst*1 site of pKAM2 in frame with *cps*2T to generate pKAM3 (Figure 4-5), the plant binary vector carrying all the four set-A polysaccharide genes. Cloning of these genes is detailed in the following sections.

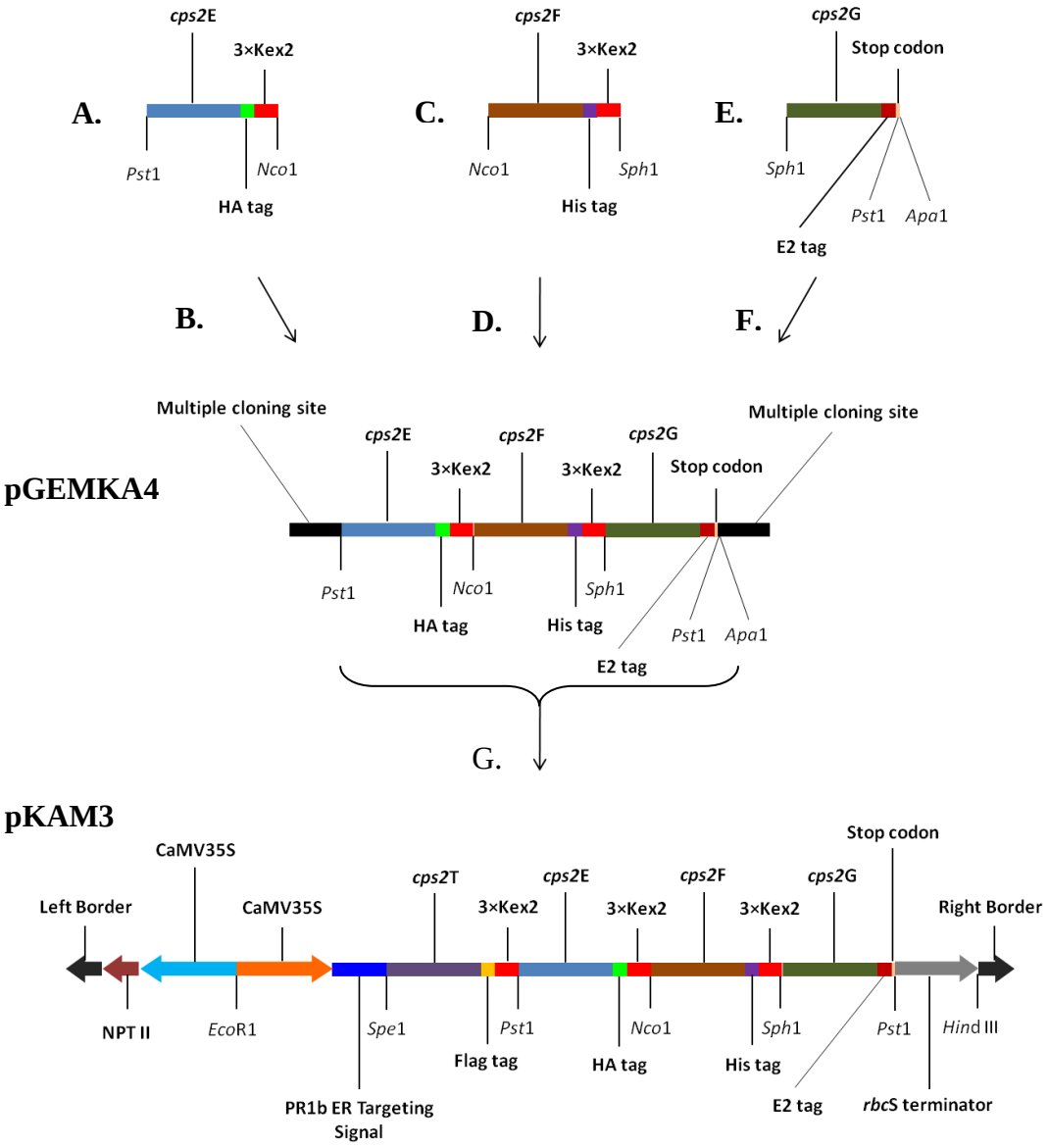


Figure 4-5 Cloning of *cps*2E, *cps*2F and *cps*2G: The remaining three set-A polysaccharide genes (*cps*2E, *cps*2F and *cps*2G) were amplified from *Streptococcus pneumoniae* strain D39 genomic DNA. During amplification the additional required sequences were also attached (**A.** *cps*2E with HA tag, 3×Kex2 sites and *Nco*1 site at 3′ end and *Pst*1site at 5′ end; **C.** *cps*2F with His tag, 3×Kex2 sites and *Sph*1 site at 3′ end, and *Nco*1 site at 5′ end; **E.** *cps*2G with E2 tag, 3×Kex2 sites, stop codon, *Pst*1 and *Apa*1 sites at 3′ end, and *Sph*1 site at 5′ end). These three fusion gene products were sequentially cloned into

pGEMKA1 (section 3.3.1) to generate pGEMKA4 (**B.** *cps2*E gene PCR product was cloned into the *Pst*1/*Nco*1 sites of pGEMKA1to generate pGEMKA2 (not shown); **D.** *cps2*F gene PCR product was cloned into the *Nco*1/*Sph*1 sites of pGEMKA2 in frame with *cps2*E to generate pGEMKA3 (not shown); **F.** *cps2*G gene PCR product was cloned into the *Sph*1/*Apa*1 sites of pGEMKA3 in frame with *cps2*F to generate pGEMKA4). **G.** The DNA segment consisting of the three genes was then excised from pGEMKA4 at the two flanking *Pst*1 sites and cloned into the *Pst*1 site of pKAM2 (section 4.3.1) in frame with *cps2*T to generate the binary vector pKAM3. In pKAM3, the set-A type 2 polysaccharide expression cassette is controlled by CaMV35S promoter, The N-terminal PR1b signal is for targeting of the set-A polysaccharide polyprotein to endoplasmic reticulum, from where the polyprotein will pass onto the Golgi apparatus for the Kex2 protease processing to generate four processed proteins encoded by these four genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G). Ultimately the protein will be excreted to the apoplast. The *rbc*S terminator sequence is attached after the stop codon of the polygene for transcription termination of the polycistronic mRNA.

4.3.2.1 Cloning of *cps*2E

*cps2*E was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using phusion DNA polymerase using forward primer KA020 (with a *Pst*1 site at 5′ end) and the mini reverse primer KA021. The resulting PCR product was used as template to re-amplify *cps2*E using forward primer KA020 and full-length reverse primer KA017. This amplification resulted in the attachment of the HA tag, 3×Kex2 cleavage sites and an *Nco*1 restriction site to the 3′ end (Figure 4-5A). PCR products were analysed on 1.5% agarose gel (Figure 4-6).

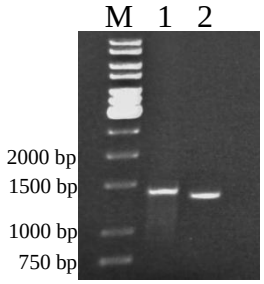


Figure 4-6 PCR amplification of *cps2*E: *cps2*E was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using phusion DNA polymerase. *cps2*E was first amplified with forward primer KA020 (with a *Pst*1 site at 5′ end) and mini reverse primer KA021 (lane 2, 1371 bp). The PCR product was re-amplified with forward primer KA020 and full length reverse primer KA017. This amplification incorporated the HA tag, 3×Kex2 cleavage sites and *Nco*1 restriction site to the 3′ end (lane 1, 1449 bp). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

The full-length *cps2*E fusion gene and pGEMKA1 (section 3.3.1) were restricted with *Pst*1/*Nco*1 restriction enzymes. Digested *cps2*E fusion gene was then cloned into the *Pst*1/*Nco*1 sites of pGEMKA1. The resultant vector (pGEMKA2) was used to transform DH10B cells, which were selected on LB agar containing 100 µg/ml ampicillin. Plasmid minipreps were *Pst*1/*Nco*1 digested to confirm *cps2*E integration. Restriction digests were analyzed on 1.5% agarose gel (Figure 4-7). A fragment of correct size

(1449bp) was observed in 4 DNA minipreps. Clone 5 and 6 were selected for sequencing. Sequencing was carried out with the universal primers T7 and SP6 that flank the multiple cloning site of pGEMKA2. Clone 6 was selected for further manipulation after confirmation of correct sequence.

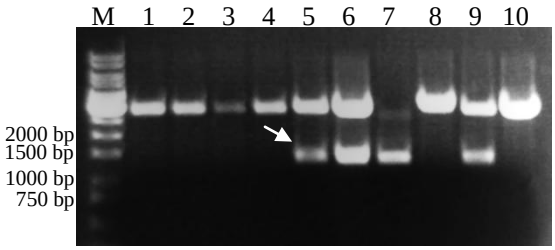


Figure 4-7 Restriction digestion of pGEMKA2 with *Pst*1/*Nco*1 restriction enzymes to confirm *cps2*E integration: Ten mini prep samples (lane 1-10) were restricted with *Pst*1/*Nco*1 restriction enzymes and analysed on 1.5 % agarose gel. A band of expected size (1449 bp, indicated by an arrow) corresponding to *cps2*E was observed in sample 5 (lane 5), sample 6 (lane 6), sample 7 (lane 7) and sample 9 (lane 9). The remaining samples could contain the self-ligated vector. Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

4.3.2.2 Cloning of *cps2*F

*cps2*F was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA with forward primer KA025 (containing a *Nco*1 site at 5′ end) and the short reverse primer KA027. The PCR product was re-amplified with forward primer KA025 and full-length reverse primer KA026. This amplification attached a His tag, 3×Kex2 cleavage sites and a *Sph*1 site to the 3′ end of the gene (Figure 4-5C). The PCR product was analysed on 1.5% gel (Figure 4-8).

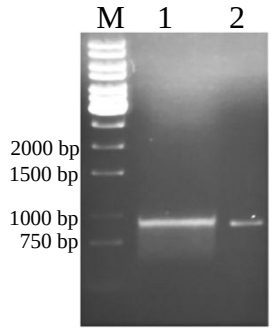


Figure 4-8 PCR amplification of *cps2*F: *cps2*F was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using phusion DNA polymerase. The gene was first amplified with forward primer KA025 (with an *Nco*1 site at 5′ end) and mini reverse primer KA027 (lane 2, 921 bp). The PCR product was re-amplified with forward primer KA025 and the full length reverse primer KA026 incorporating the

His tag, 3×Kex2 sites and *Sph*1 site to the 3′ end (lane 1, 990 bp). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

The *cps2*F fusion gene and the construct, pGEMKA2 (section 4.3.2.2) were digested with *Nco*1/*Sph*1. Digested *cps2*F fusion gene product was cloned into the *Nco*1/*Sph*1 sites of pGEMKA2 to generate pGEMKA3. The resultant vector was used to transform XL1-Blue cells, which were isolated on LB agar plates containing 100 µg/ml ampicillin. Plasmid minipreps isolated from isolated from ampicillin resistant colonies were digested with *Nco*1/*Sph*1 and were analysed on 1.5% agarose gel. The results are shown in Figure 4-9, a fragment of correct size (990 bp) was seen in 5 clones. Clones 1 and 3 were sequenced with reverse primers KA031 and KA032. Clone 1 was selected for further cloning after sequence screening.

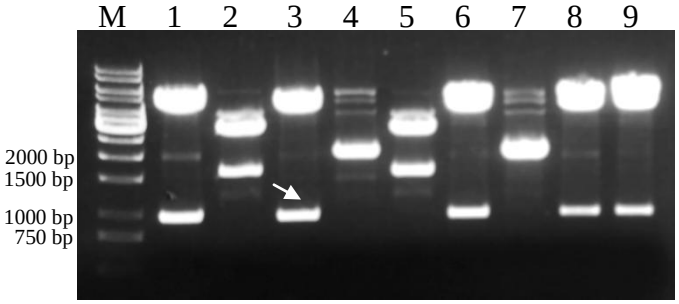


Figure 4-9 Restriction digestion of pGEMKA3 with *Nco*1/*Sph*1 restriction enzymes to confirm *cps2*E integration: DNA mini prep samples) were digested with *Nco*1/*Sph*1 restriction enzymes and the products were analysed on 1.5 % agarose gel. A band of expected size (990 bp, indicated by an arrow) corresponding to *cps2*F was seen in lane 1 (clone 1), lane 3 (clone 3), lane 6 (clone 6), lane 8 (clone 8) and lane 9 (clone 9). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

4.3.2.3 Cloning of *cps2*G

Attempts at amplification of the *cps2*G sequence from *Streptococcus pneumoniae* strain D39 genomic DNA using primer KA028 and a short reverse primer KA030 were unsuccessful. Amplification of *cps2*G from *Streptococcus pneumoniae* strain D39 genomic DNA was therefore carried out with forward primer KA028 (with *Sph*1 site at 5′ end) and full-length reverse primer KA029 (carrying an E2 tag, stop codon, a *Pst*1 and an *Apa*1 site in addition to the sequence complementary to *cps2*G). The *Apa*1 site was introduced into the gene product to enable cloning of the sequence into the pGEMKA3 construct (section 4.3.2.2). The *Pst*1 site was introduced for the subsequent cloning of this sequence into the construct pKAM2 (section 4.3.1) to generate pKAM3

(Figure 4-5G). Figure 4-5E shows a diagram of the amplified *cps2*G fusion gene with the attached sequences. Figure 4-10 shows the agarose gel analysis of the amplified PCR product.

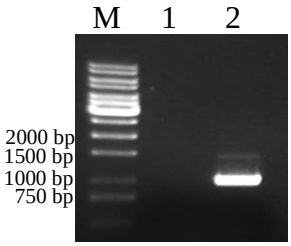


Figure 4-10 PCR amplification of *cps2*G: *cps2*G was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using phusion DNA polymerase. Amplification was first tried with forward primer KA028 (with *Sph*1 site at 5′ end) and short reverse primer KA030. However, the PCR did not work (lane 1). The amplification was then tried with forward primer KA028 and the full length reverse primer KA029 (carrying an E2 tag, a stop codon, a *Pst*1 site and an *Apa*1 site). The PCR worked this time and a fragment (1062 bp) corresponding to *cps2*G was amplified (lane 2). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

The *cps2*G PCR product and the construct pGEMKA3 (section 4.3.2.2) were digested with *Sph*1/*Apa*1 restriction enzymes. The digested gene product was cloned into the *Sph*1/*Apa*1 sites of pGEMKA3 (section 4.3.2.2) to generate pGEMKA4 (Figure 4-5F) which was transformed into XL1-Blue electrocompetent cells. Transformed cells were selected on LB agar plates containing 100 µg/ml ampicillin. In pGEMKA4 the 3483 bp DNA segment consisting of the three genes (*cps2*E, *cps2*F and *cps2*G) is flanked by two *Pst*1 sites (Figure 4-5). DNA minipreps were therefore digested with *Pst*1 enzyme to identify clones containing this DNA fragment and analysed on 1.5 % gel (Figure 4-11). A band of correct size (3483 bp) was demonstrated in 9 samples. Clone 3 was sequencined using forward primer KA033 and reverse primer KA032. The sequencing results demonstrated the correct sequence of the insert and this clone was used for further manipulations**.**

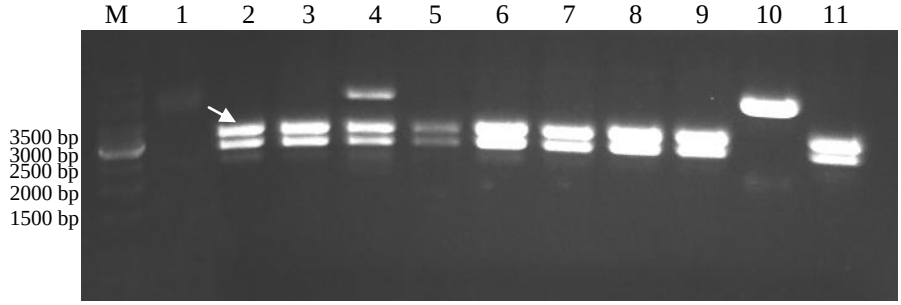


Figure 4-11 Restriction digestion of pGEMKA4 with *Pst*1 restriction enzymes to confirm *cps2*G integration: Ten mini prep samples (lane 2-11) were digested with *Pst*1 restriction enzyme. The three genes (*cps2*E, *cps2*F and *cps2*G) in pGEMKA4 are flanked by two *Pst*1 sites. The restriction digestion resulted in the generation of a 3483 bp fragment (indicated by an arrow) corresponding to these three genes. The band was present in all samples (lane 2, clone 1; lane 3, clone 2; lane 4, clone 3; lane 5, clone 4; lane 6, clone 5; lane 7, clone 6; lane 8, clone 7, lane 9, clone 8 and lane 11, clone 10) except sample 9 (lane 10). Lane 1 contains the uncut clone 1. Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

4.3.2.4 Cloning of *cps2*E, *cps2*F and *cps2*G into pKAM2 to generate pKAM3

The next step in cloning of the set-A type 2 polysaccharide genes was to cut pGEMKA4 (section 4.3.2.3) with *Pst*1 restriction enzyme to excise the DNA fragment consisting of the three genes (*cps2*E, *cps2*F and *cps2*G) and clone into *Pst*1 site of pKAM2 in frame with *cps2*T to generate the construct pKAM3 (Figure 4-12 and Figure 4-5G). pKAM3 was used to transform DH10B cells which were selected on LB agar plates containing 25 µg/ml kanamycin. DNA minipreps from resistant colonies were digested with *Pst*1 restriction enzyme and the products were run on 1.5% agarose gel (Figure 4-13). A band of correct size of 3483 bp corresponding to the three genes (*cps2*E, *cps2*F and *cps2*G) was observed in most of the samples.

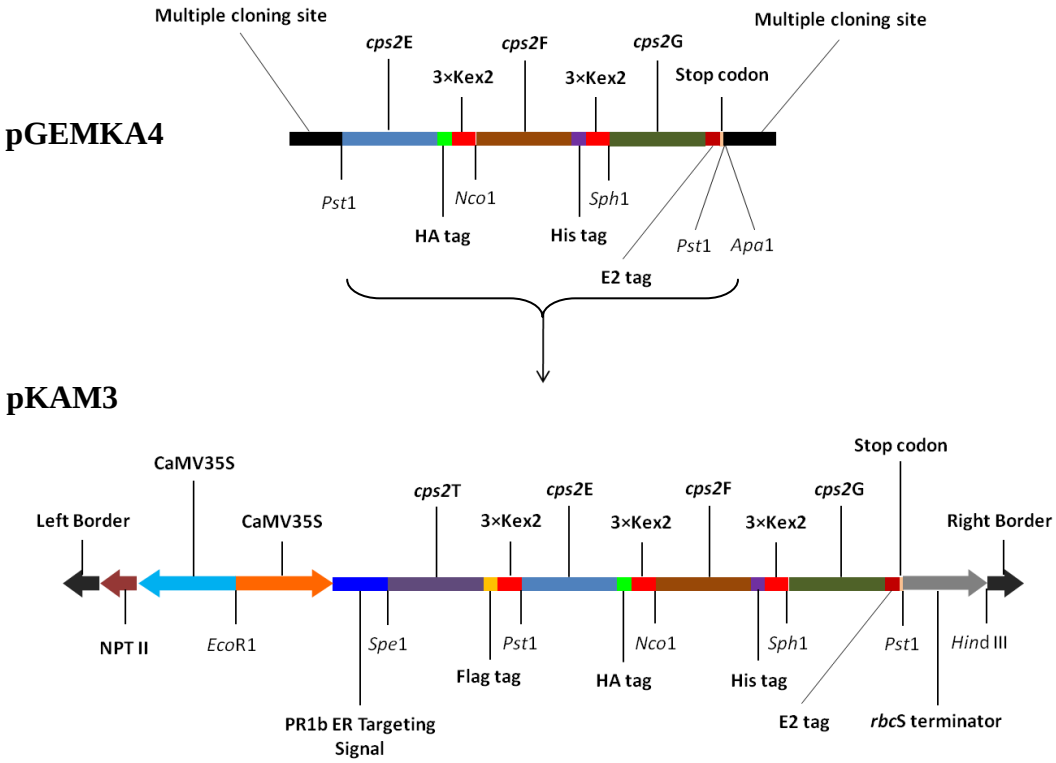


Figure 4-12 Cloning of *cps2*E, *cps2*F and *cps2*G into pKAM2 to generate pKAM3: The DNA segment consisting of the three genes (*cps2*E, *cps2*F and *cps2*G) was excised from pGEMKA4 and cloned into the *Pst*1 site of pKAM2 (section 4.3.1) in frame with *cps2*T to generate the final binary vector pKAM3. In pKAM3, all the four set-A type 2 polysaccharide genes are under the control of a single CaMV35S promoter in the form of a single polygene, The N-terminal PR1b signal serves to target the set-A polysaccharide polyprotein to endoplasmic reticulum. The *rbc*S terminator sequence is attached after the stop codon of the polygene for transcription termination.

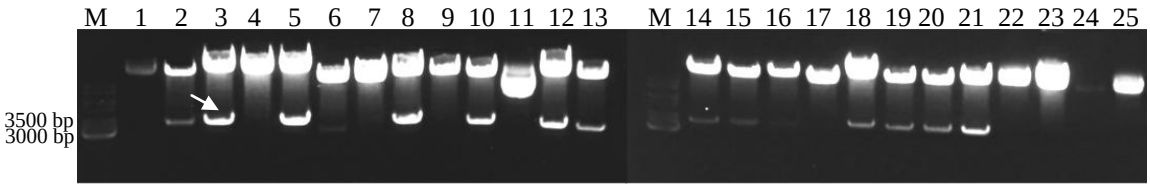


Figure 4-13 Restriction digestion of pKAM3 with *Pst*1 restriction enzyme to confirm integration of the three genes (*cps2*E, *cps2*F and *cps2*G): Twenty-four mini prep samples (lane 2-25) were digested with *Pst*1 restriction enzyme. Lane 1 contained the uncut clone 1 DNA. A band of expected size of 3483 bp (indicated by an arrow) corresponding to the three genes (*cps2*E, *cps2*F and *cps2*G) was observed in most of the samples. Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

Since, in pKAM3 the DNA segment consisting of the three genes (*cps2*E, *cps2*F and *cps2*G) is flanked by two *Pst*1 sites, the insert DNA can either be integrated in the correct orientation (*Pst*1-*cps2*E- *cps2*F-*cps2*G- *Pst*1) or the opposite orientation (*Pst*1-*cps2*G- *cps2*F-*cps2*E- *Pst*1). Therefore, positive pKAM3 clones from *Pst*1 digestion (Figure 4-13) were further analysed by restriction digestion with *Spe*1 restriction

enzyme. There are two *Spe*1 sites in pKAM3; one at the start of *cps2*T and another in *cps2*E. If the insert had been integrated in the correct orientation (*Pst*1-*cps2*E- *cps2*F-*cps2*G- *Pst*1) then the *Spe*1 fragment would be 2034 bp. However, if the insert DNA is integrated in the reverse orientation (*Pst*1-*cps2*G- *cps2*F-*cps2*E- *Pst*1) then the *Spe*1 fragment would be 3963 bp. Figure 4-14 shows the *Spe*1 restriction digestion of selected pKAM3 clones. A band of correct size of 2034 bp (indicated by an arrow) corresponding to the correct integration of these genes was detected in four samples. A band of 3963 bp (denoted by asterisks) corresponding to incorrect integration of the insert was observed in the remaining samples except clone 5, 7 and 9 (in which the digestion did not work).

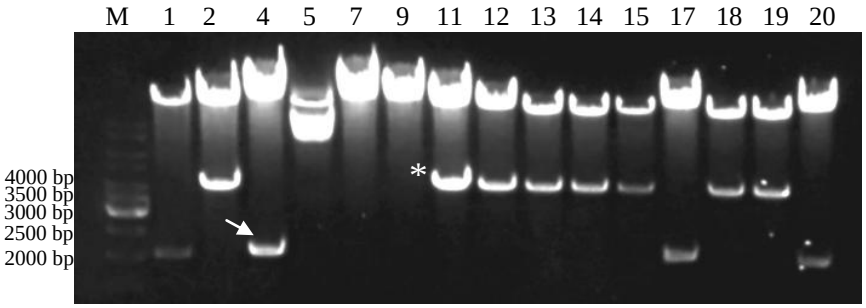


Figure 4-14 Confirmation of the orientation of the DNA segment consisting of three genes (*cps2*E, *cps2*F and *cps2*G) in pKAM3: The DNA segment consisting of three genes (*cps2*E, *cps2*F and *cps2*G) is flanked by two *Pst*1 sites. Hence, positive pKAM3 clones from *Pst*1 digestion (Figure 4-13) were further analysed by restriction digestion with *Spe*1 restriction enzyme to determine orientation of the insert. There are two *Spe*1 sites in pKAM3; one at the start of *cps2*T and another 2034 bp downstream of the start codon of *cps2*T located in *cps2*E if the insert has been integrated in the correct orientation. However, if the insert has been integrated in reverse orientation (*Pst*1-*cps2*G- *cps2*F-*cps2*E- *Pst*1), then the gap between the two *Spe*1 sites would be 3963 bp. Lane number refers to clone number. A band of correct size (2034 bp, indicated by an arrow) detected in clone 1, clone 4, clone 17 and clone 20 confirms the correct orientation of the integrated DNA segment in these four clones. Incorrect orientation of the insert was observed in the remaining samples. The incorrect band (3963 bp) in these samples is denoted by asterisks. Digestion in clone 5, 7 and 9 did not work. Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

pKAM3 clone 5 was sequenced: *cps2*E was sequenced with forward primer KA024, *cps2*E forward PCR primer KA020 and *cps2*E reverse short PCR primer KA021. *cps2*F was sequenced with *cps2*F forward PCR primer KA025, reverse sequencing primer KA031 and *cps2*F short reverse PCR primer KA027. *cps2*G was sequenced with forward primer KA033, *cps2*G short reverse primer KA030 and reverse primer KA009. After sequence confirmation, a midi scale pKAM3 plasmid DNA was isolated for *Agrobacterium*-mediated transformation of tobacco.

4.4 Cloning of set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H)

4.4.1 Strategy-1 for cloning set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H)

An initial strategy designed for cloning of the three set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H) is outlined in Figure 4-15. The strategy involved amplifying the three genes from *Streptococcus pneumoniae* strain D39 genomic DNA and attaching the required tag sequences, Kex2 sites and restriction enzyme sites (Figure 4-15 A, C and E). The next step was to sequentially clone these genes into pGEMKA1 (section 3.3.1) in the order; 5′*Spe*1-*cps2*H-c-MYC tag-3Kex2-*Nco*1-*cps2*I-VSV-G tag-3Kex2-*Sph*1-*Cps2*J-Strep 2 tag-stop-*Pst*1-*Apa*13′ (Figure 4-15 B, D and F). The DNA segment consisting of the three genes would then be excised from the resultant vector pGEMKA7 and cloned into the *Spe*1/*Pst*1 sites of pKAM1 (section 3.3.3) to generate the final construct (Figure 4-15 G) replacing the mini-insulin fusion gene in pKAM1. However, a number of cloning anomalies were experienced and this strategy was not pursued. These are detailed in the following sections. To overcome these cloning difficulties a new strategy was designed that is detailed in section 4.4.2.

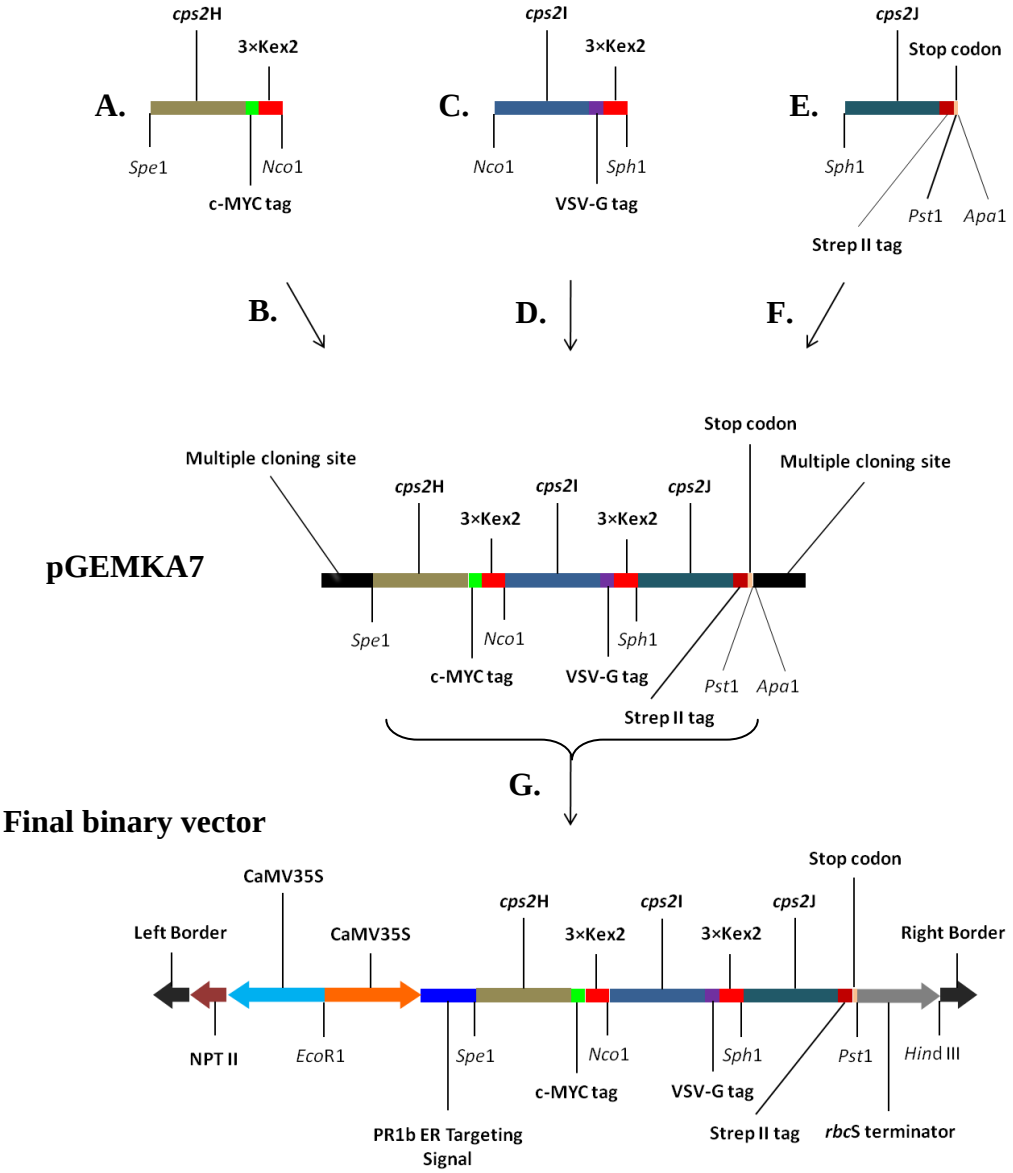


Figure 4-15 Initial cloning strategy of set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H): The three set-B polysaccharide genes (*cps2*I, *cps2*J and *cps2*H) were amplified from *Streptococcus pneumoniae* strain D39 genomic DNA and the additional required sequences were incorporated (**A.** 5′*Spe*1-*cps2*H-c-MYC tag-3Kex2-*Nco*13′; **C.** 5′*Nco*1-*cps2*I-VSV-G tag-3Kex2-*Sph*13′; **E.** 5′*Sph*1-*Cps2*J-Strep 2 tag-stop-*Pst*1-*Apa*13′). These three gene products would be sequentially cloned into pGEMKA1 (section 3.3.1) to generate pGEMKA7 (**B, D** and **F**). **(G)** The DNA segment consisting of the three genes would then be excised from pGEMKA7 and cloned into the *Spe*1/*Pst*1 sites of pKAM1 (section 3.3.3) to generate the final plant binary vector. Due to cloning anomalies (details are given in the following sections), the strategy was not successful; hence, a revised strategy (strategy-2) was designed (for detail see section 4.4.2).

4.4.1.1 Cloning of *cps2*H with strategy-1

*cps2*H was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using the forward primer KA035 (with an *Spe*1 site at 5′ end) and the short reverse primer KA037. The PCR product was re-amplified with forward primer KA035 and the full length reverse primer KA036 (carrying a c-Myc tag, 3×Kex2 cleavage sites and an *Nco*1restriction site). Figure 4-15A shows diagram of the amplified *cps2*H PCR product. PCR products were analysed on 1.5% agarose gel (Figure 4-16). A band of correct size (1254 bp) corresponding to the full-length *cps2*H fusion gene can be seen in lane 1 in Figure 4-16.

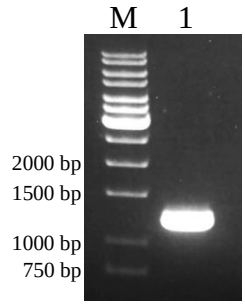


Figure 4-16 PCR amplification of *cps2*H: *cps2*H was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA. The gene was first amplified with forward primer KA035 (with a *Spe*1 site at 5′ end) and short reverse primer KA037. The PCR product was re-amplified with forward primer KA035 and the full length reverse primer KA036 (carrying a c-Myc tag, 3×Kex2 cleavage sites and an *Nco*1restriction site) to get the full length (1254 bp) *cps2*H gene product (lane 1). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

*cps2*H PCR product and the construct pGEMKA1(section 3.3.1) were digested with *Spe*1/*Nco*1 restriction enzymes. Digested *cps2*H gene product was cloned into the *Spe*1/*Nco*1 sites of pGEMKA1 to generate pGEMKA5. pGMEKA5 was used to transform XL1-Blue cells which were selected on LB agar plates containing 100 µg/ml ampicillin. Plasmid minipreps from these resistant clones were digested with *Spe*1/*Nco*1 and were analysed on a 1.5% agarose gel. Figure 4-17 shows results of the restriction digestion. A 1248 bp fragment corresponding to *cps2*H fusion gene was observed in all the 6 samples. Clone 2 and clone 5 were selected for sequencing. Sequencing was carried out with commercially available forward primer SP6 and reverse primer KA032. Clone 5 was selected for further cloning after correct sequencing result.

