FINAL REPORT to SUGAR RESEARCH AND DEVELOPMENT CORPORATION

Project code: UQ6S

Project title: Molecular tagging of economically important genes and promoters in sugarcane

Field of Research: The Cane Plant **Field Code:** 1

Organisation: The University of Queensland

Address: St. Lucia Qld 4072

Project Supervisor: Dr. R.G. Birch, B.Sc.Hons (Qld), M.S., Ph.D. (Hawaii)

Position: Senior Lecturer

Location: Department of Botany

Telephone: (07) 365 3347 **Facsimile:** (07) 365 1699

Commencement date: July 1990 **Completion date:** June 1992

Funding: SRDC 1990-91 $34,020
1991-92 $9,000

The University of Queensland provided the salary of the project supervisor, input by graduate students, including a scholarship, and costs of laboratory equipment and specialised facilities necessary for the project (estimated costs: 1990-91 $26,000; 1991-92 $40,000).

Project Objectives : The overall aim of this project was to develop techniques for the isolation of useful gene control sequences (promoters) in sugarcane. Early in the project we also undertook preliminary studies to lay some groundwork for techniques to tag and isolate useful genes in sugarcane. Specific objectives were:

1. Develop techniques for the isolation of specific sugarcane gene control sequences. This led us to evaluate the applicability of differential cDNA hybridisation to recover tissue specific genes from sugarcane, followed by genomic library screening to recover corresponding tissue-specific promoters.

2. Prepare genetic constructs for transfer into sugarcane to assist identification and isolation of useful genes and promoters in sugarcane. This led us to test the effectiveness of a range of well characterised foreign promoters and novel artificial promoters in sugarcane.

3. Probe the sugarcane genome for endogenous transposable elements homologous to those isolated from other Gramineae, to help evaluate the potential for transposon tagging of useful sugarcane genes.

All project objectives were achieved.

Non-Technical Summary

An important aspect of practical sugarcane genetic engineering work is the ability to tailor expression of introduced genes. This ensures that the desired gene product is produced where it is needed in the plant. For example we would aim to express novel genes for cane grub resistance in the sugarcane roots, but not in harvested portions of the crop. Tailored gene expression also reduces metabolic cost to the plant and helps avoid possible safety restraints, by blocking expression where the foreign gene product is not required in the plant.

The timing and strength of gene expression in plants is determined by regulatory sequences associated with each gene. The promoter region immediately adjacent to each gene typically exerts a major influence on the pattern of expression of that gene. Other more distant regulatory sequences known as enhancer elements can also influence the level and developmental or tissue specificity of expression.

The technology for isolation of promoters has advanced rapidly in the past few years. It is likely that some promoters isolated from other plants will prove useful when tested in sugarcane. In other cases, particularly when stringent control of a complex regulation pattern is required, it may be more effective to isolate the desired promoter from sugarcane. In this project we set out to develop techniques for the isolation and testing of economically important gene control sequences in sugarcane.

The project led to the identification of a set of foreign and artificial promoters which result in various levels of expression of introduced genes in sugarcane. These promoters will probably result in more or less continuous gene expression in most cell types, that is "constitutive" expression, and should serve all foreseeable needs for strong constitutive gene expression in sugarcane cells.

To obtain more specific patterns of gene expression, we set out to develop techniques for isolation of specific promoter types from sugarcane itself. This work showed that an experimental approach known as differential cDNA hybridisation can be used to recover copies of genes that are expressed in specific sugarcane tissues, and subsequently to recover promoter sequences regulating the corresponding gene expression pattern. The basic molecular methodology developed for other plant species proved applicable to sugarcane. However, because sugarcane is a genetically complex, heterozygous polyploid, some of the techniques are more demanding than in genetically simpler "model" plant species.

As a practical test of this methodology, we chose to isolate a root-specific gene and its promoter from sugarcane. This appears to have been successful, although final confirmation will require functional testing of the isolated promoter in a future project. If confirmed, the root-specific promoter will be of immediate interest to couple with genes for control of sugarcane root pests and diseases.

The methods developed in this project should be of continuing use for isolation of other desired promoter types from sugarcane. Several gene libraries from sugarcane were produced during the project, and are available for future use in a range of projects aiming to apply molecular approaches to sugarcane improvement. Preliminary results from the project also laid some groundwork for alternative approaches to the isolation of useful sugarcane genes and gene control sequences.

Technical Summary

Overview

Control of transgene expression is a basic requirement for practical genetic engineering of sugarcane. This requirement was addressed in work to develop artificial promoters for expression of introduced genes in sugarcane cells and to evaluate and apply differential cDNA hybridisation for isolating sugarcane promoters.

In collaborative work we have tested several promoter constructs for constitutive expression in sugarcane cells. Very strong expression was obtained using an artificial promoter (EMU) prepared by Dr David Last of CSIRO Plant Industry. This strong monocot promoter incorporates multiple enhancer elements from the maize AdhI gene and the *Agrobacterium* OCS promoter, an intron and a minimal segment of the CaMV 35S promoter. Carl Rathus prepared a set of constructs including different combinations of these elements, for quantitative determination of promoter strength in transient gene expression assays following electroporation into sugarcane protoplasts. Our results indicate that these available constructs should suffice for foreseeable needs for strong constitutive gene expression in sugarcane cells.

The applicability of a differential cDNA hybridisation approach for isolating specific sugarcane gene regulatory sequences was investigated by Tricia Franks, taking the isolation of the promoter sequence from a gene expressed primarily in sugarcane roots as a model. Using cDNA probes derived from sugarcane wounded shoot, shoot, wounded root and root mRNAs, 6000 clones from an amplified sugarcane cDNA library were screened for non-constitutive patterns of gene expression. No wound-inducible sequences were isolated, but a root specific clone and a shoot specific clone were selected. In northern analysis, the shoot specific clone hybridised shoot transcripts exclusively and strongly and the root specific clone detected low level shoot expression and high level root expression of a transcript which was approximately equal in length to the cDNA. In Southern analysis of sugarcane DNA probed with the root specific and shoot specific cDNAs, both probes hybridised multiple restriction fragments, consistent with the presence of various allelic forms of the genes in the polyploid genome. Furthermore, the probes readily detected extensive polymorphism amongst several sugarcane cultivars. The root specific transcript and its putative translation product have homology with metallothionein-like genes previously isolated from the roots of other plant species. An 813 bp promoter region was rapidly isolated from a genomic clone which was homologous to the root specific transcript. The promoter sequence contains short motifs which are common to other plant promoters, but experimental confirmation of promoter functionality is still required. Differential cDNA hybridisation has proved applicable to isolating moderately abundant, nonconstitutively expressed sequences from sugarcane. More research is required to determine if the complex polyploid genome imposes additional constraints to isolating other functional promoter sequences by this approach.

Differential cDNA hybridisation for isolating nonconstitutively expressed sugarcane genes

The differential cDNA hybridisation approach was found useful for isolating transcripts of genes which are expressed nonconstitutively in sugarcane. In practice, logistics of the screening process limited the approach to the detection of transcripts which appear at moderate to high levels in at least one of the tissues used for comparison. In addition, sensitivity of hybridisation with heterogeneous cDNA probes is so low that northern analysis is required to more accurately assess patterns of expression of selected clones. Two clones which were both apparently tissue specific were readily isolated from sugarcane by

differential cDNA hybridisation. Subsequent northern analysis found that one clone was absolutely shoot specific and the second was primarily root specific and expressed at low levels in shoots. With more painstaking or computer-assisted rescreening and reassessment of differential hybridisation patterns, the existing cDNA libraries would almost certainly permit isolation of sequences with moderate expression levels, and this may include wound inducible transcripts. To apply the approach to isolating lower copy sequences, it would be necessary to proceed via the technically demanding construction of subtractive cDNA libraries and application of subtractive probes.

The recovered primarily root specific cDNA clone was fortuitously near full-length and this facilitated the isolation of the upstream promoter sequence. However, shortened cDNA clones (which are characteristic of the existing libraries) need not be a barrier to recovery of promoter sequences from homologous genomic clones. An *in vitro* capped mRNA probe applied directly to a genomic homologue of a selected cDNA would quickly identify restriction fragments bearing likely gene transcription start sites (Shaw *et al.*, 1989).

From a technical perspective, methods previously applied in other plants for mRNA purification and production of screenable cDNA libraries were similarly applicable to sugarcane, which appeared to pose no unusual difficulties with tissue processing, nuclease levels or genetic complexity.

Constraints to isolation of sugarcane promoter sequences via differential cDNA hybridisation

In any species, recovery of functional promoter sequences from the genome presumes their linkage 5' to the transcription start site of the gene with the desired expression pattern. Remote enhancers contributing to the transcription response are not easily recovered by this route and their participation in directing gene expression could lead to an incorrect suggestion that cloned proximal sequences have specific promoter activity. With virtually no information regarding sugarcane molecular genetics available when the project commenced, it was anticipated that the large and complex polyploid sugarcane genome and heterozygosity of the crop would impose additional constraints to promoter sequence isolation compared to the genome of a simple diploid, homozygous species. For every isolated cDNA transcript, at least one ploidy number (unknown) of corresponding genes would be present in the genome and a heterozygous set of alleles may be present at each locus. Additionally the gene may be a member of a multigene family further extending the number of homologous loci. Southern analysis indicated that several genes in the sugarcane genome have homology to the primarily root specific transcript. Preliminary data from Southern analysis of sugarcane DNA probed first with the coding region and then with the cloned promoter sequence, suggests that the promoter regions of these homologous genes may also be conserved. Sequence comparison of promoters of these genes would indicate their functional relatedness. Isolation of these promoters would require screening the entire sugarcane genome, which is a feasible but formidable task.

Confirmation of functionality of the isolated putative root specific promoter in stably transformed sugarcane plants is now required. The success in isolating this likely promoter indicates that the large and complex polyploid sugarcane genome is not necessarily an obstacle to the recovery of functional promoters from sugarcane via the differential cDNA hybridisation approach. Analysis of additional sugarcane promoters is required in order to establish the frequency and extent of promoter divergence on homologous transcribed

sequences.

Other approaches for generating promoter sequences for use in sugarcane

Cis-acting wound responsive elements will soon be defined in heterologous systems (Keil *et al.*, 1989; Lamb and Dixon, 1990; Meier *et al.*, 1991; Sheng *et al.*, 1991; Katagiri and Chua, 1992) and function of these sequences may now be tested in sugarcane. Alternatively, wound responsive *cis* acting promoter elements could be isolated from the existing sugarcane genomic library using wound inducible sequences from other species (Bradshaw *et al.*, 1989; Logemann and Schell, 1989 and Rumeau *et al.*, 1990) as heterologous probes. Promoters of various function will also be readily isolated by this approach.

Now that gene transfer to sugarcane is possible, methods for promoter isolation using transgenic plants are relevant (Herman *et al.*, 1986; Teeri *et al.*, 1986; Koncz *et al.*, 1989; Claes *et al.*, 1991). Indications that low copy numbers of the transgene appear in sugarcane plants transformed via microprojectiles (Bower and Birch, 1992) will favour the success of "promoter-tagging" strategies.

Characterisation of the root specific promoter

Sequence analysis of the root specific promoter has allowed identification of regions likely to be involved in *trans* interactions to mediate initiation of transcription. Based on that information, a calculated deletion analysis of the promoter can proceed immediately. For example, using PCR, the existing three oligonucleotides (which were designed for sequencing purposes) will easily generate two deletion mutations which roughly divide the promoter in half.

The cloned sugarcane promoter adds to the infrequent monocot examples in a growing list of isolated plant promoters and its comparison to analogous sequences isolated from maize and pea may provide some insight into the mechanisms which distinguish monocot and dicot transcriptional control.

Prospects for engineering sugarcane promoters

Differential cDNA hybridisation proved efficient for isolation of a sugarcane promoter sequence which may direct root specific gene expression. More research is required to determine if this technique for recovery of promoters will be generally applicable and efficient in sugarcane. The isolated promoter may be applied to drive primarily root specific gene expression and is a candidate for modification to generate more precise and alternative patterns of transcriptional control. Characterisation of the root specific promoter is likely to reveal modules involved in specific patterns of gene expression. It may therefore be possible to modify the existing promoter, for example to eliminate the observed marginal levels of shoot expression, by deleting or mutating an element identified as shoot specific. Additionally, wound responsive elements may be added to the promoter. In principle it should be possible to generate a promoter of any desired function in this fashion, although much remains to be learned about the mechanisms for promoter module interactions before this approach to "designer promoters" becomes practical and reliable.

Constructs likely to drive constitutive gene expression in sugarcane

Several chimaeric promoter regions coupled to the *uid A* β-glucuronidase gene coding region were evaluated for expression strength following electroporation into sugarcane (monocot) and carrot (dicot) protoplasts. In transient gene expression assays, no significant difference was

found between the nopaline synthase promoter (NOP) and the Cauliflower mosaic virus 35S promoter (35SP) in sugarcane. However, in carrot the NOP construct resulted in 4-fold higher expression than the 35SP construct. Insertion of intron 1 from the maize alcohol dehydrogenase 1 gene (*Adh 1*) immediately after a promoter reduced expression in sugarcane cells. In carrot cells this intron reduced expression with NOP but increased or did not affect expression with 35SP. The octopine synthase (OCS) enhancer in either orientation increased expression, as did a duplication of 35SP in sugarcane. A promoter region containing multiple copies of both the OCS enhancer and the anaerobic regulatory element of *Adh1* proved 400-fold stronger than 35SP in sugarcane, but only equalled 35SP in carrot. These results show that various levels of expression of introduced genes in sugarcane or carrot cells can be achieved by using appropriate combinations of promoter and enhancer elements, and that significant differences exist between the responses of these species. This is consistent with previous observations of different responses to some gene regulatory elements in monocots and dicots, but sugarcane differs in some respects even from other species in the *Poaceae.*

Groundwork for sugarcane gene tagging

No sequences homologous to maize transposons Ac and Ds were detected using non-radioactive probes on Southern blots of sugarcane genomic DNA. Genetic constructs which allow detection of transposition by maize element Ac following introduction into sugarcane cells, based on activation of a selectable antibiotic resistance gene and a reporter gene "GUS", were provided by George Coupland of the Max-Planck Institute and Jean Finnegan of CSIRO Plant Industry. An experiment by Robert Bower and Carl Rathus to determine the time-course and frequency of Ac excision in sugarcane *vs* carrot cells showed no detectable excision in our transient assay system. Further work on this aspect was beyond the resources of the current project.

Origin of the project (UQ1S / UQ6S)

A new project application titled "Molecular tagging of economically important genes and promoters in sugarcane", with a budget of $65,000 to $75,000 per annum for 3 years, was submitted to SRC in 1988. The overall aim of the proposed project was to develop techniques to introduce suitable genetic elements into sugarcane to allow detection and isolation of economically important genes and gene control sequences (promoters). SRC responded that it was interested to fund a project (UQ1S) at $35,000 to $45,000 per annum. Discussions were held with SRC on an appropriate subset of objectives. In the light of available funding and recent technological advances it was agreed that the project should be directed primarily at development of techniques for gene transfer into sugarcane, and detection and isolation of sugarcane gene control sequences. Objectives related to characterisation of genetic elements which could transpose in the sugarcane genome, and subsequent isolation of a sugarcane gene by transposon tagging, were relinquished. Specific objectives adopted were:

1. Test microprojectile and embryo transformation techniques for gene transfer into intact sugarcane cells.

2. Develop techniques for the isolation of specific sugarcane gene control sequences.

3. Prepare genetic constructs for transfer into sugarcane to assist identification and isolation of useful genes and promoters in sugarcane.

4. Probe the sugarcane genome for endogenous transposable elements homologous to those isolated from other Gramineae, to help evaluate the potential for transposon tagging of useful sugarcane genes.

Our results during 1989 indicated that the microprojectile technique had great potential as a practical transformation system for sugarcane, and should be developed with resources additional to those originally intended for the work on sugarcane promoters. Project UQ1S was reviewed by SRC, and it was agreed to divide the work into two projects from 1990-91. Apparently for administrative reasons, the following project designations were provided by SRC:

Project UQ1S continued the work on development of a transformation system, and was retitled "Microprojectile technique to produce transgenic sugarcane plants".

