Durham E-Theses

Isolation of phloem specific gene promoters for use in genetic engineering of insect resistance in rice

Ming-Bo Wang

How to cite:

Wang, Ming-Bo (1994) Isolation of phloem specific gene promoters for use in genetic engineering of insect resistance in rice. Doctoral thesis, Durham University.

Use policy

Isolation of Phloem Specific Gene Promoters for Use in Genetic Engineering of Insect Resistance in Rice

By

Ming-Bo Wang

(Graduate Society; B. Sc., Beijing University, China;

M. Sc., Chinese Academy of Sciences, China)

A thesis submitted for the Degree of Doctor
of Philosophy in the University of Durham

Department of Biological Sciences

June, 1994

2 7 JUN 1994

To my wife and my late parents

MEMORANDUM

Parts of this work have been included in the following publications:

Ming-Bo Wang, Donald Boulter and John A. Gatehouse (1994) Characterization and sequencing of cDNA clone encoding the phloem protein PP2 of *Cucurbita pepo*. *Plant Molecular Biology* 24: 159-170.

Ming-Bo Wang, Donald Boulter and John A. Gatehouse (1992) A complete sequence of the rice sucrose synthase-1 (*RSs1*) gene. *Plant Molecular Biology* 19: 881-885.

Ying Shi, Ming-Bo Wang, Kevin S. Powell, E.Van Damme, Vaughan A. Hilder, Angharad M.R. Gatehouse, Donald Boulter and John A. Gatehouse (1994) Use of the rice sucrose synthase-1 promoter to direct phloem-specific expression of β-glucuronidase and snowdrop lectin genes in transgenic tobacco plants. *Journal of Experimental Botany* 45: 623-631.

ABSTRACT

Towards the aim of producing transgenic rice with enhanced resistance to one of its phloem sap-sucking insect pests, the brown planthopper (BPH), two potential phloem-specific promoters, of the rice sucrose synthase-1 (*RSs1*) and the cucurbit phloem protein PP2 genes, were isolated and investigated.

Using a PCR fragment of the maize sucrose synthase-1 (*Sh1*) gene, genomic clones containing the *RSs1* and *RSs2* (rice sucrose synthase-2) genes were isolated from a genomic library of rice (*Oryza sativa* L. Japonica) and characterised. A full-length *RSs1* gene from one of the genomic clones was sequenced, including 1756 bp of 5'-flanking sequence and 710 bp of 3'-flanking sequence. The gene had an identical intron-exon structure (16 exons and 15 introns) to the maize *Sh1* gene. The *RSs1* 5'-flanking region contained a number of promoter-like sequences, including putative *cis*-acting elements homologous to those found in several endosperm-specific, anaerobiosis-inducible, or phloem-specific promoters.

The *RSs1* promoter region, including 1.9 kb 5'-flanking sequence, the first intron, the first exon, and the translational start codon, was fused with coding sequences for β-glucuronidase (GUS) and snowdrop (*Galanthus navalis*) lectin (GNA). Tobacco plants were transformed with these chimaeric genes in order to determine the expression pattern directed by the *RSs1* promoter. Histochemical and immunochemical assays demonstrated that the expression of both GUS and GNA was restricted to phloem cells in various tissues (i.e. stem, root, leaf, petiole) and in different transformants. In addition, GNA was detected by immunological assay in the honeydew excreted by peach potato aphids (*Myzus persicae*), also a phloem sap-sucking insect, feeding on *RSs1*-GNA transgenic tobacco plants. This provided direct evidence that GNA was not only expressed in the phloem tissue, but was also present in the phloem sap of transgenic tobacco plants. Since GNA has been shown to have antimetabolic effect against BPH, the *RSs1*-GNA construct is being used to transform rice plants by collaborating groups elsewhere.

In order to isolate the phloem protein PP2 gene promoter, the PP2 polypeptide from 3 months old *Cucurbita pepo* plants was partially sequenced after *in situ* cleavage with cyanogen bromide vapour, giving 75 residues of sequence on two cyanogen bromide fragments. Using an oligonucleotide probe based on this amino acid sequence, cDNA clones encoding PP2 were isolated from a *C. pepo* cDNA library constructed from mRNA of 3-5 days old seedling hypocotyls. A cDNA clone was used as probe to screen a *C. pepo* genomic library, and several genomic clones were isolated. Restriction mapping showed that these clones contained different genes, consistent with results from Southern blots of *C. pepo* genomic DNA probed with PP2 cDNA, in which multiple bands were detected in all restriction endonuclease digests. One of the clones was partially sequenced, and was shown to contain a gene encoding PP2. A 1.2 kb 5'-flanking region of this clone was fused with a GUS reporter gene, and this construct was used to transform tobacco. Initial histochemical analysis showed that this chimaeric gene was not expressed in the putative transgenic plants examined. Possible reasons for this failure are discussed.

ACKNOWLEDGEMENTS

ABBREVIATIONS

ATP	adenosine-5'-triphosphate
6-BAP	6-benzylaminopurine
bp	base pair
BSA	bovine serum albumin
cDNA	complementary DNA
C-terminus	carboxy terminus
dATP	deoxyadenosine-5'-triphosphate
dCTP	deoxycytidine-5'-triphosphate
dGTP	deoxyguanosine-5'-triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dTTP	deoxythymidine-5'-triphosphate
EDTA	ethylene diamine tetra-acetic acid
g	gram
hr(s)	hour(s)
HRP	Horseradish peroxidase (conjugate)
In.	inch
IPTG	isopropyl-β-thiogalactopyranoside
kb	kilobasepair
kD	kilodalton
mA	milliampere
mg	milligram
μg	microgram
min	minute
ml	millilitre
μl	microlitre
Mr	relative molecular weight
mRNA	messenger RNA
NAA	α-naphthaleneacetic acid
nm	nanometre
nt	nucleotide
N-terminus	amino terminus
NZ amine	casein hydrolysate enzymatic
O.D.	optical density
oligo	oligonucleotide
ORF	open reading frame

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
poly (A)	polyadenylic acid
RNA	ribonucleic acid
RNase	ribonuclease
SDS	sodium dodecyl sulfate
Tris	tris(hydroxymethyl)methylamine
UV	ultra violet light
UDP	uridine diphosphate
UDPGlc	uridine diphosphoglucose
V	volt
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1.   INTRODUCTION

1.1. Rice production and the rice brown planthopper

Rice (*Oryza sativa* L.) is a cultivated genus of the family *Oryzeae*, Order

Gramineae, the majority of which are grown in standing water. It is a diploid plant

with a haploid nuclear DNA content of about 5.5×10^8 bp (Bennett *et. al*, 1982;

Walbot and Gallie 1991). Three ecogeographic races of *Oryza sativa* have evolved;

Indica is the tropical race, Japonica the temperate race, and Javanica, believed to have

evolved in Indonesia, is intermediate between the other two (Chang 1976).

Rice is the primary food crop for more than a third of the world's population (David

1991), second only to wheat in its importance with respect to world total production.

It is now grown on over 148 million hectares, representing more than 10 percent of the

earth's arable land (IRRI 1994). Current global production is about 520 million tons of

unmilled, rough rice (Bonman *et al.* 1992; IRRI 1994), or approximately 57 kg of milled rice per capita, sufficient to provide 20% of the total global human caloric requirement. Over 90% of all rice is
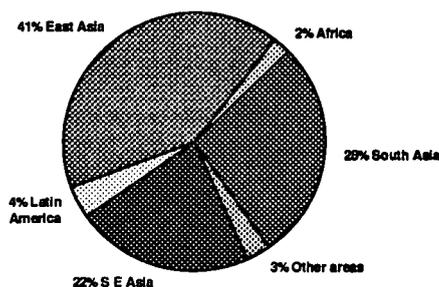
Figure 1.1. Regional distribution of world rough rice production in 1988. (Source: IRRI 1994; David 1991)

produced and consumed in Asia (Fig.1.1), home to 59 percent of the world's

population and rice is also becoming increasingly popular in other geographic regions.

As human population growth in the rice growing regions is the fastest, the need

for an increase in rice production is therefore inevitable. Recent projections indicate a

world rice food need of about 758 million tons in 2025, representing 70 percent more

rice than is consumed today (IRRI 1994). Before the 1970's the required increase in

rice production could be achieved largely by expansion of the area used to grow the

crop. However, with fast population growth and diminished opportunities for expanding rice-growing areas, the growth in rice production has now become mainly dependent on increased rice yields (David 1991). This increased future demand can only be met by improvement of rice productivity.

Some of this increased production might be obtained by appropriate insect pest management, as insect damage is one of the major constraints to rice production. There are two major reasons for this. Firstly, crop damage due to insect attack is severe. Early estimates put the loss in rice yield due to insect pests at about 32% in Asia and 21% in North and Central America (Cramer 1967), and these figures have been reflected by more recent data, for example, 35-44% general losses from insects (Pathak and Dhaliwal 1981) and 24% from insects in East and Southeast Asia (Ahrens *et al.* 1982). Secondly, changes in production practices, and more intensive rice cultivation, which are both necessary to achieve increase in production, can worsen insect pest problems (Norton and Way 1990; Bonman *et al.* 1992).

More than 100 species of insects infest and feed on rice in different rice-growing environments, causing varying degrees of crop damage (Saxena and Khan 1989; Pathak and Saxena 1980). Of these, 30 are considered to be economically important (Norton and Way 1990). Among the most economically important rice insect pests are the planthoppers (Bonman *et al.* 1992; Saxena and Khan 1989). These insects belong to the suborder Homoptera, a group of insects which feed on their host by sucking sap. Planthoppers attack rice plants at various growth stages, but preferentially at the reproductive stage of the plants (Norton and Way 1990). Light infestations by planthoppers reduce tillering, plant height, number of productive tillers per plant, and general vigour of the crop, and increase the number of unfilled grains. Heavy infestations can destroy the crop completely by producing the symptom known as "hopperburn" (Bae and Pathak 1970). Apart from the direct feeding damage, planthoppers are virus vectors (Fig.1.2), and, together with rice leafhoppers, are responsible for the transmission of the majority of the presently known virus diseases in rice (Khush 1977; Saxena and Khan 1989).
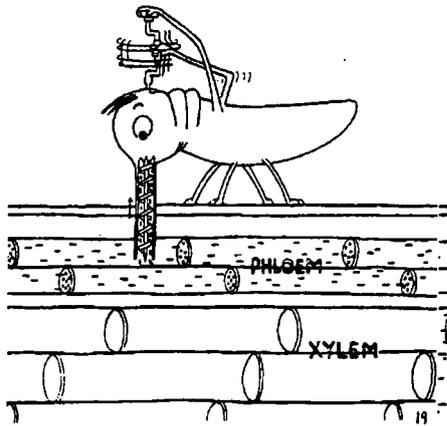
Figure 1.2. Virus transmission by phloem sap-feeding insects. The insect obtains the food from the phloem tissue and with it picks up the virus and transmits it to other plants again by puncturing the phloem (according to Esau 1956).

Several species of planthoppers are major rice pests, including *Nilaparvata lugens* (Stal) (brown planthopper), *Sogatella furcifera* Horvath (whitebacked planthopper) and *Laodelphax striatellus* Fallen (small brown planthopper) (Heinrichs 1994). However, the most destructive of rice planthoppers is the brown planthopper (BPH) (Bonman *et al.* 1992; Gallagher *et al.* 1994). BPH became a major rice pest in the early 1970's, coinciding with the wide-scale adoption of modern high yielding varieties and intensive management practices (Smith 1972; Nickel 1973; Khush 1977), and currently this planthopper remains a major threat in several of the largest rice producing countries, such as China, Thailand, Vietnam and Japan (Mochida 1992; Gallagher *et al.* 1994).

Throughout the year BPH is widely distributed in non-temperate rice growing regions of South, Southeast, and East Asia, the South Pacific Islands, and Australia (Dyck and Thomas 1979; Benrey and Lamp 1994), and adult planthoppers migrate into temperate regions such as Japan and Korea, presumably from mainland China each summer (Dyck and Thomas 1979; Heinrichs 1994). The host plants of BPH are restricted to cultivated rice and a few wild *Oryza* species (Sogawa 1976). BPH has a life cycle of about 25 days (Mochida and Okada 1979) and preferentially infests and multiplies in the basal part of the host plant (Khush 1977). Its eggs are usually laid within the tissue, mainly in sheaths but also in leaf blades. BPH has a specialised feeding behaviour, using its piercing and sucking mouthpart, the stylet, to penetrate the plant tissues and ingest sap from the phloem (Fig.1.3). During sustained feeding, BPH excretes large amounts of liquid called honeydew (Sogawa 1976). The stylet

penetration pathway has been shown to be intracellular, and is accompanied by disruption and degeneration of the plant cell organelles, presumably resulting in cell death (Spiller 1990). Consequently, BPH feeding behaviour is very damaging to the host plant. However, damage caused by BPH is mainly due to high intake of phloem sap and transmission of the stunt virus diseases. The rate of honeydew excretion by a single adult BPH is estimated at 0.5 to 5.0 μl per hour, and it has been shown that 1000 adult females feeding on a rice plant excretes 10 to 15 ml of honeydew per day (Sogawa 1976). BPH is the vector of grassy stunt and ragged stunt virus diseases (Rivera *et al.* 1966; Ling 1977). Stunt disease is the most important virus disease of rice (Khush 1977), and plant damage due to the virus disease can be as

Figure 1.3. The "stylet-sheath" feeding behaviour of BPH. The feeding process can be divided into two main behavioural phases - stylet probing and sucking. The probing is done in the parenchyma outside the vascular bundles through a single point of entry, and ceases when the stylet enters the phloem. During the course a coagulable saliva is secreted along with the stylet penetration to form a tubular lining (the stylet sheath) to the path taken by the stylets, which remains within the plant tissue after withdrawal of the stylets (according to Sogawa and Cheng 1979).

serious as the direct feeding damage. Data on the economic losses due to BPH attack are scarce, but it was estimated that the total loss in the late 1970's due to BPH and the grassy stunt disease was at least US$ 300 million (Dyck and Thomas 1979), and losses for 1990/1991 alone were US$ 30 million in Thailand and Vietnam (Gallagher *et al.* 1994).

To date, control of BPH has been mainly through the application of insecticides and use of resistant rice varieties produced by conventional breeding. Other control methods include biological control and cultural control. Apart from high cost and potential environmental damage, insecticide use has serious consequences. BPH is difficult to kill with insecticides. Reasons for this include the failure of sprays to reach the plant base where the BPH feeds, difficulty in killing eggs inside the plant, presence

of different biotypes of BPH which differ in their susceptibility to insecticides, and overlapping of different generations of BPH during the season (Heinrichs 1979). Insecticide application can result in the resurgence of BPH outbreaks. Resurgence, a significant increase in BPH populations after insecticide treatment, has been widely observed (Heinrichs 1979; Heinrichs and Mochida 1984; Kenmore et al. 1984; Whalon et al. 1990), and it has been a major factor contributing to the increase and severity of BPH outbreaks (Heinrichs and Mochida 1984; Whalon et al. 1990; Gallagher et al. 1994). Several factors have been implicated in the causes of insecticide-induced BPH resurgence (Heinrichs 1979; Heinrichs and Mochida 1984; Gallagher et al. 1994; Roderick 1994), including a decrease in natural enemy populations due to the toxic effects of broad spectrum pesticides on parasites and predators, development of resistance by the BPH to the insecticide through various mechanisms such as enzymatic degradation by oxidases and esterases, changes in the chemical nature of the rice plant affecting BPH nutrition, and insecticide stimulation of BPH oviposition, feeding, or hatching. The destruction of natural enemies and reproductive stimulation are regarded as of major importance (Heinrichs and Mochida 1984; Gallagher et al. 1994). The degree of resurgence is affected by insecticide management practices, such as rate, number, method and timing of application (Heinrichs and Mochida 1984; Roderick 1994). Judicial use of insecticides can reduce planthopper outbreaks and extend the life of a resistant variety in the field (Gallagher et al. 1994).

Biological control refers to the use of natural enemies to reduce field population of insects. At least 200 natural enemies have been recorded for BPH (Chiu 1979; Benrey and Lamp 1994), some of which have shown great potential for controlling pest populations. However, only a limited practical effectiveness has been demonstrated in the field (Benrey and Lamp 1994).

Cultural control, such as simultaneous cropping with suitable rotation is another alternative approach (Oka 1979). However, this would need continuous supervision and legislation, and the wide-scale migration and short life cycle of BPH

5

(Mochida and Okada 1979) and the fact that most rice is grown in small farms (Bonman *et al.* 1992), make cultural control difficult in practice.

The most successful control measure to date for the BPH has been the use of resistant rice varieties produced by conventional breeding. Transfer of resistance into high-yielding rice varieties through crossing with resistant lines and wild rices has been a major research objective in rice growing countries (Saxena and Khan 1989), and the resultant resistant varieties are now grown widely. Major BPH outbreaks which occurred throughout Southeast Asia have now been partly alleviated by this control measure (Norton and Way 1990; Claridge 1990) and in some countries, such as the Philippines, BPH is currently no longer a major rice pest. However, several potential problems exist in the conventional breeding. First, BPH is capable of developing resistance-breaking biotypes that can survive on previously resistant varieties (Claridge and Den Hollander 1980; Roderick 1994), sometimes within a few years of their release (Pathak and Heinrichs 1982; Gallagher *et al.* 1994; Hare 1994) or within even shorter periods (Whalon *et al.* 1990). Secondly, currently used BPH-resistant varieties have a rather narrow genetic base (Saxena and Khan 1989), and despite massive screening programmes, there are only a limited number of genes for BPH resistance identified (Claridge 1990; Hare 1994). Thirdly, there is a shortage of resistant breeding lines in some countries or in some rice varieties. In India and Sri Lanka fewer resistant varieties are available, making this control measure less successful (Claridge 1990). In Japan, a major rice producer, and where most of the rice grown is Japonica varieties, no commercially viable BPH-resistant rice cultivar has appeared until now due to the lack of resistant breeding lines (Mochida 1992). It has been forecast that BPH will be among the top five of rice insect pests in Japan in 2000 (Mochida 1992).

In conclusion, the brown planthopper is one of the most important pests in Asia. To date control of BPH has been mainly through the use of pesticides and resistant rice varieties. However, BPH is capable of adapting to both insecticides and host plant resistance and control of BPH based solely on the use of pesticides or resistant varieties will not be effective in the long term. There is thus a need to find a

6

new control approach. As will be described in the next section, genetic engineering will bring new tools to insect pest control, and if used wisely, will enable other tools to be used more effectively. As has been anticipated, effective, long-term, BPH control could only be achieved by integrated pest management (IPM), a strategy which utilises a set of different control approaches (Hare 1994). Genetic engineering can potentially play an important role in an IPM strategy by creating a new type of host plant resistance and by augmenting other control approaches such as biological control and chemical control.

1.2. Genetic engineering of insect resistance

The development of a new type of control approach, namely the production of plants with built-in insect resistance, has become a priority target in recent years for several reasons. Pest control is a major expense. Currently, the world insecticide sales are around US$7.5 billion (Gatehouse *et al.* 1993) and even with this expenditure, farmers still report crop losses of up to 13% due to insect pests (Gatehouse *et al.* 1993). Substitutes for pesticides are increasingly being considered environmentally desirable. Plants with "built-in" pest resistance will leave minimum pesticide residues, will eliminate problems of spray drift application, will have minimum adverse effect on beneficial organisms, and will not have problems of limited worker re-entry and groundwater or stream contamination. A third factor is that "built-in" pest control is perhaps the ultimate application strategy. It is convenient for the farmer, will eliminate the task and cost of field scouting and the choices inherent with timing of application of pesticides. Finally, plants that carry their own pest defence system would probably overcome much of the problem of onset of pest resistance, as is common when using chemical insecticides. Introduction of multigene- or multimechanism-resistance into the plant, expression of resistance genes in a specific tissue or at a specific developmental stage, use of antifeedant genes rather than toxin genes, delicate regulation of gene expression levels to give an optimal balance between resistance and selection pressure, and the use of genes against particular insects without affecting the

7

population of their natural enemies (Hilder *et al*. 1992), could all be used to delay the onset of pest resistance.

Despite several potential problems with genetic engineering for resistance, such as the technical difficulty in attaining sufficient level of resistance and the possibility of resistance development by insects (Boulter 1993), the advantages of using this technique in conjunction with conventional breeding in developing "built-in" pest resistance are well-known (Boulter 1993). Genetic engineering can augment conventional breeding by decreasing the time to produce resistant varieties, and by widening the potential gene pool available as genes from any source can be used. The latter can be a crucial factor in breeding resistance to BPH, as only a limited number of BPH resistance genes are available for conventional breeding, and some of these genes have already become ineffective due to the development of resistance-breaking biotypes by BPH (Hare 1994). Whereas classical breeding programmes have usually set goals for a single new resistance factor, one of the advantages of genetic engineering is that it may be possible to engineer crops with several combined new potent resistance factors, since this technology allows the efficient transfer of several new genes to the same host crop variety. The relative precision of recombinant DNA techniques means that it is possible to create delicately regulated expression of insect resistance genes in plants.

There are three major steps of genetic engineering for insect pest resistance, i.e. identification of insect resistance genes, transformation and regeneration of plants, and expression of the intended trait by the new gene. In addition to these, at some stage, field evaluation of the new plant must be carried out, and, if the plant variety used to obtain tissue for laboratory manipulations was not from a line that will be desirable ultimately for the farmers, the genetically engineered seed derived from the research must be crossed with appropriate breeding stock in order to attain a marketable product.

One way to identify desired resistance genes is to screen currently available single gene products using an artificial diet system into which proteins encoded by the

resistance genes are incorporated. An alternative way is to over-express the genes in systems like tobacco and challenge the plants with the insects to be tested which feed on them. Both have proven successful, although problems exist in that these systems sometimes do not reflect the real situation of commercially used plant lines in the field.

The gene that has received the most attention in the genetically engineered insect resistance area is the gene encoding the crystal toxin (δ-endotoxin) produced by the bacterium *Bacillus thuringiensis* (*Bt*) on its sporulation. *Bt* has been used as a biopesticide in field spray form for nearly 30 years. In addition, transgenic plants (tobacco and tomato) expressing the *Bt* toxin gene, first reported in 1987, showed substantial resistance to lepidopteran insects (Fischhoff *et al.* 1987; Vaeck *et al.* 1987; Barton *et al.* 1987). Many more *Bt*-expressing transgenic crop plants have been generated since. Some of them, including cotton, maize, tobacco, tomato, potato and walnut plants expressing engineered *Bt* genes have been tested in field trials (Kareiva 1993; Boulter 1993), and several of these transgenic plants are expected to be released commercially in the next few years (Horsch 1993).

Although the *Bt* gene has shown great potential in genetic engineering of insect resistance, it has some limitations. *Bt* toxin genes with specificity against a wide variety of insects have been identified, but to date these have been restricted to lepidopteran, coleopteran and dipteran insects. *Bt* genes with toxicity towards sap-sucking insects (i.e. homopteran insects) have not been identified. The homopteran insects, which include the rice brown planthopper, are responsible for a large proportion of insecticide use worldwide (Fig.1.4). Although a *Bt* gene against rice yellow stemborers has been identified and expressed in transgenic rice (Fujimoto *et al.* 1993), no other *Bt* genes had been discovered that show adequate toxicity to the remaining major rice insect pests (Norton and Way 1990). Furthermore, it has been reported that some strains of insects could develop resistance to *Bt* (McGoughey 1985; Stone *et al.* 1989; McGoughey and Whalon 1992). Therefore, dependence on this single source of insect resistance genes will not be adequate in engineering resistance to the wide range of insect pests and in acquiring durable resistance. There are also

technical difficulties in using *Bt* genes. The bacterial version of *Bt* genes cannot be efficiently expressed in plants, and modifications of the genes such as truncation and changing the genetic codon usage (Fischhoff *et al.* 1987; Perlak *et al.* 1990) have been necessary, which has proved to be costly and time-consuming.
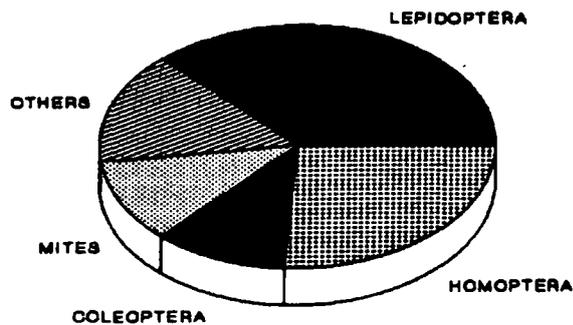
Figure 1.4. Insecticide usage by pest insect order. (Source: Hilder *et al.* 1992)

Another group of genes with great potential in engineering insect resistance are those derived from the plant kingdom. Several types of plant proteins show antimetabolic effects towards insects and some of them, such as the proteinase inhibitors, are believed to play a protective role (Green and Ryan 1972; Ryan 1978; Boulter *et al.* 1990). Genes encoding these proteins are therefore being explored for use in genetic engineering of insect resistance. Several genes, such as CpTI, encoding a cowpea trypsin inhibitor (Hilder *et al.* 1987), and the genes encoding proteinase inhibitor II from tomato and potato (Johnson *et al.* 1989), have been successfully expressed in tobacco and the resultant plants exhibited substantial resistance to insect pests. Many more plant-derived insect resistance genes have been identified using the artificial diet and/or model plant systems. These include genes encoding various proteinase inhibitors, α-amylase inhibitors, lectins, and ribosome inactivating proteins as well as enzymes lipoxygenases and chitinases (Gatehouse *et al.* 1992; Hilder *et al.* 1992; Boulter 1993). Recently, several plant proteins have been shown to exhibit significant antimetabolic effects towards some homopteran insects, including BPH, GLH (rice green leafhopper), aphids, and potato leafhoppers (Powell *et al.* 1993; Rahbe and Febvay 1993; Habibi *et al.* 1993). They are the first proteins found to be effective against homopterans. Plant-derived resistance genes are usually less toxic than *Bt* genes, but they show great potential as control agents for several reasons. They provide a wide

range of mechanisms of resistance, as exemplified by the discovery of a range of gene products toxic to BPH (Powell *et al*. 1993). In contrast to the *Bt* gene, plant-derived genes have minimal problems associated with codon usage, mRNA stability, and protein processing, and they can be readily expressed in transgenic plants (Hilder *et al*. 1987; Edwards *et al*. 1991). These genes have evolved in the plant concomitantly with insects (Boulter *et al*. 1990), and their resistance is based on various different mechanisms (Hilder *et al*. 1992), so they may be less susceptible to adaptation by insects and therefore the introduction of different plant-derived resistance genes into one crop could possibly result in durable resistance (Hilder *et al*. 1992). Finally, the low toxicity can be advantageous in some respects. They will pose less damage to animals and beneficial predator insects, and low lethal effect may be important in balancing the resistance level and selection pressure. Nevertheless, the level of gene expression could be enhanced to improve resistance by engineering gene regulation.

A major step in genetic engineering has been plant transformation. To date, most crop plants can be transformed although wheat transformation is still not routine. Presently available plant transformation systems include *Agrobacterium tumefaciens* Ti-plasmid mediated transformation, electroporation, PEG-mediated transformation, microinjection, and particle bombardment. The first is the oldest and best established system and has proven versatile and reliable in transforming dicotyledonous plants such as tobacco, tomato, potato, petunia and sunflower, and tobacco transformation using this system has been widely used in the study of new genes and promoters. However, this system has proven generally inapplicable to transform monocotyledonous plants. Electroporation and PEG-mediated transformation, both involving the introduction of DNA into protoplasts and regeneration of transformed cells into whole plants, are the second most successful transformation techniques used so far. They have been used in transforming both dicot and monocot plants, and the majority of transgenic monocotyledonous plants such as rice have been produced using these techniques. However, there are two principal limitations to the use of these techniques. Firstly, it is difficult to produce the viable protoplasts that are critical for later recovery and

11

regeneration. Secondly, the transformation in many cases has only been transient. Therefore, there are still considerable difficulties in routinely using these techniques. The microinjection technique, which may also require the use of protoplasts, was initially tried but has been rarely used in plant transformation. Particle bombardment, which involves coating DNA onto tiny particles and then accelerating them into plant tissue, is a newly developed plant transformation technique. A major advantage of this technique appears to be its applicability to a wide range of species and varieties because it is not dependent on *Agrobacterium* and it avoids many of the most time- and labour-intensive parts of tissue culture regeneration. Some plant species have already been transformed by this technique, and recently, successful transformation of barley (Wan and Lemaux 1994) and wheat (Vasil *et al.* 1994), species that are very difficult to transform using other techniques, has been reported. However, further development is still needed for this technique. Overall, although plant transformation techniques have been a constraint in genetic engineering, it appears that, with the development of currently available techniques, the barriers which have to date prevented many plant species from being transformed routinely will be overcome in the near future.

Another major goal in genetic engineering is to achieve desired expressions of new genes. In the case of genetic engineering of insect resistance, the level, timing, and localisation of the gene expression are sometimes essential if desired resistance to target insects is to be attained in transgenic plants. This requires appropriate promoters along with some other factors such as compatible codon usage. Compared with the isolation of insect resistance genes, much less effort has been devoted to this area.

1.3. Promoters

From the technical viewpoint, genetic engineering requires two immediate elements: genes which encode desired traits, and promoters that lead to desired expression of the traits in the target organisms. The term "promoter" referred to here

represents all the non-translational regulatory DNA sequence which is essential for controlling a gene's expression. The promoter responsible for the production of the 35S RNA from cauliflower mosaic virus (CaMV) is the most widely used in plant genetic engineering. The CaMV 35S promoter has been extensively studied and its enhancer sequence is found to contain different functional domains which exhibit combinatorial and synergistic properties (Kay *et al.* 1987; Benfey *et al.* 1989; Fang *et al.* 1989; Benfey *et al.* 1990a, 1990b). It is functional in both dicot and monocot plants, and is a relatively strong promoter in many species. However, the efficiency of this promoter is generally higher in dicot than in monocot plants (Peterhans *et al.* 1990; Bhattacharyya-Pakrasi *et al.* 1993). Histochemical analyses show that the 35S promoter is relatively constitutive, with certain levels of tissue preference (Jefferson *et al.* 1987; Battraw and Hall 1990; Terada and Shimamoto 1990). The expression directed by 35S promoter is relatively high in leaf tissues, so it is favoured for expression of proteins with insecticidal activity towards chewing insects. Some other relatively constitutive promoters have also been studied, including the *Agrobacterium* nopaline synthase (nos) and octopine synthase (ocs) promoter (Dekeyser *et al.* 1989), the rice light harvesting chlorophyll a/b-binding protein (LHCP) promoter (Tada *et al.* 1991), and a rice actin promoter (Zhang *et al.* 1991). All these have the potential for use in engineering resistance to chewing insects.

For engineering of insect resistance, however, a finely-tuned expression of a resistance gene might be more desirable. It would be ideal if an insecticidal gene could be modulated in such a way that it is expressed in response to particular signals, such as wounding or light, in a particular tissue where the insects attack, or at a particular developmental stage when the plant is most vulnerable to insect attack. These types of highly regulated expression has been suggested as one of the potential strategies for developing durable resistance (Bonman *et al.* 1992). Such promoters would also be more desirable than constitutive ones if the latter were shown to result in a yield penalty.

13

One way of obtaining such highly regulatory promoters is to modify already-available promoters, as demonstrated by Benfey *et al.* (1990a, 1990b) who constructed promoters with different tissue specificity by various combinations of the CaMV 35S promoter subdomains. However, to date the main source for such promoters is the plant itself. Some have already been isolated, such as the wound-inducible (Thornburg *et al.* 1987; Keil *et al.* 1989; Johnson *et al.* 1990) and the root-specific promoters (Hall *et al.* 1993). A major problem with the use of plant-derived promoters is that so far very little has been known about their regulatory mechanisms. Unlike the 35S promoter, which appears to function quite consistently in different species, plant-derived promoters appear to be rather variable between monocot and dicot species or between homologous and heterologous systems in terms of the level and pattern of gene expression. On the one hand, many plant promoters can be efficiently expressed in heterologous species as well as in homologous species, and on the other, the situation can be entirely different. A dicot promoter can lose its whole activity in a monocot plant (Zheng *et al.* 1991) or *vice versa* (Keith and Chua 1986; Ellis *et al.* 1987); the expression pattern of a monocot promoter can be changed when introduced into a heterologous monocot plant (Kyozuka *et al.* 1991); the activity of a dicot promoter can be much lower than that of a monocot promoter in its homologous monocot species (Kyozuka *et al.* 1993). Another factor associated with the efficiency of a plant-derived promoter is the effect of introns on gene expression. Results from transient expression have shown that insertion of monocot introns into a transcription unit can greatly enhance the expression of reporter genes with a variety of promoters in the protoplasts of various monocot species (Callis *et al.* 1987; Vasil *et al.* 1989; Maas *et al.* 1991; Hensgens *et al.* 1993), but insertion of the maize *Sh1* first intron inhibits the expression of a reporter gene in the protoplasts of tobacco, a dicot plant (Maas *et al.* 1991). A recent report shows that incorporation of the first intron from the rice actin 1 gene significantly enhanced the activity of a potato inhibitor II gene promoter in transgenic rice plants without affecting the wound-inducible property of the promoter (Xu *et al.* 1993). The mechanism of such enhancement by introns is still

unknown, although it has been observed that introns act post-transcriptionally (Maas *et al.* 1991) and it is suggested that pre-mRNAs containing introns may be shunted through a processing pathway which, as a result of interaction with the spliceosome or splicing *per se*, or both, may result in more efficient capping, polyadenylation, or nucleocytoplasmic transport of these transcripts, thus increasing the level of mature cytoplasmic mRNA (Walbot and Gallie 1991; Maas *et al.* 1991). It is also not clear whether introns always have a similar enhancing effect *in planta,* or whether they exert an inhibiting effect when introduced into a heterologous dicot plant, since to date most of the results come from transient expression with monocot species. Stimulatory effects on gene expression have also been observed with exon sequences (Maas *et al.* 1991; Maas *et al.* 1990), and in addition, promoter elements have been identified from inside the transcription units of some genes, such as the 5S RNA genes of *Xenopus laevis* (Geiduschek and Tocchini-Valentini 1988). Furthermore, it is suggested that the 3' non-coding sequence of a gene may also be important in the normal expression pattern of a plant-derived promoter (Hall *et al.* 1993; Larkin *et al.* 1993).

The complexity of regulatory mechanisms, and the lack of understanding of them at the present time, make it preferable, in some people's view, to use homologous rather than heterologous promoters for genetic engineering. Indeed, accumulated data on transgenic studies appear to show that homologous promoters function rather consistently and efficiently (Zhang *et al.* 1991; Hensgens *et al.* 1993; Tada *et al.* 1991). A typical example is the rice LHCP promoter (Tada *et al.* 1991). The spatial pattern and the light inducible property of the LHCP-GUS (β-glucuronidase) chimaeric gene in transgenic rice is characteristic of the rice LHCP gene itself, and the expression level in green tissue of the transgenic plant is 10 times higher than that with the generally-strong CaMV 35S promoter.

However, in the long term, the incompatibility among different species with regard to promoter activity may not be a problem. As shown by the example given above, the dicot potato proteinase inhibitor-II promoter alone can not be efficiently expressed in rice, but after incorporation of the rice actin intron its activity is enhanced

to a level comparable to that of the strong actin promoter while still maintaining the wound-inducible expression pattern (Xu *et al.* 1993). As more knowledge on gene regulation is obtained in the future, it will be possible to construct a desired promoter by modification of those already available promoters.

As far as this particular project is concerned, two factors have to be considered in choosing promoters. First, the insecticidal gene needs to be efficiently expressed in rice; secondly and more importantly, the insecticidal product encoded by the gene should be present in the phloem sieve elements, as BPH is a phloem sap feeder. A strong and phloem-specific promoter is therefore most desirable. As already mentioned, confining high-level expression of resistance genes to the tissues of actual insect attack will not only conserve the metabolic energy and building blocks of plant cells but may also minimise the selection pressure on insects. Several promoters have been used in transgenic rice, including the CaMV 35S, maize alcohol dehydrogenase (*Adh-1*) (Kyozuka *et al.* 1991), rice actin (Zhang *et al.* 1991), rice small-subunit ribulose-1,5-bisphosphate carboxylase (*rbcS*) (Kyozuka *et al.* 1993), rice LHCP (Tada *et al.* 1991) and maize zein (Dekeyser *et al.* 1989). Of these, the 35S promoter appears to give the best and most reliable expression. These promoters, however, do not have the required specificity for the present project. Although the 35S promoter is shown to be more active in vascular tissue including the phloem as a whole, histochemical studies indicate that its activity is low in the particular type of cell targeted for expression, i.e. the sieve elements in transgenic rice plants (Battraw and Hall 1990; Terada and Shimamoto 1990). Even if only the expression level rather than expression pattern is of concern, the 35S promoter is far from being the most suitable. Several plant-derived promoters have already shown activities in transgenic rice or rice protoplasts higher than, or equivalent to, the 35S promoter (Marcotte *et al.* 1988; Tada *et al.* 1991; Hensgens *et al.* 1993). It may therefore be possible to isolate a highly expressed, and phloem-specific, promoter from the plant. Among the potential candidates are the sucrose synthase-1 and the phloem protein PP2 gene promoters, as they are either already shown to direct highly phloem cell-specific expression in

transgenic plants (in the case of sucrose synthase promoter) or the protein encoded by the gene is highly abundant specifically in the phloem (in the case of PP2).

1.4. Sucrose synthase and its genes

The enzyme

Sucrose synthase (SS) (UDP-glucose: D-fructose 2-glucosyltransferase, EC 2.4.1.13) was first characterised in wheat germ (Cardini *et al.* 1955) and later in many other plant species such as maize (Chourey and Nelson 1976, Chourey 1981; McCormick *et al.* 1982), rice (Chan *et al.* 1990; Elling and Kula 1993), tomato (Sun *et al.* 1992) and *Vicia faba* (Ross and Davies 1992). Whereas in the monocotyledonous plants examined, such as maize (Chourey and Nelson 1976, Chourey 1981; McCormick *et al.* 1982; Nguyen-Quoc *et al.* 1990), *Sorghum* (Chourey *et al.* 1991b), sugarcane (Buczynski *et al.* 1993), and rice (Chan *et al.* 1990), two or different isoforms of SS with similar biochemical characteristics have been reported, enzyme purification indicates only one isozyme in soybean nodules (Morell and Copeland 1985), *Phaseolus* seedlings (Delmer 1972), *Vicia faba* cotyledons (Ross and Davies 1992) and wild tomato fruit (Sun *et al.* 1992). Both *in vitro* assay (Cardini *et al.* 1955) and *in vivo* analysis (Geigenberger and Stitt 1993) show that SS catalyses a reversible reaction,

$$\text{sucrose} + \text{UDP} \leftrightarrow \text{UDPGlc} + \text{fructose},$$
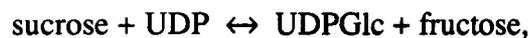
but it is suggested that SS functions mainly in the direction of sucrose breakdown *in vivo* (Avigad 1982; Geigenberger *et al.* 1993; Misra *et al.* 1994). The physiological roles of SS are still not clear, but the enzyme has been implicated in several important metabolic processes in plants, including starch synthesis (Chourey and Nelson 1976; Heim *et al.* 1993; Wang *et al.* 1993), cell wall synthesis (Chourey *et al.* 1991a; Hendrix 1990) and phloem transport of photoassimilate, and it is used as an indicator of sink strength (Sun *et al.* 1992; Wang *et al.* 1993). All of these are associated with its basic catalytic role in the breakdown of sucrose in plants. The clearest evidence for its role in starch synthesis came from the investigations on maize SS. The loss of the

maize SS1 protein, as in the case of the *shrunken* (*Sh*) mutant, resulted in a sharp (60%) decrease in starch content in mature maize kernels (Chourey and Nelson 1976). Other evidence which shows that the increase in the amount of starch coincides with the increase in SS activitiy (Heim *et al*. 1993; Wang *et al*. 1993) and that products of the SS-catalysed reaction could be utilised directly in starch synthesis (Wang *et al*. 1992), also support this hypothesis. Apart from acting as precursors for starch synthesis, the products from the cleavage of sucrose by SS may also be used for cell wall synthesis (Hendrix 1990; Maas *et al*. 1990). SS is present in all plant tissues, but its activity is found to be present in low quantities in mature green leaves (source organ) (Nguyen-Quoc *et al*. 1990; Buczynski *et al*. 1993) and higher in sink tissues, such as endosperm (Chourey and Nelson 1976), fruit (Sun *et al*. 1992), young expanding leaves (Nguyen-Quoc *et al*. 1990; Claussen *et al*. 1985b; Buczynski *et al*. 1993), and nodules (Zammit and Copeland 1993; Gordon *et al*. 1992; Kuster *et al*. 1993). Furthermore, SS activity is found to be positively correlated with the growth· rate of tomato fruit (Sun *et al*. 1992; Wang *et al*. 1993). These tissues are sucrose utilisation sinks in which hexose-P is obtained from the breakdown of imported sucrose. Although acid invertase (EC 3.2.1.26) and SS both catalyse the cleavage of sucrose into hexose, SS is found to be the dominant enzyme in metabolising imported sucrose, at least in some plants (Sun *et al*. 1992; Wang *et al*. 1993). Therefore, SS activity is regarded by some as an indicator of sink strength (Sun *et al*. 1992). In recent years, much interest has been given to the potential role of SS in the transport of sucrose in phloem. Sucrose is the major transport form of assimilate on its way through phloem from photosynthetic organs to sink organs. The uptake and transport of sucrose in the phloem requires energy. It has been suggested that SS is involved in the supply of energy for phloem loading in source tissues (Martin *et al*. 1993) and of energy probably needed along the entire transport route (Geigenberger *et al*. 1993; Nolte and Koch 1993b). It has also been suggested that SS may be needed in the phloem to generate uridine diphosphoglucose (UDPGlc) for callose synthesis in response to wounding (Geigenberger *et al*. 1993; Nolte and Koch 1993b). Callose has

18

been implicated in the plugging of sieve plates following injury to phloem (Wolf *et al.* 1991). Another related hypothesis is that SS is required for the maintenance of an equilibrium between phloem sucrose and its breakdown products (Claussen *et al.* 1985a, 1985b) and for the maintenance of a descending sucrose gradient considered favourable for sucrose exit from the phloem (Ho 1988). The involvement of SS in phloem transport is supported by consistent observations that SS is predominantly localised in the vascular tissues (Claussen *et al.* 1985b; Chen and Chourey 1989; Lowell *et al.* 1989; Tomlinson *et al.* 1991; Nolte and Koch 1993b), and by the fact that SS activities have been detected in the phloem exudate (Lehmann 1973). Recently, Geigenberger and Stitt (1993) suggested that SS is particularly suitable for cells which receive sucrose directly from the phloem by showing that SS catalyses a readily reversible reaction *in vivo*. Their arguments are that, firstly, the sucrose concentration in the phloem is similar to the affinity of SS enzyme; and secondly, the high reversibility of sucrose breakdown reaction catalysed by SS is well suited to allow the rate of sucrose breakdown in the phloem to respond rapidly to changes in the metabolic requirement for ATP (source of energy for phloem loading and transport), and for UDPGlc during callose production, while constantly maintaining high and stable level of sucrose in the phloem for transportation requirement. More recently, Geigenberger *et al.* (1993) showed that artificial supply of different sugars (i.e. glucose, fructose, sucrose) to the cotyledons of germinating castor-bean resulted in changes in the levels of various carbohydrates and glycolytic metabolites, including sucrose, fructose, UDPGlc and UDP, the reactants of SS, in the phloem exudate from cut hypocotyls; but the molar mass action ratio estimated from the measured levels of these reactants remains very close to the theoretical equilibrium constant (K_{eq}) of the SS-catalysed reaction. This clearly indicates that SS-catalysed sucrose breakdown reaction operates in the phloem or the SS is present in the phloem complex.

The genes

The genes encoding the sucrose synthase enzyme have been extensively studied in recent years, and many of them have been isolated and sequenced (Werr *et al.* 1985; Salanoubat and Belliard 1987; Maraña *et al.* 1990; Yu *et al.* 1992; Wang *et al.* 1992; Chopra *et al.* 1992; Arai *et al.* 1992; Heim *et al.* 1993; Martin *et al.* 1993). These data show that there are two SS genes in monocotyledonous plants such as maize (Chourey 1981; McCormick *et al.* 1982), wheat (Maraña *et al.* 1988a), barley (de Ilarduya *et al.* 1993), and *Sorghum* (Chourey *et al.* 1991b). These two genes, generally designated as *Ss1* and *Ss2*, are located at different loci within the same or different chromosomes (McCarty *et al.* 1986; Maraña *et al.* 1988a, 1988b; de la Hoz *et al.* 1992), and the *Ss1* sequence is more closely related to the *Ss1* sequence from other species than to the *Ss2* sequences from the same species, and *vice versa*. Based on this it is suggested that a gene duplication followed by divergent evolution of an ancestral gene took place prior to the evolutionary separation of these plant species (de Ilarduya *et al.* 1993). An exception among the examined monocot species is the rice plant, from which three different SS cDNAs have been isolated (Wang *et al.* 1992). From initial studies on the SS genes from dicot species it is suggested that only one type of SS gene is present (Heim *et al.* 1993). However, the results on which this suggestion was made were not unambiguous (Chopra *et al.* 1992; Heim *et al.* 1993). Recently, a second SS gene has been isolated from *Arabidopsis* (Martin *et al.* 1993); this gene exhibits higher homology to the maize *Sh1* (*Ss1*) and the potato SS gene (Salanoubat and Belliard 1987) than to the previously isolated *Arabidopsis* SS gene (Chopra *et al.* 1992), which is very similar to the monocot SS genes. This clearly indicates that, at least in some dicot species, two or more SS genes are present.

Both the monocot and the dicot SS genes investigated contain a multi-intron/exon structure with some variation in the number of introns and exons and generally with a relatively large first intron (Werr *et al.* 1985; Yu *et al.* 1992; Chopra *et al.* 1992; Martin *et al.* 1993). The SS genes from different species contain open reading frames with similar sizes, ranging from 2406 bp in maize *Sh1* gene (Werr *et al.*

1985), 2412 bp to 2439 bp in the dicot SS genes studied so far (Salanoubat and Belliard 1987; Chopra *et al.* 1992; Arai *et al.* 1992; Heim *et al.* 1993; Martin *et al.* 1993), to 2448 bp in *Ss2* of rice (Yu *et al.* 1992) and barley (de Ilarduya *et al.* 1993).

The SS genes in general are subject to regulation by various developmental and environmental stimuli, such as anaerobiosis, cold, and sugar. The two SS genes in monocot species are also differentially regulated. Many of the studies on SS gene expression has been focused on the maize *Sh1* and *Sus1* (*Ss2*) genes. The expression of both genes is dependent on endogenous and environmental factors. As opposed to *Sus1*, the *Sh1* gene is strongly induced under anaerobic conditions (Rowland *et al.* 1989). The induction occurs at the transcriptional level (Frommer and Starlinger 1988) and is subject to posttranscriptional regulation, i.e. anaerobic treatment leads to significant elevations of *Sh1* transcripts without a concomitant increase in the SS1 protein (Taliercio and Chourey 1989). Similar regulation is found in the case of homologous genes in rice (Richard *et al.* 1991), wheat (Maraña *et al.* 1990), potato (Salanoubat and Belliard 1989), *Sorghum* (Chourey *et al.* 1991b), and *Arabidopsis* (Martin *et al.* 1993). However, in *Sorghum* significant induction by anaerobic treatment is observed for both *Ss1* and *Ss2*, and in rice and *Arabidopsis*, the increase in RNA levels leads to increased SS activity or SS protein concentration. Apart from anaerobic conditions, the SS genes can be induced by other environmental factors such as exposure to low temperature (Maraña *et al.* 1990; Crespsi *et al.* 1991; Martin *et al.* 1993). It has also been found that the level of SS enzyme and mRNA is dependent on light/dark regime (Springer *et al.* 1986; Nguyen-Quoc *et al.* 1990; Maraña *et al.* 1990; Gordon *et al.* 1993; Misra *et al.* 1994) and wounding (Claussen *et al.* 1985a; Salanoubat and Belliard 1989). Furthermore, SS genes are regulated by carbohydrate supply. It has been shown that the maize *Sh1* gene expression is repressed by increasing concentration of sugars and enhanced under conditions of limited sugar supply, whereas the *Sus1* gene is upregulated by high sugar concentration (Maas *et al.* 1990; Koch *et al.* 1992; Nolte and Koch 1993a). Through the use of a *Ss2* cDNA probe of maize, the rice SS gene expression in embryo is found to be strongly induced

by both the hexoses glucose and fructose, and by the disaccharides maltose and sucrose (Karrer and Rodriguez 1992). The results appear to be more complex for the dicot SS genes investigated. The *Asus1* promoter from *Arabidopsis* is down-regulated in transgenic *Arabidopsis* by high sucrose concentration (Martin *et al.* 1993), similar to the *Sh1* gene. However, the transcript levels of the potato (Salanoubat and Belliard 1989) and the *Vicia faba* (Heim *et al.* 1993) SS genes increase with increasing sucrose concentrations, which seems different from the *Arabidopsis Asus1* gene. A possible explanation for this difference is that the potato and *Vicia faba* SS genes (the only type of SS genes identified to date in these species) may be more closely related to type-2 SS genes of monocot species with regard to carbohydrate regulation, whereas the *Asus1* gene happens to be *Ss1* type (two SS genes have been isolated from *Arabidopsis*, see above). However, another experiment on the *V. faba* SS (Heim *et al.* 1993) showed that its activity increased concomitantly with decreasing concentrations of soluble fructose and glucose measured in plant extracts (Heim *et al.* 1993), which appears to be contradictory to results for the monocot *Ss2* genes. This may be interpreted by the likely different cellular distribution of fructose and glucose in tissues from normally grown plants to that in tissues under *in vitro* treatment (i.e. in the study of rice and maize *Ss2* genes, the tissues were excised from the whole plant and incubated in culture media containing different concentrations of sugars before being examined for SS expression).

The differentiated regulation of SS genes leads to distinct temporal and spatial expression patterns. Again, the maize genes have been more extensively studied than others in this respect. Within developing maize kernels, the *Sh1* gene is expressed only in the endosperm, whereas *Sus1* is found in almost all tissues of the kernel with expression levels specific for cell type and developmental stage (Heinlein and Starlinger 1989). Although both of the two SS enzymes are abundant in endosperm tissue, the *Sh1* and *Sus1* expression are spatially and/or temporally separated, and only SS1 is associated with starch deposition (Chourey *et al.* 1986; Chen and Chourey 1989). A recent report shows that in sugarcane the SS1 enzyme is present in all tissues

examined but SS2 is undetectable in mature internodes and mature green leaves (Buczynski *et al.* 1993). In another two reports, the expression of SS genes in mature leaves (Nolte and Koch 1993b) and young roots (Chen and Chourey 1989) of maize is found to be predominant in the vascular tissues including the phloem. This is consistent with recent evidence demonstrating phloem-specific expression of both the maize *Sh1* promoter and the *Arabidopsis Asus1* promoter in transgenic tobacco (Yang and Russell 1990; Martin *et al.* 1993).

Sucrose synthase promoter

Data from the studies on both SS enzymes and SS genes indicate that SS is predominantly expressed in the phloem tissue. As mentioned above, the maize *Sh1* promoter has been shown to direct highly phloem cell-specific expression of a GUS reporter gene in transgenic tobacco (Yang and Russel 1990). Under its control, GUS expression is only observed in the sieve elements and companion cells of the vegetative tissues, including roots, stems, and leaves. Such phloem-specific expression is also observed in flower and fruit tissues. A more recent report shows that the *Arabidopsis* SS gene *Asus1* promoter also directs a highly phloem-specific expression in transgenic tobacco as well as in transgenic *Arabidopsis* (Martin *et al.* 1993).

The maize *Sh1* promoter used in Yang and Russel's study (1990) includes a 1347 bp upstream nontranscribed sequence, the first exon, first intron and the translational start site of the maize *Sh1* gene (Werr *et al.* 1985). The first intron and first exon appear to be indispensable with regard to expression level. It has been shown by transient expression that the insertion of the *Sh1* first intron into the transcription unit can increase reporter gene expression by up to 100-fold when driven by the CaMV 35S promoter and up to 58 times when driven by intron-less *Sh1* promoter (Vasil *et al.* 1989). It has also been shown that both exon 1 and intron 1 of the maize *Sh1* gene individually stimulate expression of reporter genes in the protoplasts of monocot species including rice if present within the transcription unit (Maas *et al.* 1991). Their stimulatory effects are multiplicative and the combination of

them can result in an enhancement of chimaeric gene expression by up to 1000-fold (Maas *et al.* 1991).

1.5. Phloem protein PP2 and its gene

Phloem

The phloem, together with the xylem, constitutes the continuous conducting system of vascular plant. The xylem functions principally in the conduction of water, and the phloem of products of photosynthesis from the leaves to places where it can be utilised.

Phloem tissue contains sieve tubes (in angiosperm) or sieve cells (in gymnosperm) and unspecialised phloem parenchyma; often thick-walled phloem fibre cells are present; and in majority of species, specialised parenchyma cells are associated with the sieve tubes. In angiosperms these are known as companion cells and in gymnosperms they are termed albuminous cells.

The sieve tube is a longitudinal file of cells, the individual members of which are known as sieve elements. The end walls of the elements are modified to form sieve plates, which are perforated so as to allow the massive rapid flow of nutrient material. The sieve cells in gymnosperms are not aligned in vertical files, and intercommunication takes place through sieve areas in the lateral walls, rarely the end walls.

Sieve elements develop from the procambial and cambial cells. During sieve element development the Golgi bodies, microtubules, and ribosomes disappear after a period of intensive wall synthesis which results in considerable thickening of the wall. The nucleus gradually loses its content and usually no trace of it remains when the sieve tube is mature. Gymnosperms and a few angiosperms retain a very shrunken necrotic nucleus against the wall throughout the life of the sieve tube. During the development, the tonoplast, the membrane surrounding the central vacuole, breaks down and disappears. However, the endoplasmic reticulum, mitochondria, and plastids all have characteristic structures in the mature sieve tube and are usually found against

the walls. Another major aspect of sieve element development is the appearance of slime bodies (P-protein) in young cells followed by dispersal of the P-protein fibrils which form a meshwork in the lumen of the mature sieve tube with a longitudinal orientation.