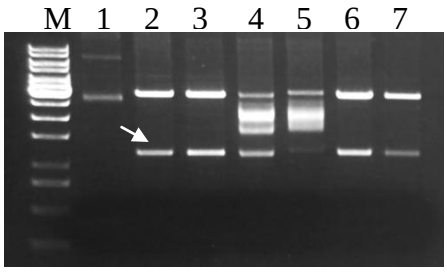


Figure 4-17 Restriction digestion of pGEMKA5 with *Spe*1/*Nco*1 to confirm *cps2*H fusion gene product integration: Six mini prep samples (lane 2-7) were digested with *Spe*1/*Nco*1 restriction enzymes. The restriction digestion resulted in the generation of a 1248 bp fragment (indicated by an arrow) corresponding to *cps2*H fusion gene. The band was present in all the 6 samples (lane 2-7). Lane 1 contained the non-restricted clone 1 DNA. Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

4.4.1.2 Cloning of *cps2*I with strategy-1

Amplification of *cps2*I from *Streptococcus pneumoniae* strain D39 genomic DNA was first attempted with forward primer KA038 and the short reverse primer KA040, this however was unsuccessful. Amplification with forward primer KA038 and full-length reverse primer KA039 was successful. Figure 4-15C shows diagram of the amplified *cps2*I fusion gene product with the attached sequences. Figure 4-18 shows results of agarose gel analysis of amplified *cps2*I fusion gene. No amplified band can be seen in lane 1(PCR result from shorter version of reverse primer i.e. KA038). A fragment of correct size (1251 bp) can be seen in lane 2 containing the full- length PCR product.

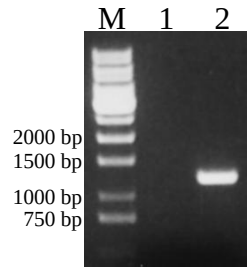


Figure 4-18 PCR amplification of *cps2*I: *cps2*I was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA. The gene was first test-amplified with forward primer KA038 (with an *Nco*1 site at 5′ end) and short reverse primer KA040. However the PCR did not work (lane 1). The amplification was then tried with forward primer KA038 and the full-length reverse primer KA039 (carrying a VSV-G tag, 3×Kex2 cleavage sites and a *sph*1 site). The PCR was successful this time and a fragment corresponding to *cps2*I fusion gene product was amplified (lane 2, 1251 bp). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

*cps2*I PCR product and the construct pGEMKA5 (section 4.4.1.1) were digested with *Nco*1/*Sph*1. The digested fusion gene was cloned into the *Nco*1/*Sph*1 sites of pGEMKA5 (section 4.4.1.1) to generate pGEMKA6. Transformation of electrocompetent *E. coli* cells with pGEMKA6 suffered from a series of cloning anomalies that did not allow the assembly of the three set-B type 2 polysaccharide genes in the form of a single polygene. After transformation of electrocompetent *E. coli* cells with pGEMKA6, the transformed cells were grown overnight on LB agar plates containing 100 µg/ml ampicillin. Though resistant colonies were obtained in some of these experiments, in most cases the restriction digestion analyses (with *Nco*1/*Sph*1 restriction enzymes) of mini prep samples to confirm integration of *cps2*I fusion gene did not produce any fragment specific to *cps2*I fusion gene (1245 bp). Furthermore, attempts to clone *cps2*I in frame with *cps2*H resulted in the deletion of the already cloned gene *cps2*H (section 4.4.1.1). The following parameters were checked to overcome/point out the cloning anomalies: (1) Two different UV transilluminators (Syngene Gel Vue transilluminator, model GVS30 and UVP Dual Intensity transilluminator, model TM-40) were used for cutting the gel slices containing the digested insert and vector DNA segments so as to avoid degradation of DNA with minimum exposure of the digested DNA segments to UV light. (2) Ligation of the insert into the vector was carried out with and without dephosphorylation of the vector after restriction digestion. Dephosphorylation of the vector can overcome problems arising due to self-ligation of the vector. (3) Vector/insert ratio for ligation ranged from 2:1, 1:1, 1:2 and 1:3 (vector: insert) to determine the optimum ratios. (4) Ligation of the insert into the vector was performed at 16°C overnight and at room temperature for 30 min to estimate the optimum ligation temperature. (5) Ligation mixture was cleaned up either by dialysis or ethanol precipitation to find the most efficient desalting method, to avoid arcing during electroporation and get maximum yield of purified ligation product. (6) Amount of ligation mixture used for transformation involved 0.5 µl, 1 µl, 1.5 µl, 2 µl and 3 µl to obtain enough resistant colonies as in some cases very few or no colonies were obtained. (7) Two different mini prep DNA extraction protocols were tried; diatomaceous earth mini prep and Qiagen plasmid DNA mini prep so as to get maximum plasmid DNA yield and much cleaner DNA for digestion analysis as in some cases the amount of DNA obtained was very low (8) Electrocompetent cells used for transformation included NEB 5-alpha and MegaX DH10B to analyse the effect of host

cells on cloning of *cps2*I. Figure 4-19A shows an example of *Nco*1/*Sph*1 restriction digestion results of 10 pGEMKA6 mini prep samples that were obtained after transformation of NEB 5-alpha cells with pGEMKA6. The enzymes did not appear to cut these samples as no difference was observed between the uncut sample 1 (lane 1) and the ten digested plasmid DNA samples (lane 2-11). However, size variation in the plasmid DNA of these clones can be seen as sample 3 (lane 4), sample 8 (lane 9) and sample 9 (lane 11) appear to be a higher molecular weight than the rest of the samples. Clone 3 was selected for further digestion with *Spe*1/*Nco*1 restriction enzymes to confirm whether the *cps2*H fusion gene was still intact (if it is a self-ligated vector) or not (in that case the new vector would be a recombinant that has lost the *cps2*H fusion gene). Figure 4-19B shows results of this digestion. If clone 3 was a self-ligated vector, then *Spe*1/*Nco*1 digestion should have released a 1248 bp fragment specific to *cps2*H fusion gene. No such band was observed in the digested clone 3 DNA sample (lane 2). In fact the uncut sample (lane 1) and the digested sample (lane 2) appear to be similar. It can be inferred that clone 3 was actually a recombinant that has lost the previously cloned *cps2*H gene during cloning of *cps2*I (which also did not integrate into the vector).

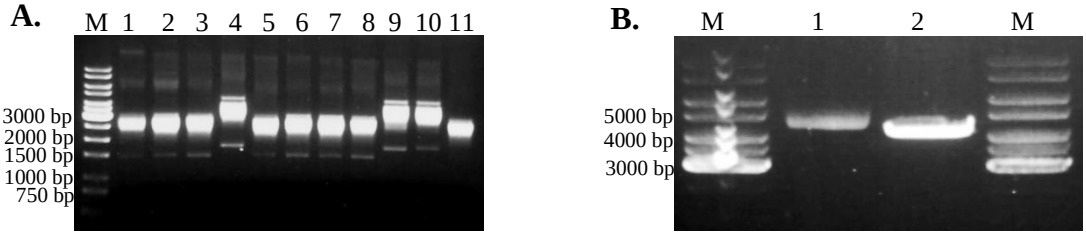


Figure 4-19 Restriction digestion of pGEMKA6 to confirm cps2I fusion gene integration (NEB 5-alpha electrocompetent cells were used in the generation of these clones): A. Ten mini prep samples (lane 2-11) were digested with *Nco*1/*Sph*1 restriction enzymes. Lane 1 contained uncut clone 1DNA. The restriction enzymes did not cut any of these plasmid DNA samples as there is no difference between the uncut clone 1 (lane 1) and the digested samples except size variation in the original plasmid DNA as can be seen in lane 4 (sample 3), lane 9 (sample 8) and lane 11 (sample 9). **B.** Clone 3 was further restricted with *Spe*1/*Nco*1 restriction enzymes to confirm whether it is just a self-ligated pGEMKA5 plasmid (section 4.4.1.1) or it is a recombinant.

Another example of the *cps2*I cloning and restriction digestion analysis of mini prep DNA with *Nco*1/*Sph*1 restriction enzymes is shown in Figure 4-20A. In this experiment pGEMKA6 was used to transform DH10B cells. This time a band of expected size of

1245 bp corresponding to *cps2*I was observed in 5 clones confirming *cps2*I integration into pGEMKA6. Clone 3 and clone 5 DNA was further digested with *Spe*1/*Nco*1 restriction enzymes to make sure that the already cloned *cps2*H fusion gene is still present in pGEMKA6. The results of this digestion are shown in Figure 4-20B. Surprisingly, no band corresponding to *cps2*H (the gene that was cloned before *cps2*I) was observed in either of the two digested DNA samples; clone 3 (lane 2) and clone 5 (lane 5). Again, it appeared that some recombination event had caused the deletion of the previously cloned *cps2*H gene. In fact the uncut (lane 1 and lane 4) and cut samples (lane 2 and lane 5) appear to be similar. Deletion of *cps2*H was further confirmed by PCR using pGEMKA6 clone 3 and clone 5 DNA as template and using pGEMKA5 (section 4.4.1.1) as positive control. PCR was carried out with *cps2*H forward PCR primer KA035 and reverse primer KA037. No amplification was observed in both of the pGEMKA6 samples, however, amplification of correct size was observed in the positive control pGEMKA5 (data not shown).

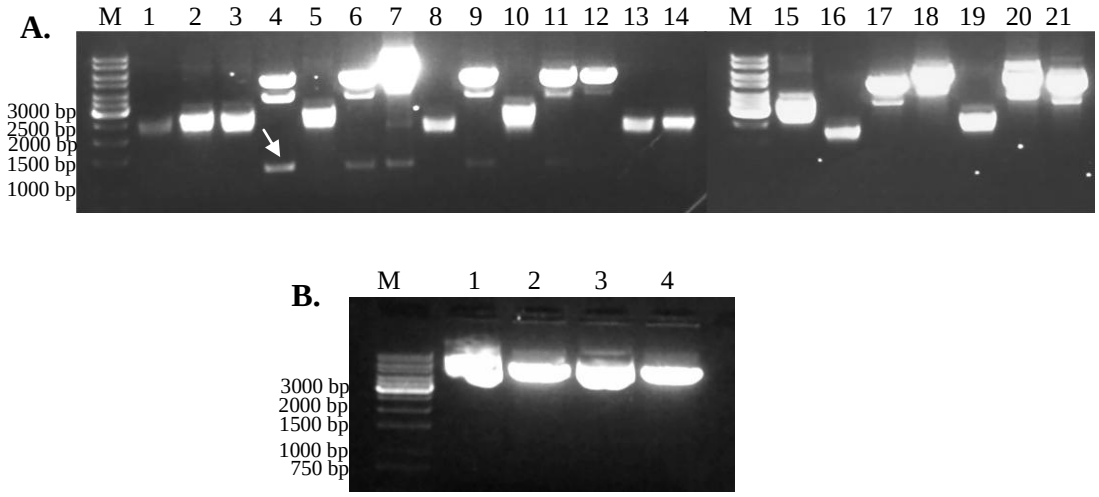


Figure 4-20 Restriction digestion of pGEMKA6 to confirm *cps2*I integration (ElectroMAX DH10B electrocompetent cells were used in the generation of these clones): A. Twenty mini preps (lane 2-21) were digested with *Nco*1/*Sph*1 restriction enzymes. Lane 1 contained uncut sample 1. A band of expected size (1245 bp, indicated by an arrow) corresponding to *cps2*I was observed in lane 4 (clone 3), lane 6 (clone 5), lane 7 (clone 6), lane 9 (clone 8) and lane 11 (clone 10). **B.** Clone 3 and clone 5 were further digested with *Spe*1/*Nco*1 restriction enzymes to make sure that the previously cloned gene *cps2*H is still present in pGEMKA6. No specific band corresponding to *cps2*H was observed in any of the digested DNA samples (clone 3, lane 2; clone 5, lane 4). Lane 1 contained the uncut clone 3 DNA sample and lane 4 contained the uncut clone 5 DNA sample. Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

To further investigate the nature of these consistent cloning problems, pGEMKA6 clone 3 and clone 5 DNA were sequenced across the multiple cloning site flanking the cloned

genes. Sequencing was carried out with universal forward primer SP6, *cps2*H forward PCR primer KA035, forward primer KA044 and reverse primer KA032. Sequencing with primers KA035 and KA044 did not work, which was further evidence of the deletion of *cps2*H from pGEMKA6 as these primers have binding sites in *cps2*H region. Sequencing with forward primer SP6 and reverse primer KA032 flanking the multiple cloning site enabled the complete sequencing of *cps2*I which after alignment with the original sequence was found to be 100% correct. However, no sequence of *cps2*H was found which confirmed its deletion from pGEMKA6. Unexpectedly, in place of *cps2*H sequence a 433 bp DNA fragment was found upstream of the start codon of *cps2*I which shares a six bases (CCATGG) homology with the target site (5′ end of *cps2*H) . This 6 bp DNA fragment is actually the *Nco*1 restriction site that is found at junction of *cps2*H and *cps2*I. This 433 bp DNA segment is not present in the original parental vector pGEM-T Easy. The sequence was present in both of the pGEMKA6 clones that were sequenced (clone 3 and clone 5). This 433 bp sequence was analysed for homology against the NCBI gene bank nucleotide sequences using the NCBI online package BLASTN 2.2.23 that analyses homology of the query sequence with the gene bank nucleotide sequences (Zhang *et al*., 2000).

A total of 125 gene bank sequences were found to have a 99% to 100% identity with the query sequence. Most of these sequences were from eukaryotic cloned DNA segments in *E. coli* hosts. Interestingly, the query sequence also matched 100% to the genomic DNA sequence of *E. coli* strain DH10 B (GenBank no. CP000948.1). Since DH10B cells were used as hosts for the generation of these pGEMKA6 clones (clones 3 and clones 5, see above for detail) and this foreign DNA sequence does not exist in the parental cloning vector pGEM-T Easy, it pointed to the conclusion that a recombination of DH10 B genomic DNA with pGEMKA6 DNA had occurred. The query sequence was further compared with DH10B genomic DNA sequence to find out which region of DH10B genome is homologous with the query sequence. The 433 bp query sequence matched exactly to the first 433 bp of the insertion element IS10R of DH10B (Durfee *et al.*, 2008). Alignment of the two sequences is shown in Figure 4-21.

Figure 4-21 Alignment of the 433 bp transposed DNA in PGEMKA6 with the IS10R insertion element of DH10B: The cloning of *cps2*I was successful after a number of attempts; however, it resulted in the deletion of the previously cloned gene *cps2*H. After sequencing, it was found that a 433 bp DNA segment had got inserted into pGEMKA6 in place of *cps2*H leading to the deletion of *cps2*H. When this 433 bp sequence was compared with NCBI gene bank sequences for homology, among the 125 matched sequences was the genomic DNA sequence of DH10B (the *E. coli* host used in cloning of *cps2*I) that had a 100% homology with the query sequence. On further analyses of the query sequence for homology with the DH10B genome, it was clear that the 433 bp sequence had a 100% homology with the first 433 bp of IS10R insertion element of DH10B. This led to the conclusion that transposition of IS10R from DH10B genomic DNA was involved in the deletion of *cps2*H from the construct pGEMKA5.

IS10R is a prokaryotic insertion element that is found in the genomic DNA of *E. coli* DH10B (Durfee *et al.*, 2008). However, its occurrence in eukaryotes has not been reported (Kovařík *et al.*, 2001). This suggests that the occurrence of this sequence in the remaining 124 matched gene bank eukaryotic sequences might be the result of IS10R transposition from the *E. coli* hosts to these cloned sequences (Kovařík *et al.*, 2001). The insertion element IS10R is the right side of the bacterial composite transposon tn10 (Kleckner, 1981). Tn10 transposon has a total length of 9147 bp (Figure 4-22; Chalmers *et al.*, 2000). The transposon caries two insertion elements IS10L and IS10R at its ends each of 1329 bp (Chalmers *et al.*, 2000). The central region of Tn10 carries a number of genes (Kleckner *et al.*, 1975, Jorgensen & Reznikoff, 1979), of which the tetracycline resistance genes make the transposon an important marker tool for the

generation of various *E.coli* host cells (Kleckner *et al.*, 1977). The transposition functions of Tn10 predominantly lie in IS10R which carries its own transposase gene and thus it can move independently (Foster *et al.*, 1981). On the other hand, IS10L is defective and is rarely able to transpose (Foster *et al.*, 1981).

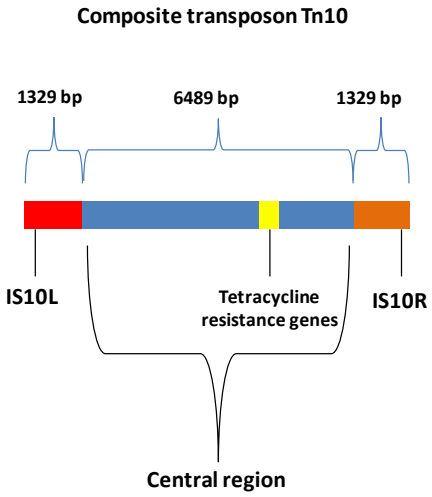


Figure 4-22 Schematic diagram of Tn10 transposon (adapted from Chalmers *et al.*, 2000)**:** Total length of Tn10 transposon is 9147 bp. The central region of the transposon carries a number of genes, most important of which are the tetracycline resistance genes. The central region is flanked by two insertion elements IS10L and IS10R. Each of the insertion elements is 1329 bp in length.

Tn10 transposon has been used extensively in the creation of various bacterial host strains including DH10B (Kleckner *et al.*, 1977). After development of the bacterial strain, tn10 transposon can be removed from the bacterium genome (Bochner *et al.*, 1980). However, parts of the transposon might still remain in the genome after its removal as was the case with DH10B where only the insertion element IS10R was found to be present in the genomic DNA and not the whole Tn10 transposon (Kovařík *et al.*, 2001) . Contamination of eukaryotic cloned DNA sequences by IS10R transposition from DH10B genome has been reported previously (Kovařík *et al.*, 2001). However, deletion of a cloned gene by IS10R transposition has not been reported. A schematic diagram of the insertion of 433 bp IS10R fragment into pGEMKA6 and deletion of *cps2*H from the vector is given in Figure 4-23.

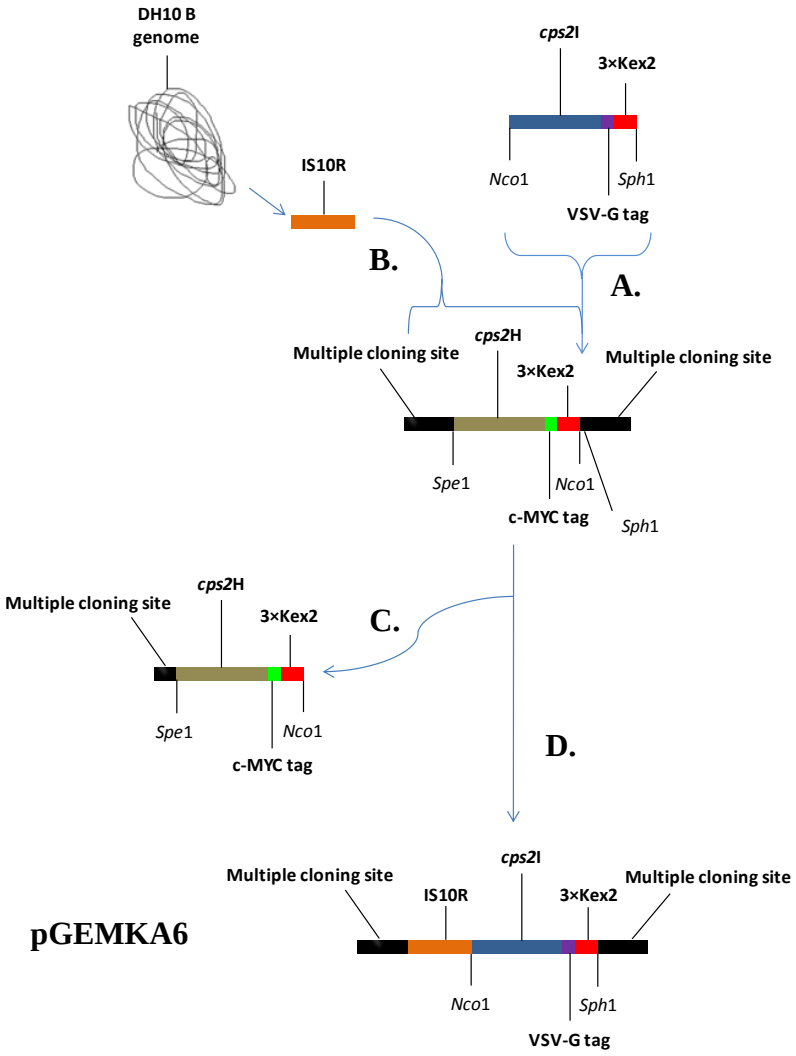


Figure 4-23 Transposition of partial sequence (433 bp) of the insertion element IS10R from DH10B genomic DNA to pGEMKA6: (**A**) After a number of attempts, the type 2 polysaccharide gene *cps2*I was finally cloned into *Nco*1/*Sph*1 sites of pGEMKA5 to generate pGEMKA6. (**B**) However, during this cloning step a 433 bp fragment of the IS10R insertion element from DH10B host genome transposed into the new vector pGEMKA6. (**C**) This 433 bp DNA fragment consistently caused the deletion of the previously cloned gene *cps2*H. (**D**) As a result the newly generated vector pGEMKA6 contained the *cps2*I gene and the 433 bp fragment of IS10R upstream of start codon of *cps2*I in place of *cps2*H. The sequence was found to be present in both of the pGEMKA6 sequenced clones (clone 3 and clone 5).

IS10 transposition has been described a rather rare event (Charlebois, 1999). In contrast, in the present instance the incidence was found to be occurring at a higher rate as IS10R partial sequence was found in both of the pGEMKA6 clones (clone 3 and clone 5) that were sequenced. The question now arises why the rate of IS10R transposition is so high in case of cloning the pneumococcal genes. Difficulties in cloning of *Streptococcus pneumoniae* genes have been previously reported (Stassi & Lacks, 1982). These difficulties have been attributed to the occurrence of strong promoter activity arising

from the cloned pneumococcal sequences (Chen & Morrison, 1987) and lethal/toxic effects of the cloned pneumococcal gene products on the *E. coli* host (Prats *et al.*, 1985). It could be inferred that the pneumococcal genes *cps2*H and *cps2*I when cloned end to end in *E. coli* have lethal/toxic effects on the host. To cope with the lethal/toxic effects, the *E. coli* genome might become more plastic and the rate of IS10R transposition might increase leading to consistent deletion of cps2H. Increased rate of IS10R transposition under stress conditions has been reported previously (Skaliter *et al.*, 1992). Stress in general activates transposon movement (McClintock, 1984).

4.4.1.3 Cloning of *cps2*J with strategy-1

After the deletion of *cps2*H during cloning of *cps*2I was confirmed, further investigations were carried out to see whether cloning of *cps2*J in frame with *cps2*I would be possible or it will also lead to difficulties. For this purpose, *cps2*J was first amplified from *Streptococcus pneumoniae* strain D39 genomic DNA using the forward primer KA041 and the mini reverse primer KA043. The resulting PCR product was re-amplified with forward primer KA041 and the full length reverse primer KA042. The *Apa1* site at 3′ end of the gene was introduced for cloning into pGEMKA6 (pGEMKA6 does not contain a *pst*1 site) and *Pst*1 site was introduced for subsequent cloning into the binary vector pKAM1. Figure 4-15E shows a diagram of the amplified *cps2*J PCR product.

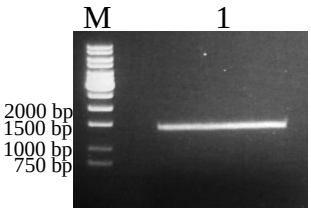


Figure 4-24 PCR amplification of *cps2*J: *cps2*J was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA. The gene was first amplified with forward primer KA041 (with a *Sph*1 site at 5′ end) and short reverse primer KA043. The PCR product was re-amplified with forward primer KA041 and the full length reverse primer KA042 (carrying a strep II tag, stop codon, *Pst*1 and *Apa*1 sites) to get the full length (1458 bp) *cps2*J fusion gene product (lane 1). Fermentas GeneRuler 1kb ladder (lane M) was used for size comparison.

*cps2*J fusion gene and the construct pGEMKA6 (section 4.4.1.2) were restricted with *Sph*1/*Apa*1 restriction enzymes. The digested *cps2*J fusion gene was cloned into the *Sph*1/*Apa*1 sites of pGEMKA6 next to *cps2*I to generate pGEMKA7. Figure 4-25 shows

a diagram of pGEMKA7. As can be seen in Figure 4-15, the originally proposed construct would have contained all of the three set-B type 2 polysaccharide genes (*cps2*H, *cps2*I and *cps2*J). However, as a result of transposition of the 433 bp IS10R fragment, *cps2*H has been deleted from the vector (section 4.4.1.2). Instead the vector contained the 433 bp IS10R fragment and the last two set-B genes (*cps2*I and *cps2*J, Figure 4-25).

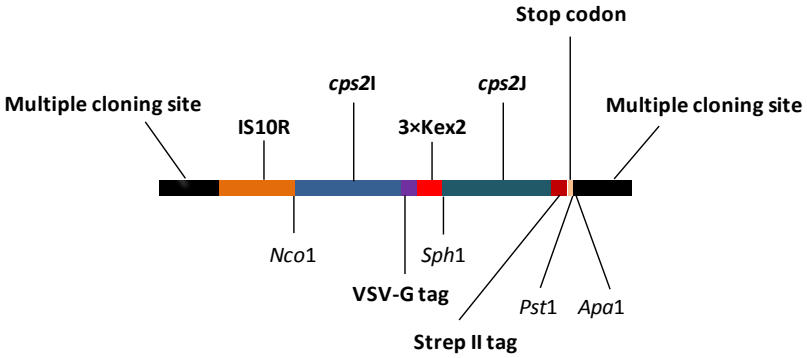


Figure 4-25 Diagram of pGEMKA7: *cps2*J fusion gene was cloned into the *Sph*1/*Apa*1 sites of pGEMKA6 (Figure 4-23) to generate pGEMKA7. The vector differs from the originally proposed diagram of pGEMKA7 (section 4.4.1, Figure 4-15) as the vector has lost the cloned gene *cps2*H and instead a 433 bp IS10R segment from DH10B genome has been inserted replacing *cps2*H. The vector now contains the last two set-B genes (*cps2*I and *cps2*J) instead of the three type 2 polysaccharide genes (*cps2*H, *cps2*I and *cps2*J).

pGMEKA7 was used to transform DH10B cells. Transformed cells were selected on LB agar plates containing 100 µg/ml ampicillin. To confirm *cps2*J integration and also to confirm that *cps2*I was intact, pGEMKA7 mini prep samples were digested with *Nco*1/*Pst*1 restriction enzymes as the two genes are flanked by these two restriction sites. The digestion was predicted to result in the release of a 2691 bp fragment relating to the two genes. The products of digestion were analysed on 1.5 % gel. Figure 4-26 shows results of the restriction digestion. A fragment of correct size (2691 bp, indicated by an arrow) was observed in most digested samples. Clone 1 was selected for sequencing to confirm the correct sequences of the two genes. Sequencing was carried out with the commercially available forward primer SP6, *cps2*I primer KA040, forward primer KA045, *cps2*J forward PCR primer KA041 and reverse primer KA032. Sequences of both the genes (*cps2*I and *cps2*J) were correct. The transposed 433 bp fragment of IS10R that was inserted during cloning of *cps2*I was still present in the vector.

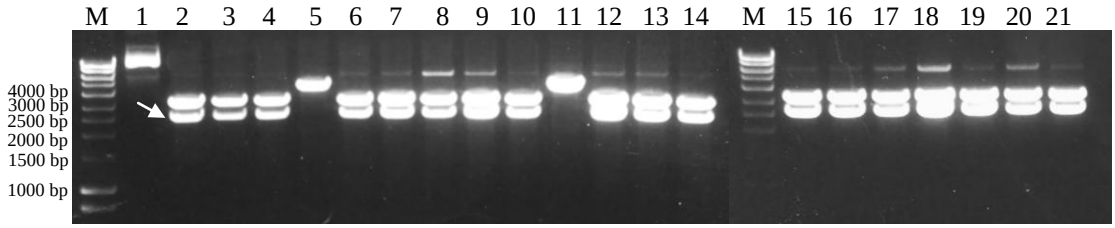


Figure 4-26 Restriction digestion of pGEMKA7 with *Nco*1/*Pst*1 to confirm *cps2*J integration and intactness of *cps2*I: Twenty mini prep samples (lane 2-21) were digested with *Nco*1/*Pst*1 restriction enzymes. The restriction digestion resulted in the generation of a 2691 bp fragment (indicated by an arrow) corresponding to the combined length of *cps2*I and *cps2*J. The band was observed in all samples except sample 4 (lane 5) and sample 10 (lane 11). Lane 1 contained the non-restricted clone 1 DNA sample. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

Once genes *cps2*I and *cps2*J had been cloned adjacent to each other, efforts were made to overcome the cloning artefacts that did not allow putting all the three set-B genes (*cps2*H, *cps2*I and *cps2*J) together in the same cassette. In one attempt, the DNA segment consisting of *cps2*I and *cps2*J fusion genes was excised from pGEMKA7 and cloned into the *Nco*1/*Apa*1 sites of pGEMKA5 in frame with *cps2*H (section 4.4.1.1). The resultant ligation was used to transform DH10B. After selection on LB agar medium containg 100 µg/ml ampicillin, plasmid miniprep DNA from resulting clones were restricted with *Spe*1/*Apa*1 restriction enzymes. Results of this digestion are shown in Figure 4-27. An expected band of 3927bp was not observed in any of the samples. Furthermore, it was reasoned that if these samples contained the self-ligated pGEMKA5 vector then they should have produced a band specific to the cloned genes *cps2*H as *cps2*H is flanked by these two restriction sites in the parental vector pGEMKA5. It again pointed to the deletion of *cps2*H from the parental vector.

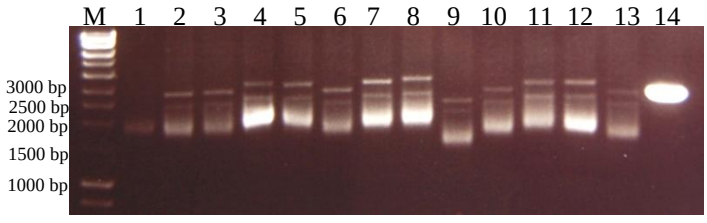


Figure 4-27 Restriction digestion analysis of *cps2*I and *cps2*J integration into pGEMKA5: Twenty four mini prep samples (only 13 samples shown here, lane 2-lane 14; lane 1 contained uncut sample 1) were digested with *Spe*1/*Apa*1 restriction enzymes. The digestion should have resulted in the release of a 3927 bp fragment corresponding to the combined length of the three genes (*cps2*H, *cps2*I and *cps2*J). However, no such band was seen. In case of self-ligation of the original vector, the digestion should have released a band (~1248 bp) specific to *cps2*H fusion gene that is already present in the original vector and flanked by these restriction sites. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

In another experiment, the DNA segment consisting of the two fusion genes (*cps2*I and *cps2*J) was excised from pGEMKA7 and cloned into the *Nco*1/*Apa*1 sites of pGEMKA1 (section 3.3.1) that did not contain *cps2*H. The resultant vector was used to transform MegaX DH10B electrocompetent cells. Transformed cells were selected on LB agar plates containing 100 µg/ml ampicillin. DNA miniprep and restriction digestion analysis of a number of clones was carried out for confirmation of these two genes in the target vector. This was carried out using a number of available unique restriction sites flanking the two genes. No specific products were observed in any case (data not shown).

4.4.1.4 *In vitro* ligation of type 2 polysaccharide set-B genes (*cps2*H, *cps2*I and *cps2*J) and cloning as a single DNA fragment

To bypass the multiple steps involved in cloning of type 2 polysaccharide set-B genes (*cps2*H, *cps2*I and *cps2*J**)** sequentially, an *in vitro* ligation strategy was adapted for the three genes. To achieve this, *cps2*H was amplified from *Streptococcus pneumoniae* strain D39 genomic DNA as described in section 4.4.1.1. *cps2*H PCR product was restricted with *Nco*1 restriction enzyme. Similarly, the DNA segment consisting of *cps2*I and *cps2*J fusion genes was excised from pGEMKA7 (section 4.4.1.3) by restriction digestion with *Nco*1/*Apa*1. The digested *cps2*I-*cps2*J product was dephosphorylated to avoid self ligation. Ligation of digested *cps2*H PCR product and digested *cps2*I-*cps2*J product was then carried out with T4 DNA ligase. Figure 4-28A shows result of the ligation. In lane 1, band 1 shows the ligated product *cps2*H-*cps2*I-*cps2*J (3939 bp), band 2 shows the non-ligated *cps2*I-*cps2*J digested product (2691 bp) and band 3 shows the non-ligated *cps2*H digested PCR product (1248 bp). The ligation product was purified and then amplified by PCR. PCR amplification of the ligation product was carried out with *cps2*H forward PCR prime KA035 and *cps2*J full-length reverse primer KA042 to produce a 3945 bp product (lane 1 in Figure 4-28B).