Project UQ6S was created to continue the work on sugarcane promoters, and retained the title originally used for UQ1S.

This proved somewhat confusing, as UQ1S underwent a change in title and had a corresponding new aim, while UQ6S inherited an old title which was no longer indicative of its major aim. A better title for the promoter work (UQ6S) would be "Isolation of economically important gene control sequences in sugarcane".

This report will be concerned with work on objectives 2, 3, and 4 above, which commenced during 1988-90 under project UQ1S, and continued during 1990-92 under project UQ6S, with a primary focus on differential hybridisation for tissue-specific promoter isolation as part of

PhD studies by Tricia Franks. Constructs relevant to objective 3 above were prepared in PhD studies by Carl Rathus, working primarily in BSES laboratories. Work on objective 1 above, and subsequent development of the sugarcane transformation system, is covered in a separate report on project UQ1S.

Project Objectives (UQ6S)

The overall aim of this project was to develop techniques for the isolation of useful gene control sequences (promoters) in sugarcane. Early in the project we also undertook preliminary studies to lay some groundwork for techniques to tag and isolate useful genes in sugarcane. Specific objectives were:

1. Develop techniques for the isolation of specific sugarcane gene control sequences. This led us to evaluate the applicability of differential cDNA hybridisation to recover tissue specific genes from sugarcane, followed by genomic library screening to recover corresponding tissue-specific promoters.

2. Prepare genetic constructs for transfer into sugarcane to assist identification and isolation of useful genes and promoters in sugarcane. This led us to test the effectiveness of a range of well characterised foreign promoters and novel artificial promoters in sugarcane.

3. Probe the sugarcane genome for endogenous transposable elements homologous to those isolated from other Gramineae, to help evaluate the potential for transposon tagging of useful sugarcane genes.

All project objectives were achieved.

Background and Significance

With the development of reliable techniques for genetic transformation of sugarcane, key factors limiting application of this technology to hasten achievement of breeding objectives will be the limited availability of isolated genes affecting important crop characteristics, and the lack of isolated sugarcane promoters controlling desired patterns of foreign gene expression. When this project commenced we had virtually no knowledge of gene control sequences effective in sugarcane, and only a few promoters had been characterised in other "model" plant species. There were, however indications in the literature that several new techniques could be effective for the isolation of specific types of plant promoters. We set out to test whether the most promising of these techniques would be applicable in a complex polyploid like sugarcane. We expected that this would first require the identification of some basic methodologies (e.g. RNA and DNA extraction methods) which worked for sugarcane, and would continue to be useful in future sugarcane molecular biology research. We also expected to produce several cDNA and genomic libraries from sugarcane which could be maintained as a valuable resource for future researchers. If the tested techniques proved applicable to sugarcane, we hoped to demonstrate this in a concrete way by isolating one sugarcane promoter of potential economic importance. This would identify the technical limitations, and lay the groundwork for subsequent isolation of other desired promoter types for use in sugarcane improvement.

Genes of potential agronomic importance in sugarcane are already available. For example, genes which confer resistance to insects, herbicides, phytotoxins and viruses and which encode ribozymes and antisense RNA to endogenous plant genes have been cloned. For practical application, most introduced genes will need to be expressed in transgenic plants in a highly controlled manner. Precisely defined environmental, temporal and spatial patterns of transgene expression will often be essential for appropriate physiological responses, and will generally be more energy efficient for the plant.

Initiation of gene transcription is subject to precise patterns of control by a "promoter" sequence which lies 5' to the transcription start site, and this provides a suitable mechanism for regulating transgene expression. Directed expression of transgenes by promoters has been tested in transgenic plants in field trials. In a striking example, expression of a reporter gene fused downstream of a wound inducible promoter in transgenic tobacco plants was triggered by insects attacking the field grown plants. A promoter such as this has obvious applications to couple with novel genes for enhanced crop resistance to insect pests (Thornburg *et al.*, 1990). Considerable research effort has been directed towards the isolation and characterisation of a diverse range of gene regulatory promoter sequences for potential application to plant genetic engineering (see for example: Parsons *et al.*, 1989; Rumeau *et al.*, 1990; Roberts and Okita, 1991; Tada *et al.*, 1991).

Terminology and overview of promoter function

Transcriptional control in eukaryotes is known to be positively and negatively responsive to promoter sequences which are present on the same DNA molecule (ie. in *cis*) as the gene. These *cis* acting sequences (or motifs) generally occur at sites nearby (within 100 bp) and 5' of the region of transcription initiation but also appear at extended distances upstream and downstream of the transcription start site (enhancers, repressors), in introns and in the 3' noncoding region of the gene (for example, Callis *et al.*, 1987; Basler *et al.*, 1989; Dean *et al.*, 1989). *Cis* motifs interact with the core promoter (TATA consensus) and RNA polymerase II via *trans* acting factors (DNA binding molecules) to elicit the transcription response (for reviews see Levine and Manley, 1989; Latchman, 1990; Martin, 1991; Waugh and Brown, 1991; Guarente and Bermingham-McDonogh, 1992; Katagiri and Chua, 1992). The site of interaction of a *trans* acting factor with the DNA molecule is determined by the structural properties of the *trans* acting factor and by the sequence of the *cis* motif. Sequence motifs (or

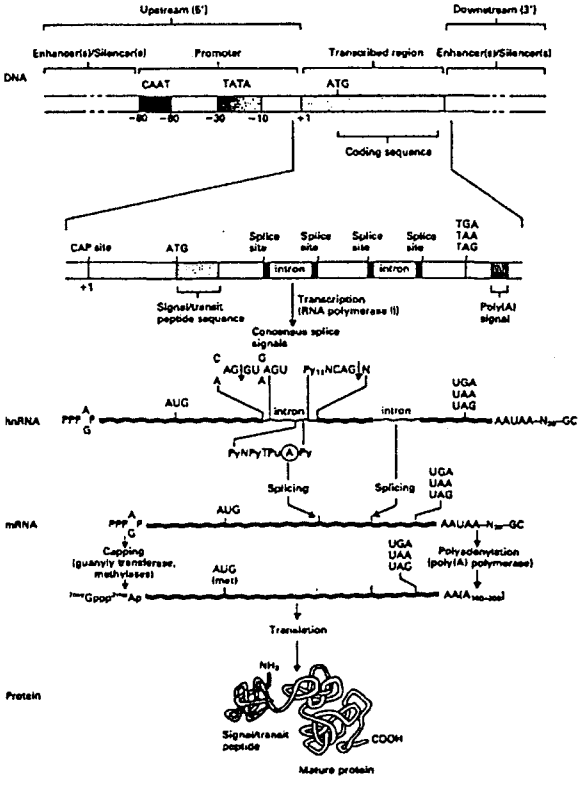


Figure 1. Structure and expression of nuclear-encoded plant genes (from Draper *et al.* 1988).

"modules") which range in length from 4 bp to 20 bp have been associated with light regulated, tissue specific, temperature regulated, anaerobiosis induced, hormone stimulated/repressed and wound or pathogen induced/repressed transcription responses in various plant species (Katagiri and Chua, 1992; Waugh and Brown, 1991).

Gene promoters have been shown to consist of a composite of interacting modules. Each module contributes to at least one expression pattern and the combination of all modules gives rise to the actual expression profile of the gene (see Dynan, 1989).

Isolation of promoters with specific expression patterns

Specific promoters have been isolated using three different approaches. One approach is to introduce a promoterless selectable marker/reporter gene into the plant genome. Any ensuing patterns of reporter gene expression are determined by "tagged" *cis* acting sequences which may then be recovered from the transgenic plant via their probable linkage 5' of the promoterless transgene (Teeri *et al.*, 1986; Koncz *et al.*, 1989). In a second approach, sequences of a genome are screened for promoter activity by randomly cloning the fragmented genome 5' to a reporter gene. The resultant pool of genomic clones are transformed into plants and promoter activity of any one of the cloned fragments is indicated by expression of the reporter gene (Herman *et al.*, 1986; Claes *et al.*, 1991). In this case, there is no need to rescue the promoter from the plant genome provided an account is kept of the constructs used for transformation. Both of these approaches to promoter isolation are confounded when multiple copies of the transforming DNA integrate into the genome, thereby generating a composite expression pattern and multiple candidate promoter sequences. Furthermore, since it is not possible to predetermine the transgene expression pattern, numerous transgenics must be screened in order to identify a specific pattern of expression.

The third and most widely and long used means for promoter isolation is *via* differential hybridisation of genomic or cDNA clones to labelled cDNA. Differential hybridisation was originally used by St. John and Davis (1979) to identify inducible genes in *Saccharomyces* (see Sambrook *et al.*, 1989). Unlike the approaches previously described here, promoter isolation via differential cDNA hybridisation is directed specifically to a gene of predefined expression pattern. Differential representation of cloned genomic or cDNA sequences in different mRNA populations are indicated by quantitative differences in hybridisation of these clones to labelled cDNA probes reverse transcribed from the mRNA populations in question (for example, Figure 2). A genomic clone exhibiting a desired pattern of expression provides an immediate starting point for the recovery of promoter sequence. Promoter isolation from complex genomes which contain much untranscribed DNA is simplified when a cDNA library is first screened for clones showing the desired expression pattern. Selected cDNA clones are then used as homologous probes to screen a genomic library for recovering promoter sequences. In addition to transcriptional control, post-transcriptional mechanisms including rates of transcript export from the nucleus; differential splicing of the primary transcript (Bingham *et al.*, 1988); and transcript stability mediated by 5' capping and 3' polyadenylation (Drummond *et al.*, 1985; Gallie *et al.*, 1989) may all contribute to the steady state levels of a particular mRNA transcript. Before a cloned cDNA which has been selected on the basis of its differential expression is used for promoter isolation, it may therefore be appropriate to confirm that it is primarily subject to transcriptional regulation. The possibility of post-transcriptional mechanisms for control is best excluded using transcriptional-run-off experiments (Ausubel *et al.*, 1987).

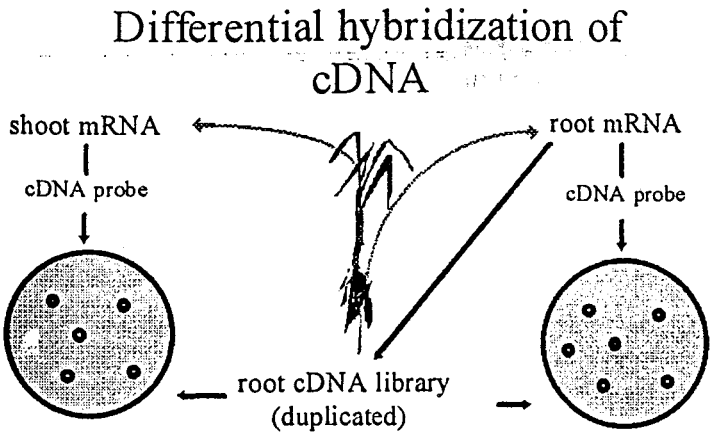


Figure 2. Explanation of differential cDNA hybridisation. Duplicate copies of cDNA (or genomic) clones are probed with single stranded cDNA probes reverse transcribed from mRNAs which were isolated from the tissues for comparison.

The efficiency and sensitivity of comparison of gene populations at both the DNA and RNA levels has been enhanced by various means. Using subtractive hybridisation, sequences which are common to the two mRNA populations to be compared are depleted from the cDNA to be cloned for library construction. Prior to hybridisation, the probes are similarly depleted of sequences common to both mRNA populations, and this enhances the identification of very low copy sequences by the differential screening process (Timberlake, 1980; Zimmerman *et al.*, 1980; see Maniatis *et al.*, 1982). Cloning subtle sequence differences has been enhanced by PCR based sequence amplification following subtractive hybridisation and prior to cloning for library construction (Wieland *et al.*, 1990). In another variation of the differential hybridisation approach, promoter sequences have been identified directly from genomic libraries by screening with mRNA probes labelled at the cap site (Shaw *et al.*, 1989). Since the probe hybridises in the vicinity of the likely transcription start site, the genomic clones identified thus are immediate sources of promoter elements.

Once cloned, a promoter sequence associated with differentially regulated gene transcription requires further characterisation. First, ability of the sequence to act as a functional promoter must be verified. Accurate *in vitro* transcription systems are not easily established for plant systems (see Roberts and Okita, 1991). Promoters are therefore best assessed for their ability to direct transcription of a reporter gene following gene transfer into plant cells (see Willmitzer, 1988).

Research Methodology

Overview

This project was initiated to contribute to practical sugarcane genetic engineering by the development of a practical system for isolation of promoter sequences which will be useful for regulating transgene expression in sugarcane. At the outset, in the absence of a fully established and efficient system for sugarcane transformation, the most appropriate strategy for sugarcane promoter isolation was via differential hybridisation of cDNA. Very little published characterisation of the sugarcane genome at the molecular level was available. In addition, no information on sugarcane promoters and no tests of foreign promoters in sugarcane had yet been reported. However, sugarcane is known to posses a large and complex polyploid genome and therefore an element of the work was to evaluate the applicability (to sugarcane) of techniques which were established on a limited range of principally model plant species. High priority promoters for isolation were designated as those which direct root specific and wound inducible expression, because significant yield decline of Australian sugarcane crops and substantial preventative expenditure has been associated with combating various bacterial and fungal pathogens and insect pests of the plant roots.

These expression patterns provided a suitable test of the differential cDNA strategy for sugarcane, and any resulting promoters had potential for practical application in combination with appropriate pathogen and pest deterring genes for improved sugarcane agriculture.

Using tobacco as a model, it has previously been shown that although each plant organ (leaf, root, stem, petal, ovary and anther) has similar nuclear RNA complexity, each RNA population has a set of transcripts that are undetectable in the RNA of heterologous organs (Kamalay and Goldberg, 1984). Such tissue specific transcripts are amenable to isolation by differential cDNA hybridisation and have lead to recovery of some corresponding promoters (eg. shoot: Eckes *et al.*, 1986; pollen: Hamilton *et al.*, 1989 and root: Yamamoto *et al.*, 1991). Mechanically wound induced transcripts and their corresponding promoters have also been isolated using a differential cDNA approach (eg. potato: Keil *et al.*, 1989; poplar: Bradshaw *et al.*, 1989; bean: Sheng *et al.*, 1991). Differential cDNA hybridisation was therefore used to compare mRNA populations from sugarcane shoots, roots, wounded roots and wounded shoots. Outcomes of the differential hybridisation technique in sugarcane were consistent with those from other systems and permitted recovery of tissue specific transcripts. A gene expressed primarily in sugarcane roots was selected for further characterisation, and a number of sequences with homology to the gene were identified in the polyploid sugarcane genome. One of the homologous sequences was recovered from a genomic library, and its promoter was cloned.

In collaborative studies supported largely by BSES, Carl Rathus prepared and tested a series of artificial promoter constructs expected to provide a range of constitutive expression strengths in sugarcane cells.

Differential hybridisation of cDNA for isolating nonconstitutively expressed sugarcane genes

Plant Growth Conditions and Wounding

Pindar sugarcane variety was used in all experiments due to its ability to quickly send out prolific set roots and plants were grown in hydroponics, to allow easy access to the roots for wounding and mRNA extraction. Sets were dipped briefly in Shirtan 120 fungicide (Consolidated Fertilizers Limited, Brisbane) at the recommended rate and were floated in a well aerated one half strength Aquasol (Hortico) solution to germinate and grow for six weeks in the glasshouse. At 24 hours and then again at four hours prior to RNA extraction, the plants were wounded by successive needle stabs through the entire organ at 2 cm intervals along the stems, leaves and roots.