The companion cells have dense cytoplasm containing a nucleus, endoplasmic reticulum, vacuoles, numerous mitochondria, ribosomes, plastids, and Golgi bodies, and are the only cells to have numerous plasmodesmatal connections to the sieve elements. The companion cells are derived from the same cambial mother cell as the sieve element. Each sieve element has one or more companion cells; they usually run the whole length of their sieve element, but not in all species. It is generally considered that one of the major functions of the companion cell is maintenance of the structures and osmotic integrity of the sieve element, which at maturity has neither nucleus nor ribosome to allow protein synthesis. There are large number of specialised pores between sieve element and companion cell, suggesting that massive symplastic transfer of materials between the two cells may occur through them.

The sieve (plate) pore develops by the enlargement of a simple plasmodesma by a process involving the endoplasmic reticulum and callose. This produces a pore of up to 50 times the original diameter with a callose lining between the plasmalemma and the wall. The pore plasmalemma is continuous with the sieve tube plasmalemma on both side of the pore.

Phloem protein (P-protein)

Phloem cells contain specific proteins, some of which are present in large quantities and visible with both the light and electron microscopes, while others can be demonstrated only with chemical methods such as gel electrophoresis. These proteins are collectively referred to as phloem-specific proteins (Cronshaw and Sabnis 1990). The term "phloem protein" or "P-protein" was originally used to designate all types of specific protein components in the phloem (Cronshaw and Esau 1967; Esau and Cronshaw 1967), but later defined to represent the proteinaceous material in the

phloem which is sufficiently characteristic when observed with the electron microscope (Cronshaw 1975). Such cell inclusions have long been observed and are initially known as slime bodies (Hartig 1854; Wilhelm 1880), but their proteinaceous nature was confirmed rather later (Esau and Cheadle 1965; Cronshaw and Esau 1968a; Kleinig *et al.* 1971). P-protein has been found in all dicotyledonous and many monocotyledonous species examined, but is absent from the gymnosperms and lower vascular plants (Cronshaw and Sabnis 1990). P-proteins are most abundant in the sieve elements, but are also found in companion cells and phloem parenchyma cells (Cronshaw and Esau 1968a, 1968b; Cronshaw 1975). There are several morphological forms of P-protein. In young, differentiating sieve elements, the P-protein component arising in the cytoplasm is present mainly in the form of tightly packed aggregates known as P-protein bodies in which P-protein can be amorphous, tubular, crystalline or filamentous. During the later stages of differentiation of the sieve element, the P-protein bodies usually disperse and eventually fuse in the tonoplast-free lumen of the sieve element, and the different forms of P-protein are generally transformed into filaments or fibrils (Cronshaw and Esau 1968a, 1968b; Kollmann *et al.* 1970; Cronshaw and Sabnis 1990). Between different species there is considerable variation in both the morphology of the P-protein components and the number of P-protein bodies per cell, as well as in the arrangement of the components in the P-protein bodies and in the changes that occur during differentiation (Cronshaw and Sabnis 1990). Although the sieve tubes allow a relatively free movement of the photoassimilate, it has been concluded that P-protein is relatively stationary in the sieve element protoplast and does not move in the assimilate stream (Cronshaw 1975). The origin of P-protein in the cell is not clear, but it has been observed that the initial synthesis of P-protein is closely linked to the presence of free helical polysomes and the formation of P-protein filaments coincides with the disappearance of ribosomes (Behnke 1974).

Until recently, biochemical studies on phloem-specific proteins were restricted to those species from which exudate containing proteins in quantities adequate for

analysis can be collected. Many species of the family Cucurbitaceae bleed relatively copiously from cut stems and incisions into the fruit, and the exudate generally contains a high concentration of proteins (e.g. 20-30 mg/ml) (Walker and Thaine 1971; Read and Northcote 1983a, 1983b), much higher than from other species (e.g. 0.3-2.2 mg/ml) (Esau and Cronshaw 1967; Kennecke *et al*. 1971; Hall and Baker 1972; Fisher *et al*. 1992). Some of these proteins are components of P-protein filaments within the sieve element (Kleinig *et al*. 1971). The cucurbit species are thus particularly suitable subjects for the study of P-protein.

Numerous proteins have been identified from the phloem exudate of cucurbit species by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Weber *et al*. 1974; Sabnis and Hart 1976; Sloan *et al*. 1976). Marked differences have been observed among different genera, species, or even among different cultivars of the same species in terms of numbers, relative quantities and properties of exudate polypeptides and of their molecular mass and isoelectric point (Sloan *et al*. 1976; Sabnis and Hart 1976). There are, however, certain features common to the exudate proteins of most cucurbit species: the number of major proteins are relatively small but they comprise most of the protein content in the exudate; several proteins were of relatively large subunit size (Mr values of 80-130 kD or more); the major components are highly basic with high isoelectric point (Sabnis and Hart 1976); upon exposure to air, the exudate proteins cross react through disulphide bonds, resulting in the gellation of phloem exudate, which can be prevented with thiol reagents (Read and Northcote 1983a).

More detailed biochemical studies have been carried out with the *Cucurbita* species. These species generally contain two major exudate proteins, designated as PP1 and PP2, respectively, both found to be structural components of the P-protein filaments *in vivo* (Read and Northcote 1983a, 1983b; Smith *et al*. 1987). The monomer of PP1 has a Mr of 80 kD by SDS-PAGE and 96 kD by sedimentation-equilibrium ultracentrifugation (Read and Northcote 1983a, 1983b). This protein is capable of reconstituting filaments on oxidation and is responsible for the gellation of the phloem exudate (Kleinig *et al*. 1975). PP2 is present as a dimer of Mr of 48 kD

27

held together by buried disulphide bridges (Read and Northcote 1983b). Like PP1, PP2 also precipitates upon oxidation, but it does not form filaments (Kleinig *et al.* 1975) and is not required for the gellation of the exudate (Read and Northcote 1983a). It is assumed that PP2 is incorporated by disulphide bridges into the filaments observed *in vivo* (Read and Northcote 1983a). PP2 from *Cucurbita* normally appears on SDS-PAGE as a single band, but at high pH (pH 9.2), the PP2 dimer can be resolved into two bands (Mr of 26.5 kD and 25 kD) of approximately equal intensity (Read and Northcote 1983b). Similar doublets of PP2 polypeptides were identified at pH 8.8 (normal buffer condition) in *Cucumis* and *Citrullus* species (Read and Northcote 1983a). PP2 comprises 13-40% of the exudate proteins (Sabnis and Hart 1978; Read and Northcote 1983a), and together with PP1, represents more than 80% of the total exudate proteins in *Cucurbita* (Weber *et al.* 1974).

PP1 and PP2 share striking similarities: they are both highly basic, have similar amino acid compositions (Walker 1972; Beyenbach *et al.* 1974; Weber *et al.* 1974; Allen 1979; Read and Northcote 1983a), and in some cases show immunological cross-reactivity (Cronshaw and Sabnis 1990); both have readily oxidised cysteine residues (Read and Northcote 1983a); their Mr value appears to be in simple ratio (Read and Northcote 1983a; Beyenbach *et al.* 1974) and both exhibit nearly identical band patterns on SDS-PAGE after being treated with cyanogen bromide (Weber *et al.* 1974).

Since PP1 is assumed to be responsible for exudate gellation and filament formation, it is surprising that is not ubiquitous among the different species of Cucurbitaceae as examined by SDS-PAGE. It is a major protein in the *Cucurbita* species, but is absent or is a minor protein in species of other genera (Sabnis and Hart 1976; Sloan *et al.* 1976; Cronshaw and Sabnis 1990). Its native molecular size as given by different workers is variable, ranging from 80 to 136 kD (Walker 1972; Beyenbach *et al.* 1974; Nuske and Eschrich 1976; Sabnis and Hart 1979; Read and Northcote 1983b), which suggests that the PP1 subunit structure can depend on the conditions of exudate isolation and of preparation of the protein sample for

electrophoresis. On SDS-PAGE, PP1 is usually seen in a region where frequently the proteins appear to be incompletely dissociated under the experimental conditions (Beyenbach *et al.* 1974; Sloan *et al.* 1976; Cronshaw and Sabnis 1990; Bostwick *et al.* 1992). It has been speculated that high molecular weight proteins observed on SDS-PAGE might come from irreversible cross-linking of smaller polypeptides by bridges others than of the reversible disulphide types (Weber *et al.* 1974).

PP2 is a lectin specific for oligomers of N-acetylglucosamine (Sabnis and Hart 1978; Allen 1979; Gietl *et al.* 1979; Read and Northcote 1983b). By contrast to PP1, PP2-like phloem-specific lectin appears to be ubiquitous, not only among the cucurbit species but also among other species. Kauss and Ziegler (1974) first reported lectin activity in the sieve-tube sap of *Robinia pseudoacacia*. Similar high lectin activity was also evident within phloem sap from *Cucumis sativus* and *Cucumis melo*. Allen (1979) then showed that a lectin from *Cucurbita pepo* exudate strongly interacted with chitin oligosaccharides. Later a lectin was identified from *Cucurbita maxima*, which was shown to be PP2 (Sabnis and Hart 1978; Read and Northcote 1983b).

PP2-like lectins have also been identified from many other plant species. Gietl *et al.* (1979) isolated a lectin protein from *Robinia pseudoacacia* which exhibited a major band on SDS-PAGE with similar molecular size (25 kD) to the PP2 of cucurbits. They also examined the sieve-tube sap of 63 different tree species belonging to 21 families for lectin activity. With the exception of the members of the families Aceraceae and Oleaceae (in which Cronshaw and Sabnis [1990] suggested that the hemagglutination may have been masked), all analysed species agglutinated trypsinised rabbit erythrocytes, and the sugar specificity of the 15 species they examined was similar to that of cucurbit lectin. Recently, Schulz *et al.* (1989) isolated two phloem-specific lectins from *Pinus sabiniana*, a gymnosperm. These lectins are characteristic of those isolated from angiosperm species mentioned above in that they bind strongly to N-acetylglucosamine, are sensitive to thiol reagents, and have similar molecular sizes (23 and 25 kD) to the doublets of cucurbit PP2 polypeptides. More recently, a lectin was isolated from *Sechium edule* fruit exudate which also shows marked

similarities to the cucurbit lectin (Volzari-Hampe *et al.* 1992). This lectin is a 44-kD dimer consisting of two identical subunits, constituting high proportion of exudate protein (9.5%), and exhibits high specificity for N-acetylglucosamine.

Although the phloem-specific lectin appears to be ubiquitous, it exhibits substantial genus-specificity as determined by immunological methods. The antibody raised against the PP2 lectin of *Cucurbita maxima* reacted with sieve tube proteins from other species of the genus *Cucurbita* but showed only limited reaction with other genera (Smith *et al.* 1987). Similar results were obtained with other phloem-specific lectins (Cronshaw and Sabnis 1990).

The cell type-specific localisation of PP2 lectin in *Cucurbita maxima* was examined by Smith *et al.* (1987) by immunocytochemical methods. The lectin was found to be restricted to the sieve elements and companion cells.

The physiological and biochemical function of the P-protein is still unknown. Several hypotheses on its physiological role have long been proposed, such as involvement in phloem translocation (Thaine 1969; MacRobbie 1971; Lee 1972; Weber *et al.* 1974; Schulz *et al.* 1989) and providing a structural framework on which non-translocated phloem components can be anchored (Allen 1979; Sabnis and Hart 1978; Smith *et al.* 1987). Currently the most favoured proposal is that P-protein plays a protective role in the plant. The gellation of the P-protein is assumed to act as a sealing mechanism for wounded sites whereas the carbohydrate-binding lectin contained in the filaments serves to immobilise invading pathogens to the cross-linked filaments, thus maintaining sterility (Sabnis and Hart 1978; Allen 1979; Gietl *et al.* 1979; Read and Northcote 1983b). Early in the 1970's, there was interest in the possibility that one or more components of P-protein might be related to contractile proteins such as actin or tubulin, but attempts to find convincing evidence for this hypothesis were not successful (Sabnis and Hart 1974) although interest in it continues (Kulikova 1992).

Some other properties on the P-protein were also observed in early studies. Nucleoside phosphatase and acid phosphatase activity have been demonstrated in

association with the P-protein (Gilder and Cronshaw 1974; Esau and Charvat 1975; Cronshaw 1980). Weber *et al.* (1974) reported that the P-proteins of *Cucurbita* showed in several properties (amino acid compositions, molecular weights of the smaller components, IEPs, retention on CM-cellulose, antigenic sites) a strong similarity to proteins extracted from a ribosomal fraction from *Cucurbita* seedlings, in particular to a weakly basic subfraction of ribosomal proteins. Based on these results, they proposed that P-proteins are formed during sieve element maturation by aggregation and oxidative disulphide cross-linking of pre-existing proteins which, at least in Cucurbitaceae, are basic and may include ribosomal or ribosome-associated proteins.

The PP2 gene and its expression

In contrast to the protein biochemistry described above, studies on the P-protein encoding genes were not done until recently, and so far have been restricted to the PP2 gene. Sham and Northcote (1987) first reported the isolation of two cDNA clones encoding the PP2 of *Cucurbita maxima*. They found that the levels of both the PP2 mRNA and PP2 protein increased in the first few days of seedling growth. In older seedlings, however, the protein level remained constant but the mRNA level decreased dramatically to a very low level. Based on their observations they suggested that the synthesis of PP2 was subject to both transcriptional and post-transcriptional regulation.

No sequence data for the cDNA clones used by Sham and Northcote have been published. The first report containing a deduced PP2 amino acid sequence appeared at the end of 1992 (Bostwick *et al.* 1992), and the corresponding cDNA sequence was published in the middle of the following year (Bostwick *et al.* 1993). The cDNA was isolated from *Cucurbita maxima* by screening an expression cDNA library using polyclonal antibodies raised against the whole exudate of pumpkin phloem. Using this cDNA clone as a probe, they identified a single mRNA species of about 1 kb in size from pumpkin seedlings, and they found that PP2 mRNA was localised specifically

within the companion cells of the phloem tissue (Bostwick *et al.* 1992). According to this result they suggested that the PP2 may be synthesised in the companion cells and subsequently transported into the sieve element *via* plasmodesmata. A similar suggestion was also made by Fisher *et al.* (1992) after detecting substantial amount of newly synthesised proteins in the nucleus-free sieve elements of wheat. However, Bostwick *et al.* (1992) also pointed out the possibility that PP2 synthesis also occurs in sieve elements during early stage of cell differentiation. The latter view is supported by the previous observations that P-protein bodies are formed at early stage of sieve element differentiation (Cronshaw and Esau 1968a; Behnke 1974).

Apart from the PP2 cDNA, Bostwick *et al.* (1992) isolated another cDNA which hybridised to a single mRNA band with molecular size of about 2.5 kb. They suggested that this was a partial sequence of the PP1 gene. However, no further sequence data or evidence have been given to confirm this suggestion.

1.6. Specific aim of this project

A research program was started in 1990 in this department sponsored by the Rockefeller Foundation which aimed at producing transgenic rice with enhanced resistance to a major insect pest, the rice brown planthopper (BPH). The whole program was divided into three parts: identification of plant-derived proteins with toxicity to BPH using an artificial diet system, which has led to the discovery of several proteins with antimetabolic effects towards BPH, including snowdrop (*Galanthus navalis*) lectin (GNA) (Powell *et al.* 1993); isolation of promoters, and construction of expression vectors for plant transformation using the promoters and the genes that encode insecticidal proteins; and transformation of rice, which was to be carried out in other laboratories in the Rockefeller Program. The second part was the subject of this thesis.

As BPH is a phloem sap feeder, an appropriate promoter was needed which could direct expression of BPH resistance genes in a way that the gene products would be present in the phloem sap. As described earlier, the promoters of the sucrose

synthase gene and the phloem protein PP2 gene of *Cucurbita* were potentially good candidates. When this project began, the sucrose synthase promoter from the maize *Sh1* gene had been shown to direct phloem-specific expression of a reporter gene in transgenic tobacco, and the expression was observed in the sieve elements. Other evidence available then also suggested that the sucrose synthase promoter could be phloem-specific. Although the maize *Sh1* promoter should function in rice, in this work the rice sucrose synthase-1 promoter was pursued instead, in the belief that a homologous promoter might function better. Since the PP2 protein is highly abundant in the phloem exudate of *Cucurbita*, it was expected that the promoter might be useful in directing high level expression of resistance genes solely in the phloem. It was also expected that other (protein) sequences necessary (or useful) to target protein to phloem sap might be found in PP2 genes.

Therefore, the objectives of this project included:

1. Isolation and sequencing of the genomic DNA of the rice sucrose synthase-1 gene and the *Cucurbita pepo* PP2 gene;

2. Identification of the promoter sequences;

3. Construction of expression vectors using the promoters and, transformation of tobacco plants with the constructs in order to determine the activity of the promoters, before they are used by collaborators to transform rice.

CHAPTER 2. MATERIALS AND METHODS

2.1. Materials

Bacterial strains (Escherichia coli) and their genotype

DH5α F-, hsdR17 ($r_k^-m_k^-$), supE44, ΔlacU169, φ80dlacZΔ M15, hsdR17, recA1, endA1, gyrA96, thi-1, relA1, λ-

HB101 supE44, hsdS20($r_B^-m_B^-$), recA13, ara-14, proA2, lacY1, galK2, rpsL20, xyl-5, mtl-1

JM101 supE, thi, Δ(lac-proAB), F′[traD36, proAB+, lacI^q, lacZΔM15]

KW251 F-, supE44, supF58, galK2, galT22, metB1, hsdR2, mcrB1, mcrA-, argA81:Tn10, recD1014

LE392 F-, hsdR514(r_k^-,m_k^-), supE44, supF58, lacY1 or Δ(lacIZY)6, galK2, galT22, metB1, trpR55, λ-

NM538 supF, hsdR, trpR, lacY

SOLR e14-(mcrA), Δ(mcrCB-hsdSMR-mrr)171, sbcC, recB, recJ, umuC::Tn5(kan^r), uvrC, lac, gyrA96, relA1, thi-1, endA1, λ^R, [F′ proAB, lacI^{q}ZΔM15], Su- (non-suppressing)

SURE e14-(mcrA), Δ(mcrCB-hsdSMR-mrr)171, sbcC, recB, recJ, umuC::Tn5(kan^r), uvrC, supE44, lac, gyrA96, relA1, thi-1, endA1, [F′ proAB, lacI^{q}ZΔM15]

XL1-Blue recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac, [F′, proAB, lacI^{q}ZΔM15, Tn10(tet^r)]

Chemicals and reagents

Agarose and generally-used chemicals (analytical grade) were obtained from BDH Chemicals Ltd., Poole, Dorset. Acrylamide, Alumina (Type A-5), ethidium bromide, DTT (dithiothreitol), BSA (fraction V), CM Sepharose CL-6B, 3-aminopropyltriethoxysilane, thioglycolic acid (sodium salt), diethyldithiocarbamic acid (sodium salt), 3,3'-diaminobenzidine (DAB), Herring sperm DNA, 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc), NAA, and BAP were purchased from Sigma Chemical Company Ltd., Poole, Dorset. Antibiotics, DNA electrophoresis markers, X-gal and IPTG were supplied by NBL (Northumbria Biologicals Ltd.), Cramlington, Northumberland. M13 mp18 & 19 RF DNA and glycogen were obtained from Boehringer Mannheim UK Ltd., Lewes, E. Sussex and DH 5α competent cells and

pUC 19 monomer from Gibco BRL, Life Technologies Ltd., Paisley, Scotland.

Murashige and Skoog (MS) salts was purchased from Flow Laboratories, Irvine,

Scotland. Bacto-agar was supplied by Difco Ltd., W. Molesey, Surrey, Bacto-

tryptone by Becton Dickinson, Cowley Oxon, and OXOID Bacto-yeast extract by

Unipath Ltd., Basingstoke, Hants.. Radiochemicals were purchased from Amersham

International plc., Aylesbury, Bucks.. Film developer (Ilford phenisol) was supplied by

Ilford Ltd., Mobberly, Ches. and fixer (Kodak Unifax) by Phase Separations Ltd.,

Deeside, Clwyd.. Clontech pBI 101 vector including HB101:pRK3013 and

Agrobacterium tumefaciens LBA 4404 were obtained from Cambridge BioScience,

Cambridge. λGem-12 half-site arms including fill-in reaction buffer and KW 251

bacteria and Packagene packaging extract were supplied by Promega Ltd.,

Southampton.

Antibodies

Maize sucrose synthase polyclonal antibody was kindly provided by Dr. Karen E.

Koch, Horticultural Sciences Department, University of Florida, USA and GNA

(*Galanthus nivalis* [snowdrop] agglutinin) antibodies were donated by Drs. W.

Peumans and E. van Damme, Catholic University of Leuven, Belgium. The Goat anti-

rabbit IgG (H+L) Horseradish peroxidase conjugate was purchased from Biorad

Laboratories Ltd., Hemel Hempstead, Herts..

Enzymes

Various restriction endonucleases and modifying enzymes were supplied by NBL or

Boehringer Mannheim. Proteinase K, RNase A and DNase 1 were obtained from

Sigma and Taq DNA polymerase and Mung Bean nuclease from Promega.

Equipment

Medicell dialysis tubing and Nescofilm were obtained from BDH and nitrocellulose

membranes (Schleicher & Schuell grade BA-85) from Anderman and Company Ltd.,

Kingston-Upon Thames, Surry. Whatman membrane filters (0.2 μm pore size) and 3MM paper were purchased from Whatman Labsales Ltd., Maidstone, Kent. Petri dish was supplied by Bibby Sterilin Ltd., Stone, Staffs.. X-ray films (Fuji-RX) were obtained from Fuji Photo Film (UK) Ltd., London. X-ray cassettes, ATTO protein gel and blotting apparatus were purchased from Genetic Research Instrumentation Ltd., Dunmow, Essex. DNA synthesiser, DNA sequencer and PVDF membrane (ProBlott) were supplied by Applied Biosystems Inc., Warrington, Ches.. Spectrofluorimeter was purchased from Bairo Atomic Ltd., Essex.

DNA library, kits, and others

A rice genomic library (Clontech) was supplied by Cambridge BioScience, Cambridge and Stratagene ZAP-cDNA synthesis kit by NBL. Quigen DNA purification kit was obtained from Hybaid Ltd., Teddington, Middlesex and GENECLEAN kit from Stratech Scientific, Luton. DNA sequencing kit and oligonucleotide synthesising reagents were supplied by Applied Biosystems. The GNA coding sequence fragment was kindly provided by L. Gatehouse and the 35S-GNA transgenic tobacco plant by the Agricultural Genetics Company.

Commonly used buffers and media

50× TAE (agarose gel-running) buffer	per liter: Tris base — 242 g glacial acetic acid — 57.1 ml 0.5 M EDTA pH 8.0 — 100 ml
2× SDS-PAGE sample buffer	100 mM Tris-HCl pH 6.8 2% β-mercaptoethanol 4% SDS 0.2% bromophenol blue 20% glycerol
5× Formaldehyde agarose gel-running buffer	0.1 M 3-(N-morpholino) propanesulfonic acid (MOPS) pH 7.0 40 mM sodium acetate 5 mM EDTA

6× Agarose gel-loading buffer	0.25% bromophenol blue 40% (w/v) sucrose

20× SSC

per liter:
NaCl	175.3 g
Sodium citrate	88.2 g
pH 7.0	

20× SSPE

per liter:
NaCl	175.3 g
$NaH_2PO_4.H_2O$	27.6 g
EDTA	7.4 g
pH 7.4	

TE buffer

10 mM Tris-HCl pH 8.0
1 mM EDTA pH 8.0

PBS buffer

0.15 M NaCl
0.1 M NaH_2PO_4-Na_2HPO_4 pH 7.4

Phage buffer

20 mM Tris-HCl pH 7.4
100 mM NaCl
10 mM $MgSO_4$

SM phage buffer

per liter:
NaCl	5.8 g
$MgSO_4.7H_2O$	2 g
Tris-HCl pH 7.5	50 ml
2% gelatin solution	5 ml

Denhardt's solution, 50×

per liter:
Ficoll	5 g
polyvinylpyrrolidone	5 g
BSA (fraction V)	5 g

Phenol:chloroform:isoamyl alcohol (25:24:1)

Redistilled phenol was extracted several times with TE buffer (or 0.1 M Tris-HCl pH7.6) and then 1 part of the lower, phenol phase mixed with 1 part of chloroform:isoamyl alcohol (24:1).

LB medium

per liter:
Bacto-tryptone	10 g
Bacto-yeast extract	5 g
NaCl	10 g
pH 7.0	

LB-agar	Add 15 g Bacto-agar to 1 liter of above

LB top-agar	Add 0.65 g Bacto-agar to 100 ml of LB medium

TB top-agar	per 100 ml:	
	Bacto-tryptone	1.0 g
	NaCl	0.5 g
	Bacto-agar	0.8 g

NZY medium	per liter:	
	NZ amine	10 g
	NaCl	5 g
	Bacto-yeast extract	5 g
	$MgSO_4.7H_2O$	2 g
	pH 7.5	

NZY top-agar	Add 0.65 g Bacto-agar to 100 ml of above

2× YT medium	per liter:	
	Bacto-tryptone	16 g
	Bacto-yeast extract	10 g
	NaCl	5 g
	pH 7.0	

YEB medium	per liter:	
	Bacto-tryptone	5 g
	Bacto-yeast extract	1 g
	Brain Heart Infusion	5 g
	1 M $MgSO_4$	2 ml
	sucrose	5 g

MSO medium (for plant growth)	per liter:	
	MS salts	4.7 g
	sucrose	30 g
	pH 5.8	
	Add 8 g Bacto-agar to make plates	

MSD4×2 medium (for plant growth)	MSO medium plus:	
	NAA	0.1 mg/l
	6-BAP	1 mg/l
	Add 8 g Bacto-agar to make plates	

Sterilisation of glassware, plasticware, solutions and media

Glassware, plasticware, culture media and most of solutions were sterilised by autoclaving for 20 min at 120 °C or 15 lb./sq. in.. Solutions which contained chemicals unstable at high temperature, such as antibiotics and maltose, were sterilised by filtration through a 0.2-micron filter.

Plant materials

Growth of rice plant

Seeds of rice plant (*Oryza sativa* L. cv. Taichung Native 1, kindly supplied by the International Rice Research Institute, Manila, Philippines), were sowed in loam-based compost (John Innes Centre, Norwich) in propagation trays which were immersed half way in water-containing trays. The rice plants were grown in a 25 °C growth room with an illumination cycle of 14-hr light and 10-hr dark.

Dark growth of maize seedlings

Seeds of maize (Sweet Corn, F1 Kodlak, from Samuel Yates Ltd., Ches.) were soaked in water at room temperature for about 8 hrs and then placed on wet tissue papers in a plastic box. This box was then sealed from light with aluminium foil and kept in a 25 ° C incubator to allow the seeds to germinate and grow.

Growth of Cucurbita pepo plants

Plants of *Cucurbita pepo* L. 'Autumn Gold' (pumpkin) were grown from seeds (Suttons Seeds Ltd., Torquay) in Levington potting compost (Fisons, Loughborough) under growth room conditions. One day after the hypocotyls emerged, the seedlings were regarded as 1 day-old.

Growth of sterile tobacco plants

Sterile plants used for transformation were kindly provided by Ms. Julia Bartley. To grow such plants, seeds of tobacco (*Nicotiana tobacum* cv. SRI) were sterilised by treatment with 10% hyperchlorite solution (containing ~1.4% chlorine) in a syringe for 30 min and washed with 3 changes of sterile distilled water, and then transferred onto MSO plates. The plates were sealed with Nescofilm and incubated in a 25 °C

incubator with 14-hr photoperiod. After about 10 days the young seedlings were transferred to Magenta pots containing MSO medium (1 seedlings in each pot) and allowed to grow in a growth room.

E. coli competent cells

Competent cells of *E. coli* DH 5α and JM 101 used for transformation were either commercially obtained from GIBCO BRL or prepared by the one-step preparation method (Chung *et al.* 1989). For preparation, *E. coli* cells were grown to early exponential phase (O.D.$_{600}$ = 0.3-0.4), pelleted by centrifugation at 1000g for 10 min, and resuspended at one-tenth of their original volume in ice-cold transformation and storage solution (TSS, which consisted of LB broth with 10% (w/v) PEG 6000, 5% (v/v) dimethyl sulfoxide and 35 mM MgSO$_4$, at a final pH of 6.5). This suspension was divided into small aliquots, frozen in dry ice/ethanol bath, and stored at -70 °C before use.

Storage of bacterial strains

Appropriate bacterial culture medium (5 ml) containing certain antibiotics (e.g. 100 μ g/ml ampicilin for pUC vector, 50 μg/ml kanamycin for pBI vector) was inoculated with a single, well-isolated bacterial colony and shaken at 37 °C overnight. To 0.85 ml of the bacterial culture in a labelled storage tube, 0.15 ml of sterile glycerol was added and the tube tightly capped. The culture was gently mixed, frozen in liquid nitrogen and then stored at -70 °C.

Oligonucleotides

Oligonucleotide synthesis was done by Mr. John S. Gilroy on an Applied Biosystems model 381 DNA synthesiser, based on solid-phase phosphoramidite chemistry. Basically, this involves 5 steps, "detritylation", which removes the dimethoxy trityl group from the support-bound nucleoside to free the 5' -OH for reaction; "addition", in which activated phosphoramidite rapidly reacts with the 5' -OH; "capping", which

terminates or 'caps' any chains which did not react in the addition step by acetylating the unreacted 5' -OH; "oxidation", which converts the labile trivalent phosphorous linkage formed in the addition step to the stable, pentavalent phosphorous linkage of biologically active DNA; and "deprotection", in which the phosphate protecting groups are removed and the chains separated from the solid support, and the base protecting groups are removed. The synthesised oligonucleotide was purified by ethanol precipitation and stored at -20 °C either dried or in deionised water. Concentration of the oligonucleotides was determined spectrophotometrically.

2.2. Methods

2.2.1. Isolation of nucleic acids

Minipreparation of plasmid DNA for routine restriction

Plasmid DNA of pUC18, pUC19 and their recombinants used for endonuclease restriction was prepared according to the alkaline lysis method described by Maniatis *et al.* and Mierendorf and Pfeffer (1987). For each preparation, 5-ml LB-ampicilin (100 μg/ml) medium was inoculated with a single colony of plasmid-containing *E. coli* cells and shaken overnight at 37 °C. The cells were pelleted in an Eppendorf tube and resuspended by vortexing in 100 μl of ice-cold solution containing 50 mM glucose, 25 mM Tris-HCl (pH 8.0) and 10 mM EDTA (pH 8.0). Then 200 μl of lysis solution (0.2 M NaOH, 1% SDS) was added, and the tube was inverted immediately several times and placed on ice. After 5-10 min when the suspension had cleared significantly, 150 μl of ice-cold potassium acetate (pH 4.8) (prepared by adding 11.5 ml glacial acetic acid and 28.5 ml water to 60 ml of 5 M potassium acetate) was added, and mixed immediately by inversion. The tube was placed on ice for 10 min and centrifuged at 10,000g for 5 min. The clear supernatant was transferred to a fresh tube, treated with RNase A at 0.1 μg/μl at 37 °C for 30 min, and extracted with equal volume of phenol:chloroform:isoamyl alcohol. To precipitate the DNA, 0.6 volume of

isopropanol was added to the aqueous phase and mixed thoroughly. After 15 min at room temperature, the precipitate was pelleted by centrifugation at 10,000g for 10 min. The pellet was washed with 70% ethanol, dried under vacuum and resuspended in 50-100 μl of TE (pH 8.0) or deionised water.

Minipreparation of plasmid DNA for sequencing

For nucleotide sequencing, plasmid DNA isolated by the alkaline lysis method described above was further purified by PEG precipitation according to Mierendorf and Pfeffer (1987). Nucleic acid prepared from 5 ml of overnight cell culture was dissolved in 48 μl of deionised water and then mixed with 12 μl of 4 M NaCl. Equal volume (60 μl) of 13% polyethylene glycol (MW 8000) (PEG-8000) was added to the above and the mixture incubated on ice for 20 min, followed by centrifugation at 10,000g for 10 min. The supernatant was carefully removed and discarded with a micropipette tip, and the DNA pellet washed twice with 75% ethanol, dried under vacuum and resuspended in 10 μl deionised water.

Minipreparation of single-stranded M13 DNA for sequencing

The procedure was a modification of the PEG procedure described by Messing (1983). An *E. coli* JM101 culture was prepared by inoculating 5 ml of LB medium with a single colony and incubating at 37 °C until O.D.$_{600}$ reached 0.3-0.4. Fifty μl of the culture was mixed with 5 ml of LB medium in a sterile McCartney bottle. A newly-grown, well-isolated phage plaque was transferred to the above 5-ml culture using a sterile toothpick. The bottle was rotated on a wheel in a 37 °C incubation room for 5-7 hrs. The infected cells were transferred to 1.5-ml Eppendorf tubes and centrifuged at 10,000g for 10 min. The supernatant was stored at 4 °C, and centrifuged again for 10 min at 10,000g immediately prior to use. To a fresh tube 1.25 ml of the clear supernatant was added and then mixed with 250 μl of 20% PEG-6000 in 2.5 M NaCl. The tube was left at room temperature for 20 min and then centrifuged at 10,000g for 5 min, and the supernatant was discarded. The tube was spun again and the residual

fluid removed with pipette tip. The pellet was then resuspended in 200 μl of TE

buffer, and the solution extracted twice with 100 μl of Tris-saturated phenol, once with

200 μl of Tris-saturated chloroform:isoamyl alcohol (24:1) and twice with 200 μl of

water-saturated ether. For each extraction the tube was vortexed for 5 second (phenol

extraction), 15 second (chloroform extraction), or until the cloudiness disappeared

(ether extraction). To the final aqueous phase 20 μl of 3 M sodium acetate (pH 5.2)

and 500 μl of ethanol were added, mixed, and the mixture was left at -20 °C for at

least 1 hr. The single-stranded DNA was pelleted by centrifugation at 10,000g for 15

min, washed with ice-cold 70% ethanol, dried in vacuum and dissolved in 10 μl TE.

Preparation of phage λ DNA

Preparation of phage lysate

Purified phage was re-plated as in "Titration of DNA libraries" section. A single, well-

isolated plaque was picked from the plate by removing a plug of agar containing the

plaque. The plug was expelled into a tube containing 500 μl of SM phage buffer and

the tube kept at room temperature for 40 min or at 4 °C overnight. The whole of this

phage solution was mixed with 500 μl of overnight host cell cultures (NM 538 or KW

251, prepared as in "Titration of DNA libraries" section), and the suspension incubated

at 37 °C for 20-30 min, which was then used to inoculate 50 ml of prewarmed LB

medium containing 10 mM MgSO$_4$. The culture was shaken at 37 °C for about 6.5

hrs, then 250 μl of chloroform was added and the culture was shaken at 37 °C for

further 10 min. The cell debris was eliminated by centrifugation at 4000g for 10 min.

The clear supernatant was stored at 4 °C or used immediately for DNA preparation.

Isolation of λ DNA

Lambda DNA was isolated from above phage lysate using the QIAGEN λDNA midi-

preparation kit following the supplier's instruction. Basically, the phage lysate was

treated with DNase and RNase to eliminate the nucleic acid released by the lysed

bacterial cells, and then the phage particles were pelleted with PEG. Phage λDNA was

released from the phage particles by treatment with 2% SDS at 70 °C, and most of the

proteins were removed by precipitation with 0.85 M potassium acetate (pH 4.8). The released DNA was purified by chromatography on a QIAGEN column (QIAGEN-tip 100), concentrated by isopropanol precipitation and dissolved in 200 µl of TE buffer.

Purification of DNA fragments from agarose gel

DNA separated by agarose gel electrophoresis was purified using GENECLEAN II kit. DNA bands cut out from the gels were incubated at 50 °C in 3 volumes of the 6 M NaI solution provided with the kit until the gel slices were melted. The solution was then mixed by vortexing with 5 µl of the "glassmilk" supplied and incubated on ice for 10 min. The "glassmilk" was recovered by brief centrifugation, washed four times with 200 µl of the supplied "New Wash" and air dried. DNA was eluted off the "glassmilk" by incubating the "glassmilk" in 2 changes of deionised water (20 µl totally) at 50 °C for 3 min and collected by centrifugation at 10,000g for 3 min.

Minipreparation of maize genomic DNA for PCR

Isolation of a small amount of high molecular weight DNA was based on the SDS-proteinase K method designed for DNA isolation from small amount of fresh tissues (Draper and Scott 1988). Seven day-old dark-grown maize seedlings were harvested and ground into fine powder in a mortar and pestle with the assistance of liquid nitrogen. About 0.2 g of the frozen powder was transferred into a precooled 1.5-ml Eppendorf tube, and immediately 0.7 ml of SDS-extraction buffer (100 mM Tris-HCl pH 8.5, 100 mM NaCl, 50 mM EDTA and 2% SDS, preheated to 80 °C) was added and mixed by gentle inversion of the tube. The tube was kept at 65 °C for 10 min and then cooled down to 37 °C. To the tube 10 µl of proteinase K (10 mg/ml) was added, and the mixture was incubated at 37 °C for 1 hr and then extracted twice by gentle mixing with equal volume of phenol:chloroform:isoamyl alcohol and centrifugation at 10,000g for 2 min. The upper aqueous phase was transferred to a fresh tube and the DNA precipitated with 0.6 volume of isopropanol at -20 °C for 15 min. The DNA pellet was washed with 70% ethanol, vacuum-dried and resuspended in 100 µl of 0.1 ×

TE. The sample was treated with RNase A at 0.1 μg/μl at 37 °C for 30 min and purified by phenol:chloroform extraction and ethanol precipitation.

Isolation of rice and maize genomic DNA for Southern blotting

Maize genomic DNA used for genomic Southern blotting was prepared by a scaled-up procedure of the SDS-proteinase K method described above. Six grams instead of 0.2 g of maize seedlings was used and the extraction was performed in a 38-ml screw-capped centrifuge tube.

Rice genomic DNA was isolated by the SDS-proteinase K method (Draper and Scott 1988) designed for DNA isolation from freeze-dried tissues. Four grams of 4-5 week-old leaves were harvested and freeze-dried, and then pulverised in a mortar and pestle with assistance of alumina. The powder was transferred into a centrifuge tube and mixed by gentle inversion with 30 ml of SDS-proteinase K solution (see above section, with 0.1 mg/ml proteinase K, prewarmed to 37°C). The tube was stored at 37 °C for 1.5 hrs with occasional gentle inversion. The slurry was extracted twice with phenol:chloroform:isoamyl alcohol, and the upper aqueous phase was mixed with 0.6 volume of isopropanol and left at -20 °C for 1 hr. The DNA precipitate was spooled out with pipette tips and transferred to Eppendorf tubes containing 70% ethanol. The tubes were spun at 10,000g for 2 min and the DNA pellets were washed again with 70% ethanol, vacuum-dried, redissolved in 150 μl 0.1× TE and treated with RNase A as above.

Isolation of genomic DNA from Cucurbita pepo

Genomic DNA of *Cucurbita pepo* (pumpkin) used in construction of genomic library and Southern blotting was isolated from 3-5 day-old seedlings by CTAB method according to Draper and Scott (1988) and Rogers and Bendich (1988). About 0.5 g of freeze-dried pumpkin seedlings was pulverised in a mortar and pestle with assistance of alumina. The powder was transferred to a 38-ml centrifuge tube containing 9 ml of 1× CTAB extraction buffer (50 mM Tris-HCl pH 8.0, 0.7 M NaCl, 10 mM EDTA, 1%

CTAB [cetyl triethylammonium bromide], 20 mM 2-mercaptoethanol) and gently dispersed with a sterile glass rod. The slurry was incubated at 56 °C for 20 min with occasional very gentle inversion of the tube and then mixed with 8 ml of chloroform:isoamyl alcohol (24:1) by gently inverting the tube for 10 min. The tube was centrifuged at 2500g for 6 min and the aqueous phase transferred with a wide-bore pipette tip to a fresh tube, and then 1.5 ml of 1× CTAB buffer was added to the original tube, mixed gently and the tube was centrifuged again. The aqueous phase was pooled with the first one, mixed with 1/5 volume of 5% CTAB (5% [w/v] CTAB, 0.35 M NaCl) and extracted again with chloroform:isoamyl alcohol. The DNA in the aqueous phase was precipitated with equal volume of CTAB precipitation buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% [w/v] CTAB) at room temperature for 20 min. The precipitate was pelleted by centrifugation at 2000g for 7 min and resuspended in 3.5 ml of high-salt TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 M NaCl), and the suspension was centrifuged to eliminate residual undissolved pellets. To the clear solution two volumes of ethanol was added and the mixture kept overnight at -20 °C. The DNA was pelleted by centrifugation at 2000g for 4 min, vacuum-dried, resuspended in 4 ml of 0.1 × TE, and treated with RNase A as above.

Extraction of total Cucurbita pepo RNA

Plant tissues (leaves, stems, hypocotyls, cotyledons) were harvested, frozen immediately in liquid nitrogen and stored at -70 °C before use. Total RNA was extracted by a modification of the phenol extraction method (de Vries *et al.* 1988). Plant materials (4.2 g) were ground into fine powder as above and the frozen powder was quickly transferred into a precooled centrifuge tube. As the liquid nitrogen evaporated completely, 8 ml of phenol:RNA extraction buffer (1:1)(preheated in boiling water) was added and the powder dispersed immediately with a sterile spatula and with occasional heating of the tube in hot water. The extraction buffer consists of 100 mM LiCl, 1% SDS, 100 mM Tris-NaOH pH 9.0, and 10 mM EDTA. The slurry was extracted with 4 ml of chloroform:isoamyl alcohol by gently inverting the tube for

30 min, and the phases were separated by centrifugation at 10,000g for 15 min. The upper aqueous phase was transferred to a fresh tube and extracted again with chloroform:isoamyl alcohol for 15 min. To the final aqueous phase 1/3 volume of 8 M LiCl was added and the mixture left at 4 °C for 20 hrs. The RNA was pelleted by centrifugation at 4 °C, 10,000g for 25 min, washed once with ice-cold 2 M LiCl and twice with 80% ethanol, air-dried and redissolved in 100 μl of deionised water. The RNA samples were stored at -70 °C or in liquid nitrogen.

Purification of Poly (A)$^+$ RNA

Poly(A)+ RNA used for construction of *Cucurbita pepo* cDNA library was purified from 1 mg of the total RNA of 3-5 day-old hypocotyls using the PolyATract mRNA isolation system III (Promega). The system used a biotinylated oligo (dT) primer to hybridise in 0.5× SSC to the 3' Poly(A) region of the mRNA species. The hybrids were captured and washed with 0.1× SSC using streptavidin coupled to paramagnetic particles and a magnetic separation stand. The mRNA was eluted from the solid phase by addition of RNase-free deionised water and concentrated by ethanol precipitation..

2.2.2. Cloning and transformation

General subcloning procedure

Restricted DNA fragments were ligated into restricted vector DNA of pUC18, pUC19, or M13mp18, M13mp19 RF DNA. Vector DNA restricted with single enzymes was usually dephosphorylated with alkaline phosphatase before use. Ligation reactions containing vector and DNA fragments were usually set up in 10-μl volume with 1 unit of T4 DNA ligase, and incubated overnight at 15 °C. The ligated DNA was used to transform *E. coli* cells and the resulting bacterial colonies or phage plaques were selected by the blue-white colour test (Maniatis *et al*. 1989) and verified by restriction and electrophoretic analysis of mini-preparations of the recombinant DNA.

Digestion of DNA with restriction endonucleases

Digestion was set up in Eppendorf tubes on ice and carried out for a minimum of 1 hr at 37 °C or at other required temperatures. Amounts of enzymes used were at least 1 unit per μg of DNA, and the final dilution of enzymes was no less than 10 times that of the original commercial form. The buffers used were supplied with the enzymes.

Dephosphorylation of vector DNA

Linearised vector DNA restricted with a single enzyme was treated with alkaline phosphatase at 37 °C for 30 min in 1 mM $ZnCl_2$, 1 mM $MgCl_2$ and 10 mM Tris-HCl (pH 8.3) according to Maniatis et al.(1989). The treated DNA was then purified by phenol:chloroform extraction and ethanol precipitation.

Transformation of E. coli competent cells with plasmid DNA

Competent cells commercially obtained or self-prepared were transformed following the supplier's instruction or according to Maniatis et al.(1989). The only difference between them lay in the heat-shock step: 37 °C, 20 seconds for the commercially obtained DH 5α cells and 42 °C , 90 seconds for the self-prepared cells. For each transformation, an aliquot of 1-3 μl of ligated plasmid was gently mixed with 50 μl of freshly thawed competent cells and incubated on ice for 30 min. The mixture was heat-shocked, chilled on ice for 2 min, and then 0.95 ml of LB broth added. The cells were incubated by shaking at 37 °C for 1 hr to recover and to express the ampicilin resistance. Aliquots were then spread over LB-ampicilin X-gal plates and incubated overnight at 37 °C.

Transformation of E. coli competent cells with M13 DNA

M13 DNA transformation followed the same procedure as plasmid DNA transformation up to the heat-shock step described above. After heat shock, aliquots of the transformed cells were quickly mixed with 3 ml of 47 °C molten SOB top-agar (Maniatis et al. 1989), 200 μl of exponential JM101 cells, 4 μl of IPTG (200 mg/ml)

and 8 μl of X-gal (100 mg/ml in dimethylformamide), and spread over LB plates. The plates were placed at room temperature for 10 min to allow the top-agar to harden and then incubated inverted at 37 °C for overnight.

Cloning of PCR fragment

The PCR products were fractionated by electrophoresis on a 1.8% agarose gel. The 0.3 kb band was isolated from the gel, blunt-ended by treatment with T4 DNA polymerase and then cloned into the SmaI site of pUC 18.

Isolation of maize Sh-1 exon 13 sequence by polymerase chain reaction (PCR)

The following components were mixed up in an 0.5-ml Eppendorf tube: deionised water, 59 μl; 10× PCR reaction buffer (500 mM KCl, 100 mM Tris-HCl, pH 8.8; 15 mM $MgCl_2$ and 1% Triton X-100), 10 μl; dNTPs (1.25 mM of each), 16 μl; primer A (0.16 μg/μl), 7 μl; primer B (0.23 μg/μl), 5 μl; maize genomic DNA (0.5 μg/μl), 2 μl. The mixture was heated at 94 °C for 5 min, and cooled to 55 °C then 1 μl of Taq DNA polymerase (2 u/μl) was added, mixed with a pipette and overlaid with 30 μl of mineral oil. The PCR reaction was performed 32 cycles on a thermocycler (Model PHC-3, Techne Inc., USA) using the following temperature profile: denaturing, 94 °C, 1 min; primer annealing, 55 °C, 1.5 min; primer extension, 72 °C, 3 min. The cycling concluded with a final extension at 72 °C for 6 min. Primer A is a 19-nt sequence 5'-TTCGTGCTGAAGGACAAGA-3' and primer B a 18-nt oligonucleotide 5'-CTGCACGAATGCGCCCTT-3'.

Blunting of PCR-amplified DNA ends with T4 DNA polymerase

The DNA fragment purified from 100 μl of the PCR reaction was mixed with 2 units of T4 DNA polymerase in 20 μl of 50 mM Tris-HCl, pH 8.5, 15 mM $(NH_4)_2SO_4$, 7 mM $MgCl_2$, 0.1 mM EDTA, 10 mM 2-mercaptoethanol and 33 μM of each dNTP and then incubated at 37 °C for 30 min. The treated DNA was purified by

phenol:chloroform extraction and ethanol precipitation, and resuspended in 10 µl of deionised water.

Blunting of 3'-overhang DNA ends with Mung Bean nuclease

DNA (~1 µg) digested with HaeII was mixed with 15 units of Mung Bean nuclease in 200 µl of 30 mM NaAc pH 4.6, 50 mM NaCl, 1 mM ZnCl$_2$, and 5% glycerol, and incubated at 30 °C for 30 min. The reaction mixture was then neutralised with 20 µl of 1 M Tris-HCl pH 8.0 and extracted with phenol:chloroform. The treated DNA was ethanol precipitated and redissolved in deionised water.

2.2.3. Analysis of DNA and RNA

Agarose gel electrophoresis of DNA

Agarose gel was prepared according to Maniatis et al.(1989). Agarose powder was weighed, added to certain volume of TAE buffer containing 0.5 µg/ml ethidium bromide, and melted in a microwave oven. The gels were cast in the plastic plates provided with the electrophoresis apparatus or on glass plates. Electrophoresis was carried out in a Pharmacia GNA-100 mini-gel system, in a NBL medium-size (128 × 149 mm) gel apparatus or on a 185 × 152 mm gel in a self-made tank containing 2 litres of running buffer. DNA samples were mixed at one-sixth of its volume with gel loading buffer and then slowly loaded into slots of submerged gels. The gels were run at 1-5 volts per centimetre at room temperature in TAE buffer containing 0.5 µg/ml ethidium bromide. HindIII-digested or PstI-digested λDNA was used as molecular weight marker. DNA bands were visualised and photographed with a short-wave UV transilluminator.

Formaldehyde-agarose gel electrophoresis of RNA

Agarose (0.6 g) was melted in 31 ml of water, cooled to 70 °C in a chemical hood, and then were added 10 ml of 5× gel running buffer and 9 ml of 37% deionised

formaldehyde, mixed and poured in a gel former. The gel was allowed to set for a minimum of 1 hr and pre-electrophoresed at 5 V/cm for 5-20 min. RNA sample was mixed with 2 μl of 5× running buffer, 3.5 μl of formamide, 10 μl of formaldehyde, and distilled water to make final volume of 20 μl. The mixture was heated at 65 °C for 15 min, chilled on ice, spun briefly to deposit all of the fluid, mixed with 2 μl of gel-loading buffer (50% glycerol, 1 mM EDTA pH 8.0, 0.25% bromophenol blue, 0.25% xylene cyanol FF), and then loaded to the gel. The gel was run in 1× gel-running buffer at 60-80 V using the Pharmacia GNA-100 electrophoretic apparatus. After completion of electrophoresis, the lanes for viewing were cut out and stained for 5 min in 5 μg/ml ethidium bromide, destained in several changes of water for 2 hrs or longer, and visualised and photographed as above.

Alkaline agarose electrophoresis and autoradiography

Alkaline agarose gel electrophoresis was performed to analyse cDNA synthesis quality *via* examining the sizes of DNA and the incorporation of radioactive dCTP. Agarose (0.8 g) was melted in 72 ml of water and then cooled to 55 °C. To the cooled agarose 8 ml of 10× alkaline agarose buffer (0.3 M NaOH, 20 mM EDTA) was added, and this was swirled to mix and poured immediately to a plastic plate to form a thin (~3 mm thick) gel. DNA sample with radioactivity was mixed with equal volume of 2× loading buffer (20% glycerol, 4.8% bromophenol, 25 mM NaOH) and then loaded. The gel was run at 100 mA in 1× alkaline agarose buffer using the Pharmacia GNA-100 electrophoresis apparatus.

After completion of electrophoresis, the gel was washed with several changes of water for a period of 40 min, blotted on filter paper to remove excess liquid, sealed with Clingo-rap film, and exposed to X-ray film at -70 °C overnight.

Determination of insert orientation in M13 transformants

When a DNA fragment with two same restriction ends was cloned into M13 vector, its orientation with respect to each other in different clones was determined by the

complimentary or "C" test (Messing 1983). One μl each of single-stranded recombinant M13 DNA from every two clones to be tested was mixed, together with 11.5 μl TE buffer, 1.5 μl 5 M NaCl, and 5 μl formamide/SDS/dye mix (3% SDS, 0.1% bromophenol blue, 60% deionised formamide, 25 mM EDTA). The mixture was incubated at 65 °C for 1 hr, and then electrophoresed on a 0.8% agarose gel. If the orientation of inserts was opposite to each other, their sequences were complimentary and hybridised molecules will be formed which had lower mobility on the gel than *vice versa*.

Determination of nucleic acid concentration

Nucleic acid concentration was determined spectrophotometrically on a Philips PU8740 UV/Vis scanning spectrophotometer. An O.D. of 1 at 260 nm was assumed to correspond to 50 μg/ml for double-stranded DNA, 40 μg/ml for single-stranded DNA and RNA, and 20 μg/ml for single-stranded oligonucleotides. The ratio between the reading at 260 nm and 280 nm ($O.D._{260}/O.D._{280}$) was used to estimate the purity of the nucleic acid samples. Pure preparations of DNA and RNA have $O.D._{260}/O.D._{280}$ values of 1.8 and 2.0, respectively.

Northern blotting

After completion of electrophoresis, the formaldehyde-agarose gel was soaked in 0.05 N NaOH for 15 min with gentle shaking, rinsed in several changes of distilled water for a period of 2 hrs, and then soaked in transfer buffer (20× SSC) for 1.5 hrs. RNA was transferred from the gel to nitrocellulose membrane following the procedure described by Maniatis *et al.* (1989).

Labelling of DNA probe

Restricted DNA fragments used as DNA hybridisation probes were purified from agarose gel and labelled with α-^{32}P by the random priming labelling technique (Feinberg and Vogelstein 1983, 1984). The labelling reaction was carried out at room

temperature for 2.5 hrs or longer. A typical reaction was set up as follows: H_2O, if

needed to 50 μl; OLB (Feinberg and Vogelstein 1984), 10 μl; BSA (10 μg/μl), 2 μl;

DNA, 30-50 ng; $[\alpha\text{-}^{32}P]dCTP$, 5 μl (50 μCi or 10 pmole); the Klenow fragment of *E.*

coli DNA polymerase I, 2 units. Before being added to the reaction, DNA fragments

were mixed with H_2O, heated in boiling water for 2 min and chilled on ice. The

reaction was stopped by addition of 1/10 volumes of 10% SDS. To purify the labelled

DNA, the reaction was mixed with 1/10 volumes of Dextran blue 100 and then

separated by gel filtration through a 2-ml Sephadex G-50 column (run in 0.15 M NaCl,

10 mM EDTA, 0.1% SDS, 50 mM Tris-HCl, pH 7.5). Fractions containing Dextran

blue (in blue colour) were collected and used for hybridisation.