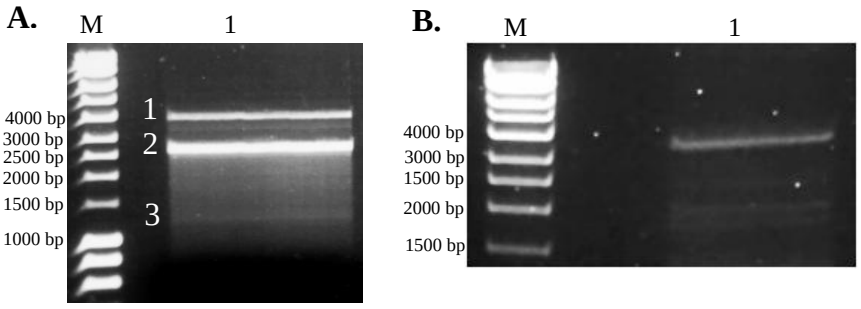


Figure 4-28 Ligation of *cps2*H PCR product with *cps2*I-*cps2*J fusion genes: A. *cps2*H PCR product was ligated with *cps2*I-*cps2*J fusion genes. In lane 1, band 1 shows the ligated product *cps2*H-*cps2*I-*cps2*J (3939 bp), band 2 shows the non-ligated *cps2*I-*cps2*J product (2691 bp) and band 3 shows the non-ligated *cps2*H PCR product (1248 bp). **B.** The ligated product was amplified by PCR using *cps2*H forward PCR prime KA035 and full-length *cps2*J reverse primer KA042 to generate a 3945 bp product (lane 1). BIOLINE hyper ladder 1 (lane M) was used for size comparison.

The PCR amplified product (*cps2*H-*cps2*I-*cps2*J) was restricted with *Spe*1/*Pst*1 restriction enzymes. The digested product was cloned into the *Spe*1/*Pst*1 sites of plant binary vector pKAM1 (section 3.3.3). The resultant vector was used to transform DH10B cells. Transformed cells were selected on LB agar plates containing 25 µg/ml kanamycin. DNA mini prep samples of a number of clones were digested with *Spe*1/*Pst*1 restriction enzymes. No digestion products related to the DNA segment consisting of the three genes *(cps2*H-*cps2*I-*cps2*J) were observed in the mini prep samples (figure not shown).

A similar experiment was carried out by adapting the much simpler TA cloning strategy avoiding any restriction enzyme sites. *In vitro* ligation of *cps2*H PCR product and *cps2*I-*cps2*J was carried out as above. Enough ligation product was obtained eliminating the need for PCR amplification. The ligation product was made suitable for TA cloning by A-tailing of the ligation product with taq polymerase after treatment with Klenow DNA polymerase I large fragment to remove the 3′ overhangs present from the *Apa*1 digestion. The A-tailed *cps2*H-*cps2*I-*cps2*J product was cloned into the pGEM-T Easy vector. The resulting vector was used to transform DH10B. Transformed cells were selected on LB agar plates containing 100 µg/ml ampicillin. DNA mini preps were restricted with *Spe*1/*Pst*1 restriction enzymes and the products of digestion were analyzed on 1.5% agarose gel. Digestion did not result in the release of any products related to the DNA segment consisting of the three genes (*cps2*H-*cps2*I-*cps2*J) (figure not shown).

4.4.2 Strategy-2 for cloning type 2 polysaccharide set-B genes (*cps2*I, *cps2*J and *cps2*H)

A number of attempts were carried out to assemble the set-B type 2 polysaccharide genes in the order 5′*Spe*1-*cps2*H-c-MYC tag-3Kex2-*Nco*1-*cps2*I-VSV-G tag-3Kex2-*Sph*1-*cps2*J-Strep 2 tag-stop-*Pst*1-*Apa*13′ (strategy-1, section 4.4.1). An alternate strategy was designed for cloning these genes. This involved rearranging the order of the three genes to be cloned. This rearrangement also eliminated the use of *Nco*1 restriction enzyme site that could be a potential target site for the 433 bp IS10R fragment (section 4.4.1.2). The revised order of the three genes was; 5′*Spe*1-*cps2*I-VSV-G tag-3Kex2-*Sph*1-*Cps2*J-Strep 2 tag-3Kex2- *Pst*1- *cps2*H-c-MYC tag-stop-*Pst*1-*Apa*13′. Cloning of the three set-B genes in the above order is described in the following sections.

4.4.2.1 Cloning of *cps2*I and *cps2*J (strategy-2)

The initial strategy for expressing all the seven type 2 pneumococcal genes in a single tobacco plant involved the production of two independent homozygous transgenic tobacco lines; one expressing the set-A polysaccharide genes (c*ps2*ET, c*ps2*E, *cps2*F and *cps2*G) and another expressing the set-B polysaccharide genes (*cps2*I, *cps2*J and *cps2*H). The next step was then to cross the two transgenic tobacco lines to develop a hybrid transgenic line that expresses all the seven type 2 pneumococcal genes responsible for type 2 polysaccharide production. However, at the time the second strategy for cloning set-B type 2 polysaccharide genes was being carried out, transgenic tobacco lines expressing set-A genes had already been produced. The time constraints did not allow developing homozygous transgenic tobacco lines expressing set-B genes and then crossing the two lines to assemble all the seven genes in a single plant. As a result, a new strategy was devised which involved double transformation of the already developed set- A transgenic lines (expressing set-A genes) with the set-B genes. To achieve this, another plant binary vector was needed for cloning of set-B genes that carries a plant selection marker other than the kanamycin resistance gene as the gene is already expressed in the set-A transgenic lines and selection of the double transformants would not be possible with the same marker gene. Hence, the plant binary vector pCAMBIA1302 (section 2.1.5.4) that carries a hygromycin plant selection gene was

selected for cloning of set-B genes. Now the issue with pCAMBIA1302 was that it contained a *Spe*1 site in its backbone and the *Spe*1site was necessary to be added to the 5′ end of the DNA fragment consisting of the three set-B genes to enable the subsequent cloning of these genes in frame with the PR1b signal. One possibility was to clone the DNA segment consisting of set-B genes into pKAM1 or pKAM2 (section 3.3.3 and 4.3.1 respectively) that contained the remaining elements of the expression cassette (i.e. CaMV35S promoter, PR1b endoplasmic reticulum targeting signal and the *rbc*S terminator) and then excise the whole expression cassette at the *EcoR*1/*Hind*III sites and clone into pCAMBIA1302. However, the issue now was that the *Hind*III site was also present in *cps2*I and *cps2*H sequences so this approach was not practicable. To solve these issues the following step by step cloning methodology was adopted.

4.4.2.1.1 Cloning of CaMV35S promoter, PR1b signal and *rbc*S terminator sequences in pKAMBA1302

The insulin expression cassette was excised from pKAM1 (section 3.3.3) and cloned into *EcoR*1/*Hind* III sites of pCAMBIA1302 (section 2.1.5.4) to generate pKAM5 (Figure 4-29). This resulted in the incorporation of the desired sequences of CaMV35S promoter, PR1b signal and *rbc*S terminator into pKAM5 for the generation of set-B genes expression cassette. pKAM5 contained unique *EcoR*1/*Pst*1 sites (Figure 4-29) that were compatible with all the three type2 polysaccharide set-B genes to be cloned subsequently. pKAM5 was used to transform DH10B cells. Transformed cells were selected on LB agar plates containing 25 µg/ml kanamycin. DNA mini prep samples were digested with *EcoR*1/*Pst*1 restriction enzymes and analysed on 1.5% agarose gel to confirm integration of the insulin gene expression cassette (Figure 4-30). A DNA fragment of expected size (1157 bp) corresponding to the inserted DNA was detected.

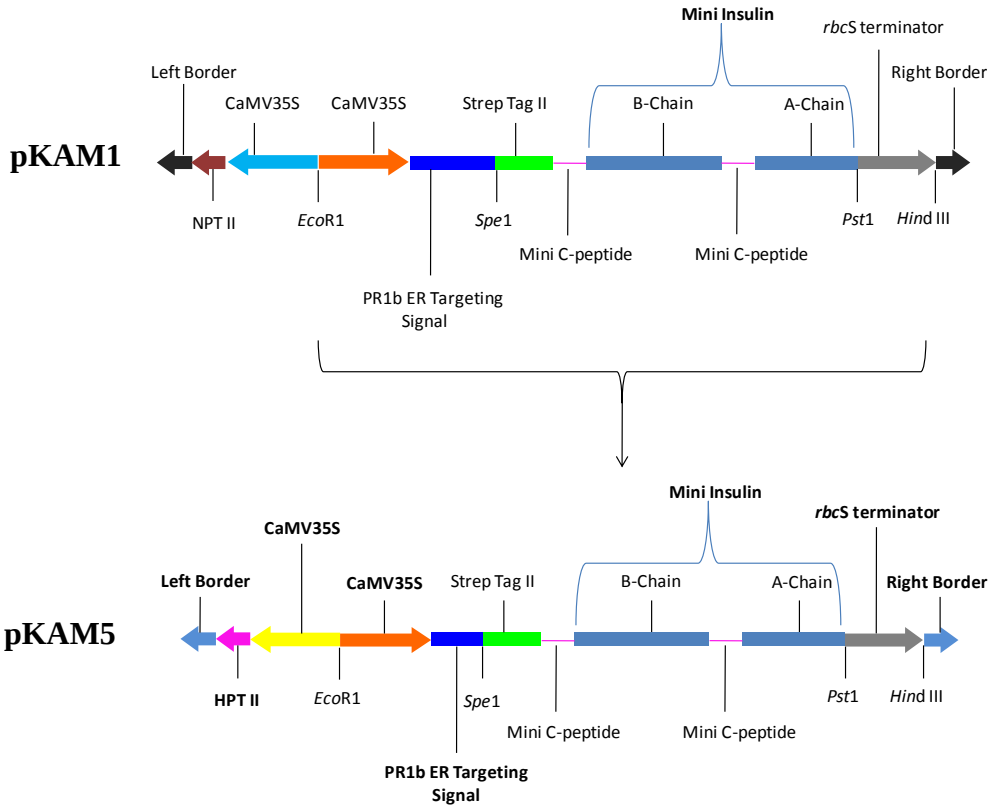


Figure 4-29 Manipulation of pCAMBIA1302 to incorporate CaMV35S promoter, PR1b signal and *rbc*S terminator: The insulin expression cassette was excised from pKAM1 (section 3.3.3) and cloned into the *EcoR*1/*Hind* III sites of pCAMBIA1302 (section 2.1.5.4) to generate pKAM5. This cloning incorporated the required sequences of CaMV35S promoter, PR1b signal and *rbc*S terminator.

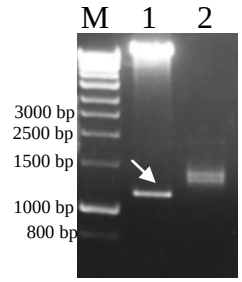


Figure 4-30 Restriction digestion of pKAM5 with *EcoR1*/*Pst1* restriction enzymes to confirm integration of insulin expression cassette: Two mini prep samples (lane1 and lane 2) were digested with *EcoR*1/*Pst*1 restriction enzymes. A fragment of correct size (1157 bp, indicated by an arrow) corresponding to the CaMV35S promoter, PR1b signal and insulin fusion gene was observed in lane 1 (sample 1) but not in lane 2 (sample 2). BIOLINE hyper ladder 1 (lane M) was used for size comparison.

4.4.2.1.2 Cloning of *cps2*I and *cps2*J into pKAM2

The DNA segment consisting of *cps2*I and *cps2*J fusion genes was PCR amplified from pGEMKA7 plasmid DNA (section 4.4.1.3) with the revised *cps2*I forward primer KA046 and the revised *cps2*J full- length reverse primer KA047 A diagram of the

modified *cps2*I and *cps2*J DNA fragment is shown in Figure 4-32A. Figure 4-31 shows the results of the amplified *cps2*I and *cps2*J fusion genes PCR product.

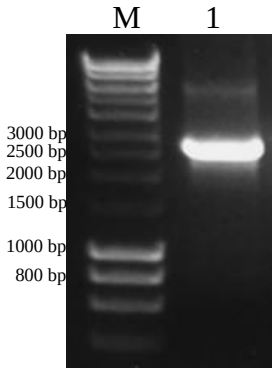


Figure 4-31 PCR amplification of *cps2*I and *cps2*J fusion genes using pGEMKA7 DNA as template: DNA segment consisting of *cps2*I and *cps2*J fusion genes was amplified using pGEMKA7 DNA as template. Amplification was carried out with the revised *cps2*I forward primer KA046 (with a *Spe*1 site at 5′ end) and the revised *cps2*J full- length reverse primer KA047 (carrying a strep-tag II, 3×Kex2 cleavage sites and a *Pst*1 site). This amplification introduced the desirable restriction sites at the 5′ and 3′ ends of the fusion genes product. The amplified product (2733 bp) can be seen in lane 1. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

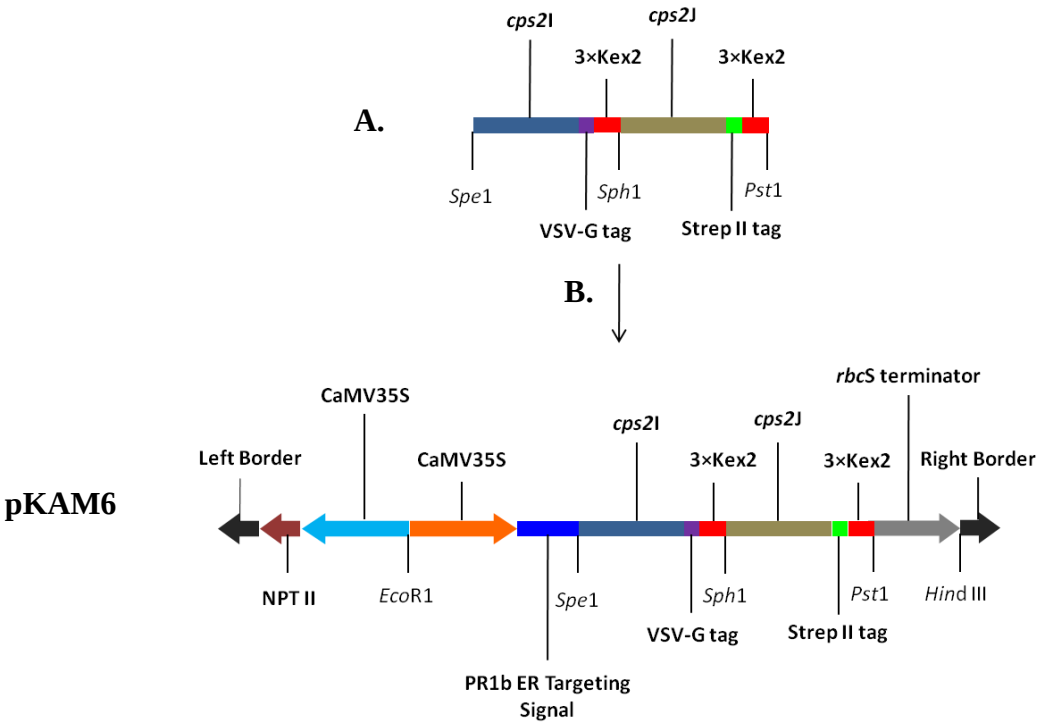


Figure 4-32 Cloning of *cps2*I and *cps2*J with strategy-2: A. The DNA segment consisting of *cps2*I and *cps2*J was amplified from vector pGEMKA7 (section 4.4.1.3) using *cps2*I forward primer KA046 (with a *Spe*1 site at 5′ end) and *cps2*J full- length reverse primer KA047 (carrying a strep-tag II, 3×Kex2 cleavage

sites and a *Pst*1 site). **B.** The amplified product was cloned into the *Spe*1/ *Pst*1 sites of pKAM2 (section 4.3.1) to generate pKAM6.

The amplified PCR product consisting of the two genes and the construct pKAM2 (section 4.3.1) were digested with *Spe*1/*Pst*1 restriction enzymes. Digested PCR product was then cloned into *Spe*1/*Pst*1 sites of pKAM2 to generate pKAM6 (Figure 4-32B). pKAM6 was used to transform DH10B cells. Transformed cells were selected on LB agar plates containing 25 µg/ml kanamycin. DNA minipreps were digested with *Spe*1/*Pst*1 restriction enzymes for confirmation of the two genes integration. Figure 4-33 shows results of this digestion. A DNA fragment of correct size (2727 bp) corresponding to the two genes combined product can be seen in lane 1 (clone 7), lane 2 (clone 8) and lane 3 (clone 9). Lane 4 (clone 10) might contain the self-ligated vector. Clone 7 (lane 1) was selected for further manipulations.

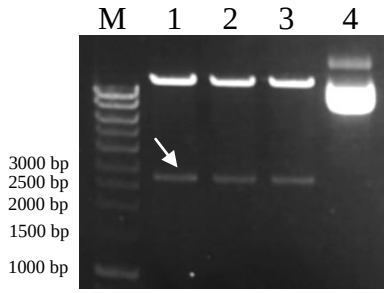


Figure 4-33 Restriction digestion of pKAM6 for confirmation of *cps2*I and *cps2*J integration: DNA miniprep samples were restricted with *Spe*1/*Pst*1 restriction enzymes and analysed on 1.5 % agarose gel. A band of correct size (2727 bp, indicated by an arrow) can be seen in lane 1(clone 7), lane 2 (clone 8) and lane 3 (clone 9). Lane 4 (clone 10) might contain the self-ligated vector. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

4.4.2.1.3 Cloning of *cps2*I and *cps2*J into pKAM5

To clone the DNA segment consisting of *cps2*I and *cps2*J from pKAM6 (section 4.4.2.1.2) into pKAM5 (section 4.4.2.1.1), the *Spe*1/*Pst*1 restriction sites flanking these genes in pKAM6 could not be used as *Spe*1 site is also present in the construct pKAM5 so *EcoR*1/*Pst*1 restriction sites were used. The *EcoR*1 site is located at the 5′ end of CaMV35S promoter in both of the vectors pKAM6 and pKAM5 (Figure 4-32). So the DNA segment consisting of CaMV35S promoter, PR1b signal, *cps2*I and *cps2*J was excised from pKAM6 and cloned into the *EcoR*1/*Pst*1 sites of pKAM5 to generate pKAM7 (Figure 4-34).

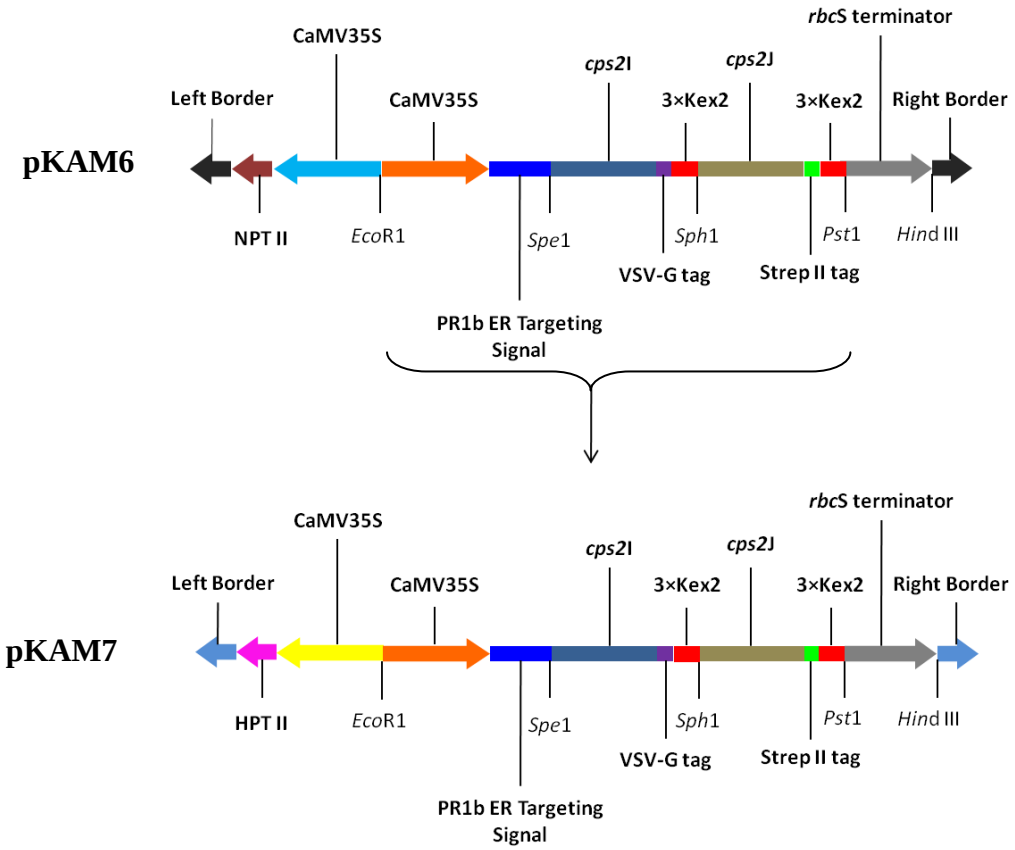


Figure 4-34 Cloning of *cps2*I and *cps2*J into pKAM5: pKAM6 was cut at the *EcoR*1/ *Pst*1 sites to release the DNA fragment consisting of CaMV35S promoter, PR1b signal and the two cloned genes (*cps2*I and *cps2*J). The excised DNA fragment was cloned into the *EcoR*1/ *Pst*1 sites of pKAM5 (section 4.4.2.1.1) to generate pKAM7.

pKAM7 was used to transform DH10B. Transformed cells were selected on LB agar plates containing 25 µg/ml kanamycin. Mini prep samples were digested with *EcoR*1/*Pst*1 restriction enzymes to confirm integration of the insert into the vector. Figure 4-35 shows results of this digestion. A DNA fragment of expected size (3683 bp) corresponding to CaMV35S promoter, PR1b signal, *cps2*I and *cps2*J was observed in most of the samples (lane 1, lane 2, lane 3, lane 4, lane 5, lane 6, lane 7, lane 8 and lane 10). Clone 3(lane 3) was selected for further cloning.

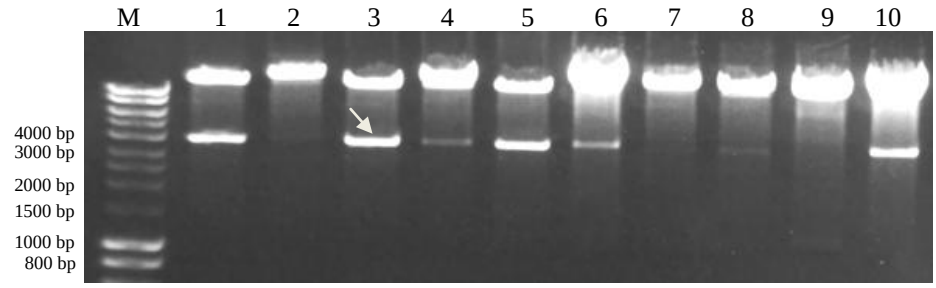


Figure 4-35 Restriction digestion of pKAM7 with *EcoR*1/*Pst*1 restriction enzymes for confirmation of *cps2*I and *cps2*J integration: Ten mini prep DNA samples (lane 1- lane 10) were restricted with *EcoR*1/*Pst*1 restriction enzymes and analysed on 1.5 % agarose gel. A band of correct size (3683 bp, indicated by an arrow) corresponding to the DNA segment consisting of CaMV35S promoter, PR1b signal, *cps2*I and *cps2*J was observed in lane 1, lane 2, lane 3, lane 4, lane 5, lane 6, lane 7, lane 8 and lane 10. Lane 9 might contain the self-ligated vector. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

4.4.2.2 Cloning of *cps2*H with strategy-2 and creation of pKAM8

*cps2*H, the last of the three set-B type 2 polysaccharide genes, was amplified from pGEMKA5 (section 4.4.1.1) with revised forward primer KA048 (with a *Pst*1 site at 5′ end) and revised full-length reverse primer KA049 (carrying a c-Myc tag, stop codon, a *Pst*1 site and an *Apa*1 site). This amplification attached the required sequences and restriction sites to *cps2*H for cloning with strategy-2. A diagram of the modified *cps2*H fusion gene is shown in Figure 4-37A. *cps2*H PCR product was analysed on 1.5% gel. The results are shown in Figure 4-36. A DNA fragment of correct size (1218 bp) can be seen in lane 1.

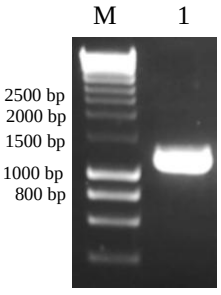


Figure 4-36 PCR amplification of *cps2*H for cloning with strategy-2: *cps2*H was amplified from pGEMKA5 (section 4.4.1.1) using revised forward primer KA048 (with a *Pst*1 site at 5′ end) and full-length reverse primer KA049 (carrying a c-Myc tag, stop codon, a *Pst*1 site and an *Apa*1 site). A band of 1218 bp corresponding to modified *cps2*H fusion gene product can be seen in lane 1. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

*cps2*H PCR product and the construct pKAM7 (section 4.4.2.1.3) were digested with *Pst*1 restriction enzyme. The digested insert was cloned into the *Pst*1site of pKAM7 to generate the construct pKAM8 (Figure 4-37B). Like expression cassette-A (section 4.3.2.4), the type 2 pneumococcal polysaccharide expression cassette-B is under the control of the constitutive CaMV35S promoter for expression in tobacco leaves. The N-terminal PR1b signal has been introduced for targeting to endoplasmic reticulum and for ultimate excretion of the three proteins to the apoplast after the Kex2 protease processing in the Golgi apparatus. The *rbc*S terminator has been attached to the 3′ end of the polygene for transcription termination.

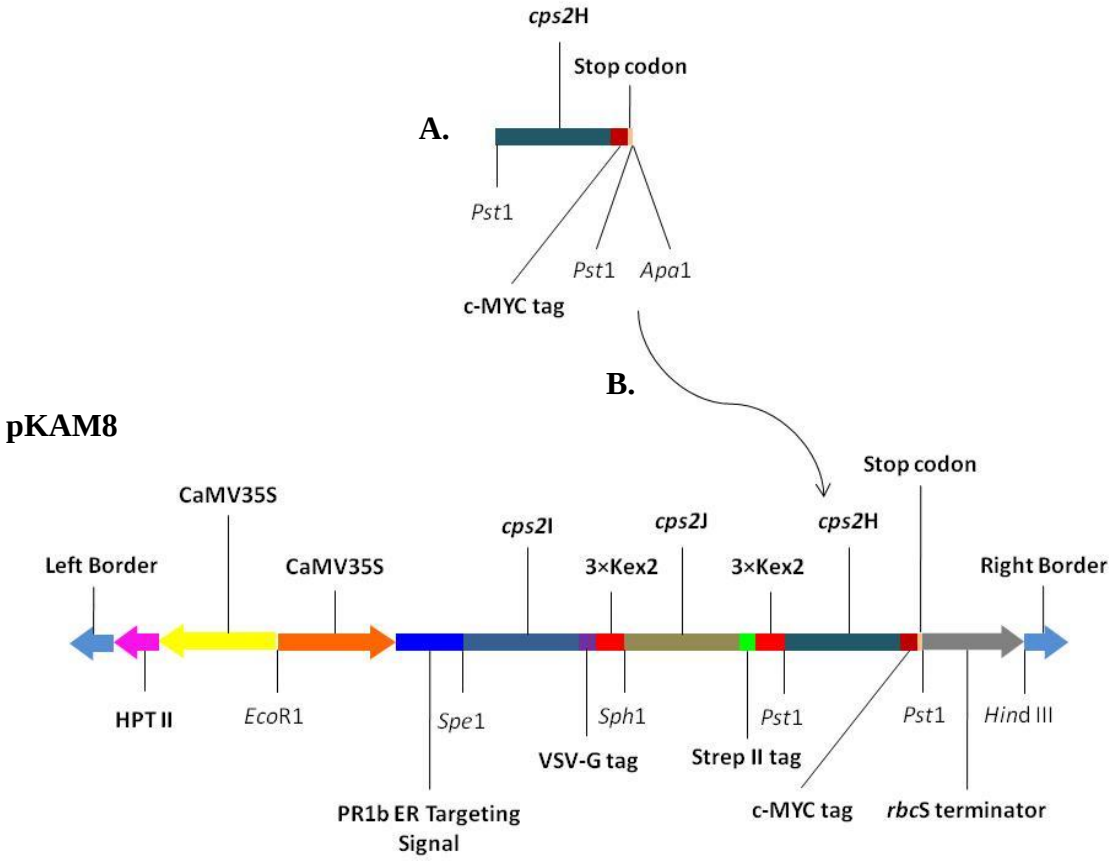


Figure 4-37 Cloning of *cps2*H with strategy-2: A. *cps2*H was amplified from pGEMKA5 (section 4.4.1.1) using revised forward primer KA048 (with a *Pst*1 site at 5′ end) and full-length reverse primer KA049 (carrying a c-Myc tag, stop codon, a *Pst*1 site and an *Apa*1 site). This amplification attached the required sequences to *cps2*H for cloning with strategy-2. **B.** Modified *cps2*H fusion gene was cloned into the *Pst*1 site of pKAM7 (section 4.4.2.1.3) in frame with the previously cloned genes *cps2*I and *cps2*J to generate pKAM8 (the final binary destination vector). In pKAM8, the type 2 polysaccharide expression cassette-B is driven by CaMV35S promoter, The N-terminal PR1b signal serves to target the set-B polysaccharide polyprotein to endoplasmic reticulum, from where the polyprotein would be passed onto

Golgi apparatus for the Kex2 protease processing to generate three processed proteins (*cps2*I, *cps2*J and *cps2*H). Ultimately the mature proteins would be excreted to the apoplast. The *rbc*S terminator sequence is attached after the stop codon of the polygene for transcription termination.

pKAM8 was used to transform DH10B cells. Transformed cells were selected on LB agar plates containing 25 µg/ml kanamycin. DNA minipreps were digested with *Pst*1 restriction enzyme for confirmation of *cps2*H fusion gene integration. Figure 4-38 shows results of the restriction digestion. A 1206 bp DNA fragment corresponding to *cps2*H fusion gene was observed in all mini prep samples (lane 1-lane8).

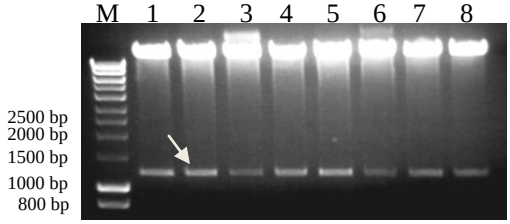


Figure 4-38 Restriction digestion of pKAM8 with *Pst*1 restriction enzyme for confirmation of *cps2*H integration: Eight mini prep samples (lane 1- lane 8) were restricted with *Pst1* restriction enzyme and analysed on 1.5% agarose gel. A band of correct size (1206 bp, indicated by an arrow) corresponding to *cps2*H fusion gene was observed in all samples (lane 1-lane 8). BIOLINE hyper ladder 1 (lane M) was used for size comparison.

Since *cps2*H fusion gene is flanked by two *Pst*1 sites, it can be integrated in the correct orientation or the reverse orientation. Therefore, all the positive pKAM8 clones from *Pst*1 digestion (Figure 4-38) were further analysed by restriction digestion with *Hind* III restriction enzyme. There are three *Hind* III sites in pKAM8; the first one is located 874 bp downstream of start codon of *cps2*I, second site is located 1164 bp downstream of start codon of *cps2*H and the third site is located at the 3′ end of the expression cassette after the *rbc*S terminator. If the insert had been integrated in the correct orientation then restriction digestion with *Hind* III should release two fragments from the vector; a 3017 bp fragment corresponding to the 1st and 2nd *Hind* III sites and a 679 bp fragment corresponding to the 2nd and 3rd *Hind* III sites. However, if the insert DNA has been integrated in the reverse orientation then the sizes of the two released fragments should be 1871 bp and 1825 bp respectively. Figure 4-39 shows results of the *Hind* III restriction digestion of the eight mini preps. Correct orientation of *cps2*H was observed in lane 1 (clone1) and lane 3 (clone 3). The 3017 bp fragment is shown by an asterisk and the 679 bp fragment is shown by double asterisks. The remaining samples (lane 2, lane 4, lane 5, lane 6, lane 7 and lane 8) showed the incorrectly integrated *cps2*H fusion

gene. Since the difference between the two released bands corresponding to incorrect orientation is 46 bp (see above), the two bands could not be resolved on 1.5% agarose gel. These fragments are shown by an arrow in Figure 4-39.

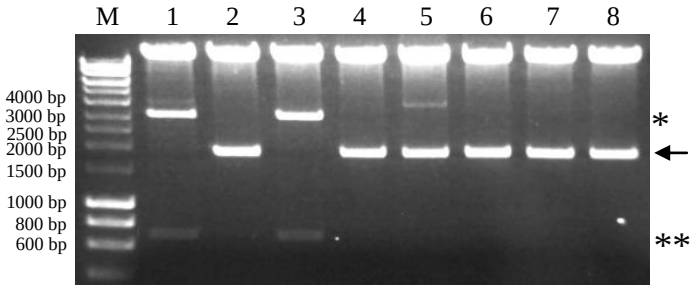


Figure 4-39 Confirmation of the orientation of *cps2*H fusion gene in pKAM8: The positive pKAM8 clones (Figure 4-38) were restricted with *Hind* III restriction enzyme to confirm orientation of *cps2*H in pKAM8 (for details see text). Correct orientation of the integrated *cps2*H was observed in lane 1 (clone1) and lane 3 (clone 3) with the release of two correct size bands (3017 bp, indicated by an asterisk and 679 bp indicated by double asterisks). The remaining samples (lane 2, lane 4, lane 5, lane 6, lane 7 and lane 8) showed two fragments related to incorrect orientation of *cps2*H fusion gene in pKAM8. The sizes of the two bands are 1875 and 1825 bp respectively. These fragments could be resolved on 1.5% agarose gel as the difference between the two bands is only 46 bp. These two bands are indicated by an arrow collectively. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

Clone 3 was selected for DNA sequencing. The whole expression cassette-B was sequenced with forward primer KA008, reverse primer KA050, *cps2*I revised forward PCR primer KA046, *cps2* I mini reverse PCR primer KA040, forward primer KA045, *cps2*J forward PCR primer KA041, *cps2*J mini reverse PCR primer KA043, forward primer KA051, *cps2*H revised forward PCR primer KA048, *cps2*H mini reverse PCR primer KA037, forward primer KA044 and reverse primer KA009. After sequence confirmation, pKAM8 midi scale plasmid DNA was isolated for *Agrobacterium*-mediated transformation of tobacco.