RNA Extraction and mRNA Purification

Total RNA was extracted from shoot and root tissues using a guanidine thiocyanate/caesium chloride method which is essentially that of Chirgwin *et al.*, (1979). All solutions were made with DEPC treated water. Tissue samples of 3 to 4 g were frozen and ground with a mortar and pestle in liquid nitrogen and thawed in 5 volumes of homogenisation buffer (4 M guanidine thiocyanate; 25 mM trisodium citrate; 100 mM ß-mercaptoethanol, pH 7.0 with 1 M NaOH and filter sterilised). After further brief homogenisation, sodium sarkosyl (10%) was added to a final concentration of 0.5%. The homogenate was filtered through miracloth into centrifuge tubes and spun at 10,000 xg for 15 minutes at 20°C. The supernatant was refiltered through miracloth and layered carefully onto a 6 ml CsCl cushion (5.7 M CsCl; 100 mM EDTA) to fill a 36 ml open top centrifuge tube. The tubes were spun at 25,000 rpm for 18 hours at 20°C in a Beckman SW27 swing out rotor. The supernatant was aspirated from

the pellet which, once well drained, was resuspended in DEPC treated water containing 1 mM DTT and 80 units of RNAsin (recombinant ribonuclease inhibitor, Promega).

PolyA mRNA was purified by passing the RNA twice through a poly (dT)-cellulose spun column (Pharmacia) according to the manufacturer's instructions.

In vitro Translation and Protein Gel Electrophoresis

In vitro translation of mRNAs extracted from wounded shoots, shoots, wounded roots and roots of sugarcane were performed using a Promega wheat germ system in the presence of ^{35}S-methionine (Amersham), exactly as described by the manufacturer. At the same time, *in vitro* translation of mouse liver mRNA (provided in the kit) and a reaction devoid of mRNA served as positive and negative controls, respectively.

Protein LDS (lithium dodecylsulphate) gel electrophoresis of the labelled translation products, staining, drying and exposure to X-ray film were all carried out by Dr. Adrian Clarke using routine techniques (Ryrie, 1983).

Construction of cDNA Libraries

All libraries were constructed starting with the same mRNA stock as those which were sampled for use as *in vitro* translation templates. Using a Pharmacia cDNA synthesis kit, double strand cDNA was synthesized from a wounded shoot mRNA template. Reverse transcription of the first cDNA strand was primed with a polydT oligonucleotide and the double stranded cDNA was made ready for cloning by blunt end ligation to an *Eco* RI adapter which also carried an internal *Not* I site. The adapted cDNA was ligated into *Eco* RI cut λgt10 (Clontech) in a 10 μl mix containing polyethylene glycol at 13.6%. Following precipitation with 0.6 M NaCl, the ligated phage genomes were resuspended in TE and packaged using the Promega "Pack-a-gene" kit according to the manufacturer's instructions.

Wounded root and root cDNA libraries were made with an Amersham cDNA synthesis kit using a polydT primer for reverse transcriptase synthesis of the first cDNA strand. Double stranded cDNA was cloned using an Amersham cDNA cloning kit.

All three libraries were plated using routine methods (see Sambrook *et al.*, 1989) on *E. coli* NM514 [*hsd*R514($r_k^-m_k^-$),ArgH,galE,galX,StrA,*lyc*B7,(Hfl$^+$)] which is most efficiently lysed by recombinant phage. The entire wounded shoot library was amplified by plating the $9x10^4$ clones, flooding the plates with SM medium (1 M NaCl; 0.1 M Tris.Cl pH 7.4; 10 mM $MgSO_4$; 0.01% gelatin), rocking overnight at 4°C and collecting the lysis products (a total of $1.1x10^6$ clones).

DNA Amplification and Purification

Small scale plasmid purification was by a boiling method (see Sambrook *et al.*, 1989). Small scale phage DNA isolation was by a modification of the method provided in the "Amersham cDNA cloning system λgt10" manual. Two 90 mm plates were grown to complete lysis of the NM514 host by the appropriate phage clone, then 4 ml of SM medium was layered over each plate and rocked gently at 4°C for at least 3 hours. The lysate was collected in a 15 ml polypropylene centrifuge tube, centrifuged at 300 xg for 10 minutes, after which the supernatant was recovered. An equal volume of 20% PEG (8000MW)/20 mM NaCl solution (in SM) was added and the mix was set on ice for at least one hour. The tube was spun for 30 minutes at 300 xg in the swing out rotor of a bench top centrifuge. The pellet was drained well and resuspended in 750 μl Luria broth. The phage were purified twice by mixing with

an equal volume of preswollen Whatman DE52 cellulose (prepared by suspending first in 0.05 M HCl and adjusting to pH 7 with NaOH, then resuspending and equilibrating in Luria broth), pelleting the cellulose by centrifugation and recovering the supernatant. To each 1 ml of supernatant, 17.5 μl of Proteinase K (25 mg ml^{-1}) and 42.5 μl of SDS (10%) were added and the mix was allowed to stand at room temerature for 5 minutes before addition of 20 μl of 0.5 M EDTA and further incubation at 68°C for 10 minutes. The mix was extracted once each with phenol/chloroform and chloroform; precipitated with one volume of isopropanol; RNase (10 μl ml^{-1}) treated and then reprecipitated with o.1 volume of sodium acetate (3M) plus 2 volumes of ethanol.

λgt10 clone inserts were amplified using the Clontech λgt10 PCR primer set exactly as described by the manufacturer. These primers (5'AGCAAGTTCAGCCTGGTTAAG and 5'TTATGAGTATTTCTTCCAGGG) bind the vector on opposite strands at sites either side (8 bp and 12 bp, respectively) of the λgt10 *Eco* RI cloning site. The region between the two primer binding sites (including the cloned insert) was amplified from crude preparations of phage clones in a mix (10 mM Tris.Cl pH 8.3; 50 mM KCl; 1.5 mM $MgCl_2$; 0.01% gelatin; 0.2 mM of each dNTP; 1 μM of each primer; 1.25 units Amplitaq^R DNA polymerase) in a Perkin-Elmer-Cetus 480 thermal cycler. Temperature cycling was for 31 cycles: 94°C (denaturing), 30 seconds for the first cycle and 15 seconds for subsequent cycles; 50°C (annealing) for 15 seconds; 72°C extension for 1 minute. A final period of extension at 72°C was for 7 minutes. DNA fragments were purified from agarose gel using the BioRad "Prep-a-gene" kit.

Nucleic Acid Electrophoresis and Transfer of Nucleic Acids to Nylon Membranes
Phage were plated at appropriate densities (1500 plaques per 150 mm diameter plate and then at successively lower densities on 90 mm plates as purification and rescreening proceeded) and following incubation overnight at 37°C, lifts were taken from plates onto Hybond-N circular filters (Amersham). Denaturation (0.5 M NaOH, 1.5 M NaCl) and neutralization (0.5 M Tris.Cl pH 7.0, 1.5 M NaCl) were for five minutes each and were followed by a brief rinse in 2x SSC, air dried and UV crosslinked, essentially as for Benton and Davis (1977).

Agarose gel electrophoresis of DNA fragments and ethidium bromide staining of DNA was by routine techniques. Transfer of DNA fragments from agarose gel was by capillary blotting (Southern, 1975) to Hybond-N nylon membrane (Amersham) followed by UV crosslinking.

RNA gel electrophoresis was done through a 1.4% agarose gel containing 5% formaldehyde in MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA) as described by Sambrook *et al.*, (1989). Lanes were loaded with 20 μg of total RNA. Gels were made ready for transfer by soaking in 0.05 M NaOH for twenty minutes, followed by a rinse in deionized water then soaking in 20x SSC for 45 minutes. RNA was transferred from the gel to Hybond-N nylon membrane (Amersham) overnight by capillary blotting in 20x SSC.

Hybridization of Radioactively Labelled DNA to Membrane Bound Nucleic Acids
Preparation of probes
DNA templates to be used as probes were labelled with [α-^{32}P]dCTP using a multiprime DNA labelling kit (Amersham) and purified by passage through a Biospin 30 Chromotography column (Bio-Rad). The probes were denatured by boiling for five minutes before addition to the hybridization mix.

To prepare cDNA probes reverse transcribed from mRNA templates, 2 μg to 4 μg of mRNA in 3 μl of DEPC treated water was heated to 65°C for 5 minutes and transferred to ice for 5 minutes. After briefly microfuging the sample the following components were added in order: 1 μl DTT (10 mM); 0.5 μl RNAsin (recombinant ribonuclease inhibitor at 40 U μl^{-1}, Promega); 3.1 μl Tris.Cl pH8.3 (400 mM); 1.2 μl KCl (625 mM); 0.5 μl $MgCl_2$ (500 mM); 1.5 μl oligodT (1.6 μl μl^{-1}, Amersham); 1.25 μl dATP (10 mM); 1.2 μl dTTP (10 mM); 1.2 μl dGTP (10 mM); 120 μCi [α-^{32}P]dCTP; 2 μl reverse transcriptase (2.5 U μl^{-1}, Bethesda Research Laboratories). The mix was incubated at 41°C for 2 hours and spun through a Biospin 30 chromotography column (Bio-Rad) and directly through a 0.45 μm Ultrafree-MC Durapore filter (Millipore). Just prior to addition of the probe to the hybridization mix, it was denatured, and the RNA was hydrolysed by the addition of 100 μl water and 60 μl NaOH (1M). The mix was left at room temperature for 10 minutes and then neutralized with the addition of 60 μl HCl (1M) and 60 μl 20x SSPE. The probe was added to the hybridization mix to a concentration of 10^6 to 10^7 cpm ml^{-1} (measured by counting ß particle emission from two replicates of 1 μl purified labelling reaction mix in 1 ml of Packard Scintillator 299™).

Hybridization conditions

In all hybridization experiments, the prehybridization and hybridization mix was: 5x SSPE; 5x Denhardt's solution (0.1% Ficoll; 0.1% polyvinylpirrolidone; 0.1% gelatine); 8% dextran sulphate; 0.2% SDS; 100 μg ml^{-1} sheared, denatured salmon sperm DNA. For hybridization to cDNA clones and Northern blots, polyA DNA (Sigma) was added to a final concentration of 10 μg ml^{-1}.

Prehybridization periods ranged from 1 hour (for Southern blots) to overnight for plaque lifts and Northern blots. Hybridizations were always at 65°C and were for periods ranging from 16 hours (for Southern and Northern blots) to 48 hours (for differential cDNA hybridization experiments).

Following hybridization, filters were washed twice at room temperature in 5x SSPE for five minutes; once at 65°C in 1x SSPE, 0.1% SDS for 15 minutes and twice at 65°C in 0.1x SSPE, 0.1% SDS for 15 minutes each. Filters were exposed to X-ray film (Kodak X-Omat X-K1) at -70°C for the appropriate period.

Video Densitometric Analysis of Autoradiograph Images

Densitometric analysis of autoradiograph images of a Northern blot hybridized to cDNA probe #B6cii was done by Mr. Yuan Zhi Zheng using a "Bio-Rad model 620 video densitometer one dimensional analyst".

Solutions, Restriction Enzyme Digestion, Ligation Reactions and Subcloning Strategies

Solutions (TE: 10 mM Tris.Cl pH 7.4; 1 mM EDTA; SSPE: 0.15 M NaCl, 1 mM $NaH_2PO_4.H_2O$, 1.25 mM EDTA, pH 7.4 and SSC: 0.15 M NaCl; 15 mM trisodium citrate pH 7.2) were prepared exactly as described by Maniatis *et al.* (1982).

Restriction enzymes and T_4 DNA ligase were purchased from New England Biolabs. The enzymes were used with the manufacturer supplied buffers exactly as directed.

The λgt10 clone #32 insert was PCR amplified as described previously, phenol/chloroform extracted, digested with *Eco* RI to remove the cloning adapters, purified by Centricon 100 (Amicon) and ligated to *Eco* RI digested pBluescriptIISK+ (Stratagene). The 600 bp *Not* I

fragment from λgt10 clone #B6cii was obtained from a *Not* I digest of clone DNA by gel electrophoresis and purification using a "Prep-a-gene" kit (Bio-Rad) and then ligated to *Not* I digested pBluescriptIISK+ (Figure 3).

Sequencing

All sequencing was by the dideoxynucleotide strategy of Sanger *et al.* (1977) using the Pharmacia T_7 DNA polymerase sequencing kit. Completed sequencing reactions were electrophoresed through 6% acrylamide, 7 M urea gel; fixed in 10% acetic acid, 15% ethanol; dried and autoradiographed using routine techniques.

Restriction site analysis of the #B6cii clone showed that a single *Pst* I site approximately halved the cDNA insert. Complete double strand sequence of the #B6cii cDNA was obtained from T_3 and T_7 primed sequence of two *Not* I/*Pst* I subclones in pBluescriptIISK+ which combined, represent the entire cDNA (Figure 3).

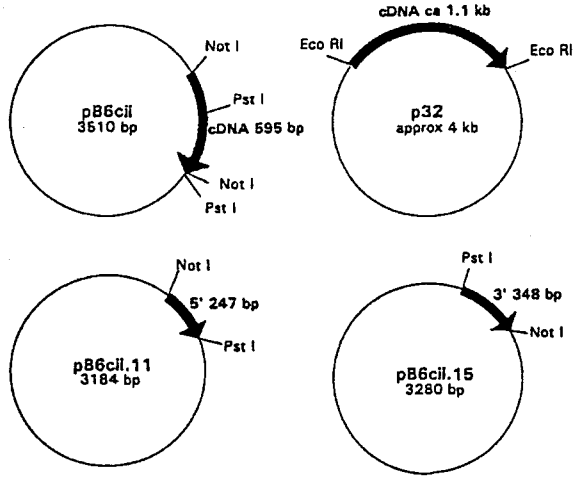


Figure 3. pBluescriptIISK+ subclones of the shoot specific (#32) and the root specific (#B6cii) cDNAs selected from the sugarcane wounded shoot cDNA library.

Clone p32 was generated from an *Eco* RI digested amplicon of the corresponding λgt10 cloned insert.

Clone pB6cii contains the only root specific cDNA of three cDNAs housed in a single λgt10 clone.

Clones pB6cii.11 and pB6cii.15 are derivatives of pB6cii which were generated for sequencing purposes (see text).

Sequence Comparison With Databases

The "Fasta" computer program (Pearson and Lipman, 1988) was used to screen the GenBank and PIR Protein database accessions for homology to the cloned #B6cii nucleic acid and derived protein sequences, respectively.

Cloning and Characterization of the promoter of a gene transcribed primarily in sugarcane roots

Solutions, Restriction Enzymes, T_4DNA Ligase, DNA Polymerase

Solutions (TE: 10 mM Tris.Cl pH 7.4; 1 mM EDTA; SSPE: 0.15 M NaCl, 1 mM $NaH_2PO_4.H_2O$, 1.25 mM EDTA, pH 7.4 and SSC: 0.15 M NaCl; 15 mM trisodium citrate pH 7.2) were prepared as described by Maniatis *et al.* (1982).

Restriction enzymes, T_4 DNA ligase and $Vent_R^{TM}$ DNA polymerase were all purchased from New England Biolabs and used with the manufacturer supplied buffers.

DNA Extraction and Purification

Extraction of total sugarcane DNA

DNA was isolated from young leaf tissue of mature plants (Pindar variety) using the rapid CTAB (hexadecyltrimethylammonium bromide) method described by Rogers and Bendich (1988).

Large scale phage DNA purification
A method for purification of bacteriophage λ DNA which is essentially that of Yamamoto *et al.* (1970), was performed as described by Sambrook *et al.* (1989). Briefly, phage were collected from 4 x 150mm confluently lysed plates and purified by equilibrium centrifugation of 0.75 g of CsCL per ml of bacteriophage suspension, in a 2.6 ml TL100 heat sealed centrifuge tube, overnight at 100,000 rpm. DNA was extracted from the banded phage using formamide.

Plasmid DNA preparation
Mini plasmid DNA preparation was by a boiling method (Maniatis *et al.*, 1982). Large scale plasmid purification for microprojectile bombardment experiments was by alkaline lysis (Maniatis *et al.*, 1982). Following treatment with RNase, the preparation was purified by phenol/chloroform extraction. DNA concentrations were calculated from A_{260nm} and concentrations of serially diluted samples of the plasmid preparation were visually verified alongside a known DNA concentration standard in ethidium bromide stained agarose gel.

DNA purification from agarose gel
Electrophoresed DNA fragments were purified from low melting temperature agarose gel using the BioRad "Prep-A-Gene" kit.

DNA Transfer to Nylon Membrane and Hybridization to Labelled Probes
DNA fragments separated by agarose gel electrophoresis were transferred to Hybond-N nylon membrane (Amersham) by capillary blotting (Southern, 1978) and were UV crosslinked to the filter.