Labelling of oligonucleotide probe

Synthetic oligonucleotide probes were labelled with $\gamma\text{-}^{32}P$ following the procedure in

Maniatis *et al.* (1989). About 10 pmoles of oligonucleotide was mixed with 5 μl of $[\gamma\text{-}^{32}P]$ ATP (50 μCi or 10 pmoles) and 4 units of bacteriophage T4 polynucleotide

kinase in 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM

spermidine-HCl and 0.1 mM EDTA (pH 8.0). The mixture was incubated at 37 °C for

30 min and stopped by addition of 1/10 volumes of 0.5 M EDTA. The labelled

oligonucleotide was purified by ethanol precipitation (Maniatis *et al.* 1989).

Hybridisation of Northern blot with degenerate oligonucleotide probe

Total RNAs (50 μg each) from leaves and stems of 20 days old pumpkin plants and

from the cotyledons and hypocotyls of 3-5 days old seedlings were separately loaded

and separated on a formaldehyde-agarose gel (1.2% agarose) and transferred to

nitrocellulose membrane as described above. The filter was prehybridised for 4 hrs and

hybridised for 17 hrs at 40 °C in 25 ml of buffer containing 6× SSC, 2× Denhardt's

solution, 0.25% SDS and 100 μg/ml denatured herring sperm DNA. The end-labelled

oligonucleotide, 5'-GC(C/T)TC(C/T)TT(C/T)TC(C/T)TT(A/G)TT(A/G)TCCAT-3',

was used as hybridisation probe. The filter was washed in 2× SSC, 0.05% SDS for 1

hr at room temperature, 20 min at 45 °C and 30 min at 48 °C. The blot was exposed for 1 hr.

Southern hybridisation of rice genomic DNA

Portions of genomic DNA (20 µg) were digested with restriction endonucleases EcoRI, HindIII and BamHI, and the resulting fragments were separated by electrophoresis on a 0.7% agarose gel. DNA was transferred to nitrocellulose membrane following the capillary transfer procedure described by Maniatis *et al.* (1989). Hybridisation was carried out under the same conditions as for screening of the genomic libraries, using PCR-amplified maize *Sh-1* exon 13 sequence as DNA probe. After hybridisation, the filter was washed with 2× SSC, 0.05% SDS at 42 °C for 1 hr and with 1× SSC, 0.1% SDS at 65 °C for 15 min, and then exposed at -70 °C with intensifying screen.

Southern hybridisation of Cucurbita pepo genomic DNA

Portions of genomic DNA were digested with various restriction endonucleases, separated by electrophoresis on appropriate concentrations of agarose gels, and transferred to nitrocellulose membranes as above. The nitrocellulose filters were prehybridised for about 4 hrs at 42 °C in 15-25 ml of 6× SSPE, 5× Denhardt's solution, 0.5% SDS, 100 µg/ml denatured herring testes DNA and 50% formamide (Maniatis *et al.* 1989). Radioactive probes prepared by random labelling of PP2 cDNA were heated at 100 °C for 10 min, chilled on ice, and then added to the prehybridisation solution. Hybridisations were carried out at 42 °C overnight. DNA blots were rinsed briefly in 2× SSC, 0.5% SDS and then washed with 2× SSC, 0.1% SDS at room temperature for 15 min, in 0.1× SSC, 0.5% SDS at 37 °C for 30 min, and finally in the same buffer at 65 °C for 30-45 min. The blots were then exposed at -70 °C.

Nucleotide sequencing

Automated nucleotide sequencing was carried out by Ms. Julia Bartley on an Applied Biosystems model 373A DNA sequencer, based on the dideoxy chain termination method (Sanger *et al.* 1977) and PCR, using the Taq Dye Primer Cycle sequencing kit.

2.2.4. Library construction and screening

Construction of Cucurbita pepo genomic library

Partial digestion and purification of genomic DNA

Cucurbita pepo genomic DNA was partially digested with Sau3A I enzyme to yield DNA fragments with average size of approximately 17 kb. Portions (75 µg) of genomic DNA were mixed with 0.17, 0.15, 0.12 and 0.1 units of Sau3A I, respectively, in 1× NBL buffer No.2 in 200 µl volume, and incubated at 37 °C for 30 min. The reaction was stopped by adding 10 µl of 0.5 M EDTA (pH 8.0). To remove small amounts of contaminating carbohydrate, each reaction was mixed with equal volume of 2× CTAB buffer (see "Isolation of genomic DNA from *Cucurbita pepo*" section) and extracted with equal volume of chloroform:isoamyl alcohol. The aqueous phase was mixed with 1/5 volume of 5% CTAB and extracted again with chloroform:isoamyl alcohol. The DNA was precipitated with 2 volume of ethanol at -20 °C for 1 hr, pelleted, air-dried and resuspended in 20 µl distilled water. The DNA was re-pelleted with 7.5 µl of 10 M NH4Ac and 55 µl of ethanol at -20 °C for 1 hr, washed twice with 75% ethanol, air-dried and dissolved in 15 µl deionised water. Equal amounts of DNA from each reaction were mixed and used for fill-in reaction.

Cloning of genomic DNA into lambda vector

The partially digested and purified genomic DNA was cloned into the Promega λ GEM-12 XhoI half site arms vector and packaged using Promega Packagene extract following the supplier's instructions. The Sau3A I digested genomic DNA was partially filled-in using dATP and dGTP to yield GA-5' sticky ends. The arms of the lambda vector had been cut out with XhoI, and the ends partially filled-in using dTTP

and dCTP, leaving TC-5' cohesive ends which is compatible with the genomic DNA ends. This helps to avoid the religation of arms to stuffer fragment and multiple inserts of genomic DNA cloned. This also eliminates the process of size fractionation of Sau3A I digested DNA.

Fill-in reaction was set up by mixing 10 μl (13.6 μg) Sau3A I-digested DNA with 10× buffer containing 10 mM dATP and dGTP (supplied with the vector), 30 μl deionised water and 5 μl (5 units) of Klenow enzyme. This was incubated at 37 °C for 35 min, and then extracted twice with phenol:chloroform:isoamyl alcohol and twice with chloroform:isoamyl alcohol. The DNA was precipitated with 0.375 volume of 10 M NH_4Ac and 2 volumes of ethanol at -20 °C for 1 hr, pelleted by centrifugation at 1000g for 10 min, washed twice with 75% ethanol, air dried and resuspended in 10 μl deionised water. The concentration of the DNA was determined by UV spectrophotometry and made to 0.5 μg/μl for subsequent use.

Two test ligations were carried out using 0.5 μg vector DNA each and 0.5, 1 μg of the genomic DNA in 5 μl volume. The mixtures were incubated with 2.5 units T4 DNA ligase each at 4 °C for 2 days and 1 μl aliquot packaged with 10 μl of the Packagene extract at 22 °C for 2 hrs. The resulting packaging mixture was diluted to 100 μl with phage buffer and titrated using *E. coli* KW251 cells.

Scaled-up ligation was then set up with 0.75 μg vector DNA and 0.75 μg genomic DNA in 7.5 μl volume using 3.75 units of T4 DNA ligase, and this was incubated at 4 °C for 3 days, pooled with the remaining test ligation mixture and packaged using 2 × 50 μl Packagene extract, diluted and titrated as above. The library was amplified and stored.

Construction of Cucurbita pepo cDNA library

Cucurbita pepo cDNA library was constructed using Stratagene's ZAP-cDNA synthesis kit following the supplier's protocol. About 5 μg of poly (A) RNA from 3-5 days old pumpkin hypocotyls was used. Basically, the first strand cDNA was synthesised using a 50-base linker-primer containing a 18-base poly dT sequence and

an XhoI site, moloney-murine leukemia virus reverse transcriptase, and a nucleotide

mixture with 5-methyl dCTP which makes the first strand cDNA have a methyl group

on each cytosine base and thus protects the cDNA from restriction enzymes used in

subsequent cloning steps. Second strand cDNA synthesis employed RNase H to nick

the RNA bound to the first strand cDNA to produce a multitude of fragments which

serve as primers for DNA polymerase I, and the latter nick translates these RNA

fragments into second strand cDNA. The uneven termini of the double-stranded

cDNA were filled-in with T4 DNA polymerase and EcoRI adaptor was then ligated to

the blunt ends. The subsequent XhoI digestion released the EcoRI adaptor and

residual linker primer from 3' end of the cDNA, thus producing DNA fragments with

EcoRI and XhoI cohesive ends. These DNA fragments were purified on a Sephacryl

spin-column to remove the adaptor, the residual linker-primer and short cDNA species.

The size-fractionated cDNA was then concentrated by ethanol precipitation and ligated

to Uni-ZAP XR vector arms, and the resulting lambda library packaged in Gigapack II

Gold packaging extract and plated on *E. coli* cell line SURE.

Titration of DNA libraries

A host cell culture of appropriate *E. coli* strains (LE 392, NM 538, KW 251 for

genomic libraries; SURE for cDNA library) was prepared by inoculating with a single

bacterial colony 10 ml of LB medium supplemented with 0.2% maltose and 0.01 M

$MgSO_4$ and shaking at 37 °C overnight. Alternatively, host cell cultures could be

prepared by inoculating LB medium containing 0.2% maltose and 0.01 M $MgSO_4$ with

1/10 volume of a overnight cell culture and shaking at 37 °C for about 5 hrs. The cell

cultures were used directly or first pelleted by centrifugation at 4000g for 10 min at

room temperature and then resuspended in 0.01 M $MgSO_4$ to give an appropriate cell

density (O.D.$_{600} \approx 2$). Appropriate dilutions of DNA libraries (in bacteriophage λ

vectors) were made using SM phage buffer. Aliquots of 100 μl of each dilution were

mixed with 100 μl of the host cell culture and incubated at 37 °C (or 39 °C for SURE

cells) for 15 to 30 min. For plating, the infected cells were mixed gently but quickly

with 3 ml molten (47 °C) LB, TB, or NZY top-agar and immediately poured onto LB or NZY plates (9 cm diameter). The top agar was allowed to harden and then the plates were incubated inverted at 37 °C overnight. The titers were calculated from the number of phage plaques on every plate.

Amplification of DNA libraries

Host cell cultures were prepared as above. Aliquots of packaged mixtures or primary libraries containing about 10^5 bacteriophages were mixed in a tube with 0.6 ml of the cell culture (SURE cells for cDNA library; KW251 cells for genomic library) and incubated at 37 °C for 15 min (cDNA library) or 30 min (genomic library). The infected cells in each tube were plated as for titration onto a 22 × 22 cm square plate using 50 ml LB or TB top agar, and the plates incubated inverted at 37 °C for 6-7 hrs. After phage growth, the plates were moved to a biological hood, and the top agar was scraped off the plates and placed into sterile centrifuge tubes. For the scraped agar of each plate, 25 ml of SM phage buffer was added and the tubes were kept at 4 °C overnight. The agar debris was pelleted by centrifugation at 4000g for 10 min and the supernatant decanted into a fresh sterile tube. Chloroform was added to make a final concentration of 5% and the amplified libraries were kept at 4 °C. For longer storage, aliquots of the libraries were supplemented with 7% dimethyl sulphoxide, frozen in dry ice-ethanol bath and kept at -70 °C.

Screening of genomic libraries

Genomic libraries to be screened were plated as in the "Amplification of DNA libraries" section. On every 22 × 22 cm square LB plate, about $2 × 10^5$ phages were plated. The plates were incubated inverted at 37 °C for 6-10 hrs, then stored at 4 °C overnight. Nitrocellulose filters were prepared according to Slightom and Drong (1988). The filters (20 × 20 cm) were soaked in distilled water then in 1 M NaCl solution, and blot-dried with 3MM paper. They were then carefully placed on top of the chilled plates, marked with a needle and water-resistant ink, and allowed to remain

in contact with the plaques for about 4 min (or 5 min for replicate filters). The membranes were removed carefully from the plate and placed phage side up on a 3MM paper saturated with denaturing solution (1.5 M NaCl, 0.5 M NaOH) and left for 4 min. The membranes were blotted with a 3MM paper to remove excess moisture, and then placed phage side up for 4 min on a 3MM paper saturated with neutralisation solution (1.5 M NaCl, 0.5 M Tris-HCl pH 7.4). The filters were then soaked for 30 second in 3× SSC, blot-dried with 3MM paper and then baked in a 80 °C vacuum oven for 1 hr.

The filters were incubated at 42 °C for 2-5 hrs in prehybridisation solution (50% formamide, 5× SSPE, 5× Denhardt's, 0.1% SDS, 100 μg/ml denatured herring testes DNA). DNA probe labelled with ^{32}P was then added to the prehybridisation solution and the membranes incubated at 42 °C overnight. After hybridisation, the filters were washed at 42 °C for 1 hr in 2× SSC, 0.05% SDS, and, if the background was still high, washed at 60 °C in 1× SSC, 0.1% SDS for about 5 min, then exposed to X-ray film at -70 °C overnight or longer.

After exposure, the films were marked by alignment with the filters. Phage plaques responsible for hybridisation signals were picked by aligning the films and the original plates on a light box and removing an agar plug (about 0.5 cm in diameter) surrounding the area of each hybridisation signal with wide-bore blue tips. Each of the agar plugs was placed in 1 ml of SM phage buffer and kept at 4 °C overnight.

Phage responsible for the hybridisation signals was purified away from all the surrounding contaminating phage by a series of successive plating and hybridisation as described for the larger plates, however, on a smaller scale (9-cm plates).

Screening of Cucurbita pepo cDNA library

Plating of the cDNA library (~38,000 plaques) on SURE cells and preparation of nitrocellulose filters were carried out as described for screening of the genomic libraries. The nitrocellulose filters containing phage DNA were incubated for 5 hrs at 40 °C in 25 ml of prehybridisation solution (6× SSC, 2× Denhardt's solution, 0.25%

SDS, 100 µg/ml denatured herring sperm DNA). The end-labelled oligo probe (see "Hybridisation of Northern blot using degenerate oligonucleotide probe" section) was heated for 2 min in a boiling water bath, chilled on ice, and then added to the prehybridisation solution. The filters were incubated for 18 hrs at 40 °C, and then washed with 4 changes of washing buffer (6× SSC, 0.1% SDS) at room temperature. The hybridisation was visualised by autoradiography. Plaques corresponding to strong hybridisation signals were purified by a further round of screening as described for screening genomic libraries using the same hybridisation and washing conditions.

In vivo self excision of pBluescript from Uni-ZAP XR vector

Positive phages isolated from the *Cucurbita pepo* cDNA library were subjected to *in vivo* self excision to form insert-containing pBluescript phagemids, which were then used in subcloning and sequencing. This was carried out using the ExAssist helper phage following the procedure in the Stratagene's ZAP-cDNA synthesis kit. One hundred µl of phage stock was mixed with 200 µl of overnight, 30 °C-grown *E. coli* XL1-blue cells and 1 µl of ExAssist helper phage, and incubated at 37 °C for 20 min. The mixture was transferred to 3 ml of 2× YT medium and incubated at 37 °C for 2.5 hrs with shaking. The *E. coli* cells were killed by heating at 70 °C for 20 min and removed by centrifugation at 4000g for 5 min. One µl of the saved supernatant (containing phagemids) was mixed with 200 µl of overnight, 30 °C-grown *E. coli* SOLR cells and incubated at 37 °C for 15 min. Aliquot (100 µl) of the cell culture was plated on LB-ampicilin (50 µg/ml) plate and incubated overnight at 37 °C.

2.2.5. Protein analysis and histochemical assay

Extraction of proteins from plant tissue for immuno dot blotting

Young leaf tissue (300 mg) was ground with a glass rod in an Eppendorf tube with 200 µl of PBS buffer. The extract was centrifuged at 10,000g for 5 min and the

supernatant transferred to a fresh tube. The concentration of protein in the supernatant was determined by Bradford assay (Bradford 1976).

Extraction of proteins from plant tissue for fluorometric assay

This was carried out as above using extraction buffer (Shaw 1988) containing 50 mM NaH_2PO_4, 10 mM β-mercaptoethanol, 10 mM EDTA, 0.1% N-laurylsarcosine and 0.1% Triton X-100 (pH 7.0). For every 100 mg of plant tissue 100 μl of the buffer was used.

Extraction of proteins from plant tissue for Western blotting

About 100 mg of plant tissues was harvested, frozen in an Eppendorf with liquid nitrogen and ground into fine powder using a glass rod. Before the frozen powder was thawed, 20 μl of protein extraction buffer (Scott *et al.* 1988) (10% glycerol, 40 mM EDTA, 150 mM NaCl, 100 mM Tris-HCl pH 7.5, 3 mg/ml dithiothreitol) was added to the tube and the suspension homogenised with the glass rod. Plant debris was removed by centrifugation and the supernatant was transferred to a new tube and stored at -20 °C.

Collection of phloem exudate of C. pepo

Stems or petioles of *C. pepo* plants were cut with razor blade, and, to minimise possible contamination from non-sieve element cells, the cut surfaces were blotted immediately with filter paper and the first droplet of exudate was discarded. Exudate was then collected into appropriate buffers for subsequent use.

Dot blotting of proteins

A nitrocellulose membrane was cut to appropriate size, wetted in distilled water and then assembled in a dot blotting apparatus following the manufacturer's instructions. Two hundred μl of protein samples (10 μg total protein) was applied to each sample cells of the blotter and the liquid sucked gently through the nitrocellulose membrane

under small vacuum. The cells were washed with 200 μl PBS buffer by sucking the buffer through the membrane. The apparatus was then dissembled following the instruction and the nitrocellulose membrane used for immunological detection of proteins.

Phloem exudate of pumpkin plants and honeydew from rice brown planthopper (donated by Dr. Kevin Powell) were blotted using micropipette tips. About 1-2 μl of samples was applied to nitrocellulose membrane and allowed to dry, and then another 1-2 μl of samples applied to the same spot on the membrane and allowed to dry again. This was continued until all the sample was blotted.

Preparation of non-denaturing polyacrylamide gel

Clean glass plates were assembled with spacers held in position with fold-back clips and sealed by applying grease on the spacers or with sealing gasket. The following components were mixed in order in a flask and then degassed for 1 min: acrylamide/bisacrylamide (30:0.8%), 6.7 ml; 1 M Tris-HCl pH 8.8, 7.5 ml; distilled water, 3.5 ml. After degassing, 10 μl of TEMED (NNN'N'-tetramethylethylenediamine) and 1 ml of 1.5% ammonium persulphate were added, mixed briefly and the solution was poured between the glass plates leaving enough space for the comb and stacking gel. The gel was carefully overlaid with distilled water and left to polymerise at room temperature until a sharp line was seen between the gel and overlay. The overlaying water was then tipped off and the excess liquid removed with filter paper. The stacking gel was prepared immediately prior to use following the same procedure as for separating gel with following components: acrylamide/bisacrylamide (30:0.8%), 1.25 ml; 1 M Tris-HCl pH 6.8, 1.25 ml; distilled water, 6.1 ml; TEMED, 2.5 μl; 1.5% ammonium persulphate, 0.5 ml. The gel solution was filled into the glass plates, well-former (comb) inserted, and allowed to polymerise for about 15 min.

Preparation of SDS-polyacrylamide gel

SDS-polyacrylamide gels were prepared following the same procedure as for non-denaturing gels. The only difference was that both the separating and stacking gels contained SDS (sodium dodecyl sulphate) at a final concentration of 0.1%.

Polyacrylamide gel electrophoresis of proteins

Comb, bottom spacer and sealing gasket were removed from gel plates, and the sample wells rinsed with 1× electrophoresis buffer (25 mM Tris-HCl, 250 mM glycine, pH 8.3, plus 0.1% SDS for SDS gel). The gel plates were then mounted in an ATTO AE-6450 mini-gel (8 × 10 cm) electrophoresis apparatus or in a self-made apparatus for larger gel (15 × 17.5 cm), and the two reservoir tanks were filled with the Tris-glycine buffer. For non-denaturing gel electrophoresis, the protein sample described in "Extraction of proteins from plant tissues for Western blotting" section was mixed with 1 μl of bromophenol blue solution (0.1% in 25% glycerol) and loaded directly to the gels; pumpkin phloem exudate (36 μl) was collected into 12 μl of buffer containing 40% glycerol, 0.2 M Tris-HCl pH 6.8, 5% β-mercaptoethanol and 0.4% bromophenol blue, and loaded. For SDS gel, protein samples were mixed with equal volumes of 2× SDS-PAGE sample buffer and heated at 100 °C for 2-3 min before being loaded to the wells of the gel. Electrophoresis was carried out at 100 or 200 constant volts for the mini-gel or larger gel, respectively. The molecular weight marker used consisted of 7 proteins (SDS-7, Sigma): bovine albumin, 66,000; ovalbumin, 45,000; glyceraldehyde-3-phosphate dehydrogenase, 36,000; carbonic anhydrase, 29,000; trypsinogen (bovine pancreas), 24,000; trypsin inhibitor (soybean), 20,000; α-lactalbumin, 14,200. After electrophoresis, the gels were either stained with Coomassie blue or used for Western blotting.

Staining of proteins on polyacrylamide gel

After electrophoresis, gels were incubated at room temperature with gentle shaking for 5-15 min in staining solution (0.5% Coomassie blue in acetic acid-isopropanol-water,

1:3:6, v/v/v) and then in several changes of destaining solution (acetic acid-methanol-water, 10:45:45) for 1-2 hrs. Stained gels were stored in 20% glycerol.

Western blotting of SDS polyacrylamide gel

Proteins separated on SDS-polyacrylamide gel were transferred to nitrocellulose membrane using an ATTO semi-dry electroblotting apparatus. Onto the anode of the apparatus the following were placed in order: 2 sheets of 3MM paper soaked in 0.3 M Tris pH 10.4, 20% methanol, 0.1% SDS; 1 sheet of 3MM paper soaked in 25 mM Tris pH 10.4, 20% methanol, 0.1% SDS; 1 sheet of nitrocellulose membrane (cut to size and wetted in distilled water); the gel; 1 sheet of 3MM paper soaked in 25 mM Tris, 40 mM 6-amino-N-hexanoic acid pH 9.4, 20% methanol, 0.1% SDS; 1 sheet of dialysis membrane (wetted in distilled water) and 2 sheets of 3MM paper soaked in 25 mM Tris, 40 mM 6-amino-N-hexanoic acid pH 9.4, 20% methanol, 0.1% SDS. The cathode was then placed on top of this and the proteins electroblotted at constant current ($\sim$2.5 mA.cm^{-2}) for 1 hr. The efficiency of protein transfer was checked by staining the gel with Coomassie blue after blotting.

Western blotting of non-denaturing polyacrylamide gel

Proteins were transferred from non-denaturing polyacrylamide gels to nitrocellulose membrane according to Scott et al. (1988) using the same apparatus as for SDS gel. After electrophoresis, the gel was equilibrated in TGM buffer (1 litre contains 3.0 g Tris-base, 14.4 g glycine and 200 ml methanol) for 30 min at 4 °C. All the filter papers and nitrocellulose membrane were thoroughly soaked in TGM buffer, and these and the gel were then assembled to the blotting apparatus in order: anode; 2 sheets of 3MM paper; 1 sheet of nitrocellulose membrane; the gel; 1 sheet of 3MM paper; 1 sheet of dialysis membrane; 1 sheet of 3MM paper and cathode. Air bubbles trapped between layers were rolled out using a large pipette. Electroblotting was performed at 200-250 V for 1 hr.

Immunological detection of snowdrop lectin (GNA) on nitrocellulose membrane

Nitrocellulose filters from dot and Western blotting were incubated in PBS buffer containing 5% fat-free milk powder (Marvel) and 0.1% Tween 20 overnight at 4 °C or at least 1 hr at room temperature to block non-specific protein binding sites. The filters were rinsed with PBS buffer and treated for 1 hr at room temperature with anti-GNA antibody at 1:5000 dilution in PBS buffer containing 5% fat-free milk powder and 0.01% Tween 20. The excess antibody was washed off with 4 changes of antibody dilution buffer at room temperature for a total of 20 min. Goat anti-rabbit IgG horseradish peroxidase conjugate (Bio-Rad) was used as secondary antibody (1:3000 dilution, as above) to treat the membrane for 1 hour at room temperature followed by 4×5 min washes in antibody dilution buffer, and a rinse in distilled water. Enhanced chemiluminescence (ECL) reagents (Amersham) were used to detect the specifically bound secondary antibody as instructed by the manufacturer.

Immunological detection of sucrose synthase protein on nitrocellulose membrane

The nitrocellulose membranes containing proteins were immersed in PBS buffer containing 1% BSA at 4 °C overnight to block non-specific binding sites, and then incubated in 2.28% periodic acid at room temperature for 25 min to deactivate endogenous peroxidase. The membranes were washed with PBS for 15 min and then treated with anti-sucrose synthase antibody (1:1000 in PBS-1% BSA) at room temperature for 1 hr. Free and nonspecifically bound antibody was washed off the membrane with 6 changes of PBS-0.5% Triton X-100 for 25 min. The blots were then treated with horseradish peroxidase-conjugated goat anti-rabbit antibody (1:2000 in PBS-1% BSA) at room temperature for 50 min with shaking, and washed with PBS-0.5% Triton X-100 as above. The protein-antibody binding was visualised by staining the blots for 5-10 min in DAB-colbat-hydrogen peroxide solution (freshly made by dissolving 6 mg of 3,3'-diaminobenzidine-HCl [DAB] in 9.7 ml of 50 mM Tris-HCl pH 7.6, adding to above 0.3 ml of 1 M cobalt chloride and 10 μl of 30% hydrogen peroxide, and filtering through filter paper).

Immunochemical localisation of GNA

Plant tissue from *RSs1*-GNA tobacco plants was cut with a razor blade into 3-5 mm sections, quickly frozen by CO_2, and mounted on a metal block with cryoembedding medium. The tissue was cut into 5 mm sections on a cryomicrotone. Sections were placed on slides which had been precoated in 2% 3-aminopropyltriethoxysilane and air dried. Sections were incubated in 1% hydrogen peroxide in absolute methanol for 15 min at room temperature to quench any endogenous peroxidase activity. The slides then were rinsed in PBS buffer for 4×5 min, and treated with primary and secondary antibodies as described above, except that DAB was used as substrate to detect the antigen (see above section). After staining the slides were rinsed with distilled water and air dried, and mounted for viewing and photography.

Histochemical localisation of GUS expression

Histochemical staining was performed according to the method described by Jefferson *et al.* (1987). Fresh sections were hand cut from leaf, petiole and stem of transgenic tobacco plants. These sections, together with fractions of root, were incubated in a staining solution, containing 1 mM X-gluc and 50 mM sodium phosphate buffer (pH 7.0) for 12 hrs at 37 °C, then cleared of chlorophyll by incubation in a solution of 5% formaldehyde, 5% acetic acid and 20% ethanol for 10 minutes, followed by incubation in 50% and 100% ethanol, 2 minutes each. The sections were mounted on microscope slides for photography.

Fluorometric assay of GUS expression

Aliquots (10 µl) of protein extracts were transferred to a clean Eppendorf tube and mixed thoroughly with 190 µl of reaction buffer (Shaw 1988) (prepared by dissolving 2.2 mg of 4-methyl β-D-umbelliferyl glucuronide [MUG] directly in 10 ml of the protein extraction buffer [see above section], prewarmed to 37 °C). The reaction was carried out at 37 °C for 40 min (variable) and stopped by addition of 9 volumes of stop

buffer (0.2 M Na_2CO_3). The amount of reaction products was determined with a spectrofluorimeter, using 365 nm as excitation wavelength and measuring emission at 455 nm.

Immuno-detection of GNA in honeydew produced by aphids

RSs1-GNA transgenic tobacco plants were infested with peach potato aphids *(Myzus persicae)* and kept in a controlled environment growth chamber for a week. The small amount of honeydew produced by these aphids was collected as follows: a 3 cm diameter petri dish was lined with a moist filter paper, and another 3MM filter paper or nitrocellulose membrane was taped about 3-5 mm away from the top edge of the container to ensure a close contact between the filter and the aphids. Fifty aphids were gently placed on the filter and a layer of stretched parafilm was used to seal the container. A portion of young leaf from *RSs1*-GNA (experimental) or untransformed (control) plant was placed on the parafilm and sealed by another layer of the same film. Aphids fed through the parafilm on the leaf and the honeydew produced was absorbed by the filter or nitrocellulose membrane. The membrane and filter were removed after 24 hrs. The 3MM filters with absorbed honeydew were sprayed briefly with ninhydrin in 0.2% acetone to detect spots of honeydew by the presence of free amino acids (Rosen, 1957). The nitrocellulose membranes were treated with primary and secondary antibodies as described for dot immunoblotting, using ECL reagents to detect specific binding of the labelled secondary antibody.

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of phloem proteins

Phloem exudate was collected from stems of 3 months old pumpkin plants and transferred with a pipette directly into equal volume of 2× SDS-PAGE sample buffer containing 5% mercaptoethanol. SDS-PAGE was carried out as described above. The separating gel (12% acrylamide, 15 × 17.5 cm, 0.75 mm or 1.5 mm thick) was allowed to "cure" for at least 14 hrs prior to use, and was subjected to pre-electrophoresis at 150-200 V for 1 hr. The running buffer was always supplemented with 0.1 mM

sodium thioglycolate (Hunkapiller *et al.* 1983; Moos *et al.* 1988). Before loading, the exudate sample was heated at 60 °C for 10-20 min. Electrophoresis was conducted at 150-200 V.

Purification of phloem protein PP2 by SDS-PAGE

Phloem protein PP2 separated by SDS-PAGE was purified by electrophoretic elution using a self-made electroeluter. The eluter consisted of two electrode chambers electrically connected by a bridge made of two Eppendorf tubes (as elution cells), bottom cut, sealed with dialysis membrane at the bottom and connected with a plastic tube filled with elution buffer. After SDS-PAGE, the gel was stained with Coomassie blue at room temperature for 5 min and destained at 4 °C for 5 min, and the PP2 band excised with a razor blade. Electrophoretic elution was performed following the procedure described by Hunkapiller *et al.* (1983). After electroelution, the protein solution in the elution cell in the anode chamber was transferred to a fresh Eppendorf tube and freeze-dried.

Electroblotting of phloem proteins to PVDF membrane

For peptide sequencing, proteins separated by SDS-PAGE were electrophoretically transferred to polyvinylidene fluoride (PVDF) membranes according to LeGendre and Matsudaira (1989) using the ATTO semi-dry electroblotting apparatus. The PVDF membrane was prewetted by brief immersion in 100% methanol (2-3 sec), rinsed with water and equilibrated in transfer buffer (10 mM CAPS [3-(cyclohexylamino)-1-propanesulfonic acid], 10% methanol, pH 11) for a minimum of 15 min. All the 3MM paper and dialysis membrane were cut to the same size as the gel and soaked in the transfer buffer before use. After electrophoresis, the gel was rinsed in the transfer buffer for 5 min and assembled into the blotting apparatus in following order: anode; 3 sheets of 3MM paper; 1 sheet of ProBlott membrane; the gel; 1 sheet of 3MM paper; dialysis membrane; 4 sheets of 3MM paper and cathode. Electroblotting was conducted at 150-200 mA for 1 hr. After blotting, the ProBlott membrane was rinsed

in distilled water, stained briefly (about 1 min) in 0.1% Coomassie blue R in 50% methanol, destained with several changes of 50% methanol, 10% acetic acid for about 25 min, rinsed with several changes of distilled water for a period of 15 min, air-dried and stored at -20 °C.

In situ cleavage of PP2 protein in gel slice or on PVDF membrane with cyanogen bromide (CNBr) vapour

In situ cleavage of PP2 in gel slices was done by a modification of the procedure by Zingde *et al.* (1986). PP2 protein band on SDS gel was cut out with a razor blade and lyophilised in a small glass container, which was then placed inside a small vacuum flask containing 1 ml of 50% trifluoroacetic acid and 20 mg of CNBr at its bottom. The flask was flushed with argon gas, evacuated briefly until the liquid just started to boil, and then kept sealed in the dark at room temperature for 24 hrs. PP2 blotted on PVDF membrane was treated with CNBr vapour following the same procedure as for the lyophilised gel slices.

To separate the cleaved peptides by SDS-PAGE, the CNBr-treated gel slices was removed from the vacuum flask and lyophilised, and then suspended in several changes of 1× SDS-PAGE sample buffer until the bromophenol blue dye in the buffer remained blue. The gel slice was finally suspended in 0.4 ml of the sample buffer, heated for 10 minutes at 60 °C, loaded directly into the well of a 17% SDS-PAGE slab gel, overlaid with sample buffer, and electrophoresed as described in previous section.

Elution of proteins from PVDF membrane

The following three methods were recommended by Applied Biosystems, and the first two were carried out by John Gilroy: (1) PVDF blot was soaked in 75% hexafluoro-2-propanol for 1 hr at room temperature; (2) PVDF blot was incubated in 20% H_2O + 75% isopropanol + 5% trifluoroacetic acid for 1 hr at room temperature; (3) the blot was incubated for 20 min at 100 °C in 62 mM Tris-HCl pH 6.8, 2% Triton X-100, 3% SDS, 6 M urea, 2% β-mercaptoethanol and 10% glycerol. Proteins in (1) and (2) were

concentrated by spin-drying, whereas the eluates in (3) were directly loaded to SDS-polyacrylamide gels for protein separation.

2.2.6. Plant transformation

Construction of expression vectors

Using RSs1 promoter

An NheI fragment, including the 5' flanking region, the first intron and the first 16 nucleotides of coding sequence of the *RSs1* gene was isolated from clone RSS2.4 (Fig. 3.5) and cloned into the XbaI site of pUC19. The fragment was excised by SalI + BamHI double digestion, and inserted into pBI101.2 and the resulting plasmid was designated as pBRSS7. A cDNA sequence encoding LECGNA2, one of the GNA isoforms (van Damme *et al.*, 1991a; 1991b) was amplified by PCR from plasmid pT7T318U, using the primers: (5' CGGATCCATGGCTAAGGCAAGTCTCCTC 3'; 3' GAACACTGCCGTTTCATTACTCCATGGC 5'). The primers were designed in such a way that the preprotein coding sequence was amplified. The PCR fragment was cloned into the HincII site of pUC19 and sequenced to confirm no PCR errors were present. The fragment was excised by SmaI and PstI digestion, followed by inserting into a vector derived from pBRSS7 by removing GUS coding sequence with SmaI and SacI. The SmaI and one of the two BamHI sites between the *RSs1* promoter and GNA in the intermediate vector were eliminated by BamHI digestion and subsequent self-ligation. The resulting vector, containing a chimaeric *RSs1*-GNA gene, was designated as pBRSSLT. The constructs were checked by restriction mapping and DNA sequencing.

Using PP2 promoter

A plasmid, pCP1.3B12A containing the 1.3 kb BamHI fragment (from the BamHI site in the left arm of the λGem-12 vector to the first BamHI site inside the coding sequence) of the PP2 genomic clone CPP1.3 (Fig. 4.10 and Fig. 4.11) was digested with HaeII (the HaeII site is inside the 5' leader sequence of the PP2 gene, as shown in

70

Fig. 4.11) and blunt-ended with Mung Bean nuclease, and then digested with HindIII. The resulting 1.2 kb HindIII/blunt-end fragment was purified and cloned into HindIII + HincII double digested pUC19. The insert of the resulting plasmid (pCPM3) was excised by BamHI digestion and ligated into the BamHI site of pBI101.1. The orientation of the insert in the vector was determined by EcoRI or XbaI digestion, and the fusion site between the PP2 promoter and the GUS coding sequence was checked by nucleotide sequencing using a primer complementary to the GUS coding sequence from +32 to +13 bp from the A of the GUS ATG initiation codon. The constructs was designated as pCPGUS2.

Conjugation of recombinant plasmids into Agrobacterium through triparental mating

A helper strain of *E. coli* (HB101:pRK2013) and a bacterial strain containing a recombinant plasmid (e.g. *E. coli* JM101:pCPGUS2) were grown overnight at 37 °C from a single colony in 5 ml of LB-kanamycin (50 µg/ml) and YEB-kanamycin (50 µg/ml), respectively. Agrobacterial strain *A. tumefaciens* LBA4404 was grown overnight at 28 °C from a single colony in 5 ml of YEB-rifampicin (100 µg/ml). One hundred µl of each of the 3 strains to be mated was pipetted onto one YEB plate and mixed by spreading. This was repeated to give triparental mating in duplicate. The plates were incubated overnight at 28 °C. A loop of the cells was removed from each plate and streaked onto a YEB plate containing both kanamycin (50 µg/ml) and rifampicin (100 µg/ml), and the plate incubated at 28 °C until colonies of transconjugant had grown to substantial size (~3 days).

Transformation of tobacco using leaf disks

Preparation of leaf disks

Sterile young (just fully expanded) tobacco leaves from 3-month old plants grown in MSO medium were placed on sterile filter papers and cut into 0.5-1 cm squares using a new sterile scalpel blade. About 12-18 leaf pieces were placed onto each of 20 MSD4

x2 plates, which were then sealed with Nescofilm and incubated at low light intensity (2000 lux, 16-h photoperiod) for 2 days at 25 °C.

Inoculation

The leaf pieces were dipped into diluted 2-day culture of an appropriate transconjugant *Agrobacterium* strain (1:50 dilution with MSO liquid medium) and then returned to original MSD4x2 plates, which were sealed and incubated as above for 2 days. Three plates were set up for each strain of *Agrobacterium*.

Shoot and callus formation

After the incubation with *Agrobacterium*, the inoculated leaf pieces were transferred to the following medium (one set for each strain): 3 plates of MSD4x2 + augmentin (200 μg/ml) + kanamycin (100 μg/ml); 1 plate of MSD4x2 + augmentin (200 μg/ml) (as control). These plates were sealed and incubated as above and observed periodically for development of shoots and callus.

Rooting

After 4-6 weeks of incubation, the leaf pieces were removed from the plates and placed in a Petri dish; the shoots were excised, trimmed leaving the apical bud and 2-3 leaves with axillary buds, transferred rapidly to rooting medium (MSO + augmentin [200 μg/ml] + kanamycin [100 μg/ml]), and incubated at medium light intensity (4000 lux, 16-hr photoperiod).

Re-establishment of cultured plantlets in compost

Rooted shoots were gently removed from medium and the agar sticking to the roots was washed off. The plantlets were then placed in appropriately sized holes in compost-filled small plant pots, and the roots covered with compost. The potted plantlets were first kept in a small plastic growth chamber to avoid desiccation and then transferred to a normal growth room environment when they became hardened.

CHAPTER 3.   ISOLATION OF THE RICE SUCROSE SYNTHASE-1 GENE PROMOTER

3.1. Results

3.1.1. DNA probe

In order to obtain a DNA probe for the isolation of the rice sucrose synthase genes, polymerase chain reaction (PCR) was carried out to amplify exon 13, the longest exon sequence of the maize sucrose synthase-1 (*Sh1*) gene. The two primers for the PCR were a 19-nt (primer A) and an 18-nt (primer B) synthetic oligonucleotides (see section 2.2.2), corresponding respectively to nucleotides +4264 to +4282 of the non-coding strand and +4564 to +4581 of the coding strand of the published *Sh1* sequence (Werr *et al*. 1985). The amplified products (Mr approximately 320 bp) were purified from agarose gel and cloned into SmaI site of pUC18, and two of the resulting clones were sequenced. One product had an identical sequence to the *Sh1* exon 13, although deletion of a few bases occurred at both the 5'- and 3'-primer regions (Appendix A), possibly due to the treatment with T4 DNA polymerase. This clone was designated pSHEX and its DNA insert used as a heterologous probe for identification of rice sucrose synthase genes.

3.1.2. Southern blot analysis

When a Southern blot of rice genomic DNA was probed with the insert of pSHEX, at least three hybridisation fragments were detected (Fig. 3.1), suggesting the presence of a family of three or more different sucrose synthase genes in the rice genome. One band in every digest (i.e. ~2.2 kb in EcoRI, ~16.8 kb in HindIII and ~ 6.8 kb in BamHI digests) hybridised more strongly than the others, indicating that one
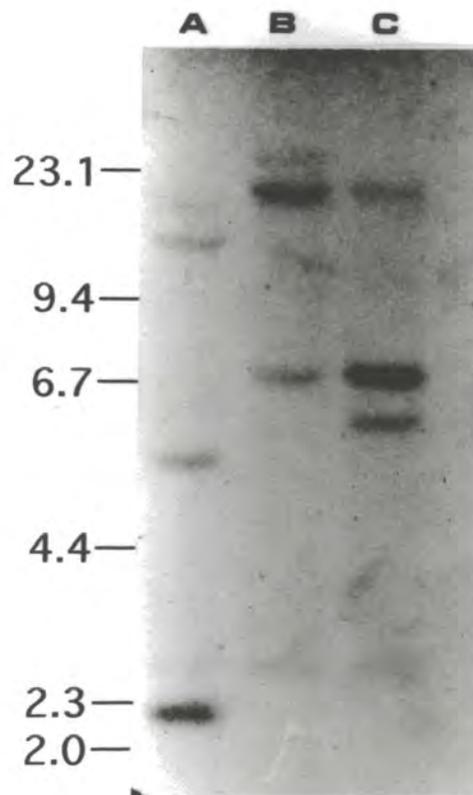
Figure 3.1. Southern blot of rice genomic DNA (20 µg per track) digested with EcoRI (trackA), HindIII (track B), and BamHI (track C), probed with exon 13 of the maize *Sh1* gene. Molecular weight markers (HindIII-digested λDNA, kb) are shown at the left.

member of the gene family was more homologous to the *Sh1* gene or had higher copy number in the genome.

3.1.3. Isolation and preliminary characterisation of rice sucrose synthase genomic clones

The DNA insert of pSHEX was used to screen a λEMBL3 library (commercially available from Clontech), which was constructed from genomic DNA of dark-grown seedlings of rice *Oryza sativa* L. Japonica. Out of 6.8×10^5 phage plaques (approximately 3.4×10^5 recombinants) ten were hybridised positively to the probe, of which eight were purified by two rounds of rescreening. The eight clones, designated as RSS1.1, RSS1.2, RSS1.3, RSS2.1, RSS2.2, RSS2.4, RSS2.5, and RSS4.1, were analysed with restriction endonucleases SalI, EcoRI and BamHI, and the resulting rough restriction maps (Fig. 3.2) showed that they could be divided into two groups. One group comprised of three clones (RSS1.2, RSS2.1 and RSS2.5), in which no BamHI site was found in the DNA inserts, and every insert gave a 4.7 kb EcoRI/SalI fragment which was hybridised to the pSHEX probe (Fig. 3.3). The other group, comprising the remaining 5 clones (in which RSS4.1 had an identical restriction map as RSS1.3), was characterised by having several BamHI and EcoRI sites in the inserts and giving several EcoRI-digested fragments with identical sizes (e.g. 2.5, 2.1, 1.7 and 0.9 kb), of which the 2.1 kb fragment was hybridised with the probe. The only differences among the members in each group lies in the different length or orientation of the inserts in the λ vector, or in the covering of different regions of the genes, thus each of the two groups should represent a single but different gene.

The two clones, RSS2.1 and RSS2.4, representing the two different groups and most likely containing full-length sucrose synthase genes, judging by the positions of the hybridisation fragments (4.7 kb EcoRI/SalI and 2.1 kb EcoRI/EcoRI) in their inserts, were further analysed by restriction mapping and partial DNA sequencing. A 0.7 kb PstI/PstI fragment, within the 2.1 kb EcoRI/EcoRI sequence of RSS2.4; and a
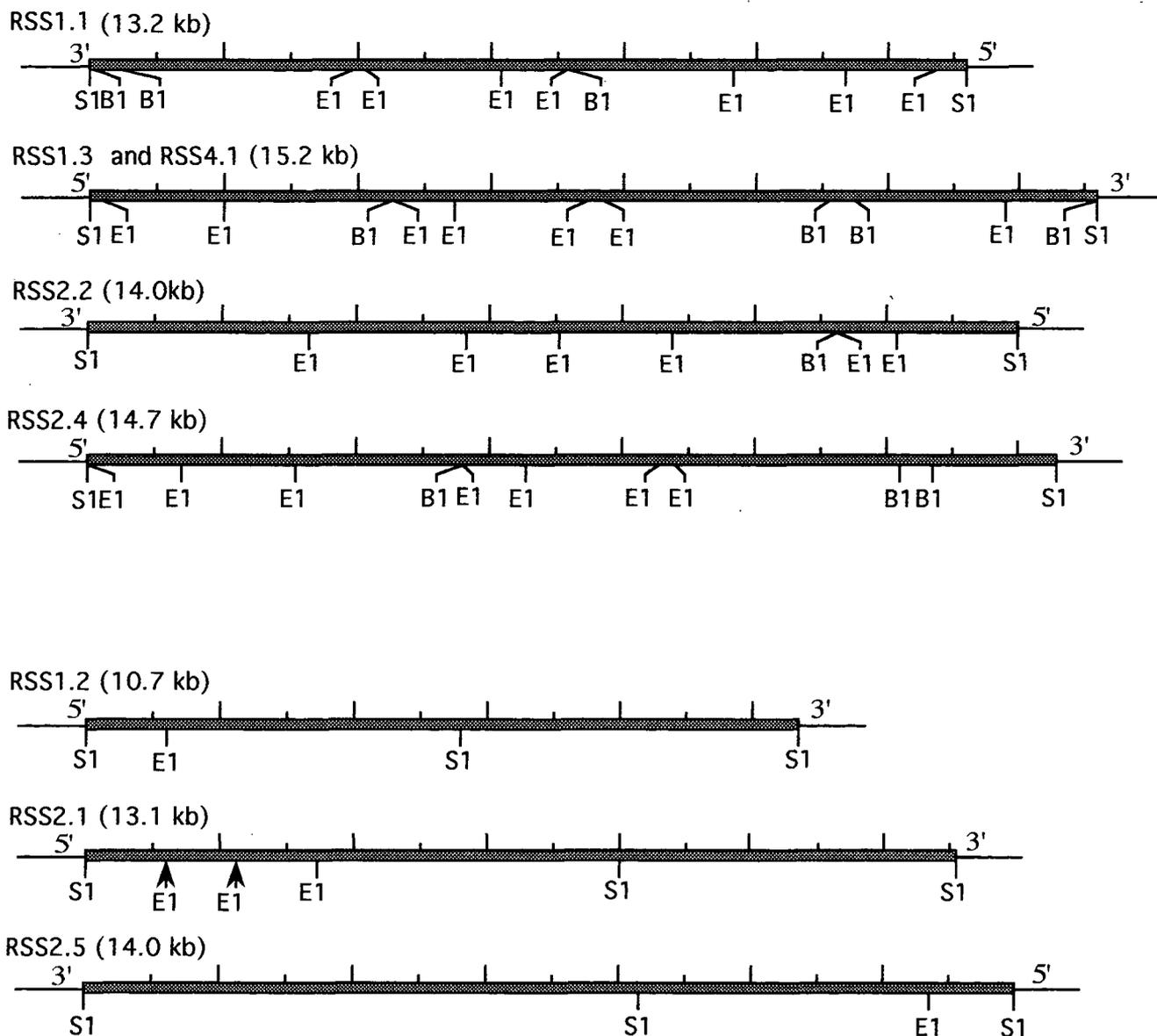
Figure 3.2. Restriction enzyme maps of the rice sucrose synthase genomic clones. E1 = EcoRI, B1 = BamHI, S1 = SalI. The SalI sites at the two ends of the inserts come from the λEMBL3 cloning vector. The arrows indicate the uncertain locations of the EcoRI sites.
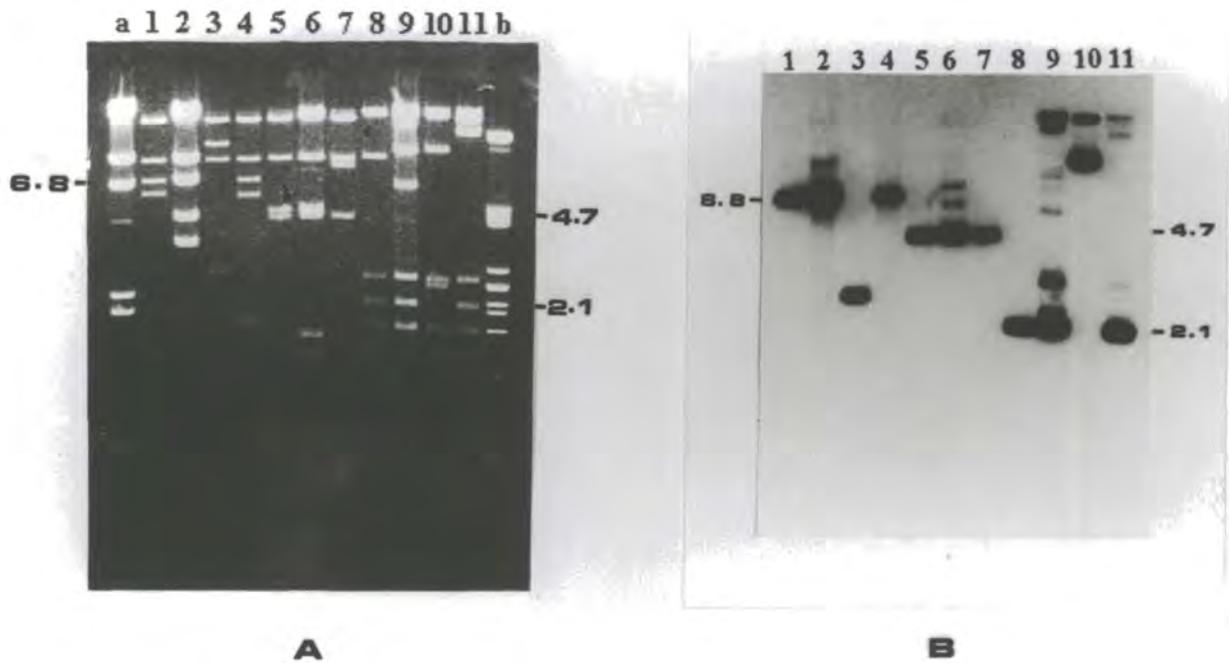
Figure 3.3. Sucrose synthase genomic clones RSS 1.1 (lane 1 and 8), RSS1.3 (lanes 2 an 9), RSS2.2 (lanes 3 and 10), RSS2.4 (lanes 4 and 11), RSS1.2 (lane 5), RSS2.1 (lane 6) and RSS2.5 (lane 7) were digested with SalI/BamHI (lanes 1-4), EcoRI/SalI (lanes 5-7), and EcoRI (lanes 8-11), separated on a 0.65% agarose gel (A), and hybridised to exon 13 sequence of the maize *Sh1* gene (B). Lanes a and b, molecular weight markers (HindIII- and PstI-digetsed λ DNA, respectively, see Appendix B for size of the DNA fragments). Molecular size (kb) is given to fragments referred in the text.

1.0 kb SacI-digested fragment, within the 4.7 kb EcoRI/SalI sequence of RSS2.1 (Fig. 3.4), were subcloned into a M13 vector and sequenced. The resulting sequence revealed that both fragments contained open reading frames corresponding to exon sequences (i.e. exons 13 and 14 for the 0.7 kb PstI fragment and exons 10, 11 and 12 for the other) of the *Sh1* gene. These partial coding sequences exhibited high homology to their counterparts in the *Sh1* gene in both nucleotide sequence (83.6% between RSS2.4 and *Sh1*, 77.2% between RSS2.1 and *Sh1*) and deduced amino acid sequence (92.2% between RSS2.4 and *Sh1*, 84.7% between RSS2.1 and *Sh1*), confirming that both RSS2.4 and RSS2.1 were sucrose synthase genes. The 2.1 kb EcoRI/EcoRI and the 6.8 kb BamHI/BamHI hybridisation bands of RSS2.4 in Fig. 3.3 correspond to the major hybridisation bands on the Southern blot (see Fig. 3.1). This and its high sequence homology with the maize *Sh1* gene indicated that RSS2.4 was the sucrose synthase 1 gene (*RSs1*). The partial sequence of RSS2.1 were later found to be identical to that of the recently published rice sucrose synthase-2 gene (*RSs2*) (Yu *et al.* 1992).

The orientations of the rice sucrose synthase genes in RSS2.4 and RSS2.1 were determined by sequencing the 0.35 kb PstI/EcoRI fragment of RSS2.4 and the 0.35 kb SacI/BglII fragment of RSS2.1 (Fig. 3.4), which contained coding sequences corresponding to exon 15 and exon 14 of the *Sh1* (or *RSs2*) gene, respectively. Based on the locations of these two fragments in the DNA inserts and the structure of the published maize *Sh1* and *RSs2* genes, it was estimated that the DNA insert of RSS2.4 extended far enough in both upstream and downstream region to contain a full-length gene including a full-length promoter, whereas RSS2.1 and the other two clones of its group were of insufficient length to include a full-length promoter.