4.5 Tobacco transformation

4.5.1 Tobacco transformation with pKAM3 (carrying set- A type 2 polysaccharide genes)

Plant binary vector pKAM3 carrying the set-A type 2 polysaccharide genes (section 4.3.2.4) was introduced into *Agrobacterium tumefaciens* strain GV3101 by heat shock transformation. The resulting *Agrobacterium* was used to transform tobacco (*Nicotiana tabacum cv* SAMSUN) leaf disks. Regeneration of shoots (on selective MS agar medium containing 3% sucrose, 250 µg/ml cefotaxime and 100 µg/ml kanamycin) from

the wounded areas of infected leaf disks was evident 1-2 weeks post infection (Figure 4-40A). Non-infected wild type leaf disks developed no callus or shoots and dried out (Figure 4-40B). After 6-8 weeks, regenerated shoots were excised from leaf discs and transferred to MS agar rooting medium containing 3% (w/v) sucrose, 250 µg/ml cefotaxime and 100 µg/ml kanamycin with no additional hormones. Root development started after 7-12 days (Figure 4-40C). Some escape shoots were also observed that failed to develop any roots on selective rooting medium (Figure 4-40D). After root development, 34 healthy transgenic plantlets (designated as PolA1 to PolA34) were transferred to soil in pots and were allowed to grow in the green house (Figure 4-40E). Transgenic lines PolA1, PolA4, PolA11, PolA23, PolA29 and PolA33 died in the green house before maturity (due to wilting during hot weather). The remaining 28 plants reached maturity. To avoid cross-fertilization, flower buds of the transgenic lines were covered with paper bags. After maturation, seeds from transgenic lines were harvested and stored in bijou vials.

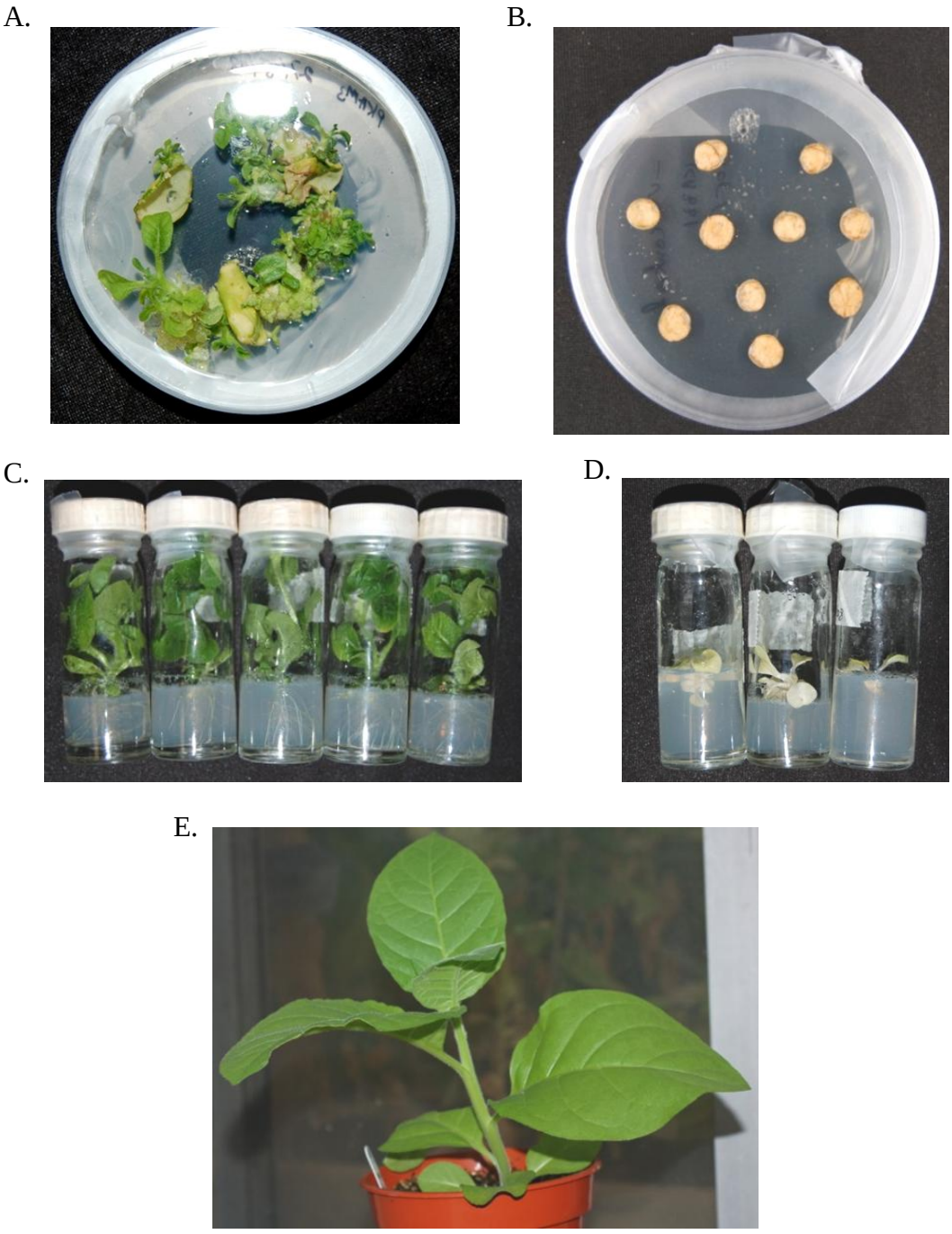


Figure 4-40 Development of transgenic tobacco plants expressing set-A pneumococcal type 2 polysaccharide genes: A. Four-week old regenerated shoots from transformed leaf disks growing on kanamycin (100 µg/ml) selective shoot medium. **B.** Non-transformed (negative control) leaf disks failed to regenerate any shoots/callus on selective shoot medium and dried; **C.** Regenerated transformed shoots developed roots on selective medium containing 100 µg/ml kanamycin. **D.** Escapes (non-transformed regenerated shoots) failed to develop any roots and became chlorotic on kanamycin selective rooting medium. **E.** Transgenic tobacco plant growing in the green house.

4.5.2 Re-transformation of set-A type 2 polysaccharide transgenic tobacco lines with pKAM8 (carrying set-B type 2 polysaccharide genes)

As mentioned in section 4.4.2.1, the original strategy of developing two independent homozygous tobacco lines (expressing set-A and set-B type 2 polysaccharide genes respectively) and then crossing the two lines to assemble all the seven type 2 polysaccharide genes in a single plant was changed due to time constraints. The revised strategy involved re-transformation of the already developed transgenic tobacco lines expressing the four set-A type 2 polysaccharide genes (section 4.5.1) with pKAM8 (section 4.4.2) carrying the three set-B type 2 polysaccharide genes. pKAM8 (section 4.4.2) has a hygromycin plant selection gene. Hence the double transformation of set-A transgenic lines (set-A lines already have a kanamycin plant selection gene) was feasible.

pKAM8 (section 4.4.2) was introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation. The resulting *Agrobacterium* was used to transform tobacco (*Nicotiana tabacum cv* SAMSUN) leaf disks. Leaf disks from set-A transgenic lines PolA7, PolA14 and PolA24 (these lines were selected for transformation on the basis of expression of the various set-A genes as determined by RT-PCR and western blotting results) were transformed with pKAM8. Wild type plant leaf disks were also transformed with pKAM8 for comparison purposes and as a backup in case the double transformation did not work. Regeneration of shoots (on selective MS agar medium containing 3% sucrose, 250 µg/ml cefotaxime and 30 µg/ml hygromycin B) from the wounded areas of infected leaf disks was evident 1-2 weeks post infection (Figure 4-41A and Figure 4-41B). The double transformed regenerated shoots grew slowly (Figure 4-41B) compared to regenerated shoots from wild type transformed leaf disks (Figure 4-41A). Non-infected wild type leaf disks developed no callus or shoots and dried out (Figure 4-41C). After 6-8 weeks, regenerated shoots were excised from leaf discs and transferred to MS agar rooting medium containing 3% (w/v) sucrose, 250 µg/ml cefotaxime and 30 µg/ml hygromycin B with no additional hormones. In addition to 30µg/ml hygromycin B, 100 µg/ml kanamycin was also added to rooting medium of double transformed regenerated shoots as a second selection marker. Root development started after 7-12 days (Figure 4-41D and Figure 4-41E). Again, the shoots from wild

type transformed leaf disks grew more quickly (Figure 4-41D) than the double transformed shoots (Figure 4-41E). Some escape shoots were also observed that failed to develop any roots on selective rooting medium (Figure 4-41F). After root development, several healthy double-transgenic plantlets and 10 mono-transgenic descendants of wild type plant (designated as PolB1 to PolB10) were transferred to soil in pots and were allowed to grow in green house conditions (Figure 4-41G and Figure 4-41H). Double transformed descendants of PolA7 line were designated as PolA7B1 to PolA7B4, double transformed descendants of PolA14 line were designated as PolA14B1 to PolA14B4 and double transformed descendants of PolA24 line were designated as PolA24B1 to PolA24B10. To avoid cross-fertilization, flower buds of the transgenic lines were covered with paper bags. After maturation, seeds from transgenic lines were harvested and stored in bijou vials.

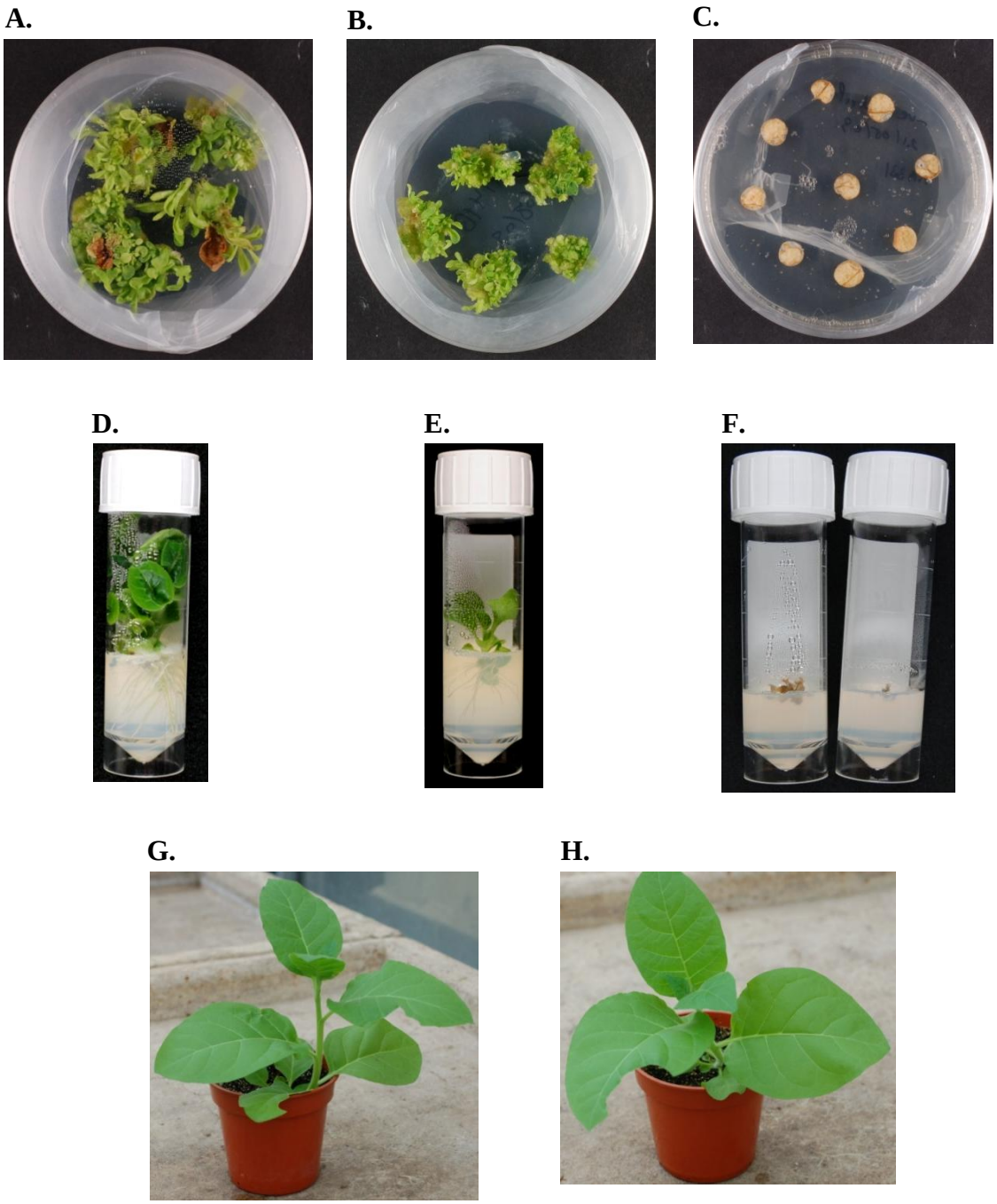


Figure 4-41 Development of transgenic tobacco plants expressing set-B pneumococcal type 2 polysaccharide genes: A. Three-four week old regenerated shoots from transformed leaf disks growing on selective shoot medium containing 30 µg/ml hygromycin. Wild type leaf disks were used for creation of these transformants. **B.** Three-four week old regenerated shoots from double transformed leaf disks growing on selective shoot medium containing 30 µg/ml hygromycin. The leaf disks used for transformation were selected from set-A transgenic lines PolA7, PolA14 and PolA24. These regenerated shoots grew slowly and less vigorously as compared to the regenerated shoots obtained from wild type leaf disks transformation. **C.** Non-transformed (negative control) leaf disks failed to regenerate any shoots/callus on selective shoot medium and dried. **D.** Regenerated transformed shoots from wild type leaf disks transformation developed roots on selective medium containing 30 µg/ml hygromycin and grew vigorously. **E.** Regenerated double transformed shoots developed roots on selective medium containing 30µg/ml hygromycin and 100 µg/ml kanamycin. Again, the double transformed shoots grew less vigorously as compared to shoots from wild type transformed leaf disks. **F.** Escapes (non-transformed

regenerated shoots) failed to develop any roots and became chlorotic on hygromycin selective rooting medium. **G.** A single-transgenic plant is growing in the green house. **H.** A double-transgenic tobacco plant is growing in the green house. Once transferred to the green house, the double-transgenic plants grew normally, similar to the single-transgenic lines.

4.6 Analysis of transgenic plants

4.6.1 PCR analysis

4.6.1.1 PCR analysis of the set-A type 2 polysaccharide genes integration in the plant genome

The integration of the four set-A type 2 polysaccharide genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G) in the plant genome was demonstrated by PCR. Genomic DNA was isolated from 28 transgenic lines and one wild type plant. pKAM3 DNA was used as positive control. PCR was carried out with *cps2*E forward primer KA020 and reverse primer KA021. PCR products were analyzed on 1.5% agarose gel. Figure 4-42 shows results of PCR analysis. A band of 1371 bp corresponding to the *cps2*E was observed in most of the transgenic tobacco lines (lane 3-30) and positive control pKAM8 sample (lane 2). No amplification was observed in the wild type plant sample (lane 1). Absence of any amplification in some of the transgenic lines could be due to PCR error as the genomic DNA template was of low quality. Some of these lines were additionally analysed by RT-PCR and were found positive. All of the transgenic lines were found positive on the basis of T1 seed segregation analysis showing either a 3:1 or 15:1 segregation ratios (section 4.6.2). As the four set-A type 2 polysaccharide genes belong to the same expression cassette-A, therefore, the PCR results also confirm the integration of the remaining three set-A type 2 polysaccharide genes (*cps2*T, *cps2*F and *cps2*G).

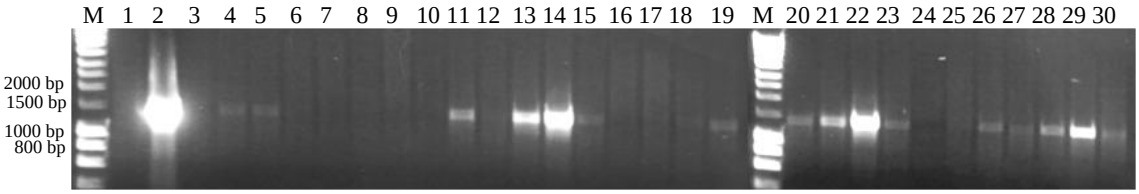


Figure 4-42 PCR analysis of the four set-A type 2 polysaccharide genes ((*cps2*T, *cps2*E, *cps2*F and *cps2*G) integration in the plant genome: lane 1, negative control; lane 2, positive control pKAM3; lane 3-lane 30 contained 28 set-A transgenic lines. PCR was carried out with *cps2*E forward primer KA020 and mini reverse primer KA021. A band of 1371 bp corresponding to the *cps2*E can be seen in many of the transgenic tobacco lines (lane 3-30) and positive control pKAM3 sample (lane 2). No amplification was observed in the wild type plant sample (lane 1). BIOLINE hyper ladder 1 (lane M) was used for size comparison.

4.6.1.2 PCR analysis of the set-B type 2 polysaccharide genes integration in the plant genome

Genomic integration of the three set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H) in transgenic tobacco was demonstrated by PCR. Genomic DNA was isolated from four transgenic lines (PolA7B1, PolA24B1, PolA24B2 and PolA24B3) and one wild type plant. pKAM8 DNA (section 4.4.2.2) was used as positive control. PCR was carried out with *cps2*I forward primer KA046 and *cps2*I reverse primer KA040. PCR products were analyzed on 1.5% agarose gel. Figure 4-43 shows results of PCR analysis. An 1164 bp fragment corresponding to *cps2*I was observed in PolA7B1 (lane 1), PolA24B2 (lane 3) and positive control pKAM8 sample (lane 2). No amplification was observed in the wild type plant sample (lane 5). Although amplification was not observed in PolA24B3 (lane 4), however, RT-PCR analysis confirmed it as a positive transgenic line (section 4.6.3.2). Amplification was also not observed in PolA24B1 (lane 2).

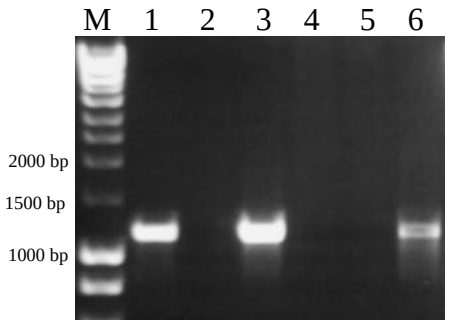


Figure 4-43 PCR analysis of the three set-B type 2 polysaccharide genes ((*cps2*I, *cps2*J and *cps2*H) integration in the plant genome: lane 1 (PolA7B1), lane 2 (PolA24B1), lane 3 (PolA24B2), lane 4 (PolA24B3), lane 5 (negative control), lane 6 (positive control pKAM8). PCR was carried out with *cps2*I revised forward primer KA046 and *cps2*I mini reverse primer KA040. A band of 1164 bp corresponding to the *cps2*I was observed in PolA7B1 (lane 1), PolA24B2 (lane 3) and positive control pKAM8 sample (lane 6). No amplification was observed in the wild type plant sample (lane 5). BIOLINE hyper ladder 1 (lane M) was used for size comparison.

4.6.2 T1 seed screening of set-A type 2 polysaccharide transgenic tobacco lines

T1 seeds from 28 set-A type 2 polysaccharide transgenic tobacco lines were screened on selective MS agar medium (containing 3% sucrose and 150 µg/ml kanamycin) for segregation analysis. Susceptible seedlings became chlorotic and died a few days after germination while resistant lines remained healthy and green. After 3-4 weeks of germination, T1 progenies were scored as resistant and susceptible for segregation analysis. Segregation ratios were calculated and compared with the expected Mendelian segregation ratios of 3:1(resistant: susceptible) for a single copy insertion or 15:1(resistant: susceptible) for a double copy insertion using the chi-square test (Table 4-1). Based on the observed segregation ratio, all the set-A type 2 polysaccharide lines except PolA30 were tested against an expected segregation ratio of 3:1 (resistant: susceptible). PolA30 line was tested against an expected segregation ratio of 15:1 (resistant: susceptible). The null-hypothesis that the observed segregation ratios do not differ significantly from Mendelian segregation ratio of either a 3:1 or 15:1 was not rejected at 10% level ($p>0.100$) (Pearson and Hartley, 1966) for each of the lines tested. All the 27 lines tested for a 3:1 segregation ratio showed the standard Mendelian segregation. However, PolA2 and PolA3 line showed a p-value of 0.106, which is just above the critical p-value of 0.100. PolA30 line that was tested for a 15:1 segregation ratio also followed standard Mendelian segregation ratio. Transgenic tobacco lines PolA7, PolA14 and PolA24 that showed a single copy insertion were selected for double-transformation with pKAM8 carrying the set-B polysaccharide genes (section 4.5.2).

Table 4-1 Segregation analysis of T1 seed of set-A type 2 polysaccharide transgenic tobacco lines

Line	Number of seedlings screened	Resistant seedlings	Susceptible Seedlings	Observed segregation ratio	Expected segregation ratios	Chi square Value	P value
PolA2	142	116	26	4.5:1	3:1	2.613	0.106
PolA3	144	118	26	4.5:1	3:1	2.613	0.106
PolA5	129	98	31	3.2:1	3:1	0.053	0.817
PolA6	122	93	29	3.2:1	3:1	0.053	0.817
PolA7	180	145	35	4.1:1	3:1	1.92	0.166
PolA8	137	105	32	3.3:1	3:1	0.213	0.644
PolA9	186	138	48	2.9:1	3:1	0.053	0.817
PolA10	161	135	35	3.5:1	3:1	0.48	0.488
PolA12	128	98	30	3.3:1	3:1	0.213	0.644
PolA13	151	117	34	3.4:1	3:1	0.213	0.644
PolA14	157	119	38	3.1:1	3:1	0.053	0.817
PolA15	193	147	46	3.2:1	3:1	0.053	0.817
PolA16	156	120	36	3.3:1	3:1	0.213	0.644
PolA17	210	160	50	3.2:1	3:1	0.053	0.817
PolA18	232	177	55	3.2:1	3:1	0.053	0.817
PolA19	166	130	36	3.6:1	3:1	0.48	0.488
PolA20	115	85	30	2.8:1	3:1	0.053	0.817
PolA21	171	127	44	2.9:1	3:1	0.053	0.817
PolA22	187	144	43	3.3:1	3:1	0.213	0.644
PolA24	148	115	33	3.5:1	3:1	0.48	0.488
PolA25	100	75	25	3:1	3:1	0	1
PolA26	118	86	32	2.7:1	3:1	0.213	0.644
PolA27	156	120	36	3.3:1	3:1	0.213	0.644
PolA28	125	96	29	3.3:1	3:1	0.213	0.644
PolA30	145	138	7	19.7:1	15:1	0.177	0.674
PolA31	138	96	42	2.3:1	3:1	1.333	0.248
PolA32	130	90	40	2.3:1	3:1	1.92	0.166
PolA34	141	105	36	2.9:1	3:1	0.053	0.817

4.6.3 RT-PCR analysis

4.6.3.1 RT- PCR to analyse expression of set-A type 2 polysaccharide genes by transgenic tobacco

Total RNA was isolated from three T1 set-A type 2 polysaccharide transgenic tobacco lines (PolA7, PolA14 and PolA24) and one wild type plant. RNA was subjected to RT-PCR analyses. Since the four set-A type 2 polysaccharide genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G) are expressed as a single polycistronic mRNA, any region of the polycistronic mRNA can be used for RT-PCR analysis. Consequently, first strand cDNA was synthesized using *cps2*T mini reverse PCR primer KA018. The resultant cDNA was used to amplify; a shorter segment of *cps2*T and the full-length *cps2*T. The shorter segment (391 bp) of *cps2*T was amplified using forward primer KA024 and *cps2*T reverse PCR primer KA018. The full-length *cps2*T was amplified using *cps2*T forward PCR primer KA014 and *cps2*T reverse PCR primer KA018. To demonstrate chance DNA contamination was not amplified from each sample, pure RNA extracts (without RT reaction) from each transgenic line and wild type plant were also included in the PCR analysis. pKAM3 DNA (section 4.3.2.4) was used as positive control. The results of RT-PCR are shown in Figure 4-44A and Figure 4-44B. Figure 4-44A shows results of RT-PCR analysis for the shorter segment of *cps2*T. A 391 bp fragment corresponding to the amplified region of *cps2*T was observed in all the three transgenic line samples that underwent reverse transcription reaction prior to PCR (lane 3, PolA7; lane 5, PolA14 and lane 7, PolA24). The fragment was also present in the positive control pKAM3 DNA sample (lane 9) but not in the wild type RNA sample (lane 1) that underwent RT reaction prior to PCR. Amplification was not observed in any of the three transgenic lines RNA samples (lane 4, PolA7; lane 6, PolA14 and lane 8, PolA24) and the wild type RNA sample (lane 2) that were not reverse transcribed prior to PCR. Absence of amplification from the RNA samples eliminates the possibility that DNA contaminants are being amplified. Figure 4-44B shows the result of RT-PCR analysis of the full-length *cps2*T gene. A 1191 bp fragment corresponding to *cps2*T was observed in two transgenic line samples (lane 3, PolA7 ; lane 7, PolA24) that underwent reverse transcription reaction prior to PCR and the positive control pKAM3 DNA sample (lane 9) but not in the wild type RNA sample (lane 1) that underwent RT reaction prior to PCR. Amplification was not observed in transgenic line PolA14 sample (lane 5) that

underwent RT reaction prior to PCR. Absence of amplification in this sample could be due to PCR error or RNA degradation as the line was confirmed as positive with RT-PCR analysis for the shorter segment of *cps2*T (see above). No amplification was observed in any of the three transgenic lines RNA samples (lane 4, PolA7; lane 6, PolA14 and lane 8, PolA24) and wild type RNA sample (lane 2) that had not undergone the RT step prior to PCR. Absence of amplification in the pure RNA samples eliminates the possibility of DNA contamination.

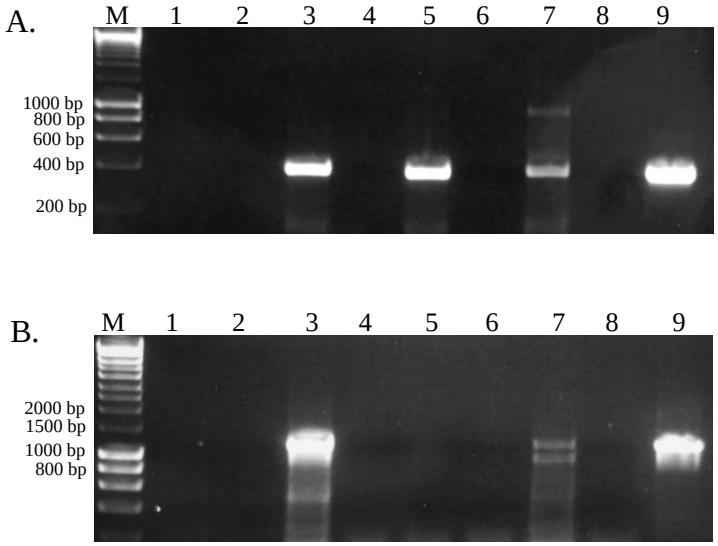


Figure 4-44 RT-PCR to detect expression of type 2 polysaccharide set-A genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G) by transgenic tobacco plants: First strand cDNA was synthesized with *cps2*T mini reverse PCR primer KA018**. A.** The resultant cDNA was used to amplify a shorter segment (391 bp) of *cps2*T using forward primer KA024 and *cps2*T mini reverse PCR primer KA018. A band of correct size (391 bp) was observed in all the three transgenic line samples that underwent reverse transcription reaction prior to PCR (lane 3, PolA7; lane 5, PolA14 and lane 7, PolA24) and the positive control pKAM3 DNA sample (lane 9) but not in the wild type RNA sample that underwent RT reaction prior to PCR (lane 1). No amplification was observed in any of the three transgenic lines RNA samples (lane 4, PolA7; lane 6, PolA14 and lane 8, PolA24) and the wild type RNA sample (lane 2) that did not underwent RT reaction prior to PCR. **B.** The resultant cDNA was also used to amplify the full-length *cps2*T gene using *cps2*T forward PCR primer KA014 and mini reverse PCR primer KA018. A band of correct size (1191 bp) was observed in two transgenic line samples that underwent reverse transcription reaction prior to PCR (lane 3, PolA7 and lane 7, PolA24) and the positive control pKAM3 DNA sample (lane 9) but not in the wild type RNA sample that underwent RT reaction prior to PCR (lane 1). Amplification was not observed in transgenic line PolA14 sample with RT reaction prior to PCR (lane 5). No amplification was observed in any of the three transgenic lines RNA samples (lane 4, PolA7; lane 6, PolA14 and lane 8, PolA24) and wild type RNA sample (lane 2) that did not underwent RT reaction prior to PCR. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

4.6.3.2 RT- PCR to analyse expression of set-B type 2 polysaccharide genes by transgenic tobacco

Total RNA was isolated from two double transformed type 2 polysaccharide transgenic tobacco lines (PolA24B2 and PolA24B3) and one wild type plant. RNA was subjected to RT-PCR analyses. Since the three set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H) are expressed as single polycistronic mRNA, any region of the polycistronic mRNA can be used for RT-PCR analysis. Consequently, first strand cDNA was synthesized using *cps2*I reverse PCR primer KA040. The resultant cDNA was used to amplify *cps2*I. Amplification was carried out with *cps2*I forward PCR primer KA046 and reverse primer KA040. To eliminate chances of DNA contamination, RNA extracts (without RT reaction) from each transgenic line and wild type plant were also included in the PCR. pKAM8 DNA (section 4.4.2.2) was used as positive control. The results of RT-PCR are shown in Figure 4-45. A 1164 bp fragment corresponding to *cps2*I was observed in transgenic line samples that underwent reverse transcription reaction prior to PCR (lane 3, PolA24B2; lane 5, PolA24B3). The fragment was also present in the positive control pKAM8 sample (lane 9). No amplification was observed in the wild type RNA sample (lane 1) that underwent RT reaction prior to PCR. Amplification was not observed in any of the two transgenic lines RNA samples (lane 4, PolA24B2; lane 6, PolA24B3) and the wild type RNA sample (lane 2) that had not undergone RT prior to PCR. Absence of amplification in the pure RNA samples eliminates the possibility of low amounts of genomic DNA contamination giving positive PCR results.

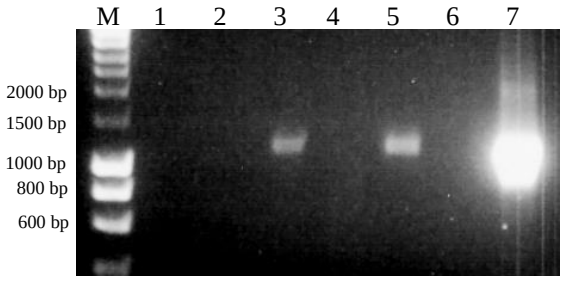


Figure 4-45 RT-PCR to detect expression of type 2 polysaccharide set-B genes (*cps2*I, *cps2*J and *cps2*H) by transgenic tobacco plants: First strand cDNA was synthesized using *cps2*I mini reverse PCR primer KA040. The resultant cDNA was used to amplify *cps2*I using *cps2*I forward primer KA046 and mini reverse PCR primer KA040. A band of correct size (1164 bp) was observed in both of the transgenic line samples that underwent reverse transcription reaction prior to PCR (lane 3, PolA24B2; lane 5, PolA24B3) and the positive control pKAM8 DNA sample (lane 7) but not in the wild type RNA sample

that underwent RT reaction prior to PCR (lane 1). No amplification was observed in any of the two transgenic lines RNA samples (lane 4, PolA24B2; lane 6, PolA24B3) and the wild type RNA sample (lane 2) that did not underwent RT reaction prior to PCR. BIOLINE hyper ladder 1 (lane M) was used for size comparison.

4.6.4 Western blot analysis

Plant proteins (total crude protein and apoplastic fluid protein) were extracted and separated by Tris- Glycine SDS-PAGE. The protein gel was either subjected to Coomassie staining for protein visualization, or the proteins were transferred to a PVDF membrane for western blot detection. Coomassie staining did not show any unique bands related to the seven expressed proteins (data not shown). Western blot detection of the type 2 polysaccharide biosynthetic proteins was carried out using primary antibodies targeted against each peptide tag (Table 2-8) and detection of bound primary antibody using a goat anti-mouse horse radish peroxidise (HRP) conjugate secondary antibody (DakoCytomation). Expressed proteins were visualized with EZ-ECL based detection system. Protein extraction and western blotting protocols were optimized for each of the primary antibodies used.