Plated phage were lifted in duplicate onto Hybond-N circular filters (Amersham), denatured (0.5 M NaOH pH7.0, 1.5 M NaCl) and neutralized (0.5 M Tris.Cl pH7.0, 1.5 M NaCl) for five minutes each, essentially as for Benton and Davis (1977). The filters were air dried, rinsed briefly in 2 x SSC and the DNA was UV crosslinked to the membrane.

Probes were labelled with [α-^{32}P]dCTP using a multiprime DNA labelling kit (Amersham), purified by passage through a Biospin 30 chromatography column (Bio-Rad) and, prior addition to the hybridization mix, denatured by boiling for 5 minutes.

The prehybridization and hybridization solution contained 5 x SSPE; 5 x Denhardt's solution (0.1 % Ficoll; 0.1 % polyvinylpirrolidone; 0.1 % gelatine); 8 % dextran sulphate; 0.2 % SDS; 100 μg ml sheared, denatured salmon sperm DNA. Prehybridization periods were 1 hour (for Southern blots) and 16 hours (for plaque lifts). Hybridization was always at 65°C for 16 hours. Following hybridization, the filters were washed twice at room temperature for five minutes in 5 x SSPE; once at 65°C in 1 x SSPE, 0.1 % SDS for 15 minutes and twice at 65°C in 0.1 x SSPE, 0.1 % SDS for 15 minutes each. Filters were exposed to X-ray film (Kodak X-Omat X-K1) at -70°C for the required time.

Construction and Screening of Sugarcane Genomic Library
Sugarcane DNA (*cv.* Pindar) was partially digested with *Sau* 3AI and size fractionated by

sucrose gradient centrifugation as described by Ausubel *et al.* (1989) except that fractions were collected through a hole pierced at the base of the centrifuge tube in which the gradient had been established. Sucrose gradient fractions containing DNA fragments in the size range 15 kb to 24 kb (mean size: 19.5 kb) were pooled and the DNA was ligated to *Bam* HI digested, dephosphorylated λgem11 arms (Promega) in a 10 μl ligation mix containing PEG at 13.6%. The recombinant phage were precipitated in 0.6 M NaCl, resuspended in TE and packaged using the Promega "Pack-A-Gene" kit (which is derived from a single bacterial strain lacking all known DNA restriction systems) exactly as described by the manufacturer.

Library clones were plated on *E. coli* KW251 (F^-, *sup*E44, *gal*K2, *gal*T22, *met*B1, *hsd*R2, *mcr*B1, *mcr*A, [*arg*A81: Tn10], *rec*D1014), LE392 (F^- e14$^-$(*mcr*A$^-$)*hsd*R514(r_k^- m_k^+)*sup*E44*sup*F58*lac*Y*gal*k2*gal*T22*met*B1*trp*R55) or ER1648 (F^-*fhu*A2 (*lac*Z)r1*sup*E44*trp*31*mcr*A1272::Tn10(Tet^r)*his*$^-$1*rps*L104(Str^r)*xyl*$^-$7*mtl*$^-$2*met*B1 (*mcr*C-*mrr*)102::Tn10(Tet^r)) hosts using routine protocols (see Sambrook *et al.*, 1989). For first round hybridization screening of the library, phage were plated at very high density on KW251 (15,000 pfu per 150mm diameter plate), and subsequent rounds of screening for stock purification were at successively lower plaque densities.

Oligonucleotide Design and Synthesis

Three oligonucleotides (oligo1, oligo2 and oligo3) were designed for use as PCR and/or sequencing primers (Figure 4). Oligo1 has homology to the sense strand of #B6cii cDNA at a site between the putative translation and transcription start sites. Oligo2 was designed from initial sequence generated directly from the λgem11#B6cii genomic clone and oligo3 was designed from sequence generated from the amplicon clone (pBSRP) described below. Based on the criteria established by Lowe *et al.* (1990) and using the "Primer Detective" computer program (Clontech), oligonucleotides oligo1 and oligo3 were each designed to be suitable for use in PCR with the Reverse-M13 sequencing primer. Additionally, oligo3 is suitable for use with oligo1 for PCR amplification.

Oligonucleotides 1, 2 and 3 were synthesized and purified by the Queensland Institute of Medical Research facility. M13 Reverse primer, T_3 and T_7 primers were purchased from Pharmacia.

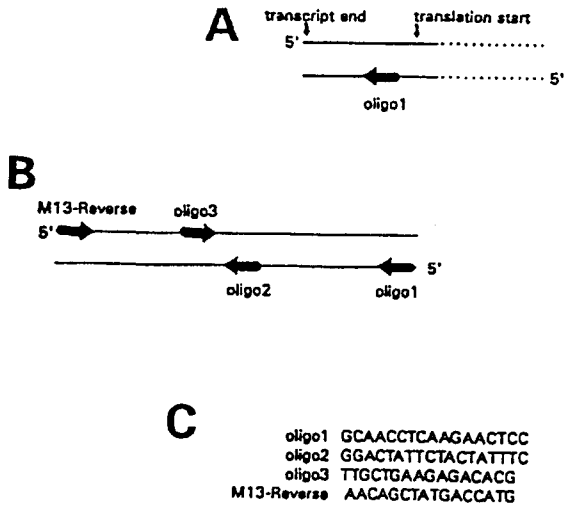


Figure 4. Oligonucleotides used for PCR amplification and sequencing of the B6cii (root specific transcript) promoter.

A shows the location of the oligo1 sequence at the 5' end of the B6cii cDNA sequence. The primer binds to its complementary sequence on the upper, sense strand.

B is the promoter amplicon generated from the genomic subclone (pB6ciiERI) by PCR amplification between the M13-Reverse and oligo1 primers. The positions of oligo2 and oligo3 which were used for sequencing the promoter, are also shown.

C lists the primer sequences 5' to 3'.

PCR Amplification and Amplicon Cloning

DNA fragments were generated by PCR amplification of the region between oligo1 and the Reverse-M13 sequencing primer, using as templates various 2.5 kb *Eco* RI subclones of the λgem11#B6cii genomic clone in pBluescriptIISK+ (eg. pB6ciiERI, Figure 5). The reaction mix contained 1 ng of template DNA; 250 ng each of oligo1 and M13-Reverse sequencing primers; 3 mM $MgSO_4$; 400 μM of each dNTP; 1x$Vent_R^{TM}$ DNA polymerase buffer; BSA at 0.1 mg ml^{-1} in a final total volume of 50μl, in a 0.5 ml microfuge tube. A drop of paraffin oil was added to the tube which was then transferred to a Perkin-Elmer-Cetus 480 thermal cycler and heated for 7 minutes at 94°C. One unit of $Vent_R^{TM}$ DNA polymerase was added to the mix which was then subjected to 35 cycles of: 94°C (denaturing) for 45 seconds; 37°C (annealing) for 1 minute and 73°C (extension) for 2 minutes. After the final cycle, the reaction mix was incubated at 72°C for 7 minutes. The annealing temperature depended on the length and the base composition of the primers and was the lowest value calculated by the formula: annealing temperature °C=4x(G+C)+2x(A+T)-5.

An approximately 1 kb amplicon generated from pB6ciiERI (Figure 5) was purified from the reaction mix by 2 rounds of phenol/chloroform and 2 rounds of chloroform/isoamyl alcohol extraction followed by centrifugation in a Centricon 100 apparatus (Amicon). $Vent_R^{TM}$ DNA polymerase generates blunt ended PCR products and the amplicon was easily cloned by ligation with *Eco* RV cut pBluescriptIISK+. A clone which contained the amplicon oriented with oligo1 closest to the vector *Kpn* I site, was selected by restriction analysis (pRPamp, Figure 5). An *Eco* RI/ *Kpn* I digest of this clone generated a "promoter" fragment which no longer contained the region amplified between the M13-Reverse primer binding site and the *Eco* RI site of the pBluecriptIISK+ polylinker, but did contain an extra 44 bp derived from the vector polylinker between the original Eco RV site and the *Kpn* I site. This fragment was cloned into the corresponding (*Eco* RI/ *Kpn* I) sites of pBluescriptIISK+ to generate clone pBSRP for sequencing (Figure 5).

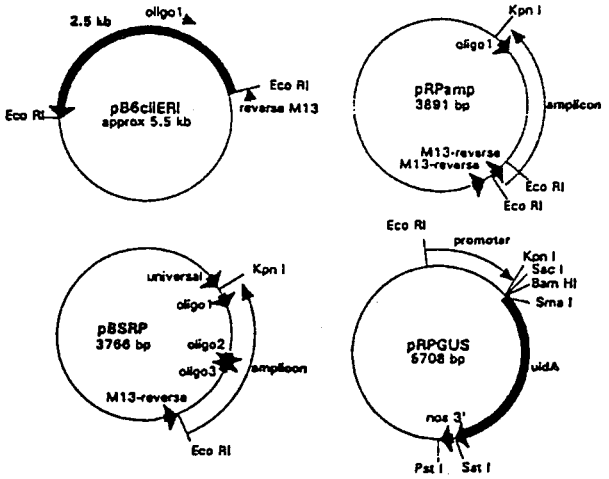


Figure 5. Explanation of the constructs used for isolating and characterizing the B6cii (root specific transcript) promoter region.

All clones are described in detail in the text and are based on pBluescriptIISK+ except for pRPGUS which is derived from pGN100 (promoterless GUS construct). Clone pB6ciiERI contains a 2.5 kb genomic fragment which includes the entire B6cii coding region. Sequence upstream of the translation initiation site in pB6ciiERI was PCR amplified between oligo1 and M13-Reverse primers, and is cloned in pRPamp. The amplicon was excised from pRPamp, between *Eco* RI and *Kpn* I, and that fragment was used to generate pBSRP for sequencing and pRPGUS for gene expression studies.

Sequencing

All sequencing was by the dideoxynucleotide strategy of Sanger *et al.* (1977) using the Pharmacia T_7 DNA Polymerase sequencing kit. Completed sequencing reactions were electrophoresed through 6% acrylamide, 7M urea gel; fixed in 10% acetic acid, 15% ethanol; vacuum dried and autoradiographed using routine techniques.

The sense strand of pBSRP was sequenced using the T_3 and oligo3 primers and the antisense strand was sequenced using the T_7, oligo1 and oligo2 primers (see Figure 5).

DNA Sequence Analysis

Computer programs from the Australian Genome Information Service (ANGIS) suite were used for sequence analysis. The "Signal Scan" computer program (Prestridge, 1991) was used to screen sequence for known eukaryotic DNA sequence motifs.

Constructs likely to drive constitutive gene expression in sugarcane

Plasmid construction

Restriction enzyme digests were performed according to Maniatis *et al.* (1982). pBI221 (obtained from Clontech Laboratories Inc., Palo Alto, CA) contains 35SP followed by *uid A*, with polyadenylation signals provided by the 3' region of the nopaline synthase gene (NOS3') [18]. Plasmid 911 (a gift from Dr V. Walbot, Stanford University) contains *Adh1* intron 1 as a 560 bp *Bcl* I-*Bam*H I fragment in pUC8. pNOSCAT contains NOP followed by the chloramphenicol acetyl transferase (*cat*) coding region from Tn9 and NOS3'. pOACN-100, p35SIGN and pEmuGN were gifts from Drs E. Dennis and D. Last (CSIRO, Division of Plant Industry, Canberra). pOACN-100 contains a 180 bp *Eco*R I-*Bam*H I fragment of the OCS enhancer in front of a crippled *Adh1* promoter (ΔAdh1), followed by *cat* and NOS3' [12]. p35SIGN contains 35SP linked to *Adh1* intron 1, followed by *uidA* and NOS3'. pEmuGN contains six copies of the *Adh1* anaerobic regulatory region and four copies of the OCS enhancer region upstream of the ΔAdh1 promoter linked to *Adh1* intron 1 followed by *uidA* and NOS3' [24]. pGN1 contains back to back 35S promoters driving *uid A* and neomycin phosphotransferase II (*neo*) coding regions followed by NOS3' regions.

The following plasmids were constructed:

pCR1- The 35S promoter in pBI221 was replaced with the 1.3 kb *Hin*d III-*Bam*H I NOS promoter fragment of pNOSCAT, to obtain a NOP-*uidA*-NOS3' construct.

pCR2C and **pCR2R**- The 560 bp *Bcl* I-*Bam*H I *Adh1* intron 1 fragment of plasmid 911 was inserted into the *Bam*H I site of pCR1 in both orientations, to obtain NOP-intron 1-*uid A*-NOS3' constructs.

pCR6C and **pCR6R**- *Adh1* intron 1 was inserted (as above) into pBI221 to obtain 35SP-intron 1-*uidA*-NOS3' constructs.

pCR7C, **pCR7R** and **pCR7RR**- pBI221 was linearised with *Hin*d III, at the 5' end of the 35S promoter. This was followed by dephosphorylation of the 5' ends. The OCS enhancer was isolated as an *Eco*R I-*Sal* I fragment from pOACN-100, blunt-ended and *Hin*d III linkers added. This fragment was then ligated into the *Hin*d III site of pBI221. Three plasmids were selected as OCS enhancer-35SP-*uid A*-NOS3' constructs: pCR7C, with a single OCS enhancer in the native orientation; pCR7R, with a single OCS enhancer in reverse orientation; and pCR7RR, which contains two copies of the OCS enhancer in the reverse orientation.

pCR8- The 560 bp *Xba* I-*Sma* I *Adh1* intron 1 fragment of pCR6C was ligated into the *Xba* I-*Sma* I site of pCR7R to provide an OCS enhancer-35SP-intron 1-*uid A*-NOS3' construct, with correct orientation of the intron.

pCR9- The *Hind* III site of pBI221 was converted to a *Sal* I site, and the 200 bp *Sal* I-*Bam*HI ΔAdh1 promoter fragment of pOACN-100 was ligated into the 4.9 kb *Sal* I-*Bam*H I fragment of the modified pBI221, to provide a ΔAdh1-*uid A*-NOS3' construct.

Protoplast isolation and electroporation
Protoplasts of sugarcane cultivar Q63 and carrot cultivar Chantenay were isolated, and electroporated as described previously (Rathus and Birch, 1992; Bower and Birch,1990).

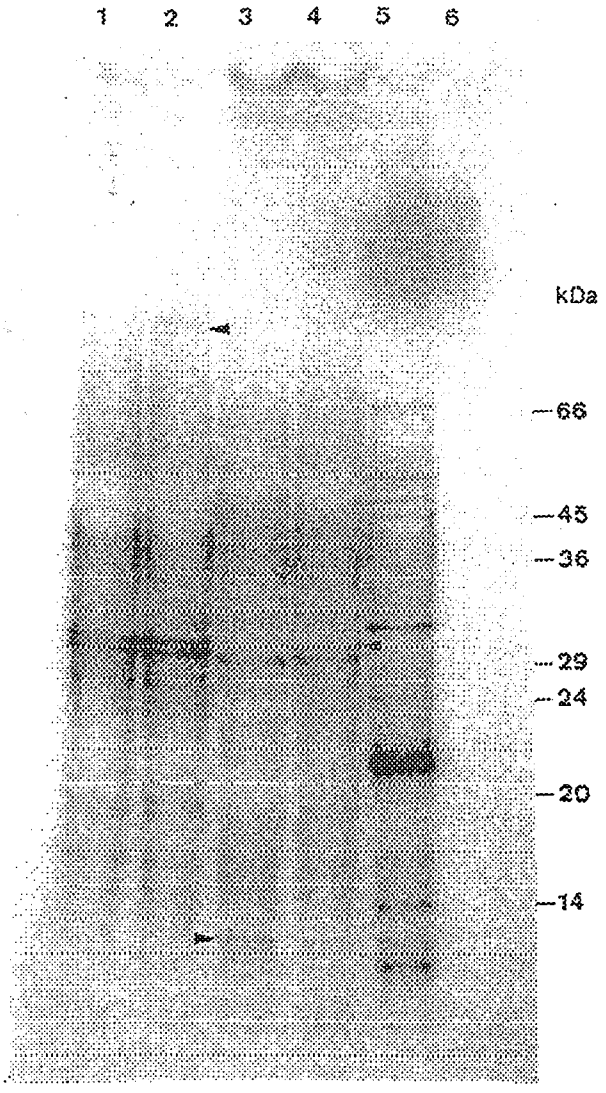
Fluorimetric GUS assays were performed as described by Jefferson *et al.* (1987). Protein concentrations in the supernatants from the cell extracts used in the GUS assays were determined by the dye-binding method of Bradford (1976) with a Bio-Rad Protein Assay kit (Bio-Rad Laboratories, Richmond, CA), using bovine serum albumin for calibration.