3.1.4. Sequence of the rice sucrose synthase-1 gene

A detailed restriction map of RSS2.4 was generated (Fig. 3.5c) by subcloning and restriction digestion of the large restriction fragments (i.e. the 7.0 kb BamHI/BamHI, 5.7 kb SalI/BamHI, 2.1 kb EcoRI/EcoRI fragment, see Fig. 3.2) and
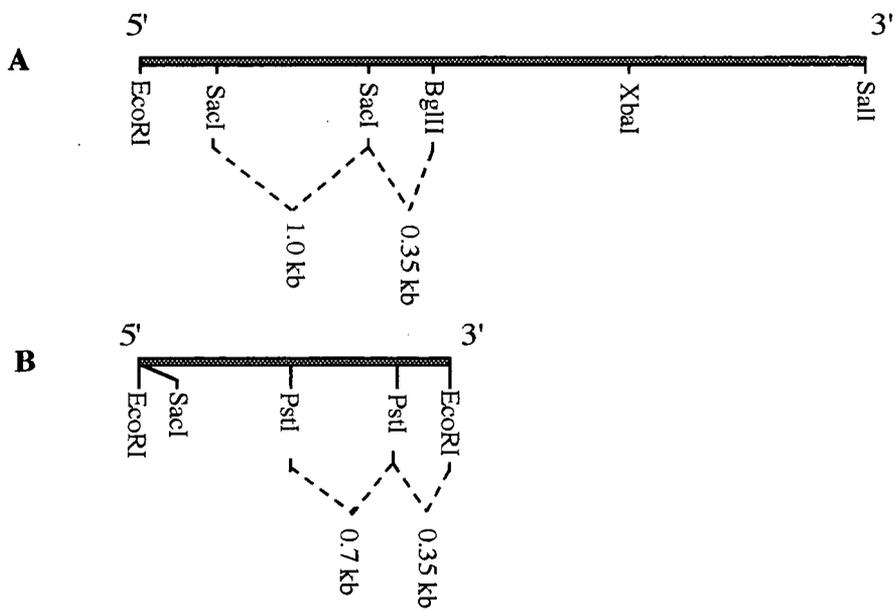
Figure 3.4. Restriction enzyme maps of the 4.7 kb EcoRI/SalI fragment of RSS2.1 (A) and the 2.1 kb EcoRI/EcoRI fragment of RSS2.4 (B).
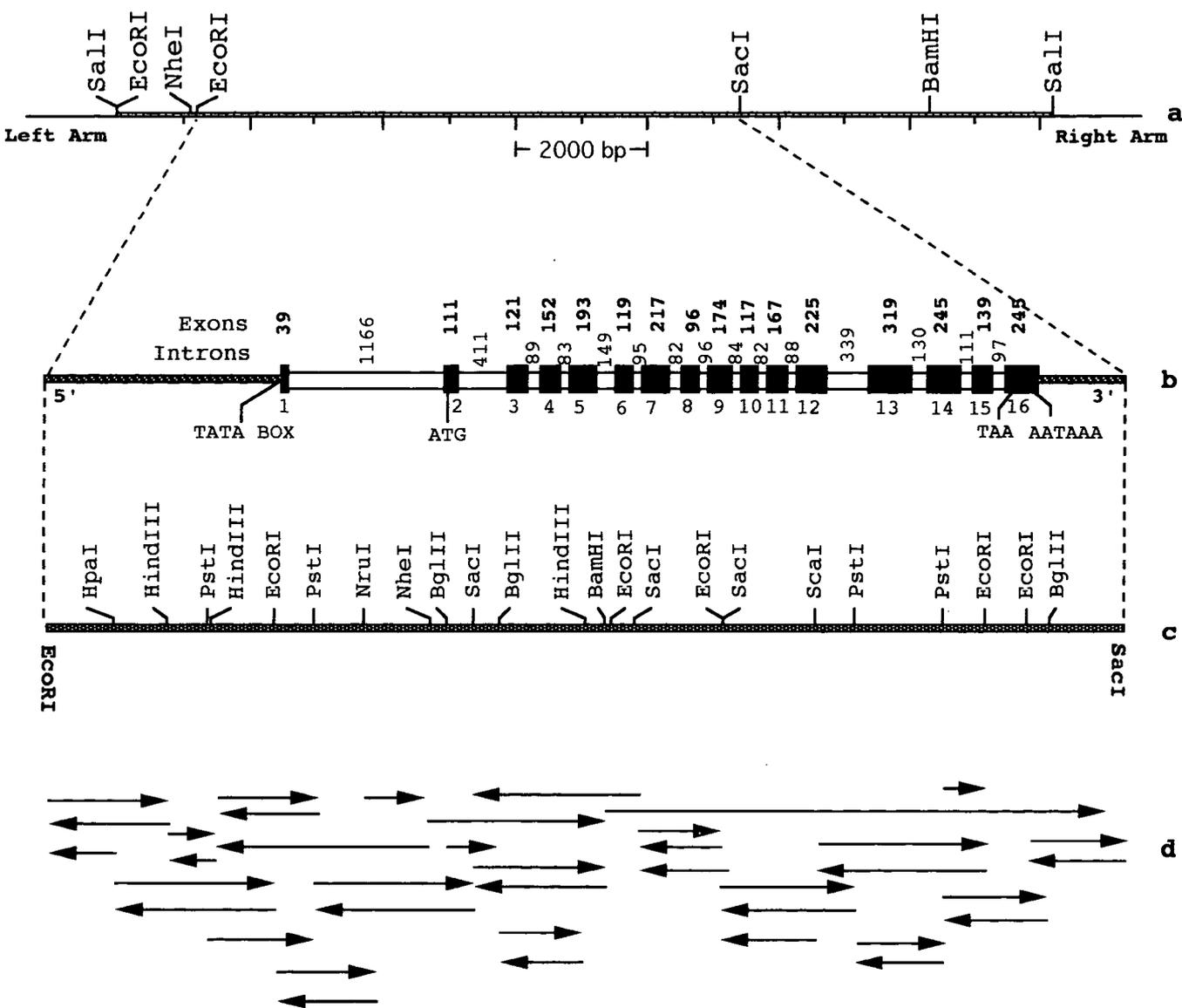
Figure 3.5. Restriction enzyme maps of the rice sucrose synthase genomic clone RSS2.4 (a and c), strategy used for sequencing (d), and structure of the *RSs1* gene (b). The exons are shown by solid boxes and the introns by empty ones. The size of the exons and introns is given above the drawing. The arrows in (d) indicate the sequencing direction.

by computer analysis of the partial sequences obtained earlier. The 8.2 kb region between the EcoRI and SacI sites shown in Fig. 3.5a was digested with various restriction endonucleases in order to generate small or overlapping fragments, which were then subcloned into pUC or M13 vectors and sequenced (Fig. 5d). Using this sequencing strategy, 85% of the nucleotide sequences were determined from both DNA strands, whereas the remaining 15% were sequenced more than once from the same DNA strand. The full sequence is shown in Fig. 3.6.

3.1.5. Structure of *RSs1* transcription unit

Alignment of the *RSs1* and maize *Sh1* sequences (Werr *et al.* 1985) shows that the *RSs1* gene has the same number of exons (Fig. 3.5b) as the maize *Sh1* gene, adding up to an open reading frame (ORF) of 2424 bp or 808 amino acids which predicts a molecular weight for the protein of 92.1 kD. All the exons, except for exons 1, 2 and 16, are identical in size to their counterparts in the maize *Sh1* gene. Exon 1 of *RSs1* is 39 bp long, 13 bp smaller in size than that of *Sh1*, but the 12 bp sequence (-3 GAGAAACCCTCC +9) around its 5' region and the 9 bp sequence (+32 ACCATTGGG +40) in the 3' region are identical to the sequences of the same regions (i.e. -3 to +9 and +44 to +52) in *Sh1*, assuring the identification of the transcription start site. The first ATG inside the exons is found in exon 2, 55 bp downstream of the transcription start. The sequence surrounding this translation initiation codon (5'-AGTCATGGC) are well conserved between *RSs1* and *Sh1*, and matches six of nine nucleotides from the consensus sequence for plant genes (5'-AACAATGGC-3') (Lutcke *et al.*, 1987).

A potential polyadenylation signal sequence (AATAAA) is located 159 bp downstream of the translation stop codon TAA. The position of the putative polyadenylation site shown in Fig. 3.6 is established from the sequence identity around the region to the *Sh1* gene, which gives rise to a 245 bp-long exon 16 with a 200 bp non-translated sequence.

```
TGAATGACAGAATTCAGAGCCTTCGTGGGCTCCAATCATCCCTTAGAAAGGCAGAAGAATATCTGATGGGCATTCCTCAAGACACGCCCTACTCGGAGTTCAACCACAGGTGATTACTTT +2644
 L  N  D  R  I  Q  S  L  R  G  L  Q  S  S  L  R  K  A  E  E  Y  L  M  G  I  P  Q  D  T  P  Y  S  E  F  N  H  R<---------

ATTATTTCTCCTTATCATATAGCCCATGTTATACTACTGCAAATGGAATAATTCCTTACAATGTGATTCATGATTTGAATACAGGTTCCAAGAGCTCGGTTTGGAGAAGGGTTGGGGTGA +2764
--------------------------------------------------IVS-6---------------------------------------------------> F  Q  E  L  G  L  E  K  G  W  G  D

CTGTGCAAAGCGTGTGCTTGACACCATCCACTTGCTTCTTGACCTTCTTGAGGCCCCTGATCCGGCCAACTTGGAGAAGTTCCTTGGAACTATTCCAATGATGTTCAATGTTGTTATCCT +2884
 C  A  K  R  V  L  D  T  I  H  L  L  L  D  L  L  E  A  P  D  P  A  N  L  E  K  F  L  G  T  I  P  M  M  F  N  V  V  I  L

GTCTCCGCATGGATACTTTGCCCAATCCAATGTGTTGGGATACCCTGATACTGGTGGTCAGGTAGAGCTTTTTAACAAAATTTTCCCCTTATATACAGTACTTCAAATATTTCTAGTTTTG +3004
 S  P  H  G  Y  F  A  Q  S  N  V  L  G  Y  P  D  T  G  G  Q <-----------------------------------------IVS-7----------------

ACCTGTTATTCTGTTGTTTGCAGGTTGTGTACATTTTGGACCAAGTCCGCGCTTTGGAGAATGAGATGCTTTTGAGGATCAAGCAGCAAGGCCTTGATATCACACCTAAGATCCTCATTG +3124
---------------------->V  V  Y  I  L  D  Q  V  R  A  L  E  N  E  M  L  L  R  I  K  Q  Q  G  L  D  I  T  P  K  I  L  I<

TATGTTCAAGATGTTGGTCTTGAATGTTTCAAGTTTTAAACACTTCTCAGTGCGAATTGTAGAGTAACGTGCGGACATGATTTTACTTGATGCAGGTAACCAGGCTGTTGCCTGATGCTG +3244
--------------------------------------IVS-8-------------------------------------------------------->V  T  R  L  L  P  D  A

TTGGTACTACATGCGGCCAGCGTGTGGAGAAGGTTATTGGAACTGAGCACACTGACATTCTTCGTGTTCCATTCAGGAGTGAGAATGGTATCCTCCGCAAGTGGATCTCCCGTTTTGATG +3364
 V  G  T  T  C  G  Q  R  V  E  K  V  I  G  T  E  H  T  D  I  L  R  V  P  F  R  S  E  N  G  I  L  R  K  W  I  S  R  F  D

TCTGGCCATTCCTGGAAACATACACTGAGGTAAATGGTTTATCTGACAACGTCCTTCAGAGCTGAGCACAAACCATGAATTCTGAGAGCTCATCTTTTTTTTTTTTGGTTGCAGGATGTTG +3484
 V  W  P  F  L  E  T  Y  T  E <----------------------------------------IVS-9----------------------------------------> D  V

CAAACGAAATTATGAGGGAAATGCAAGCCAAACCTGATCTCATCATTGGCAATTACAGTGATGGAAACCTTGTTGCCACTCTGCTGGCTCACAAATTAGGAGTTACCCAGGTATGTTTGG +3604
 A  N  E  I  M  R  E  M  Q  A  K  P  D  L  I  I  G  N  Y  S  D  G  N  L  V  A  T  L  L  A  H  K  L  G  V  T  Q <---------

TTGTAAACCACTCCTTTAAATATTTCTTTCATTAAGTCAATTACTTATGATTTCTCTTCTACATGGATACAGTGTACCATTGCTCATGCCTTGGAGAAAACCAAATACCCCAACTCAGAC +3724
-------------------------------IVS-10------------------------------> C  T  I  A  H  A  L  E  K  T  K  Y  P  N  S  D

ATATACTTGGACAAGTTTGACAGCCAGTACCACTTCTCATGCCAATTCACTGCTGATCTTATCGCCATGAATCACACTGATTTCATCATCACCAGTACATTCCAAGAAATTGCTGGAAGG +3844
 I  Y  L  D  K  F  D  S  Q  Y  H  F  S  C  Q  F  T  A  D  L  I  A  M  N  H  T  D  F  I  I  T  S  T  F  Q  E  I  A  G  S<

TAAAAATTTTCTAAAGAGCAGTAGAATATTGCTGCGTTGTGCCAACACCTAAACGGGTATCTGATGTTTGTTTCCTCTGTTTCTTCAGCAAGGACACTGTGGGGCAGTATGAATCACACAT +3964
-------------------------------------IVS-11--------------------------------------------------> K  D  T  V  G  Q  Y  E  S  H  I

TGCATTCACCCTTCCTGGGCTTTACCGAGTTGTGCATGGCATAGATGTTTTTGATCCCAAGTTCAACATTGTCTCTCCTGGAGCTGACATGAGTGTCTACTTCCCGTACACCGAGGCTGA +4084
 A  F  T  L  P  G  L  Y  R  V  V  H  G  I  D  V  F  D  P  K  F  N  I  V  S  P  G  A  D  M  S  V  Y  F  P  Y  T  E  A  D

CAAGAGGCTCACTGCTTTCCACCCTGAAATTGAGGAGCTTCTCTACAGTGAAGTCGAGAACGATGAACACAAGTGAGTACTGAACTGCTGGTTCCGTCCCTGTTATTCACCTACCATGTA +4204
 K  R  L  T  A  F  H  P  E  I  E  E  L  L  Y  S  E  V  E  N  D  E  H  K<-------------------------------------------------

ACAAATAATTAAACACTGATTTGTGTGTTTGGTCTAGCTTTGGCTCTTCAGTCTACTAAGTGCCATGCACTAGAACTTTGTAGATTTGAACTGCAATTCTTAATTTTCTCACTCCTTAGT +4324
-----------------------------------------------------------------------------------------IVS-12-----------

AGACCACAATAATAATTATTGTGTCTGACCCAGCTAGTTTACAATAGTGACTTCAATGCTTGTTTTTTCCCCTTGAACAATTGTTTACTTTCTAGCAGTGGAGTTTGGCATCTAAACTTG +4444
---------------------------------------------------------------------------------------------------------------------

AATACACGCTGTTCTTTTTTTGTGTATAAAAATGTTTATTTAATGTTCTGCAGGTTTGTATTGAAGGACAAGAACAAGCCAATCATCTTCTCCATGGCTCGTCTTGACCGAGTGAAGAACAT +4564
-----------------------------------------------------> F  V  L  K  D  K  N  K  P  I  I  F  S  M  A  R  L  D  R  V  K  N  M

GACAGGTCTGGTTGAGATGTATGGTAAGAATGCACATCTCAGGGATTTGGCAAACCTTGTGATTGTTTGTGGTGACCACGGCAATCAGTCCAAGGACAGGGAGGAGCAGGCTGAGTTCAA +4684
 T  G  L  V  E  M  Y  G  K  N  A  H  L  R  D  L  A  N  L  V  I  V  C  G  D  H  G  N  Q  S  K  D  R  E  E  Q  A  E  F  K

GAAGATGTACGGTCTCATTGACCAGTACAAGTTGAAGGGGCATATCCGCTGGATCTCAGCTCAGATGAACCGTGTTCGTAACGGGGAGTTGTACCGATACATTTGTGACACCAAGGGAGT +4804
 K  M  Y  G  L  I  D  Q  Y  K  L  K  G  H  I  R  W  I  S  A  Q  M  N  R  V  R  N  G  E  L  Y  R  Y  I  C  D  T  K  G  V

CTTTGTCCAGGTATATACATACGCTGTCATATTCTACTGCTATATTTCTATTCTGAGCTTCCAGTAGCAAATGATTTCTACTATTGTAAAGAACAGCCTTGTGCTTGAGACTAAAACTG +4924
 F  V  Q <-----------------------------------------------------------------IVS-13----------------------------------------

CTTTTAAAATTGTTTTTTTAGCCTGCATTCTATGAAGCGTTTGGTCTGACTGTCATCGAAGCCATGACATGTGGTTTGCCAACAATCGCAACATGCCATGGTGGCCCTGCTGAGATTATTG +5044
------------------->P  A  F  Y  E  A  F  G  L  T  V  I  E  A  M  T  C  G  L  P  T  I  A  T  C  H  G  G  P  A  E  I  I

TTGATGGGGTGTCTGGTCTGCACATTGATCCTTACCACAGTGACAAGGCTGCTGATATCTTGGTCAACTTCTTTGAGAAGTGCAAGCAGGATTCAACCTACTGGGACAATATTTCACAGG +5164
 V  D  G  V  S  G  L  H  I  D  P  Y  H  S  D  K  A  A  D  I  L  V  N  F  F  E  K  C  K  Q  D  S  T  Y  W  D  N  I  S  Q

GAGGTCTGCAGAGGATTTACAGAAGTATGTATTTCATGATCTGCAAATATATAACATTTTTTTTTTTACGATTGACATATAACTGTAGAAATCATGTGTCCAATCTATATTGATCCTACA +5284
 G  G  L  Q  R  I  Y  E  K<----------------------------------------------IVS-14------------------------------------------

AATTTGGACTTGTCAGGTACACCTGGAAGCGTGTACTCTGAGAGGCTGATGACCTTGACTGGTGTATACGGATTCTGGAAGTACGTAAGCAACCTTGAGAGGCGCGAGACTCGCCGTTACA +5404
---------------->Y  T  W  K  L  Y  S  E  R  L  M  T  L  T  G  V  Y  G  F  W  K  Y  V  S  N  L  E  R  R  E  T  R  R  Y

TTGAGATGTTCTATGCTCTGAAATACCGCAGCCTGGTAAGTTACTCTCGCTGTTCAACAGCTTGTTAATTTCAGTATTATTGCACTGAGCATATCTCTGAATTCTGTTGCTACATTCTT +5524
 I  E  M  F  Y  A  L  K  Y  R  S  L <--------------------------------------IVS-15--------------------------

TGTTGAATCCAGGCCAGCGCCGTCCCATTGGCTGTCGATGGAGAGAGCACATCCAAGTAATGGAGGGGAAAATATGCATCTTCAGCAGGAGAAGCCGTCAGCTGCATTCGAATTTGATAA +5644
-----------> A  S  A  V  P  L  A  V  D  G  E  S  T  S  K

TTTTCTGTAGTTGTCATTTGGCATCCATGTTTGCCATGGATGTGTACTATGTCTAAGGTTTCAGTACTTTTGCGAGATTTTGGGCAGTGCTTGCCTCAAATAAACGCCGGTTCCTGGTGT +5764
                                                                             PolyA+

TTATCGTTCAGAATTCAATGCAATGTTTTATTGCCGCGCTTCCTCAATCTGTCATGTTGAATGACCACTCCCTAGTGCCCTACTGTTAAAAAAGGTTGGTTCAAAGCCATGGATGAACCC +5884
              ^

AACATTTTTGTTACCTTGTTCTTTTCAGAATTTCCTTAAGAGTTGTGATAGTAGGCCCAGTGCGACCAAAATGTTGGGTTCCTTGAGTTTTCAGTATAAAAATTTCTGATATTAAATGCAG +6004

ATCTTGGCTCAGTACAGAGTCTTGCAAACCATGTGGTTGCATCTTTTCTACTGCTGTAGTAGATGAAACTGAAGATACTACTTATTACTCTTTAGACAAAGTTTCATCATATTCCCAAAA +6124

TCTTGGTCTAGGATTGGGAGACATCAAGAAGATATAAACTTCACACTCAGGCCCTGTTTGGCAACCCTGAATAATTTTCTTTTGTGTAATGCAAATGCACCTGTAAATTTGGGTGACCCA +6244

AAGATGCCAGTGAAATATAAATCCCATGTAGCCAAACTGTTTGTTAGAAGAACTAGAACACTCAGTAGAGGTTTTAGCTAATTACAGTTCATTGCTGATCAAGGTTGCATCAGACATAGA +6364

AGTCAGACCGAAACATCAATTTGGAGCTATAATACAAAATTCGATAAACCAATACGAAATTCAAATAAAATAACATTGTTCATCAAAAATTTCTGGCCAGAACCATCCAAGAGCAGTTGTA +6484

AGCGAGA
```

The *RSs1* gene has a large intron 1, and 14 other introns with a size range of 82 to 411 bp (Fig. 3.5b), similar to the maize *Sh1* gene, with the exception of intron 12, which is of 339 bp in length, much longer than intron 12 (134 bp) of the *Sh1* gene. The borders between exons and introns (Table 3.1) are in full agreement with the GT-AG rule (Breathnach and Chambon, 1981). The 9 bp sequence at the donor site and the 16 bp sequence at the acceptor site also agree well with the consensus sequences of plant intron splice junctions, (C/A)AG:GTAAGT and TTT(T/Pu)TT(T/Pu)(T/Pu)(T/Pu)(T/Pu)TGCAG:G (Brown, 1986), respectively.

Table 3.1. Exon/intron boundaries of *RSs1*

5' Splice junction	Introns	3' Splice junctions
T G G: G T A T G T	1	A A C C T T T C T T T G C A G: T
T A G: G T G A A G	2	T A T C A C T C T A T G C A G: G
C A G: G T A A A T	3	C T T G A C T G T A T G T A G: G
A C A: G T A A G T	4	G T T T T T C T T A A A C A G: C
A C G: G T A A G C	5	A A T T T T C T A T T A C A G: A
C A G: G T G A T T	6	A T G A T T T G A A T A C A G: G
C A G: G T A G A G	7	T T C T G T T G T T T G C A G: G
A T T: G T A T G T	8	T T T T A C T T G A T G C A G: G
G A G: G T A A A T	9	T T T T T T T G G T T G C A G: G
C A G: G T A T G T	10	T C T A C A T G G A T A C A G: T
A A G: G T A A A A	11	C C T C T G T T T C T T C A G: C
C A A: G T G A G T	12	T T T A A T G T T C T G C A G: G
C A G: G T A T A T	13	A A A A T T G T T T T T A G: C
G A A: G T A T G T	14	A T T T G G A C T T G T C A G: G
C T G: G T A A G T	15	C T T T G T T G A A T C C A G: G
G 2 1 11 15 0 3 1 8 2		1 0 1 1 3 2 2 6 3 0 1 7 0 0 15 9
A 4 10 3 0 0 12 9 6 1		5 4 1 4 4 1 1 0 3 7 1 4 0 15 0 1
C 7 2 0 0 0 0 0 0 1		3 2 2 3 1 3 2 3 0 2 0 1 13 0 0 3
T 2 2 1 0 15 0 5 1 11		6 9 11 7 7 9 10 6 9 6 13 3 2 0 0 2

3.1.6. The 5'-flanking region

A 1756 bp region upstream of the transcription start site was sequenced.
Putative TATA and CAAT boxes are located at positions -85 and -28 respectively
(Fig. 3.6), a common distance in eukaryotic genes. The TATA box is flanked by GC-
rich sequences, typical of other promoters, and the 21-nt sequence around it is nearly
identical to the TATA box region in the maize *Sh1* gene (Table 3.2). Five directly
repeated sequences (RP) are present at several positions, including the 48-bp long
repeats at -1032 and -979 (RP2, Fig. 3.6). The 5' upstream region also contains four
potential hairpin loop-forming sequences (HL). These palindromes, especially the
three at positions -429 (HL2), -350 (HL3) and -189 (HL4), share a similar 10-nt
sequence, which is analogous to the 10-bp direct palindrome sequences found in maize
zein gene -300 element and almost identical to the sequences present in two 13-bp
direct repeats found in the 5'-flanking region of the gene encoding a rice endosperm-
specific ADPglucose pyrophosphorylase subunit (Anderson *et al.* 1991) (Fig. 3.7). A
17-bp sequence of *RSs1* is similar to a sequence near the TATA box of the CaMV 35S
promoter (Odell *et al.* 1988):

```
RSs1        -104    GAAAGGCCCTACCACTA  -88
CaMV 35S    -46     GCAAGACCCTTCCTCTA  -30.
```

```
HL2        -419    ATATATATTT    -410
HL3        -340    -C--------    -331
HL4        -186    -C----G---    -177
M-zein     -331    -C-C--G-G-    -322
M-zein     -352    -C-C----G-    -343
R-AGPP     -628    -------A--    -619
R-AGPP     -531    ----------    -522
```

Figure 3.7. Homologies of the 10-nt sequences inside the
palindromes of *RSs1* 5'-flanking region with those found in the
maize zein (M-zein) gene -300 region and in the two direct
repeats of the 5'-flanking region of a rice endosperm-specific
ADPglucose pyrophosphorylase gene (R-AGPP). The zein
sequence between -331 to -322 is part of a consensus sequence of
all zein genes and of a 22-bp nuclear factor-binding sequence
(Maier *et al.* 1987).

Another two potential regulatory sequences are found at positions -1011 and +9 (inside exon1), which are homologous to one of the two core sequences essential for anaerobic induction of maize alcohol dehydrogenase 1 gene (*Adh1*) (Walker *et al.* 1987). A similar sequence is also present in the maize *Sh1* gene (Fig. 3.8).

```
Adh1   -114   C C G T G G T T T G C T T G C C   -99
Adh2   -119   - - - A - C C - - T - - - C - -   -104
Sh1    -837   - - - A - - - - - - G - - - T G   -822
RSs1  -1011   - - A - - - - - - T - - G C A A   -996
Sh1     +20   - A T - - - A C - - - - - - - T   +36
RSs1     +9   - A T - C - - - - - - - C - T -   +24
```

Figure 3.8. Homologies of sequences in the 5'-flanking regions of *Sh1* and *RSs1* with one of the 2 core sequences essential for anaerobic induction in maize alcohol dehydrogenase gene (Adh) promoters (Walker *et al.* 1987). Clustered point mutation experiments show that this core sequence is more important than the other. Mutation of this core sequence abolishes all expression, whereas mutation of the other reduces anaerobic expression to the level observed under aerobic conditions. Positions of the point mutations which abolish whole expression are underlined.

A comparison of the *RSs1* and the *Sh1* promoters was made using a computer program (DNA Strider), and the major homologous sequences found are summarised in Table 3.2. Sequences homologous to them reported from several other phloem-specific promoters are also included, i.e. the promoter region of the *Agrobacterium rhizogenes* Ri plasmid ORF12 gene (*rolC*) (Slightom 1986; Sugaya *et al.* 1989), the major promoter region for the transcription of the genome of rice tungro bacilliform virus (RTBV) (Qu *et al.* 1991; Bhattacharyya-Pakrasi *et al.* 1993), and the 2' promoter in the TR-DNA of *Agrobacterium tumefaciens* (TR2') (Velten *et al.* 1984; Saito *et al.* 1991).

Table 3.2. Major homologous sequences between the *RSs1* and maize *Sh1* (MSh1) 5'-flanking regions. DNA sequences analogous to these sequences from several reported phloem-specific promoters, *rolC* (the promoter region of the *Agrobacterium rhizogenes* Ri plasmid ORF12 gene), RTBV (the major promoter region for the transcription of the genome of rice tungro bacilliform virus), and TR2' (the 2' promoter in the TR-DNA of *Agrobacterium tumefaciens*), are also included. The TATA boxes in *RSs1* and *Sh1* are underlined.

```
RSs1      -31     CAACTATTTATTGGTCCTTCT    -12
MSh1      -33     CGTCTATTTATTGGTCCCTCT    -13

RSs1     -482     GCCGGAATATGCTAATAGGAT    -462
MSh1     -235     GCCAGCATATGCTAATTGGAT    -215
rolC      -61     GCTCCAATACGCTAAGAG       -44

RSs1     -145     AATAATTTGGGGTTATCCATTGC  -123
MSh1     -152        TAATTTGGCATTATCCATTGC  -132
RTBV     -342     AAGAATTTGGGTTT           -328

RSs1     -874     TTTTCTGAAAGACCTATGTTTC   -853
MSh1     -730     TTTTCTGAGAGACCCATGTTTC   -711

RSs1    -1266     TTTGTTCTTGTTTGGAAC      -1249
MSh1    -1035     TCTGTTCTTGTTTGGAGC      -1018

RSs1     -971     AGAGCATAAAAGAATTAC       -954
MSh1     -449     AGAGCTCAAAAGAATTAC       -432

RSs1    -1322     CTTTTTCTTTTTTACC        -1307
MSh1     -420     CTTTTTCTTTTCTTCC         -405

RSs1    -1246     ACGGAATGCCCGATGC        -1230
MSh1    -1002        CGGCACGCCCGATGC       -988
TR2'    -83~-73   ACGGAATGCGCG           -72~-62
```

3.1.7. Sequence comparison with other plant species

The predicted amino acid sequence of *RSs1* was compared with nine other published sequences of sucrose synthase enzymes, and the percentages of identical residues for all binary comparisons are summarised in Table 3.3. These data show that the RSs1 sequence is more closely related to the Ss1 sequences from other species (>93%) than to the Ss2 sequence from the same species (80.7%) and *vice versa*, in a similar manner as previously described for other Ss1 and Ss2 enzymes (Sanchez de la Hoz *et al*. 1992; de Ilarduya *et al*. 1993). Sucrose synthase sequences from dicot species are about equidistant to the monocot Ss1 (66-75%) and Ss2 (64-74%).

Table 3.3. Homologies of the predicted amino acid sequences among sucrose synthases from different plant species. MSh1= maize Sh1 (Werr *et al.* 1985), BSs1= barley Ss1 (Sanchez de la Hoz *et al.* 1992), BSs2= barley Ss2 (de Ilarduya *et al.* 1993), MbSs= the only type of Ss described so far for mung bean (Arai *et al.* 1992), ASus and ASsA= two different Ss in *Arabidopsis* (Martin *et al.* 1993; Chopra *et al.* 1992), PtSs= the only type of Ss reported so far for potato (Salanoubat and Belliard 1987), VfSs= the only type of Ss described so far for *Vicia faba* (Heim *et al.* 1993).

	MSh1	BSs1	RSs2	BSs2	MbSs	ASus	ASsA	PtSs	VfSs
RSs1	94.3	93.3	80.7	79.6	74.8	71.6	66.8	73.8	73.9
MSh1	----	93.1	79.1	78.4	76.6	73.4	66.4	75.1	75.1
BSs1	----	----	78.9	78.4	74.7	71.5	66.0	73.3	73.7
RSs2	----	----	----	88.8	79.0	70.5	65.7	73.8	74.4
BSs2	----	----	----	----	74.0	70.3	63.7	72.7	72.8
MbSs	----	----	----	----	----	81.1	65.8	81.5	94.3
ASus	----	----	----	----	----	----	64.6	78.4	81.9
ASsA	----	----	----	----	----	----	----	65.9	64.9
PtSs	----	----	----	----	----	----	----	----	80.8

Alignment of the predicted amino acid sequences reveals that differences between the proteins are not equally distributed. Regions of low homology are mainly located in the N- and C-termini of the proteins (Fig. 3.9).

3.1.8. Base composition

Similarly to other eukaryotic genes, the introns of *RSs1* are more AT-rich (76.3-58.4%) than the exons (57.5-45.5%) but both the introns and the exons in the upstream region of the gene are more GC-rich than the downstream ones. Unlike maize *Sh1* gene whose exons 13, 14 and 15 are significantly more GC-rich (Werr *et al.* 1985), no large deviation in G+C content is observed for *RSs1* exons.

Figure 3.9. Comparison of the predicted amino acid sequences of RSs1 and sucrose synthases from different species. Identical amino acids are given as dashes and deletions indicated by asterisks.

```
RSs1 M**AA***KLARLHSLRERLGATFSSHPNELIALFSRYVNQGKGMLQRHQLLAEFDALI*EADKEK 60
MSh1 -**~-**~--T-----------------------------H-----------------F*DS----  .
BSs1 -**--**~--T-----------------------------H-----------------F*-S----
RSs2 -GE--GDRV-S----V------DSL-A----V-V-T-L--L------A--I-YNNA-S---R--
BSs2 -GET-GERA-S-V--V----I-HSL-A-T---V-V-L--L------P---T--YN-A-P--ER-
MbSs -***-TDR*-T-V-------DE-L-ANR--IL--L--IEGK---I--H--VI---EE*-P-ESRQ-
ASus -**-NAERMITRV--Q----NE-LV-ER--VL--L--VEAK---I--QN-II---E--*P-QTQK-
ASsA -***-S**FFDLVYHK--IGFLNK-L-FWVFWVC-V*---AH---I--S---ID--LKTV*KV-GTL
PtSs -**-**ERV-T-V------VD--LAA-R--ILLFL--IESH---I-KP-EI------IR*QD--N-
VfSs -***-TER*-T-V------DE-LTANR--IL--L--IEAK---I--H--V----EEI*P-ENRQ-

RSs1 **YAPFEDILRAAQEAIVLPPWVALAIRPRPGVWDYIRVNVSELAVEELSVSEYLAFKEQLV 120
MSh1 **-------------------------------------------------------------
BSs1 **-------------------------T--------------------T-------------
RSs2 LKDGA---V--S---G--IS--------------E-V----------L-T-P--Q------
BSs2 LKNT----L--G------I--------------E-V------LR--Q-----A
MbSs LTDGA-GEV--ST-----------V------E-L---HA-V--V-QPA----R---E--
ASus LEGG--F-L-KST-------------V------E-L---LHA-L----QPA-F-H---E--
ASsA *EDLNKSPFMKVL--------F----------RE-V----Y--S-DH-T-----R---E--
PtSs LNEHA--EL-KST--------------L-----E-----NA-V-----P---Q--E--
VfSs LTDGA-GEV--ST-----------V------E-----VHA-V--N-QPA-F-K---E--

RSs1 DGHTNSNFVLELDFEPFNASFPRPSMSKSIGNGVQFLNRHLSSKLFQDKESLYPLLNFLK 180
MSh1 --QS-----------------------------------------------------------
BSs1 -E-ASRK-------------------Y-K-------------------------------
RSs2 EEG--N---------------L----------------H----M------R
BSs2 N-S-DN--------G----------L----------------H----M--------R
MbSs --SS-------T------TLN-----------A--H-----H---E--R
ASus --VK-G--T----------I---TLH-Y-----D------A---H-----L--K--R
ASsA N--A-GDYL---H-------TL--TR-S-------LV------IM-RN---ME---E--R
PtSs --AS-G------------T----K-TLT------E------A-M-H-----MT---E--R
VfSs --SA-G------------TLN--------A--HD---H----E--R

RSs1 AHNHKGTTMMLNDRIQSLRGLQSSLRKAEEYLMGIPQDTPYSEFNHRFQELGLEKGWGDC 240
MSh1 ---Y----------------------LSV---------------------------T
BSs1 ---Y-----I-------------A--------VS--E---S----------------T
RSs2 ---Y-M--------R--SA--GA------H-S-LSA--------H-----------
BSs2 ---Y-M--------R--GT--GA------TH-S-L-A----T--H----------
MbSs L-SV--K-L-------NPDA--HV--------GTV-PE----A-E-K---I---R-----N
ASus L-S-Q-KNL--SEK--N-NT--HT-------AELKPE-L-E--EAK-E-I---R--N
ASsA T-K-D-RP-------NIPI--GA-AR---F-SKL-LA------EFEL-GM-F-R----T
PtSs --HY--K-------NSNT--NV--------IML-PE---F--E-K---------T
VfSs L-SY--K-L-------NPDS--HV--------STVDPE------E-------R----S

RSs1 AKRVLDTIHLLLDLLEAPDPANLEKFLGTIPMMFNVVILSPHGYFAQSNVLGYPDTGGQV 300
MSh1 -------L-----------------------------------------------------
BSs1 ---H------------------A-------------------------------------
RSs2 ---SQE-------ST----------V-----M--M----A-------------
BSs2 -Q-ASE--------------SS-------VL--------------A------------
MbSs -E---ES-Q-------CT--T---R---V---------------D---------------
ASus -E---M-R--------CT--T--RV--V---------------D---------------
ASsA -QK-SEMV-----I-Q----SV--T------RY-----A---L----A--
PtSs -E---EMVCM--------SCT------R---V--------E------------
VfSs -E---ES-Q------CT--T--DR---V------------DD----------

RSs1 VYILDQVRALENEMLLRIKQQGLDITPKILIVTRLLPDAVGTTCGQRVEKVIGTEHTDIL 360
MSh1 -----------------------------------------------L-----------
BSs1 -----------------------------------------------L-----------
RSs2 ---------M--------------N---R--------T-------L-------H--
BSs2 ---------M-----------------M---H--------L------
MbSs ------------H--------V-R---I------------------L--------SH--
ASus ----------I---Q------N-K-R---L---------------E-L-R-YDS--YC---
ASsA ----------QK--EVI----------E-K----N----R-S-T--A--
PtSs -----------R---K------I-R------------I---Y-A--SH--
VfSs -----------S--N---K-----V-R---I---------L---Y----CH--

RSs1 RVPFRSENGILRKWISRFDVWPFLETYTEDVANEIMREMQAKPDLIIGNYSDGNLVATLL 420
MSh1 -----N---------------Y--------SS---K-----------------
BSs1 ----T---*-------------Y---------------T----------
RSs2 ----T----V------E----Y--F-D---H--AG-L--N--------------C--
BSs2 ----KT-D--V------E----Y--A-D---H-LAK-L-G-----V---------C--
MbSs -----T-K--V------E----Y-------A-V-LSK-LDG-------------S--
ASus -I---T-K------------Y--FA-AS--L-G-V-N-------------S-
ASsA -----T-K--V------E----Y--FI----K--SA-L-------E---A-S-
PtSs -----DQK--V-------E----Y--F-----K---SA-L-----V-------I--S--
```

```
RSs1 AHKLGVTQ*CTIAHALEKTKYPNSDIYLDKFDSQYHFSCQFTADLIAMNHTDFIITSTFQE 480
MSh1 --------*-------------------~-----------------------------------
BSs1 --------*--------------------------------------------------------
RSs2 ---M--H*------------A-----WK--EDH--------T-------A----------
BSs2 -------H*------------L-WK--EDH-----------------A----------
MbSs --------*------------E----WK-LEER----------F-------------
ASus --------Q-------------WK--L-DKY--------IF---------------
ASsA -S------*-N-----------E----WRNHEDK----S-----------NA-------Y--
PtSs --------Q-------------D----WK---EK----S----------
VfSs --------*-----------E----WK--EEK-----------F------------

RSs1 IAGSKDTVGQYESHIAFTLPGLYRVVHGIDVFDPKFNIVSPGADMSVYFPYTEADKRLTA 540
MSh1 --------------------------------------------------Y----T-----
BSs1 ------S---------D----------------------T------T------
RSs2 ---N---------M--M----------------------I----S-SR----S
BSs2 ---N---------M--M------------------------QQ----S
MbSs -----------T-----------------QTI--H--TSR---S
ASus -----E--------T----------------------I-----EKR--K
ASsA ------NN-------T--M---------------TI---SDKER----
PtSs ---------M--M-------------N----------INL-S-S-TE----
VfSs ------------T--------------------QTI---TSR---S

RSs1 FHPEIEELLYSEVENDEHKFVLKDKNKPIIFSMARLDRVKNMTGLVEMYGKNAHLRDLAN 600
MSh1 --------------D---S--------K-------------------R--E----
BSs1 --S--------D--------------K------------------K----
RSs2 --------------D-N----M--R-----------------L-----L--R-PR-QE-V-
BSs2 L-T------F-D---A---------K----------------M---------R-PR-QE-V-
MbSs --T--------S---E--IC----RS---T----------I---W----K-E-V-
ASus --S--------D--K--LC------K--L-T-----------LS-------TR---V-
ASsA L-ES-----F-AEQ-D--VGL-S-QS----------------L----C-A--SK--E---
PtSs -----D------DV-----CC----RT---L-T---------I-------A--PR--G-V-
VfSs --Y--------T---E--IC----RS---T----------I----W----K-E-V-

RSs1 LVIVCGD*HGNQSKDREEQAEFKKMYGLIDQYKLKGHIRWISAQMNRVRNGELYRYICDTK 660
MSh1 ----A--*--KE-----------S--E--------------
BSs1 ----A--*--KE-----------R--S--EE---------
RSs2 --V----*---P---K-------FD---E--N-----------------
BSs2 --V----*--KV---K---V------FD--EK-N-S----------------M-
MbSs --V-A--*RRKE---L--K--M----S--ET---N-QF----S-----------V-A---
ASus ----G--*RRKE---N-K-----D--D-EE---N-QF----S--D----------
ASsA ----G-YIDE---R-----M--IQ--HS--E--D-H-EF---A-----------A---
PtSs --V-G--*RRKE---L-----M----E-ETHN-N-QF----S----------A--
VfSs --V-A--*RRKE---L--K--M--E--ET---N-QF--S----------V----

RSs1 GVFVQPAFYEAFGLTVIEAMTCGLPTIATCHGGPAEIIVDGVSGLHIDPYHSDKAADILV 720
MSh1 ----------------S--------------------------------------------
BSs1 ------------------------------------------------------------
RSs2 -A--------------V-S-------F-AY--------N---F-----QG--SAL--
BSs2 -A-------------------F--AY-------N---Y---QN--SAL--
MbSs -A-----V--------V---------F--N-----H-K--F------G-R---L--
ASus -A-----L---------V--------F---K------H-K--F-----G-Q---T-A
ASsA ---------------V-S--A--F----------EN----F------P-QV-GS-A
PtSs -A--------------V--------F--N-----H-K--F------GEQ---L-A
VfSs -A----V--------V---AT----F--LN------H-K--F-----G-R---L--

RSs1 NFFEKCKQDSTYWDNISQGGLQRIYEKYTWKLYSERLMTLTGVYGFWKYVSNLERRETRR 780
MSh1 ---D---A-PS---E------------------------------------
BSs1 -----STA-PS---K-----------K----------------
RSs2 E-----QE-PSH-TK---------E---------------
BSs2 G-G---QE-PSH-NK--------E-----------S-----------D------
MbSs E----V-V-PSH--K---A----E----QI-Q--L---------H----D---S-
ASus D-T---E-PSH--E--K------EDE---QI-Q--L---------H----D-L-A--
ASsA *L--T-NTNPNH-VK--E---K----R----K---L--A---A---H--K-------
PtSs D-----K-PSH-ET--M---K--E----Q---S-L--AA--------H--K-D-L-A--
VfSs E----V-A-PSH--K--L-----E-----Q---Q--L---------H----D-L-S--
```

```
RSs1 AHKLGVTQ*CTIAHALEKTKYPNSDIYLDKFDSQYHFSCQFTADLIAMNHTDFIITSTFQE 480
                                                      (continued as above)

RSs1 YIEMFYALKYRSLASAVPLAVDGESTSK 808
MSh1 ------------Q---SF-        802
BSs1 ------------A-------SGN     810
RSs2 ----L------TM--T----E--PSN- 816
BSs2 -L--L------KM-AT----E--TSGE 816
MbSs -L---------K--ES----E       805
ASus -L---------P--*----Q-D      813
ASsA ~L---S--F-D--NSI---T-EN     804
PtSs ----------KM-E----AE        806
VfSs ----------K--ES-----EE      806
```

3.2. Discussion

Two classes of sucrose synthase genomic clones have been identified, which correspond to two non-allelic genes in rice, *RSs1* and *RSs2*, respectively. However, in rice three different sucrose synthase cDNAs have been reported (Wang *et al.* 1992) and the Southern blot in Fig. 3.1, which detected more than three DNA fragments, also suggests the presence of three or more sucrose synthase genes. The failure to identify a type 3 sucrose synthase gene (*RSs3*) from the eight genomic clones may be coincidental, or may be due to the much higher homology between *RSs3* and *RSs2* (90% in deduced amino acid sequence) than between *RSs3* and *RSs1* (76%) (Wang *et al.* 1992) that may lead to homogeneity of the restriction endonuclease sites (i.e. SalI and EcoRI in Fig. 3.2) in *RSs2* and *RSs3* genomic DNA.

A DNA fragment of the genomic clone RSS2.4, encoding RSs1, was fully sequenced, and the gene organisation established by alignment with the maize *Sh1* gene. Isolation of *RSs1*, rather than of *RSs2* or *RSs3*, was the target of this project, since the promoter of its counterpart in maize, the *Sh1* gene, has been shown to direct phloem-specific expression in transgenic tobacco (Yang and Russell, 1990). The sequenced RSS2.4 fragment is 8248 bp in length and contains 1756 bp of 5'-flanking sequence and 710 bp of 3'-flanking sequence. The predicted amino acid sequence of *RSs1* is more homologous to those of other cereal *Ss1* genes than to that of *RSs2*, supporting the view that a duplication followed by divergent evolution of an ancestral gene took place prior to the evolutionary branching-out between rice and other monocot species (de Ilarduya *et al.* 1993). The equidistance of all the dicot sucrose synthase proteins compared with both Ss1 and Ss2 of cereals argues for the presence of a single SS gene, at least in some dicot species, as suggested by some workers (Chopra *et al.* 1992; Heim *et al.* 1993). The *RSs1* gene predicts six extra amino acids than maize *Sh1* gene at the C-terminus (Fig. 3.9). The same extra amino acid residues also occur at the C-termini of other reported monocot SS proteins, including the wheat

Ss2 type (de Ilarduya *et al.* 1993), implying that maize *Sh1* may have lost six amino acids-coding capacity during its evolution.

Comparison of *RSs1* with *Sh1* shows that the location of introns and exons is highly conserved. Both *RSs1* and *Sh1* have a long first intron, located upstream of the translation start. The first intron of the *Sh1* gene has been shown to be involved in the regulation of gene expression with enhancing effect (Vasil *et al.* 1989; Maas *et al.* 1990). Although exon1 is a non-translated sequence, it remains substantially conserved between *RSs1* and *Sh1*, suggesting that exon1 of *RSs1* may also be important in gene regulation as observed for that of *Sh1* (Maas *et al.* 1991).

Besides usual features of promoter regions, such as the TATA box, CAAT box and five direct repeats, several features of the 5'-flanking region of *RSs1* deserve attention. The three palindrome sequences at positions -429 (HL2), -350 (HL3), and -189 (HL4) contain a region homologous to the 10-bp palindrome sequence in maize zein gene promoter -300 region and almost identical to parts of the two direct repeats found in the 5' region of a rice endosperm-specific ADPglucose pyrophosphorylase gene (AGPP) (Anderson *et al.* 1991) (Fig. 3.7). The maize zein gene sequence, ACACATGTGT, is within a 22-nt specific binding-site for a nuclear factor from maize endosperm, which has been determined by filter binding, gel retardation, and DNase I footprinting assays (Maier *et al.* 1987). It is also part of the 15-bp sequence conserved in all zein genes (Brown *et al.* 1986), which overlaps the -300 box described for cereal storage proteins (Forde *et al.* 1985). Furthermore, a 43-bp region containing this 10-nt sequence has been demonstrated, using both *in vitro* and *in vivo* techniques, to have enhancing effect on CaMV 35S promoter-driven gene expression, and this enhancing effect is endosperm-specific and orientation- and position-dependent (Quayle and Feix 1992). The homology of the *RSs1* palindrome sequences to the zein gene and the AGPP sequences suggests that they may play a role in the regulation of *RSs1* expression, possibly an endosperm-specific type of expression. Although not solely expressed in the endosperm, as in the case of maize zein genes, maize *Sh1* is preferentially expressed in the endosperm of the immature kernel (Chourey 1981).

However, computer analysis found no significantly homologous sequences in the *Sh1* promoter to these palindrome regions of *RSs1*. Instead, it identified several highly homologous sequences in other regions (Table 3.2). The 21-bp sequence at position -143 of *RSs1* may be a potential regulatory element. It is located at a similar position to its homologous sequence in *Sh1* gene with respect to the transcription start site, and it contains the sequence TGGG, which is one of the five tandem repeats in *Sh1* promoter (Werr *et al.* 1985) and occurs at more than one position within a small region of *RSs1* promoter, i.e. -164, -155, -138, and -42. An analogous sequence is found in one of the reported phloem-specific promoters, RTBV (Table 3.2). Another sequence of *RSs1* at position -482 (Table 3.2) may also be important, since its homologous sequence in *Sh1* is within the nuclear protein-binding region (-235 to transcription start) of the *Sh1* gene that has been demonstrated by gel retardation and footprinting experiments (Springer *et al.* 1990). It is also homologous to a sequence at position -61 in the phloem-specific *rolC* promoter. The *rolC* sequence is within a region shown by deletion experiment to be necessary for the phloem-specific expression of *rolC* promoter in transgenic tobacco (Sugaya and Uchimiya 1992). Besides the RTBV and *rolC* promoters, the *RSs1* 5'-flanking region is also compared with another reported phloem-specific promoter, TR2', but no common sequence to all of them has been found.

Another two sequences at positions -1011 and +9 (inside exon 1) are also potentially important. They are highly homologous to one of the two core sequences in maize alcohol dehydrogenase 1 gene (*Adh1*) promoter (Fig. 3.8), demonstrated by base substitution experiments to be essential for anaerobic induction of *Adh1* and for background expression under aerobic conditions (Walker *et al.* 1987). The maize *Sh1* promoter also contains such sequences at similar positions. Since both *RSs1* and *Sh1* are subject to anaerobic induction (Taliercio and Chourey 1989; Richard *et al.* 1991), these sequences may be also involved in gene regulation by anaerobic treatment. The presence of a regulatory *cis*-element in exon1 of *RSs1* and *Sh1* is likely, since apart from the enhancing effect of the *Sh1* first exon on gene expression observed by Maas

et al. (1991), a deletion experiment has shown that a 20 bp sequence immediately downstream of the *Sh1* TATA box plus 42 bp of the untranslated exon1 (including the +9 sequence) is sufficient to reproduce the expression profile and activity of the full *Sh1* promoter in maize suspension protoplasts (Maas *et al.* 1990). The presence of promoter-like elements inside a transcription unit has been reported for other genes (Geiduschek and Tocchini-Valentini 1988).

Sucrose synthase 1 genes are subject to down-regulation by sugars (Koch *et al.* 1992; Nolte and Koch 1993a; Maas *et al.* 1990). However, the DNA sequences common to the promoters of several sugar-repressed plant genes including *Sh1* (Karrer and Rodriguez 1992) is not present in either the 5'-flanking region or the first exon of the *RSs1* gene. Through the analysis of promoter deletions, Maas *et al.* (1990) identified a 26 bp region (-20 to +6) in the maize *Sh1* gene which they found responded negatively to high sucrose concentrations and positively to cell wall regeneration in a transient expression system, suggesting the presence of sugar- or/and cell wall synthesis-responsive *cis*-elements within this 26 bp sequence. The potential importance of this region is supported by the high sequence similarity between *RSs1* and *Sh1* (Fig. 3.10), which also suggests that *RSs1* may contain similar sugar- or/and cell wall synthesis-responsive elements around this region.

```
RSs1   -20  GGTCCTTCTGGA-TCTCGGAGAAACCCTCC  +9
Sh1    -21  GGTCCCTCTCCCGTCCCAGAGAAACCCTCC  +9
```

Figure 3.10. Comparison of sequences near the transcription start sites immediately downstream of the TATA boxes in *RSs1* and *Sh1*. Highly conserved regions are underlined.

The *RSs1* sequence obtained includes 1756 bp 5'-flanking sequence, longer than in the maize *Sh1* promoter (1347 bp) used by Yang and Russell (1990) to direct phloem cell-specific gene expression in transgenic tobacco, so it should contain all the necessary regulatory information required for phloem-specific activity.

In summary, the rice sucrose synthase 1 (*RSs1*) gene has been isolated and fully sequenced, and a full-length promoter identified. The structure of *RSs1*, including the location and number of introns and exons, is similar to that of the *Sh1* gene. The *RSs1* 5'-flanking region contains sequences homologous to those found in several endosperm-specific, anaerobiosis-inducible, or phloem-specific promoters, but whether these sequences are involved in gene regulation is as yet undetermined. Sucrose synthase has been interesting to plant physiologists as it plays a potential role in several important metabolic processes in plants, including starch synthesis (Chourey and Nelson 1976; Heim *et al*. 1993; Wang *et al*. 1993), cell wall synthesis (Hendrix 1990; Chourey *et al*. 1991a) and phloem transport of photoassimilate (Nolte and Koch 1993b; Geigenberger *et al*. 1993). However, the most detailed studies on both the SS enzymes and genes are still restricted to maize. Therefore, besides the potential usefulness of the promoter sequence, the rice sucrose synthase genes isolated in this work can be used for the study of SS gene expression in rice plants.

CHAPTER 4. ISOLATION OF PHLOEM PROTEIN
PP2 PROMOTER

4.1. Results

4.1.1. Sequencing of PP2 protein

No sequence data on either the PP2 protein or genes were available in the literature, so the isolation of the PP2 gene promoter was initiated by sequencing the protein. The PP2 protein polypeptide was purified by SDS-PAGE of phloem exudate from 3 months old *Cucurbita pepo* L. "Autumn Gold" plants, after initial experiments had shown that conventional protein purification techniques were hindered by gelling of the protein on contact with air, and that SDS-PAGE was consequently the quickest and most efficient method of obtaining small amounts of pure material. A typical separation is shown in Fig. 4.1. The SDS-PAGE pattern of the exudate proteins was similar to those previously reported for *Cucurbita* species with PP2 as a major protein at 22-26 kD (Sabnis and Hart 1978; Read and Northcote 1983a; Cronshaw and Sabnis 1990). As shown in Fig. 4.1, the exudate protein was apparently not fully dissociated under the conditions for sample treatment (100 °C 4 min or 60 °C 10-20 min), similar to the results in many previous reports (Sloan *et al.* 1976; Cronshaw and Sabnis 1990; Bostwick *et al.* 1992). The presence and intensities of the prominent bands at about 45 kD and 66 kD seen in Fig. 4.1 are variable in different reports (Sloan *et al.* 1976; Cronshaw and Sabnis 1990) and dependent on experimental conditions (e.g. concentrations of β-mercaptoethanol) (Sloan *et al.* 1976; Read and Northcote 1983b), suggesting that they are all cross-linking products containing the PP2 polypeptide. The molecular weights of the two prominent bands at about 45 kD were similar to those of PP2 dimers of *Cucurbita maxima* (40 kD and 48 kD) (Read and Northcote 1983b).