4.6.4.1 Western blot detection of set-A type 2 polysaccharide genes (*cps2*T, *cps2*E, *cps2*F and *cps2*G) products

4.6.4.1.1 Western blot detection of *cps2*T gene product

Crude protein was isolated from three type 2 polysaccharide set-A transgenic lines (PolA7, PolA14 and PolA24) and a wild type plant. A 52 kDa Multiple Tag protein that contains 16 commonly used tags (GenScript, Cat. No. M101) was used as Western blot positive control. The Multiple Tag protein was also spiked into wild type plant extract to analyse the effect of plant background on its detection. Detection of *cps2*T gene product was carried with mouse Anti-Flag M2 monoclonal Antibody at a dilution of 0.2 µg/ml and polyclonal goat anti-mouse HRP conjugate secondary antibody at a dilution of 1/2000 using the EZ-ECL based detection system. Figure 4-46 shows results of western blot detection of *cps2*T gene product in crude plant protein extract. A specific band (with apparent molecular weight of ~ 41kDa, indicated by an arrow) related to *cps2*T gene product (expected molecular weight ~ 46.3 kDa) was observed in all the three transgenic lines (lane 1, PolA24; lane 2, PolA14 and lane 3, PolA7). The band was absent in wild type plant sample (lane 6). A non-specific band (indicated by an asterisk)

with an apparent molecular weight of about 48 kDa was observed in all the samples. Lane 4 contained the positive control and lane 5 contained the positive control spiked into wild type plant extract.

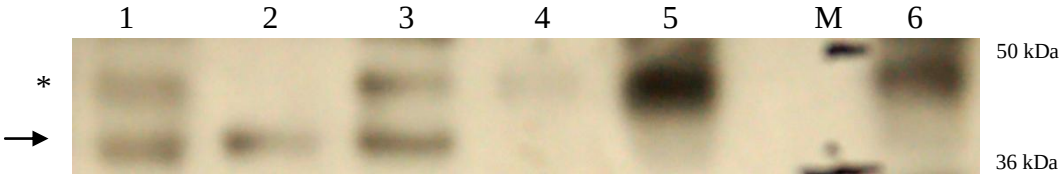


Figure 4-46 Western blot detection of *cps2*T gene product in crude plant extract: Total plant proteins from three set-A transgenic lines (PolA7, PolA14 and PolA24) and one wild type plant were separated through Tris-Glycine SDS-PAGE and analysed by western blotting. A 52 kDa Multiple Tag protein (GenScript, Cat. No. M101) was used as positive control (lane 4). A specific band (with apparent molecular weight of ~ 41kDa, indicated by an arrow) related to *cps2*T gene product was observed in all the three transgenic lines (lane 1, PolA24; lane 2, PolA14 and lane 3, PolA7). The band was absent in wild type plant sample (lane 6). The Multiple Tag protein was also spiked (8 ng) into the wild type plant extract (lane 5). A non-specific band (indicated by an asterisk) with apparent molecular weight similar to the positive control Multiple Tag protein was also observed. Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

4.6.4.1.2 Western blot detection of *cps2*E gene product

Crude protein was isolated from five type 2 polysaccharide set-A transgenic lines (PolA7, PolA13, PolA14, PolA21 and PolA24) and a wild type plant. Detection of *cps2*E gene product was carried with mouse monoclonal Anti-HA antibody (Sigma, Cat. No. H9658) at a dilution of 1/10000 and polyclonal goat anti-mouse HRP conjugate secondary antibody (DakoCytomation, Cat. No. P0447) at a dilution of 1/2000 using the EZ-ECL based detection system. Figure 4-47 shows results of western blot detection of *cps2*E gene product in crude plant protein extract. A specific band (with apparent molecular weight of ~ 45 kDa) related to *cps2*E gene product (expected molecular weight ~ 52 kDa) was observed in all the five transgenic lines (lane 1, PolA24; lane 2, PolA21; lane 3, PolA14; lane 4, PolA13 and lane 3, PolA7). The band was absent in wild type plant sample (lane 6).

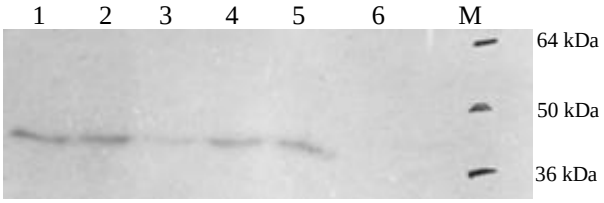


Figure 4-47 Western blot detection of *cps2*E gene product in crude plant extract: Total plant proteins from five set-A transgenic lines (PolA7, PolA13, PolA14, PolA21 and PolA24) and one wild type plant were separated through Tris-Glycine SDS-PAGE and analyzed by western blotting using mouse monoclonal Anti-HA antibody. A specific band related to *cps2*E gene product was observed in all the five transgenic lines (lane 1, PolA24; lane 2, PolA21; lane 3, PolA14; lane 4, PolA13 and lane 3, PolA7). The band was absent in wild type plant sample (lane 6). Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

Apoplastic fluid was also analysed to determine whether *cps2*E protein product was present. For this purpose, apoplast proteins were extracted from three transgenic plants (PolA7, PolA14 and PolA24) and a wild type plant. The extracts were concentrated with methanol and analysed by western blotting as above. Figure 4-48A shows results of the western blotting. A specific band (with apparent molecular weight of ~ 45 kDa, indicated by an asterisk) related to *cps2*E gene product (expected molecular weight ~ 52 kDa) was observed in all the three transgenic lines (lane 1, PolA24; lane 2, PolA14 and lane 3, PolA7) but not in the wild type plant sample (lane 4). Another high molecular weight specific band (indicated by double asterisks) was observed in all the three transgenic lines (lane 1, PolA24; lane 2, PolA14 and lane 3, PolA7). The band was absent in wild type plant sample (lane 4). This high molecular weight band could be related to the aggregated unprocessed set-A polyprotein (~180.5 kDa). A non-specific band (~ 36 kDa, indicated by an arrow) was observed in all samples including the wild type plant.

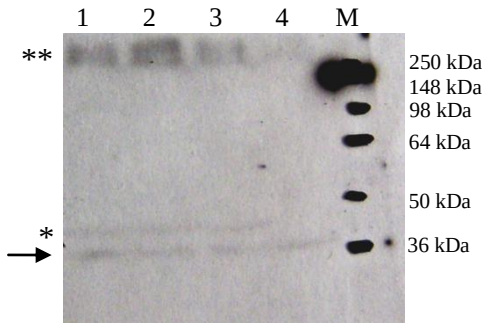


Figure 4-48 Western blot detection of *cps2*E gene product in apoplastic fluids: Apoplast proteins from three set-A transgenic lines (PolA7, PolA14 and PolA24) and one wild type plant were analysed by western blotting using mouse monoclonal Anti-HA antibody. A specific band (with apparent molecular weight of ~ 45 kDa, indicated by an asterisk) related to *cps2*E gene product was observed in all the three transgenic lines (lane 1, PolA24; lane 2, PolA14 and lane 3, PolA7). The band was absent in wild type plant sample (lane 4). Another high molecular weight specific band (indicated by double asterisks) was observed in all the three transgenic lines (lane 1, PolA24; lane 2, PolA14 and lane 3, PolA7) but not in the wild type plant sample (lane 4). A non-specific band (~ 36 kDa, indicated by an arrow) was observed in all samples including the wild type plant. Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

4.6.4.1.3 Western blot detection of *cps2*F gene product

Crude protein was isolated from four type 2 polysaccharide set-A transgenic lines (PolA7, PolA10, PolA14 and PolA24) and a wild type plant and analysed by western blotting. Detection of *cps2*F gene product was carried with mouse Tetra-His antibody at a dilution of 1/5000 and polyclonal goat anti-mouse HRP conjugate secondary antibody at a dilution of 1/2000 using the EZ-ECL based detection system. The results of western blotting are shown in Figure 4-49. A specific band (with apparent molecular weight of ~ 30 kDa) related to *cps2*F gene product (expected weight ~35.6 kDa) was detected in all transgenic lines samples (lane 1, PolA24; lane 2, PolA14; lane 3, PolA10 and lane 4, PolA7). The band was absent in the negative control (lane 5).

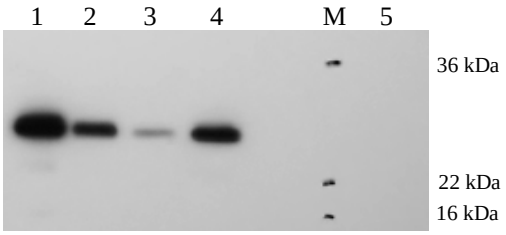


Figure 4-49 Western blot detection of *cps2*F gene product in crude plant extract: Total plant proteins from four set-A transgenic lines (PolA7, PolA14 and PolA24) and one wild type plant were subjected to western blotting using mouse Tetra-His antibody. A specific band (with apparent molecular weight of ~ 30 kDa) was detected in all the four transgenic lines (lane 1, PolA24; lane 2, PolA14; lane 3, PolA10 and lane 4, PolA7) that was absent in the wild type plant sample (lane 5). Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

4.6.4.1.4 Western blot detection of *cps2*G gene product

Crude proteins from transgenic line PolA7 and a wild type plant were analysed by western blotting. Detection of *cps2*G gene product was carried with mouse monoclonal E2 tag antibody at a dilution of 1/500 and polyclonal goat anti-mouse HRP conjugate secondary antibody at a dilution of 1/2000 using the EZ-ECL based detection system. The results of western blotting are shown in Figure 4-50. A specific band (indicated by an arrow) related to *cps2*G gene product (~ 40.3 kDa) was detected in the transgenic line PolA7 (lane 1) but not in the negative control (lane 3). A non-specific band (~ 60 kDa, indicated by an asterisk) was also observed both in PolA7 (lane 1) and negative control (lane 3). Lane 2 contained the 52 kDa Multiple Tag protein.

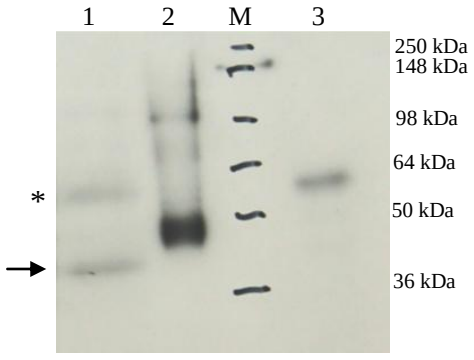


Figure 4-50 Western blot detection of *cps2*G gene product in crude plant extract: Total plant proteins from transgenic line PolA7 and one wild type plant were subjected to western blotting using mouse monoclonal E2 tag antibody. A 52 kDa Multiple Tag protein (GenScript, Cat. No. M101) was used as positive control (lane 2). A specific band (indicated by an arrow) related to *cps2*G gene product (~ 40.3 kDa) was detected in the transgenic line PolA7 (lane 1). The band was absent in the wild type plant

sample (lane 3). Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

4.6.4.2 Western blot detection of set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H) products

4.6.4.2.1 Western blot detection of *cps2*H gene product

Crude proteins from three type 2 polysaccharide double-transgenic lines (PolA7B3, PolA24B2 and PolA24B3), one set-B transgenic line (PolB1) and a wild type plant were analysed by western blotting. Detection of *cps2*H gene product (45.6 kDa) was carried out with culture supernatant containing mouse anti c-MYC (9E10) antibody at a dilution of 1/200 and polyclonal goat anti-mouse HRP conjugate secondary antibody at a dilution of 1/2000 using the EZ-ECL based detection system. The results of western blotting are shown in Figure 4-51. A specific band (with apparent molecular weight of ~ 40 kDa, indicated by an arrow) was detected in PolA24B2 (lane 1) and PolA24B3 (lane 2) but not in PolA7B3 (lane 3) and PolB1 (lane 4). The band was also absent in negative control (lane 5). A non-specific band (~ 52 kDa, indicated by an asterisk) was also observed that was strong in the negative control (lane 5).

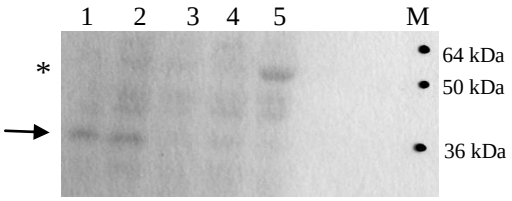


Figure 4-51 Western blot detection of *cps2*H gene product in crude plant extract: Total plant proteins from three type 2 polysaccharide double-transgenic lines (PolA7B3, PolA24B2 and PolA24B3), one set-B transgenic line (PolB1) and a wild type plant were subjected to western blotting using mouse anti c-MYC (9E10) antibody. A specific band (~ 40 kDa, indicated by an arrow) was observed in PolA24B2 (lane 1) and PolA24B3 (lane 2) but not in PolA7B3 (lane 3) and PolB1 (lane 4) and the wild type plant sample (lane 5). A non-specific band (~ 52 kDa, indicated by an asterisk) was also observed. Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

4.6.4.2.2 Western blot detection of *cps2*J gene product

Crude proteins from three type 2 polysaccharide double-transgenic transgenic lines (PolA7B3, PolA24B2 and PolA24B3), one set-B transgenic line (PolB1) and a wild type plant was extracted and analysed by western blotting. *cps2*J gene product was detected with mouse Strep-tag II monoclonal antibody (Novagen, Cat. No. 71590-3) at a

dilution of 1/1000 and polyclonal goat anti-mouse HRP conjugate secondary antibody (DakoCytomation, Cat. No. P0447) at a dilution of 1/2000 using the EZ-ECL based detection system. Figure 4-52 shows results of the western blotting. A very high background signal in the range of 40-98 kDa was observed in all the transgenic lines (PolA24B2, lane 1; PolA24B3, lane 2; PolA7B3, lane 3 and PolB1, lane 4) and the negative control (lane 5). The background signal interfered with the detection of *cps2*J gene product (~ 54.7 kDa). In subsequent experiments, efforts were made to reduce the background noise by reducing the concentration of primary antibody, reducing amount of protein loaded per well, increasing blocking time, increasing the amount blocking agent (skimmed milk protein) in the blocking solution, reducing incubation time with primary antibody, increasing washing steps, increasing amount of Tween in washing buffer, using nitrocellulose membrane instead of PVDF membrane and analysing all the remaining transgenic lines developed. However, no specific band related to *cps2*J gene product was observed (data not shown).

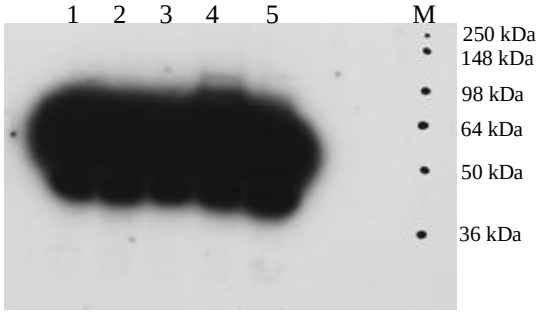


Figure 4-52 Western blot detection of *cps2*J gene product in crude plant extract: A. Total plant proteins from three type 2 polysaccharide double-transgenic lines (PolA7B3, PolA24B2 and PolA24B3), one set-B transgenic line (PolB1) and a wild type plant were subjected to western blotting. A strong background signal in the range of 40-98 kDa was observed in all the transgenic lines samples (PolA24B2, lane 1; PolA24B3, lane 2; PolA7B3, lane 3 and PolB1, lane 4) as well as the negative control (lane 5). No band specific to *cps2*J gene product was observed. Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

4.6.4.2.3 Western blot detection of *cps2*F gene product in the double-transgenic tobacco

Double-transgenic tobacco plants carrying all the 7 type 2 polysaccharide genes were also analyzed for expression of *cps2*F gene product (one of the four set-A type 2 polysaccharide genes) to confirm stable expression of the set-A genes. Crude protein was isolated from double-transgenic line PolA24B2 and two wild type plants and

analysed by western blotting. Detection of *cps2*F gene product was carried out as described in section 4.6.4.1.3. The results of western blotting are shown in Figure 4-53. A specific band (with apparent molecular weight of ~ 30 kDa) related to *cps2*E gene product was detected in PolA24B2 sample (lane 1). The band was absent in the two negative controls (lane 2 and lane 3).

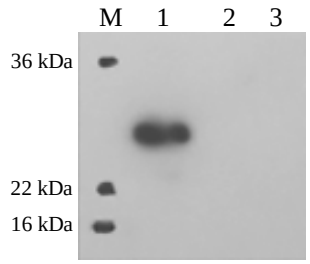


Figure 4-53 Western blot detection of *cps2*F gene product in double-transgenic tobacco: Total plant protein from double-transgenic line PolA24B2 and two wild type plants were subjected to western blotting. *cps2*F gene product was detected with mouse Anti-Flag M2 monoclonal Antibody (Sigma, Cat. No. F3165) at a dilution of 1/5000 and polyclonal goat anti-mouse HRP conjugate secondary antibody (DakoCytomation, Cat. No. P0447), at a dilution of 1/2000. A specific band (~ 30 kDa) related to *cps2*F gene product was detected in PolA24B2 (lane 1) but not in the two wild type plant samples (lane 2 and lane 3). Invitrogen SeeBlue Plus2 protein ladder (lane M) was used for molecular weight comparison.

4.6.5 Detection of pneumococcal type 2 polysaccharide production in transgenic tobacco plants

Attempts to detect Pneumococcal type 2 polysaccharide production in transgenic tobacco plants was carried out using Ouchterlony assay, polysaccharide ELISA, tissue printing, polysaccharide dot blot, polysaccharide gel electrophoresis and polysaccharide immunolocalization.

4.6.5.1 Ouchterlony assay

Ouchterlony double-diffusion immuno assay (Ouchterlony and Nilsson, 1973) is based on the diffusion of both the antigen and antibody molecules through agar and formation of a precipitate (antibody-antigen complex) at the site where the antibody and antigen meet in the agar. The precipitate can be seen as a white arc.

Total plant polysaccharide was extracted from three double-transgenic lines PolA7B3, PolA24B2 and PolA24B3 and a wild type plant. Type 2 pneumococcal polysaccharide

(LGC Standards, Cat. No. ATCC-165-X) was used as positive control. Detection was carried out using polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) as described in section 2.7.4.1. Figure 4-54 shows results of the assay. The outer wells on the microscope slide were filled with 20 µl of polysaccharide extracts from three transgenic plants (PolA7B3, well 2; PolA24B2, well 3 and PolA24B3, well 4), 20 µl of wild type plant extract (well 5) and 20 µl (0.5 µg/ µl) of type 2 pneumococcal polysaccharide (LGC Standards, Cat. No. ATCC-165-X) respectively. The type 2 polysaccharide sample was mixed with wild type plant extract (well 1). To the central well, 20 µl of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) was added. The slide was then incubated at 4°C for 1-2 weeks in a humidity box to prevent drying of samples. This assay proved very insensitive as it was only after 9-10 days of incubation that a white arc was visible (indicated by an arrow) with the positive control type 2 pneumococcal polysaccharide sample and the type 2 antiserum (well 1). This signal developed only when a total of 90-100 µg of type 2 pneumococcal polysaccharide had been added to the well (adding 10 µg each day to keep a constant supply of the antigen). Also 20 µl of the type 2 antiserum had to be added to the central well each day to provide a continuous supply. Purified plant polysaccharide samples were also analyzed by the assay but no signal was detected (Figure not shown).

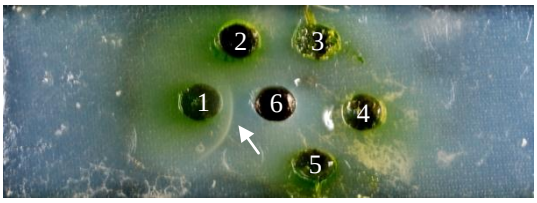


Figure 4-54 Ouchterlony double diffusion immuno assay: To the outer wells, 20 µl of each of the three transgenic plants polysaccharide extracts (PolA7B3, well 2; PolA24B2, well 3 and PolA24B3, well 4) and one wild type plant extract (well 5) were added. 20 µl (0.5 µg/ µl) of type 2 pneumococcal polysaccharide (LGC Standards, Cat. No. ATCC-165-X) mixed in wild type plant extract was added to well 1 as positive control. To the central well, 20 µl of polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) was added. A white arc (zone of antigen-antibody precipitation) was only visible (indicated by an arrow) in the positive control (well 1) on day 9-10 post incubation.

4.6.5.2 Polysaccharide ELISA

A sensitive indirect ELISA assay was developed for detection of type 2 polysaccharide as described in section 2.7.4.2. Serial dilutions of type 2 pneumococcal polysaccharide

were used as standards for the development of the assay. Polyclonal rabbit type 2 antiserum at a dilution of 1/1000 was used as primary antibody and goat Anti-Rabbit IgG (H+L) alkaline phosphatase conjugate at a dilution of 1/2000 was used as secondary antibody. Detection was carried out with SIGMA*FAST* p-Nitrophenyl phosphate substrate. Absorbance readings were recorded at 405 nm. The OD_{405} readings of the type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg) were used to generate a standard curve after subtracting the 0 µg sample OD_{405} value from all the serial dilutions OD_{405} values. The standard curve is shown in Figure 4-55A and Figure 4-55B. Figure 4-55A shows the standard curve for the full range of the type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg) while Figure 4-55B shows the standard curve for the lower range of the type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg).

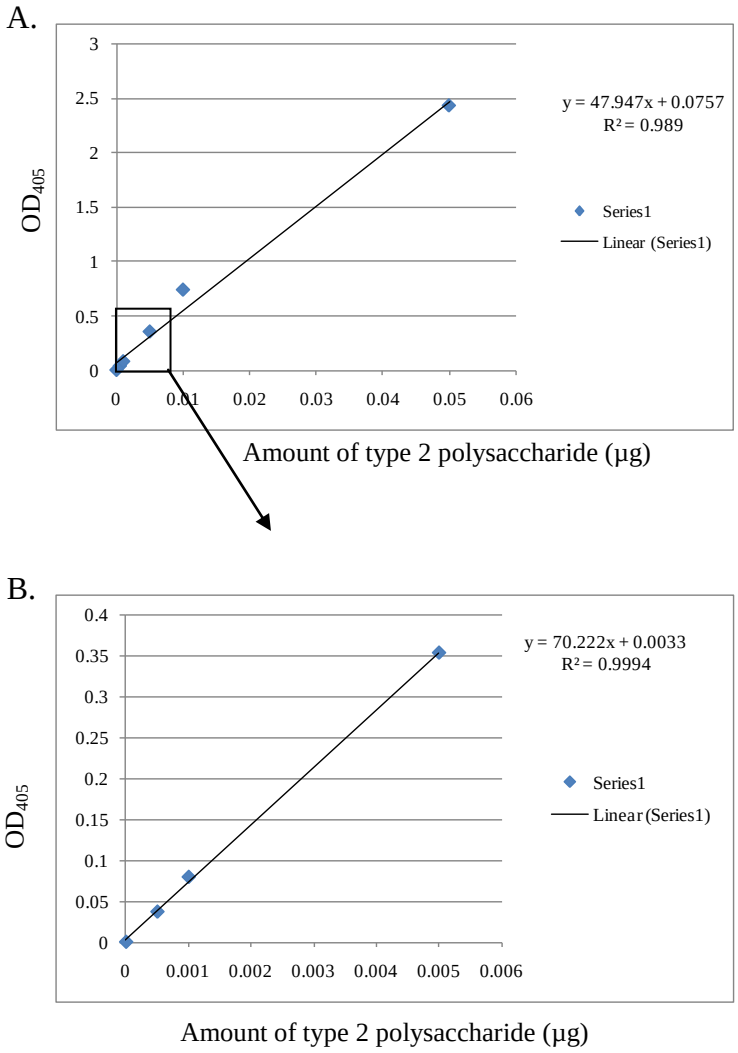


Figure 4-55 Type 2 polysaccharide ELISA standard curve: Serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg respectively) of type 2 pneumococcal polysaccharide were assayed and OD_{405} values were plotted against amount of type 2 polysaccharide in each serial dilution (after subtracting the type 2 polysaccharide 0 µg sample OD_{405} value from all OD_{405} values). **A.** Standard curve for all the type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg), **B.** Standard curve for the lower range of type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg).

To analyse the effects of plant matrix effects in the polysaccharide ELISA assay, different quantities (100 µl, 50 µl, 10 µl, 5 µl, 2.5 µl, 1 µl, 0.5 µl and 0 µl respectively) of wild type plant polysaccharide extract in PBS were subjected to the assay. The results are shown in Figure 4-56. Mean OD_{405} values were plotted against the respective quantity of plant extract used. An increase in ELISA absorbance readings was observed for all the samples containing plant extract (100 µl, 50 µl, 10 µl, 5 µl, 2.5 µl, 1 µl and

0.5 µl respectively) compared to the sample containing no plant extract (containing only PBS). .

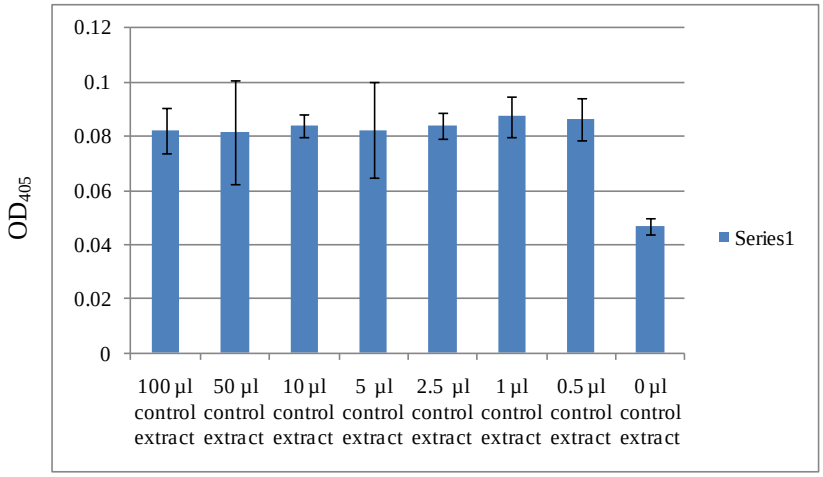


Figure 4-56 Cross-reactivity of plant polysaccharide extract with type 2 polysaccharide assay: Different quantities (100 µl, 50 µl, 10 µl, 5 µl, 2.5 µl, 1 µl, 0.5 µl and 0 µl respectively) of wild type plant extract were assayed by the type 2 polysaccharide ELISA. Mean absorbance readings at OD_{405} were recorded for each sample and plotted against the respective quantities of plant extract used. Error bars show standard error of mean of each sample.

In next experiment, serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg) of pneumococcal type 2 polysaccharide were each mixed with 50 µl of wild type plant polysaccharide extract and analysed by the type 2 polysaccharide ELISA. OD_{405} ELISA readings obtained were plotted against the respective amount of type 2 polysaccharide in each serial dilution (Figure 4-57). The overall signals were reduced compared to the signals obtained from type 2 polysaccharide serial dilution samples containing no plant extract (see above). The data could be used to generate a standard curve in the presence of plant extract (Figure 4-58A and Figure 4-58B). Figure 4-58A shows the standard curve for the full range of the type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg) while Figure 4-58B shows the standard curve for the lower range of the type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg).

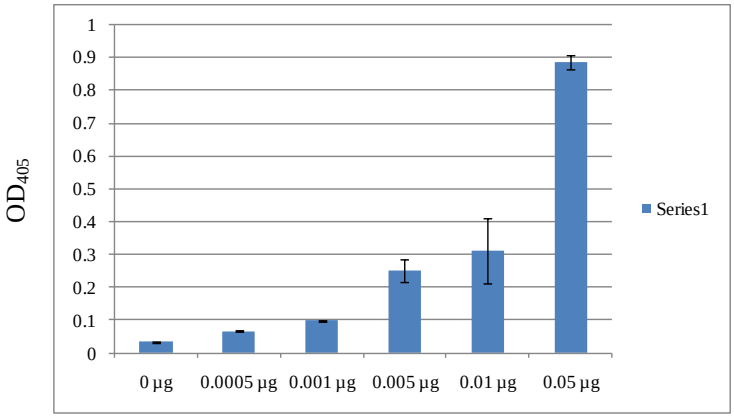


Figure 4-57 Effect of plant extract on detection of type 2 polysaccharide: Serial dilutions of pneumococcal type 2 polysaccharide (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg) were each mixed with 50 µl of wild type polysaccharide extract and subjected to type 2 polysaccharide ELISA. Mean absorbance readings were recorded at 405 nm and plotted against amount of type 2 polysaccharide in each serial dilution. Although the signal was reduced in each serial dilution as compared to the serial dilutions containing no plant extract (see text above) but still a reasonable and linear range of absorbance readings was obtained to help produce a reliable standard curve (see below). Error bars show standard error of mean of each sample.

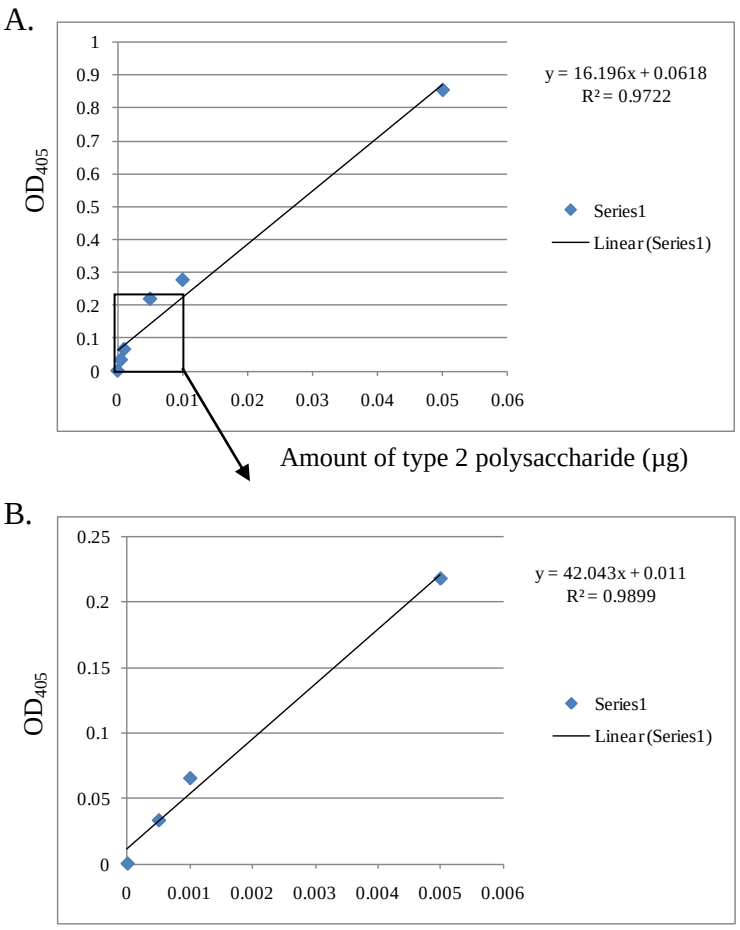


Figure 4-58 Standard curve for type 2 polysaccharide serial dilutions in a plant background: Serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg respectively) of type 2 pneumococcal polysaccharide were each mixed with 50 µl of wild type plant extract and assayed. OD_{405} values were plotted against amount of type 2 polysaccharide in each serial dilution (after subtracting the type 2 polysaccharide 0 µg sample OD_{405} value from all OD_{405} values). **A.** Standard curve for all the type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg, 0.01 µg and 0.05 µg), **B.** Standard curve for the lower range of type 2 polysaccharide serial dilutions (0 µg, 0.0005 µg, 0.001 µg, 0.005 µg).

The optimised type 2 polysaccharide ELISA was used to detect production of type 2 polysaccharide in transgenic tobacco. Total plant polysaccharide (50 µl) from three transgenic lines (PolA7B3, PolA24B2 and PolA24B3) and a wild type plant was isolated and subjected to type 2 polysaccharide ELISA. Absorbance readings were taken at 405 nm and used to plot a bar chart (Figure 4-59). As can be seen from the bar

chart, higher signals were recorded in two transgenic lines PolA24B2 and PolA24B3 compared to the negative control. However, the signals were not significantly different from the negative control as the differences in signals of transgenic lines from the negative control (PolA24B2, 0.039; PolA24B3, 0.054) were less than 2 standard deviations (0.066) of negative control. A weaker signal was recorded in PolA7B3.

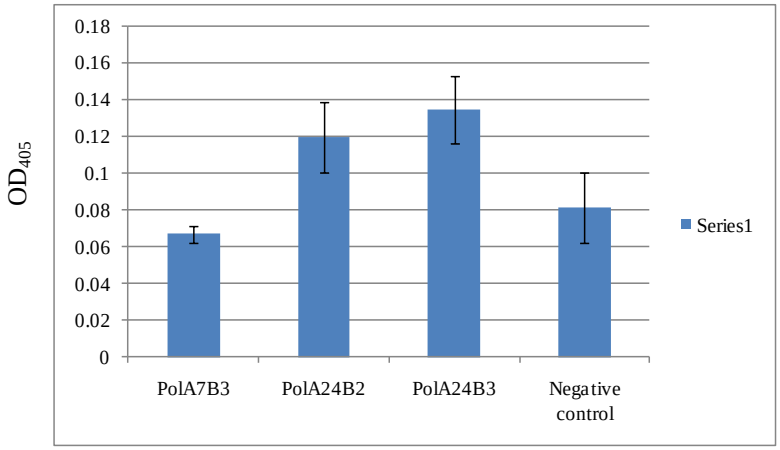


Figure 4-59 Detection of type 2 polysaccharide in transgenic tobacco by type 2 polysaccharide ELISA: Total plant polysaccharide from three transgenic tobacco lines (PolA7B3, PolA24B2 and PolA24B3) and a wild type plant were assayed by type 2 polysaccharide ELISA. OD_{405} values obtained were used to plot a bar chart. Though higher signals were recorded for PolA24B2 and PolA24B3, however, the signals were not significantly different from the negative control as the differences in signals of transgenic lines (PolA24B2, 0.039; PolA24B3, 0.054) from the negative control were less than 2 standard deviations (0.066) of negative control. PolA7B3 showed a weaker signal than the negative control. Error bars show standard error of mean of each sample.