Analysis of Data
GUS activity was calculated as nmol 4-methyl umbelliferone (MU) per min per mg protein. Expression relative to the internal control, pBI221, was calculated per mole of plasmid to correct for size differences between plasmids.

Detailed Results

Differential hybridization of cDNA for isolating nonconstitutively expressed sugarcane genes

In Vitro Translation of mRNA from Sugarcane Tissues



High molecular weight polypeptides greater than 66 kDa were generated from each of the four sugarcane mRNA templates and the banding pattern was sharp (Figure 6). There were distinct differences in translation profiles of the shoot samples compared to the root samples. More subtle differences were apparent for the wounded versus the unwounded shoot polypeptide profiles (arrowed in Figure 6).

Figure 6. Autoradiograph of products from *in vitro* translation of sugarcane mRNA.
Lanes are: **1** wounded shoot; **2** shoot; **3** wounded root; **4** root; **5** mouse liver (positive control); **6** no template (negative control.

Characterization of cDNA Libraries Constructed from Wounded Shoot, Root and Wounded Root mRNAs

Three libraries were constructed and characterized with respect to insert a size and frequency. The total number of library clones, the calculated effective library size and the insert size range for the three libraries appear in Table 1. The entire wounded shoot library was amplified to a final total of $1.1x10^6$ clones.

Table 1. Characterization of the wounded root, root and wounded shoot sugarcane cDNA libraries

cDNA library	Total n° pfu on *E. coli* NM514	Frequency of inserts[a]	Calculated library size[b]	Insert size range
wounded shoot	0.9×10^5	9/14	0.5×10^5	500 bp - 5 kb
wounded root	0.83×10^6	2/10	0.17×10^6	<100 bp - 400 bp
root	0.51×10^6	7/10	0.36×10^6	<100 bp - 300 bp

[a]DNA was isolated from 14 clones selected at random from the wounded shoot library and digested with Eco RI to release the insert from the vector. The presence and size of cloned cDNA inserts were determined. Ten clones randomly selected from the wounded root library and ten clones from the root library were assayed for insert presence and insert size using PCR amplification of DNA between the cloning site.

[b]Calculated by multiplying the total number of pfu (plaque forming units) on NM514 by the frequency of clones which have inserts.

Differential Hybridization Screening of the Wounded Shoot cDNA Library

Quadruplicate lifts of 6000 clones from the amplified wounded shoot library were hybridized to cDNA probes reverse transcribed from either wounded shoot, wounded root, shoot or root mRNA templates (for example, Figure 7).

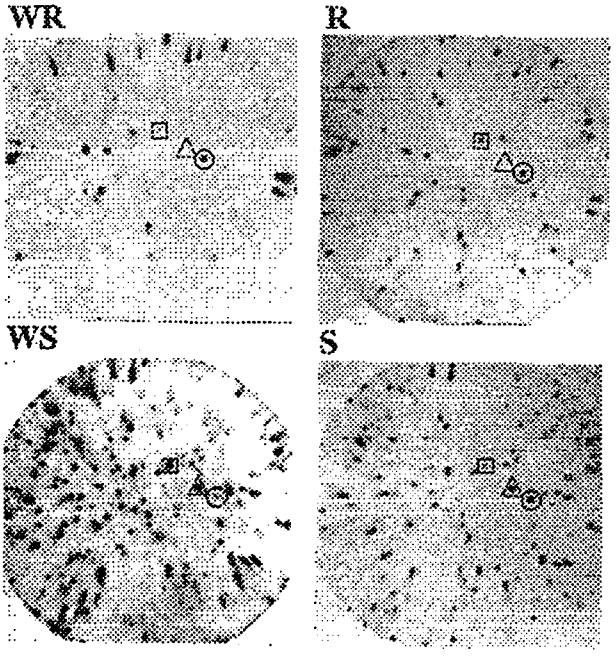


Figure 7. An example of results of the first round screen of wounded shoot cDNA library hybridized to wounded shoot (WS), shoot (S), wounded root (WR) and root (R) cDNA probes. Marked clones hybridised the probes differentially

Rapid visual screening of the first round clones found hybridization patterns consistent with low level constitutive, high level constitutive and various patterns of differential expression, as shown in Table 2. Twenty clones were selected, with an emphasis on root specific clones, for second round purification and rescreening (Table 3). Although clones included in that group were scored wound inducible, their first round pattern of expression proved to be an artefact because as screening continued for stock purification, no wound inducible sequences were recovered. Fortuitously, a shoot specific clone (#32) was confirmed to be a pure stock after only one round of purification, but other impure stocks yielded clones with different patterns of expression. After three rounds of purification and rescreening, all phage stocks were homogeneous, and 26 clones (derived from 7 phage stocks collected after the first round screen) were selected for testing by restriction analysis. The clones fell into 6 classes based on the restriction pattern generated by *Eco* RI digestion. Clones which yielded multiple insert fragments with *Eco* RI were digested with *Not* I (the second restriction enzyme site in the linkers used for cDNA cloning). In all cases *Not* I digestion resulted in restriction patterns identical to those generated by *Eco* RI, suggesting that the clones were likely to contain multiple cDNA inserts rather than multiple *Eco* RI and *Not* I sites within a single insert. Southern analysis of quadruplicate blots of the 26 *Eco* RI digested clones, probed with the four cDNA types, indicated that at the most, one size class of shoot specific clones and two size classes of root specific clones had been isolated (Figure 8).

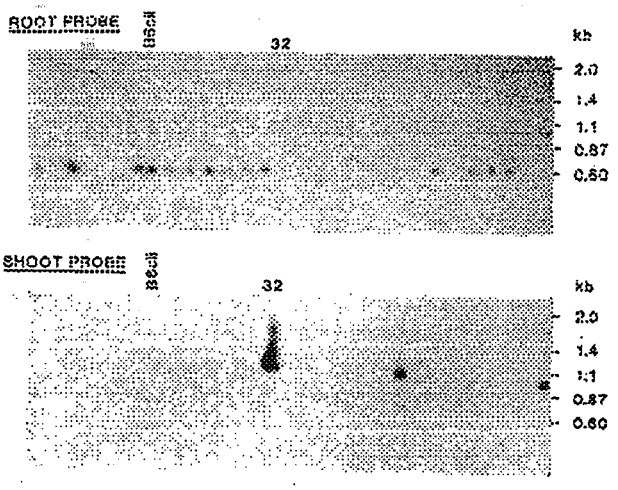


Figure 8. Southern analysis of 26 *Eco* RI digested wounded shoot library clones. Quadruplicate blots were hybridized to wounded shoot, shoot, wounded root and root cDNA probes. Hybridization was identical for wounded shoot and shoot probes and wounded root and root probes.

Using high stringency conditions, one shoot specific and one root specific insert were back hybridized to the same filter. Results of that experiment indicated that all of the cloned shoot specific inserts were homologous and likewise for the root specific inserts. Hence, two unique tissue specific sequences (one shoot specific and one root specific) had been identified. In both cases, these sequences occurred in two distinct clone classes based on *Eco* RI and *Not* I digest patterns and insert size. Inserts from the longest root specific clone (#B6cii) and shoot specific clone (#32) were subcloned in pBluescriptIISK+. The subcloned root specific insert was about 600 bp and the shoot specific insert was about 1.1 kb.

The frequencies of appearance of clone #B6cii and clone #32 homologous sequences in the amplified wounded shoot cDNA library were determined by hybridization of the cloned inserts to plaque lifts of clones in that library. Twelve clones in 1500 hybridized the #B6cii probe and 20 clones in 1500 hybridized the #32 probe. Similarly, 1500 clones from the nonamplified root cDNA library were probed with #B6cii and two homologous clones were identified.

Table 2. Example of results for "rapid visual screening" of clones from first round differential hybridization.

Hybridizing probes[a]				N° clones
wounded shoot	shoot	wounded root	root	
-	-	+	+	20
-	-	+	-	4
+	+	+	-	7
+	+	-	-	15
-	+	+	+	2
+	-	+	+	5
+	+	-	+	6
-	+	+	-	1
-	+	-	+	1
-	+	-	-	2
+	-	-	+	1
+	-	-	-	4
				68

[a]Clones recorded here were only scored if they were expressed differentially and hybridized strongly to at least one of the probes.

Table 3. First round hybridization pattern of 20 clones selected to enter second round screen.

Hybridizing probes				N° clones
wounded shoot	shoot	wounded root	root	
+	-	-	+	1
-	-	+	+	10
+	-	+	+	2
-	-	+	-	3
-	+	-	-	1
+	-	-	-	1
+	+	-	+	1
+	+	-	-	1

Northern Analysis of Two Tissue Specific cDNA Clones

Inserts from clones #32 and #B6cii were labelled to the same specific activity and hybridized to a Northern transfer of electrophoretically separated total RNA isolated from wounded roots, roots, wounded shoots and shoots of sugarcane plants. The clone #32 probed Northern was exposed to X-ray film for 8 days and the #B6cii probed Northern was exposed for 14 days. The pattern of hybridization was identical for wounded roots compared to roots and similarly, the pattern of hybridization was identical for wounded shoots and shoots. The #B6cii cDNA hybridized strongly to a root transcript of about 600 bp and very weakly to a shoot transcript which had migrated through agarose gel slightly slower than the root transcript. The probe also hybridized weakly to other larger and smaller transcripts in the root RNA lane. Cloned cDNA #32 hybridized very strongly to a number of transcripts which only appear in the shoot RNA population and centre around a very strongly hybridizing transcript of about 3 kb (Figure 9).

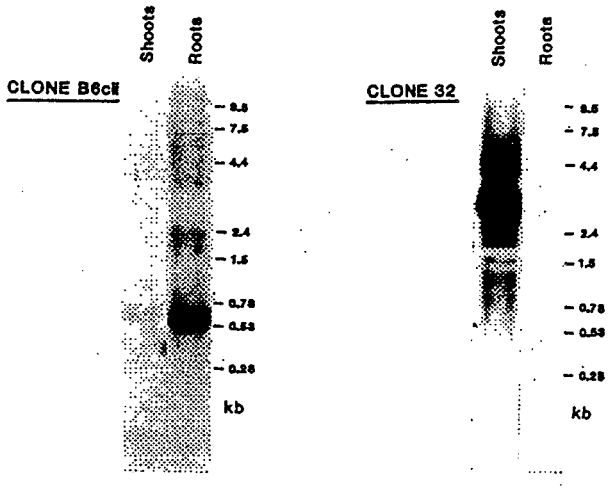


Figure 9. Northern analysis of RNA isolated from sugarcane tissues probed with clone #32 and clone #B6cii. The pattern of hybridization was identical for wounded shoot and shoot lanes and wounded root and root lanes.

Video densitometric analysis of transcript hybridization to clone #B6cii in Northern analysis (Figure 9) measured 7 bands (or peaks) in the root RNA lane (Table 4). The sensitivity of the apparatus was insufficient for detecting the single visible band of hybridization of shoot RNA to #B6cii.

Table 4. Video densitometric quantification of root[a] transcript hybridization to cDNA #B6cii (Figure 9)

Peak (hybridizing band[b])	Height (optical density)	Area (optical density x mm[c])	Relative area (% total)
1	0.22	0.45	4.96
2	0.19	0.39	4.31
3	0.11	0.29	3.25
4	0.02	0.04	0.43
5	2.06	7.66	84.6
6	0.09	0.18	2.00
7	0.02	0.04	0.40

[a]shoot transcript hybridization to #B6cii was undetected by the apparatus
[b]hybridizing band, counting from top of gel
[c]hybridizing band width

Sequence of the Root-Specific cDNA Clone #B6cii

The entire root-specific cDNA clone #B6cii was sequenced and a short open reading frame of 231 bp which encoded 76 amino acids was preceded by a 62 bp leader sequence and followed by a 303 bp noncoding region. The translation product of the open reading frame consists of a central length of 43 amino acids flanked at the amino and carboxy termini by cysteine rich regions of 19 and 15 amino acids, respectively and terminates with a TGA codon. A 13 bp inverted repeat motif appears at position 39 bp upstream of the ATG start (Figure 10).

CTCAATCTTC	CCCTCTCTGC	ATAGGAGTTC	TTGAGGTTGC	AGGAGACCGA
GAAGTTCAGA	AGATGTCTTG	CGGTGGAAGC	TGCAACTGCG	GTTCCAGCTG
CGGCTGCGGC	GGCGGATGTG	GAAAGATGTA	CCCTGACTTG	GCCGAGAAGA
GCGCCGCCGC	CGCCCCGGCC	ACGTCCTCGG	CGTTGCACCT	GAGAAGGGGC
ACCTCGGGAC	CGGGTTGGAG	AAGCCGCGGA	GTCGAGGCCG	GCCATGGCTG
CAGCTGCGGC	TCCGGGTGCA	AGTGCAACCC	TTGCAACTGC	TGATGGAGAA
AAGGCGATTT	GCGTCATTGC	GTGCACGCTC	GTCGTCGTCA	GATCGTCACC
CTGTATGTGT	ACGTGTATGT	GAGTATGTGA	CCGTGTGAGT	AAATAAGGAG
CTATCCGGAC	ACCCAGTGTG	AGCTGTCAGG	TATGATGGAC	ACCTTGTGAA
GTTGTGATGG	TGCGAATTGT	GTGTTGTTGG	GAATCATGTG	GCTACCAAGC
TCTTGTTCAC	CCAATCTGTT	ATCTGTGTGT	TCTATGCTGT	AACAGTGTGC
TGTTGGAGGA	ATACAACGTG	GTTTGCTTGT	GCATCTTCAA	AAAAAA

Met·Ser·Cys·Gly·Gly·Ser·Cys·Asn·Cys·Gly·Ser·Ser·Cys·Gly·Cys·Gly·
Gly·Gly·Cys·Gly·Lys·Met·Tyr·Pro·Asp·Leu·Ala·Glu·Lys·Ser·Ala·Ala·
Ala·Ala·Pro·Ala·Thr·Ser·Ser·Ala·Leu·His·Leu·Arg·Arg·Gly·Thr·Ser·
Gly·Pro·Gly·Trp·Arg·Ser·Arg·Gly·Val·Glu·Ala·Gly·His·Gly·Cys·Ser·
Cys·Gly·Ser·Gly·Cys·Lys·Cys·Asn·Pro·Cys·Asn·Cys·*

Figure 10. Complete 596bp sequence of root specific cDNA (#B6cii) and the 76 amino acid sequence of the hypothetical gene translation product. An inverted repeat motif at the 5' end and a putative polyadenylation signal at the 3' end are stippled. The 231bp open reading frame, which generates the hypothetical polypeptide, is underlined.

Cloning and Characterization of the promoter of a gene transcribed primarily in sugarcane roots

Southern Analysis of Sugarcane DNA probed with Shoot and Root Specific cDNA Clones

Inserts from the shoot specific (#32) and then the root specific (#B6cii) clones were hybridized to a Southern blot of *Eco* RI and *Bam* HI digested DNA isolated from Q63, Pindar, Q44, NCo310, Q87 and NG51-142 (Marker) sugarcane cultivars. At least three different sized fragments hybridized the probes for each cultivar and the patterns of hybridization varied between cultivars. For each specific digest (either *Eco* RI or *Bam* HI) and for each probe (either #B6cii or #32) at least one band was shared by all cultivars (arrowed in Figure 11).