Initial attempts to directly sequence the 24 kD PP2 polypeptide, either purified by electrophoretic elution from a polyacrylamide gel, or blotted onto polyvinylidene

97
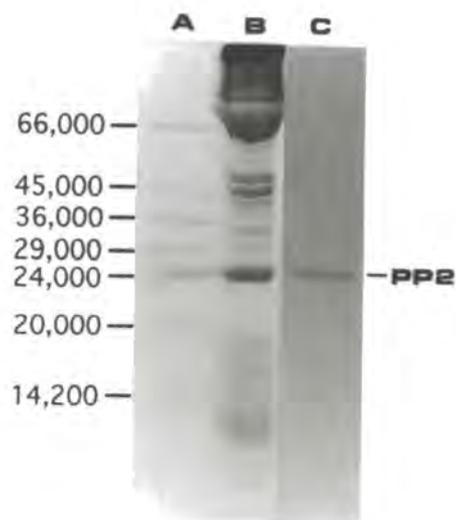
Figure 4.1. SDS-PAGE analysis of pumpkin phloem exudate and purified PP2 protein. Gel is 12% acrylamide. Track A, protein markers (SDS-7, Sigma), whose molecular weights (in dalton) are given at the left. Track B, 10 µl of pumpkin phloem exudate. Track C, PP2 protein purified by SDS-PAGE and electrophoretic elution. Proteins were stained with Coomassie blue.

fluoride (PVDF) membrane were unsuccessful, because the N-terminus of the protein was totally blocked to Edman degradation. No amino acid residues were obtained in amounts above background, demonstrating that the purified PP2 polypeptide was not contaminated with other sequenceable polypeptides. The PP2 polypeptide blotted onto PVDF membrane could be efficiently cleaved with cyanogen bromide (CNBr) vapour, generating two sequenceable peptides, which gave 18 residues of mixed sequence data after 9 sequencing cycles (Table 4.1). Since the chemical cleavage of a protein by CNBr occurs at methionine residues, this result suggested that there were two methionine residues in the PP2 polypeptide (unless two different PP2 polypeptides were present), consistent with estimation from previously reported amino acid compositions of PP2 (Weber et al. 1974; Allen 1979). Elution of these cleaved peptides from the membrane, and separation by SDS-PAGE or HPLC on a reverse-phase column was not successful, due to low recovery (less than 10%, according to analysis of amino acid residues by the automated sequencer), making it difficult to distinguish sequence data from background. In situ cleavage of PP2 in gel slices by CNBr vapour proved a successful technique; the cleaved polypeptides

Table 4.1. Mixed amino acid sequence of unseparated CNBr-cleaved PP2 on PVDF membrane

Sequencing cycles	1	2	3	4	5	6	7	8	9
Amino acid residues in each cycle	D	N	K	E	K	E	A	R	E
	L	S	P	N	V	V	Y	E	V

produced were separated by a second SDS-PAGE step and blotted onto PVDF membrane (Fig. 4.2). Two peptides, of about 14 kD and 15 kD, were obtained, and both were successfully sequenced; the sequences obtained are shown in Fig. 4.3. The first nine residues at the N-termini of these two peptides were consistent with the mixed sequence data given in Table 4.1. The expected third peptide, corresponding to the blocked N-terminus, was not seen as a band on SDS-PAGE after CNBr cleavage, suggesting that it has a low molecular weight. The predicted amino acid sequence of
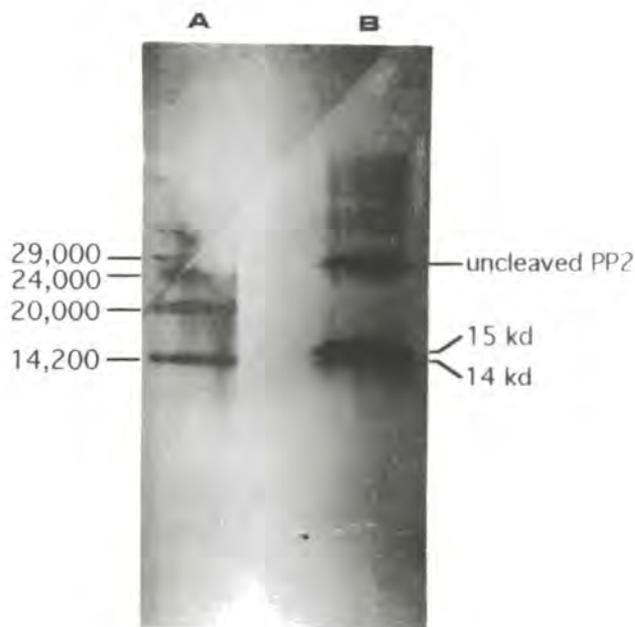
Figure 4.2. Electroblot of CNBr-cleaved PP2 polypeptides.

Pumpkin phloem exudate (50µl) was electrophoresed in a 12% SDS-polyacrylamide gel. The PP2 band was sliced and treated with CNBr vapour, and the cleaved peptides were separated on a 17% SDS-polyacrylamide gel and electroblotted onto a polyvinylidene fluoride (PVDF) membrane (Track B). Track A, protein markers. Proteins were stained with Coomassie blue.

15 kd peptide:

```
1                       10                      20                      30
D-N-K-E-K-E-A-R-E-K-L-G-G-E-V-K-L-G-H-X-L-D-V-I-L-K-N-A-D-V-

31                      40
A-L-X-Y-P-S-F-L-K-L-Y-D-
```

14 kd peptide:

```
1                       10                      20                      30
L-S-P-N-V-V-Y-E-V-A-L-Q-V-Q-L-N-S-R-A-S-G-W-N-A-P-L-N-I-E-L-

31
K-K-P-D-G-
```

Figure 4.3. N-terminal amino acid sequences of purified peptides produced by CNBr-cleavage of PP2 polypeptide. X = uncertain residue.

PP2 (Fig. 4.5, next section), from cDNA clones, suggests that this third peptide is a single methionine residue. A 28 kD band on the PVDF blot was uncleaved PP2. Fig. 4.2 shows that there are other faint high-molecular-weight bands on the PVDF blot, which must represent oligomers of PP2 or cleaved PP2 polypeptides, since the PP2 monomers had been separated from other proteins by being excised out of a gel before being treated with CNBr. Interestingly, the pattern of these high-molecular-weight bands in Fig. 4.2 was similar to that of crude phloem exudate proteins shown in Fig. 4.1, especially when the latter were separated and blotted under the same conditions as the cleaved peptides. Other workers also found that purified PP2 could be oxidised to polymeric material (Read and Northcote 1983b).

4.1.2. Isolation of cDNA clones and nucleotide sequencing

Total RNAs isolated from different tissues of *Cucurbita pepo* were examined by Northern blotting using a fully degenerate oligonucleotide probe, 5'-GC(C/T)TC(C/T)TT(C/T)TC(C/T)TT(A/G)TT(A/G)TCCAT-3', corresponding to amino acids 1-7, and the implied methionine residue before amino acid 1, of the 15 kD CNBr fragment of PP2 (Fig. 4.3). The result shown in Fig. 4.4 indicated that 3-5 days old hypocotyls of *Cucurbita pepo* contained the highest level of PP2 mRNA, consistent with the result for *C. maxima* (Sham and Northcote 1987). Poly(A)$^+$ RNA was therefore isolated from 1 mg of total RNA extracted from 3-5 days old hypocotyls and used as a template for cDNA synthesis. A library containing 9×10^5 independent clones in the lambda phage vector ZAP II was obtained using one-twelfth of the resulting double-stranded cDNA. Approximately 38,000 recombinants of this primary λZAP cDNA library were screened using the same oligonucleotide probe (see above), and approximately 150 plaques (0.4%) gave strong hybridisation signals.

After plaque purification, 12 positive clones were subjected to *in vivo* excision to form insert-containing pBluescript phagemids. Seven of the 12 clones had similar size of inserts (~900 bp), and two of them, p9A and p11A, were fully sequenced; they

~ 0.9 kb
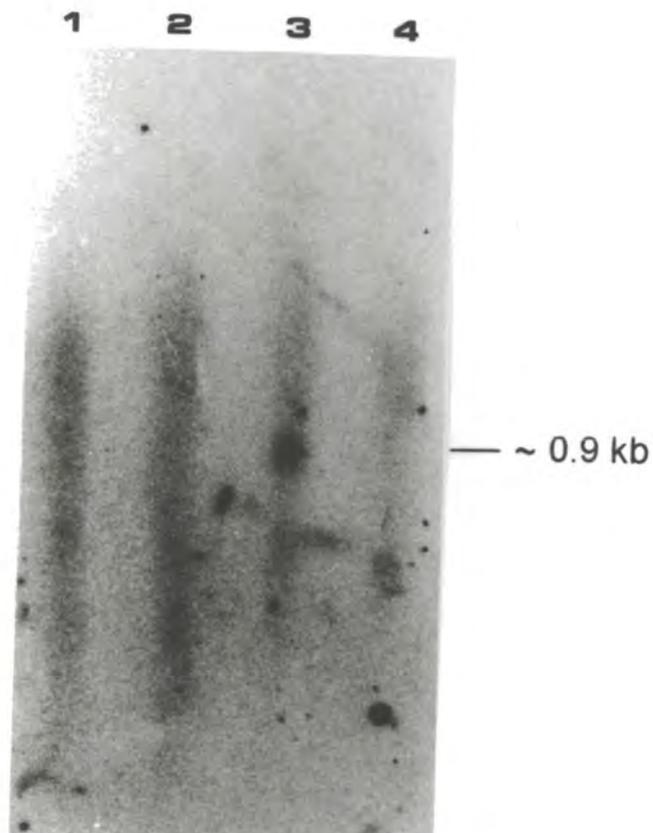
Figure 4.4. Northern blot of *Cucurbita pepo* RNA from different tissues. Total RNA (50 μg per track) from leave (track 1) and stem (track 2) of 20 days old plants, and from hypocotyls (track 3) and cotyledons (track 4) of 3-5 days old seedlings were separated on a 1.2% formaldehyde-agarose gel, transferred onto a nitrocellulose membrane and hybridised with the 23-nt oligonucleotide probe based on the PP2 amino acid sequence. The blot was exposed for 1 hr.

```
AAAAAAGAAAGCGCTATAATGGACAACAAAGAGAAGGAAGCGAGAGAGAAATTAGGAGGA        60
                  M  D  N  K  E  K  E  A  R  E  K  L  G  G        14

GAAGTGAAGCTTGGTCATTGCTTGGATGTTATTTTGAAGAATGCTGACGTAGCACTGCAC       120
 E  V  K  L  G  H  C  L  D  V  I  L  K  N  A  D  V  A  L  H        34

TATCCCTCCTTCCTTAAGCTTTATGACCAACTTGTTGCTGGGATCCTCTTGAACAAGGGA       180
 Y  P  S  F  L  K  L  Y  D  Q  L  V  A  G  I  L  L  N  K  G        54

GCTATAAAGTACATCTTTGATAAGAAGTTAAACAGCCATTGGTACTTTATATTTGCGAGA       240
 A  I  K  Y  I  F  D  K  K  L  N  S  H  W  Y  F  I  F  A  R        74

GCTCTCTCAATCGCCTGGATTGAAGATAAAAGATACTGGAAATGGGGATCCTGTGGCAAT       300
 A  L  S  I  A  W  I  E  D  K  R  Y  W  K  W  G  S  C  G  N        94

AGCGAAGTTGCAGAACTTATTGAAGTATCTTGGCTGAACATTCGTGGAAAGATCAATGAG       360
 S  E  V  A  E  L  I  E  V  S  W  L  N  I  R  G  K  I  N  E       114

TCTATGCTCTCACCAAATGTTGTGTATGAAGTAGCACTTCAGGTACAGCTAAATAGTAGA       420
 S  M  L  S  P  N  V  V  Y  E  V  A  L  Q  V  Q  L  N  S  R       134

GCCTCCGGGTGGAATGCTCCTTTGAACATCGAATTGAAGAAGCCAGATGGGAGCAAGATA       480
 A  S  G  W  N  A  P  L  N  I  E  L  K  K  P  D  G  S  K  I       154

GTGCGCCAGGAATGCCTGTTGGGGAAGCCACAAAACCAGTGGTTTGAGATTGTTGTTGAG       540
 V  R  Q  E  C  L  L  G  K  P  Q  N  Q  W  F  E  I  V  V  E       174

TTCAAGGTAGGTAACCATGGCTGTGGAAGTAGTGGCGAGATCGAGTTTTCCTTTTTTGAA       600
 F  K  V  G  N  H  G  C  G  S  S  G  E  I  E  F  S  F  F  E       194

CATGGAGGGCATTGGAAGAGAGGGCTGCTCGTGAAAGGTGTTCGGATTGGAGCAAAGGGA       660
 H  G  G  H  W  K  R  G  L  L  V  K  G  V  R  I  G  A  K  G       214

TGTGGTTGCTCATGATCAAAATAGTCTTGGTGTACACAATCTCTCTTTCTCAAACTCAAA       720
 C  G  C  S  *

GTACACTTATTTTGATGTTGAGGAACGAGAGTTTGTGTTATAATCAAATATGAAAGAATG       780

TACTAGTATGAAACAATAAATAATAGACCTTTTGCTTACGGGCAAGAAGGTTAAGTTTT(A)$_n$
```

Figure 4.5. Nucleotide sequence of the PP2 cDNA, with the deduced amino acid sequence for the PP2 polypeptide. Amino acid sequence regions covered by PP2 CNBr peptide sequences (Fig. 3) are underlined. A potential polyadenylation signal sequence in the cDNA is also underlined. The stop codon is marked by an asterisk.

contained identical coding and non-coding sequence with similar extent of 5' extension (Fig. 4.5). Another clone, which contained a larger insert (2.2 kb; p27A), was also sequenced. The insert in this clone contained the entire PP2 cDNA (identical in sequence to the inserts of p9A and p11A) fused with a partial cDNA sequence, which was identified on the basis of sequence homology to entries in the EMBL database as encoding endo-β-1,4-glucanase precursor (Appendix D). Another two of the 12 clones, containing 1.5 kb (p4A) and 1.9 kb (p30A) inserts respectively, were also investigated; their partial sequences (Appendix D) in the 5' regions did not encode the PP2 protein and matched no entries in the EMBL sequence database. The strong hybridisation of these two clones to the oligo probe might be due to presence of homologous sequences to the probe, or to presence of a fused PP2 cDNA fragment as in p27A.

The nucleotide sequences of the PP2 cDNA clones contain a region (bases 19-41) exactly complementary to one of the oligonucleotides in the degenerate probe. About 840 bases of sequence were determined, plus a poly(A) tail of variable length; the predicted mRNA size is thus in agreement with the species observed on the Northern blot (Fig. 4.6). A polyadenylation signal sequence is present 45 bases 5' to the poly(A) tail in all the clones sequenced. The first ATG in the sequence has been assumed to be the start codon, and then predicts a correct amino acid sequence; this assumption is based on the similar 5' extensions of the three sequences determined, which suggest that the cDNAs are near full-length. Results from CNBr cleavage of PP2 make it unlikely that the coding sequence could extend significantly further in the 5' direction, although the nucleotide sequence determined does not rule this possibility out.

The deduced amino acid sequence in the cDNAs is of 218 amino acid residues, and predicts a polypeptide of molecular mass 24,550, in agreement with the observed size of the PP2 polypeptide (24 kD). The predicted sequence contains regions identical to the sequences determined for the CNBr fragments of PP2 (Fig. 4.3); the N-terminal region of the 15 kD fragment follows the methionine residue at position 1, and

the N-terminal region of the 14 kD fragment follows the methionine residue at position 116. There are two undetermined residues in the former sequence; one is predicted to be cysteine, which would not have been determined as the cysteine residues were not carboxymethylated; the other is predicted to be histidine, which is consistent with the result obtained at this position, although the trace from the sequencer is not clear enough to allow the residue to be unambiguously designated as histidine. The predicted molecular masses of the fragments are 13,277 and 11,159 respectively; these fragments are smaller than the sizes determined by SDS-PAGE, but calibration of the gels is not reliable in this molecular weight range. No other methionine residues are present in the sequence, and the N-terminal methionine residue must thus be blocked in the mature protein. Comparison of the 15 kD peptide sequence with the deduced amino acid sequence suggests that this is the initial methionine residue. The correspondence of the sequence predicted by the cDNAs and the sequences determined for the CNBr fragments of PP2 establish that these cDNAs contain the coding sequence of a PP2 gene.

The insert from p11A was used as a probe on a Northern blot of total seedling hypocotyl RNA. As shown in Fig. 4.6, p11A hybridised to a single mRNA band of Approximately 0.9 kb, and the PP2 mRNAs were abundant in the young hypocotyls, which agrees with previous measurements of the mRNA encoding PP2 (Sham and Northcote 1987; Bostwick *et al.* 1992).

4.1.3. Sequence comparison

The predicted amino acid composition of the PP2 protein is in good agreement with that determined by other workers (Table 4.2), being rich in glycine (10.1%), leucine (10.6%), glutamic acid/glutamine (10.6%), aspartic acid/asparagine (9.2%), and lysine (9.6%), with no threonine present, and six cysteine residues per polypeptide. Overall, the PP2 polypeptide is of basic property (Fig. 4.7), consistent with a previous report that the isoelectric point for the PP2 of *C. maxima* was pH 10.2-10.4 (Weber *et al.* 1974). When the PP2 sequence was run against the GenBank and EMBL

databases, no nucleotide or protein sequences with significant homology (e.g. >52%) was found, except for the PP2 sequence of *C. maxima* (Bostwick and Thompson 1993). The 654 bp PP2-coding sequence in p11A (or p9A, p27A) shares 95.7% identity with that of *C. maxima*. The 3' non-coding regions, however, have very little homology between them. The deduced amino acid sequence for the PP2 protein differs from that of Bostwick *et al*. (1992) at nine positions (Table 4.3), and all but two (residues 67 and 96) of the changes in amino acid residue are produced by single base changes. In codon usage of the PP2 cDNA, A and T are preferred over G and C compared with the PP2 cDNA of *C. maxima*, as among the 28 base substitutions in the coding region, nine C→T and seven G→A transitions occur in contrast to one T→C and two A→G changes. The other two entries with the best scores of sequence match to the PP2 cDNA in the EMBL database are the genes encoding translation elongation factor 3 from *Candida albicans* and the gene encoding an yeast ubiquitin-specific processing protease (UBP3). The elongation factor 3 genes exhibit 52.1% (Myers *et al*. 1992) or 49% (DiDomenica *et al*. 1992) nucleotide identities to the PP2 cDNA within two approximately 900 bp coding regions. The sequence identity between the PP2 cDNA and a 890 bp region of UBP3 coding sequence (Baker *et al*. 1992) is 51.7%. However, there is no substantial homology at protein level between PP2 and the other two genes.

The predicted sequence of PP2 does not contain sequence with significant homology to the chitin-binding domains found in chitinases and chitin-binding lectins such as wheat germ agglutinin and nettle lectin (Beintema and Peumans 1992). However, PP2 shares a sequence, KWGSCG with the wheat germ agglutinin isolectins (Smith and Raikhel 1989) and a root-specific lectin from barley (Lerner and Raikhel 1989). This sequence is the longest match of PP2 to the entries in the EMBL protein sequence database.

Table 4.2. Predicted PP2 amino acid compositions (mol %) of *C. pepo* and *C. maxima* and previously published biochemical quantitations

Amino acid	p11A	Bostwick *et al.* 1992	Allen 1979	Weber *et al.* 1974
Ala	6.0	7.3	6.6	7.5
Arg	3.7	3.7	3.8	3.5
Asp	3.2	4.1	9.9*	11.1*
Asn	6.0	6.0		
Cys	2.8	2.8	2.7	2.0
Glu	7.8	7.3	10.4**	10.2**
Gln	2.8	3.2		
Gly	10.1	10.1	11.0	10.2
His	2.8	2.3	2.1	2.4
Ile	6.9	6.9	6.6	5.3
Leu	10.6	10.1	10.4	10.1
Lys	9.6	9.6	8.8	9.7
Met	0.9	0.9	1.1	1.0
Phe	4.1	4.1	3.8	4.0
Pro	2.3	1.8	2.2	3.6
Ser	6.9	6.4	6.6	5.6
Thr	0.0	0.0	0.0	2.9
Trp	3.7	3.7	2.2	2.6
Tyr	2.8	2.8	3.3	2.7
Val	7.3	6.9	8.2	5.2

* and ** Combined values for Asp plus Asn and Glu plus Gln

Table 4.3. Amino acid substitutions between PP2 and the published sequence.

Positions	64	67	94	96	107	119	155	191	218
p11A	L	H	N	E	N	P	V	S	S
Bostwick *et al.* 1992	N	N	D	N	D	Q	A	A	A
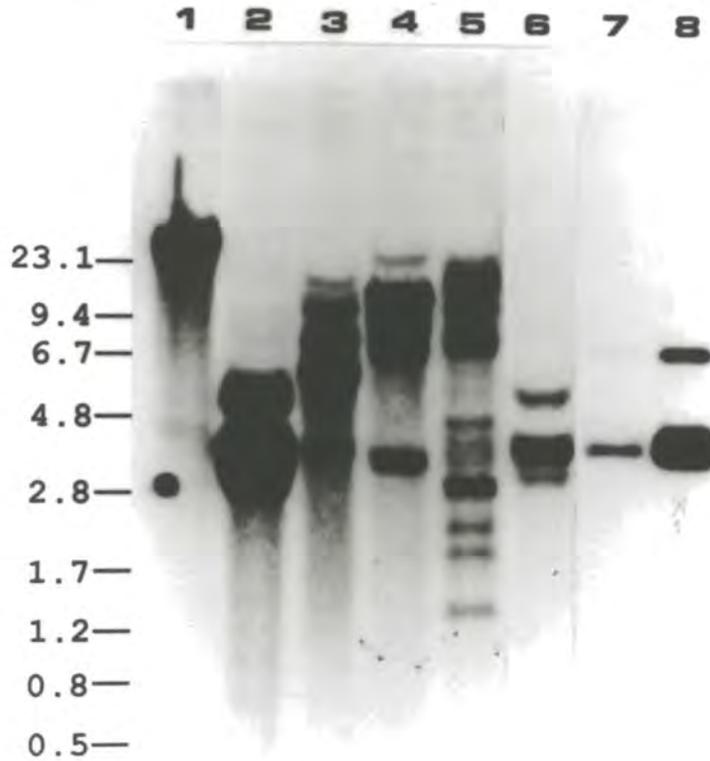
Figure 4.8. Southern blot of restriction endonuclease digests of *Cucurbita pepo* genomic DNA hybridised with ^{32}P-labelled cDNA insert of PP2 cDNA clone p11A. Track 1-5, *C. pepo* genomic DNA (10 μg per track), undigested (track 1) and digested with HindIII (track 2), KpnI (track 3), BamHI (track 4) and XbaI (track 5). Track 6, same as track 2 but with less exposure (5 hrs; 17 hrs for all the others). Track 7 and 8, XhoI-digested PP2 cDNA (p11A, 3.8 kb): 30 pg in track 7, 300 pg in track 8 (1 and 10 gene copy equivalent, respectively). Gel is 0.7% agarose. Molecular weight markers (kb) are shown at the left.

4.1.4. Southern blotting of genomic DNA

Genomic DNA from *Cucurbita pepo* was restricted with various enzymes, separated by gel electrophoresis, blotted, and probed with the PP2 cDNA (insert from clone p11A). Results are shown in Fig. 4.8. The hybridisation pattern obtained suggested that PP2 was encoded by a gene family with a relatively large number of members (within the approximate range 7-15, comparing the intensities of hybridisation to gene copy equivalents run on the same gel). Digestion with HindIII gave a major band of 4.0 kb, with fainter bands at Approximately 5.0 kb and 3.1 kb. KpnI, BamHI and XbaI gave 5-10 hybridising fragments, at intensities corresponding to approximately single and multiple copies. With longer exposure, a small band (~200 bp) can be seen in the BamHI-digest (picture not included), indicating that there are BamHI sites within the coding region, consistent with the PP2 cDNA sequence.

4.1.5. Isolation and characterisation of PP2 genomic DNA

A genomic library containing 3.6×10^5 recombinants was constructed by cloning partially digested (with Sau3AI) and partially filled-in genomic DNA of *C. pepo* into the Promega λGEM-12 XhoI half-site arms vector. An important modification on the cloning procedure, which led to a significant increase in ligation efficiency, was the removal of residual carbohydrates from the genomic DNA by CTAB precipitation following partial digestion. Approximately 4.4×10^5 plaques of an amplified library was screened using the PP2 cDNA clone p11A as a probe, and about 56 strong hybridisations were obtained. The high population of positive plaques was consistent to the result of Southern analysis (Fig. 4.8) which shows that the PP2 gene belongs to a multi-member family. Four of the clones, arbitrarily selected and designated CPP1.2, CPP1.3, CPP2.1 and CPP2.3, were purified and analysed. Preliminary results from restriction endonuclease digestion and DNA hybridisations indicated that these four clones were different from each other with regard to the patterns of their restriction fragments, making it unlikely that the high population of positive clones in the genomic library is the result of uneven amplification due to

112
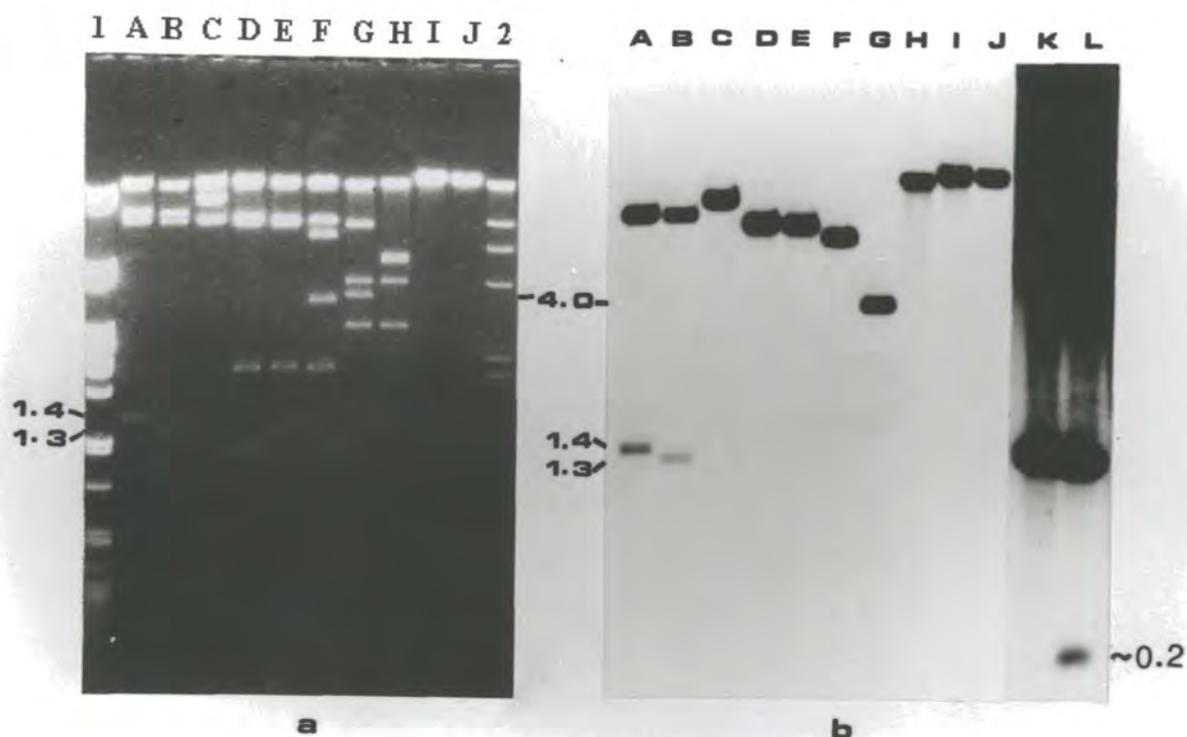
Figure 4.9. (a) PP2 genomic clones CPP1.2 (tracks A and D), CPP1.3 (tracks B, E, and G), and CPP2.3 (tracks C, F and H) were digested with BamHI (tracks A, B and C), EcoRI (tracks D, E and F) and HindIII (tracks G and H), and separated on a 0.8% agarose gel. Tracks I and J, undigested CPP1.2 and CPP2.3 respectively. Tracks 1 and 2, molecular weight markers (PstI- and HindIII-digested λDNA, respectively). (b) The separated DNA digests were transferred from the gel onto a nitrocellulose membrane and hybridised with ^{32}P-labelled cDNA insert of the PP2 cDNA clone p11A. Tracks K and L, a longer exposure of tracks A and B, respectively. Molecular size (kb) is given to the fragments referred in the text.

preferential growth of some particular phages. Only CPP1.3 gave a 4.0 kb HindIII and a ~200 bp BamHI fragments hybridised to the PP2 cDNA (lanes B, G and L; Fig. 4.9). Based on the PP2 cDNA sequence which contains two BamHI sites (125 bp apart) and the Southern blot where the 4.0 kb HindIII fragment was the major hybridisation band, CPP1.3 was most likely the expected PP2 genomic clone and therefore further characterised. CPP1.2 gave a large hybridisation fragment and a small band at 1.4 kb (lane A) (similar in size to the 1.3 kb BamHI fragment of CPP1.3), but it did not contain the expected ~200 bp BamHI fragment (lane K), indicating that there was only one BamHI site inside the PP2 gene sequence. CPP1.2 also lacked the expected 4.0 kb HindIII fragment (lane D). This result along with other restriction endonuclease analysis (results not included) suggests that CPP1.2 is a different member of the PP2 gene family.

A putative restriction map of CPP1.3 is drawn based on restriction endonuclease digestion and the PP2 cDNA sequence (Fig. 4.10). The 1.3 kb BamHI and the 4.0 kb HindIII fragments were subcloned and partially sequenced. The partial sequence shown in Fig. 4.11 contains an open reading frame nearly identical to the corresponding sequence of the PP2 cDNA, confirming that CPP1.3 was a PP2 genomic clone. The putative translation start codon ATG is the only one in frame within the contiguously sequenced 265 bp upstream region. There are two intervening sequences, 101 bp and 87 bp in length, respectively, which are consistent with the GT-AG rule of introns. The same number of introns was also reported for the PP2 genomic DNA from *C. maxima* (Bostwick *et al*. 1993). The exact transcription start site is not determined, but the similar extension in the 5' primer region of all the three PP2 cDNA clones sequenced, as shown in Fig. 4.5, suggests that this site should not be far away from the adenosine 18 bp upstream of the ATG codon. It is therefore of little likelihood that an intron is present upstream of the translation start site as in the case of rice sucrose synthase-1 gene (*RSs1*). The partial coding sequence of the genomic clone contains three different nucleotides to the corresponding cDNA sequence, and one of the variations results in an amino acid substitution (from a leucine
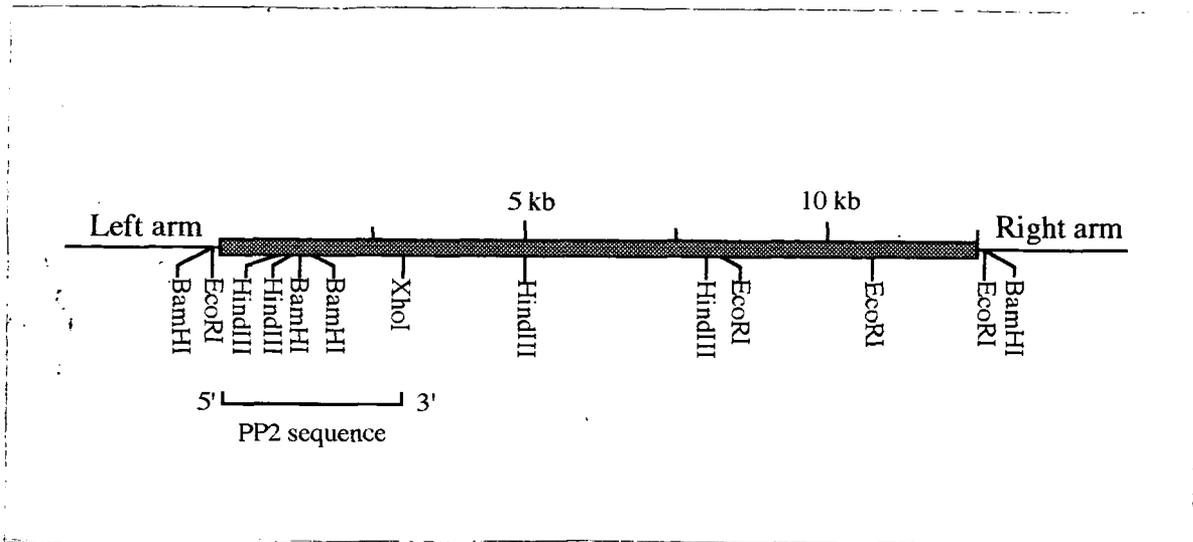
Figure 4.10. Restriction enzyme map of the PP2 genomic clone CPP1.3.

```
GATCTTTCTATTTAACTTTAAAAAAGTATCAAAAATCACATTGTCAGAAAAAAAACAAAAAATCTTAATA

TCTGATCAACTTTTATAAATTTTTAAGACTAAATTTATTAAAAGTAGGTTTGTTGAAAAATCGATAGTCT

AAATAAAACTTAAGTGTTCGAATATGAATAAATAAATGTGATTAATGTACAATATCGACGGATAAATAAT

TATAGGAAAAAAATTATAACTTTCAAAAAAATAACGGTTCTAATTTTATCTTTTTCATATGTCAAATAGG

TTTTTTATTTAATATATATTT........ATATATATTTTTCCTGATTTTAAAATTTAAATTAAAAAGAA

TATATCAAAAGAAGAGATTATTATGGAAGCTTAAAAAATTCCTTGCTTTTGTATAAAATAGTTAAAAAA

ATTAAAAATTAAAAAAATAAAAAAGTGAAAATAACAAATGTGCATGACTTAGGTTCAAAATGTTGTAGCA

          **************                                            *
AAGGAGGGGTTATATATATATCCATTCTTCCCTCTCACATTAACTCATATCTAAGTTCAGTTCATAAAAA
          TATA box

    HaeII                                                       HindIII
AGAAAGCGCTATAATGGACAACAAAGAGAAAGAAGCGAGAGAGAAATTAGGAGGAGAAGTGAAGCTTGGT
              M  D  N  K  E  K  E  A  R  E  K  L  G  G  E  V  K  L  G

                                                          HindIII
CATTGCTTGGATGTTATTTTGAAGAATGCTGACGTAGCGCTGCACTATCCCTCCTTCGTTAAGCTTTATG
 H  C  L  D  V  I  L  K  N  A  D  V  A  L  H  Y  P  S  F  V  K  L  Y

          BamHI
ACCAACTTGTTGCTGGGATCCTCTTGAACAAGGGAGCTATAGTAAGTGCAACTATATATATATATATNNN
 D  Q  L  V  A  G  I  L  L  N  K  G  A  I <-----------------------------

NCNNCCAACTTCACCTTTGTATATCATAAACATATTAAATCAAAATATTTGCTTTCTTTTCTTTTTACGT
--------------------INV-1----------------------------------------------

AGAAGTACATCTTTGATAAGAAGTTAAACAGCCATTGGTACTTTATATTTGCGAGAGCTCTCTCAATCGC
-> K  Y  I  F  D  K  K  L  N  S  H  W  Y  F  I  F  A  R  A  L  S  I  A

                              BamHI
CTGGATTGAAGATAAAAGATACTGGAAATGGGGATCCTGGTATAATTTTTAACTAATTTCTCAACGAAAA
  W  I  E  D  K  R  Y  W  K  W  G  S  C<------------------------------

AAAATGATAAAAAGTTGATTTCTTGATCACTCACTCGGGCGTCTAAAAACTTGCAGTGGCAATAGCGAAG
---------------------INV-2--------------------------> G   N   S   E

TTGCAGAACTTATTGAAGTATCTTGGCTGAACATTCGTGGAAAGATCAATGAGTCTATGCTCTCACCAAA
V  A  E  L  I  E  V  S  W  L  N  I  R  G  K  I  N  E  S  M  L  S  P  N

T.......................
```

Figure 4.11. Partial nucleotide sequence and deduced amino acid sequence of PP2 genomic clone CPP1.3. The putative TATA box, the translation start ATG, the HindIII and BamHI sites present in the cDNA sequence, and the HaeII site used in cloning the promoter, are underlined. The asterisks mark the palindrome sequence. The three nucleotides and one amino acid residue different from those of the cDNA sequence are double-underlined. The nucleotide corresponding to the first adenosine of the cDNA sequence is marked by a bold asterisk. The dashed lines indicate the unsequenced regions.

to a valine) at position 39. However, the rest of the reading frame and the 18 bp

sequence before the ATG is identical to the corresponding sequence in the PP2 cDNA.

The partial upstream sequence of CPP1.3 is highly AT-rich (79%), a common

feature of the 5'-flanking region of a gene, which suggests that this region is devoid of

other protein-encoding genes. A TATA box-like sequence flanked by two GC-rich

regions is located 73 bp upstream of the translation start codon ATG. It is also a part

of a 16-nt palindrome sequence. However, no CAAT or GGGCGG boxes are found

near this region.

4.2. Discussion

The determination of a significant proportion of the total amino acid sequence

of the PP2 protein (75 out of 218 amino acid residues) through sequencing of CNBr

fragments has allowed the identity of the PP2 cDNA to be unambiguously assigned;

there are no differences between the predicted and determined amino acid sequences.

PP2 does not have a precursor with a cleavable signal peptide, and retains its initial

methionine residue in the mature protein. It is not unusual that the initial methionine is

retained in mature proteins. A plant chloroplast ribosomal protein S16 and its

counterpart in *E. coli* both contain the initial methionine residue although it is not

blocked in the mature proteins (Schmidt *et al.* 1992). The nature of the blockage at

the N-terminus that prevents direct sequencing of PP2 remains uncertain; N-terminal

acetylation is the most likely possibility, as many cytosolic proteins are modified in this

way, putatively to increase their *in vivo* lifetime by interfering with the ubiquitin-

mediated degradation system (Mayer *et al.* 1989). Treatment of the blotted PP2

polypeptide with pyroglutamate aminopeptidase did not result in a detectable N-

terminal residue being produced, in agreement with these conclusions. The abundance

of PP2 mRNA in *Cucurbita pepo* seedling hypocotyls, and its low level in older plant

tissues, as detected by an oligonucleotide hybridisation experiment (Fig. 4.4) (in

agreement with results of Sham and Northcote [1987]), suggests that PP2 synthesis

declines as tissue development proceeds. Under these conditions, the abundance of
PP2 in the phloem would require that this protein was long-lived.

The predicted amino acid sequence for the PP2 protein differs from that given
by Bostwick *et al.* (1992) at nine positions (Table 4.3). It is not clear whether these
differences represent genuine differences in the sequence of PP2 between different
species of *Cucurbita*, or inherent heterogeneity in the sequence of PP2, due to the
presence of multiple expressed genes, with differing coding sequences, in all *Cucurbita*
species. Read and Northcote (1983a, 1983b) found that the structure of PP2 protein
was a disulphide-linked dimer, which was resolved into a doublet of monomer bands
(26,500 and 25,000 respectively) by SDS-PAGE under alkaline conditions (pH 9.5).
While this would account for the sequence differences observed between the present
study and that of Bostwick *et al.* (1992), it would predict that the two polypeptides
deduced from cDNA sequences should contain different numbers of amino acids,
which is not the case. However, since PP2 is modified at its N-terminus (or possibly at
some other positions), the occurrence of the double monomer bands could possibly be
due to partial release of the modifying group(s) caused by the alkaline condition of the
SDS-gel rather than the presence of two different peptides. One such example is the
E. coli ribosomal protein L7/L12. These two proteins are identical in sequence except
that L7 possesses an N-terminal acetyl group (Terhorst and Möller 1973), but they
show different electrophoretic mobility on both SDS- and urea-containing
polyacrylamide gel (Möller *et al.* 1972). Possible presence of intramolecular
disulphide bridges in the PP2 monomer and their partial breakdown under certain
conditions may also lead to different mobility on the gel. A supportive evidence for
this is the observation that oxidation of denatured PP2 monomers gave a third band on
SDS-PAGE at Mr of 20,000, which moved to the original Mr of 26,000 on reduction
(Read and Northcote 1983b). Furthermore, the high similarities in amino acid
compositions of PP2 between the predicted and the biochemically measured make it
unlikely the presence of a PP2 polypeptide with significant differences to the predicted
ones in molecular size or primary structure. The presence of a relatively large gene

family encoding PP2, as suggested by results from Southern blot analysis and from isolation and preliminary characterisation of the PP2 genomic clones, does, however, give the potential for considerable sequence variation in this protein, although the failure to observe sequence variation in the determined amino acid residues of PP2, and the predicted amino acid sequences in the cDNA clones examined, is unexpected.

Since the results reported here show that the *C. pepo* genome contains a multigene family potentially encoding PP2, the observation that the sequences of the three cDNA clones encoding PP2 were identical is also unexpected. It is possible that this has happened by chance, and that a number of different cDNA sequences are present; alternatively, it may be due to sequence homogenisation in the genes encoding PP2, or may indicate that although many genes potentially encoding PP2 are present, only one is active. The latter view is supported by the results on mammalian ribosomal protein genes which have been found to be present in multiple copies (Dudov and Perry 1984; Wagner and Perry 1985; Batra *et al.* 1991; Suzuki and Wool 1993;). However, only one gene in each case has been found to be functional and the other members are assumed to be processed pseudogenes (Dudov and Perry 1984; Wagner and Perry 1985)..

Partial sequencing of a PP2 genomic clone reveals the presence of genes potentially encoding a different PP2 protein to that predicted by the cDNA sequence. It is unlikely that the sequence differences between the cDNA and the genomic DNA arise from random mistakes of the reverse transcriptase (during cDNA synthesis) or the Taq DNA polymerase (during nucleotide sequencing), because the three cDNA clones gave an identical sequence and the genomic DNA had been sequenced twice around the region. Despite these variations, the high sequence identity between the genomic DNA and the cDNA, particularly in the 5' non-coding region, suggests that this genomic clone represents an active gene. The presence of introns is also an indication that the gene is active. In the case of mouse ribosomal protein genes mentioned above, the only highly active member is an intron-containing gene, while the rest of the approximately 16 members contain no intron (Dudov and Perry 1984). The

failure to identify a corresponding cDNA or PP2 polypeptide may be a result of lower expression of this particular member which leads to under-representation of its mRNA species or the proteins, or may be due to a differentiated expression at a specific developmental stage or in particular tissues different to those from which the PP2 mRNA or protein was isolated. However, the possibility of CPP1.3 being a pseudogene can not be ruled out. As discussed earlier, the protein sequencing and cDNA characterisation strongly suggest the presence of a single type of PP2 polypeptide. The substituted amino acid residue at position 39 is conserved between the PP2 protein of the species used here (*C. pepo*) and that of *C. maxima* (Bostwick *et al*. 1992), a different species, so it would be expected that PP2 proteins of the same family in the same species are also conserved at this position. Fully sequencing the genomic clone, including the 3' noncoding region, may reveal strong evidences, but deactivation of a gene may be a result of promoter silencing or heavy methylation. In this case, the activity of the gene can possibly be determined by expressing the promoter in a transient expression system or in transgenic plants.

The function of P-proteins is still a subject of speculation. Currently the most favoured hypothesis is that these proteins are involved in plant defence mechanisms, e.g. sealing of wounds and trapping of invading pathogens (Read and Northcote 1983b; Smith *et al*. 1987). However this hypothesis lacks supportive evidence. Expression of plant defence-related genes, such as the proteinase inhibitor genes (Graham *et al*. 1986), the peroxidase gene (Mohan *et al*. 1993) and the β-1,3-glucanase gene (Simmons *et al*. 1992), is generally subject to certain delicate regulations, e.g. wound-inducible, elicitor-inducible and infection-inducible. There has been no report indicating that P-protein genes are under similar regulations. The long-lasting and high concentration of P-proteins in the sieve elements of *Cucurbita*, particularly of mature fruits (Read and Northcote 1983a), seems unnecessary and uneconomical. Current studies on molecular biology of plant defence have not identified P-proteins as defence-related proteins. The suggestion of P-proteins anti-invasive role is based mainly on the previously observed chitin-binding lectin activity of

PP2 (Sabnis and Hart 1978; Allen 1979; Read and Northcote 1983b). However, the predicted sequence of PP2 does not contain sequence with significant homology to the chitin-binding domain found in chitinases and chitin-binding lectins such as wheat germ agglutinin and nettle lectin (Beintema and Peumans 1992). The sharing of a six-amino-acid sequence (KWGSCG) by PP2 and several other plant lectins, and the presence of six half-cystine residues potentially forming disulphide bridges that are essential for the function of some lectins (Beintema and Peumans 1992), though, do suggest that PP2 may have similar sugar specificities to other defence related lectins.

Although no significant sequence homology has been found between PP2 and the ribosomal proteins, comparison of them reveals striking resemblance at both protein and gene level. Weber *et al.* (1974) found that P-protein shared many biochemical and immunological similarities with ribosomal proteins. Some other properties of P-proteins, such as the presence of phosphatase activity (Cronshaw 1980) and the formation of stable dimers through interchain disulphide bonds (Read and Northcote 1983b), are also similar to those of ribosomal proteins. As discussed earlier, the PP2 is represented by a multigene family which might contain one or only a few functional members and possibly some pseudogenes, similar to the mammalian ribosomal protein genes. Recent investigations on plant ribosomal protein gene structure and expression showed that these genes are preferentially expressed in young and proliferating tissues rather than in old and expanded tissues (Lebrun and Freyssinet 1991; Marty and Meyer 1992; Boham-Smith *et al.* 1992; Joanin *et al.* 1993; Gao *et al.* 1993), and generally present in a small multigene family (Lebrun and Freyssinet 1991; Bonham-Smith *et al.* 1992; Joanin *et al.* 1993; Sanwan *et al.* 1993; Gao *et al.* 1993). The PP2 mRNA level was high in hypocotyls of 3-5 days old *Cucurbita pepo* seedlings but almost undetectable in the stems and leaves of 3 months old plants when probed with the degenerate oligonucleotide probe, which is similar to the plant ribosomal gene (Marty and Meyer 1992). Sham and Northcote (1987) reported that the total PP2 mRNA level increased in the first 4 days of seedling growth but decreased to lower level in older seedlings, resembling Joanin *et al.*'s observation (1993) that the maize

cytoplasmic ribosomal protein S13 mRNA level in germinating seeds reached a maximum level after 4 days incubation.

From the proceeding discussion it seems that P-proteins are closely related to ribosomal proteins. It has long been observed that the origin of P-proteins was closely linked to the presence of helical free polysomes and the formation of P-protein filaments coincided with the disappearance of ribosomes (Behnke 1974). However, the significance of this remains unknown. Weber *et al.* (1974) suggest that P-proteins are not specifically synthesised during phloem differentiation but are aggregation of a related class of pre-existing, relatively stable, cross-linked proteins, possibly containing ribosomal or ribosome-related proteins. This view does not necessarily imply that P-proteins are intracellular debris without a function in translocation (Weber *et al.* 1974). The isolation of PP2 cDNA and genomic clones will be very useful in elucidating the function of P-proteins.

The 5'-flanking sequence in the PP2 genomic clone CPP1.3 is about 1.3 kb in length. Usually this is enough to contain necessary *cis*-elements for normal expression of a gene. A promoter of the PP2 gene has thus been isolated.

CHAPTER 5. DETERMINATION OF PROMOTER ACTIVITIES IN TRANSGENIC PLANTS

5.1. Results

5.1.1. Construct assembly

Constructs containing translational fusions between the *RSs1* promoter (containing approximately 1.8 kb of 5' flanking sequence, the transcription start site, the first intron, the translation start codon and the first 5 amino acids encoded by *RSs1*) and the coding sequences of GUS and GNA were assembled by restriction and ligation of appropriate fragments. Constructs were checked by sequencing to ensure that the reading frame had been preserved. Sequences across the fusion regions of *RSs1*-GUS and *RSs1*-GNA are shown in Fig. 5.1a and 5.1b.

The fusion site in the PP2 promoter-GUS construct (pCPGUS2), as shown in Fig. 5.1c, is inside the transcribed but non-translated region of the PP2 gene (Fig. 4.11). Mung Bean nuclease treatment has resulted in deletion of a few bases (GAAAGCGC) around the HaeII site (Fig. 4.11), so in total ten nucleotides of the 5' leader sequence immediately upstream of the ATG codon in the PP2 gene were absent in the construct, but 31 extra nucleotides had arisen from the cloning vectors between the fusion site and the translation start ATG of the GUS gene. There was no second ATG codon inside this region, so the GUS open reading frame was not affected.

5.1.2. Tissue-specific expression of *RSs1*-GUS chimaeric gene

After transformation of tobacco leaf discs with *Agrobacterium tumefaciens,* containing the *RSs1*-GUS chimaeric gene construct, twenty four putatively transformed plants were regenerated from kanamycin resistant calli. Thirteen plants were assayed by histochemical staining with X-gluc four weeks after potting. In transverse sections of stem, the internal and external phloem tissue along both sides of the xylem were

Figure 5.1. Structure of chimaeric gene constructs used in tobacco transformation; restriction sites relevant for cloning are included. (a) pBRSS7. (b) pBRSSLT. (c) pCPGUS2. RB, T-DNA right border; LB, T-DNA left border; KanR, kanamycin resistance cassette consisting of the neomycin phosphotransferase II (NPTII) gene fused to nopaline synthase promoter (Nos-Pro) and terminator (NOS-ter or NOS); RSs1, rice sucrose synthase-1 promoter including 5' flanking region, transcription start site, the first intron and translation start codon; GUS, β-glucuronidase coding sequence; GNA, snowdrop lectin coding sequence. The sequences shown are the fusion regions between the *RSs1* promoter and GUS, and GNA, and between the PP2 promoter and GUS.

(a) pBRSS7


```
        ├─ 3089 bp ─┤    ├─1870 bp─┤
    Sall            BamHI Smal          SstI
RB                                              LB
    ┌─────┐   ┌─────────┐    ┌───────┐  ┌────┐
────┤ KanR├───┤  RSs1   ├────┤  GUS  ├──┤NOS ├──
    └─────┘   └─────────┘    └───────┘  └────┘
                     ATG       ATG
```

GCT GCC AAG CTA GAG GAT CCC CGG GTA GGT CAG TCC CCT

(b) pBRSSLT


```
        ├─ 3089 bp ─┤    ├474 bp┤
    Sall            BamHI Xbal
RB                                          LB
    ┌─────┐   ┌─────────┐    ┌─────┐  ┌────┐
────┤ KanR├───┤  RSs1   ├────┤ GNA ├──┤NOS ├──
    └─────┘   └─────────┘    └─────┘  └────┘
                     ATG      ATG
```

GCT GCC AAG CTA GAG GAT CCT CTA GAG TCC GGT TCC
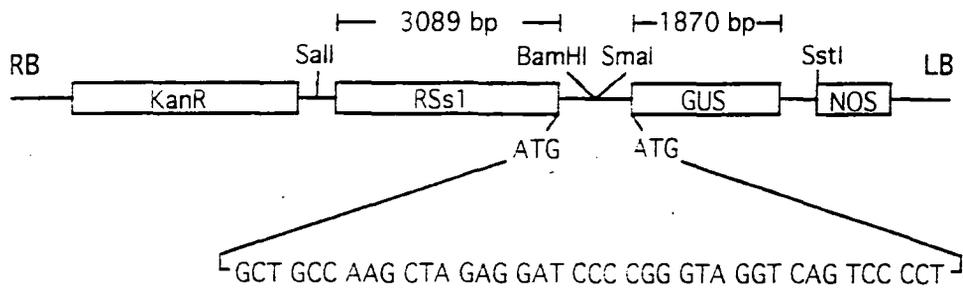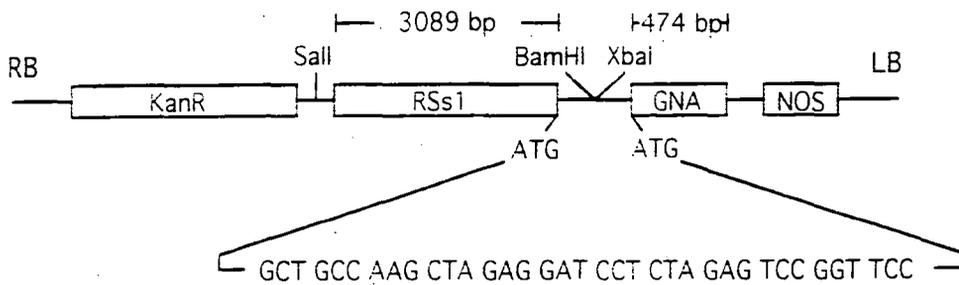
(c) pCPGUS2

Sequence around PP2 promoter/GUS fusion site:

BamHI

...TCATAAAAAA GACTCTAGAGGATCCCCGGGTGGTCAGTCCCTT ATG TTA...

From PP2 promoter ◄──┴──► From pUC and pBI vector

125

intensely stained, whereas epidermis, cortex, xylem and pith tissue of the stem remained unstained (Fig. 5.2a). This phloem specific expression pattern was also observed in petioles although the picture was not as clear (Fig. 5.2e). In leaf the specificity of GUS activity was illustrated by the vein reticulum and midrib which was blue after X-gluc staining (Fig. 5.2b). It is clear that only part of the midrib was stained which corresponds to phloem tissue, whereas other tissues including xylem, mesophyll and epidermis remained unstained. Expression of GUS in roots of transgenic plants was consistent with the pattern shown by stem and leaf, i.e. only phloem tissue was stained blue (Fig. 5.2c).

Plants transformed with a chimaeric gene construct containing the cauliflower mosaic virus (CaMV) 35S promoter fused to GUS (35S-GUS) were used as a positive control. In contrast to the tissue-specific expression directed by the RSs1 promoter, GUS activity in the 35S-GUS transgenic plants was detected in epidermis, cortex, phloem, xylem, pith tissue, and mesophyll cells as well as in the trichome (Fig. 5.2d), though the most intensely stained tissues were phloem and xylem. This is the expression pattern expected from the constitutive CaMV 35S promoter (Odell et al., 1985). When untransformed plant tissue was incubated in the staining solution under the same conditions, no blue staining was observed in any tissue of the plant. Seeds from the selfed R_0 plants (primary transformants) were germinated, and the seedlings were grown in a greenhouse. In order to examine the transmission of the chimaeric RSs1-GUS gene, eight plants from two different primary transformants, pBRSS7.4 and pBRSS7.7, were assayed for GUS expression. GUS activity was only observed in six plants. The patterns of histochemical staining for GUS activity in transverse sections of stem, petiole, leaf and root from R_1 plants (grown from self-fertilised seeds of R_0 plants) were identical to those of the R_0 plants. These results demonstrated that chimaeric RSs1-GUS gene was stably inherited to R_1 plants.

