ISOLATION AND CHARACTERISATION OF A CULM-SPECIFIC PROMOTER ELEMENT FROM SUGARCANE

By

Tsion Goshu Abraha

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science at the Stellenbosch University

Supervisor: Professor F. C. Botha
Co-Supervisor: J H. Groenewald

Institute of Plant Biotechnology, Department of Botany and Zoology
Stellenbosch University

April 2005

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or part been submitted at any university for a degree.

…….. ……

Tsion Goshu Abraha Date

ABSTRACT

Sugarcane (*Saccharum* spp) is an important crop worldwide and is cultivated for the high level of sucrose in its mature internodes. Because of the exhaustion of the genetic potential in the commercial sugarcane germplasm conventional breeding has not lately been able to enhance sucrose content. Currently there is a concerted effort to improve culm sucrose content by genetic engineering which will require appropriate transgenes and promoters. One of the major constraints to genetic engineering of sugarcane is the lack of stable promoters required to drive tissue- or organ-specific expression of transgenes. Tissue and developmental stage specific promoters allow targeting of transgene activity and in doing so reduce the impact on non-target tissues. These promoters could also be advantageous to manipulate certain aspects of sucrose metabolism specifically in mature culm tissue. In addition, no promoters are currently freely available to the South African Sugar Industry for use in their transgenic program. The primary goal of this project was therefore to isolate a mature tissue-specific promoter for use in transgenic sugarcane plants.

The approach followed was firstly, to identify an endogenous gene expressed in the desired pattern, and then to isolate the corresponding promoter from the sugarcane genome. cDNA macroarrays were initially used to identify differentially expressed sequences. The tissue specificity of potential clones was confirmed using RNA blot analysis. Two clones (c23-a and c22-a) were isolated and confirmed to be mature culm specific. Clone c22-a (putative dirigent-like protein) was selected for promoter isolation based on its culm tissue specific expression pattern and its proximity to the 5' end of the gene. Furthermore, to confirm the activity of this promoter in the storage parenchyma cells, the exact cellular localisation of the transcript in the mature tissue was determined through *in situ* hybridisation. *In situ* hybridisation results confirmed the presence of the transcript in the parenchyma cells of mature culm tissue only. Moreover, the transcript is present in high concentrations in the parenchyma tissues surrounding the vascular bundles and parenchyma cells of the vascular complex.

The selected dirigent-like gene was sequenced to allow the design of primers that could be used for the isolation of the corresponding promoter region using a long-range inverse PCR (LR-

iPCR) method. Using these we have successfully isolated two highly homologous promoter regions of the dirigent like gene of respectively 1151 and 985 base pairs. *In silico* analyses confirmed the presence of various transcription motifs, including a TATA-box. However, experimental verification is needed to fully assess the functionality of these promoter regions. Verifying the activity of the isolated promoters through transient expression analysis proved to be problematic because of their highly mature culm specificity. Both constructs are therefore being used to obtain stable transformants in which promoter activity can be evaluated in mature internodal tissues.

OPSOMMING

Suikerriet (*Saccharum* spp) is wêreldwyd ‘n belangrike gewas wat verbou word vir die hoë sukrose inhoud in volwasse internodes. As gevolg van die uitputting van die genetiese potensiaal in die kommersiële suikerriet kiemplasma, is konvensionele teling onsuksesvol om sukrose opbrengs te verhoog nie. Tans val die fokus op genetiese manipulering om die suiker opbrengs te verhoog en een van hoof fokusse val op ‘n soektog na geskikte transgene en promoters vir hierdie doel. Een van die grootste beperkings vir genetiese manipulering van suikerriet is die tekort aan stabiele promoters wat vereis word vir weefsel- of orgaanspesifieke uitdrukking van transgene. Weefsel- en ontwikkelingstadium-spesifieke promoters laat die teikening van transgeen aktiwiteit toe en so kan die impak op nie-geteikende weefsels verminder word. Hierdie promoters kan ook voordelig wees om sekere aspekte van sukrose metabolisme spesifiek in volwasse stam weefsel te manipuleer. Daar is tans geen promoters beskikbaar vir gebruik deur die Suid-Afrikaanse suikerbedryf in hulle transgeniese suikerriet program nie. Die hoofdoel van hierdie projek was dus om ‘n volwasse weefsel spesifieke promoter, wat in transgeniese suikerriet gebruik kan word, te isoleer.