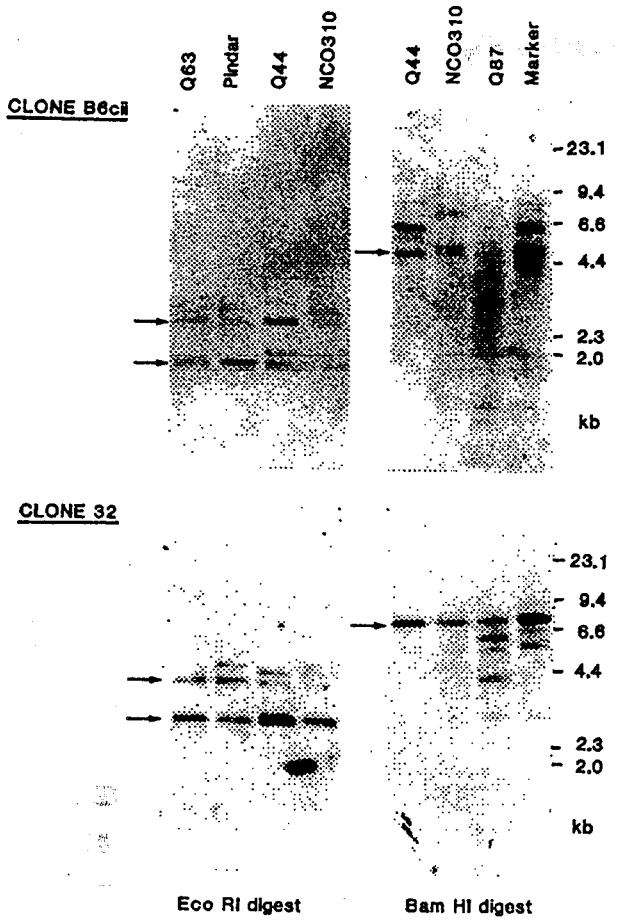


Figure 11. Southern analysis of DNA from sugarcane varieties probed with cDNAs #32 and #B6cii. Arrowed bands are common to all varieties.

Characterization and Screening of the Genomic Library

Efficiency of cloning sugarcane genomic DNA was influenced by the identity of the bacterial host (Table 5).

Table 5. Bacterial host dependent efficiency of cloning sugarcane DNA

Bacterial host	Number of plaques[a]
LE392	8
KW251	110
ER1648	103

[a]100 μl of a 10 fold dilution of a pilot ligation (1 μg *Bam* HI digested λgem11 with 50 ng *Sau* 3AI partially digested sugarcane DNA) was plated onto each host.

The genomic library produced from sugarcane *cv.* Pindar consisted of 0.6×10^6 clones when plated on KW251. A total of 90,000 clones from the primary genomic library DNA was screened for hybridization with cDNA #32 and #B6cii and for each case a single homologous genomic clone was identified and recovered (λgem11#32 and λgem11#B6cii).

Initial Sequence Analysis of Genomic Clone #B6cii and PCR Amplification of the Region 5' to the Transcription Start Site

The oligo1 primer was used to generate sequence directly from the 15 kb insert of λgem11#B6cii genomic clone, from a site just upstream of the putative translation start codon, to approximately 350 bp upstream of the probable transcription start site. This sequence was subsequently confirmed by double stranded sequencing of a plasmid clone (see Figure 12).

Southern analysis of various restriction digests of the λgem11#B6cii genomic clone probed with #B6cii cDNA established that a single 2.5 kb *Eco* RI fragment contained the entire #B6cii cDNA sequence. An *Eco* RI digest of the #B6cii λgem11 genomic clone was ligated with *Eco* RI digested pBluescriptIISK+ and a fraction of the ligation mix was used as template for PCR amplification between the M13-Reverse primer and oligo1. One major band of about 1.0 kb and 2 minor bands of about 1.1 kb and 1.4 kb resulted. A portion of the ligation mix was transformed into a bacterial host and 8 individual recombinant clones were selected at random. Each of the clones were used as templates for PCR amplification between oligo1 and the M13-Reverse sequencing primer. Four of the 8 clones gave single amplicons, whereas the other 4 clones were not amplified, presumably because either the oligo1 homologous sequence was absent from the insert, or the oligo1 containing insert was in an orientation unsuitable for the PCR amplification. Three of the four amplified products were about 1.0 kb and the fourth was about 1.1 kb. One of the smallest amplicons (approximately 1.0 kb) which corresponded to the major amplicon of the ligation mix, was selected for further analysis. The specificity of the amplicon was tested by using it as the template for amplification between oligo1 and the T3 sequencing primer. The binding site of the latter primer lies between the M13-Reverse sequencing primer site and the vector *Eco* RI cloning site and the resulting amplicon was marginally smaller than the original M13-Reverse/oligo1 primed amplicon, as expected.

Southern Analysis of Sugarcane DNA Probed with the #B6cii Promoter

Using high stringency conditions, the selected amplicon was probed to a Southern blot of DNA from several sugarcane varieties digested with *Bam* HI and *Eco* RI. In each case the pattern of hybridization was identical to that observed when the same blot was probed with the #B6cii cDNA. That is, no additional hybridizing bands were observed and no bands were excluded (see Figure 11).

Sequence of the #B6cii Promoter

The promoter amplicon was cloned and sequenced. The complete 813 bp sequence in Figure 12 corresponds to that amplified from the original genomic clone between and including both the native *Eco* RI site and the oligo1 binding site.

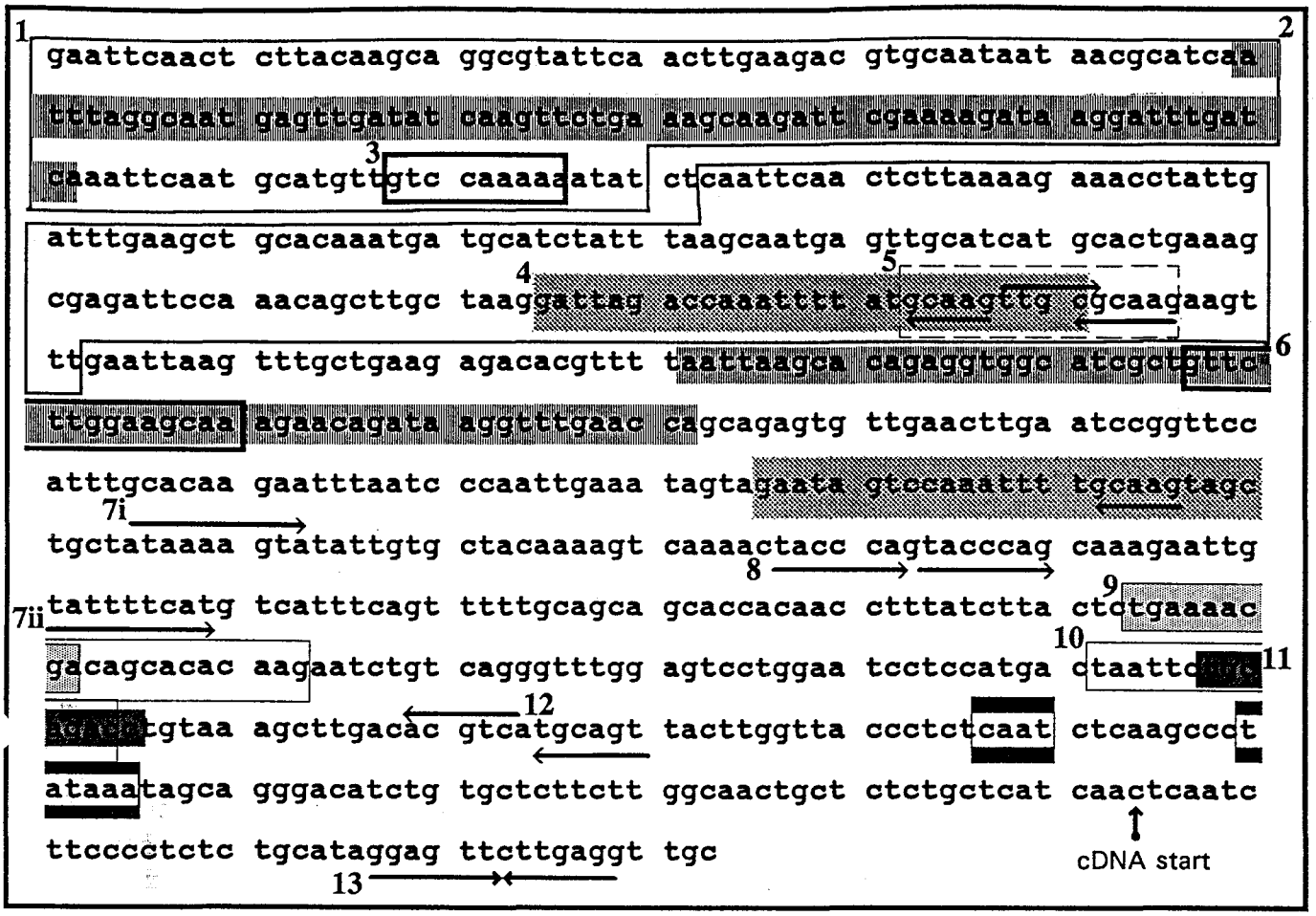


Figure 12. Sequence of the promoter of a gene encoding a root abundant transcript. Delimited motifs are explained below.

1. 150 bp element repeated directly with 79% similarity.
2. 63 bp subelement of 1, repeated with 73% similarity, 209 bp downstream.
3. appears in the promoter of a metallothionein-like gene which is expressed at high basal levels in the roots of maize (de Framond, 1991).
4. 26 bp element repeated with 88% similarity, 164 bp downstream.
5. 14 bp sequence which includes a 5 bp repeated element and an 8 bp palindrome.

6*. 14 bp element appears in maize with one base pair change.

7i. appears in the promoter of a metallothionein-like gene which is expressed at high basal levels in the roots of pea (Evans *et al.*, 1991) and has been located in enhancers of several viral, amphibian and animal genes and may therefore be a ubiquitous transcription factor (See Maedo *et al.*, 1987).

7ii. is a reversed complementary sequence to 7a.

8. 7 bp direct, perfect direct repeat.

9*. 9 bp element also appears with some changes in pea and maize.

10*. 14 bp element appears with some changes in pea and maize.

11. 8 bp sequence similar to animal metal regulatory element (see text).

12. CREB/as-1 sequence arrangement (see text).

13. reversed repeat unit within the transcribed region.

*appear in the promoters of corresponding genes isolated from pea (Evans *et al.*, 1991) and maize (de Framond, 1991) and are detailed in Figure 15.

Constructs likely to drive constitutive gene expression in sugarcane

Figure 13 shows the levels of GUS activity obtained in sugarcane and carrot protoplasts 18-24hr after electroporation with the various plasmid constructs.

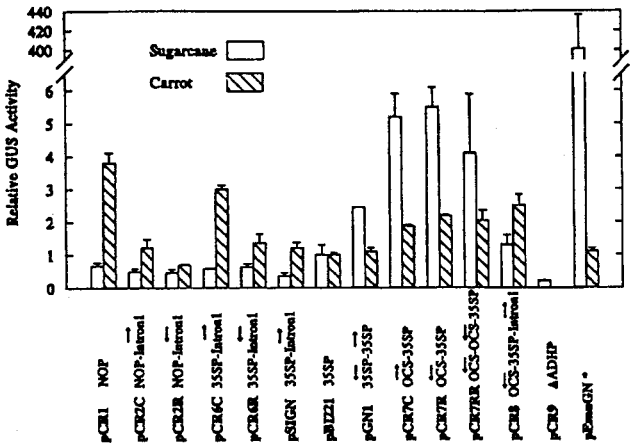


Figure 13. Relative GUS activity in sugarcane and carrot protoplasts after electroporation. All results are given relative to activity using pBI221. Absolute value for sugarcane using 100 μg pBI221 per ml ($1x10^6$ protoplasts) was 13.4 pmol MU min^{-1} mg $protein^{-1}$. Absolute value for carrot using 20 μg pBI221 per ml ($2x10^6$ protoplasts) was 1,450 pmol of MU min^{-1} mg $protein^{-1}$. Results are means (with standard error bars) from three replicates, based on equimolar quantities of each plasmid for each plant species. * See Materials and Methods for details of the Emu promoter.

As expected, pCR9, which contained the crippled ΔAdh1 promoter, gave GUS activity not significantly above background. This promoter lacks the upstream sequences (-140 to -100bp) necessary for both constitutive expression and anaerobic regulation in maize.

35S and NOS promoters

Plasmids containing either 35SP or NOP gave expression levels well above background. The 800 bp 35SP region used here contains at least two regulatory domains within the first 350bp upstream of the transcription initiation site, both of which are necessary for constitutive expression. For NOP, all essential elements are contained within the first 130 bp upstream of the transcription initiation site in the 1300 bp region used here. In sugarcane, the 35SP construct (pBI221) and the corresponding NOP construct (pCR1) gave similar expression levels. In carrot protoplasts, NOP drove a 4-fold higher level of expression than 35SP.

In previous studies using *cat* or *neo* reporter genes, 35SP constructs resulted in much higher expression than NOP constructs in barley, rice and rye protoplasts, 70-fold higher expression in maize, and approximately equal expression in carrot protoplasts. Sequence differences at the junctions of the regulatory elements assembled in chimaeric promoter-reporter gene constructs may affect recognition by trans-acting factors, nucleic acid folding or translation of the transcript. This possibly accounts for minor (several-fold) differences in relative expression strengths of NOP and 35SP constructs tested in various laboratories. With the exception of pGN1 and pEMUGN, all comparisons reported here are based on constructs in which alternative regulatory regions replace 35SP in pBI221.

Enhancers

pGN1, with a duplication of the 35S enhancer element, gave a 2-fold increase in expression relative to pBI221 in sugarcane, but not in carrot. Duplication of 35S enhancer has previously been shown to have a stimulatory effect of expression in a dicot, tobacco. Our results indicate that duplication of the 35S enhancer also stimulates expression in a monocot, sugarcane.

Insertion of the OCS enhancer element, in either orientation, immediately upstream of 35SP (pCR7C, pCR7R) led to increased expression in both sugarcane (5-fold) and carrot (2-fold). Thus the OCS enhancer stimulated expression even when in conjunction with a strong promoter, and at a considerable distance (approximately 800 bp) from the transcription initiation site. Inserting a tandem repeat of the OCS element in reverse orientation (pCR7RR) did not lead to any further increase in expression. The OCS element, when placed immediately upstream of the ΔAdh1 promoter, was shown by Ellis *et al.* (1987) to stimulate expression in both maize and tobacco. The enhancement was independent of orientation,but was reduced to 15% of maximum when distance from the transcription initiation site was increased by 275 bp. It would therefore be of interest to test the OCS enhancer closer to the transcription initiation site for stronger enhancement in sugarcane.

Adh1 intron 1

Inclusion of *Adh1* intron 1, in either orientation, between 35SP (pCR6C and R, p35SIGN) or NOP (pCR2C and R) and *uid A* caused a 30-70% reduction in expression in sugarcane protoplasts. This effect was confirmed in pCR8, where inclusion of intron 1 effectively negated increased expression due to the OCS element upstream of the 35S promoter in pCR7R.

In maize protoplasts, inserting *Adh1* intron 1 behind 35SP and NOP increased transient *cat* expression by 8 to 20-fold and 170-fold respectively. In the same system, the maize Bronze 1 intron behind 35SP increased expression 10-fold. Rice actin 1 intron 1 inserted behind 35SP increased *uid A* expression by 40-fold and 56-fold in rice and maize protoplasts respectively [30]. When inserted between 35SP and *cat*, maize *Adh1* intron 6 increased expression 30-fold in breadwheat protoplasts. The first exon and intron of the maize shrunken 1 gene increased expression from 35SP by up to 1000-fold in maize and rice.

This evidence of enhanced expression due to presence of an intron has led to wide use of constructs including an a 5' intron, usually *Adh1* intron 1, in studies aimed at strong expression of introduced genes in monocots. However, the intron enhancement effect may be nullified or reversed in some constructs, depending on adjacent sequences. Last *et al.* (1991) found no enhancement due to inclusion of *Adh1* intron 1 in 35SP constructs expressed in protoplasts of maize, wheat, rice, eikhorn, *Lolium* or *Nicotiana.* Our results show that inclusion of *Adh1* intron 1 consistently reduces expression from NOP and 35SP constructs in sugarcane.