Figure 5.2. Histochemical localisation of GUS expression in *RSs1*-GUS plants (a, b, c, e) and *35S*-GUS plant (d). (a) Transverse section of stem. (b) Transverse section of leaf blade. (c) Root. (d) Transverse section of stem of *35S*-GUS plant. (e) Transverse section of petiole. Co, cortex; E, epidermis; EP, external phloem; IP, internal phloem; M, midrib; Pi, pith; T, trichome; X, xylem. Bars=100 μm.
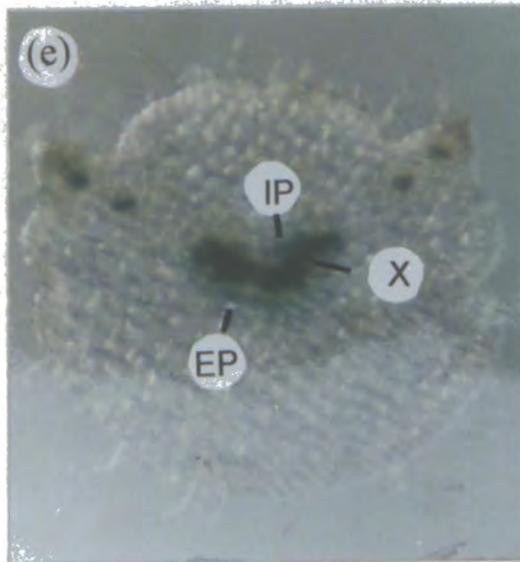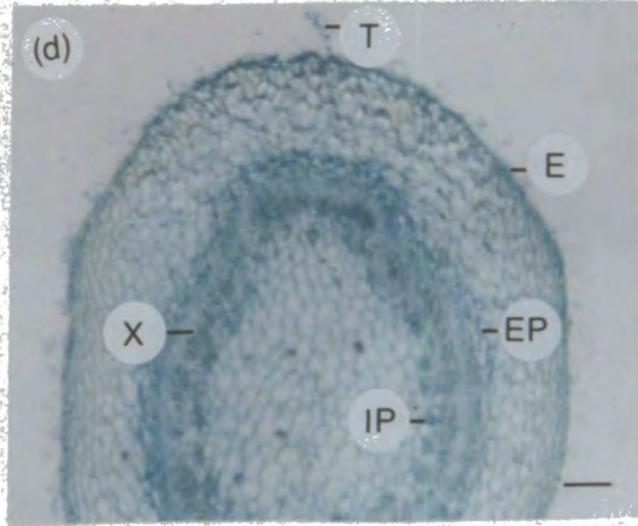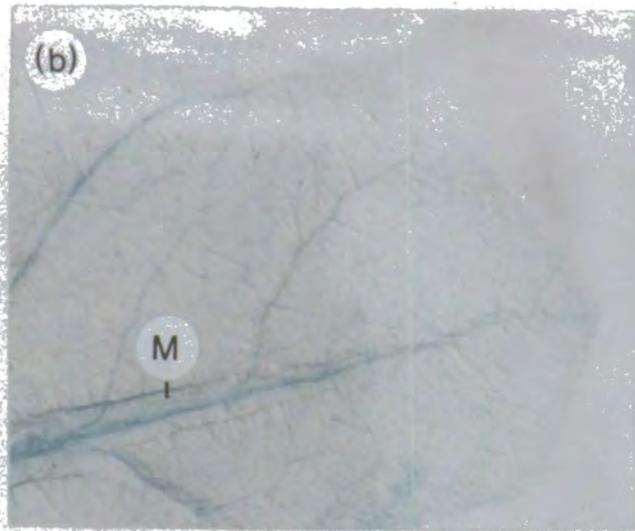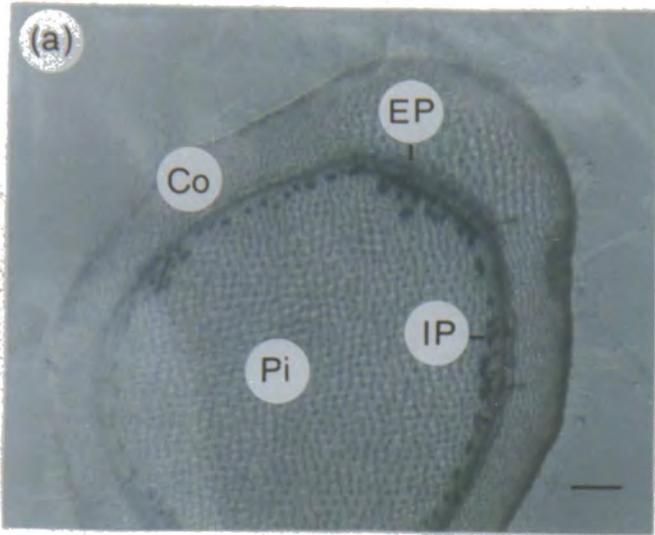
5.1.3. Dot immuno-blot and immuno-localisation of GNA in transgenic plants

Transgenic tobacco plants containing the chimaeric *RSs1*-GNA gene (pBRSSLT, Fig. 5.1b) were produced. Five and four samples from R_0 and R_1 plants, respectively, were assayed for GNA expression by dot immuno-blotting using anti-GNA antibody as a probe. GNA was detected in both R_0 and R_1 plants, but no positive signal was observed in the same samples treated under similar conditions with the primary antibody omitted, nor in the untransformed plants. The resulting blot was scanned using a densitometer to quantify the relative expression level of *RSs1*-GNA in different transformants, in comparison with the expression level of *35S*-GNA. The results are shown in Table 5.1. The relatively low expression level of GNA directed by *RSs1* promoter (0.02-0.4 of the level given by the 35S promoter) is related to the low abundance of phloem cells in different parts of the plants, and does not reflect an inherently low activity of this promoter.

Table 5.1. The expression level of *RSs1*-GNA relative to *35S*-GNA in transgenic plants, measured by densitometric scanning of protein dot-blots probed with anti-GNA antiserum. The first five accessions were primary transformants with chimaeric gene *RSs1*-GNA; the following four were R_1 transgenic plants with the same construct; 15GNA79 was transformed with the *35S*-GNA construct.

Accession	Densitometer reading (peak area, AU*mm)/ mg leaf protein	Relative level of GNA
pBRSSLT1	0.053	0.02
pBRSSLT2	0.362	0.16
pBRSSLT3	0.928	0.40
pBRSSLT5	0.522	0.23
pBRSSLT7	0.600	0.26
pBRSSLT2.3	0.373	0.16
pBRSSLT2.4	0.199	0.09
pBRSSLT7.3	0.394	0.17
pBRSSLT7.4	0.165	0.07
15GNA79	2.297	1.00

Figure 5.3. Immuno-localisation of GNA protein in *RSs1*-GNA plant (a-c). Untransformed plant (d) was included as control. (a) Transverse section of stem. (b) High magnification of the phloem tissue. (c) A similar section to (a) without primary antibody treatment. (d) Transverse section of stem from an untransformed plant. CC, companion cell; SE, sieve element; also see Figure 5.2 legend. Bars=30 μm
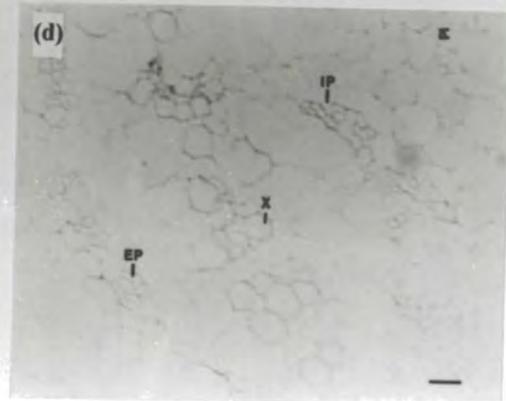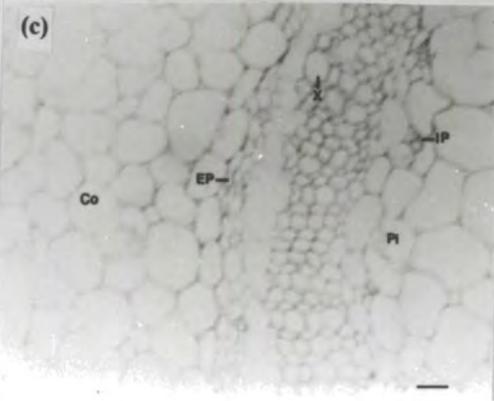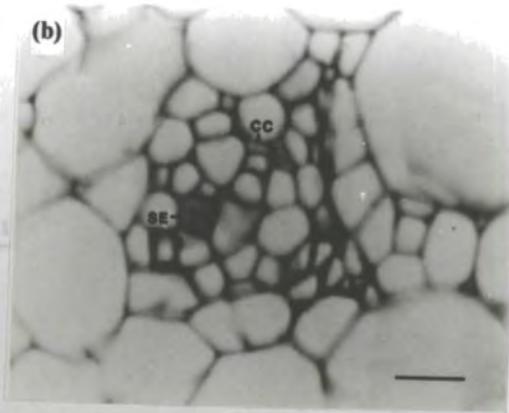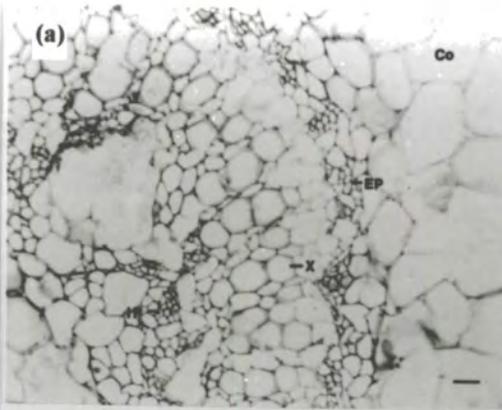
Three *RSs1*-GNA R_0 plants were assayed by immunolocalisation using anti-GNA antibodies, to detect the site of GNA protein accumulation. Untransformed tobacco plants, as controls, were assayed under the same conditions. The *in situ* immunochemical staining results show that in *RSs1*-GNA plants, GNA was only detected in phloem tissues (Fig. 5.3a), similar to the results obtained for expression of *RSs1*-GUS. However, the higher resolution possible with immunolocalisation allowed the GNA to be localised to sieve elements, companion cells and phloem parenchyma, and not in any other types of tissue and cell (Fig. 5.3b). No specific binding was seen when the primary antibody was omitted from the treatment of a stem section of an *RSs1*-GNA plant (Fig. 5.3c). When the same treatment and conditions were applied to sections of untransformed plant no staining was observed (Fig. 5.3d). These results demonstrate that the specificity of *RSs1* promoter in *RSs1*-GNA plants remains the same as in *RSs1*-GUS plants, and, as expected, is independent of the gene which it directs.

5.1.4. Immuno-detection of GNA in honeydew excreted by aphids

Peach-potato aphids were allowed to feed on a *RSs1*-GNA plant and an untransformed plant. The amount of honeydew produced by these aphids is very small and cannot be measured or collected directly, as it does not form droplets, but instead forms a moist coating over the rear of the aphid's body. The honeydew was thus collected by absorption onto filter papers and nitrocellulose membranes in contact with the aphids. The filter papers were stained with ninhydrin, visualising drops of honeydew as small purple spots (Fig. 5.4a and b), due to the presence of free amino acids derived from the sap. The device set up to collect the honeydew was thus successful. A difference in the amount of honeydew produced by aphids feeding on *RSs1*-GNA and untransformed plants was evident from the number of stained spots; less honeydew was produced by the aphids on the *RSs1*-GNA plant, indicating that GNA has an effect on aphids. The nitrocellulose membranes with absorbed honeydew were treated with

Figure 5.4. Immuno-detection of GNA protein in honeydew. (a-b) 3MM discs with absorbed honeydew produced by peach potato aphids feeding on a *RSs1*-GNA plant (a) and an untransformed plant (b), sprayed with ninhydrin. (c-d) Nitrocellulose discs with absorbed honeydew produced by aphids feeding on *RSs1*-GNA plant (c) and an untransformed plant (d), treated with anti-GNA antibody followed by HRP secondary antibody.
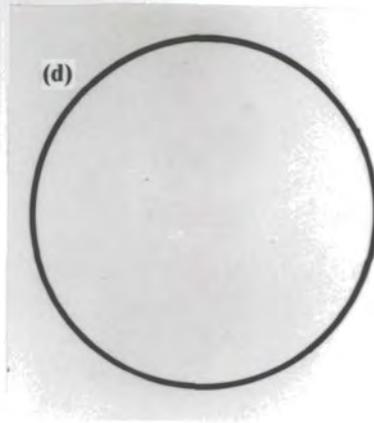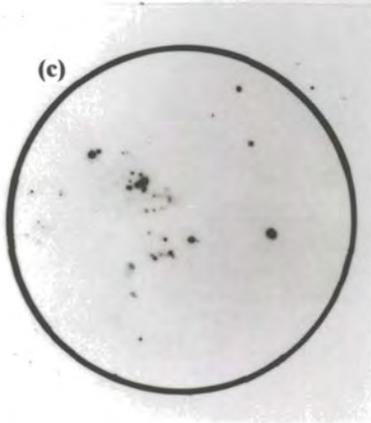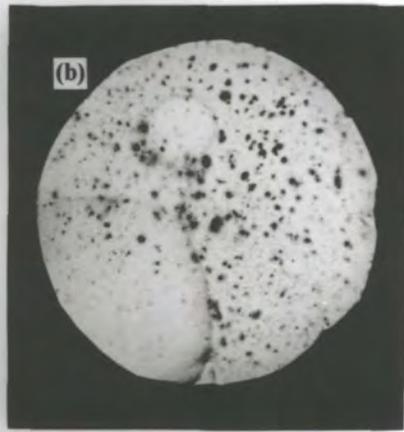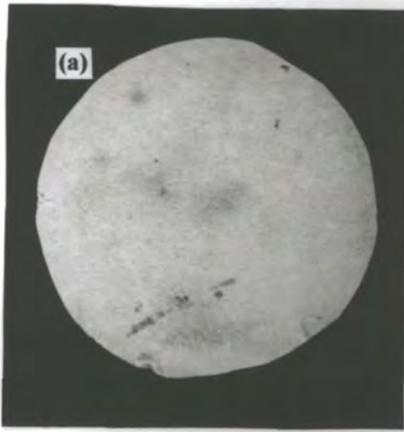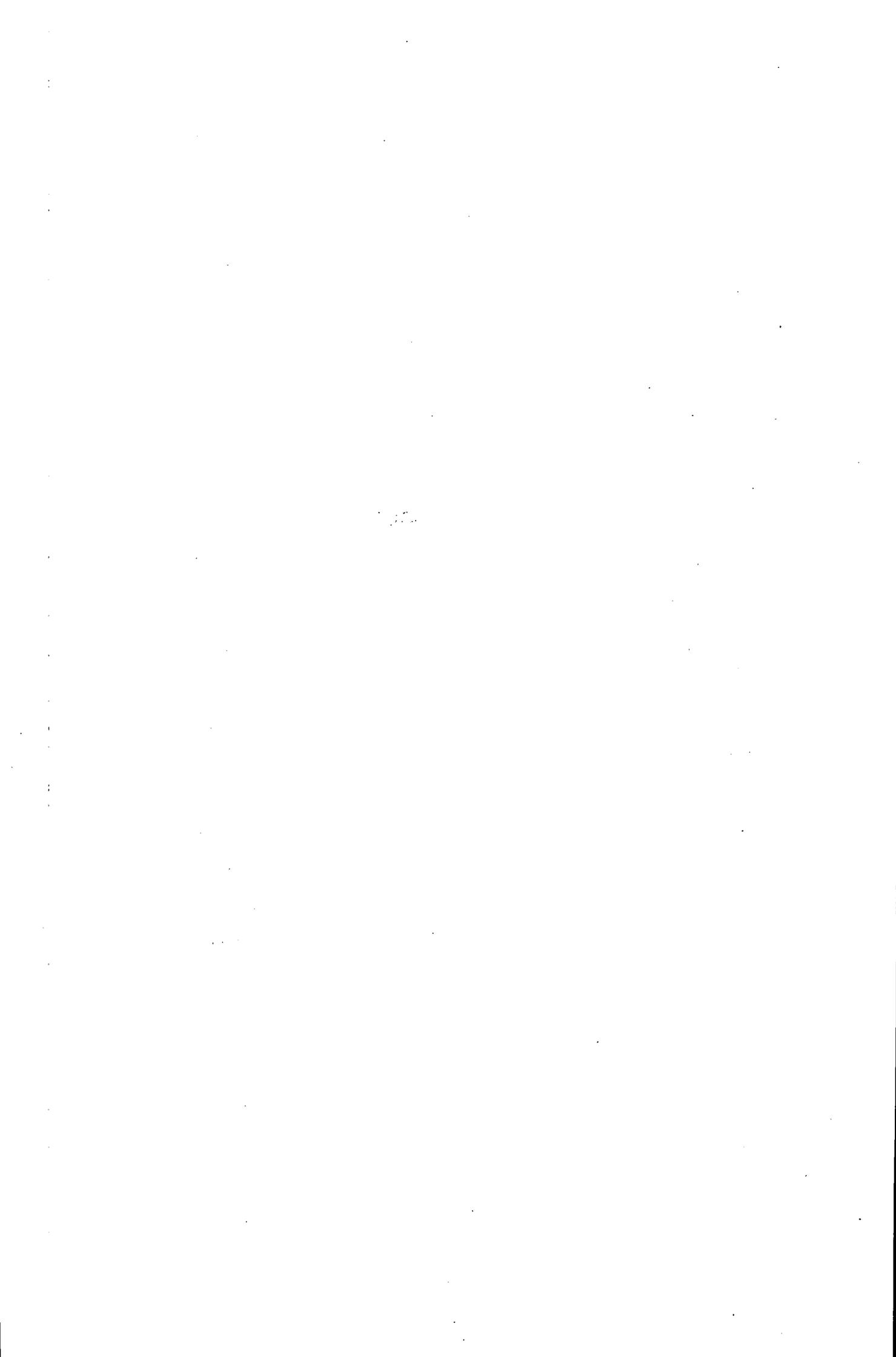
anti-GNA antibody. Honeydew produced by aphids fed on the *RSs1*-GNA transgenic plant reacted with the antibody, giving small black spots after ECL detection of bound antibody, corresponding to the drops of the honeydew, whereas honeydew from aphids fed on an untransformed plant produced no signal (Fig. 5.4c and d). Results from these experiments demonstrated that GNA was passed out in the honeydew by insects fed on plants containing the *RSs1*-GNA chimaeric gene, and provide direct evidence of the presence of GNA in phloem sap. Therefore, the *RSs1* promoter directed phloem-specific expression of GNA in such a way that the protein was present in the sap.

5.1.5. Activity of PP2 promoter

Tobacco leaf disk transformation with the PP2-GUS construct (Fig. 5.1c) produced 33 independent plants with kanamycin resistance. Seven of these plants were assayed fluorometrically for GUS expression, and only one plant showed fluorescence approximately 3 times that of the control plant (non-transformed tobacco) but about 130 times lower than that of a 35S-GUS plant. However, histochemical analysis of a young leaf and petiole of this plant showed no blue staining in phloem or in other tissues. The reason for this was undetermined, but since this plant was the only one among the examined showing phenotypic abnormality (i.e. it stopped normal growth and appeared to start flowering one week after being moved to soil) when tissues were collected for GUS assay, the fluorescence detected might be due to a higher endogenous background than in other normal plants. Petioles and leaves of another 7 plants were also analysed histochemically, but again no stain was observed.

5.2. Discussion

5.2.1. *RSs1* promoter

The results described above demonstrate that the *RSs1* promoter is active and phloem-cell specific, and that the expression pattern of GUS and GNA directed by the *RSs1* promoter in transgenic tobacco plants was similar to the expression of GUS directed

by the maize *Sh1* promoter (Yang and Russell, 1990) and the *Arabidopsis* sucrose

synthase gene *Asus1* promoter (Martin *et al.* 1993). At high magnification, Yang and

Russell (1990) observed that GUS was only expressed in sieve elements and

companion cells in transgenic tobacco. By *in situ* immunolocalisation the GNA protein

was shown to be present in sieve elements and companion cells, demonstrating that the

protein accumulates in these cells, and is not degraded so rapidly that its steady-state

level is undetectable. The expression pattern of the sucrose synthase promoters in

transgenic plants is consistent with the widespread observations that sucrose synthase

is associated with vascular bundles or phloem tissues (Claussen *et al.* 1985b;

Tomlinson *et al.* 1991; Nolte and Koch 1993b; Geigenberger *et al.* 1993; Wang *et al.*

1994), supporting the hypotheses for the physiological significance of this enzyme in

phloem (Nolte and Koch 1993b; Geigenberger *et al.* 1993). The phloem-specific

expression of all the three sucrose synthase promoters examined seems to suggest that

the phloem specificity is an intrinsic property of the sucrose synthase genes.

The constructs used in this work contained the first untranslated intron of the

RSs1 gene, which, by analogy to its maize homologue, would be expected to enhance

expression of the gene in monocots (Vasil *et al.* 1989; Maas *et al.* 1991). However,

this intron has a lower A/T content than is typical for dicot introns, and may therefore

be spliced inefficiently in transgenic tobacco (Goodall and Filipowicz, 1991). If this is

the case, the expression detected from the *RSs1* promoter in this work may not

represent its maximum potential in a dicot system. The combination of exon 1 and

intron 1 of the maize sucrose synthase gene *Sh1* (as in the present *RSs1* promoter

constructs) had little effect on gene expression in tobacco protoplasts, and, without a

full-length exon 1, the *Sh1* intron 1 reduced expression by more than 10-fold (Maas *et*

al., 1991). The first intron of *RSs1* contains a number of start codons, but no open

reading frame extending outside the intron; there is thus no possibility of protein

products other than that predicted being detected. The observed expression of both

enzymatically active GUS and immunologically reactive GNA suggests that correct

intron processing, to at least some extent, does occur. The *Arabidopsis Asus1*

promoter does not contain introns, and yet like the *RSs1* and *Sh1* promoters it directs

phloem-specific expression of GUS in both transgenic tobacco and transgenic

Arabidopsis plants (Martin *et al.* 1993). This suggests that the *cis*-element(s)

conferring the phloem specificity of the *RSs1* and *Sh1* promoters is (are) not located in

the first introns.

The detection of GNA in honeydew produced by aphids feeding on an *RSs1*-

GNA transgenic plant provides direct evidence that GNA is not only expressed in

phloem tissue, but more importantly, is also present in phloem sap. Furthermore, it

suggests that some GNA, at least, remains relatively intact after passing through the

insect gut. In a similar experiment, sucrose synthase was detected in the honeydew

excreted by the brown planthopper feeding on rice plants (Fig. 5.5b). The presence of

sucrose synthase in phloem sap was also demonstrated by Western blotting which

detected a specific high-molecular-weight protein band in the phloem exudate of *C.*

pepo using maize SS antibodies (Fig. 5.5a). Because mature sieve elements lack both

a nucleus and ribosomes (Cronshaw 1981), they are incapable of protein synthesis.

The presence of GNA in phloem sap may thus require the transport of the protein from

adjacent nucleate cells, i.e. the companion cells. Plasmodesmatal transport of proteins

between sieve elements and companion cells has been suggested by Fisher *et al.* (1992)

after clearly demonstrating the presence of newly synthesised proteins in phloem

exudates of wheat. A similar hypothesis was made by Bostwick *et al.* when they found

that mRNA encoding the phloem protein PP2 was localised in the companion cells of

pumpkin hypocotyls. Fisher *et al.* (1992) observed that the sieve tube-companion cell

movement *via* plasmodesmata appeared to be protein specific and not based simply on

passive molecular sieving. If this is the case, some sort of recognition factor, e.g. a

specific protein domain, must be involved. The translational fusion made in

constructing the *RSs1*-GUS and *RSs1*-GNA chimaeric genes is unlikely to add new

targeting information to the GNA protein, since although thirteen extra amino acids

were attached at the N-termini of the proteins, only six of these come from sucrose

synthase. Whether the GNA polypetide contains its own targeting information remains
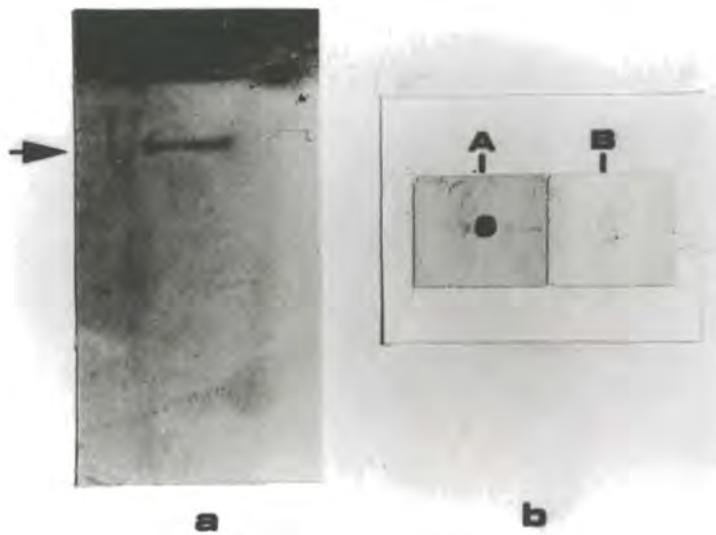
Figure 5.5. Immunological detection of sucrose synthase protein using maize sucrose synthase antibody. (a) Phloem exudate (36 µl) was collected from stems of *Cucurbita pepo*, separated by non-denaturing PAGE, and blotted to a nitrocellulose membrane. The arrow indicates the antibody-binding band. (b) Honeydew (4 µl on each dot) excreted by brown planthopper feeding on rice (*Oryza sativa* L cv. Taichung Native 1) was blotted to a nitrocellulose membrane. A, treated with both sucrose synthase antibody and HRP secondary antibody; B, treated only with HRP secondary antibody.

to be elucidated. Nonetheless, since the GNA polypeptide is relatively small (~15 kD), a passive flow through plasmodesmata is not unlikely. In Fisher *et al*.'s report (1992), the majority of the exudate proteins with molecular size smaller than 36 kD were labelled with [^{35}S]-methionine, indicating that a large number of newly synthesised proteins can pass through the intercellular barrier between companion cells and sieve elements. The GNA protein detected in phloem sap may alternatively be pre-existing proteins originally synthesised in young, developing sieve elements. However, long-lasting presence of soluble proteins in the translocation stream of mature sieve elements without continual supplementation seems less likely, and, such proteins, like the P-proteins (which do not appear to be mobile or soluble components in the translocation stream [Cronshaw 1975]), would need to be highly stable to survive the degenerating process accompanying the maturation of sieve elements. Furthermore, the ongoing involvement of sucrose synthase in phloem translocation requires the continual presence of active enzymes.

The GNA concentration in the phloem sap of transgenic plants was not determined due to the difficulty in collecting phloem exudate from tobacco plants. From the fact that protein concentrations in exudates from non-cucurbit species are generally low (approximately 0.3-2.2 mg/ml) (Hall and Baker 1972; Esau and Cronshaw 1967; Kennecke *et al*. 1971; Fisher *et al*. 1992) it would be expected that the concentration of a specific protein, such as the sucrose synthase, will not be high. However, it is possible that a phloem function-associated protein in its active form is bound to, or sequestered in, certain subcellular sites of sieve elements or adjacent cells, which is thus separated from the translocation stream, and the protein detected in the exudate is an inactive form replaced by newly synthesised ones. If this is the case, a foreign protein unrelated to a phloem function may enter the translocation stream immediately after being synthesised, making it relatively abundant in the phloem sap. Since phloem exudate is shown to be basically free of proteolytic enzymes (Fisher *et al*. 1992), a relatively high steady-state level of a foreign protein can thus be attained. However, the concentration and composition of proteins in phloem exudate remain

basically constant with time (Becker *et al.* 1971; Fisher *et al.* 1992) despite their continual synthesis, suggesting that protein turn-over occurs in the sieve tube. As suggested by Fisher *et al.* (1992), these proteins are continually removed from the soluble protein pool in the sieve element *via* plasmodesmata. If this process is non-selective for different proteins, the concentration of a specific protein can only reach a certain level in the phloem sap, which will be dependent on the rate of its synthesis in the companion cells. If, instead, selective removal or proteolysis is involved in regulating the concentrations of these soluble proteins, possibly a way can then be found to increase the steady-state level of a foreign protein in phloem sap.

In order to control homopteran insects such as BPH using transgenic plants, an insecticidal gene must first be expressed in phloem tissue, and then the gene product must be transferred to, or be present in, the phloem sap to ensure some effect on the insects. The results described above suggest that the *RSs1* promoter should prove useful in satisfying the first of these requirements, although some modification of the promoter may be necessary to optimise expression in a target plant. GNA has been found to be insecticidal against BPH, so the *RSs1*-GNA construct can be directly used in transforming rice plants. The presence of GNA in the honeydew of a homopteran insect feeding on the transgenic plant clearly indicates that the insecticidal protein can be ingested by the insect, which thus satisfies the second requirement mentioned above.

5.2.2. The PP2 promoter

Although the fluorometric analysis showed above-background GUS activity in one of the seven plants examined, the histochemical assay, which detected no GUS activity in any of the plants examined, appears to suggest that the chimaeric PP2 promoter-GUS gene was not expressed in transgenic tobacco, at least at the protein level. Several possible reasons for this failure can be suggested. The PP2 gene family contains multiple copies of genes, and the genomic DNA from which the promoter was isolated may be a pseudogene with a non-functional promoter. This is possible given the fact

that the sequence of the genomic clone is not identical to that of the cDNA clones. Another possibility is that the PP2 gene is strictly species-specific and its expression highly regulated, and the tobacco plant is not suitable for its expression. The lack of P-proteins in gymnosperms and in some monocotyledonous plants such as wheat (Cronshaw 1981), and the genus-specificity of the PP2 protein in immunological cross-reaction observed by some workers (Smith *et al.* 1987; Cronshaw and Sabnis 1990), argue for this possibility. It is also possible that the GUS coding sequence is transcribed under the PP2 promoter, but due to the lack of the 10-nt 5' leader sequence in the PP2 cDNA and the addition of extra nucleotides between the promoter and the translation start ATG, the transcript can not be recognised by protein translation factors for such a phloem-specific translation as that of an inherent PP2 gene. The P-proteins are stable in the sieve element, possibly due to the formation of cross-linked, insoluble products immediately after their synthesis or due to specific subcellular localisation of the P-protein bodies. Sham and Northcote (1989) observed that as the level of PP2-mRNA synthesis was decreased at later stage of *C. maxima* seedling growth, the proportion of PP2 in the total amount of protein synthesised remains constant. This suggests that there is possibly little breakdown of the PP2 already synthesised and present in the sieve elements. If this is the case, a foreign protein without such a high stability as PP2 may not reach a detectable level even if it is synthesised in the phloem. Although PP2 mRNA has been shown to be localised in the companion cells of young cucurbit hypocotyls (Bostwick *et al.* 1992) and has been detected in other plant organs as well (Dannenhoffer *et al.* 1993), the subcellular site of its translation may be different from that of other phloem-specific genes like *RSs1* and its gene product may be immediately exposed to the degeneration process of the sieve element after synthesis. Substantial expression of a foreign gene may therefore be obtained by translationally fusing its coding sequence downstream with a PP2 coding sequence, since such a transcript will retain the 5' leader sequence of the PP2 mRNA and the resulting protein can be possibly anchored to the P-protein bodies *via* the PP2 polypeptide section avoiding its proteolytic degradation. The PP2 polypeptide

141

may contain certain type of recognition signals for subcellular localisation or for transport of the protein into phloem sap after its synthesis.

CHAPTER 6.   GENERAL DISCUSSION AND
FUTURE PROSPECTS

This Rockefeller Rice Biotechnology Program based at the University of Durham has as its aim the production of transgenic rice with enhanced resistance to the rice brown planthopper (BPH), a phloem sap-sucking insect. Towards this aim, two promoters, of rice *RSs1* and a cucurbit PP2 gene, have been isolated and tested in transgenic tobacco plants. The *RSs1* promoter directs phloem-specific expression of both a reporter gene (β-glucuronidase, GUS) and a BPH resistance gene (*Galanthus nivalis* agglutinin, GNA) in transgenic tobacco. Most importantly, the insecticidal protein produced under this promoter can be delivered directly into phloem sap and shown to be ingested by aphids (*Myzus persicae*), also a sap-sucking homopteran insect. In a parallel experiment carried out in the Rockefeller program in this laboratory, GNA has been identified as the most toxic protein towards BPH among several plant-derived proteins tested using an artificial diet system (Powell *et al.* 1993). The *RSs1*-GNA construct is accordingly being used to transform rice plants by collaborating groups elsewhere in the Rockefeller program.

Whether an insecticidal effect can be achieved when the GNA protein is expressed in rice is yet to be proven, but the key to this will be a steady level of GNA expression in the phloem sap of rice. As discussed in the last chapter, the *RSs1* promoter may have not reached its maximum potential in transgenic tobacco, and yet transgenic plants containing the *RSs1*-GNA construct showed some antimetabolic effect to the aphids feeding on them (i.e. the aphids feeding on *RSs1*-GNA plants excreted substantially less honeydew than those feeding on nontransformed plants, as shown in Fig. 5.4). It is reasonable to expect an equivalent or higher expression of this chimaeric gene in transgenic rice as rice is the homologous system with regard to the promoter. In Powell *et al.*'s report (1993), the GNA concentration used in the bioassay against BPH was 0.1% [w/v], but a later report indicate that as low as 0.02% [w/v] (equivalent to 0.2 mg/ml) GNA in the artificial diet can have substantial

antimetabolic effect on this insect (Powell *et al.* 1994, submitted). As already
discussed, a foreign protein expressed under the *RSs1* promoter may or may not be
able to reach high level in phloem sap, but considering the general concentration of
total exudate proteins in non-cucurbit species (0.3-2.2 mg/ml), such levels are not
unachievable. The sucrose synthase gene is expressed in rice through most of its
growth stages (Chan *et al.* 1990). Such a lasting presence of GNA in transgenic rice
will ensure persistent ingestion of this toxic protein by BPH throughout its life cycle.
The level of GNA required in transgenic plants to attain an antimetabolic effect may be
lower than that observed in bioassays, since the bioassay employs an optimal diet.

The expression level and pattern of a promoter in a transgenic homologous
plant usually reflect the intrinsic properties of its corresponding gene, as in the case of
the rice actin promoter (Zhang *et al.* 1991) and the rice light harvesting chlorophyll
a/b-binding protein gene promoter (Tada *et al.* 1991). A recent report shows that in
sugarcane, the SS1 protein is more abundant in stem than in leaf and other tissues
(Buczynski *et al.* 1993). A similar pattern of GNA expression would be desirable, as
BPH preferentially feeds and multiplies in the basal part of the plant (Khush 1977).

Total resistance to BPH is most desirable, but partial resistance may be
sufficient when used in an IPM (integrated pest management) program, utilising a
combination of different practices such as biological, cultural, physical and chemical
control measures. A major factor contributing to BPH outbreaks is the excessive use
of wide-spectrum chemical pesticides which can damage the populations of natural
predators and parasites, and can lead to the development of insecticide resistance in the
pest insect (Gallogher *et al.* 1994; Heinrichs 1994). Plants with a certain level of
inherent resistance to BPH can possibly keep the BPH field population at a
controllable level through its natural enemies while minimising the use of chemical
pesticides, which in turn can foster these natural enemies on the host plant. An
additional benefit of using such plants would be a reduction in virus transmission by
BPH, as this insect is a virus vector of two major rice dwarf viruses (Rivera *et al.*
1966; Ling 1977; Hare 1994). Reduced feeding, short feeding time and lower mobility

of the insect will certainly reduce the chance of transmitting the stunt viruses, which are transmitted in a persistent manner (i.e. circulative and usually propagative) and require a relatively long insect feeding time (Walkey 1985).

The *RSs1* promoter used here has a size of approximately 3.1 kb, including approximately 1.9 kb of the 5'-flanking region, the first exon, first intron and the translation star site of *RSs1*. Such a relatively long promoter may have adverse effects on rice transformation due to possible homologous recombination and the potential difficulty in inserting a large DNA fragment into the rice genome. More studies might therefore be needed to define the promoter and identify the *cis*-elements essential for its normal expression in order to optimise the size of the promoter used for making expression vectors.

Unfortunately, the PP2 promoter isolated in this work has not been expressed in transgenic tobacco to date. Since PP2 is highly abundant in the phloem exudate of cucurbit species, it is expected that its promoter is phloem-specific and the products of gene expression under this promoter can be delivered directly into phloem sap. Therefore, it is worthwhile to carry out more investigations on the PP2 gene in order to elucidate reasons for the failure in expressing the promoter in transgenic tobacco or to identify an active PP2 promoter, if the available one is from a pseudogene. It may be necessary to sequence the 3' non-coding region of the genomic clone CPP1.3, since this region is usually less conserved and therefore the difference between a genomic DNA and a cDNA species, or between different members of a gene family can be more readily revealed through comparison of the sequences in this region. The *C. pepo* genomic library and the PP2 cDNA and genomic clones obtained in this work can be used to isolate other members of the PP2 gene family. However, even if an active PP2 promoter can be obtained or the available one can be made to function in transgenic tobacco, other problems may occur, such as that an active PP2 promoter may fail to function properly and effectively in rice (a monocot). In this case, it may be necessary to isolate a gene homologous to that encoding PP2 from rice itself using the PP2 gene as a heterologous probe.

Besides the sucrose synthase and the PP2 promoters, several other promoters have been reported to be phloem-specific, such as the major promoter region for the transcription of rice tungro bacilliform virus (Bhattacharyya-Pakrasi *et al.* 1993), a combination of CaMV 35S enhancer subdomains (4×B2 + A) (Benfey *et al.* 1990b), and the promoter of the ORF12 gene (*rolC*) of the Ri plasmid (Sugaya *et al.* 1989; Sugaya and Uchimiya 1992). These reports lack defined histochemical results to show the cell-type specificity of the expression in phloem, but these promoters are potential alternatives to the *RSs1* or PP2 promoters. The 5' flanking regions of some other genes, such as that encoding the sucrose transporter (Riesmeier *et al.* 1993), may also be suitable candidates.

However, as mentioned earlier, the mature sieve element is incapable of protein synthesis and there may be an upper limit for the concentration of a specific protein in phloem sap regardless of the promoter used. If no promoter can express GNA at a level in the phloem sap which gives resistance to BPH, it would be necessary to identify a new BPH resistance gene with higher toxicity than GNA. Alternatively a gene pyramiding strategy can be adopted where more than one BPH resistance gene with different insecticidal mechanisms are introduced into a single plant genotype to produce an additive or synergistic effect against the insect. In any case this strategy would be desirable, since with various mechanisms, it would be more difficult for a pest to overcome several resistance genes in the host than to overcome a single resistance gene. An example of this additive effect is transgenic tobacco expressing two insecticidal genes, the cowpea trypsin inhibitor gene (CpTI) and a pea lectin gene, which exhibits higher resistance to tobacco budworm than transgenic tobacco expressing the respective single genes (Boulter *et al.* 1990).

Apart from BPH, many other insects belong to the suborder Homoptera, such as aphids, whiteflies, planthoppers and leafhoppers, all of which mainly feed on phloem sap. Crop damage is caused by these insects in a number of ways (Hilder *et al.* 1992). Extraction of sap deprives the plant of nutrients and water, leading to loss of vigour and wilting. Phytotoxic substances present in the saliva of some species, and

mechanical blockage of the phloem by feeding, may result in distortion and necrosis of foliage and in blindness or shrunken kernels in grain crops. Injury caused by insertion of the mouthparts leaves lesions through which plant pathogens may enter. Production of copious honeydew may allow sooty moulds to develop leading to reduction in leaf area for photosynthesis, and its stickiness may interfere with the harvesting of cereals and cotton. Some of the most serious damage caused by these pests is indirect, due to their role as vectors of plant viruses. The strategy used here to genetically engineer BPH-resistance, if successful, may be applied in attempts to control all of these insect pests.

Another area in which phloem-specific promoters can be useful is the control of plant virus diseases. Plant viruses cause major losses in crop production in almost every sector of agriculture. Unlike insects and other plant pathogens such as fungi and bacteria, there are presently only very limited means for the protection of plants that are susceptible to infection by a particular virus strain. Phloem tissues are the major target for virus infection. Many viruses such as those transmitted by some homopteran insects are phloem limited, and even for those non-phloem limited mosaic viruses, their long-distance systemic infection is also through the phloem (Esau 1956; Leisner and Turgeon 1993). Genetic engineering of virus resistance in plants has recently become a major research interest, and several strategies have been used, such as to engineer plants to express an anti-virus antibody gene (Tavladoraki *et al*. 1993) or to express a virus-derived gene such as a coat protein gene (Powell Abel *et al*. 1986), a satellite RNA gene (Gerlach *et al*. 1987), a RNA replicase gene (Golemboski *et al*. 1990; Donson *et al*. 1993), a defective virus movement protein gene (Malyshenko *et al*. 1993), an antisense gene (Kawchuk *et al*. 1991; Powell Abel *et al*. 1989), or a ribozyme gene (Edington *et al*. 1994). A phloem-specific promoter could be very useful in these strategies.

In conclusion, the initial objectives of this research project have largely been achieved, i.e. the isolation and characterisation of the *RSs1* and a PP2 genes, identification of their promoter sequences and determination of these promoter

activities in transgenic tobacco, and the construction of an expression vector using the *RSs1* promoter for rice transformation. The practical significance of this research, however, is yet to be proven. Previous experiences from BPH management indicate that dependence of any single approach will probably be ineffective in the long term. The future of BPH management will continue to depend to some extent on the use of conventional control methods, including biological control, host plant resistance, cultural manipulations, or even chemical control. A common consensus on the role of genetic engineering is that it will not replace these conventional approaches but bring new tools to assist insect pest control. The strategy adopted here, if proven, will provide an efficient way of creating a new type of host plant resistance for the control of not only BPH but probably also other phloem sap-sucking insect pests or plant viruses.

LITERATURE CITED

Ahrens C, Cramer HH, Mogk M, Peschel H (1982) Economic impact of crop losses. *Proc. 10th Int. Con. of Plant Prot.*, pp.65-73.

Allen AK (1979) A lectin from the exudate of the fruit of the vegetable marrow (*Cucurbita pepo*) that has specificity for β-1,4-linked N-acetylglucosamine oligosaccharides. *Biochem. J.* 183: 133-137.

Anderson JM, Larsen R, Laudencia D, Kim WT, Morrow D, Okita TW, Preiss J (1991) Molecular characterization of the gene encoding a rice endosperm-specific ADPglucose pyrophosphorylase subunit and its developmental pattern of transcription. *Gene* 97: 199-205.

Arai M, Mori H, Imaseki H (1992) Expression of the gene for sucrose synthase during growth of mung bean seedlings. *Plant Cell Physiol.* 33: 503-506.

Avigad G (1982) Sucrose and other disaccharides. In Loewus FA, Tanner W (eds) *Encyclopedia of Plant Physiology New Series, 13A*, pp.217-347. Berlin: Spring Verlag.

Bae SH, Pathak MD (1970) Life history of *Nilaparvata lugens* (Homoptera: Delphacidae) and susceptibility of rice varieties to its attack. *Ann. Entomol. Soc. Amer.* 63:149-155.

Baker RT, Tobias JW, Varshavsky A (1992) Ubiquitin-specific proteases of *Saccharomyces cerevisiae*: cloning of UBP2 and UBP3, and functional analysis of the UBP gene family. *J. Biol. Chem.* 267: 23364-23375.

Barton KA, Whiteley HR, Yang NS (1987) *Bacillus thuringiensis* δ-endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lipidopteran insects. *Plant Physiol.* 85: 1103-1109.

Batra SK, Metzgar RS, Hollingsworth MA (1991) Molecular cloning and sequence analysis of the human ribosomal protein S16. *J. Biol. Chem.* 266: 6830-6833.

Battraw MJ, Hall TC (1990) Histochemical analysis of CaMV 35S promoter-β-glucuronidase gene expression in transgenic rice plants. *Plant Mol. Biol.* 15: 527-538.

Becker D, Kluge M, Ziegler H (1971) The incorporation of $^{32}PO_4^{3-}$ into organic compounds in sieve tube sap. *Planta* 99: 154-162.

Behnke HD (1974) Comparative ultrastructural investigations of angiosperm sieve elements: aspects of the origin and early development of P-protein. *Z. Pflanzenphysiol.* 74: 22-34.

Beintema JJ, Peumans WJ (1992) The primary structure of stinging nettle (*Urtica dioica*) agglutinin: a two-domain member of the hevein family. *FEBS Lett.* 299: 131-134.

Benfey PN, Ren L, Chua NH (1989) The CaMV 35S enhancer contains at least two domains which confer different developmental and tissue-specific expression patterns. *EMBO J.* 8: 2195-2202.

Benfey PN, Ren L, Chua NH (1990a) Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* 9: 1677-1684.

Benfey PN, Ren L, Chua NH (1990b) Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. *EMBO J.* 9: 1685-1696.

Bennett MD, Smith JB, Heslop-Harrison JS (1982) Nuclear DNA amounts in angiosperms. *Proc. R. Soc. Lond.* B 216:179-199.

Benrey B, Lamp WO (1994) Biological control in the management of planthopper populations. In Denno RF, Perfect TJ (eds) *Planthoppers, Their Ecology and Management*, pp. 519-549. Chapman & Hall, New York, USA.

Beyenbach J, Weber C, Kleinig H (1974) Sieve-tube proteins from *Cucurbita maxima*. *Planta* 119: 113-124.

Bhattacharyya-Pakrasi M, Peng J, Elmer JS, Laco G, Shen P, Kaniewska MB, Kononowicz H, Wen F, Hodges TK, Beachy RN (1993) Specificity of a promoter from the rice tungro bacilliform virus for expression in phloem tissues. *Plant J.* 4: 71-79.

Bonham-Smith PC, Oancia TL, Moloney MM (1992) Cytoplasmic ribosomal protein S15a from *Brassica napus*: molecular cloning and developmental expression in mitotically active tissues. *Plant Mol. Biol.* 18: 909-919.

Bonman JM, Khush GS, Nelson RJ (1992) Breeding rice for resistance to pests. *Annu. Rev. Phytopathol.* 30: 507-528.

Bostwick DE, Dannenhoffer JM, Skaggs MI, Lister RM, Larkins BA, Thompson GA (1992) Pumpkin phloem lectin genes are specifically expressed in companion cells. *The Plant Cell* 4: 1539-1548.

Bostwick DE, Skaggs MI, Thompson GA (1993) Genomic organization of phloem lectin genes. *Plant Physiol.* 102 (Supplement Issue): 73. Meeting abstract.

Bostwick DE, Thompson GA (1993) Nucleotide sequence of a pumpkin phloem lectin cDNA. *Plant Physiol.* 102: 693-694.

Boulter D (1993) Insect pest control by copying nature using genetically engineered crops. *Phytochemistry* 34: 1453-1466.

Boulter D, Gatehouse JA, Gatehouse AMR, Hilder VA (1990) Genetic engineering of plants for insect resistance. *Endeavour-New Series* 4: 185-190.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* 72: 248-254.

Breathnach R, Chambon P (1981) Organization and expression of eukaryotic split genes encoding for proteins. *Ann. Rev. Biochem.* 50: 349-383.

Brown JWS (1986) A catalogue of splice junction and putative branch point sequences from plant introns. *Nucl. Acids Res.* 14: 9545-9559.

Brown JWS, Wandelt C, Feix G, Neuhaus G, Schweiger HG (1986) The upstream region of zein genes: sequence analysis and expression in the unicellular green-alga *Aacetabularla*. *Eur. J. Cell. Biol.* 42: 161-170.

Buczynski SR, Chourey TP, Maretzki A (1993) Tissue distribution and characterization of sucrose synthase isozymes in sugarcane. *J. Plant. Physiol.* 142: 641-646.

Callis J, Fromm M, Walbot V (1987) Introns increase gene expression in cultured maize cells. *Genes Develop.* 1: 1183-1200.

Cardini CE, Leloir LF, Chriboga J (1955) The biosynthesis of sucrose. *J. Biol. Chem.* 214: 149-155.

Chan HY, Ling TY, Juang RH, Ting IN, Sung HY, Su JC (1990) Sucrose synthase in rice plants: growth-associated changes in tissue specific distributions. *Plant Physiol.* 94: 1456-1461.

Chang TT (1976) The origin, evolution, cultivation, dissemination, and deversification of Asian and African rice. *Euphytica* 25: 425-441.

Chen YC, Chourey PS (1989) Spatial and temporal expression of the two sucrose synthase genes in maize: immunohistological evidence. *Theor. Appl. Genet.* 78: 553-559.

Chiu SC (1979) Biological control of the brown planthopper. In *Brown Planthopper - Threat to Rice Production in Asia*, pp. 335-355. International Rice Research Institute, Los Banos, Philippines.

Chopra S, Delfavero J, Dolferus R, Jacobs M (1992) Sucrose synthase of *Arabidopsis*: genomic cloning and sequence characterization. *Plant Mol. Biol.* 18: 131-134.

Chourey PS (1981) Genetic control of sucrose synthetase in maize endosperm. *Mol. Gen. Genet.* 184: 372-376.

Chourey PS, Chen YC, Miller ME (1991a) Early cell degeneration in developing endosperm is unique to shrunken mutation in maize. *Maydica* 36: 141-146.

Chourey PS, Latham MD, Still PE (1986) Expression of two sucrose synthetase genes in endosperm and seedling cells of maize: evidence of tissue specific polymeriztion of promoters. *Mol. Gen. Genet.* 203: 251-255.

Chourey PS, Nelson OE (1976) The enzymatic deficiency conditioned by the shrunken 1 mutation in maize. *Biochem. Genet.* 14: 1041-1055.

Chourey PS, Taliercio EW, Kane EJ (1991b) Tissue-specific expression and anaerobically induced posttranscriptional modulation of sucrose synthase genes in *Sorghum bicolor* M. *Plant Physiol.* 96: 485-490.

Chung CT, Niemela SL, Miller RH (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* 86: 2172-2175.

Claridge MF (1990) Variation in pest and natural enemy populations: relevance to brown planthopper control strategies. In Grayson BT, Green MB, Copping LG (eds) *Pest Management in Rice*, pp.143-154. University Press, Cambridge.

Claridge MF, Den Hollander J (1980) The biotypes of the rice brown planthopper, *Nilaparvata lugens*. *Ent. Exp. Appl.* 27: 23-30.

Claussen W, Hawker JS, Loveys BR (1985a) Sucrose synthase activity, invertase activity, net photosynthetic rates and carbohydrate content of detached leaves of eggplants as affected by attached stems and shoots (sinks). *J. Plant Physiol.* 119: 123-131.

Claussen W, Loveys BR, Hawker JS (1985b) Comparative investigations on the distribution of sucrose synthase activity and invertase activity within growing, mature and old leaves of some C3 and C4 plant species. *Physiol. Plant* 65: 275-280.

Cramer HH (1967) Plant protection and world crop protection. *Pflanzenschutz-Nachrichten* 20: 1-524.

Crespi MD, Zabaleta EJ, Pontis HG, Salerno GL (1991) Sucrose synthase expression during cold acclimation in wheat. *Plant Physiol.* 96: 887-891.

Cronshaw J (1975) P-proteins. In Aronoff S, Dainty J, Gorham PR, Srivastava LM, Swanson CA (eds) *Phloem Transport*, pp.79-115. Plenum, London.

Cronshaw J (1980) Histochemical localization of enzymes in the phloem. *Ber. Dtsch. Bot. Ges.* 93: 123-139.

Cronshaw J (1981) Phloem structure and function. *Ann. Rev. Plant Physiol.* 32: 465-484.

Cronshaw J, Esau K (1968a) P protein in the phloem of *Cucurbita*. 1. The development of P-protein bodies. *J. Cell Biol.* 38: 25-39.

Cronshaw J, Esau K (1968b) P protein in the phloem of *Cucurbita*. 2. The P protein of mature sieve elements. *J. Cell Biol.* 38: 292-303.

Cronshaw J, Esau K (1967) Tubular and fibrillar components of mature and differentiating sieve elements. *J. Cell. Biol.* 34: 801-816.

Cronshaw J, Sabnis DD (1990) Phloem proteins. In Behnke HD, Sjolund RD (eds) *Sieve Elements. Comparative Structure, Induction, and Development*, pp. 257-283. Springer-Verlag.

Dannenhoffer JM, Bostwick DE, Thompson GA (1993) Developmental expression of phloem lectin genes. *Plant Physiol.* 102 (Supplement Issue): 124. Meeting abstract.

David CC (1991) The world rice economy: Challenges ahead. In Khush GS and Toenniessen (eds) *Rice Biotechnology*, pp.1. CAB Int., UK, in association with IRRI.

de Ilarduya OM, Vicente-Carbajosa J, Sanchez de la Hoz P, Carbonero P (1993) Sucrose synthase genes in barley: cDNA cloning of the Ss2 type and tissue-specific expression of Ss1 and Ss2. *FEBS Letters* 320: 177-181.

Dekeyser R, Claes B, Marichal M, Van Montagu, Caplan A (1989) Evaluation of selectable markers for rice transformation. *Plant Physiol.* 90: 217-223.

Delmer DP (1972) The purification and properties of sucrose synthetase from etiolated *Phaseolus aureus* seedlings. *J. Biol. Chem.* 247: 3822-3828.

de Vries S, Hoge H, Bisseling T (1988) Isolation of total and polysomal RNA from plant tissues. In Gelvin SB, Schilperoot RA, Verma DPS (eds) *Plant Molecular Biology Manual*, pp.B6/1-B6/13. Kluwer Academic Publishers, Dordrecht.

DiDomenico B, Lupisella JA, Sandbaken MG, Chakbraburtty K (1992) Isolation and sequence analysis of translation elongation factor 3 from *Candida albicans*. *Yeast* 8: 337-352.

Donson J, Kearney CM, Turpen TH, Khan IA, Kurath G, Turpen AM, Jones GE, Dawson WO, Lewandowski DJ (1993) Broad resistance to tobamoviruses is mediated by a modified tobacco mosaic-virus replicase transgene. *Mol. Plant-Microbe Interact.* 6: 635-642.

Draper J, Scott R (1988) The isolation of plant nucleic acids. In Draper J, Scott R, Armitage P, Walden R (eds) *Plant Genetic Transformation and Gene Expression - A Laboratory Manual*, pp.199-236. The Alden Press, Oxford.