The optimised type 2 polysaccharide ELISA was also used to analyse purified total plant polysaccharides. Purified total plant polysaccharide (50 µl) from three transgenic lines (PolA7B3, PolA24B2 and PolA24B3) and three wild type plants was isolated and subjected to type 2 polysaccharide ELISA. OD_{405} values obtained were used to plot a bar chart (Figure 4-60). No significant differences in signal were detected between the two transgenic lines (PolA24B2 and PolA24B3) and the three negative controls. However, PolA7B3 showed a higher signal, which was significantly different from the average signal of three negative controls. The difference in signal of PolA7B3 from the average signal of negative controls was 0.0270, which is more than the average 2 standard deviations (0.0014) of three negative controls. However, in subsequent experiments reproducible significantly higher signals for PolA7B could not be obtained.

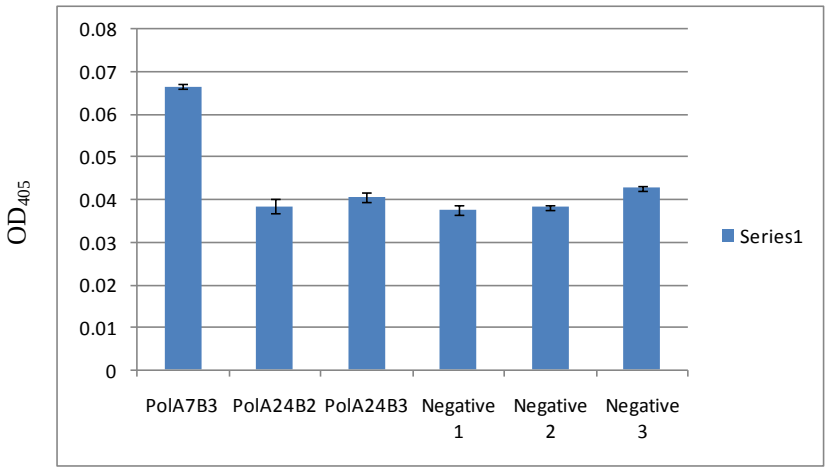


Figure 4-60 Detection of type 2 polysaccharide in purified transgenic plants extracts by type 2 polysaccharide ELISA: Purified plant polysaccharide from three transgenic tobacco lines (PolA7B3, PolA24B2 and PolA24B3) and three wild type plants were assayed by type 2 polysaccharide ELISA. OD_{405} values obtained were use to plot a bar chart. PolA24B2 and PolA24B3 showed signals similar to the three wild type plants. However, PolA7B showed a higher signal which was significantly different from the average signal of three negative controls. The difference in signal of PolA7B from the average signal of negative controls was 0.0270 which is more than the average 2 standard deviations (0.0014) of negative controls. Error bars show standard error of mean of each sample.

4.6.5.3 Tissue printing

Leaf petiole sections (~1 mm thick) were obtained from three double-transgenic lines (PolA7B3, PolA24B2 and PolA24B3) and four wild type plants and used in tissue printing analysis as described in section 2.7.4.3. Three different samples of each plant type were used. Polyclonal rabbit type 2 antiserum was used as primary antibody at a dilution of 1/1000 to 1/15000 and goat Anti-Rabbit IgG (H+L) alkaline phosphatase conjugate was used as secondary antibody at a dilution of 1/2000. SIGMA*FAST* BCIP/NBT alkaline phosphatase substrate was used for signal development. Figure 4-61 shows results of tissue printing. Signal was detected on the outer edges of tissue prints of the three transgenic lines (column 1, PolA7B3; column 2, PolA24B2 and column 3, PolA24B3). The signal was stronger in PolA24B2 (indicated by an arrow, column 2). Signal development on the outer sides of the tissue prints suggests that the reacting

substance is found in the cell sap or apoplastic fluid (expected site of type 2 polysaccharide production), which is released during pressing of the leaf sections on the nitrocellulose membrane. Signal was also observed in some of the wild type plant samples (indicated by an arrow in sample 5c). To separate the background noise from the signal various optimization experiments were conducted using different concentrations of primary antibody (1/2000, 1/4000, 1/8000, 1/12000, 1/15000), no decrease in background signal was observed.

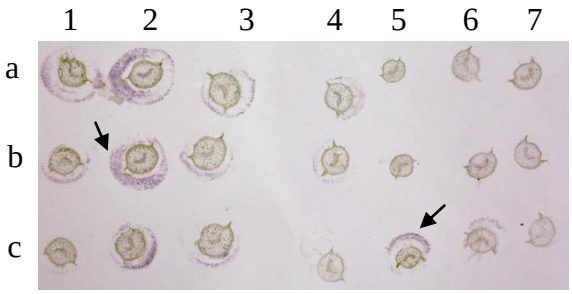


Figure 4-61 Tissue printing of leaf petiole sections to analyze production of type 2 polysaccharide by transgenic tobacco: Tobacco leaf petiole sections of ~1 mm thickness were obtained from three transgenic lines (column 1, PolA7B3; column 2, PolA24B2 and column 3, PolA24B3) and four wild type plants (column 4, column 5, column 6 and column 7) were analysed by tissue printing. A signal was developed on the outer edges of tissue prints of the three transgenic lines (column 1, PolA7B3; column 2, PolA24B2 and column 3, PolA24B3). The signal was stronger in PolA24B2 (indicated by an arrow in replicate 2b). However, the signal was also observed in some of the wild type plant samples (indicated by an arrow in sample 5c).

To increase the sensitivity of the assay, EZ-ECL based detection system was also utilized. For this purpose, polyclonal rabbit type 2 antiserum was used as primary antibody at a dilution of 1/4000 and HRP-Goat Anti-Rabbit IgG (H+L) antibody was used as secondary antibody at a dilution of 1/2000. Three conditions for the experiments were carried out. In condition 1, no primary or secondary antibodies were used. In condition 2, only secondary antibody was used. In condition 3, both primary and secondary antibodies were used. Endogenous peroxidase activity caused large background signals (condition 1, Figure 4-62). Hence, no further experiments were carried out with this detection system.

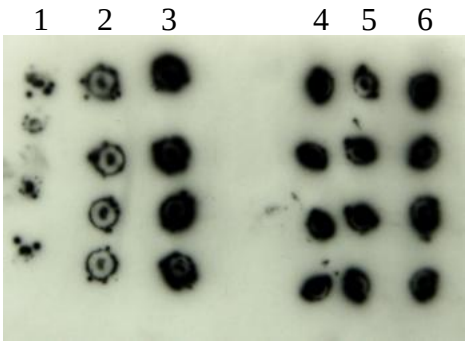


Figure 4-62 Tissue printing of leaf petiole sections to analyze production of type 2 polysaccharide by transgenic plants using EZ-ECL based detection system (experiment 1): Tobacco leaf petiole sections from three transgenic lines (column 1, PolA7B3; column 2, PolA24B2 and column 3, PolA24B3) and three wild type plants (4, 5 and 6) were used for tissue printing. The prints (in this particular experiment) were developed without using primary or secondary antibody. Detection was carried out using EZ-ECL based detection system. Strong background signals were observed in all samples including the negative controls which suggests that endogenous peroxidase activity is present in the plant extracts.

4.6.5.4 Type 2 polysaccharide dot blot assay

The tissue print experiments (section 4.6.5.3), suggested that something in the tissue fluids rather than the cell walls or internal cell structures react with the polyclonal rabbit type 2 polysaccharide antiserum that produced a signal around the tissue print margins. Therefore, a sensitive type 2 polysaccharide dot blot assay (section 2.7.4.4) was developed to detect type 2 polysaccharide in plant tissue fluids (apoplastic fluid and cell sap). The assay can be used to detect nano gram quantities of type 2 polysaccharide. Serial dilutions (0.5 ng, 1 ng, 2 ng, 5 ng, 10 ng, 20 ng) of type 2 pneumococcal polysaccharide were made both in ddH2O and in wild type plant leaf juice (apoplastic fluid and cell sap, section 2.7.1.2). All samples were spotted on nitrocellulose membrane in duplicate. The greenish colour of plant leaf juice sometimes interferes with signal development. Therefore, a further experiment was carried out, using a bleach solution to remove the background colour on the membrane (Uddin *et al.*, 2003). Under condition A (Figure 4-63A), the samples were applied to the nitrocellulose membrane as before. Under condition B (Figure 4-63B), the nitrocellulose membrane was washed with bleach before blocking to remove the greenish colour from samples

that were mixed with plant leaf juice. Polyclonal rabbit type 2 antiserum was used as primary antibody at a dilution of 1/2000 and goat Anti-Rabbit IgG (H+L) alkaline phosphatase conjugate was used as secondary antibody at a dilution of 1/2000. SIGMA*FAST* BCIP/NBT alkaline phosphatase substrate was used for signal development. Figure 4-63A and B show results of this experiment. Signal was detected in all the samples. However, the signal was weaker in samples that had been treated with bleach (Figure 4-63B), which could be due to the loss of some of the antigen during washing with bleach. Detection of type 2 pneumococcal polysaccharide was not affected by plant background as the standards gave a similar pattern of signals both for samples mixed in ddH2O (upper two rows in Figure 4-63A and Figure 4-63B) and sample mixed with plant juice (lower two rows in Figure 4-63A and Figure 4-63B). The assay can be used to detect as low as 0.5 ng of type 2 polysaccharide.

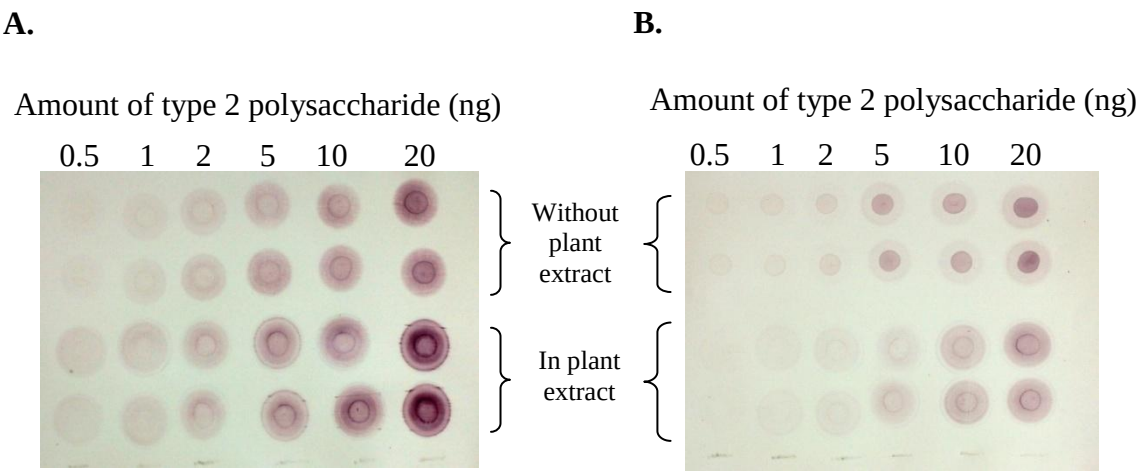


Figure 4-63 Type 2 polysaccharide dot blot assay development: Serial dilutions (0.5 ng, 1 ng, 2 ng, 5 ng, 10 ng and 20 ng) of pneumococcal type 2 polysaccharide were used as standards. In (**A**) the nitrocellulose membrane was not washed with bleach after applying the samples while in (**B)** the nitrocellulose membrane was washed with bleach for 2 min after applying the samples (before blocking) to get rid of the greenish colour of plant extract (the green colour interfered with signal in some cases). Detection of type 2 polysaccharide was carried out alone (upper two rows both in **A** and **B**, each sample applied in duplicate) and in a plant background (lower two rows both in **A** and **B**, each sample applied in duplicate). Signal was detected in all the samples though it was weaker in **B** which could be due to the loss of some of the antigen during washing with bleach.

The type 2 polysaccharide assay was used to detect production of type 2 polysaccharide in transgenic plants (Figure 4-64). Plant leaf juice from two double-transgenic lines (PolA24B2, column 1 and PolA24B3, column 2), two set-A transgenic lines (PolA24-1, column 6 and PolA24-2, column 7), two set-B transgenic lines (PolB5, column 4 and

PolB6, column 5) and a wild type plant (column 3) was assayed. All the samples were applied in triplicate. Column 8 (Figure 4-64A and Figure 4-64B) contained type 2 polysaccharide as positive control. In condition A (Figure 4-64A), the nitrocellulose membrane was not washed with bleach after applying the samples while in condition B (Figure 4-64B), the nitrocellulose membrane was washed with bleach for 2 min after applying the samples (before blocking) to get rid of the green colour of plant extract. As can be seen in Figure 4-64A (condition A), the greenish colour of plant extract interfered with signal detection. While in condition B (Figure 4-64B) the interference was not noted, however, compared to the double-transgenic lines PolA24B2 and PolA24B3 (column 1 and 2) a stronger signal was detected in the negative control (column 3). A weak signal was also observed in the set-A (column 7) and set-B (column 4) transgenic lines. The results suggest that the polyclonal rabbit type 2 antiserum has cross reactivity with plant polysaccharide extract making it difficult to detect very low quantities (sub nano gram quantities) of type 2 polysaccharide.

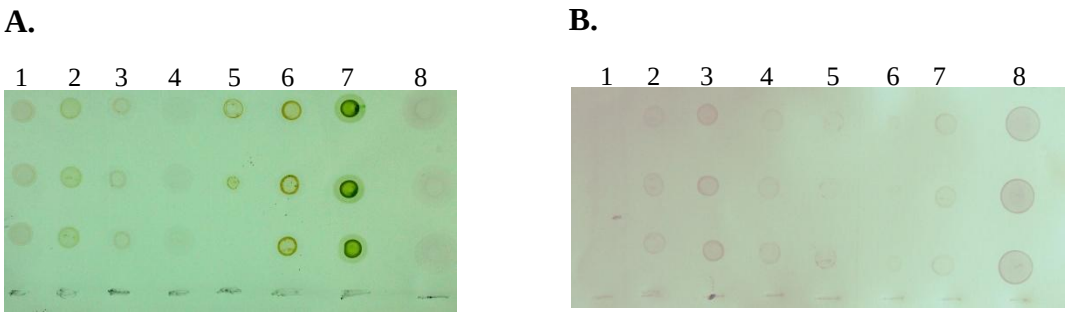


Figure 4-64 detection of type 2 polysaccharide production in transgenic tobacco: Plant leaf juice from 2 double-transgenic lines (PolA24B2, column 1 and PolA24B3, column 2), two set-A transgenic lines (PolA24-1, column 6 and PolA24-2, column 7), two set-B transgenic lines (PolB5, column 4 and PolB6, column 5) and a wild type plant (column 3) was subjected to type 2 polysaccharide dot blot. All the samples were applied in triplet. Column 8 (both in A and B) contained type 2 polysaccharide as positive control. In **(A)** the nitrocellulose membrane was not washed with bleach after applying the samples while in **(B)** the nitrocellulose membrane was washed with bleach for 2 min after applying the samples (before blocking) to get rid of the greenish colour of plant extract.

Experiments were repeated with equalized amount of total plant polysaccharide (5 µg total plant polysaccharide per sample) in each sample (Figure 4-65) to standardise the background signals in the negative controls, the presence of any type 2 polysaccharide in the transgenic lines might be above this 'background'. Signal was detected in the double transgenic lines (PolA24B2, column 1 and PolA24B3, column 2). No signal was detected in the wild type plant (column 3). Signal was also absent in the set-A lines

(PolA24-1, column 6 and PolA24-2, column 7) and set-B transgenic line (PolB6, column 5). However, signal was detected in set-B transgenic line PolB5 (column 4). A strong signal was detected in the type 2 polysaccharide positive control sample (column 8). The experiment was repeated more than once for reproducibility, but no strong evidance was obtained to conclude that the type 2 polysaccharide is being made (data not shown).

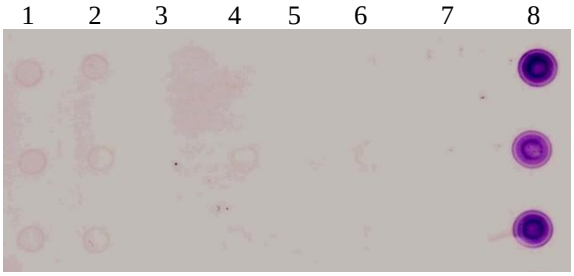


Figure 4-65 detection of type 2 polysaccharide production in transgenic tobacco using equal amounts of plant polysaccharide extract: Plant leaf juice (containing 5 µg total plant polysaccharide) from 2 double transgenic lines (PolA24B2, column 1 and PolA24B3, column 2), two set-A transgenic lines (PolA24-1, column 6 and PolA24-2, column 7), two set-B transgenic lines (PolB5, column 4 and PolB6, column 5) and a wild type plant (column 3) was subjected to type 2 polysaccharide dot blot analysis. All the samples were applied in triplet. Column 8 contained type 2 polysaccharide as positive control. The nitrocellulose membrane was washed with bleach for 2 min after applying the samples (before blocking) to get rid of the greenish colour of plant extract. Signal was detected in the double transgenic lines (PolA24B2, column 1 and PolA24B3, column 2) but not in the wild type plant (column 3). Signal was also not detected in the set-A lines (PolA24-1, column 6 and PolA24-2, column 7) and set-B transgenic line PolB6 (column 5). However, a weak signal was detected in set-B transgenic line PolB5 (column 4).

4.6.5.5 Polysaccharide gel electrophoresis and immunoblotting

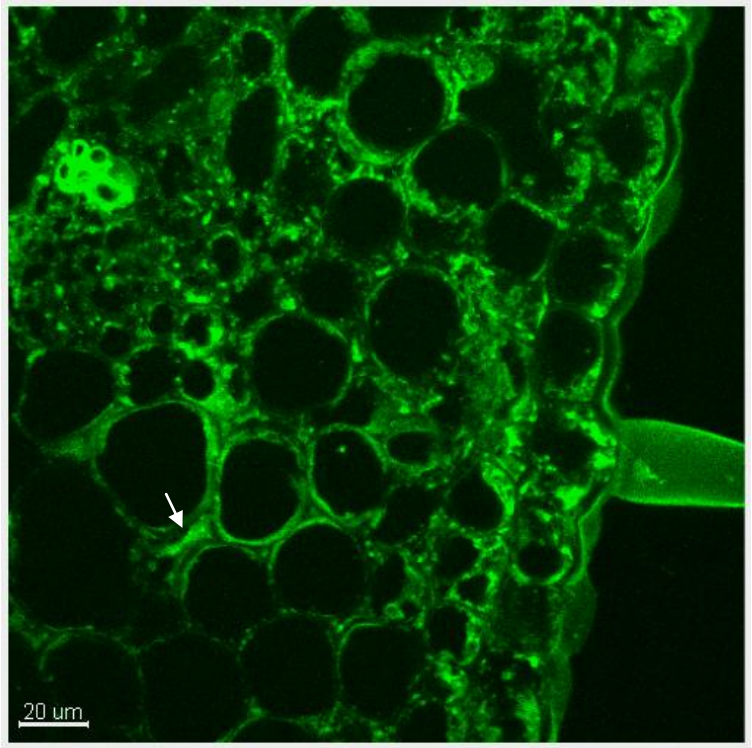
Polysaccharide immunoblotting is another technique that can be used to detect type 2 polysaccharide (Bender *et al*, 2003). Total plant polysaccharide from two transgenic lines (PolA24B2 and PolA24B3) and two wild type plants along with type 2 pneumococcal polysaccharide serial dilutions (5 ng, 10 ng, 50 ng, 100 ng, 500 ng, 1 µg, 2 µg and 5 µg) were analyzed as described in section 2.7.4.5. Detection was carried out with 1:1000 dilution of polyclonal rabbit type 2 antiserum as primary antibody and 1/2000 dilution of goat Anti-Rabbit IgG (H+L) alkaline phosphatase conjugate (Zymax, Cat. No. 81-6122) as secondary antibody. No signal was detected in any of the samples including the positive control serial dilutions except a very weak signal in the serial

dilution containing 5 µg of type 2 polysaccharide (figure not shown). Due to poor assay sensitivity this method was not investigated further.

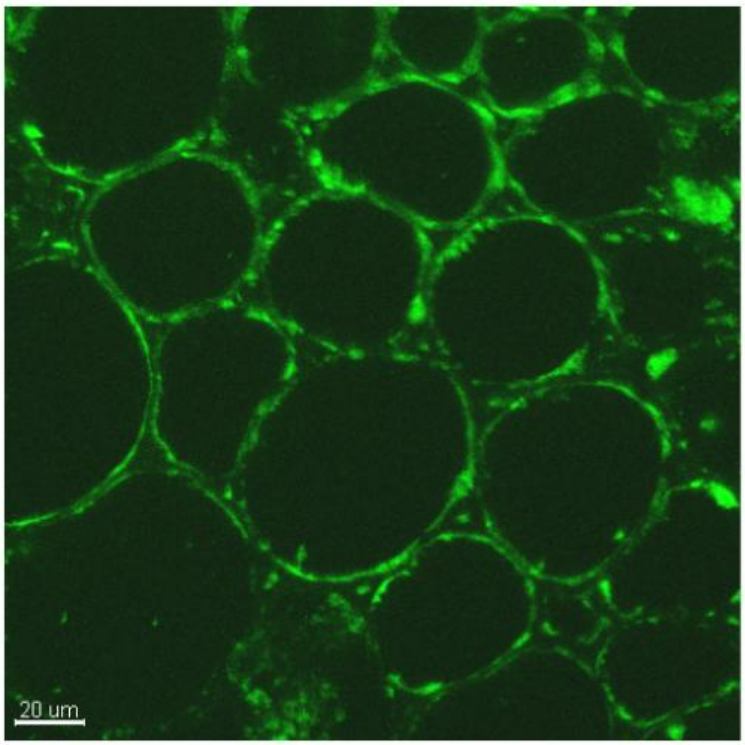
4.6.5.6 Polysaccharide immunolocalization

To localize type 2 polysaccharide production, leaf petioles sections (100-300 µm) from transgenic lines (PolA24B2, PolA24B3) and a wild type plant were analyzed by immunolocalization as described in section 2.7.4.6. Polyclonal rabbit type 2 antiserum was used as primary antibody at a dilution of 1/30 to 1/200, while goat anti rabbit IgG FITC conjugate antibody was used as secondary antibody at a dilution of 1/150. Results of immunolocalization are shown in Figure 4-66A, B and C. Figure 4-66A and Figure 4-66B shows images of two PolA24B3 samples. The type 2 polysaccharide is expected to be produced in the intercellular space. Although signals in intercellular spaces of transgenic plant leaf tissue were detected (indicated by an arrow in Figure 4-66A), however, the higher background signal in the wild type plant sample (Figure 4-66C) made it difficult to conclude the production of type 2 polysaccharide by transgenic plants.

A.



B.



C.

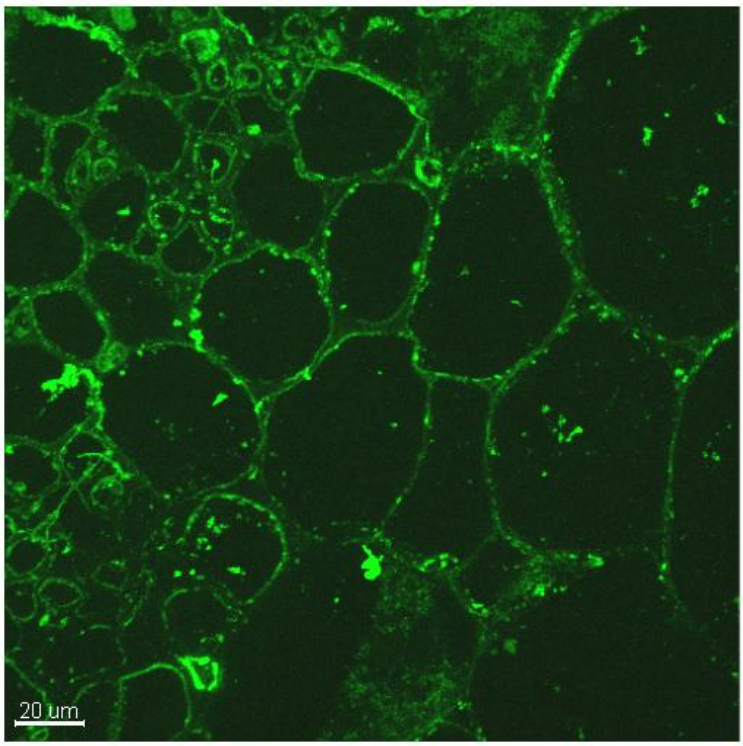


Figure 4-66 immunolocalization of type 2 polysaccharide in transgenic tobacco: Leaf petioles sections from two transgenic lines (PolA24B2, PolA24B3) and a wild type plant were mounted on microscopic slides for immunolocalization. Polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) was used as primary antibody while goat anti rabbit IgG FITC conjugate antibody was used as secondary antibody (Sigma, Cat. No. F-0511) to detect type 2 polysaccharide. (**A**) and (**B**) shows images of two PolA24B3 samples while (**C**) shows image of wild type plant sample. The type 2 polysaccharide is expected to be produced in the intercellular space. Signals can be seen in the transgenic line samples intercellular spaces (indicated by an arrow in **A**), but the higher background signal in the wild type plant sample (**C**) makes it difficult to conclude production of type 2 polysaccharide.

5 DISCUSSION

5.1 Expression of human insulin in transgenic tobacco

The mini human insulin gene was synthesized *in vitro* to encode a strep-tag II-mini-insulin fusion protein. The mini-insulin protein consisted of the B and A chains of human insulin joined by a mini C-peptide (AAK). The synthetic mini-human insulin was expressed in transgenic tobacco to investigate cheaper ways of mass production of this important pharmaceutical. Expression of mini-insulin by transgenic tobacco was confirmed by RT-PCR, western blotting and ELISA. RT-PCR analysis with insulin specific primers confirmed insulin expression in transgenic lines by amplification of an RNA fragment of the correct size (207 bp). Western blot detection was carried out using two primary antibodies; one targeted against the human insulin and another targeted against the strep-tag II attached to the mini-insulin. Both of these antibodies confirmed the production of insulin in crude protein extracts of transgenic tobacco lines. The mini-insulin also carried an N-terminally attached PR1b endoplasmic reticulum targeting signal (Cutt *et al.*, 1988; Denecke *et al.*, 1990). Endoplasmic reticulum is the place for disulphide bond formation (one intra disulphide bond in the A chain and two inter chain disulphide bonds (Frand *et al.*, 2000; Helenius and Aebi, 2001; Dias-Gunasekara & Benham, 2005). After processing the proten is subsequently excreted to the apoplast. Targeting to the apoplast was thought to result in higher accumulation (Schillberg *et al.*, 1999). Therefore, detection of insulin was also carried out in the apoplastic fluids of transgenic tobacco. The strep-tag II-mini-insulin protein could not however be detected in the apoplastic fluid with any of the two primary antibodies used. The presence of the fusion protein in the crude extract and absence in the apoplastic fluid suggests that some sequences in the fusion protein might be responsible for the ER retention of the fusion protein thus not allowing the secretion of fusion protein into the apoplast. The lack of secretion of a recombinant protein (single-chain antibody fragments specific for beet necrotic yellow vein virus coat protein) into the culture fluids of *Nicotiana benthamiana* carrying a *Phaseolus vulgaris* phytohemagglutinin (PHA) secretory signal has been reported previously (Fecker *et al.*, 1996). Another possibility could be that the mini-insulin is not stable in the apoplast and is partially degraded in the apoplastic environment. Proteases in the extracellular fluids have been reported to be involved in

degradation of recombinant proteins (Engelen *et al.*, 1994; Drake *et al.*, 2003; Delannoy *et al.*, 2008).

Purification of the strep-tag II-mini-insulin fusion protein from transgenic tobacco was carried out utilizing the strep-tag II affinity tag attached to the mini-insulin. Strep-tag II has been previously used for purification of recombinant proteins in plants (Witte *et al.*, 2004). The strep-tag II attached to the mini-insulin has affinity to the Strep-Tactin matrix and hence the mini-insulin can be purified from crude protein extracts using Strep-Tactin columns. The bound protein attached to the Strep-tag II can be eluted from the StrepTactin matrix by washing with biotin or desthiobiotin (a competitor of strep-tag II) containing buffers (Skerra and Schmidt, 2000). Purification was achieved by applying crude plant extract to a 50% suspension of Strep-Tactin Sepharose. Only a very low amount of purified insulin could be detected in some of the transgenic lines. One reason for very low recovery could be the low expression levels of this fusion protein. As stated above protein degradation might be another factor responsible for very low recovery. It could be possible that the concentration of desthiobiotin (2.5 mM; desthiobiotin competes with strep-tag II for binding with StrepTactin matrix) in the elution buffer was not high enough to allow the release of the strep tagged mini-insulin. Witte *et al.* (2004) purified a strep II tagged membrane-anchored protein kinase (NtCDPK2) from *Nicotiana benthamiana* leaf extract using a 10 mM concentration of desthiobiotin in the elution buffer. Aggregation of insulin in dimers and hexamers might be another factor that hindered the purification of insulin. Protein aggregation has been one of the impeding reasons in the purification of recombinant proteins (Bondos & Bicknell, 2003). Another factor contributing to low recovery might be the partial degradation of the attached strep-tag II. Van Esse *et al* (2006) failed to purify the C and N terminally Strep II-tagged *Cladosporium fulvum* effector proteins targeted to tomato and tobacco leaf apoplast. Based on their findings they suggested that *in planta* instability of the expressed fusion proteins causes the removal of affinity-tags independent of the nature of the tag and its position in the fusion protein, resulting in no purification of the fusion protein.

Detection of insulin in transgenic plant protein extract was also carried out by ELISA. Two types of ELISA assays were used. A direct ELISA assay was developed for

detection of insulin in transgenic tobacco protein extract. However, the assay showed poor sensitivity and very weak signals were recorded both for the bovine insulin standards and transgenic plant extracts. The assay also overestimated the amounts of insulin in transgenic lines. The detected amount of insulin in one of the transgenic lines was estimated to be 14.3% of total plant protein. The estimated higher amounts of insulin in transgenic plants extracts could be due to the fact that the assay might have different specificities for the bovine insulin used as standard and the plant made mini-insulin. To further analyse detection of insulin in plant extracts, a commercial insulin ELISA kit (Mercodia Ultrasensitive Insulin ELISA kit) was used to estimate the amount of insulin in plant extracts. The Mecrodia ELISA assay is designed for detection of mature insulin in human serum or plasma and has < 0.01 % reactivity to proinsulin (proinsulin consists of the B-chain and A-chain of insulin joined by the central C-peptide). As a result no significant differences in signals in transgenic plants crude protein extracts and wild type plants extracts were observed. It was reasoned that treating the plant protein extract with trypsin to release the mature insulin (mature insulin consists of the B and A chains of insulin joined by two disulphide bonds) by the removal of trypsin cleavable central mini C-peptide and strep-tag II would improve detectability of insulin. Therefore, plant protein extracts were treated with trypsin (section 2.5.6) to release mature insulin by cleaving the two mini-C peptides (one between the DesB30-chain and A-chain of mini-insulin and another between the strep-tag II and mini-insulin). The trypsin-treated and non-treated plant extracts were analysed by the ELISA. Trypsin-treated transgenic line samples showed significantly higher absorbance values than non-treated samples. However, no significant difference in absorbance was observed in trypsin-treated and non-treated wild type plant samples. The amount of insulin detected in one of the trypsin treated transgenic plant extract was estimated to be 4.0 pg/ 100 µg total plant protein while the amount of insulin in the non trypsin-treated extract of the same plant was estimated to be 0.3 pg/ 100 µg total plant protein. The amount of insulin detected in transgenic tobacco extract by Mercodia ELISA is much less than the amount of insulin detected by indirect ELISA (above). This could be due to the low expression levels of insulin synthesized in plant cells. However, a number of other factors could contribute to it. One reason could be the differences in specificity of the antibodies used in both detection systems to the plant made insulin. Also, the reactivity of Mercodia ELISA to commercially available insulin

detemir (DesB30 insulin is a precursor of insulin detemir, for detail see chapter 1) is <0.0000007 % (Mercodia). This could be another factor contributing to such low detectable levels of plant made insulin (which is also a DesB30 form of insulin). Insulin aggregation might be also a factor for low dectability. Ruhlman *et al.* (2007) expressed a cholera toxin B subunit-human proinsulin (CTB-Pins) fusion protein in chloroplasts of transgenic tobacco as an oral vaccine against insulitis. The fusion protein accumulated upto 16% of total soluble protein based on spot densitometry estimations. However, they could not establish expression levels of insulin in transgenic tobacco by ELISA. They attributed the aggregation of the fusion protein to be responsible for the inability to correctly quantify insulin levels by ELISA. Due to the difficulties encountered during purification of insulin, insulin function assays were not conducted.