Adh1 introns are apparently inefficiently spliced in dicots, and provide no enhancement in tobacco or bean cells, which is consistent with the lack of enhancement shown here for p35SIGN and pCR2C in carrot (Figure 13). pCR6C and p35SIGN, which are independently produced but similar 35SP-intron 1-*uidA*-NOS3' constructs, gave 2-fold different expression levels in carrot. This again emphasises the potential influence of precise sequence context around regulatory elements.

The similarity in expression levels between constructs with *Adh1* intron 1 in forward and reverse orientations (Figure 13; pCR2C *vs* pCR2R, p35SIGN and pCR6C *vs* pCR6R) was unexpected. This occurred in both sugarcane and carrot. The effect of an inverted *Adh1* intron 1 in the 5' untranslated region has not previously been reported. However an inverted Bronze 1 or Shrunken 1 intron in this position greatly reduced expression in maize. The inverted *Adh1* intron 1 would not be expected to be spliced out of the mRNA, and would introduce

10 spurious AUG codons (8 with in-frame stop codons) before the usual translational start . However, none of these spurious AUG codons are flanked by sequences resembling the consensus for plant translational starts and they may not prevent translation from a downstream start site. Alternatively, RNA processing may remove some 5' sequence before effective translation occurs. Similar considerations may explain why the sense orientation of *Adh1* intron 1 (9 spurious AUG codons, 8 with in-frame stops) does not reduce expression in *Nicotiana* or carrot (with 35SP, Figure 13), even though it is inefficiently spliced in these dicots.

Multiple elements

pEmuGN, which contains 6 copies of the anaerobic regulatory element from the maize *Adh1* promoter and 4 copies of the OCS enhancer, as well as *Adh1* intron 1, gave expression 400-fold higher than pBI221 (Figure 13). We have not attempted to quantify the contribution of the various elements in the Emu promoter. However preliminary results from microprojectile bombardment studies show that another artificial promoter with 4 copies of the OCS enhancer, a core 35S promoter region and an untranslated Ac leader (unpublished construct from R.W.Michelmore) is at least as strong as Emu in sugarcane cells (Bower and Birch, unpublished). Clearly the use of multiple enhancer elements close to the TATA box of a core promoter region can produce very strong promoters in sugarcane, as in other monocots. Given the decreased expression in sugarcane protoplasts electroporated with NOP and 35SP constructs containing *Adh1* intron 1, removal of the intron from pEmu constructs could be tested if even higher expression levels were desired in sugarcane.

Discussion

Differential hybridization of cDNA for isolating nonconstitutively expressed sugarcane genes

In Vitro Translation of mRNAs

The products of *in vitro* translation of wounded shoot, shoot, wounded root and root mRNAs (Figure 6) indicated that mRNA integrity was preserved during purification because protein bands were well defined, with no indication of smearing due to translation from shortened transcripts. The largest *in vitro* translation products were consistent with transcripts greater than 2.0 kb (assuming that 10 kDa protein requires 270 bp coding nucleic acid). It was therefore concluded that full-length cDNA clones could be generated using poly dT primed reverse transcription products from such mRNA stocks and these stocks were used subsequently for cDNA library construction.

The strikingly different *in vitro* translation profiles of root mRNA compared to shoot mRNA indicated that the corresponding libraries should include a proportion of transcripts with specificity to root or shoot tissues. The *in vitro* translation products of wounded tissues versus untreated tissues (Figure 6) provided an indication that some response to wounding occurred. Products were absent from or were reduced in the wounded tissue compared to the untreated tissue for both roots and shoots. However, there was no clear indication of products strongly induced upon wounding. Perhaps the technique which was used to wound the sugarcane tissues (needle stabs at 2 cm intervals, 24 hours and 4 hours prior to RNA extraction) was inappropriated for inducing a dramatic wound response. Certainly, needle stabs would inflict damage to a lower proportion of cells in comparison to slicing (18 hours prior to RNA extraction), which induced high levels of RNA synthesis in potato (Logemann *et al.*, 1989). However, systemically induced transcripts have also been cloned by differential

cDNA hybridization from tissues of plants subjected to comparatively less harsh wounding. In those cases, wounding was by application of dialysis clamps 24 hours and 4 hours before extraction (in potato, Sanchez-Serrano *et al.*, 1987) or by repeated plier bites in a 48 hour period (in poplar, Parsons *et al.*, 1989). No *in vitro* translation analysis of induced transcripts accompanied these studies and this prevents comparison of those wounding strategies with the one which was applied here.

Characterization of the cDNA Libraries

The wounded root and root cDNA libraries were large and the cloned inserts were short. In contrast, the wounded shoot library was relatively small and had large inserts. These differences may reflect different properties of the kits for cDNA synthesis which were used to construct the libraries (Amersham for the root libraries and Pharmacia for the wounded shoot library). Large inserts (preferably full length cDNA clones) were desirable for subsequent work to isolate genomic sequences which generally lie just 5' to the transcription start site.

Evaluation of Differential Hybridization for Isolating Specific Sequences

The differential hybridization approach permited screening of large numbers of clones at a time and both of the shoot and root specific sequences were recovered independently several times. In contrast, screening individual cDNA library clones by Northern analysis is labour intensive but extremely sensitive and transcripts which represent 2.5 x 10^{-5}% of total polyA mRNA are detectable by that approach (Evans *et al.*, 1988).

Using the formula $N=\ln(1-p)/\ln(1-1/n)$, it is estimated that a cDNA library of $1x10^6$ (N) clones is required in order to be 99% (p) certain of isolating a low abundance transcript, 1 part in 250,000 (1/n). A transcript which comprises 0.0078% of the total mRNA population has a 99% probablility of being represented in the wounded shoot library (of 0.59 x 10^5 clones). However, previous results imply that the differential cDNA hybridization is only sufficiently sensitive to isolate specific cDNAs which are of intermediate abundance (0.05% to 2% of the polyA RNA: Maniatis *et al.*, 1982; Gasser *et al.*, 1989; Conkling *et al.*, 1990). A transcript present at 0.0078% in a cDNA probe would hybridize so weakly to its corresponding cDNA clone as to go unnoticed in the differential screening process. Therefore the predicted complexity of the wounded shoot library was adequate with respect to the sensitivity of the technique because a primary library of only 9,208 clones would be sufficiently complex to include clones which represent 0.05% of total mRNA, with 99% certainty. The inability to recover wound specific sequences is likely to be due to their presence in the cDNA probes at undetectably low levels (*ie.* less than 0.05%). Transcripts of very low incidence are best isolated from a cDNA library constructed from an mRNA population depleted of repeated sequences using subtractive processes and by screening such a library with subtractive probes (Sambrook *et al.*, 1989).

In the course of screening the amplified wounded shoot cDNA library, clones carrying more than one insert were identified. These clones presumably arose as a result of inoptimal ratios of vector to cDNA insert during ligations for library construction. Although none of the clones which were assayed at random from the primary library carried more than a single insert (Table 1), the percentage of such clones in the amplified library is unknown. Multiple insert carrying clones may be misclassified by hybridization. If for example a shoot and root specific insert were contained in a single clone, their combined expression patterns would be apparently constitutive.

Inability to exactly standardize probing experiments also limited the application of differential cDNA hybridization and undoubtedly lead to the misclassification of clones, especially in early rounds of screening. A low level constitutively expressed gene isolated from sugarcane would be a suitable internal reference to equalize hybridization. Perhaps the speed and accuracy of analysis of clone hybridization patterns could be enhanced by the development and application of a computerized system.

Frequency of Appearance of cDNA Clones in Wounded Shoot and Root Libraries
It was surprising that the apparently root specific #B6cii sequence appeared at about one half the frequency of the shoot specific sequence (clone #32) in the wounded shoot library (12 versus 20 clones per 1500). Northern analysis showed that the shoot specific sequence is present in the shoot RNA population at a frequency many times that at which #B6cii appears. One possible explanation of the unpredicted high frequency of appearance of #B6cii in the wounded shoot library is that clones containing that sequence are susceptible to amplification at a greater rate than most other clones in the library (see Kaiser and Murray 1983). This possibility could be tested by determining whether the appearance of the #B6cii homologous sequence increases disproportionately (in comparison to other clones in the library) when the library is amplified.

The frequency of hybridization of #B6cii to clones of the root library gives an estimate that the sequence represents about 0.2% of the polyA mRNA population in roots and this falls central in the range of frequencies of root specific transcripts isolated from tobacco using differential hybridization (Conkling *et al.*, 1990).

Northern Analysis of Root and Shoot Specific Transcripts
Northern analysis of clones #32 and #B6cii confirmed their tissue specificity as predicted by the Southern analysis of these clones by differential hybridization. #B6cii hybridized a root transcript of a size similar to the cloned cDNA suggesting that the #B6cii cDNA represents a near full length transcript. In addition, the transcript is expressed at low levels in shoots, and is apparently slightly longer than the root transcript. Perhaps the root transcript is derived from a different gene or is a differently processed transcript of the same gene from which the shoot transcript is generated. Densitometric analysis was not sufficiently sensitive to measure the extent of shoot transcript hybridization to #B6cii. However, the area below the highest peak of root transcript hybridization to cDNA #B6cii is 191 fold greater than the area below the lowest peak (Table 4) and the #B6cii transcript must therefore be at least 191 times more abundant in root RNA compared to shoot RNA.

Clone #32 hybridized a major transcript of about 3 kb as well as numerous other less prominent transcripts, possibly indicating that the transcript is derived from a member of a multigene family. Assuming that clone #32 is derived from the major transcript, the cDNA is approximately one third the length of its progenitor transcript. This clone is very strongly shoot specific.

The tissue specificity of the two transcripts could be regulated by transcription initiation, transcript elongation, post transcriptional modification, transcript export from the nucleus or stability of the transcript. However, it is most likely that transcriptional regulation occurs primarily at initiation of transcript synthesis (see Waugh and Brown, 1991). Kamalay and Goldberg (1984) have established that 70% of transcript specificity in roots of tobacco is attributable to transcriptional initiation control and levels of the remaining 30% of transcripts

are regulated by post-transcriptional mechanisms. The near full-length cDNA clone #B6cii is therefore a good candidate as a probe for high probability of recovering potential root-specific promoter sequence from a sugarcane genomic library.

Nucleotide Sequence of Clone #B6cii

The putative start codon of the 231 bp open reading frame of #B6cii lies amongst a sequence (AAGATGTC) consistent with plant eukaryotic consensus (AACATGGC: Joshi, 1987a; Lütcke *et al.*, 1987). A polyadenylation signal (AATACAA: Joshi, 1987b) begins 29 bp upstream of the cDNA polyA tail. The palindrome which lies just 5' to the putative translation start site may be functionally significant. Primary and secondary structure of mRNA transcript leader sequences can influence translational control (see Gallie *et al.*, 1989) and it is possible that the palindrome is involved in gene regulation.

The Root Specific cDNA has Homology to Metallothioneins

When GenBank nucleic acid sequence accessions were screened for homology to the complete nucleotide sequence of #B6cii, the most closely homologous sequence was the putative coding region of a barley root abundant transcript which is induced in response to iron deficiency (Okumura *et al.*, 1991). The region of overlap of the barley sequence with #B6cii encompasses 291 bp starting at the proposed translation start site of both sequences and finishing about 45 bp downstream of the proposed translation stop site of both sequences. Nucleotide homology over that region was 67%. Sequences in the PIR Protein amino acid sequence database were screened for homology with the putative translation product of the #B6cii transcript and the best alignment was 54.8% identity with the 73 amino acid putative translation product of the same barley sequence extracted by the Genbank data base search. In addition, other polypeptides derived from published nucleic acid sequences which did not appear in the Genbank data base were found to have substantial homology (52.6% to 34.2%) to the sugarcane root specific sequence: a maize metallothionein-like 76 amino acid encoding transcript which is expressed at high levels in root and low levels in leaves, pith and seed (de Framond *et al.*, 1991); a barley 115 amino acid encoded by an aleurone specific transcript (Klemsdal *et al.*, 1991); a heavy metal repressible, metallothionein-like, root abundant transcript which encodes a 74 amino acid polypeptide isolated from the flowering plant *Mimulus guttatus* (de Miranda *et al.*, 1990) and a pea metallothionein-like, 75 amino acid encoding transcript which is expressed specifically at high basal levels in roots and at low levels in leaves (Evans *et al.*, 1990). Furthermore, high degrees of homology for short lengths were found to known metallothionein class I proteins isolated from horse, cow, rat and purple sea urchin by virtue of the short cysteine rich regions at the amino and carboxy termini of the sugarcane polypeptide. These same cysteine rich regions of the sugarcane putative polypeptide showed the highest homology with the barley, *Mimulus*, pea and maize sequences. Like the sugarcane polypeptide, the four other root specific plant polypeptides are characteristically short and consist of two cysteine rich, metallothionein-like regions which flank a cysteine-free (approximately 40 amino acid) region. The longer barley aleurone specific sequence has a cysteine rich region at only the amino terminus of the polypeptide only.

Metallothioneins are metal binding proteins found in animal systems and are thought to function in the accumulation, detoxification and metabolism of metal ions. The proteins are characteristically rich in cysteine residues which readily bind heavy metal ions and have been classified into class I and class II types (Hamer, 1986). Phytochelatins are the functional equivalent of animal metallothioneins in plants. They are classified type III metallothionein

because they are synthesized by a heavy metal responsive, enzyme catalysed, stepwise addition of dipeptidyl (-glu-cys) units (Grill *et al.* 1987). Until recently, synthesis of mRNA encoded class I metallothioneins in plants was unconfirmed (see de Miranda *et al.*, 1990; Evans *et al.*, 1990 and de Framond, 1991) and their function as such is still not verified.

Using the criteria established by Fiser *et al.* (1992), the cysteine residues of putative protein sequences derived from the sugarcane, pea, maize, *Mimulus* and barley root specific cDNAs were analysed for their potential to participate in cystine formation. The nature of the amino acids adjacent to the cysteine residues in the proteins indicate that these residues are all highly likely to be half cystine and therefore capable of bonding with other potentially half cystine residues or with heavy metal atoms.

The five plant metallothionein-like root specific proteins (sugarcane, pea, maize, *Mimulus* and barley) have been aligned to show partial homology in the region which intervenes the cysteine rich, metallothionein-like domains (Figure 14).

Mimulus guttatus	SMYPDME TNTTVTMIEGVAPLK MYSEGSEKSFGAEGG
Pea	NKRSSGLS YSEMETTETVILGVGPAK IQFEGAEMSAASEDG
Barley	GKMYPDLEEKSGATMQVTVI VLGVGSAY VQFEEAAEFGEAAHG
Maize	GKKYPDLEETSTAA QPTVV LGVAPEKKAAPEFVEAAESGEAAHG
Sugarcane	GKMYPDLAEKSAAAAPATSSALHLRRGT SGPGWRSRGVEAGHG

Figure 14. Alignment of the cysteine-free intervening regions of five plant root-specific-metallothionein-like polypeptides. Gaps were inserted in the sequence to demonstrate all regions of homology. Amino acids in bold face type appear in corresponding positions in at least three of the five polypeptides.

The conservation of amino acid sequence favours a functional role for the intervening region. Perhaps the region has macromolecule binding properties and/or is cleaved from the two cysteine rich regions in post-translational modification (de Miranda *et al.*, 1990). Either of these possibilities could be responsive to heavy metal status because the intact protein is likely to take on a different structure depending on whether the cysteines are bound to heavy metal or other cysteine residues.

As is the case for the metallothionein-like cDNAs isolated from pea (Evans *et al., 1990*) and maize (de Framond, 1991), the sugarcane cDNA is expressed at high steady state levels in the plant roots. The #B6cii cDNA clone is near full length and therefore provides a starting point for the isolation of upstream genomic elements which may be involved in the regulation of root specific transcription initiation.