Dudov KP, Perry RP (1984) The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* 37: 457-468.

Dyck VA, Thomas B (1979) The brown planthopper problem. In *Brown Planthopper - Threat to Rice Production in Asia*, pp. 3-17. International Rice Research Institute, Los Banos, Philippines.

Echt CS, Chourey PS (1985) A comparison of 2 sucrose synthetase isozymes from normal and *shrunken-1* maize. *Plant Physiol.* 79: 530-536.

Edington BV, Nelson RS, Dixon RA, Draper KG, Yang XJ (1994) Ribozyme cleavage of tobacco mosaic-virus sequences and inhibition of TMV replication in tobacco protoplasts. *J. Cell. Biochem.* S18A: 109. Meeting abstract.

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

Elling L, Kula MR (1993) Purification of sucrose synthase from rice and its protein-chemical characterisation. *J. Biotechnol.* 29: 277-286.

Ellis JG, Llewellyn DJ, Dennis ES, Peacock WJ (1987) Maize *Adh1* promoter sequences control anaerobic regulation: addition of upstream promoter elements from constitutive genes is necessary for expression in tobacco. *EMBO J.* 6: 11-16.

Esau K (1956) An anatomist's view of virus diseases. *Am. J. Bot.* 43: 739-748.

Esau K, Charvat ID (1975) An ultrastructural study of acid phosphatase localization in cells of *Phaseolus vulgaris* phloem by the use of azo dye method. *Tissue Cell* 7: 619-630.

Esau K, Cheadle VI (1965) Cytologic studies on phloem. *Univ. Cal. Publ. Bot.* 36: 253-344.

Esau K, Cronshaw J (1967) Tubular components in cells of healthy and tobacco mosaic virus-infected *Nicotiana*. *Virology* 33: 26-35.

Fang RX, Nagy F, Sivasubramaniam S, Chua NH (1989) Multiple *cis* regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *Plant Cell* 1: 141-150.

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132: 6-13.

Feinberg AP, Vogelstein B (1984) Addendum for a technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 137: 266-267.

Fischhoff DA, Bowdish KS, Perlak FJ, Marrone PG, McCormick SM, Niedermeyer JG, Dean DA, Kusano-Kretzmer K, Mayer EJ, Rochester DE, Rogers SG, Fraley RT (1987) Insect tolerant transgenic tomato plants. *Bio/Technology* 5: 807-813.

Fisher DB, Wu Y, Ku MSB (1992) Turnover of soluble proteins in the wheat sieve tube. *Plant Physiol.* 100: 1433-1441.

Forde BG, Heyworth A, Pywell J, Kreis M (1985) Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. *Nucl. Acids Res.* 13: 7327-7339.

Frommer WB, Starlinger P (1988) DNAse I hypersensitive sites in the 5'-region of the maize *Shrunken*1 gene in nuclei from different organs. *Mol. Gen. Genet.* 212: 351-359.

Fujimoto H, Itoh K, Yamamoto M, Kyozuka J, Shimamoto K (1993) Insect resistant rice generated by introduction of a modified β-endotoxin gene of *Bacillus thuringiensis*. *Bio/Technology* 11: 1151-1155.

Gallagher KD, Kenmore PE, Sogawa K (1994) Judicial use of insecticides deters planthopper outbreaks and extends the life of resistant varieties in Southeast Asian rice. In Denno RF, Perfect TJ (eds) *Planthoppers, Their Ecology and Management*, pp. 599-613. Chapman & Hall, New York, USA.

Gao J, Kim SR, Lee JM, An G (1993) Nucleotide and protein sequences of 60S ribosomal protein L17 from tobacco (*Nicotiana tabacum* L.). *Plant Physiol.* 103: 1027-1028.

Gatehouse AMR, Hilder VA, Powell KS, Boulter D, Gatehouse JA (1992) Potential of plat-derived genes in the genetic manipulation of crops for insect resistance. In Menken SBJ, Visser JH, Harrewijn (eds) *Proc. 8th Int. Symp. Insect-Plant Relationships*, pp. 221-234. Kluwer Academic Publisher, Dordrecht.

Gatehouse AMR, Shi Y, Powell KS, Brough C, Hilder VA, Hamilton WDO, Newell CA, Merryweather A, Boulter D, Gatehouse JA (1993) Approaches to insect resistance using trangenic plants. *Phil. Trans. R. Soc. Lond.* B 342: 279-286.

Geiduschek EP, Tocchini-Valentini GP (1988) Transcription by RNA polymerase III. *Annu. Rev. Biochem.* 57: 873-914.

Geigenberger P, Langenberger S, Wilke I, Heineke D, Heldt HW, Stitt M (1993) Sucrose is metabolised by sucrose synthase and glycolysis within the phloem complex of *Ricinus communis* L. seedlings. *Planta* 190: 446-453.

Geigenberger P, Stitt M (1993) Sucrose synthase catalyses a readily reversible reaction *in vivo* in developing potato tubers and other plant tissues. *Planta* 189: 329-339.

Gerlach WL, Llewellyn D, Haseloff J (1987) Construction of a plant disease resistance gene from the satallite RNA of tobacco ringspot virus. *Nature* 328: 802-805.

Gietl C, Kauss H, Ziegler H (1979) Affinity chromatography of a lectin from *Robinia pseudoacacia* L. and demonstration of lectins in sieve-tube sap from other tree species. *Planta* 144: 367-371.

Gilder J, Cronshaw J (1974) A biochemical and cytochemical study of adenosine triphosphatase activity in the phloem of *Nicotiana tabacum*. *J. Cell Biol.* 60: 221-235.

Golemboski DB, Lomonossoff GP, Zaitlin M (1990) Plant transformed with tobacco mosaic virus nonstructural gene sequence are resistant to the virus. *Proc. Natl. Acad. Sci. USA* 87: 6311-6315.

Goodall GJ. and Filipowicz W (1991) Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *EMBO J.* 10: 2635-2644.

Gordon AJ, Ougham HJ, James CL (1993) Changes in levels of gene transcripts and their corresponding proteins in nodules of soybean plants subjected to dark induced stress. *J. Exp. Bot.* 44: 1453-1460.

Gordon AJ, Thomas BJ, Reynolds PHS (1992) Localization of sucrose synthase in soybean root-nodules. *New Phytologist* 122: 35-44.

Graham JS, Hall G, Pearce G, Ryan CA (1986) Regulation of synthesis of proteinase inhibitors I and II mRNAs in leaves of wounded tomato plants. *Planta* 169: 399-405.

Green TR, Ryan CA (1972) Wound-induced proteinase inhibitor in plant leaves: a possible defence mechanism against insects. *Science* 175: 776-777.

Habibi J, Backus EA, Czapla TH (1993) Plant lectins affect survival of the potato leafhopper (Homoptera: Cicadellidae). *J. Econ. Entomol.* 86: 945-951.

Hall SM, Baker DA (1972) The chemical composition of *Ricinus* phloem exudate. *Planta* 106: 131-140.

Hall TC, Xu Y, Huntley CC, Yu H, Seay J, Connell JC, Lepetit M, Dong J, Wallage D, Way MO, Buchholz WG (1993) Transgene expression and agronomic improvement of rice. *Phil. Trans. R. Soc. Lond.* B 342: 197-201.

Hare JD (1994) Status and prospects for an integrated approach to the control of rice planthoppers. In Denno RF, Perfect TJ (eds) *Planthoppers, Their Ecology and Management*, pp. 615-632. Chapman & Hall, New York, USA.

Hartig T (1854) Uber die Quescheidewande zwischen den einzelnen Gliedern der Siebrohren in *Cucurbita pepo. Bot. Z.* 12: 51-54.

Heim U, Weber H, Bäumlein H, Wobus U (1993) A sucrose-synthase gene of *Vicia faba* L.: Expression pattern in developing seeds in relation to starch synthesis and metabolic regulation. *Planta* 191: 394-401.

Heinlein M, Starlinger P (1989) Tissue- and cell-specific expression of the two sucrose synthase isoenzymes in developing maize kernels. *Mol. Gen. Genet.* 215: 441-446.

Heinrichs EA (1979) Chemical control of the brown planthopper. In *Brown Planthopper - Threat to Rice Production in Asia*, pp.145-167. International Rice Research Institute, Los Banos, Philippines.

Heinrichs EA (1994) Impact of insecticides on the resistance and resurgence of rice planthoppers. In Denno RF, Perfect TJ (eds) *Planthoppers, Their Ecology and Management*, pp. 571-597. Chapman & Hall, New York, USA.

Heinrichs EA, Mochida O (1984) From secondary to major pest status: the case of insecticide-induced rice brown planthopper, *Nilaparvata lugens*, resurgence. *Prot. Ecol.* 7: 201-218.

Hendrix DL (1990) Carbohydrates and carbohydrate enzymes in developing cotton ovules. *Physiol. Plant.* 78: 85-92.

Hensgens LAM, de Bakker EPHM, van Os-Ruygrok EP, Rueb S, van de Mark F, van der Maas HM, van der Veen S, Kooman-Gersman M, Hart L, Schilperoot RA (1993) Transient and stable expression of *gusA* fusions with rice genes in rice, barley and perennial grass. *Plant Mol. Biol.* 22: 1101-1127.

Hilder VA, Brough C, Gatehouse AMR, Gatehouse LN, Powell KS, Shi Y (1992) Genes for protecting transgenic crops from chewing and sap-sucking insect pests. *Brit. Crop Prot. Conf.*, pp.731-740.

Hilder VA, Gatehouse AMR, Sheerman SE, Barker RF, Boulter D (1987) A novel mechanism of insect resistance engineered into tobacco. *Nature* 300: 160-163.

Ho LC (1988) Metabolism and compartmentation of imported sugars in sink organs in relation to sink strength. *Annu. Rev. Plant Physiol.* 39: 355-378.

Horsch RB (1993) Commercialization of genetically engineered crops. *Phil. Trans. R. Soc. Lond.* B: 287-291.

Hunkapiller MW, Lujan E, Ostrander F, Hood LE (1983) Isolation of microgram quantities of protein from polyacrylamide gels for amino acid sequence analysis. In Hirs CHW and Timasheff SN (eds) *Meth. Enzymol.* Vol. 91, pp.227-237. Academic Press, London.

IRRI (International Rice Research Institute) (1994) *Rice facts.* Los Baños, Philippines.

Jefferson RA, Kavanagh TA, Beven MW (1987) Gus fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6: 3901-3907.

Joanin P, Gigot C, Philipps G (1993) cDNA nucleotide sequence and expression of a maize cytoplasmic ribosomal protein S13 gene. *Plant Mol. Biol.* 21: 701-704.

Johnson R, Lee JS, Ryan CA (1990) Regulation of a wound-inducible tomato inhibitor I gene in transgenic nightshade plants. *Plant Mol. Biol.* 14: 349-356.

Johnson R, Marraez J, An G, Ryan C (1989) Expression of proteinase inhibitor I and II in transgenic tobacco plants: effects on natural defence against *Manduca sexta* larvae. *Proc. Natl. Acad. Sci. USA* 86: 9871-9875.

Kareiva P (1993) Transgenic plants on trial. *Nature* 363: 580-581.

Karrer EE, Rodriguez RL (1992) Metabolic regulation of rice α-amylase and sucrose synthase genes *in planta. The Plant J.* 2: 517-523.

Kauss H, Ziegler H (1974) Carbohydrate-binding proteins from the sieve-tube sap of *Robinia pseudoacacia. Planta* 121: 197-200.

Kawchuk LM, Martin RR, McPherson J (1991) Sense and antisense RNA-mediated resistance to potato leafroll virus in Russet Burbank potato plants. *Mol. Plant-Microbe Interact.* 4: 227-253.

Kay R, Chan A, Daly M, McPherson J (1987) Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science* 236: 1299-1302.

Keil M, Sanchez-Serrano JJ, Willmitzer L (1989) Both wound-inducible and tuber-specific expression are mediated by the promoter of a single member of the potato proteinase inhibitor-II gene family. *EMBO J.* 8: 1323-1330.

Keith B, Chua NH (1986) Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. *EMBO J.* 5: 2419-2425.

Kenmore PE, Carino FO, Perez CA, Dyck VA, Gutierrez AP (1984) Population regulation of the rice brown planthopper (*Nilaparvata lugens* Stål) within rice fields in the Philippines. *J. Plant Prot. Trop.* 1: 19-37.

Kennecke M, Ziegler H, de Fekete MAR (1971) Enzyme activities in the sieve sap of *Robinia pseudoacacia* L. and of other tree species. *Planta* 98: 330-356.

Khush GS (1977) Disease and insect resistance in rice. *Adv. Agron.* 29: 265-341.

Kimmins FM, Woodhead S, Cook AG (1987) Resistance mechanisms in rice to the brown planthopper, *Nilaparvata lugens* (Stål). In Labeyrie V, Fabres G, Lachaise D (eds) *Insect - Plants*, pp.283-288. Dr W. Funk Publishers, Dordrecht.

Kleinig H, Dörr I, Weber C, Kollman R (1971) Filamentous proteins from plant sieve tubes. *Nature New Biol.* 229: 152-153.

Kleinig H, Thoenes J, Dörr I, Kollmann R (1975) Filament formation *in vitro* of a sieve tube protein from *Cucurbita maxima* and *Cucurbita pepo*. *Planta* 127: 163-170.

Koch KE, Nolte KD, Duke ER, McCarty DR, Avigne WT (1992) Sugar levels modulate differential expression of maize sucrose synthase genes. *Plant Cell* 4: 59-69.

Kollmann R, Dörr I, Kleinig H (1970) Protein filaments - structural components of the phloem exudate. I. Observation with *Cucurbita* and *Nicotiana*. *Planta* 95: 86-94.

Kulikova AL (1992) Some properties of phloem proteins from pumpkin exudate. *Soviet Plant Physiol.* 39: 734-739.

Kuster H, Fruhling M, Perlick AM, Puhler A (1993) The sucrose synthase gene is predominantly expressed in the root-nodule tissue of *Vicia faba*. *Mol. Plant-Microbe Interact.* 6: 507-514.

Kyozuka J, Fujimoto H, Izawa T, Shimamoto K (1991) Anaerobic induction and tissue-specific expression of maize *Adh1* promoter in transgenic rice plants and their progeny. *Mol. Gen. Genet.* 228: 40-48.

Kyozuka J, McElroy D, Hayakawa T, Xie Y, Wu R, Shimamoto K (1993) Light-regulated and cell-specific expression of tomato *rbcS-gusA* and rice *rbcS-gusA* fusion genes in transgenic rice. *Plant Physiol.* 102: 991-1000.

Larkin JC, Oppenheimer DG, Pollock S, Marks MD (1993) Arabidopsis *GLABROUS1* gene requires downstream sequences for function. *Plant Cell* 5: 1739-1748.

Lebrun M, Freyssinet G (1991) Nucleotide sequence and characterization of a maize cytoplasmic ribosomal protein S11 cDNA. *Plant Mol. Biol.* 17: 265-268.

Lee DR (1972) Possible significance of filaments in sieve elements. *Nature* 235: 286.

LeGendre N and Matsudaira P (1989) Electroblotting to PVDF membranes. In Matsudaira PT (ed) *A Practical Guide to Protein and Peptide Purification for Microsequencing*, pp.53-59. Academic Press, London.

Lehmann J (1973) Studies on the phloem exudate of *Cucurbita pepo* L. *Planta* 114: 51-61.

Leisner SM, Turgeon R (1993) Movement of virus and photoassimilate in the phloem: a comparative analysis. *BioEssays* 15: 741-748.

Lerner DR, Raikhel NV (1989) Cloning and characterization of root-specific barley lectin. *Plant Physiol.* 91: 124-129.

Ling KC (1977) Rice ragged stunt disease. *Int. Rice Res. Newl.* 2: 6-7.

Lowell CA, Tomlinson PT, Koch KE (1989) Sucrose-metabolizing enzymes in transport tissues and adjacent sink structures in developing citrus fruit. *Plant Physiol.* 90: 1394-1402.

Lutcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codons differs in plants and animals. *EMBO J.* 6: 43-48.

Maas C, Laufs J, Grant S, Korfhage C, 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 Mol. Biol.* 16: 199-207.

Maas C, Schaal S, Werr W (1990) A feedback control element near the transcription start site of the maize *Shrunken* gene determines promoter activity. *EMBO J.* 9: 3447-3452.

MacRobbie EAC (1971) Phloem translocation. Facts and mechanisms: a comparative survey. *Biol. Rev.* 46: 429-481.

Maier LG, Brown JWS, Toloczyki C, Feix G (1987) Binding of a nuclear factor to a consensus sequence in the 5' flanking region of zein genes from maize. *EMBO J.* 6: 17-22.

Malyshenko SI, Kondakova OA, Nazarova YV, Rassadina GV, Krasheninnikova LV, Khromova LM, Talyansky ME, Atabekov IG (1993) Resistance of transgenic plants producing defective viral transport proteins to viruses. *Doklady Akademii Nauk* 328: 517-519.

Maniatis T, Fritsch EF, Sambrook J (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Maraña C, Garcia-Olmedo F, Carbonero P (1990) Differential expression of two types of sucrose synthase-encoding genes in wheat in response to anaerobiosis, cold shock and light. *Gene* 88: 167-172.

Maraña C, Garcia-Olmedo F, Carbonero P (1988a) Linked sucrose synthase genes in group-7 chromosomes in hexaloid wheat (*Triticum aestivum* L.). *Gene* 63: 253-260.

Maraña C, Garcia-Olmedo F, Carbonero P (1988b) Equivalent locations of sucrose synthase genes in chromosomes 7D of wheat, 7Ag of *Agropyron elongatum*, and 7H of barley. *FEBS Lett.* 234: 417-420.

Marcotte WRJr, Bayley CC, Quatrano RS (1988) Regulation of a wheat promoter by abscisic acid in rice protoplasts. *Nature* 335: 454-457.

Martin T, Frommer WB (1993) Expression of an *Arabidopsis* sucrose synthase gene indicates a role in metabolization of sucrose both during phloem loading and in sink organs. *The Plant J.* 4: 367-377.

Marty I, Meyer Y (1992) cDNA nucleotide sequence and expression of a tobacco cytoplasmic ribosomal protein L2 gene. *Nucl. Acid. Res.* 20: 1517-1522.

Mayer A, Siegel NR, Schwartz AL, Ciechanover A (1989) Degradation of proteins with acetylated amino termini by ubiquitin system. *Science* 244: 1480-1483.

McCarty DR, Shaw JR, Hannah LC (1986) The cloning, genetic-mapping, and expression of the constitutive sucrose synthase locus of maize. *Proc. Natl. Acad. Sci. USA* 83: 9099-9103.

McCormick S, Mauvais J, Fedoroff N (1982) Evidence that the two sucrose synthetase genes in maize are related. *Mol. Gen. Genet.* 187: 494-500.

McGoughey WH (1985) Insect resistance to the biological insecticide, *Bacillus thuringiensis*. *Science* 229: 193-195.

McGaughey WH, Whalon ME (1992) Managing insect resistance to *Bacillus thuringiensis* toxins. *Science* 258: 1451-1455.

Messing J (1983) New M13 vector for cloning. In Wu R, Grossman L, Moldave K (eds) *Meth. Enzymol.* Vol 101, pp.20-78. Academic Press, London.

Mierendorf RC, Pfeffer D (1987) Direct sequencing of denatured plasmid DNA. In Berger SL, Kimmel AR (eds) *Meth. Enzymol.* Vol. 152, pp.556-562. Academic Press, London.

Misra JB, Daniel EV, Premchand (1994) Sucrose synthase and sucrose phosphate synthase in potato: distribution of activities and variations in leaves. *Plant Physiol. Biochem.* 32:131-136.

Mochida O (1992) Management of insect pest of rice in 2000. In Aziz A, Kadir SA, Barlow HS (eds) *Pest Management and the Environment in 2000*, pp.67-81. CAB Int., UK.

Mochida O, Okada T (1979) Taxonomy and biology of *Nilaparvata lugens* (Hom., Delphacidae). In *Brown Planthopper - Threat to Rice Production in Asia*, pp.21-43. International Rice Research Institute, Los Banos, Philippines.

Mohan R, Bajar AM, Kolattukudy PE (1993) Induction of a tomato anionic peroxidase gene (*tap 1*) by wounding in transgenic tobacco and activation of *tap 1*/GUS and *tap 2*/GUS chimeric gene fusions in transgenic tobacco by wounding and pathogen attack. *Plant Mol. Biol.* 21: 341-354.

Möller W, Groene A, Terhorst C, Amons R (1972) 50-S ribosomal proteins: purification and partial characterization of two acidic proteins, A1 and A2, isolated from 50-S ribosomes of *Escherichia coli*. *Eur. J. Biochem.* 25: 5-12.

Moos MJ, Nguyen NY, Liu TY (1988) Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an innert support. *J. Biol. Chem.* 263: 6005-6008.

Morell M, Copeland L (1985) Sucrose synthase of soybean nodules. *Plant Physiol.* 78: 149-154.

Myers KK, Fonzi WA, Sypherd PS (1992) Isolation and sequence determination of the gene for translation elongation factor 3 from *Candida albicans*. *Nucl. Acids Res.* 20: 1705-1710.

Nguyen-Quoc B, Krivitzky M, Huber SC, Lecharny A (1990) Sucrose synthase in developing maize leaves: Regulation of activity by protein level during the import to export transition. *Plant Physiol.* 94: 516-523.

Nickel JL (1973) Pest situation in changing agricultural systems - a review. *Bull. Ent. Soc. Amer.* 19: 136-142.

Nolte KD, Koch KE (1993a) Carbohydrate supply affects the level and localization of the Sh1 sucrose synthase protein. *Plant Physiol.* 102 (1): 43. Meeting Abstract.

Nolte KD, Koch KE (1993b) Companion-cell specific localization of sucrose synthase in zones of phloem loading and unloading. *Plant Physiol.* 101: 899-905.

Norton GA, Way MJ (1990) Rice pest management systems - past and future. In Grayson BT, Green MB, Copping LG (eds) *Pest Management in Rice*, pp.1-14. University Press, Cambridge.

Nuske J, Eschrich W (1976) Synthesis of P-protein in mature phloem of *Cucurbita maxima*. *Planta* 132: 109-118.

Odell JT, Knowlton S, Lin W, Mauvais CJ (1988) Properties of an isolated transcription stimulating sequence derived from the cauliflower mosaic virus 35S promoter. *Plant Mol. Biol.* 10: 263-272.

Odell, JT, Nag, F, Chua NH (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 313: 810-812.

Ohta S, Hattori T, Morikami A, Nakamura K (1991) High-level expression of a sweet potato sporamin gene promoter: β-glucuroidase (GUS) fusion gene in the stems of transgenic tobacco plants is conferred by multiple cell-type-specific regulatory elements. *Mol. Gen. Genet.* 225: 369-378.

Oka IN (1979) Cultural control of the brown planthopper. In *Brown Planthopper - Threat to Rice Production in Asia*, pp.357-369. International Rice Research Institute, Los Banos, Philippines.

Pathak MD, Dhaliwal GS (1981) Trends and strategies for rice pest problems in tropical Asia. *IRRI Research Paper Series* No. 64: 1-15.

Pathak MD, Heinrichs EA (1982) Selection of biotype populations 2 and 3 by exposure to resistant rice varieties. *Environ. Entomol.* 11: 85-90.

Pathak MD, Saxena RC (1980) Breeding approaches in rice. In Maxwell FG, Jennings PR (eds) *Breeding Plants Resistant to Insects,* pp.421-455. John Wiley & Sons, New York.

Perlak FJ, Deaton RW, Armstrong TA, Fuchs RL, Sims SR, Greenplate JT, Fischhoff DA (1990) Insect resistant cotton plants. *Bio/Technology* 8: 939-943.

Peterhans A, Datta SK, Datta K, Goodall GJ, Potrykus I, Paszkowski J (1990) Recognition efficiency of *dicotyledeneae*-specific promoter and RNA processing signals in rice. *Mol. Gen. Genet.* 222: 361-368.

Powell KS, Gatehouse AMR, Hilder VA, Gatehouse JA (1993) Antimetabolic effects of plant lectins and plant and fungal enzymes on the nymphal stages of two omportant rice pests, *Nilaparvata lugens* and *Nephotettix cinciteps*. *Entomol. Exp. Appl.* 66: 119-126.

Powell KS, Gatehouse AMR, Hilder VA, Van Damme EJM, Peumans WJ, Boonjawat J, Horsham K, Gatehouse JA (1994) Different antimetabolic effects of related lectins towards nymphal stages of *Nilaparvata lugens* (Stål). Submitted.

Powell Abel P, Fraley RT, Beachy RN (1986) Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. *Science* 232: 738-743.

Powell Abel P, Stark DM, Sanders PR, Beachy RN (1989) Protection against tobacco mosaic virus in transgenic plants that express tobacco mosaic virus antisense RNA. *Proc. Natl. Acad. Sci. USA* 86: 6949-6952.

Qu R, Bhattacharyya M, Laco GS, de Kochko A, Rao BLS, Kaniewska MB, Elmer JS, Rochester DE, Smith CE, Beachy RN (1991) Characterization of the genome of rice tungro bacilliform virus: comparison with *Commelina* yellow mottle virus and caulimoviruses. *Virology* 185: 354-364.

Quayle T, Feix G (1992) Functional analysis of the -300 region of maize zein genes. *Mol. Gen. Genet.* 231: 369-374.

Rahbe Y, Febvay G (1993) Protein toxicity to aphids: an *in vitro* test on *Acyrthosiphon pisum*. *Entomol. Exp. Appl.* 67: 149-160.

Read SM, Northcote DH (1983a) Chemical and immunological similarities between the phloem proteins of three genera of the Cucurbitaceae. *Planta* 158: 119-127.

Read SM, Northcote DH (1983b) Subunit structure and interactions of the phloem proteins of *Cucurbita maxima* (pumpkin). *Eur. J. Biochem.* 134: 561-569.

Ricard B, Rivoal J, Spiteri A, Pradet A (1991) Anaerobic stress induces the transcription and translation of sucrose synthase in rice. *Plant Physiol.* 95: 669-674.

Riesmeier JW, Hirner B, Frommer WB (1993) Potato sucrose transporter expression in minor veins indicates a role in phloem loading. *Plant Cell* 5: 1591-1598.

Rivera CT, Qu SH, Iida TT (1966) Grassy stunt disease of rice and its transmission by the planthopper, *Nilaparvata lugens* (Stål). *Plant Disease Reporter* 50: 453-456.

Roderick GK (1994) Genetics of host plant adaptation in Delphacid planthoppers. In Denno RF, Perfect TJ (eds) *Planthoppers, Their Ecology and Management*, pp.551-570. Chapman & Hall, New York, USA.

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. In: Gelvin SB, Schilperoot RA, Verma DPS (eds) *Plant Molecular Biology Manual*, pp.A6/1-A6/10. Kluwer Academic Publishers, Dordrecht.

Rosen, H. (1957) A modified ninhydrin colorimetric analysis for amino acids. *Arch. Biochem. Biophys.* 67: 10-15.

Ross HA, Davies HV (1992) Purification and chracterisation of sucrose synthase from the cotyledons of *Vicia faba* L. *Plant Physiol.* 100: 1008-1013.

Rowland LJ, Chen YC, Chourey PS (1989) Anaerobic treatment alters the cell specific expression of *Adh*-1, *Sh*, and *Sus* genes in roots of maize seedlings. *Mol. Gen. Genet.* 218: 33-40.

Ryan CA (1978) Proteinase inhibitors in plant leaves: a biochemical model for pest-induced natural plant protection. *Trends Biochem. Sci.* 7: 148-150.

Sabnis DD, Hart JW (1974) Studies on the possible occurrence of actomyosin-like protein in phloem. *Planta* 118: 271-281.

Sabnis DD, Hart JW (1976) A comparative analysis of phloem exudate proteins from *Cucumis melo*, *Cucumis sativus* and *Cucurbita maxima* by polyacrylamide gel electrophoresis and isoelectric focusing. *Planta* 130: 211-218.

Sabnis DD, Hart JW (1978) The isolation and some properties of a lectin (haemagglutinin) from *Cucurbita* phloem exudate. *Planta* 142: 97-101.

Sabnis DD, Hart JW (1979) Heterogeneity in phloem protein complements from different species. Consequences to hypotheses concerned with P-protein function. *Planta* 145: 459-466.

Saito K, Yamazaki M, Kaneko H, Murakoshi I, Fukuda Y, Montagu MV (1991) Tissue-specific and stress-enhancing expression of the TR promoter for mannopine synthase in transgenic medicinal plants. *Planta* 184: 40-46.

Salanoubat M, Belliard G (1989) The steady-state level of potato sucrose synthase mRNA is dependent on wounding, anaerobiosis and sucrose concentration. *Gene* 84: 181-185.

Salanoubat M, Belliard G (1987) Molecular cloning and sequencing of sucrose synthase cDNA from potato (*Solanum tuberosum* L.). *Gene* 60: 47-56.

Sanchez de la Hoz S, Vicente-Carbajosa J, Mena M, Carbonero P (1992) Homologous sucrose synthase genes in barley (Hordeum vulgare) are located in chromosome 7H (syn.1) and 2H - evidence for a gene translocation? *FEBS Lett.* 310: 46-50.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.

Sangwan V, Lenvik TR, Gantt JS (1993) The *Arabidopsis thaliana* ribosomal protein S15 (*rig*). *Biochim. Biophys. Acta* 1216: 221-226.

Saxena RC, Khan ZR (1989) Factors affecting resistance of rice varieties to planthopper and leafhopper pests. *Agr. Zool. Rev.* 3: 97-132.

Schmidt J, Herfurth E, Subramanian AR (1992) Purification and characterization of seven chloroplast ribosomal proteins: evidence that organelle ribosomal protein genes are functional and that NH_2-terminal processing occurs *via* multiple pathways in chloroplasts. *Plant Mol. Biol.* 20: 459-465.

Schulz A, Alosi MC, Sabnis DD, Park RB (1989) A phloem-specific, lectin-like protein is located in pine sieve-element plastids by immunocytochemistry. *Planta* 179: 506-515.

Scott R, Draper J, Jefferson R, Dury G, Jacob L (1988) Analysis of gene organization and expression in plants. In: Draper J, Scott R, Armitage P, Walden R (eds) *Plant Genetic Transformation and Gene Expression-A Laboratory Manual*, pp.265-339. The Alden Press, Oxford.

Sham MH, Northcote DH (1987) Transcription and translation of phloem protein (PP2) during phloem differentiation in *Cucurbita maxima*. *Planta* 170: 392-399.

Shaw CH (1988) *Plant Molecular Biology: A Laboratory Manual*. IRL Press, Oxford.

Simmons CR, Litts JC, Huang N, Rodriguez RL (1992) Structure of a rice β-glucanase gene regulated by ethylene, cytokinin, wounding, salicylic acid and fungal elicitors. *Plant Mol. Biol.* 18: 33-45.

Slightom JL, Drong RF (1988) Procedures for constructing genomic clone banks. In Gelvin SB, Schilperoot RA, Verma DPS (eds) *Plant Molecular Biology Manual*, pp.A8/1-A8/42. Kluwer Academic Publishers, Dordrecht.

Slightom JL, Durand-Tardif M, Jouanin L, Tepfer D (1986) Nucleotide sequence analysis of TL-DNA of *Agrobacterium rhizogenes* agropine type plasmid. *J. Biol. Chem.* 261: 108-121.

Sloan RT, Sabnis DD, Hart JW (1976) The heterogeneity of phloem exudate proteins from different plants: a comparative survey of ten plants using polyacrylamide gel electrophoresis. *Planta* 132: 97-102.

Smith JJ, Raikhel NV (1989) Nucleotide sequences of cDNA clones encoding wheat germ agglutinin isolectins A and D. *Plant Mol. Biol.* 13: 601-603.

Smith LM, Sabnis DD, Johnson RPC (1987) Immunocytochemical localization of phloem lectin from *Cucurbita maxima* using peroxidase and colloidal-gold labels. *Planta* 170: 461-470.

Smith RF (1972) The impact of the green revolution on plant protection in tropical and subtropical areas. *Bull. Ent. Soc. Amer.* 18: 7-14.

Sogawa K (1976) Feeding physiology of the brown planthopper. Paper presented at *the International Seminar on the Rice Brown Planthopper, Asian and Pacific Council, Food and Fertilizer Technology Center, October, 1976. Tokyo, Japan.* 39 p.

Spiller NJ (1990) An ultrastructural study of the stylet pathway of the brown planthopper *Nilaparvata lugens. Entomol. Exp. Appl.* 54: 191-193.

Springer B, Bellmann R, Werr W (1990) Formation of composite nucleoprotein complexes near the transcription start of the *Shrunken* gene from maize. *Mol. Gen. Genet.* 221: 94-101.

Springer B, Werr W, Starlinger P, Bennett DC, Zokolica M (1986) The *Shrunken* gene on chromosome 9 of *Zea mays* L. is expressed in various plant tissues and encodes an anaerobic protein. *Mol. Gen. Genet.* 205: 461-468.

Stone TB, Sims SR, Marrone PG (1989) Selection of tobacco budworm for resistance to genetically engneered *Pseudomonas florescens* containing the alpha endotoxin of *Bacillus thuringiensis* sub-species Kurstaki. *J. Invert. Path.* 53: 228-234.

Sugaya S, Hayakawa K, Handa T, Uchimiya H (1989) Cell-specific expression of the *rolC* gene of the TL-DNA of Ri plasmid in transgenic tobacco plants. *Plant Cell Physiol.* 30: 649-653.

Sugaya S, Uchimiya H (1992) Deletion analysis of the 5'-upstream region of the *Agrobacterium rhizogenes* Ri plasmid *rolC* gene required for tissue-specific expression. *Plant Physiol.* 99: 464-467.

Sun JD, Loboda T, Sung SJS, Black Jr. CC (1992) Sucrose synthase in wild tomato, *Lycopersicon chmielewskii*, and tomato fruit sink strenth. *Plant Physiol.* 98: 1163-1169.

Suzuki K, Wool IG (1993) The primary structure of rat ribosomal protein L23a. *J. Biol. Chem.* 268: 2755-2761.

Tada Y, Sakamoto M, Matsuoka M, Fujimura T (1991) Expression of a monocot LHCP promoter in transgenic rice. *EMBO J.* 10: 1803-1808.

Taliercio EW, Chourey PS (1989) Post-transcriptional control of sucrose synthase expression in anaerobic seedlings of maize. *Plant Physiol.* 90: 1359-1364.

Tavladoraki P, Benvenuto E, Trinca S, De Matinis D, Cattaneo A, Galeffi P (1993) Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack. *Nature* 366: 469-472.

Terada R, Shimamoto K (1990) Expression of CaMV 35S-GUS gene in transgenic rice plants. *Mol. Gen. Genet.* 220: 389-392.

Terhorst C, Möller W (1973) The primary structure of an acidic protein from 50-S ribosomes of *Escherichia coli* is involved in GTP hydrolysis dependent on elongation factors G and T. *Eur. J. Biochem.* 34: 138-152.

Thaine R (1969) Movement of sugars through plants by cytoplasmic pumping. *Nature* 222: 873-875.

Thornburg R, An G, Cleveland T, Johnson R, Ryan CA (1987) Wound-inducible expression of a potato inhibitor II-chloraphenicol acetyl transferase gene in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA* 84: 744-748.

Tomlinson PT, Duke ER, Nolte KD, Koch KE (1991) Sucrose synthase and invertase in isolated vascular bundles. *Plant Physiol.* 97: 1249-1252.

Vaeck M, Reynaerts A, Jansens S, DeBeuckeleer M, Dean C, Zabeau M, Van Montagu M, Leemans J (1987) Transgenic plants protected from insect attack. *Nature* 328: 33-37.

Van Damme EJM, DeClercq N, Claessens F, Hemschoote K, Peeters B and Peumans WJ (1991a) Molecular cloning and characterization of multiple isoforms of the snowdrop *(Galanthus nivalis L.)* lectin. *Planta* 186: 35-43.

Van Damme EJM, Kaku H, Ferini F, Goldstein J, Peeters B, Yagi F, Decock B and Peumans WJ (1991b) Biosynthesis, primary structure and molecular cloning of snowdrop *(Galanthus nivalis)* lectin. *Eur. J. Biochem.* 202: 23-30.

Vasil V, Clancy M, Ferl RJ, Vasil IK, Hannah LC (1989) Increased gene expression by the first intron of maize *Shrunken-1* locus in grass species. *Plant Physiol.* 91: 1575-1579.

Vasil IK, Vasil V, Srivastava V, Castillo AM, Fromm ME (1994) Molecular genetic improvement of wheat. *J. Cell Biochem.* 18AS: 78. Meeting abstract.

Velten J, Velten L, Schell J (1984) Isolation of a dual promoter fragment from the Ti plasmid of *Agrobacterium tumefaciens. EMBO J.* 3: 2723-2730.

Vozari-Hampe MM, Viegas C, Saucedo C, Rosseto S, Manica GG, Hampe OG (1992) A lectin from *Sechium edule* fruit exudate. *Phytochem.* 31: 1477-1480.

Wagner M, Perry RP (1985) Charaterization of the multigene family encoding the mouse S16 ribosomal protein: strategy for distinguishing an expressed gene from its processed pseudogene counterpart by an analysis of total genomic DNA. *Mol. Cell. Biol.* 5: 3560-3576.

Walbot V and Gallie D (1991) Gene expression in rice. In Khush GS, Toenniessen GH (eds) *Rice Biotechnology*, pp.225-251, CAB Int. UK.

Walker JC, Howard EA, Dennis ES, Peacock WJ (1987) DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase 1 gene. *Proc. Natl. Acad. Sci. USA* 84: 6624-6628.

Walker TS (1972) The purification and some properties of a protein causing gelling in phloem sieve exudate from *Cucurbita pepo. Biochim. Biophys. Acta* 257: 433-444.

Walker TS, Thaine R (1971) Proteins and the structural components in exudate from sieve tubes in *Cucurbita pepo* stems. *Ann. Bot.* (London) 35: 773-790.

Walkey DGA (1985) *Applied Plant Virology*, Section 7.4.3, Redwood Burn Ltd, Trowbridge, England.

Wan Y, Lemaux PG (1994) Development and use of an efficient transformation system for barley. *J. Cell Biochem.* 18AS: 78. Meeting abstract.

Wang AY, Yu WP, Juang RH, Huang JW, Sung HY, Su JC (1992) Presence of 3 rice sucrose synthase genes as revealed by cloning and sequencing of cDNA. *Plant Mol. Biol.* 18: 1191-1194.

Wang F, Sanz A, Brenner ML, Smith A (1993) Sucrose synthase, starch accumulation, and tomato fruit sink strength. *Plant Physiol.* 101: 321-327.

Wang F, Smith AG, Brenner ML (1994) Temporal and spatial expression pattern of sucrose synthase during tomato fruit development. *Plant Physiol.* 104: 535-540.

Weber C, Franke WW, Kartenbeck J (1974) Structure and biochemistry of phloem-proteins isolated from *Cucurbita maxima*. *Exp. Cell Res.* 87: 79-106.

Werr W, Frommer WB, Maas C, Starlinger P (1985) Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L.. *The EMBO J.* 4: 1377-1380.

Whalon ME, van de Baan HE, Untung K (1990) Resistance management of brown planthopper, *Nilaparvata lugens*, in Indonesia. In Grayson BT, Green MB, Copping LG (eds) *Pest Management in Rice*, pp.155-165, Cambridge University Press.

Wilhelm K (1880) Beitrage zur Kenntnis des Siebrohrenapparates dicotyler Pflanzen. Engelmann, Leipzig.

Wolf S, Deom CM, Beachy R, Lucas WJ (1991) Plamosdesmatal function is probed using transgenic tobacco plants that express a virus movement protein. *Plant Cell* 3: 593-604.

Xu D, McElroy D, Thornburg RW, Wu R (1993) Systemic induction of a potato *pin2* promoter by wounding, methyl jasmonate, and abscisic acid in transgenic rice plants. *Plant Mol. Biol.* 22: 573-588.

Yang NS, Russell D (1990) Maize sucrose synthase-1 promoter directs phloem cell-specific expression of *Gus* gene in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA* 87: 4144-4148.

Yu WP, Wang AY, Juang RH, Sung HY, Su JC (1992) Isolation and sequences of rice sucrose synthase cDNA and genomic DNA. *Plant Mol. Biol.* 18: 139-142.

Zammit A, Copeland L (1993) Immunocytochemical localization of nodule-specific sucrose synthase in soybean nodules. *Aus. J. Plant Physiol.* 20: 25-32.

Zhang W, McElroy D, Wu R (1991) Analysis of rice *Act1* 5' region activity in transgenic plants. *Plant Cell* 3: 1155-1165.

Zheng Z, Xiao S, Murai N (1991) Promoter for phaseolin, the major bean seed protein, does not function in rice although phaseolin protein can be expressed in transgenic rice from a wheat glutelin promoter. In Hallick RB (ed) *Program and Abstracts, 3rd Int. Cong. Plant Mol. Biol.* abstract 733. Tucson, Arizona University, USA.

Zingde SM, Shirsat NV, Gothoskar BP (1986) Peptide mapping of proteins in gel bands after partial cleavage with acidic cyanogen bromide vapor. *Anal. Biochem.* 155: 10-13.

APPENDIX

Appendix A. Nucleotide sequence of PCR-amplified maize *Sh*-1 gene exon 13 (pSHEX). The nucleotide T and the 15 nucleotides GGCGCATTCGTGCAG were lost from the 5' and 3' primer regions, respectively. The remaining nucleotides of the primers are underlined.

```
 TC GTG CTG AAG GAC AAG AAG AAG CCG ATC ATC TTC TCG ATG GCG
      5' primer (primer A)

CGT CTC GAC CGC GTG AAG AAC ATG ACA GGC CTG GTC GAG ATG TAC

GGC AAG AAC GCG CGC CTG AGG GAG CTG GCG AAC CTC GTG ATC GTT

GCC GGT GAC CAC GGC AAG GAG TCC AAG GAC AGG GAG GAG CAG GCG

GAG TTC AAG AAG ATG TAC AGC CTC ATC GAC GAG TAC AAG TTG AAG

GGC CAT ATC CGG TGG ATC TCG GCG CAG ATG AAC CGT GTC CGC AAC

GGG GAG CTG TAC CGC TAC ATT TGC GAT ACC AAG
                                      3' primer (primer B)
```

Appendix B. Molecular size (bp) of λDNA markers

HindIII-digested λ DNA, consisting 7 DNA fragments:

23130
9416
6682
4361
2322
2027
564

PstI-digested λ DNA, consisting of 29 DNA fragments:

11501	2459	1093	264	150
5077	2443	805	247	94
4749	2140	514	216	87
4507	1986	468	211	72
2838	1700	448	200	15
2556	1159	339	164	

168

Appendix C. Amino acids and codons

Amino acid	Three letter code	Single letter code	Codons
Alanine	Ala	A	GCA/C/G/T
Arginine	Arg	R	AGA/G, CGA/C/G/T
Aspartate	Asp	D	GAC/T
Asparagine	Asn	N	AAC/T
Cysteine	Cys	C	TGC/T
Glutamate	Glu	E	GAA/G
Glutamine	Gln	Q	CAA/G
Glycine	Gly	G	GGA/C/G/T
Histidine	His	H	CAC/T
Isoleucine	Ile	I	ATA/C/T
Leucine	Leu	L	CTA/C/G/T, TTA/G
Lysine	Lys	K	AAA/G
Methionine	Met	M	ATG
Phenylalanine	Phe	F	TTC/T
Proline	Pro	P	CCA/C/G/T
Serine	Ser	S	TCA/C/G/T, AGT/C
Threonine	Thr	T	ACA/C/G/T
Tryptophan	Trp	W	TGG
Tyrosine	Tyr	Y	TAC/T
Valine	Val	V	GTA/C/G/T
Stop;			
Amber	Amb		TAG
Ochre	Och		TAA
Umber	Umb		TGA

Appendix D. cDNA clones p27A, p4A and p30A isolated from *Cucurbita pepo.*

Structure of p27A:
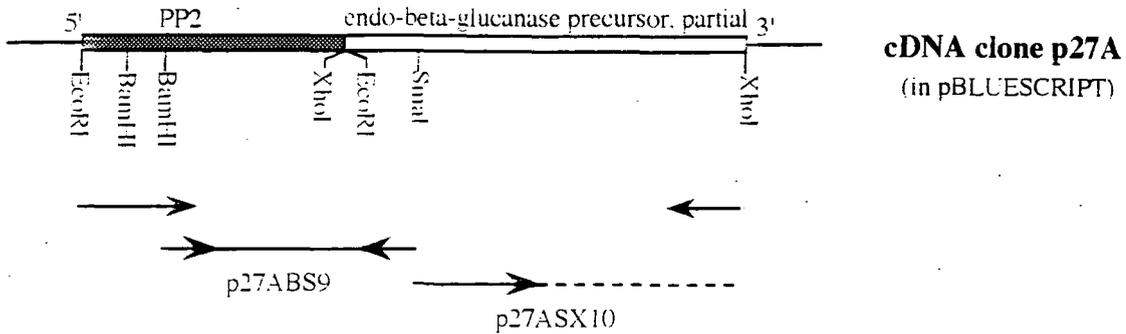
Restriction map of PP2 cDNA clone p27A and sequencing strategy. The arrows indicate the
sequencing directions: the dashed lines represent unsequenced regions: p27ABS9 and
p27ASX10 are names of subclones. The EcoRI and XhoI sites are cloning sites for the cDNA
library construction.

Partial DNA sequence of p4A near the EcoRI cloning site:

TTTTTTTTTTTCATCAATCAACTCTATATTCCACTAAATCAAAAGAATTTCT
CATGTATTACAAACTAAGAAAACAAAGAGAAAAAAAGAGAGAGGAAAACG
AAACAACAAAAGAATCAGAATACTGTTCATCTCCTGTTAGATCCAAGGCCT
TCCGACTCCATCGACGGGGAAGAAGAGGGAGAAGAAATGCCAGAACCAGA
AGCAGCAGCCATAGCCAAGCAGCCCTCAATCC

Partial sequence of p30A near the EcoRI cloning site:

GGAGAATGTTTTCTTGATCAATGTACTTGCATGAATCATGTAAGAAAACAG
AATTTTAAATTTTGTATTTTTAACATTCTGGATTATTAAACAATTGATGTTA
TGACTGTGGAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Update section

Short communication

A complete sequence of the rice sucrose synthase-1 (*RSs1*) gene

Ming-Bo Wang, Donald Boulter and John A. Gatehouse *
Department of Biological Sciences, University of Durham, South Road, Durham DH1 3LE, UK (author for correspondence)*

Abstract

Using a fragment of the maize sucrose synthase gene *Sh-1* as probe, the rice genome was shown to contain at least three genes encoding sucrose synthase. One of these genes was isolated from a genomic library, and its full sequence, including 1.7 kb of 5' flanking sequence and 0.9 kb of 3' flanking sequence, is reported. The new rice gene, designated *RSs1*, is highly homologous to maize *Sh-1* (approx. 94% identity in derived amino acid sequence), and contains an identical intron-exon structure (16 exons and 15 introns). Both *RSs1* and maize *Sh-1* show similar sequence homologies to a second rice sucrose synthase gene described recently (designated *RSs2*, Yu *et al.* (1992) Plant Mol Biol 18: 139–142), although both the rice genes predict an extra 6 amino acids at the C-terminus of the protein when compared to the maize gene. The *RSs1* 5' flanking sequence contains a number of promoter-like sequences, including putative protein-binding regions similar to maize zein genes.

A λEMBL3 library (commercially available from Clontech Laboratories) of rice (*Oryza sativa* L. Japonica) genomic DNA was screened for sucrose synthase genes, using an exon 13 fragment of the maize sucrose synthase *Sh-1* gene [1], synthesized by polymerase chain reaction (PCR), as a heterologous probe. Eight positive plaques were identified, and, after plaque purification, DNA was prepared from the clones. The eight genomic clones could be divided into two classes, apparently representing two different regions of genomic DNA, on the basis of restriction mapping and Southern hybridization. Analysis by partial sequencing showed that one of the two classes contained a sucrose synthase gene with high sequence homology to the maize *Sh-1* gene, while the other class contained a gene with much

lower homology to *Sh-1*, but which was almost completely identical in sequence to the recently published rice sucrose synthase gene *RSs2* [5]. The genes in these clones were thus characterized as *RSs1* and *RSs2*, respectively. When a Southern blot of rice genomic DNA was hybridized with the maize *Sh-1* exon 13 probe (Fig. 1), at least three fragments were detected in restriction digests, suggesting the presence of further homologous genes in the rice genome. Two sucrose synthase isozymes have been identified in maize [2, 3] and sorghum [4].

One of the clones in the *RSs1* class was fully investigated by restriction mapping, subcloning and sequencing. This clone, designated λRSS2.4, contains a full-length gene (Fig. 2). The 2.1 kb *Eco* RI/*Eco* RI and 7.0 kb *Bam* HI/*Bam* HI frag-
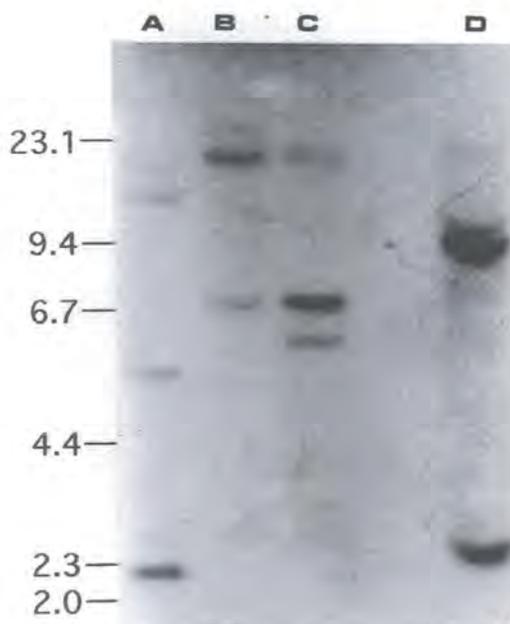
Fig. 1. Southern blot of restriction enzyme digests of rice genomic DNA (tracks A–C, 20 µg loaded per track) and maize genomic DNA (track D, 40 µg loaded) probed with exon 13 sequence from the maize sucrose synthase gene *Sh-1*. Restriction enzymes used were *Eco* RI (tracks A and D), *Hin*d III (track B) and *Bam* HI (track C). Molecular weight markers (λ *Hin*d III fragments, kb) are shown at the left.

of the first exon to 710 bp downstream of the putative polyadenylation site, is shown in Fig. 3.

The whole transcription unit of *RSs1*, including the transcription start and the polyadenylation sites, the 16 exons and 15 introns, was identified by similarity of the intron/exon structure, and the high identity of the exon sequences, to that of the maize *Sh-1* gene. The coding region of *RSs1* exhibits high homology with *Sh-1* (87% in nucleotide sequence and 94% in deduced amino acid sequence). Both *RSs1* and *Sh-1* have very similar homologies with *RSs2* (about 77% in nucleotides and 80% in deduced amino acids), suggesting that gene duplication arose prior to the evolutionary separation of rice and maize as species. A major difference between *RSs1* and *Sh-1* occurs in the last exon, where *RSs1* has 18 bp extra coding sequence compared to *Sh-1* immediately before the stop codon. The same extent of extra coding sequence also exists in this region of *RSs2*, implying that the maize *Sh-1* gene has lost 6 amino acids of its coding capacity during evolution.