5.1.1 Future prospects and ways to improve expression of insulin in plants

As mentioned in chapter 1, the incidence of diabetes is increasing quite rapidly and the demand of insulin cannot be met by the existing sources (bacterial and yeast systems) in the near future both in terms of cost and amount. A cost effective, high production system is desired to meet future demands for insulin. For these purposes, synthetic human insulin was targeted to the apoplast of tobacco leaves with a view of achieving higher accumulation and facilitating downstream processing of this protein. However, the low expression levels that were achieved necessitate further investigation into methods for improving expression levels of insulin in tobacco. One strategy could be to target the synthetic mini-insulin to sub-cellular organelles other than apoplast such as the endoplasmic reticulum (ER) and chloroplasts. Targeting to the ER could result in higher accumulation of the recombinant protein (Conrad and Fiedler, 1998) as the ER contains fewer proteases and provides a more stable environment for recombinant proteins (Bondos & Bicknell, 2003). A fusion partner could also be attached to the mini-insulin to improve stability. Boothe *et al.* (2010) analysed the effect of anti-oleosin single chain antibody (scFv) fused to human insulin that was targeted to ER. Expression levels were increased with the inclusion of the fusion partner apparently preventing insulin degradation. Another important sub-cellular organelle is the chloroplast that could help in obtaining higher expression levels of human insulin. The chloroplast has the potential to accumulate recombinant proteins several hundred fold higher than the

concentration than can be achieved by expression from the nucleus (Ruf *et al*., 2001). Ruhlman *et al.* (2007) produced cholera toxin B subunit–human proinsulin (CTB-Pins) fusion protein in chloroplasts of transgenic tobacco and lettuce for developing an oral delivery system for insulin to target insulitis. Expression levels of the fusion protein in tobacco were 16% of total soluble protein and in lettuce expression levels were 2.5% of total soluble protein. Targeting of the human insulin to seeds can improve stability and recovery (Nykiforuk *et al.*, 2006; Boothe *et al.*, 2010). The use of the full-length connecting C-peptide instead of the mini C-peptide might also have some impact on expression levels of insulin (Boothe *et al.*, 2010). Use of affinity tags other than strep-tag II could also be used to get improved recovery of insulin from crude plant protein extracts. Rhizosecretion or the secretion of recombinant protein in the hydroponic medium via roots could also be utilized to simplify downstream processing and increase protein yield (Drake *et al.*, 2009). Unfortunately, the above mentioned strategies could not be investigated due to time constraints. It is expected that a cost-effective plant expression system for insulin will be developed in the near future that could provide an economical method for meeting the future needs. Once it is achieved it would be possible to locally produce the plant made insulin in developing countries to provide insulin to those that are not able to currently afford the cost of this most important therapeutic compound.

5.2 Cloning of *Streptococcus pneumoniae* type 2 capsular polysaccharide genes and expression of type 2 polysaccharide in transgenic tobacco

The two main objectives of cloning the 7 pneumococcal type 2 polysaccharide biosynthetic genes were: (1) to investigate the use of the plant Kex2-like protease system for the expression of multiple genes in the form of a single polycistronic message and (2) to investigate the expression of type 2 pneumococcal capsular polysaccharide by transgenic tobacco engineered with the 7 type 2 polysaccharide biosynthetic genes. To accomplish these objectives, the 7 type 2 polysaccharide biosynthetic genes from *Streptococcus pneumoniae* were grouped into two sets. Set-A consisted of *cps2*E, *cps2*T, *cps2*F and *cps2*G while set-B consisted of *cps2*I, *cps2*J and *cps2*H. Two expression cassettes were designed. The genes were arranged in the form of a single polygene on each expression cassette. The genes in each expression cassette

were linked end to end by three tandemly repeated Kex2 sequences for Golgi localized processing, by the endogenous tobacco Kex2-like activity. The tobacco PR1b endoplasmic reticulum targeting signal sequence (Cutt *et al.*, 1988) was attached to the 5′ end of the polygene in each expression cassette. Double-transgenic tobacco lines carrying all the 7 type 2 polysaccharide genes were developed by sequential transformation with expression cassette -A and expression cassette-B respectively. Transgenic lines were analysed for production of correctly processed type 2 polysaccharide biosynthetic proteins and for production of pneumococcal type 2 polysaccharide. Expression of these genes in transgenic tobacco was confirmed by RT-PCR and western blotting.

5.2.1 Production of correctly processed pneumococcal type 2 polysaccharide biosynthetic proteins *in planta*

Expression and correct processing (by the plant Kex2 protease system) of the four set-A type 2 polysaccharide biosynthetic genes protein products (*cps2*T, *cps2*E, *cps2*F and *cps2*G) was confirmed by western blotting. Expression of *cps2*E was also analysed in apoplastic fluids to confirm location of the protein product accumulation. The presence of *cps2*E protein product in the apoplastic fluids was confirmed. However, a specific high molecular weight product was also observed in the apoplastic fluid. This high molecular weight band was postulated to be related to the aggregated unprocessed set-A polygene protein product (consisting of the four protein products). Among the set-B genes (*cps2*I, *cps2*J and *cps2*H), *cps2*H gene protein product was confirmed by western blotting analysis. *cps2*J expression could not be confirmed mainly as a result of high background signal, using the anti strep-tag II antibody. It is possible that the expression levels of *cps2*J protein product are under detection limits of western blotting with this antibody/tag combination or that the *cps2*J protein product is not stable in the apoplastic environment of tobacco leaves. Detection of *cps2*I was not carried out. However, *cps2*I occurs first in the DNA sequence coding for the set-B type 2 polysaccharide genes (*cps2*I, *cps2*J and *cps2*H) while *cps2*H occurs last. As correct processing of the polyprotein generated *cps2*H it might also be assumed that *cps2*I would also be present. Detection of all the four set-A type 2 polysaccharide biosynthetic genes protein products and at least one set-B gene (*cps2*H) protein product confirms correct

processing by the Golgi localized Kex2 protease system (Jiang and Rogers, 1999). As expected some un-processed fusion protein product (set-A polygene protein product) was also detected in the apoplastic fluids, the predominant forms however were those of processed proteins, suggesting that Kex2 processing turnover is high and that the individual protein products are stable in the plant environment.

5.2.2 Detection of pneumococcal type 2 polysaccharide production in transgenic tobacco plants

Detection of pneumococcal type 2 polysaccharide production in transgenic tobacco plants was carried out using Ouchterlony assay, polysaccharide ELISA, tissue printing, polysaccharide dot blot and polysaccharide immunolocalization.

1) Ouchterlony double-diffusion immuno assay (Ouchterlony and Nilsson, 1973) proved very insensitive. Signal could only be detected in the positive control after 9-10 days of incubation with type 2 antiserum and applying a total of 90-100 µg of type 2 pneumococcal polysaccharide.

2) A sensitive indirect ELISA assay was developed for detection of type 2 polysaccharide that can be used to detect as little as 0.5 ng of type 2 polysaccharide both in a plant background and alone. Detection was carried out using crude plant polysaccharide extracts and purified plant polysaccharide extracts. No significant differences in signals were observed in crude plant polysaccharide extracts from transgenic lines and wild type plants. Occasional higher signals were recorded for one transgenic line PolA7B3 but these results were not reproducible from extract to extract.

3) Tissue printing of leaf petiole sections was carried out in an attempt to both detect and localize type 2 polysaccharide production in plant. Signals specific to the anti-type 2 polysaccharide antisera were detected on the outer edges of tissue prints of transgenic lines. Signal development on the outer sides of the tissue prints suggested that the reacting substance was found in the cell sap or apoplastic fluid (expected site of type 2 polysaccharide production), which was released during pressing of the leaf sections on the nitrocellulose membrane. Some background signal was also detected in wild type

plant samples. To separate the background noise from the signal various optimization experiments were conducted, however, no decrease in background signal was observed.

4) A sensitive type 2 polysaccharide dot blot assay was developed to detect type 2 polysaccharide in plant tissue fluids (apoplastic fluid and cell sap). The assay could be used to detect nano gram quantities of type 2 polysaccharide both in a plant background and alone. Though higher signals were detected in transgenic plant extracts, due to background signals in the wild type plant extracts, the production of type 2 polysaccharide in transgenic plants could not be confirmed using this method.

5) Leaf petioles sections were also analysed by immunolocalization to localize type 2 polysaccharide production *in planta*. The type 2 polysaccharide was expected to be produced in the intercellular space. Although signals in intercellular spaces of transgenic plant leaf tissue were detected, high background signal in the wild type plant sample made it difficult to conclude that type 2 polysaccharide was being produced by transgenic plants.

In planta production of type 2 polysaccharide could not be confirmed by any of the described assays. One main reason for this was the high background signals observed in the wild type plant extracts. It should be noted that all these assays utilized a polyclonal rabbit type 2 antiserum (Statens Serum Institute, product code. 16745) that is the only commercially available antibody against type 2 polysaccharide. The antibody appears to have cross-reactivity to plants extracts. If time had allowed, a monoclonal antibody could have been developed to increase specificity. Immunoprecipitation could also be used to concentrate the type 2 polysaccharide to increase the amount of polysaccharide in the assay. However, the commercially available polyclonal rabbit type 2 antiserum could not be used for immunoprecipitation due to the high cost of this reagent. A sandwich ELISA could also have been developed to increase specificity and sensitivity. Apart from the cross-reactivity issues of the detection systems that were available, it is also possible that the type 2 polysaccharide is not being made in the transgenic plants or is produced under the detection limit. It is possible that the type 2 polysaccharide is not stable *in planta* and is degraded, preventing its accumulation in the leaf apoplast. If the type 2 polysaccharide is not being made at all by the transgenic plants, the possibility is that some of the expressed proteins responsible for its production are not functional or

they are not stable in the leaf environment. Proteases found in the apoplastic fluids might be responsible for partial degradation of these proteins (Engelen *et al.*, 1994; Drake *et al.*, 2003; Delannoy *et al.*, 2008). It would have been a good idea to purify these proteins with the help of the attached protein tags. An *in vitro* functional assay could then be conducted; however published activity assays are only available for *cps*2E (Cartee *et al.*, 2005). This assay utilizes expensive radio-labelled compounds and it was beyond the scope of this project to conduct these kinds of assays. Furthermore, it is possible that the leaf apoplast environment is not suitable for type 2 polysaccharide formation although all the required sugar precursors are available in the leaf apoplast (Fry, 1988; Voitsekhovskaya *et al.*, 2002; Büttner & Sauer, 2000; Yeo *et al.*, 1998; Velíšek & Cejpek, 2005).

The investigation into type 2 polysaccharide synthesis in transgenic tobacco was undertaken in order to attempt development of cost effective sources of pneumococcal polysaccharides that are important constituents of pneumococcal vaccines. The expression of type 2 polysaccharide required cloning of 7 type 2 polysaccharide biosynthetic genes from *Streptococcus pneumoniae.* Hence a multiple gene cloning strategy was desirable. As a Golgi localized Kex2 protease system was previously shown to correctly process recombinant polypeptides in transgenic tobacco (Kinal *et al*, 1995; Jiang and Rogers, 1999), this system was used in attempt to produce multiple proteins involved in the production of *Streptococcus pneumoniae* type 2 polysaccharide. Previously, another pneumococcal polysaccharide the type 3 polysaccharide has been expressed in un-encapsulated pneumococci and *E. coli* (Arrecubieta *et al.*, 1996). The same type 3 polysaccharide has also been expressed in transgenic tobacco (Dr. Claire Smith, department of Infection, Immunity and Inflammation, University of Leicester, personal communications). However, the type 3 polysaccharide is the simplest of capsular polysaccharides consisting of cellobiuronic acid repeating units (Arrecubieta *et al.*, 1996). Four enzymes; a glucose-1-phospate uridyltransferase (*cps*3U), a putative phosphoglucomutase (*cps*3M), a glucose dehydrogenase (*cps*3D) and a type 3 synthase (*cps*3S) are involved in its synthesis in *Streptococcus pneumoniae* (Dillard *et al.*, 1995; Forsee *et al.*, 2000). However, cloning of only one gene the type 3 synthase (*cps*3S) in tobacco was enough to express type 3 polysaccharide *in planta* (Dr. Claire Smith, department of Infection, Immunity and Inflammation, University of Leicester, personal

communications) utilizing the polysaccharide precursors abundantly available in plant leaves (Fry, 1988). In comparison to type 3 polysaccharide, the type 2 polysaccharide is much more complex. The required genes for its synthesis were successfully cloned into tobacco, however, its production by transgenic plants could not be confirmed. It is hoped that future investigations to address the issues discussed above would help to achieve production of type 2 polysaccharide *in planta*.

5.2.3 Future prospects on utilization of the plant Kex2 protease system for multiple proteins expression

A homologue of yeast Kex2 protease has been shown to exist in tobacco (Kinal *et al*, 1995; Jiang and Rogers, 1999) that functions to process polypeptides at pairs of basic residues (Julius *et al.*, 1984; Fuller *et al.*, 1989). This Golgi localized Kex2 protease system has been previously utilized to correctly process a polypeptide into a maximum of two recombinant proteins in transgenic tobacco (Kinal *et al*, 1995; Jiang and Rogers, 1999). In the current example, the Kex2 linker (IGKRG IGKRG IGKRG) was used to join four pneumococcal type 2 polysaccharide biosynthetic genes in one expression cassette (expression cassette-A) and the remaining three genes in another expression cassette (expression cassette-B). All four type 2 polysaccharide biosynthetic proteins belonging to the expression cassette-A and at least one set-B type 2 polysaccharide protein were shown to be correctly processed. The targeting of these proteins to the apoplast was also confirmed. This extends the usability of Kex2 protease system and applies this to multiple heterologous protein production in tobacco. Though expression of type 2 pneumococcal polysaccharide by the transgenic tobacco could not be confirmed, the suitability of the Kex2 protease system for utilization of multiple gene expression was demonstrated. There are many other factors that could contribute to no or very little production of type 2 polysaccharide *in planta*. The Kex2 protease system could be very helpful for improving plants for multigenic traits, improving or altering metabolic pathways, expressing multimeric foreign proteins and expressing multiple enzymes involved in the synthesis of various compounds that have a pharmaceutical/commercial interest. In comparison to the existing multigene transformation procedures like IRESs, multi-transformation, sequential transformation, and co-transformation (see chapter 1 for details), use of Kex2 protease system for

multigene engineering could be cost effective, time efficient, less laborious and the proteins of interest have more chance of being produced in equimolar quantities. All the genes of interest could be integrated in one locus making the post-transformation selection and evaluation procedures simple and efficient to get stable transgenic lines expressing all the genes of interest. This strategy could be useful in the production of pharmaceutically important multimeric proteins. For example full-size IgG antibodies that consist of four polypeptide chains (two heavy and two light) (De Wilde *et al.*, 2000) could be produced in a single tobacco plant utilizing the Kex2 protease system. The system could also be used for simultaneously improving more than one characteristics of plant like disease resistance to more than one disease can be achieved in a single transformation event. Polygenic disease resistance can be incorporated into plants to provide durable resistance against pathogens and insects. Plant metabolic pathways that are of interest for genetic manipulation include carbohydrate metabolism, amino acid and polyamine metabolism, lipid metabolism, and secondary metabolic pathways like production of alkaloids, terpenoids, flavonoids, lignins and quinines (Capell & Christou, 2004). Examples of metabolic pathways engineering include changing properties of starch by changing the ratio of amylose and amylopectin units of starch, inceasing fructans levels, increasing amino acids levels and changing composition of fatty acids (Capell & Christou, 2004). Thus, there are diverse arrays of opportunities where the Kex2 protease system could be utilized for the expression of multiple genes to improve or alter plant characteristics and to develop multigenic therapeutics.

BIBLIOGRAPHY

Akerele, O., 1993. Nature's medicinal bounty: don't throw it away. World Health Forum. **14,** 390-395.

Alberti, K., Zimmet, P., Consultation, W., 1998. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. *Diabetic Medicine.* **15,** 539-553.

Allera-Moreau, C., Chomarat, P., Audinot, V., Cogé, F., Gillard, M., Martineau, Y., Boutin, J.A., Prats, A.C., 2006. The use of IRES-based bicistronic vectors allows the stable expression of recombinant G-protein coupled receptors such as NPY5 and histamine 4. *Biochimie.* **88,** 737-746.

AlonsoDeVelasco, E., Verheul, A.F., Verhoef, J., Snippe, H., 1995. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. *Microbiology and Molecular Biology Reviews.* **59,** 591-603.

Anderson, P.W., Pichichero, M.E., Stein, E.C., Porcelli, S., Betts, R.F., Connuck, D.M., Korones, D., Insel, R.A., Zahradnik, J.M., Eby, R., 1989. Effect of oligosaccharide chain length, exposed terminal group, and hapten loading on the antibody response of human adults and infants to vaccines consisting of *Haemophilus influenzae* type B capsular antigen unterminally coupled to the diphtheria protein CRM197. *The Journal of Immunology.* **142,** 2464-2468.

Arakawa, T., Yu, J., Chong, D.K.X., Hough, J., Engen, P.C., Langridge, W.H.R., 1998. A plant-based cholera toxin B subunit–insulin fusion protein protects against the development of autoimmune diabetes. *Nature biotechnology.* **16,** 934-938.

Arrecubieta, C., López, R., Garcia, E., 1996. Type 3-specific synthase of *Streptococcus pneumoniae* (Cap3B) directs type 3 polysaccharide biosynthesis in *Escherichia coli* and in pneumococcal strains of different serotypes. *The Journal of experimental medicine.* **184,** 449-454.

Astua-Monge, G., Lyznik, A., Jones, V., Mackenzie, S.A., Vallejos, C.E., 2002. Evidence for a prokaryotic insertion-sequence contamination in eukaryotic sequences registered in different databases. *TAG Theoretical and Applied Genetics.* **104,** 48-53.

Austrian, R., and Gold, J., 1964. Pneumococcal bacteremia with especial reference to bacteremic pneumococcal pneumonia. *Annals of Internal Medicine*. **60**:759–76.

Austrian, R., Douglas, R.M., Schiffman, G., Coetzee, A.M., Koornhof, H.J., Hayden-Smith, S., Reid, R.D., 1976. Prevention of pneumococcal pneumonia by vaccination. *Transactions of the Association of American Physicians.* **89,** 184-194.

Avery, O.T, and Goebel, W.F., 1929. Chemo-immunological studies on conjugated carbohydrates. II. Immunological specificity of synthetic sugar-proteins. *Journal of Experimental Medicine*. **50,** 521-533.

Avery, O. T., and Goebel, W. F., 1933. Chemo-immunological studies of the soluble specific substance of pneumococcus. I. The isolation and properties of the acetyl polysaccharide of pneumococcus type 1. *Journal of Experimental Medicine*. **58,** 731–755.

BAJAJ, M., BLUNDELL, T.L., PITTS, J.E., WOOD, S.P., TATNELL, M.A., FALKMER, S., EMDIN, S.O., GOWAN, L.K., CROW, H., SCHWABE, C., 1983. Dogfish insulin. *European Journal of Biochemistry.* **135,** 535-542.

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M.A. and Matzke, A.J.M., 1986. The expression of nopaline synthasehuman growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. *Plant Molecular Biology.* **6**: 347–357.

Bender, M.H., Cartee, R.T. and Yother, J., 2003. Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in Streptococcus pneumoniae. *Journal of Bacteriology.* **185,** 6057-6066.

Berry, A.M., Lock, R.A., Hansman, D., Paton, J.C., 1989a. Contribution of autolysin to virulence of *Streptococcus pneumoniae*. *Infection and immunity.* **57,** 2324-2330.

Berry, A.M., Yother, J., Briles, D.E., Hansman, D., Paton, J.C., 1989b. Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infection and immunity.* **57,** 2037-2042.

Bochner, B.R., Huang, H.C., Schieven, G.L., Ames, B.N., 1980. Positive selection for loss of tetracycline resistance. *Journal of Bacteriology.* **143,** 926-933.

Bochner, B.R., Huang, H.C., Schieven, G.L., Ames, B.N., 1980. Positive selection for loss of tetracycline resistance. *Journal of Bacteriology.* **143,** 926-933.

Boehm, R., 2007. Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. *Annals of the New York Academy of Sciences.* **1102,** 121-134.

Bogaert, D., Hermans, P.W.M., Adrian, P.V., Rümke, H.C., De Groot, R., 2004. Pneumococcal vaccines: an update on current strategies. *Vaccine.* **22,** 2209-2220.

Bondos, S.E. & Bicknell, A., 2003. Detection and prevention of protein aggregation before, during, and after purification. *Analytical Biochemistry.* **316,** 223-231.

Boothe, J., Nykiforuk, C., Shen, Y., Zaplachinski, S., Szarka, S., Kuhlman, P., Murray, E., Morck, D., Moloney, M.M., 2010. Seed-based expression systems for plant molecular farming. *Plant Biotechnology Journal.* **8,** 588-606.

Boulnois, G.J., 1992. Pneumococcal proteins and the pathogenesis of disease caused by Streptococcus pneumoniae. *Microbiology.* **138,** 249-259.

Bourbonnais, Y., Danoff, A., Thomas, D.Y., Shields, D., 1991. Heterologous expression of peptide hormone precursors in the yeast Saccharomyces cerevisiae. Evidence for a novel prohormone endoprotease with specificity for monobasic amino acids. *Journal of Biological Chemistry.* **266,** 13203-13209.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.* **72,** 248-254.

Brennan, S.O. & Peach, R.J., 1988. Calcium-dependent KEX2-like protease found in hepatic secretory vesicles converts proalbumin to albumin. *FEBS letters.* **229,** 167-170.

Briles, D.E., Forman, C., Horowitz, J.C., Volanakis, J.E., Benjamin Jr, W.H., McDaniel, L.S., Eldridge, J., Brooks, J., 1989. Antipneumococcal effects of C-reactive protein and monoclonal antibodies to pneumococcal cell wall and capsular antigens. *Infection and immunity.* **57,** 1457-1464.

Briles, D.E., King, J.D., Gray, M.A., McDaniel, L.S., Swiatlo, E., Benton, K.A., 1996. PspA, a protection-eliciting pneumococcal protein: immunogenicity of isolated native PspA in mice. *Vaccine.* **14,** 858-867.

Briles, D.E., Tart, R.C., Swiatlo, E., Dillard, J.P., Smith, P., Benton, K.A., Ralph, B.A., Brooks-Walter, A., Crain, M.J., Hollingshead, S.K., 1998. Pneumococcal diversity: considerations for new vaccine strategies with emphasis on pneumococcal surface protein A (PspA). *Clinical microbiology reviews.* **11,** 645-657.

Büttner, M. & Sauer, N., 2000. Monosaccharide transporters in plants: structure, function and physiology. *Biochimica et Biophysica Acta (BBA)-Biomembranes.* **1465,** 263-274.

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

Capell, T. & Christou, P., 2004. Progress in plant metabolic engineering. *Current opinion in biotechnology.* **15,** 148-154.

Cartee, R.T., Forsee, W.T., Bender, M.H., Ambrose, K.D., Yother, J., 2005. *Cps*E from type 2 *Streptococcus pneumoniae* catalyzes the reversible addition of glucose-1-phosphate to a polyprenyl phosphate acceptor, initiating type 2 capsule repeat unit formation. *Journal of Bacteriology.* **187,** 7425-7433.

Cavener, D. R., and Ray, S. C., 1991. Eukaryotic start and stop translation sites. *Nucleic Acids Research.* **19**, 3185–3192.

Cefalu, W.T., 2004. Concept, strategies, and feasibility of noninvasive insulin delivery. *Diabetes care.* **27,** 239-246.

Chalmers, R., Sewitz, S., Lipkow, K., Crellin, P., 2000. Complete nucleotide sequence of Tn10. *Journal of Bacteriology.* **182,** 2970-2972.

Chan, S.J., Weiss, J., Konrad, M., White, T., Bahl, C., Yu, S.D., Marks, D., Steiner, D.F., 1981. Biosynthesis and periplasmic segregation of human proinsulin in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America.* **78,** 5401-5404.

Chance, R.E., Ellis, R.M., Bromer, W.W., 1968. Porcine proinsulin: characterization and amino acid sequence. *Science.* **161,** 165-167.

Charlebois, R.L., 1999. Organization of the prokaryotic genome. In Charlebois, R. L., ed, ASM Press, Washington, D.C.: *The American Society for Microbiology*. 151-169.

Chen, J.D. & Morrison, D.A., 1987. Cloning of *Streptococcus pneumoniae* DNA fragments in Escherichia coli requires vectors protected by strong transcriptional terminators. *Gene.* **55,** 179-187.

Chen, M., Liu, X., Wang, Z., Song, J., Qi, Q., Wang, P.G., 2005. Modification of plant N-glycans processing: the future of producing therapeutic protein by transgenic plants. *Medicinal research reviews.* **25,** 343-360.

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

CIMMYT. (2005). Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. Third Edition. Mexico, D.F.: CIMMYT

Conrad, U., and Fiedler, U., 1998. Compartment-specific accumulation of recombinant immunoglobulins in plant cells: An essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. *Plant Molecular Bioliology.* **38,** 101–109.

Coruzzi, G., Broglie, R., Edwards, C. and Chua, N.H., 1984. Tissue-Specific and Light-Regulated Expression of a Pea Nuclear Gene Encoding the Small Subunit of Ribulose-1,5-Bisphosphate Carboxylase. *EMBO Journal.* **3,** 1671-1679.

Cowan, M.J., Ammann, A.J., Wara, D.W., Howie, V.M., Schultz, L., Doyle, N., Kaplan, M., 1978. Pneumococcal polysaccharide immunization in infants and children. *Pediatrics.* **62,** 721-727.

Cramer, C. L., Boothe, J. G., Oishi, K. K., 1999. Transgenic plants for therapeutic proteins: linking upstream and downstream strategies. *Current Topics in Microbiology and Immunology*. **240**: 95–118.

Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Emdin, S.F., Reynolds, C.D., 1979. Structure and biological activity of hagfish insulin. *Journal of Molecular Biology.* **132,** 85-100.

Cutt, J.R., Dixon, D.C., Carr, J.P., Klessig, D.F., 1988. Isolation and nucleotide-sequence of cDNA clones for the pathogenesis-related proteins PR1a, PR1b and PR1c of *Nicotiana tabacum* Cv Xanthi Nc induced by TMV Infection. *Nucleic acids research.* **16,** 9861-9861.

Daniell, H. & Dhingra, A., 2002. Multigene engineering: dawn of an exciting new era in biotechnology. *Current opinion in biotechnology.* **13,** 136-141.

Daniell, H., 2006. Production of biopharmaceuticals and vaccines in plants via the chloroplast genome. *Biotechnology journal.* **1,** 1071-1079.

Datla, R. S. S., Bekkaoui, F., Hammerlindl, J. K., Pilate, G., Dunstan, D. I., and Crosby, W. L., 1993. Improved high-level consstitutive foreign gene expression in plants using an AMV RNA4 untranslated leader sequence. *Plant Science*. **94**, 139–149.

De Jaeger, G., Scheffer, S., Jacobs, A., Zambre, M., Zobell, O., Goossens, A., Depicker, A., Angenon, G., 2002. Boosting heterologous protein production in transgenic dicotyledonous seeds using *Phaseolus vulgaris* regulatory sequences. *Nature biotechnology.* **20,** 1265-1268.

De Wilde, C., Van Houdt, H., De Buck, S., Angenon, G., De Jaeger, G., Depicker, A., 2000. Plants as bioreactors for protein production: avoiding the problem of transgene silencing. *Plant Molecular Biology.* **43,** 347-359.

Delannoy, M., Alves, G., Vertommen, D., Ma, J., Boutry, M., Navarre, C., 2008. Identification of peptidases in *Nicotiana tabacum* leaf intercellular fluid. *Proteomics.* **8,** 2285-2298.

Denecke, J., Botterman, J., Deblaere, R., 1990. Protein secretion in plant cells can occur via a default pathway. *The Plant Cell Online.* **2,** 51-59.

Diabetes atlas, IDF, 2007. http://www.eatlas.idf.org/.

Dias-Gunasekara, S. & Benham, A., 2005. Defining the protein–protein interactions of the mammalian endoplasmic reticulum oxidoreductases (EROs). *Biochemical Society transactions.* **33,** 1382-1384.

Dillard, J.P. & Yother, J., 1991. Analysis of *Streptococcus pneumoniae* sequences cloned into *Escherichia coli*: effect of promoter strength and transcription terminators. *Journal of Bacteriology.* **173,** 5105-5109.

Dillard, J.P., Vandersea, M.W., Yother, J., 1995. Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *The Journal of experimental medicine.* **181,** 973-983.

Docherty, K., Steiner, D.F., 1982. Post-translational proteolysis in polypeptide hormone biosynthesis. *Annual Review of Physiology.* **44,** 625-638.

Dong, X., Stothard, P., Forsythe, I.J. and Wishart, D.S., 2004. PlasMapper: a web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Research.,* **32,** W660.

Drake, P.M.W., Barbi, T., Sexton, A., McGowan, E., Stadlmann, J., Navarre, C., Paul, M.J., Ma, J.K.C., 2009. Development of rhizosecretion as a production system for recombinant proteins from hydroponic cultivated tobacco. *The FASEB Journal.* **23,** 3581-3589.

Drake, P.M.W., Chargelegue, D.M., Vine, N.D., van Dolleweerd, C.J., Obregon, P., Ma, J.K.C., 2003. Rhizosecretion of a monoclonal antibody protein complex from transgenic tobacco roots. *Plant Molecular Biology.* **52,** 233-241.

Draper, J., Scott, R. and Hamil, J., 1988. Transformation of dicotyledonous plant cells using the Ti plasmid of *Agrobacterium tumefaciens* and the Ri plasmid of *A. rhizogenes.* In: Plant genetic transformation and gene expression. A laboratory manual (Draper, J., Scott R., Armitage, P. and Walden, R. eds.), Blackwell Scientific Publications. pp 69-160.

Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., Smith, F., 1956. Colorimetric method for determination of sugars and related substances. *Analytical Chemistry.* **28,** 350-356.

Durfee, T., Nelson, R., Baldwin, S., Plunkett III, G., Burland, V., Mau, B., Petrosino, J.F., Qin, X., Muzny, D.M., Ayele, M., 2008. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *Journal of Bacteriology.* **190,** 2597-2606.

Edwards, K., Johnstone, C. and Thompson, C., 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research.* **19,** 1349.

Eichenbaum, Z. & Livneh, Z., 1998. UV light induces IS10 transposition in *Escherichia coli*. *Genetics.* **149,** 1173-1181.

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

European Parliament. Regulation (EC) No 1829/2003 of the European Parliament and of the council of 22 September 2003 on genetically modified food and feed. http://eurlex.europa.eu/LexUriServ/site/en/oj/2003/l_268/l_26820031018en00010023.pdf. Official Journal of the European Union 2003;L268:1-23.

Evangelista, R. L., Kusnadi, A.R., Howard, J.A. and Nikolov, Z.L., 1988. Process and economic evaluation of the extraction and purification of recombinant glucouronidase from transgenic corn. *Biotechnology Progress*. **14**, 607-614.

Fecker, L.F., Kaufmann, A., Commandeur, U., Commandeur, J., Koenig, R. and Burgermeister, W., 1996. Expression of single-chain antibody fragments (scFv) specific for beet necrotic yellow vein virus coat protein or 25 kDa protein in *Escherichia coli* and *Nicotiana benthamiana*. *Plant Molecular Bioliology*. **32,** 979-986.

Finnegan, J., and McElroy, D., 1994. Transgene inactivation: plants fight back. *Bio/Technology* **12**, 883–888.

Fischer, R., Emans, N., 2000. Molecular farming of pharmaceutical proteins. *Transgenic Research*. **9**, 279–99.

Fischer, R., Stoger, E., Schillberg, S., Christou, P., Twyman, R. M., 2004. Plant-based production of biopharmaceuticals. *Current Opinion in Plant Biology*. **7**,152–158.

Forsee, W.T., Cartee, R.T., Yother, J., 2000. Biosynthesis of type 3 capsular polysaccharide in *Streptococcus pneumoniae*. *Journal of Biological Chemistry*. **275,** 25972-25978.

Foster, T.J., Davis, M.A., Roberts, D.E., Takeshita, K., Kleckner, N., 1981. Genetic organization of transposon Tn10. *Cell.* **23,** 201-213.

Frand, A.R., Cuozzo, J.W., Kaiser, C.A., 2000. Pathways for protein disulphide bond formation. *Trends in cell biology.* **10,** 203-210.

Franken, E., Teuschel, U. and Hain, R., 1997. Recombinant proteins from transgenic plants. *Current Opinion in Biotechnology.* **8**, 411-416.

Fry, S.C., 1988. The growing plant cell wall: chemical and metabolic Analysis. New York, John Wiley and Sons, Inc.

Fuller, R.S., Brake, A., Thorner, J., 1989. Yeast prohormone processing enzyme (KEX2 gene product) is a Ca2 -dependent serine protease. *Proceedings of the National Academy of Sciences.* **86,** 1434-1438.

Fuller, R.S., Sterne, R.E., Thorner, J., 1988. Enzymes required for yeast prohormone processing. *Annual Review of Physiology.* **50,** 345-362.

Gallie, D., 1998. Controlling gene expression in transgenics. *Current Opinion in Plant Biology.* **1,** 166–172.

Garcia, E. and Lopez, R., 1997. Molecular biology of the capsular genes of *Streptococcus pneumoniae*. *Fems Microbiology Letters.* **149,**1-10.

Geelen, S., Bhattacharyya, C., Tuomanen, E., 1993. The cell wall mediates pneumococcal attachment to and cytopathology in human endothelial cells. *Infection and immunity.* **61,** 1538-1543.

Georgiou, G. and Bowden, G. A., 1991. Inclusion body formation and the recovery of aggregated recombinant proteins. pp. 333-356. In: C. Ho, Prokop, A. and Bajpaj, R. (eds.) Recombinant DNA technology and applications. McGraw-Hill, New York.

Germain, D., Zollinger, L., Racine, C., Gossard, F., Dignard, D., Thomas, D.Y., Crine, P., Boileau, G., 1990. The yeast KEX-2-processing endoprotease is active in the Golgi apparatus of transfected NIH 3T3 fibroblasts. *Molecular Endocrinology.* **4,** 1572-1579.

Giddings, G., Allison, G., Brooks, D. and Carter, C., 2000. Transgenic plants as factories for biopharmaceuticals. *Nature Biotechnology*. **18**,1151-1155.