Cloning and Characterization of the promoter of a gene transcribed primarily in sugarcane roots

Southern Analysis of Shoot and Root Specific cDNA Clones and Recovery of Homologous Genomic Clones

In Southern analysis of various sugarcane cultivars (Figure 11), bands of hybridization to the #B6cii and #32 probes varied in intensity and some bands were shared by all cultivars. This pattern of hybridization is reminiscent of previous results of RFLP analyses of sugarcane and its wild relative species using randomly cloned genomic fragments and oat cDNAs as probes (Burnquist, 1991). Extensive variability may be a consequence of heterozygosity and

polyploidy of sugarcane. A given locus exists in many copies and in various forms in the complex polyploid sugarcane genome. Disfunction of some alleles in the genome need not leave an individual with a selective disadvantage if others remain functional and dosage of a particular allele may be different between individuals (see Glaszmann *et al.*, 1989). Perhaps the persistent inheritance of at least one band in all varieties, as observed for both #B6cii and #32 probes (Figure 11), is consistent with preferential inheritance of a specific allele on a highly conserved chromosome which was derived from one of the progenitor species. It can be concluded that the cDNA libraries which were generated for the purpose of this project are a good source of RFLP markers for application to sugarcane crop improvement via genome mapping and varietal identification (see Tanksley *et al.*, 1989; Burnquist, 1991).

The various genes which hybridized the #B6cii and #32 cDNAs may all be allelic and reside at analogous sites on homologous and homoeologous chromosomes. Alternatively, some related genes may reside at different loci, as for a multigene family. They may be functionally distinct, regulated differently or be extensively modified, inactive pseudogenes. Recognition of these possibilities is important because the presence of differently regulated and pseudogene family members in the genome may confound cloning of a functional promoter (see Waugh and Brown, 1991).

Genomic Library Construction

The high frequency of methylated cytosine in plant genomes is known to adversely affect genomic cloning efficiency because of their cleavage by bacterial methylation dependent restriction systems (Raleigh *et al.*, 1988). This effect was important in production of the sugarcane genomic library, as the number of clones produced on bacterial host LE392 (McrA$^-$) was only 7% of that on either KW251 or ER1648 hosts, both of which are deficient for the McrA and McrBC methylation dependent restriction systems (see Raleigh *et al.*, 1988).

Based on an estimate for sugarcane "diploid" (2C) genome size of 12.5 pg (Bennett and Smith, 1976) which corresponds to 1.1×10^7 kb (given 1 Mbp = 1.08×10^{-3} pg), a library of 2.6×10^6 clones is required for a 99% probability of containing a given sequence amongst clones of 19.5 kb average size (from Clark and Carbon, 1976). The actual library contained only 23% of this number of clones and provided enough coverage of the genome to include a given 19.5 kb sequence with 65% probability. Sequences homologous to the #32 and #B6cii cDNAs were identified within only 90,000 library clones. The high (*ca.* 10) ploidy of sugarcane may contribute to this high rate of clone recovery due to the presence of more than one allelic pair per genome, and the screened number of library clones correspondingly represented about one haploid fraction of the sugarcane polyploid genome.

Strategy for Cloning the Promoter of the Gene Transcribed Primarily in Sugarcane Roots

As shown above, the #B6cii cDNA is likely to be near full-length and the sequence of the cDNA clone includes a putative translation start site. The sequence primed by oligo1 directly from the 15 kb genomic clone λgem11B6cii, revealed a "TATA" motif, which is the core promoter element of many eukaryote genes, and a "CAAT" motif, which is often associated with eukaryotic promoters, upstream of the putative transcription start site and compatible with previously studied plant promoters (Lütcke *et al.*, 1987; Joshi, 1987a). Since *Eco* RI restriction sites were absent from this sequence, it was reasoned that the *Eco* RI subclone (pB6ciiERI) must contain at least 350 bp of promoter sequence between oligo1 and Reverse-M13 (Figure 5). This permitted rapid isolation of the promoter sequence by PCR as shown

in Figure 5, without extensive restriction mapping or detailed sequence analysis of the 15-kb genomic clone required for alternative strategies (cf. Conkling *et al.*, 1990; Evans *et al.*, 1990; de Framond, 1991). PCR amplification experiments generated minor products which were larger than the 1 kb amplicon chosen for characterization. These minor amplicons were interpreted to be derived from spurious cloning products, concatamerized plasmids or cloned partial digest products, and did not interfere with recovery of the major 1 kb amplicon.

Southern Analysis of Genomic DNA Probed with the #B6cii Promoter

Results of Southern analysis of sugarcane genomic DNA probed with the promoter amplicon may have indicated conservation of the 5' non-coding region of the members of the #B6cii gene family, however, closer inspection of the promoter sequence (see below) found a putative transcription start site 34 bp downstream of the TATA box and 21 bp upstream of the cDNA end. Hence, up to 56 bp of transcribed sequence could be present at the 3' end of the promoter amplicon probe. This 56 bp sequence has a theoretical melting temperature of 57°C (Bolton and McCarthy, 1962) and is likely to resist dissociation from a homologous sequence under the hybridization and post-hybridization washing conditions which were used (see Sambrook *et al.*, 1989). Lack of target sequence homology in the extensive untranscribed 5' sequence of the amplicon could contribute to decreasing the disocciation temperature of the probe (see Wetmur, 1991) but empirical studies of the hybridizing properties of the promoter probe should be undertaken to determine this (see Sambrook *et al.*, 1989). At present it is the possibility that the transcribed portion of the probe contributed solely to the hybridization pattern can not be excluded. A similar experiment using as a probe, sequence exclusively 5' of the TATA box, would be valuable to establish the degree of conservation of the promoter region amongst the related genes. Conservation of promoter sequence in all of the #B6cii homologous genes would increase the likelihood that the cloned amplicon represented a functional promoter region with the expected specificity, rather than sequence adjacent to a highly modified pseudogene or a gene regulated differently to the one from which the #B6cii sequence was transcribed.

Sequence Analysis of the Root Specific Promoter

Three putative cap sites were located between the TATA box and the translation start site by the "Signal scan" computer programme (Prestridge, 1991). Cap sites in plant promoters are often surrounded by CA, followed preferentially by pyrimidines (Bucher and Trifonov, 1986) and two such sites lie 52 bp and 57 bp downstream of the TATA, and 3 bp either side of the cDNA end. A candidate cap site lies 34 bp downstream of the TATA. The three possible transcription start sites are all separated from the TATA consensus by a distance which is consistent with many other plant promoters and the cDNA clone is probably, at most, 21 bp short of full-length.

A range of repeated sequence and palindrome motifs, which are frequently indicative of *cis*-acting regulatory sequences, were located in the 813 bp promoter clone along with short sequence motifs known to bind transcription regulating *trans*-acting factors in other species (Figure 12; Prestridge, 1991). In particular, a 5'-3' TGACGT motif lies on the antisense strand, immediately adjacent to the same sequence oriented on the sense strand (*ie.* ACGTCATGCAGT on the sense strand), 41 bp upstream of the TATA box. The TGACG motif, known as the as-1 "activation sequence", is tandemly repeated in the CaMV35S promoter and contributes principally to root expression in tobacco *via* its interaction with the *trans*-acting "activator sequence factor", ASF-1 (Lam *et al.*, 1989). The sequence complement of as-1, "CGTCA", corresponds to a cyclic AMP responsive element which has

been reported in animal promoters. The motif binds a protein (CREB) which has DNA binding properties similar to ASF-1 (Lam and Chua, 1989). This CREB binding sequence appears in the sugarcane promoter (within the TGACG arrangement described above) and in the promoters of maize and pea metallothionein-like genes which show substantial sequence homology to #B6cii and are expressed at high basal levels in roots. The CGTCA motif also occurs in a 340 bp element from tobacco which determines root specific expression (Yamamoto *et al.*, 1991) but the ability of this sequence to interact with specific proteins and its role in conferring root specific expression have not been tested in pea, maize or tobacco.

Because the promoter controls a metallothionein-like gene, it was examined for sequences known to be resonsive to heavy metals. A TGCAGACC motif has one base pair mismatch with a MRE-like motif in the promoter of a pea metallothionein-like gene (Evans *et al.*, 1991). The pea MRE (TGCACACC) fits the animal metal regulatory element consensus (TGCRCNCX: R=G/A, N=G/C/T, X=G/C) except for the A residue at position N in the pea sequence which has been shown to abolish metal regulation and permit high basal levels of gene expression in animals (see Evans *et al.*, 1991). The sugarcane motif has an additional deviation (underlined) from the MRE consensus, and is therefore not expected to be metal responsive. The promoter of the corresponding maize metallothionein-like gene contains no MRE-like motif (de Framond, 1991). The sugarcane promoter contains additional sequence motifs in common with the promoters of the analogous metallothionein-like genes isolated from roots of pea and maize (see Figure 15).

While the function of the motifs in the sugarcane promoter (see Figure 12) are unknown, their recognition provides a starting point for devising strategies for deletion analysis of the promoter, once functionality of the intact sequence has been confirmed.

Element 3 is identical in sugarcane and maize, but does not occur in pea.

g t c c a a a a a

Element 6 does not appear in pea. The underlined 7 bp subelement appears a second time in sugarcane, downstream of the transcription start site.

sugarcane	g t t c t t g g a a g c a a
maize	g t t c t t g c a a g c a a

Element 9

sugarcane	t g a a a a c g a c a g c a
maize	t g a a a a a g t c a a g c
pea	t g a a a a c g a t g a t a

Element 10

sugarcane	t a a t t c t g c a g a c
maize	g a a t t c g t c a g c c
pea	t a a t t c t t c a g c t

Figure 15. Alignment of conserved sequence motifs which appear in the promoters of metallothionein-like genes isolated from sugarcane, pea and maize (see also Figure 12).

Constructs likely to drive constitutive gene expression in sugarcane

Our results using GUS assays to measure transient expression of genes introduced by electroporation into sugarcane and carrot cells lead to the following conclusions:

(i) NOP and 35SP are weak but approximately equivalent strength promoters in sugarcane (100 to 500-fold lower expression than in carrot).
(ii) A doubled (back to back) 35S promoter results in approximately doubled expression levels in sugarcane, but no change in carrot.
(iii) Inclusion of a 180bp fragment including the OCS enhancer immediately upstream of 35SP increases expression approximately 5-fold in sugarcane and 2-fold in carrot.
(iv) Inclusion of intron 1 from maize *Adh 1* between the promoter and coding region decreases expression from both NOP and 35S in sugarcane, in contrast with the trend from cereals.
(v) The Emu promoter, with multiple enhancers close to the core promoter region, is even more strongly enhanced relative to 35SP in sugarcane (400-fold) than in cereals (10 to 50-fold). It is equivalent in strength to 35SP in carrot, as previously reported in *Nicotiana*.

The only other non-cereal monocot investigated for response to various promoter elements is *Lolium multiflorum* (Italian rye grass). Interestingly, *Adh1* intron 1 did not enhance gene expression in *Lolium*. However, of the tested monocots, *Lolium* also showed least enhancement of expression from the Emu promoter (9-fold relative to 35SP). Thus conclusions about gene regulatory sequences from studies on maize or a few cereals may not be generally applicable, even in the *Poaceae*. More species must be tested before broad predictions can be made about promoter strength in the monocots.

The plasmids described here provide a range of promoter strengths in sugarcane. The very strong expression from the Emu promoter has already been useful in optimising microprojectile bombardment conditions and selecting transgenic sugarcane plants. If intermediate expression levels are required for genes of agronomic importance in sugarcane, this should be readily achievable by altering the number of enhancer elements and their proximity to the core promoter region.

Intellectual property

The knowledge gained about sugarcane promoters as a result of this project has commercial significance in the sense that it is likely to be developed and applied in future projects using genetic engineering for sugarcane improvement. Some of this new knowledge could be considered eligible for some form of intellectual property protection. For example, the information on the probable sugarcane root-specific promoter sequence could form a part of a patent claim covering application of the sequence. The primary motivation for such a patent claim would be the possibility of :

(i) a return to investors in the research (SRDC & UQ) through royalties paid by licensed users of the promoter sequence, or

(ii) appropriate cross-licensing agreements which allowed the investors (effectively the Australian Sugar Industry) to use specific other valuable patented technologies without the need to make royalty payments (because the other patent holders saw equal value in access to "our" promoter sequence).

These are not trivial possibilities, but there are also compelling arguments against the pursuit of patent protection:

(i) The technology requires additional research to make it patentable (effectively a practical demonstration of any practical applications to be claimed in the patent). To some extent this is in progress, but it will take some time, during which there may need to be strict confidentiality about the existing sequence information, to protect patentability.

(ii) Because the project has been largely undertaken by postgraduate students (which makes the research extremely cost-efficient), inclusion of the sequence information in publicly available theses may be inconsistent with the requirements for patenting.

(iii) The need for sustained confidentiality about existing results until the technology is developed to a patentable level also complicates our relationship with members of the International Consortium for Sugarcane Biotechnology, and other forms of scientific exchange (e.g. conference presentations which may be important for career development of graduates involved in the project, or for attracting other sources of support for continuation of the research).

(iv) The costs associated with applications for patent protection are high, and only a small proportion of awarded patents ever generate any royalty returns. The likely level of such returns may be lower in agriculture than, for example, in pharmaceuticals.

(v) It follows that at least one of the investors in the research needs to make a commitment to cover the direct and indirect costs associated with protecting patentability and securing patents, from a very early stage of the research. At such early stages it is very difficult to estimate the probability of patentable outcomes, or the potential value of royalties / agreements resulting from such a patent.

(vi) Without this commitment to added costs at the outset, it is perhaps inevitable that most research must be conducted by more cost-efficient means which are not compatible with the requirements for subsequent patenting.

This general issue needs further consultation with SRDC. It is not intended to pursue the possibility of patent protection of the probable sugarcane root-specific promoter unless this is requested by SRDC.

Implications and Recommendations

Further research is required to isolate and characterise additional gene regulatory sequences which result in useful patterns of expression of foreign genes introduced into sugarcane. This involves isolation and characterisation of selected promoters to the stage of applicability for practical sugarcane variety improvement; and development and testing of techniques to isolate a range of promoters as required in the future to control the expression of agronomically useful genes introduced into sugarcane.

Specific recommended objectives, which build on the successful outcomes of SRC projects UQIS and UQ6S are:

(1) Complete the testing of differential hybridisation of cDNA libraries for isolation of specific promoter types from sugarcane and thoroughly characterise the isolated probable root specific promoter for potential use in genetic engineering for pest and disease resistance.
(2) Evaluate microprojectile-mediated transfer of promoterless luciferase genes to tag sugarcane promoters controlling a range of gene expression patterns.
(3) Evaluate transient gene expression following microprojectile bombardment of various sugarcane tissues as a rapid test of function and specificity in sugarcane of promoters available from other plant species.
(4) Screen selected promoters (or promoter components) from other plant species for desired patterns of gene control in sugarcane.

This work to identify effective promoters is an essential component of a balanced research program to develop genetic engineering for sugarcane improvement. These promoters will be essential to make the step from experimental testing of useful genes to practical use in improved commercial sugarcane varieties.

It is not yet appropriate to attempt cost/benefit analysis of a strategic research project such as UQ1S. However, it is fair to say that (i) the project has been very successful in achieving its objectives, and (ii) the research has been very cost-effective because of the substantial involvement of outstanding graduate students, particularly Tricia Franks and Carl Rathus.

Publications

This report is based substantially on the publications below, which list in full all other references cited in this report.

Birch, R.G., Franks, T., Rathus, C.S.O., and Basnayake, W.V.S. (1991). Genetic transformation, promoter tagging and disease resistance in sugarcane. J. Cell. Biochem. **Suppl. 15A**, 111.

Franks, T. (1992) Microprojectile-mediated gene transfer ans isolation of promoter sequences for sugarcane genetic engineering, Ph.D. Dissertation. Department of Botany, The University of Queensland.

Franks, T., and Birch, R.G. (1991). Microprojectile techniques for direct gene transfer into intact plant cells. In Advanced Methods in Plant Breeding and Biotechnology, D. Murray, ed (Wallingford: CAB International), pp. 103-127.

Franks, T., and Birch, R.G. (1993) Isolation and characterization of a metallothionein-like gene and promoter expressed at high levels in sugarcane roots. Proceedings of the 15th Annual Conference on the Organisation and Expression of the Genome, Lorne, February 1993.

Rathus, C., Bower, R., and Birch, R.G. (1993). Effects of promoter intron and enhancer elements on transient gene expression in sugarcane and carrot protoplasts. Plant Mol. Biol. (submitted).