Putative TATA and CAAT boxes in *RSs1* were identified at −21 and −79 relative to the transcription start, respectively, by homology to the maize *Sh-1* sequence. Several other promoter-like sequences were also found in the upstream region (shown in Fig. 3), including two 48 bp direct repeats, and three palindromes which resemble one another to some extent and are similar to the protein-binding sequences in the endosperm-specific maize zein promoter [6].

ments of λRSS2.4 correspond to the major hybridization bands detected by the maize *Sh-1* probe, by Southern blotting, in the *Eco* RI and *Bam* HI digests of rice genomic DNA (Fig. 1). The 8.1 kb region shown in Fig. 2 was digested with various restriction endonucleases to generate small or overlapping fragments. These fragments were then subcloned into pUC18, pUC19 or M13 vectors and sequenced using an Applied Biosystems model 370A DNA sequencer. Both strands of the DNA were fully sequenced on overlapping fragments. The full sequence of this region, which extends from about 1.7 kb upstream

Fig. 2. Partial restriction map of the rice genomic clone λRSS2.4. The arrows indicate the 8.1 kb region containing the rice sucrose synthase gene *RSs1*, whose sequence is presented in Fig. 3.

```
CTCCTTCATTTTCAGTGCAAATGTGCAATGCTGATTAGAGTTTGCAGATGCTGTTTGGTTTAGTTTAGATGTGGCATTTTGTTAGTGGTTTCTTTGATGAAAAATTCTTGGCTATGATAA -1557

AGTTTGCTTTCTGAATATATGAATAGTGGCCATGGTTCAAGAAACTCCAGTTAGGTGGGATAATTTATGGTGATTCTGGGCGCAATTCGGGGAAATTTTTTTGGCGAGAATCTTATCAT -1437

TGAGATAAAGAGGGCAAGAATATCAACAGACTTTTAATCTTAATAAAAAGCACTCTTAGCGTAAGAGCAAAGCATTGCAATCTCGTGTGACAAGAACGTTTCTTTTTCTCCATCTTTTTC -1317

TTTTTTACCAAAAAATGAGTGTTGCCAACTGCTGCACCTTCTTAGGCCGGTTTGTTCTTGTTTGGAACGCACGGAATGCCCGATGCAAAAAAAAAAAAGAAATGCTGTTAACAAATCACT -1197

GTCCTGACACGGCTAATTAGGTGGTAATTTGGTGCATCTGCAAAGAAGCAACAGATGCTTTCTTTCACTGAAAGCATATTTGCATGATTTCTTGTTTCTGCTTGTCCTCTCTCTGATGCT -1077

GACTGTATTCCACTCTGCGCTGTAATGCCATGTTAGTGATTAATATGTTCAAAAGAGCATAAAAGAATTGCCAATTGGATGTTAGAGATTACTGTGTTGTTCAAAAGAGCATAAAAGAAT -957
                                        Direct Repeat                                    Direct

TACCAATTTGATGGTAGATGTTACTAGCACCACCTTGGTGTTTCCCCATGGTTTTCTGCAATTCTGCCCATGATCTTTCTGCTTTTCTGAAAGACCTATGTTTCAGAGGTCAAGCTTCTG -837
Repeat

GAAGGTTATTAGGAGGGATGAGTCGTCATTTTGTCTGTGGGCCCCACTAGTCAGTGTCAATAGTTGTAAAGGGTAGAAATTTTCTTGCTGTTTTTCTTGGAAACAATTTCATTGCGCCTG -717

ATCTGATGGTCGGTCTGGTAATCAAATCACCAGATCCTGAAATCCACCAAATCAAACCGTGAGATTTTTGCAGAGGCAAAACAAGAAAAGCATCTGCTTTATTTCTCTCTTGCTTTCTTT -597

TCATCCCCAACCAGTCCTTTTTTCTTCTGTTTATTTGTAGAAGTCTACCACCTGCAGTCTATTATTCTACAGAGAAAAAGATTGAAGCTTTTTTTTCTCCAAAGCTGACAATGGTGCCGGC -477

ATATGCTAATAGGATACTCCCTTCGTCTAGGAAAAAACCAACCCACTACAATTTTGAATATATATTTATTCAGATTTGTTATGCTTCCTACTCCTTCTCAGGTATGGTGAGATATTTCAT -357
                                                   PL-1

AGTATAATGAATTTGGACATATATTTGTCCAAATTCATCGCATTATGAAATGTCTCGTTCGATCTATGTTGTTATATTATAGACGGAGATAGTAGATTCGGTTATTTTTGGACAGAGAAA -237
        PL-2

GTACTCGCCTGTGCTAGTGACATGATTAGTGACACCATCAGATTAAAAAAACATATGTTTTGATTAAAAAAATGGGGAATTTGGGGGGAGCAATAATTTGGGGTTATCCATTGCTGTTTC -117
                                          PL-3

ATCATGTCAGCTGAAAGGCCCTACCACTAAACCAATATCTGTACTATTCTACCACCTATCAGAATTCAGAGCACTGGGGTTTTGCAACTATTTATTGGTCCTTCTGGATCTCGGAGAAAC +4
                          CAAT BOX                                                      TATA BOX                ^

CCTCCATTCGTTTGCTCGTCTCTGACCACCATTGGGTATGTTGCTTCCATTGCCAAACTGTTCCCTTTTACCCATAGGCTGATTGATCTTGGCTGTGTGATTTTTTGCTTGGGGTTTTGA +124
                    <---------------------------------------------------------------------------------

GCTGATTCAGCGGCGCTTGCAGCCTCTTGATCGTGGTCTTGGCTCGCCCATTTCTTGCGATTCTTTGGTGGGTCGTCAGCTGAATCTTGCAGGAGTTTTTGCTGACATGTTCTTGGGTTT +244
-----------------------------------------------------------------------------------------------------------------------

ACTGCTTTCGGTAAATCTGAACCAAGAGGGGGGTTTCTGCTGCAGTTTAGTGGGTTTACTATGAGCGGATTCGGGGTTTCGAGGAAAACCGGCAAAAAAACCTCAAATCCTCGACCTTTAG +364
-----------------------------------------------------------------------------------------------------------------------

TTTTGCTGCCACGTTGCTCCGCCCCATTGCAGAGTTCTTTTTGCCCCCAAATTTTTTTTTACTTGGTGCAGTAAGAATCGCGCCTCAGTGATTTTCTCGACTCGTAGTCCGTTGATACTG +484
-----------------------------------------------------------------------------------------------------------------------

TGTCTTGCTTATCACTTGTTCTGCTTAATCTTTTTTGCTTCCTGAGGAATGTCTTGGTGCCTGTCGGTGGATGGCGAACCAAAAATGAAGGGTTTTTTTTTTTGAACTGAGAAAAATCT +604
------------------------------------------------------------IVS-1----------------------------------------------

TTGGGTTTTTGGTTGGATTCTTTCATGGAGTCGCGACCTTCCGTATTCTTCTCTTTGATCTCCCCGCTTGCGGATTCATAATATTCGGAACTTCATGTTGGCTCTGCTTAATCTGTAGCC +724
-----------------------------------------------------------------------------------------------------------------------

AAATCTTCATATCTCCAGGGATCTTTCGCTCTGTCCTATCGGATTTAGGAATTAGGATCTAACTGGTGCTAATACTAAAGGGTAATTTGGAACCATGCCATTATAATTTTGCAAAGTTTG +844
-----------------------------------------------------------------------------------------------------------------------

AGATATGCCATCGGTATCTCAATGATACTTACTAAAACCCAACAAATCCATTTGATAAAGCTGGTTCTTTTATCCCTTTGAAAACATTGTCAGAGTATATTGGTTCAGGTTGATTTATTT +964
-----------------------------------------------------------------------------------------------------------------------

TGAATCAGTACTCGCACTCTGCTTCGTAAACCATAGATGCTTTCAGTTGTGTAGATGAAACAGCTGTTTTTAGTTATGTTTTGATCTTCCAATGCTTTTGTGTGATGTTATTAGTGTTGA +1084
-----------------------------------------------------------------------------------------------------------------------

TTTAGCATGGCTTTCCTGTTCAGAGATAGTCTTGCAATGCTTAGTGATGGCTGTTGACTAATTATTCTTGTGCAAGTGAGTGGTTTTGGTACGTGTTGCTAAGTGTAACCTTTCTTTGCA +1204
-----------------------------------------------------------------------------------------------------------------------

GTTCCTGAAATTGAGTCATGGCTGCCAAGCTAGCTCGCCTCCACAGTCTCCGCGAACGCCTCGGTGCCACCTTCTCGTCTCATCCCAATGAGTTGATTGCACTCTTCTCTAGGTGAAGAT +1324
>            M  A  A  K  L  A  R  L  H  S  L  R  E  R  L  G  A  T  F  S  S  H  P  N  E  L  I  A  L  F  S  R<-------

CTCAATATAATATCTTGTGTCTTCTTGGAACCTGCTACTAATTGTTTTGAGTTTTCAATGATTTTATCCATACAGTTGGTTAGTTCTTAGTCAACCCAGATGTGAAACGATAACATGTTA +1444
-----------------------------------------------------------------------------------------------------------------------

GTATGTTACAAATTCTTAGCTGACAAGGTTATGCATCTCTTTAGTTTTAATGATGGTATCTATAATAATCTGTGATGAGCTCTAACATTTTTGTCTTGAGTCATGAATCTTAGATGATGT +1564
----------------------------------------------------IVS-2-----------------------------------------------------------

TGAAATATTGTTTGTTCATAGCTAGATTAAGTACAGATCATGTGGTCTTAAAAATAAGAAGTAAAAATCAGGTTTATGTGGTAGTGCCATTAACATAGTTCAGCCGCAGAAAGGATGTGA +1684
-----------------------------------------------------------------------------------------------------------------------

GATCTAAGCATTGGTTCTATTTCTTAAATATCACTCTATGCAGGTATGTTAACCAGGGAAAGGGAATGCTCCAGCGTCACCAGCTGCTTGCGGAGTTCGATGCCTTGATCGAAGCTGACA +1804
------------------------------------------->  Y  V  N  Q  G  K  G  M  L  Q  R  H  Q  L  L  A  E  F  D  A  L  I  E  A  D

AAGAGAAATATGCTCCCTTTGAAGACATTCTCCGGGCTGCTCAGGTAAATTTGGCAATGAGCCAGGATGGTTATGATAGTCAACTCAAAGTCAAATACCAGTCTTTCTAACTTACTTCCT +1924
K  E  K  Y  A  P  F  E  D  I  L  R  A  A  Q <-----------------------------------------------IVS-3-----------------------

TGACTGTATGTAGGAAGCCATTGTGCTGCCGCCCTGGGTTGCACTGGCCATCAGGCCAAGGCCTGGTGTCTGGGACTACATTCGGGTGAATGTAAGTGAGTTGGCAGTGGAAGAGCTGAG +2044
------------->E  A  I  V  L  P  P  W  V  A  L  A  I  R  P  R  P  G  V  W  D  Y  I  R  V  N  V  S  E  L  A  V  E  E  L  S

TGTTTCTGAGTACTTGGCATTCAAGGAACAGCTTGTTGATGGACAGTAAGTACCAGAACGACGATTTGCAGATTATAACTGAAGCAAAATAACAGTGACTGAATTTTACTTCTGTTTTTC +2164
 V  S  E  Y  L  A  F  K  E  Q  L  V  D  G  H<-----------------------------------------IVS-4----------------------------

TTAAACAGCACCAACAGCAACTTTGTTCTTGAGCTTGATTTTGAGCCCTTCAATGCCTCCTTCCCGCGCCCGTCCATGTCCAAGTCCATCGGAAATGGGGTGCAGTTCCTTAACCGTCAC +2284
------->  T  N  S  N  F  V  L  E  L  D  F  E  P  F  N  A  S  F  P  R  P  S  M  S  K  S  I  G  N  G  V  Q  F  L  N  R  H

CTGTCGTCCAAGTTGTTCAGGACAAGGAGAGCCTCTACCCCTTGCTGAACTTCCTGAAAGCCCATAACCACAAGGGCACGGTAAGCTTCCAATTTGTGGAACATACCAAATGCTTTTAT +2404
 L  S  S  K  L  F  Q  D  K  E  S  L  Y  P  L  L  N  F  L  K  A  H  N  H  K  G  T <------------------------------------

ACTAGATGTTTTTTCTTGTGTGAATTGCTTTAAATGTTAGTCCTTTAGAAACCCACTTGACATTCTCTGTAAGTTCCACCAAATGACGGGATCCGAATTTTCTATTACAGACAATGATGC +2524
------------------------IVS-5--------------------------------------------------------------------------->  T  M  M
```
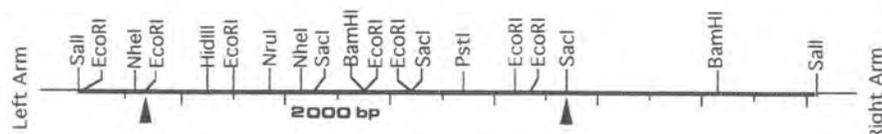
```
TGAATGACAGAATTCAGAGCCTTCGTGGGCTCCAATCATCCCTTAGAAAGGCAGAAGAATATCTGATGGGCATTCCTCAAGACACGCCCTACTCGGAGTTCAACCACAGGTGATTACTTT +2644
 L  N  D  R  I  Q  S  L  R  G  L  Q  S  S  L  R  K  A  E  E  Y  L  M  G  I  P  Q  D  T  P  Y  S  E  F  N  H  R<----------

ATTATTTCTCCTTATCATATAGCCCATGTTATACTACTGCAAATGGAATAATTCCTTACAATGTGATTCATGATTTGAATACAGGTTCCAAGAGCTCGGTTTGGAGAAGGGTTGGGGTGA +2764
-----------------------------IVS-6---------------------------------------------------------------------> F  Q  E  L  G  L  E  K  G  W  G  D

CTGTGCAAAGCGTGTGCTTGACACCATCCACTTGCTTCTTGACCTTCTTGAGGCCCCTGATCCGGCCAACTGGAGAAGTTCCTTGGAACTATTCCAATGATGTTCAATGTTGTTATCCT +2884
 C  A  K  R  V  L  D  T  I  H  L  L  L  D  L  L  E  A  P  D  P  A  N  L  E  K  F  L  G  T  I  P  M  M  F  N  V  V  I  L

GTCTCCGCATGGATACTTTGCCCAATCCAATGTGTTGGGATACCCTGATACTGGTGGTCAGGTAGAGCTTTTTAACAAAATTTTCCCCTTATATACAGTACTTCAAATATTTCTAGTTTTG +3004
 S  P  H  G  Y  F  A  Q  S  N  V  L  G  Y  P  D  T  G  G  Q <----------------------------------------IVS-7---------------

ACCTGTTATTCTGTTGTTTGCAGGTTGTGTACATTTTGGACCAAGTCCGCGCTTTGGAGAATGAGATGCTTTTGAGGATCAAGCAGCAAGGCCTTGATATCACACCTAAGATCCTCATTG +3124
--------------------->V  V  Y  I  L  D  Q  V  R  A  L  E  N  E  M  L  L  R  I  K  Q  Q  G  L  D  I  T  P  K  I  L  I<

TATGTTCAAGATGTTGGTCTTGAATGTTTCAAGTTTTAAACACTTCTCAGTGCGAATTGTAGAGTAACGTGCGGACATGATTTTACTTGATGCAGGTAACCAGGCTGTTGCCTGATGCTG +3244
-------------------------------------------IVS-8-----------------------------------------------------> V  T  R  L  L  P  D  A

TTGGTACTACATGCGGCCAGCGTGTGGAGAAGGTTATTGGAACTGAGCACACTGACATTCTTCGTGTTCCATTCAGGAGTGAGAATGGTATCCTCCGCAAGTGGATCTCCCGTTTTGATG +3364
 V  G  T  T  C  G  Q  R  V  E  K  V  I  G  T  E  H  T  D  I  L  R  V  P  F  R  S  E  N  G  I  L  R  K  W  I  S  R  F  D     .

TCTGGCCATTCCTGGAAACATACACTGAGGTAAATGGTTTATCTGACAACGTCCTTCAGAGCTGAGCACAAACCATGAATTCTGAGAGCTCATCTTTTTTTTTTTGGTTGCAGGATGTTG +3484
 V  W  P  F  L  E  T  Y  T  E <-----------------------------------IVS-9-----------------------------------> D  V

CAAACGAAATTATGAGGGAAATGCAAGCCAAACCTGATCTCATCATTGGCAATTACAGTGATGGAAACCTTGTTGCCACTCTGCTGGCTCACAAATTAGGAGTTACCCAGGTATGTTGG +3604
 A  N  E  I  M  R  E  M  Q  A  K  P  D  L  I  I  G  N  Y  S  D  G  N  L  V  A  T  L  L  A  H  K  L  G  V  T  Q <---------

TTGTAAACCACTCCTTTAAATATTTCTTTCATTAAGTCAATTACTTATGATTTCTCTTCTACATGGATACAGTGTACCATTGCTCATGCCTTGGAGAAAACCAAATACCCCAACTCAGAC +3724
--------------------------------IVS-10------------------------------> C  T  I  A  H  A  L  E  K  T  K  Y  P  N  S  D

ATATACTTGGACAAGTTTGACAGCCAGTACCACTTCTCATGCCAATTCACTGCTGATCTTATCGCCATGAATCACACTGATTTCATCATCACCAGTACATTCCAAGAAATTGCTGGAAGG +3844
 I  Y  L  D  K  F  D  S  Q  Y  H  F  S  C  Q  F  T  A  D  L  I  A  M  N  H  T  D  F  I  I  T  S  T  F  Q  E  I  A  G  S<

TAAAAATTTTCTAAAGAGCAGTAGAATATTGCTGCGTTGTGCCAACACCTAAACGGGTATCTGATGTTTGTTTCCTCTGTTTCTTCAGCAAGGACACTGTGGGGCAGTATGAATCACACAT +3964
-------------------------------------------IVS-11-------------------------------------------------> K  D  T  V  G  Q  Y  E  S  H  I

TGCATTCACCCTTCCTGGGCTTTACCGAGTTGTGCATGGCATAGATGTTTTTGATCCCAAGTTCAACATTGTCTCTCCTGGAGCTGACATGAGTGTCTACTTCCCGTACACCGAGGCTGA +4084
 A  F  T  L  P  G  L  Y  R  V  H  G  I  D  V  F  D  P  K  F  N  I  V  S  P  G  A  D  M  S  V  Y  F  P  Y  T  E  A  D

CAAGAGGCTCACTGCTTTCCACCCTGAAATTGAGGAGCTTCTCTACAGTGAAGTCGAGAACGATGAACACAAGTGAGTACTGCTGGTTCCGTCCCTGTTATTCACCTACCATGTA +4204
 K  R  L  T  A  F  H  P  E  I  E  E  L  L  Y  S  E  V  E  N  D  E  H  K<-----------------------------------------------

ACAAATAATTAAACACTGATTTGTGTGTTTGGTCTAGCTTTGGCTCTTCAGTCTACTAAGTGCCATGCACTAGAACTTTGTAGATTTGAACTGCAATTCTTAATTTTCTCACTCCTTAGT +4324
------------------------------------------------------------------------------------------------------IVS-12------------

AGACCACAATAATAATTATTGTGTCTGACCCAGCTAGTTTACAATAGTGACTTCAATGCTTGTTTTTTCCCCTTGAACAATTGTTTACTTTCTAGCAGTGGAGTTTGGCATCTAAACTTG +4444
-------------------------------------------------------------------------------------------------------------------------

AATACACGCTGTTCTTTTTTGTGTATAAAAATGTTTATTTAATGTTCTGCAGGTTTGTATTGAAGGACAAGAACAAGCCAATCATCTTCTCCATGGCTCGTCTTGACCGAGTGAAGAACAT +4564
--------------------------------------------------> F  V  L  K  D  K  N  K  P  I  I  F  S  M  A  R  L  D  R  V  K  N  M

GACAGGTCTGGTTGAGATGTATGGTAAGAATGCACATCTCAGGGATTTGGCAAACCTTGTGATTGTTTGTGGTGACCACGGCAATCAGTCCAAGGACAGGGAGGAGCAGGCTGAGTTCAA +4684
 T  G  L  V  E  M  Y  G  K  N  A  H  L  R  D  L  A  N  L  V  I  V  C  G  D  H  G  N  Q  S  K  D  R  E  E  Q  A  E  F  K

GAAGATGTACGGTCTCATTGACCAGTACAAGTTGAAGGGGCATATCCGCTGGATCTCAGCTCAGATGAACCGTGTTCGTAACGGGGAGTTGTACCGATACATTTGTGACACCAAGGGAGT +4804
 K  M  Y  G  L  I  D  Q  Y  K  L  K  G  H  I  R  W  I  S  A  Q  M  N  R  V  R  N  G  E  L  Y  R  Y  I  C  D  T  K  G  V

CTTTGTCCAGGTATATACATACGCTGTCATATTCTACTGCTATATTTCTATTCTGAGCTTCCAGTAGCAAATGATTTCTACTATTTGTAAAGAACAGCCTTGTGCTTGAGACTAAAACTG +4924
 F  V  Q <-----------------------------------------------IVS-13------------------------------------------------------

CTTTTTAAAATTGTTTTTTAGCCTGCATTCTATGAAGCGTTTGGTCTGACTGTCATCGAAGCCATGACATGTGGTTTGCCAACAATCGCAACATGCCATGGTGGCCCTGCTGAGATTATTG +5044
------------------->P  A  F  Y  E  A  F  G  L  T  V  I  E  A  M  T  C  G  L  P  T  I  A  T  C  H  G  G  P  A  E  I  I

TTGATGGGGTGTCTGGTCTGCACATTGATCCTTACCACAGTGACAAGGCTGCTGATATCTTGGTCAACTTCTTTGAGAAGTGCAAGCAGGATTCAACCTACTGGGACAATATTTCACAGG +5164
 V  D  G  V  S  G  L  H  I  D  P  Y  H  S  D  K  A  A  D  I  L  V  N  F  F  E  K  C  K  Q  D  S  T  Y  W  D  N  I  S  Q

GAGGTCTGCAGAGGATTTACGAGAAGTATGTATTTCATGATCTGCAAATATATAACATTTTTTTTTTTACGATTGACATATAACTGTAGAAATCATGTGTCCAATCTATATTGATCCTACA +5284
 G  G  L  Q  R  I  Y  E  K<------------------------------------------------IVS-14-----------------------------------------

AATTTGGACTTGTCAGGTACACCTGGAAGCTGTACTCTGAGAGGCTGATGACCTTGACTGGTGTATACGGATTCTGGAAGTACGTAAGCAACCTTGAGAGGCGCGAGACTCGCCGTTACA +5404
--------------->Y  T  W  K  L  Y  S  E  R  L  M  T  L  T  G  V  Y  G  F  W  K  Y  V  S  N  L  E  R  R  E  T  R  R  Y

TTGAGATGTTCTATGCTGTGAAATACCGCAGCCTGGTAAGTTACTCTCGCTGTTCAACAGCTTGTTAATTTCAGTATTATTGCACTGAGCATATCTCTGAATTCTGTTGCTACATTTCTT +5524
 I  E  M  F  Y  A  L  K  Y  R  S  L <----------------------------------------IVS-15------------------------------

TGTTGAATCCAGGCCAGCGCCGTCCCATTGGCTGTCGATGGAGAGAGCACATCCAAGTAATGGAGGGGAAAATATGCATCTTCAGCAGGAGAAGCCGTCAGCTGCATTCGAATTTGATAA +5644
---------->A  S  A  V  P  L  A  V  D  G  E  S  T  S  K

TTTTCTGTAGTTGTCATTTGGCATCCATGTTTGCCATGGATGTGTACTATGTCTAAGGTTTCAGTACTTTTGCGAGATTTGGGCAGTGCTTGCCTCAAATAAACGCCGGTTCCTGGTGT +5764
                                                                                        PolyA+

TTATCGTTCAGAATTCAATGCAATGTTTTATTGCCGCGCTTCCTCAATCTGTCATGTTGAATGACCACTCCCTAGTGCCCTACTGTTAAAAAAAGGTTGGTTCAAAGCCATGGATGAACCC +5884
                                   ^

AACATTTTTGTTACCTTGTTCTTTTCAGAATTTCCTTAAGAGTTGTGTGATAGTAGGCCCAGTGCGACCAAAATGTTGGGGTTCCTTGAGTTTTCAGTATAAAAATTTCTGATATTAAATGCAG +6004

ATCTTGGCTCAGTACAGAGTCTTGCAAACCATGTGGTTGCATCTTTTCTACTGCTGTAGTAGATGAAACTGAAGATACTACTTATTACTCTTTAGACAAAGTTTCATCATATTCCCAAAA +6124

TCTTGGTCTAGGATTGGGAGACATCAAGAAGATATAAACTTCACACTCAGGCCCTGTTTGGCAACCCTGAATAATTTTCTTTTGTGTAATGCAAATGCACCTGTAAATTTGGGTGACCCA +6244

AAGATGCCAGTGAAATATAAATCCCATGTAGCCAAACTGTTTGTTAGAAGAACTAGAACACTCAGTAGAGGTTTTAGCTAATTACAGTTCATTGCTGATCAAGGTTGCATCAGACATAGA +6364

AGTCAGACCGAAACATCAATTTGGAGCTATAATACAAAATTCGATAAACCAATACGAAATTCAAATAAAATAACATTGTTCATCAAAATTTCTGGCCAGAACCATCCAAGAGCAGTTGTA +6484

AGCGAGA
```

References

1. Werr W, Frommer W-B, Maas C, Starlinger P: Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. EMBO J 4: 1373–1380 (1985).
2. Chourey P: Genetic control of sucrose synthetase in maize endosperm. Mol Gen Genet 184: 372–376 (1981).
3. Echt C, Chourey P: A comparison of two sucrose synthetase isozymes from normal and *Shrunken-1* maize. Plant Physiol 79: 530–536 (1985).
4. Chourey P, Taliercio EW, Kane EJ: Tissue-specific expression and anaerobically induced posttranscriptional modulation of sucrose synthase genes in *Sorghum bicolor* M. Plant Physiol 96: 485–490 (1991).
5. Yu W-P, Wang A-Y, Juang R-H, Sung H-Y, Su J-C: Isolation and sequences of rice sucrose synthase cDNA and genomic DNA. Plant Mol Biol 18: 139–142 (1992).
6. Maier L-G, Brown JWS, Toloczyki C, Feix G: Binding of a nuclear factor to a consensus sequence in the 5′ flanking region of zein gene from maize. EMBO J 6: 17–22 (1987).

Fig. 3. Nucleotide sequence of the rice sucrose synthase gene *RSs1* and its flanking regions, with the deduced amino acid sequence for the sucrose synthase polypeptide. The full transcription region was determined by sequence alignment to the maize *Sh-1* sucrose synthase gene; the putative transcription initiation and polyadenylation sites are indicated by ' ^ '. Significant features including start and stop codons, CAAT and TATA boxes, polyadenylation signal sequences, and direct repeats and palindromic sequences (PL-1,2,3) in the 5′ flanking region, are underlined.

Characterization and sequencing of cDNA clone encoding the phloem protein PP2 of *Cucurbita pepo*

Ming-Bo Wang, Donald Boulter, and John A. Gatehouse*
Department of Biological Sciences, University of Durham, South Road, Durham DHI 3LE, UK (author for correspondence)*

Abstract

Direct N-terminal amino acid sequencing of the phloem protein 2 (PP2) from 3-month old *Cucurbita pepo* L. (pumpkin), purified by SDS-PAGE and blotted onto PVDF membrane, showed that the protein had a blocked N-terminus. However, after *in situ* cleavage of the polypeptide in a gel slice by cyanogen bromide, 75 residues of sequence on two cyanogen bromide fragments were determined. An oligonucleotide probe based on this amino acid sequence was used to screen a cDNA library, constructed from mRNA of 3–5-day old seedling hypocotyls, in λZAP II. A cDNA clone (p11A) predicted an amino acid sequence of 218 residues, in full agreement with the sequences determined for two CNBr fragments of PP2, and suggests that the N-terminus of the protein is a blocked methionine residue which is cleaved off by CNBr. Two additional cDNA clones were sequenced but no heterogeneity in the PP2 sequence was found. The deduced amino acid sequence of *C. pepo* differs in nine residues from the recently published sequence of *Cucurbita maxima* (Bostwick *et al.*, Plant Cell 4 (1992) 1539–1548). Southern blot showed that PP2 is encoded by a gene family with a relatively large number of members (estimated as 7–15 per haploid genome).

Introduction

The presence of proteins in the phloem sap of plants has been known for many years [8]; phloem protein concentrations are generally low, of the order of 0.3–2.2 mg/ml [12, 17, 20], but are much higher in *Cucurbita* spp., where concentrations as high as 110 mg/ml have been reported [13]. The cucurbits are thus particularly suitable subjects for the study of both phloem proteins and their encoding genes.

Besides a number of enzymes, such as sucrose synthase [20, 25] and peroxidases [13], the phloem sap contains a small number of proteins present in relatively large amount. In the cucurbits, these have been classified into two families,

160

designated PP1 and PP2, and are collectively referred to as P-proteins [12, 33]. The smaller of the P-proteins, PP2, is a chitin-binding lectin with a subunit molecular weight of ca. 25 000 [1, 15, 34, 35]. On exposure of phloem exudate to air, it forms a gel (with PP1) by oxidative cross-linking of cysteine residues, and is thus assumed to act as a sealing mechanism for wound sites [34, 42]. Other functions have been suggested for this protein, such as involvement in translocation [23, 26, 38, 45, 47], providing a structural framework on which untranslocated phloem components can be anchored [1, 35, 42], or as a defence against pathogens (on the basis of the carbohydrate-binding specificity) [1, 15, 34, 35].

Although characterization of the PP2 lectin at the protein level has been carried out [1, 15, 34, 35], no directly determined protein sequence data have been reported. Sequence data from cDNA species or genomic clones encoding PP2 are also unavailable. The cloning of a cDNA encoding a cucurbit P-protein cDNA, assumed to be PP2, was reported several years ago [39], but no sequence data were given. The mRNA from which this cDNA was derived was reported to be most abundant in young seedlings. In a more recent paper, the deduced amino acid sequence for cucurbit PP2 was given as part of a paper demonstrating the phloem-specific expression of its encoding gene [6]; however, once again, no sequence data for either cDNA or genomic clones was given. The present paper reports the sequence of a full-length cDNA encoding the PP2 lectin protein, identified on the basis of partial sequence data obtained from purified PP2 polypeptide, and gives preliminary characterization of the corresponding mRNA species and genes. After submission of this paper, the sequence of a cDNA encoding PP2 from *Cucurbita maxima* was published [7].

Materials and methods

Growth of plants

Plants of *Cucurbita pepo* L. cv. Autumn Gold (pumpkin) were grown from seeds (Suttons

Seeds, Torquay, UK) in Levington potting compost (Fisons, Loughborough, UK) under growth room conditions (25 °C, 70–80% RH, 14-hour photoperiod). One day after the hypocotyls emerged, the seedlings were regarded as 1 day old.

SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Phloem exudate was collected by cutting the stems of 3-month old plants and transferring the exudate with a pipette directly into an equal volume of 2× SDS-PAGE sample buffer [21] containing 5% mercaptoethanol. SDS-PAGE was carried out according to Moos *et al.* [32] and Hunkapiller *et al.* [18]. The separating gel (12% acrylamide), prepared according to Laemmli [21], was allowed to 'cure' for at least 14 h prior to use, and was subjected to pre-electrophoresis at 150–200 V for 1 h. The running buffer was always supplemented with 0.1 mM sodium thioglycolate. Before loading, the exudate sample was heated at 60 °C for 10–20 min. Electrophoresis was conducted at 150–200 V.

In situ *cleavage of PP2 protein in gel slice with cyanogen bromide (CNBr) vapour*

This was done by a modification of the procedure by Zingde *et al.* [48]. After completion of the electrophoresis, the SDS-polyacrylamide gel was stained for 5 min in 0.5% Coomassie blue in acetic acid isopropanol water (1:3:6, v/v/v) and destained briefly in 50% methanol and 10% acetic acid until the protein bands were just visible. The band corresponding to PP2 was cut out and lyophilized in a small glass container, which was then placed inside a small vacuum flask containing 1 ml of 50% trifluoroacetic acid and 20 mg of cyanogen bromide at its bottom. The flask was flushed with argon gas, evacuated briefly until the liquid just started to boil, and then kept sealed in the dark at room temperature for 24 h.

Electroblotting and sequencing of the cleaved PP2 peptides

The gel slice treated with CNBr vapour was removed from the vacuum flask and lyophilized, and then suspended in several changes of SDS-PAGE sample buffer [21] until the bromophenol blue dye in the buffer remained blue. The gel slice was finally suspended in 0.4 ml of the sample buffer, heated for 10 min at 60 °C, loaded directly into the well of a 17% SDS-PAGE slab gel, overlaid with sample buffer, and electrophoresed as above.

Electroblotting was conducted according to LeGendre and Matsudaira [24] using a semi-dry electroblotting apparatus (ATTO Corporation, Japan). The polyvinylidene fluoride (PVDF) membranes used were ProBlott from Applied Biosystems. The blotting proceeded for 1 h at 200 mA constant current. After blotting, the membrane was briefly stained (about 1 min) in 0.1% Coomassie blue in 50% methanol and destained for about 15 min as described above, washed with three changes of distilled water for a period of 15 min and air-dried. The membrane can be stored at −20 °C at this stage. Stained peptide bands were excised with a razor blade and directly sequenced on an Applied Biosystems Gas Phase Sequencer (model 470A) by sequential Edman degradation.

Isolation of poly(A)+ RNA and construction of cDNA library

Total RNA was extracted from 3- to 5-day old hypocotyls by the method described by de Vries *et al.* [9]. Poly(A)+ RNA was purified from 1 mg of the total RNA using the PolyATract mRNA isolation System III (Promega). Synthesis of double-strand cDNA from the poly(A)+ RNA and subsequent cloning into the Uni-ZAP XR vector were carried out by using the ZAP-cDNA synthesis kit (Stratagene). The resulting recombinant DNA was packaged with the Gigapack II Gold packaging extract (Stratagene).

Screening the library

A degenerate oligonucleotide, 5'-GC(C/T)TC(C/T)TT(C/T)TC(C/T)TT(A/G)TT(A/G)TCCAT-3' complementary to the DNA coding sequence predicted by amino acid sequence MDNKEKEA of the 15 kDa CNBr fragment of PP2 (Fig. 3), was synthesized on an Applied Biosystems model 381A DNA synthesizer. This oligonucleotide was end-labelled with [γ-^{32}P]ATP [27] and used as probe for the screening. Ca. 38000 plaques of the library grown in SURE *Escherichia coli* host cells were screened. The nitrocellulose membranes, prepared according to Slightom and Drong [41], were incubated for 5 h at 40 °C in 25 ml of prehybridization buffer (6 × SSC [27], 2 × Denhardt's solution [27], 0.25% SDS, 100 μg/ml denatured herring sperm DNA). The end-labelled oligo probe was heated for 2 min in a boiling water bath, chilled on ice, and then added into the prehybridization solution. The membranes were incubated for 18 h at 40 °C. After hybridization, the membranes were washed with four changes of washing buffer (6 × SSC, 0.1% SDS) at room temperature for a period of 40 min. The hybridization was visualized by autoradiography. Plaques corresponding to strong hybridization signals were purified by a further round of screening under the same hybridization and washing conditions.

Nucleotide sequencing

The purified positive phages were subjected to *in vivo* excision to form insert-containing pBluescript phagemids. This was carried out using the ExAssist helper phage according to the procedure in the Stratagene's ZAP-cDNA synthesis kit manual. The resulting phagemids were first analysed by restriction endonuclease digestion. Selected clones were further analysed by DNA sequencing. Plasmid DNA was used as template for sequencing by dideoxy chain termination method [36] on an Applied Biosystems model 370A DNA sequencer, using the Taq Dye Primer Cycle sequencing kit (Applied Biosystems). The reported

sequence was fully determined on overlapping fragments on both strands of the DNA.

Northern and Southern blotting

Total RNA (40 μg) from 3- to 5-day old hypocotyls was separated by electrophoresis on a formaldehyde gel (1.2% agarose) [27]. Genomic DNA was isolated from 3- to 5-day old pumpkin seedlings [10]. Portions of DNA (10 μg) were digested with various restriction endonucleases, under conditions as recommended by the manufacturers, and the resulting fragments were separated by electrophoresis on a 0.7% agarose gel. DNA and RNA were transferred to nitrocellulose membranes as described by Maniatis *et al.* [27]. The membranes were prehybridized for 3.5 h at 42 °C in 15 ml of hybridization solution (containing 50% formamide) [27]. The radioactive probe was prepared by random priming labelling of PP2 cDNA with [α-^{32}P]dCTP [14]. Hybridization was carried out at 42 °C for 17 h. The RNA blot was washed in 1 × SSC, 0.1% SDS at room temperature for 20 min and in 0.2 × SSC, 0.1% SDS at 65–68 °C for 50 min. The DNA blot was rinsed briefly in 2 × SSC, 0.5% SDS and then washed with 2 × SSC, 0.1% SDS at room temperature for 15 min, in 0.1 × SSC, 0.5% SDS at 37 °C for 30 min and finally in the same buffer at 65 °C for 30 min. The RNA blot and the DNA blot were both exposed for 4 and 7 h, respectively, at −70 °C with intensifying screen.

Results

Sequencing of PP2 protein

The PP2 protein polypeptide was purified by SDS-PAGE of phloem exudate from *C. pepo* plant 3 months of age, after initial experiments had shown that conventional protein purification techniques were hindered by gelling of the protein on contact with air, and that SDS-PAGE was consequently the quickest and most efficient method to obtain small amounts of pure material. A typical separation is shown in Fig. 1. Initial
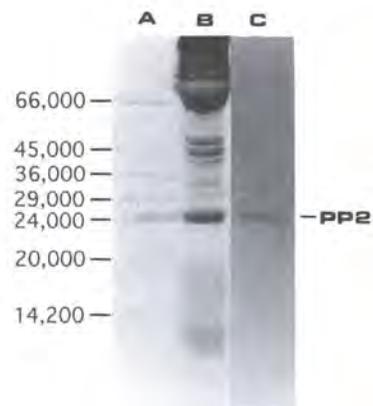
Fig. 1. SDS-PAGE analysis of pumpkin phloem exudate and purified PP2 protein. Gel is 12% acrylamide. Track A, protein markers (bovine albumin, 66000; ovalbumin, 45000; glyceraldehyde-3-phosphate dehydrogenase, 36000; carbonic anhydrase, 29000; trypsinogen (bovine pancreas), 24000; trypsin inhibitor (soybean), 20000; α-lactalbumin, 14200). Track B, 10 μl of pumpkin phloem exudate. Track C, PP2 protein purified by SDS-PAGE and electrophoretic elution. Proteins were stained with Coomassie blue.

attempts to directly sequence the 24 kDa PP2 polypeptide, either purified by electrophoretic elution from a polyacrylamide gel, or blotted onto PVDF membrane were unsuccessful, because the N-terminus of the protein was totally blocked to Edman degradation. No amino acid residues were obtained in amounts above background, demonstrating that the purified PP2 polypeptide was not contaminated with other sequenceable polypeptides. The PP2 polypeptide blotted onto PVDF membrane could be efficiently cleaved with cyanogen bromide vapor, generating two sequenceable peptides, which gave 15 residues of mixed sequence data (not presented). However, elution of these cleaved peptides from the membrane, and separation by SDS-PAGE or HPLC on a reversed-phase column was not successful, due to low recovery (less than 10%, according to analysis of amino acid residues by the automated sequencer), making it difficult to distinguish sequence data from background. *In situ* cleavage of PP2 in gel slices by CNBr vapour proved a successful technique; the cleaved polypeptides produced were separated by a second SDS-PAGE step and blotted onto PVDF membrane (Fig. 2).

Fig. 2. Electroblot of CNBr-cleaved PP2 peptides. Pumpkin phloem exudate (50 μl) was electrophoresed in a 12% SDS-polyacrylamide gel. The PP2 band was sliced and treated with CNBr vapour. The cleaved peptides were separated on a 17% SDS-polyacrylamide gel and electroblotted onto a polyvinylidene fluoride (PVDF) membrane (Track B). Track A, protein markers. Proteins were stained with Coomassie blue.

Two peptides, of about 14 and 15 kDa, were obtained, and both were successfully sequenced; the

sequences obtained are shown in Fig. 3. The expected third peptide, corresponding to the blocked N-terminus, was not seen as a band on SDS-PAGE after CNBr cleavage, suggesting that its size is small. The predicted amino acid sequence of PP2, from cDNA clones, suggests that this third peptide is a single methionine residue (q.v.). A 28 kDa band on the PVDF blot was uncleaved PP2. Figure 2 shows that there are other faint high-molecular-weight bands on the PVDF blot, which must represent oligomers of PP2, since the PP2 polypeptide had been separated from other proteins by being excised out of a gel before being treated with CNBr. Interestingly, the pattern of these high-molecular-weight bands in Fig. 2 was similar to that of crude phloem exudate proteins when the latter were separated and blotted under same conditions (data not shown). Other workers also found that PP2 could be oxidized to polymeric material [34].

Isolation of cDNA clones and nucleotide sequencing

Poly(A)$^+$ RNA was isolated from 1 mg of total RNA extracted from 3–5 day old hypocotyls of

15 kd peptide:

```
1                    10                   20                   30
D-N-K-E-K-E-A-R-E-K-L-G-G-E-V-K-L-G-H-X-L-D-V-I-L-K-N-A-D-V-

31              40
A-L-X-Y-P-S-F-L-K-L-Y-D-
```

14 kd peptide:

```
1                    10                   20                   30
L-S-P-N-V-V-Y-E-V-A-L-Q-V-Q-L-N-S-R-A-S-G-W-N-A-P-L-N-I-E-L-

31
K-K-P-D-G-
```

Fig. 3. N-terminal amino acid sequences of purified peptides produced by CNBr-cleavage of PP2 polypeptide. X = uncertain residue.

Cucurbita pepo, and used as a template for cDNA synthesis. A library containing 9×10^5 independent clones in the lambda phage vector ZAP II was obtained using one-twelfth of the resulting dscDNA. Ca. 38 000 recombinants of this primary λZAP cDNA library were screened with a fully degenerate oligonucleotide probe corresponding to amino acids 1–7, and the implied methionine residue before amino acid 1, of the 15 kDa CNBr fragment of PP2. Ca. 150 plaques (0.4%) gave strong hybridization signals.

After plaque purification, positive clones were subjected to *in vivo* excision to form insert-containing pBluescript phagemids. Two clones, p9A and p11A, were fully sequenced; they contained identical coding and non-coding sequence

(Fig. 4). Another clone, which contained a larger insert (2.2 kb; p27A), was also sequenced. The insert in this clone contained the entire PP2 cDNA (identical in sequence to the inserts of p9A and p11A) fused with a partial cDNA sequence, which was identified on the basis of sequence homology to entries in the EMBL database as encoding endo-β-1,4-glucanase precursor (data not presented).

The nucleotide sequences of the PP2 cDNA clones contain a region (bases 19–41) exactly complementary to one of the oligonucleotides in the degenerate probe. About 840 bases of sequence were determined, plus a poly(A) tail of variable length; the predicted mRNA size is thus in agreement with the species observed on the

```
AAAAAAGAAAGCGCTATAATGGACAACAAAGAGAAGGAAGCGAGAGAGAAATTAGGAGGA     60
                     M  D  N  K  E  K  E  A  R  E  K  L  G  G    14

GAAGTGAAGCTTGGTCATTGCTTGGATGTTATTTTGAAGAATGCTGACGTAGCACTGCAC    120
 E  V  K  L  G  H  C  L  D  V  I  L  K  N  A  D  V  A  L  H     34

TATCCCTCCTTCCTTAAGCTTTATGACCAACTTGTTGCTGGGATCCTCTTGAACAAGGGA    180
 Y  P  S  F  L  K  L  Y  D  Q  L  V  A  G  I  L  L  N  K  G     54

GCTATAAAGTACATCTTTGATAAGAAGTTAAACAGCCATTGGTACTTTATATTTGCGAGA    240
 A  I  K  Y  I  F  D  K  K  L  N  S  H  W  Y  F  I  F  A  R     74

GCTCTCTCAATCGCCTGGATTGAAGATAAAAGATACTGGAAATGGGGATCCTGTGGCAAT    300
 A  L  S  I  A  W  I  E  D  K  R  Y  W  K  W  G  S  C  G  N     94

AGCGAAGTTGCAGAACTTATTGAAGTATCTTGGCTGAACATTCGTGGAAAGATCAATGAG    360
 S  E  V  A  E  L  I  E  V  S  W  L  N  I  R  G  K  I  N  E    114

TCTATGCTCTCACCAAATGTTGTGTATGAAGTAGCACTTCAGGTACAGCTAAATAGTAGA    420
 S  M  L  S  P  N  V  V  Y  E  V  A  L  Q  V  Q  L  N  S  R    134

GCCTCCGGGTGGAATGCTCCTTTGAACATCGAATTGAAGAAGCCAGATGGGAGCAAGATA    480
 A  S  G  W  N  A  P  L  N  I  E  L  K  K  P  D  G  S  K  I    154

GTGCGCCAGGAATGCCTGTTGGGGAAGCCACAAAACCAGTGGTTTGAGATTGTTGTTGAG    540
 V  R  Q  E  C  L  L  G  K  P  Q  N  Q  W  F  E  I  V  V  E    174

TTCAAGGTAGGTAACCATGGCTGTGGAAGTAGTGGCGAGATCGAGTTTTCCTTTTTTGAA    600
 F  K  V  G  N  H  G  C  G  S  S  G  E  I  E  F  S  F  F  E    194

CATGGAGGGCATTGGAAGAGAGGGCTGCTCGTGAAAGGTGTTCGGATTGGAGCAAAGGGA    660
 H  G  G  H  W  K  R  G  L  L  V  K  G  V  R  I  G  A  K  G    214

TGTGGTTGCTCATGATCAAAATAGTCTTGGTGTACACAATCTCTCTTTCTCAAACTCAAA    720
 C  G  C  S  *

GTACACTTATTTTGATGTTGAGGAACGAGAGTTTGTGTTATAATCAAATATGAAAGAATG    780

TACTAGTATGAAACAATAAATAATAGACCTTTTGCTTACGGGCAAGAAGGTTAAGTTTT (A)ₙ
```

Fig. 4. Nucleotide sequence and deduced amino acid sequence of PP2 cDNA. Sequences were determined from clone p11A. Amino acid sequence regions covered by PP2 CNBr peptide sequences (Fig. 3) are underlined. A potential polyadenylation signal sequence in the cDNA is also underlined. The asterisk marks the stop codon.

168

Although no significant sequence homology has been found between PP2 and the ribosomal proteins, comparison of them reveals striking resemblance at both protein and gene level. Weber et al. [47] found that P-protein shared many biochemical and immunological similarities with ribosomal proteins. As discussed earlier in this paper, the PP2 is represented by a multigene family which might contain one or only a few functional members and possibly some pseudogenes, similar to the mammalian ribosomal protein genes. Recent investigations on plant ribosomal protein gene structure and expression showed that these genes are preferentially expressed in young and proliferating tissues rather than in old and expanded tissues [5, 19, 22, 28], and generally present in a small multigene family [5, 19, 22]. The PP2 mRNA level was high in hypocotyls of 3–5 day old *Cucurbita pepo* seedlings but almost undetectable in the stems and leaves of 3-month old plants when probed with the degenerate oligonucleotide probe (data not shown), which is similar to the plant ribosomal gene [28]. Sham and Northcote [39] reported that the total PP2 mRNA level increased in the first 4 days of seedling growth but decreased to lower level in older seedlings, resembling Joanin et al.'s observation [19] that the maize cytoplasmic ribosomal protein S13 mRNA level in germinating seeds reached a maximum level after 4 days of incubation.

From the above discussion it seems that P-proteins are closely related to ribosomal proteins. It has long been observed that the origin of P-proteins was closely linked to the presence of helical free polysomes and the formation of P-protein filaments coincided with the disappearance of ribosomes [3]. However, the significance of this remains unknown. Weber et al. [47] suggest that P-proteins are not specially synthesized during phloem differentiation but are aggregation of a related class of pre-existing, relatively stable, cross-linked proteins. This view does not necessarily imply that P-proteins are intracellular debris without a function in translocation [47]. The function of P-proteins could possibly be elucidated by expression of sense or anti-sense PP2 gene in transgenic plants.

References

1. Allen AK: A lectin from the exudate of the fruit of the vegetable marrow (*Cucurbita pepo*) that has specificity for β-1,4-linked N-acetylglucosamine oligosaccharides. Biochem J 183: 133–137 (1979).
2. Batra SK, Metzgar RS, Hollingsworth MA: Molecular cloning and sequence analysis of the human ribosomal protein S16. J Biol Chem 266: 6830–6833 (1991).
3. Behnke H-D: Comparative ultrastructural investigations of angiosperm sieve elements: aspects of the origin and early development of P-protein. Z Pflanzenphysiol 74: 22–34 (1974).
4. Beintema J, Peumans WJ: The primary structure of stinging nettle (*Urtica dioica*) agglutinin. FEBS Lett 299: 131–134 (1992).
5. Bonham-Smith PC, Oancia TL, Moloney MM: Cytoplasmic ribosomal protein S15a from *Brassica napus*: molecular cloning and developmental expression in mitotically active tissues. Plant Mol Biol 18: 909–919 (1992).
6. Bostwick DE, Dannenhoffer JM, Skaggs MI, Lister RM, Larkins BA, Thompson GA: Pumpkin phloem lectin genes are specifically expressed in companion cells. Plant Cell 4: 1539–1548 (1992).
7. Bostwick DE, Thompson GA: Nucleotide sequence of a pumpkin phloem lectin cDNA. Plant Physiol 102: 693–694 (1993).
8. Cronshaw J: P-proteins. In: Aronoff S, Dainty J, Gorham PR, Srivastava LM, Swanson CA (eds) Phloem Transport, pp. 79–127. Plenum Press, London (1974).
9. De Vries S, Hoge H, Bisseling T: Isolation of total and polysomal RNA from plant tissues. In: Gelvin SB, Schilperoort RA, Verma DPS (eds) Plant Molecular Biology Manual, pp. B6/1–B6/13. Kluwer Academic Publishers, Dordrecht (1988).
10. Doyle JJ, Doyle JL: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bulletin 19: 11–15 (1987).

11. Dudov KP, Perry RP: The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. Cell 37: 457–468 (1984).

12. Esau K, Cronshaw J: Tubular components in cells of healthy and tobacco mosaic virus-infected *Nicotiana*. Virology 33: 26–35 (1967).

13. Eschrich W, Evert RF, Heyer W: Proteins of the sieve-tube exudate of *Cucurbita maxima*. Planta 100: 208–221 (1971).

14. Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13 (1983).

15. Gietl C, Kauss H, Ziegler H: Affinity chromatography of a lectin from *Robinia pseudoacacia* L. and demonstration of lectins in sieve-tube sap from other tree species. Planta 144: 367–371 (1979).

16. Graham JS, Hall G, Pearce G, Ryan CA: Regulation of synthesis of proteinase inhibitors I and II mRNAs in leaves of wounded tomato plants. Planta 169: 399–405 (1986).

17. Hall SM, Baker DA: The chemical composition of *Ricinus* phloem exudate. Planta 106: 131–140 (1972).

18. Hunkapiller MW, Lujan E, Ostrander F, Hood LE: Isolation of microgram quantities of protein from polyacrylamide gels for amino acid sequence analysis. In: Hirs CHW, Timasheff SN (eds) Methods in Enzymology, vol. 91, pp. 227–237. Academic Press, London (1983).

19. Joanin P, Gigot C, Philipps G: cDNA nucleotide sequence and expression of a maize cytoplasmic ribosomal protein S13 gene. Plant Mol Biol 21: 701–704 (1993).

20. Kennecke M, Ziegler H, de Fekete MAR: Enzyme activities in the sieve sap of *Robinia pseudoacacia* L. and of other tree species. Planta 98: 330–356 (1971).

21. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).

22. Lebrun M, Freyssinet G: Nucleotide sequence and characterization of a maize cytoplasmic ribosomal protein S11 cDNA. Plant Mol Biol 17: 265–268 (1991).

23. Lee DR: Possible significance of filaments in sieve elements. Nature 235: 286 (1972).

24. LeGendre N, Matsudaira P: Electroblotting to PVDF membranes. In: Matsudaira PT (ed) A Practical Guide to Protein and Peptide Purification for Microsequencing, pp. 53–59. Academic Press, London (1989).

25. Lehmann J: Studies on the phloem exudate of *Cucurbita pepo* L. Planta 114: 51–61 (1973).

26. MacRobbie EAC: Phloem translocation. Facts and mechanisms: a comparative survey. Biol Rev 46: 429–481 (1971).

27. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

28. Marty I, Meyer Y: cDNA nucleotide sequence and expression of a tobacco cytoplasmic ribosomal protein L2 gene. Nucl Acid Res 20: 1517–1522 (1992).

29. Mayer A, Siegel NR, Schwartz AL, Ciechanover A: Degradation of proteins with acetylated amino termini by ubiquitin system. Science 244: 1480–1483 (1989).

30. Mohan R, Bajar AM, Kolattukudy PE: Induction of a tomato anionic peroxidase gene (*tap 1*) by wounding in transgenic tobacco and activation of *tap 1*/GUS and *tap 2*/GUS chimeric gene fusions in transgenic tobacco by wounding and pathogen attack. Plant Mol Biol 21: 341–354 (1993).

31. Möller W, Groene A, Terhorst C, Amons R: 50-S ribosomal proteins: purification and partial characterization of two acidic proteins, A1 and A2, isolated from 50-S ribosomes of *Escherichia coli*. Eur J Biochem 25: 5–12 (1972).

32. Moos MJ, Nguyen NY, Liu TY: Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an innert support. J Biol Chem 263: 6005–6008 (1988).

33. Read SM, Northcote DH: Chemical and immunological similarities between the phloem proteins of three genera of the Cucurbitaceae. Planta 158: 119–127 (1983).

34. Read SM, Northcote DH: Subunit structure and interactions of the phloem proteins of *Cucurbita maxima* (pumpkin). Eur J Biochem 134: 561–569 (1983).

35. Sabnis DD, Hart JW: The isolation and some properties of a lectin (haemagglutinin) from *Cucurbita* phloem exudate. Planta 142: 97–101 (1978).

36. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).

37. Schmidt J, Herfurth E, Subramanian AR: Purification and characterization of seven chloroplast ribosomal proteins: evidence that organelle ribosomal protein genes are functional and that NH_2-terminal processing occurs via multiple pathways in chloroplasts. Plant Mol Biol 20: 459–465 (1992).

38. Schulz A, Alosi MC, Sabnis DD, Park RB: A phloem-specific, lectin-like protein is located in pine sieve-element plastids by immunocytochemistry. Planta 179: 506–515 (1989).

39. Sham MH, Northcote DH: Transcription and translation of phloem protein (PP2) during phloem differentiation in *Cucurbita maxima*. Planta 170: 392–399 (1987).

40. Simmons CR, Litts JC, Huang N, Rodriguez RL: Structure of a rice β-glucanase gene regulated by ethylene, cytokinin, wounding, salicylic acid and fungal elicitors. Plant Mol Biol 18: 33–45 (1992).

41. Slightom JL, Drong RF: Procedures for constructing genomic clone banks. In: Gelvin SB, Schilperoort RA, Verma DPS (eds) Plant Molecular Biology Manual, pp. A8/1–A8/42. Kluwer Academic Publishers, Dordrecht (1988).

42. Smith LM, Sabnis DD, Johnson RPC: Immunocytochemical localization of phloem lectin from *Cucurbita*

170

maxima using peroxidase and colloidal-gold labels. Planta 170: 461–470 (1987).

43. Suzuki K, Wool IG: The primary structure of rat ribosomal protein L23a. J Biol Chem 268: 2755–2761 (1993).

44. Terhorst C, Möller W: The primary structure of an acidic protein from 50-S ribosomes of *Escherichia coli* is involved in GTP hydrolysis dependent on elongation factors G and T. Eur J Biochem 34: 138–152 (1973).

45. Thaine R: Movement of sugars through plants by cytoplasmic pumping. Nature 222: 873–875 (1969).

46. Wagner M, Perry RP: Characterization of the multigene family encoding the mouse S16 ribosomal protein: strategy for distinguishing an expressed gene from its processed pseudogene counterpart by an analysis of total genomic DNA. Mol Cell Biol 5: 3560–3576 (1985).

47. Weber C, Franke WW, Kartenbeck J: Structure and biochemistry of phloem-proteins isolated from *Cucurbita maxima*. Exp Cell Res 87: 79–106 (1974).

48. Zingde SM, Shirsat NV, Gothoskar BP: Peptide mapping of proteins in gel bands after partial cleavage with acidic cyanogen bromide vapor. Anal Biochem 155: 10–13 (1986).