Gilbert, C., Robinson, K., Le Page, R.W.F. and Wells, J.M., 2000. Heterologous expression of an immunogenic pneumococcal type 3 capsular polysaccharide in Lactococcus lactis. *Infect.Immun.,* **68,** 3251-3260

Goderis, I.J.W.M., De Bolle, M.F.C., Francois, I.E.J.A., Wouters, P.F.J., Broekaert, W.F., Cammue, B.P.A., 2002. A set of modular plant transformation vectors allowing flexible insertion of up to six expression units. *Plant Molecular Biology.* **50,** 17-27.

Grant, S.G., Jessee, J., Bloom, F.R., Hanahan, D., 1990. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proceedings of the National Academy of Sciences of the United States of America.* **87,** 4645-4649.

Gu, Y.Q., Anderson, O.D., Londeore, C.F., Kong, X., Chibbar, R.N., Lazo, G.R., 2003. Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. *Genome.* **46,** 1084-1097.

Guerci, B. & Sauvanet, J.P., 2005. Subcutaneous insulin: pharmacokinetic variability and glycemic variability. *Diabetes & metabolism.* **31,** 4S7-4S24.

Guidolin, A., Morona, J. K., Morona, R., Hansman, D. and Paton, J. C., 1994. Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. *Infection and Immunity*. **62,** 5384-5396.

Hajdukiewicz, P., Svab, Z. and Maliga, P., 1994. The small, versatile pPZP family of agrobacterium binary vectors for plant transformation. *Plant Molecular Biology.* **25,** 989-994.

Halling, S.M., Simons, R.W., Way, J.C., Walsh, R.B., Kleckner, N., 1982. DNA sequence organization of IS10-right of Tn10 and comparison with IS10-left. *Proceedings of the National Academy of Sciences of the United States of America.* **79,** 2608-2610.

Harashima, S., 1994. Heterologous protein production by yeast host-vector systems. *Bioprocess Technology.* **19**, 137-158.

Hayflick, J.S., Wolfgang, W.J., Forte, M.A., Thomas, G., 1992. A unique Kex2-like endoprotease from Drosophila melanogaster is expressed in the central nervous system during early embryogenesis. *Journal of Neuroscience.* **12,** 705-717.

Helenius, A., and Aebi, M., 2001. Intracellular functions of *N*-linked glycans. *Science.* **291**, 2364–2369.

Hellen, C.U.T. & Sarnow, P., 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes & development.* **15,** 1593-1612.

Henrichsen, J., 1995. Six newly recognized types of Streptococcus pneumoniae. *Journal of clinical microbiology.* **33,** 2759-2762.

Herbers, K. and Sonnewald, U., 1999. Production of new/modified proteins in transgenic plants. *Current Opinion in Biotechnology.* **10**, 163–168.

Holwerda, B.C., Galvin, N.J., Baranski, T.J., Rogers, J.C., 1990. In vitro processing of aleurain, a barley vacuolar thiol protease. *The Plant Cell Online.* **2,** 1091-1106.

Hood, E.E., (2004). Plants as enzyme factories. In Handbook of Plant Biotechnology (Christou, P. and Klee, H., eds), Wiley-VCH.

Houdebine, L.M. & Attal, J., 1999. Internal ribosome entry sites (IRESs): reality and use. *Transgenic research.* **8,** 157-177.

Houldsworth, S., Andrew, P.W., Mitchell, T.J., 1994. Pneumolysin stimulates production of tumor-necrosis-factor-alpha and interleukin-1-beta by human mononuclear phagocytes. *Infection and Immunity.* **62,** 1501-1503.

Howey, D.C., Bowsher, R.R., Brunelle, R.L., Woodworth, J.R., 1994. Lys (B28), Pro (B29). *human insulin: a rapidly absorbed analogue of human insulin.Diabetes.* **43,** 396-402.

Hunt, A. G., 1994. Messenger RNA 38 end formation in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*. **45**, 47–60.

Iannelli, F., Pearce, B.J., Pozzi, G., 1999. The type 2 capsule locus of *Streptococcus pneumoniae*. *Journal of Bacteriology.* **181,** 2652-2654.

Ingelbrecht, I. L. W., Herman, L. M. F., Dekeyser, R. A., Van Montagu, M. C., and Depicker, A. G., 1989. Different 38 end regions strongly influence the level of gene expression in plant cells. *Plant Cell*. **1**, 671–680.

Jansson, P. E., Lindberg, B., Andersson, M., Lindquist, U. and Henrichsen J., 1975. The structure of capsular polysaccharide of the pneumococcus type II. *Carbohydr. Res*. **40,** 69-75.

Jedrzejas, M.J., 2001. Pneumococcal virulence factors: structure and function. *Microbiology and molecular biology reviews.* **65,** 187-207.

Jedrzejas, M.J., 2004. Extracellular virulence factors of *Streptococcus pneumoniae*. *Frontiers in Bioscience.* **9**, 891-914.

Jiang, L. & Rogers, J.C., 1998. Integral membrane protein sorting to vacuoles in plant cells: evidence for two pathways. *The Journal of cell biology.* **143,** 1183-1199.

Jiang, L. & Rogers, J.C., 1999. Functional analysis of a Golgi-localized Kex2p-like protease in tobacco suspension culture cells. *The Plant Journal.* **18,** 23-32.

Johansen, L.K. & Carrington, J.C., 2001. Silencing on the spot. Induction and suppression of RNA silencing in the Agrobacterium-mediated transient expression system. *Plant Physiology.* **126,** 930-938.

Jorgensen, R.A. & Reznikoff, W.S., 1979. Organization of structural and regulatory genes that mediate tetracycline resistance in transposon Tn10. *Journal of Bacteriology.* **138,** 705-714.

Julius, D., Brake, A., Blair, L., Kunisawa, R. and Thorner, J., 1984. Isolation of the putative structural gene for the Lysine-Arginine-Cleaving Endopeptidase required for processing of yeast Prepro-Alpha-Factor. *Cell*. **37**, 1075-1089.

Kamerling, J.P., 1999. Pneumococcal polysaccharides: a chemical view. *Streptococcus pneumoniae: molecular biology and mechanisms of disease.* 81-114.

Kapila, J., Rycke, R. D, Montagu, M. V. and Angenon, G., 1996. An *Agrobacterium* mediated transient gene expression system for intact leaves. *Plant Science*. **122**, 101–108.

Kaufman, R.J., Davies, M.V., Wasley, L.C., Michnick, D., 1991. Improved vectors for stable expression of foreign genes in mammalian cells by use of the untranslated leader sequence from EMC virus. *Nucleic acids research.* **19,** 4485-4490.

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

Kinal, H., Park, C. M, Berry, J. O, Koltin, Y. and Bruenn, J. A., 1995. Processing and secretion of a virally encoded antifungal toxin in transgenic tobacco plants: Evidence for a Kex2p pathway in plants. *Plant Cell.* **7,** 677-688.

Kirsch, T., Paris, N., Butler, J.M., Beevers, L., Rogers, J.C., 1994. Purification and initial characterization of a potential plant vacuolar targeting receptor. *Proceedings of the National Academy of Sciences.* **91,** 3403-3407.

Kjeldsen, T., Balschmidt, P., Diers, I., Hach, M., Kaarsholm, N.C., Ludvigsen, S., 2001. Expression of insulin in yeast: the importance of molecular adaptation for secretion and conversion. *Biotechnology & genetic engineering reviews.* **18,** 89-121.

Kleckner, N., 1981. Transposable elements in prokaryotes. *Annual Review of Genetics.* **15,** 341-404.

Kleckner, N., Chan, R.K., Tye, B.K., Botstein, D., 1975. Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. *Journal of Molecular Biology.* **97,** 561-575.

Kleckner, N., Roth, J., Botstein, D., 1977. Genetic engineering in Vivo using translocatable drug-resistance elements:: New methods in bacterial genetics. *Journal of Molecular Biology.* **116,** 125-159.

Klein, J.O., 2000. The pneumococcal conjugate vaccine arrives: a big win for kids. *The Pediatric infectious disease journal.* **19,** 181-182.

Kleiner, S. M., 1995. The true nature of herbs. Physician and Sports Medicine. **23**,13-14.

Ko, K., and Koprowski, H., 2005. Plant biopharming of monoclonal antibodies. *Virus Research*. **111**, 93–100.

Kolkman, M.A.B., Van Der Zeijst, B.A.M. and Nuijten, P.J.M., 1997. Functional analysis of glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae* serotype 14. *Journal of Biological Chemistry.* **272,**19502-19508.

Korner, J., Chun, J., O'Bryan, L., Axel, R., 1991. Prohormone processing in Xenopus oocytes: characterization of cleavage signals and cleavage enzymes. *Proceedings of the National Academy of Sciences of the United States of America.* **88,** 11393-11397.

Kovařík, A., Matzke, M., Matzke, A., Koukalová, B., 2001. Transposition of IS10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. *Molecular Genetics and Genomics.* **266,** 216-222.

Krebbers, E., Bosch, D. and Vandekerckhove, J., 1992. Prospects and progress in the production of foreign proteins and peptides in transgenic plants, pp. 315–325. In: P. R. Shewry and S. Gutteridges (eds.), Plant protein engineering. Cambridge University Press, London.

Kusnadi, A. R., Hood, E. E., Witcher, D. R., Howard, J. A., and Nikolov, Z. L., 1998. Production and purification of two recombinant proteins from transgenic corn. *Biotechnology Progress.* **14**, 149–155.

Leibowitz, M.J. & Wickner, R.B., 1976. A chromosomal gene required for killer plasmid expression, mating, and spore maturation in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America.* **73,** 2061-2065.

Li, L., Zhou, Y., Cheng, X., Sun, J., Marita, J.M., Ralph, J., Chiang, V.L., 2003. Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. *Proceedings of the National Academy of Sciences.* **100,** 4939-4944.

Lock, R.A., Hansman, D., Paton, J.C., 1992. Comparative efficacy of autolysin and pneumolysin as immunogens protecting mice against infection by *Streptococcus pneumoniae* . *Microbial pathogenesis.* **12,** 137-143.

Lubon, H., Paleyanda, R. K., Velander, W. H., and Drohan, W. N., 1996. Blood proteins from transgenic animal bioreactors. *Transfusion Medicine Reviews.* **10**, 131-143.

Lutcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., Scheele, G. A., 1987. Selection of AUG initiation codons differs in plants and animals. *EMBO Journal.* **6**: 43–48.

Ma, J.K.C., Barros, E., Bock, R., Christou, P., Dale, P.J., Dix, P.J., Fischer, R., Irwin, J., Mahoney, R., 2005. Molecular farming for new drugs and vaccines. *EMBO reports.* **6,** 593-599.

Ma, J.K.C, Hiatt, A., Hein, M., Vine, N.D., Wang, F., Stabila, P., van Dolleweerd, C., Mostov, K., Lehner, T., 1995. Generation and assembly of secretory antibodies in plants. *Science.* **268,** 716-719.

Ma, J.K.C. and Hiatt, K., 1996. Expressing antibodies in plants for immunotherapy. In: M.R.L. Owen and J. Pen (Eds.) Transgenic Plants: A Production System for Industrial and Pharmaceutical Protein, John Wiley, Chichester, UK.

Ma, J.K.C., Drake, P.M.W., Chargelegue, D., Obregon, P., Prada, A., 2005. Antibody processing and engineering in plants, and new strategies for vaccine production. *Vaccine.* **23,** 1814-1818.

Ma, J.K.C., Drake, P.M.W., Christou, P., 2003. The production of recombinant pharmaceutical proteins in plants. *Nature Reviews Genetics.* **4,** 794-805.

Ma, J.K.C., Lehner, T., Stabila, P., Fux, C.I., Hiatt, A., 1994. Assembly of monoclonal antibodies with IgG1 and IgA heavy chain domains in transgenic tobacco plants. *European journal of immunology.* **24,** 131-138.

Maas, C., Laufs, J., Grant, S., Korfhage, C., and Werr, W., 1991. The combination of a novel stimulatory element in the first exon of the maize *Shrunken-1* gene with the

following intron 1 enhances reporter gene expression up to 1000-fold. *Plant Molecular Biology.* **16**: 199–207.

Markussen, J., 1980. Process for preparing insulin esters. *United Kingdom patent application GB.* **2069502**.

Markussen, J., Havelund, S., Kurtzhals, P., Andersen, A.S., Halstrøm, J., Hasselager, E., Larsen, U.D., Ribel, U., Schäffer, L., Vad, K., 1996. Soluble, fatty acid acylated insulins bind to albumin and show protracted action in pigs. *Diabetologia.* **39,** 281-288.

Martin, B., Alloing, G., Boucraut, C., Claverys, J.P., 1989. The difficulty of cloning *Streptococcus pneumoniae* mal and ami loci in *Escherichia coli*: toxicity of malX and antiA gene products. *Gene.* **80,** 227-238.

Mason, H.S., Warzecha, H., Mor, T., Arntzen, C.J., 2002. Edible plant vaccines: applications for prophylactic and therapeutic molecular medicine. *Trends in molecular medicine.* **8,** 324-329.

Matthews, B. F., and Hughes, C. A., 1993. Nutritional improvement of the aspartate family of amino acids in edible crop plants. *Amino Acids*. **4**, 21–34.

Matzke, A.J.M. & Matzke, M.A., 1998. Position effects and epigenetic silencing of plant transgenes. *Current opinion in plant biology.* **1,** 142-148.

McClintock, B. (1984). The significance of responses of the genome to challenge. Science. **226**, 792-801**.**

Mewett, O., Johnson, H.S. and Holtzapffel, R., 2007. Plant molecular farming in Australia and overseas. Bureau of Rural Sciences.

Meyer, P., and Saedler, H., 1996. Homology-dependent gene silencing in plants. *Annu Rev Plant Physiology and Plant Molecular Biology.* **47**, 23–48.

Modi, P., Mihic, M., Lewin, A., 2002. The evolving role of oral insulin in the treatment of diabetes using a novel RapidMistTM System. *Diabetes/metabolism research and reviews.* **18,** S38-S42.

Moravec, T., Schmidt, M.A., Herman, E.M., Woodford-Thomas, T., 2007. Production of *Escherichia coli* heat labile toxin (LT) B subunit in soybean seed and analysis of its immunogenicity as an oral vaccine. *Vaccine.* **25,** 1647-1657.

Morihara, K., Oka, T., Tsuzuki, H., 1979. Semi-synthesis of human insulin by trypsin-catalysed replacement of Ala-B30 by Thr in porcine insulin. *Nature.* **280,** 412-413.

Morona, J. K., Morona, R. and Paton, J. C., 1997. Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthetic pathway. *Molecular Microbiology.* **23,** 751-763.

Mudaliar, S.R., Lindberg, F.A., Joyce, M., Beerdsen, P., Strange, P., Lin, A., Henry, R.R., 1999. Insulin aspart (B28 asp-insulin): a fast-acting analog of human insulin: absorption kinetics and action profile compared with regular human insulin in healthy nondiabetic subjects. *Diabetes care.* **22,** 1501.

Mulholland, K., 1999. Magnitude of the problem of childhood pneumonia. *Lancet.* **354**, 590-2.

Munoz, R., Mollerach, M., Lopez, R. and Garcia, E., 1997. Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. *Molecular Microbiology.* **25,** 79-92.

Naas, T., Blot, M., Fitch, W.M., Arber, W., 1995. Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. *Molecular biology and evolution.* **12,** 198-207.

Nassau, P.M., Martin, S. L., Brown, R. E., Weston, A., Monsey, D., McNeil, M. R. and Duncan, K., 1996. Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase. *Journal of Bacteriology.* **178:**1047-1052.

Neuhaus, J.M. & Rogers, J.C., 1998. Sorting of proteins to vacuoles in plant cells. *Plant Molecular Biology.* **38,** 127-144.

Nolte, M.S., Karam, J.H., 2004. Pancreatic hormones and antidiabetic drugs. In: Katzung BG, ed. Basic and clinical pharmacology, 9th ed, 2004.

Nuttall, J., Julian, K.C.M., Frigerio, L., 2005. A functional antibody lacking N-linked glycans is efficiently folded, assembled and secreted by tobacco mesophyll protoplasts. *Plant Biotechnology Journal.* **3,** 497-504.

Nykiforuk, C.L., Boothe, J.G., Murray, E.W., Keon, R.G., Goren, H.J., Markley, N.A., Moloney, M.M., 2006. Transgenic expression and recovery of biologically active recombinant human insulin from *Arabidopsis thaliana* seeds. *Plant Biotechnology Journal.* **4,** 77-85.

Ouchterlony O., Nilsson L. A., 1973 - Immuno-diffusion and Immuno-electrophoresis In: Weir, D.M., Editor, 1973. Handbook of Experimental Immunology, Blackwell, Oxford, Chapter 19, pp. 39-51.

Paton, J.C., Andrew, P.W., Boulnois, G.J., Mitchell, T.J., 1993. Molecular analysis of the pathogenicity of *Streptococcus pneumoniae*: the role of pneumococcal proteins. *Annual Reviews in Microbiology.* **47,** 89-115.

Pearson, E.S. and Hartley, H.O. (1966) Biometrika Tables for Staticians, 3rd edition.

Pelletier, J. & Sonenberg, N., 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature.* **334,** 320-325.

Perrin, Y., Vaquero, C., Gerrard, I., Sack, M., Drossard, J., Stöger, E., Christou, P., Fischer, R., 2000. Transgenic pea seeds as bioreactors for the production of a single-chain Fv fragment (scFV) antibody used in cancer diagnosis and therapy. *Molecular Breeding.* **6,** 345-352.

Poland, G.A., 1999. The burden of pneumococcal disease: the role of conjugate vaccines. *Vaccine.* **17,** 1674-1679.

Pollock, H., Kimmel, J., Hamilton, J., Rouse, J., Ebner, K., Lance, V., Rawitch, A., 1987. Isolation and structures of alligator gar (Lepisosteus spatula) insulin and pancreatic polypeptide. *General and comparative endocrinology.* **67,** 375-382.

Pont-Lezica, R.F. and PONT-LEZICA, P.R., 2009. Localizing proteins by tissue printing. *Methods in molecular biology (Clifton, NJ),* **536,** 75-88.

Prats, H., Martin, B., Claverys, J.P., 1985. The hexB mismatch repair gene of *Streptococcus pneumoniae*: characterisation, cloning and identification of the product. *Molecular and General Genetics MGG.* **200,** 482-489.

Prats, H., Martin, B., Claverys, J.P., 1985. The hexB mismatch repair gene of *Streptococcus pneumoniae*: characterisation, cloning and identification of the product. *Molecular and General Genetics MGG.* **200,** 482-489.

Qian, B., Shen, H., Liang, W., Guo, X., Zhang, C., Wang, Y., Li, G., Wu, A., Cao, K., Zhang, D., 2008. Immunogenicity of recombinant hepatitis B virus surface antigen fused with preS1 epitopes expressed in rice seeds. *Transgenic research.* **17,** 621-631.

Rademacher, T., Sack, M., Arcalis, E., Stadlmann, J., Balzer, S., Altmann, F., Quendler, H., Stiegler, G., Kunert, R., Fischer, R., 2008. Recombinant antibody 2G12 produced in maize endosperm efficiently neutralizes HIV-1 and contains predominantly single-GlcNAc N-glycans. *Plant biotechnology journal.* **6,** 189-201.

Ramessar, K., Capell, T., Twyman, R.M., Quemada, H., Christou, P., 2008. Trace and traceability—a call for regulatory harmony. *Nature biotechnology.* **26,** 975-978.

Richardson, S.M., Wheelan, S.J., Yarrington, R.M. and Boeke, J.D., 2006. GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Research.,* **16,** 550-556.

Rigano, M.M. & Walmsley, A.M., 2005. Expression systems and developments in plant-made vaccines. *Immunology and cell biology.* **83,** 271-277.

Robbins, J.B., Austrian, R., Lee, C.J., Rastogi, S.C., Schiffman, G., Henrichsen, J., Mäkelä, P.H., Broome, C.V., Facklam, R.R., Tiesjema, R.H., 1983. Considerations for formulating the second-generation pneumococcal capsular polysaccharide vaccine with emphasis on the cross-reactive types within groups. *The Journal of infectious diseases.* **148,** 1136-1159.

Rockwell, N.C., Krysan, D.J., Komiyama, T., Fuller, R.S., 2002. Precursor processing by Kex2/furin proteases. *Chemical Reviews.* **102,** 4525-4548.

Rosskamp, R.H. & Park, G., 1999. Long-acting insulin analogs. *Diabetes care.* **22,** B109-B113.

Rubins, J.B. & Janoff, E.N., 1998. Pneumolysin: a multifunctional pneumococcal virulence factor. *Journal of Laboratory and Clinical Medicine.* **131,** 21-27.

Ruf, S., Hermann, M., Berger, I. J., Carrer, H., Bock, R., 2001. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nature Biotechnology.* **19**, 870–875.

Ruhlman, T., Ahangari, R., Devine, A., Samsam, M., Daniell, H., 2007. Expression of cholera toxin B–proinsulin fusion protein in lettuce and tobacco chloroplasts–oral administration protects against development of insulitis in non-obese diabetic mice. *Plant biotechnology journal.* **5,** 495-510.

Sambrook, J., Fritsch, E.F. and Maniatis, T., 1989. Molecular cloning-A laboratory Manual. Cold Spring Harbor Laboratory, New York

Sasaki, Y., Sone, T., Yahata, K., Kishine, H., Hotta, J., Chesnut, J.D., Honda, T., Imamoto, F., 2008. Multi-gene gateway clone design for expression of multiple heterologous genes in living cells: Eukaryotic clones containing two and three ORF multi-gene cassettes expressed from a single promoter. *Journal of Biotechnology.* **136,** 103-112.

Schillberg, S., Zimmermann, S., Voss, A., Fischer, R., 1999. Apoplastic and cytosolic expression of full-size antibodies and antibody fragments in *Nicotiana tabacum*. *Transgenic research.* **8,** 255-263.

Schillberga,S., Twymanb, R. M., Fischera, R., 2005. Opportunities for recombinant antigen and antibody expression in transgenic plants—technology assessment. *Vaccine.* **23**, 1764–1769.

Schillberga,S., Twymanb, R. M., Fischera, R., 2005. Opportunities for recombinant antigen and antibody expression in transgenic plants—technology assessment. *Vaccine.* **23**, 1764–1769.

Schwartz, T.W., 1990. The processing of peptide precursors. In: Okamoto, H. (ed.) Molecular Biology of the Islets of Langerhans. Cambridge University Press, Cambridge, pp 153-205.

Seidah, N.G. and Chretien, M., 1992. Proprotein and prohormone convertases of the subtilisin family Recent developments and future perspectives. *Trends in Endocrinology and Metabolism*. **3**, 133-140.

Seppala, I. & Makela, O., 1989. Antigenicity of dextran-protein conjugates in mice. Effect of molecular weight of the carbohydrate and comparison of two modes of coupling. *The Journal of Immunology.* **143,** 1259-1264.

Sexton, A., Harman, S., Shattock, R.J., Ma, J.K.C., 2009. Design, expression, and characterization of a multivalent, combination HIV microbicide. *The FASEB Journal.* **23,** 3590-3600.

Shan, F., 1986. Epidemiology of severe pneumonia in developing countries. *Pediatric Infectious Disease Journal.* **5**, 245-52.

Sharma, A.K. & Sharma, M.K., 2009. Plants as bioreactors: recent developments and emerging opportunities. *Biotechnology Advances.* **27,** 811-832.

Shelly, M.A., Jacoby, H., Riley, G.J., Graves, B.T., Pichichero, M., Treanor, J.J., 1997. Comparison of pneumococcal polysaccharide and CRM197-conjugated pneumococcal oligosaccharide vaccines in young and elderly adults. *Infection and Immunity.* **65,** 242-247.

Singh, B. K., Matthews, B. F., 1994. Molecular regulation of amino acid biosynthesis in plants. *Amino Acids*. **7**, 165–174.

Skaliter, R., Eichenbaum, Z., Shwartz, H., Ascarelli-Goell, R., Livneh, Z., 1992. Spontaneous transposition in the bacteriophage [lambda] cro gene residing on a

plasmid. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis.* **267,** 139-151.

Skerra, A. and Schmidt, T.G.M. (2000). Use of the Strep-tag and streptavidin for detection and purification of recombinant proteins. Meth. Enzymol. **326**, 271–304

Slater, S., Mitsky, T.A., Houmiel, K.L., Hao, M., Reiser, S.E., Taylor, N.B., Tran, M., Valentin, H.E., Rodriguez, D.J., Stone, D.A., 1999. Metabolic engineering of Arabidopsis and Brassica for poly (3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer production. *Nature biotechnology.* **17,** 1011-1016.

Snippe, H., Van Houte, A.J., Van Dam, J.E., De Reuver, M.J., Jansze, M., Willers, J.M., 1983. Immunogenic properties in mice of hexasaccharide from the capsular polysaccharide of *Streptococcus pneumoniae* type 3. *Infection and Immunity.* **40,** 856-861.

Sparrow, P.A.C. & Twyman, R.M., 2009. Biosafety, Risk Assessment and Regulation of Plant-Made Pharmaceuticals. *Methods in Molecular Biology.***483,** 341-353.

Spök, A., 2007. Molecular farming on the rise-GMO regulators still walking a tight rope. *Trends in Biotechnology*. **25,** 74-82.

Stassi, D.L. & Lacks, S.A., 1982. Effect of strong promoters on the cloning in *Escherichia coli* of DNA fragments from *Streptococcus pneumoniae*. *Gene*. **18,** 319-328.

Stoger, E., Sack, M., Perrin, Y., Vaquero, C., Torres, E., Twyman, R.M., Christou, P., Fischer, R., 2002. Practical considerations for pharmaceutical antibody production in different crop systems. *Molecular Breeding.* **9,** 149-158.

Stoger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P., and Fischer, R., 2000. Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. *Plant Molecular Biology.* **42**, 583–590.

Sullivan, M. L., and Green, P. J., 1993. Post-transcriptional regulation of nuclear-encoded genes in higher plants: the roles of mRNA stability and translation. *Plant Molecular Biology*. **23**, 1091–1104.

Swiatlo, E., King, J., Nabors, G.S., Mathews, B., Briles, D.E., 2003. Pneumococcal surface protein A is expressed in vivo, and antibodies to PspA are effective for therapy in a murine model of pneumococcal sepsis. *Infection and immunity*. **71,** 7149-7153.

Tao, J., Ginsberg, I., Banerjee, N., Held, W., Koltin, Y., Bruenn, J.A., 1990. Ustilago maydis KP6 killer toxin: structure, expression in *Saccharomyces cerevisiae*, and relationship to other cellular toxins. *Molecular and cellular biology.* **10,** 1373-1381.

Taylor, C. B., and Green, P. J., 1995. Identification and characterization of genes with unstable transcripts (GUTs) in tobacco. *Plant Molecular Biology.* **28**, 27–38.

Thomas, G., Thorne, B.A., Thomas, L., Allen, R.G., Hruby, D.E., Fuller, R., Thorner, J., 1988. Yeast KEX2 endopeptidase correctly cleaves a neuroendocrine prohormone in mammalian cells. *Science.* **241,** 226-230.

Thomas, L., Leduc, R., Thorne, B.A., Smeekens, S.P., Steiner, D.F., Thomas, G., 1991. Kex2-like endoproteases PC2 and PC3 accurately cleave a model prohormone in mammalian cells: evidence for a common core of neuroendocrine processing enzymes. *Proceedings of the National Academy of Sciences of the United States of America.* **88,** 5297-5301.

Tomasz, A., 1984. Building and breaking of bonds in the cell wall of bacteria— the role for autolysin, p. 3–12. *In* C. Nombela (ed.), Microbial cell wall and autolysins. Elsvier Science Publishers, Amsterdam, The Netherlands.

Trevelyan, J., 1993. Herbal medicine. *Nurs Times*. **89**, 36-38.

Tsuboi, T., Takeo, S., Iriko, H., Jin, L., Tsuchimochi, M., Matsuda, S., Han, E.T., Otsuki, H., Kaneko, O., Sattabongkot, J., 2008. Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infection and immunity.* **76,** 1702-1708.

Tuomanen, E.I. & Masure, H.R., 1997. Molecular and cellular biology of pneumococcal infection. *Microbial Drug Resistance.* **3,** 297-308.

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

Uddin, W., Viji, G., Schumann, G.L., Boyd, S.H., 2003. Detection of *Pyricularia grisea* causing gray leaf spot of perennial ryegrass turf by a rapid immuno-recognition assay. *Plant Disease.* **87,** 772-778.

Van de Wijgert, J.H., Verheul, A.F., Snippe, H., Check, I.J., Hunter, R.L., 1991. Immunogenicity of *Streptococcus pneumoniae* type 14 capsular polysaccharide: influence of carriers and adjuvants on isotype distribution. *Infection and immunity.* **59,** 2750-2757.

Van Den Elzen, P.J.M., Jongedijk, E., Melchers, L.S., Cornelissen, B.J.C., 1993. Virus and fungal resistance: From laboratory to field. *Philosophical Transactions: Biological Sciences.* **342,** 271-278.

Van Esse, H.P., Thomma, B.P.H.J., van't Klooster, J.W. and de Wit, P.J.G.M., 2006. Affinity-tags are removed from *Cladosporium fulvum* effector proteins expressed in the tomato leaf apoplast. *Journalof Experimental Botany.* **57,** 599.

Velíšek, J. & Cejpek, K., 2005. Biosynthesis of food constituents: saccharides. 1. Monosaccharides, oligosaccharides, and related compounds–a review. *Czech Journal of Food Science.* **23,** 129-144.

Verwoerd, T.C., van Paridon, P.A., van Ooyen, A.J.J., van Lent, J.W.M., Hoekema, A. and Pen, J., 1995. Stable accumulation of *Aspergillus niger* phytase in transgenic tobacco leaves. *Plant Physiology*. **109**:1199-1205.

Voitsekhovskaya, O.V., Heber, U., Wiese, C., Lohaus, G., Heldt, H.W., Gamalei, Y.V., 2002. Energized uptake of sugars from the apoplast of leaves: a study of some plants

possessing different minor vein anatomy. *Russian Journal of Plant Physiology*. **49,** 44-53.

Watson, D.A., and Musher D.M., 1999. A brief history of the pneumococcus in biomedical research. Seminars in Respiratory Infection, **14**(3), 198–208.

Whitelam, G.C., 2000. Vaccine production in plants, pp.133-142. In: Andrew, P.W., Oyston, P., Smith, G.L., and Stewart-Tull, D.E. (eds.), Fighting infection in the 21st century, Blackwell Science Ltd Uk.

Whitelam, G.C., Cockburn, B., Gandecha, A.R. and Owen, M.R.L., 1993. Heterologous protein production in transgenic plants. *Biotechnology and Genetic Engineering Reviews.* **11**, 1–29.

Whittingham, J.L., Havelund, S., Jonassens, I., 1997. Crystal structure of a prolonged-acting insulin with albumin-binding properties. *Biochemistry.* **36,** 2826-2831.

Wickner, R.B. & Leibowitz, M.J., 1976. Two chromosomal genes required for killing expression in killer strains of *Saccharomyces cerevisiae. Genetics.* **82,** 429-442.

Willats, W.G.T., Orfila, C., Limberg, G., Buchholt, H.C., van Alebeek, G.J.W.M., Voragen, A.G.J., Marcus, S.E., Christensen, T.M.I.E., Mikkelsen, J.D. and Murray, B.S., 2001. Modulation of the degree and pattern of methyl-esterification of pectic homogalacturonan in plant cell walls. *Journal of Biological Chemistry.*, **276,** 19404.

Winslow, L. C., and Kroll, D. J., 1998. Herbs as medicines. *Archives of Internal Medicine*. **158**, 2192–2199.

Witte, C.P., Noel, L., Gielbert, J., Parker, J. and Romeis, T., 2004. Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. *Plant Molecular Biology*. **55,** 135-147.

Witte, C.P., Noel, L., Gielbert, J., Parker, J., Romeis, T., 2004. Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. *Plant Molecular Biology*. **55,** 135-147.

World Health Organization. Pneumococcal vaccines. The weekly epidemiological record 2003, **14**:110-119.

X Dong, Paul Stothard, Ian J. Forsythe, and David S. Wishart "PlasMapper: a web server for drawing and auto-annotating plasmid maps" Nucleic Acids Research. 2004 Jul 1;32(Web Server issue):W660-4

Yeo, U.D., Chae, Y.K., Lee, W.K., So, S.S., Sakurai, N., 1998. Symplastic and apoplastic sugar contents in gall tissues and callus of the sumac (*Rhus chinensis* MILL.). *Journal of Plant Biology.* **41,** 135-141.

Yusibov, V., Hooper, D.C., Spitsin, S.V., Fleysh, N., Kean, R.B., Mikheeva, T., Deka, D., Karasev, A., Cox, S., Randall, J., 2002. Expression in plants and immunogenicity of plant virus-based experimental rabies vaccine. *Vaccine.* **20,** 3155-3164.

Zhang, S. & Fauquet, M., 1998. Expression and inheritance of multiple transgenes in rice plants. *Nature biotechnology.* **16,** 1060 - 1064.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W., 2000. A greedy algorithm for aligning DNA sequences. Journal of Computational Bioliology. **7, (**1-2):203-14.

Zhou, Y., Aran, J., Gottesman, M.M., Pastan, I., 1998. Co-expression of human adenosine deaminase and multidrug resistance using a bicistronic retroviral vector. *Human Gene Therapy.* **9,** 287-293.

Ziegler, M., Thomas, S., and Danna, K., 2000. Accumulation of a thermostable endo-1,4-b-D-glucanase in the apoplast of *Arabidopsis thaliana* leaves. *Molecular Breeding.* **6**, 37-46.