Die strategie wat gevolg is sluit in eerstens om ‘n endogene geen met die verlangde uitdrukkingspatroon te identifiseer en om daarna die meegaande promotor vanuit die suikerriet genoom te isoleer. cDNA **“**microarrays” is eerstens gebruik om differensieël uitgedrukte volgordes te identifiseer. Die weefselspesifisiteit van potensieële klone is bevestig deur gebruik te maak van RNA klad analises. Twee klone (c23-a en c22-a) is geïsoleer en is as spesifiek tot die volwasse stam bevestig. Kloon c22-a, wat kodeer vir ‘n putitatiewe “dirigent”-soort proteïen, is geselekteer vir promotor isolasie, gebaseer op die stamweefsel spesifisiteit van die uitdrukking van die geen en die nabyheid aan die 5’-kant van die geen. Om die aktiwiteit van die promotor in die bergingsparenkiem selle te bevestig is die presiese sellulêre uitdrukkingspatroon bepaal met behulp van *in situ* hibridisasie. *In situ* hibridisasie het bevestig dat die geen slegs in die parenkiem selle van die volwasse stam uitgedruk word. Verder is daar ook vasgestel dat die transkrip aanwesig is in hoër konsentrasies in die parenkiem selle wat die vaatbondels omring en in die parenkiem selle van die vaatbondel kompleks.

Die geselekteerde "dirigent"-soort geen se volgorde is bepaal met behulp van 'n lang-afstand omgekeerde PKR-metode (LR-iPCR), om die ontwerp van inleiers toe te laat wat gebruik kon word vir die isolasie van die promoter. Deur van hierdie metode gebruik te maak is twee promoter volgordes, 1151 en 985 basis pare elk, hoogs soortgelyk aan die "dirigent"-soort geen geïsoleer. Die teenwoordigheid van verskeie transkripsie motiewe, insluitende 'n TATA-boks, is bevestig deur gebruik te maak van *in silico* analises. Eksperimentele verifikasie is egter nodig om die funksionaliteit van hierdie motiewe te bevestig. Die volwasse stamspesifisiteit het dit egter moeilik gemaak om die aktiwiteit van hierdie promotors te toets deur gebruik te maak van tydelike uitdrukkingsanalise. Beide konstrukte word dus tans gebruik om stabiele transformante te genereer wat gebruik kan word om die aktiwiteit van die promotors in die volwasse internode weefsel te evalueer.

AKNOWLEDGMENTS

First and foremost, I would like to give all the glory and honour to almighty God, who has given me strength, determination and wisdom to complete my thesis.

I would like to thank the following people:

I am profoundly grateful to my supervisors Professor Frikkie Botha and Hennie Groenewald whose detailed criticism of earlier drafts, valuable suggestions and wise counsel have proved most invaluable for the completion of this work. It has been a most rewarding learning experience to work under their guidance.

I feel a deep sense of gratitude for my mother who formed part of my vision and taught me the good things that really matter in life. Most of all, she taught me to persevere and prepared me to face challenges with faith and humility. She has been and is still a constant source of inspiration to my life.

I reserve great thanks and appreciation to my husband, Seyoum, for his continuous encouragement, prayer, love, support and patience during the course of my studies. My son, Mikias, deserves special thanks for his love and who doesn't need to ask anymore when I will be finished with school.

I am grateful to all the staff members and students at the Institute for Plant Biotechnology, Stellenbosch University, who provided me with technical assistance, advice and encouragement throughout my studies.

Special thanks go to the Eritrean Human Resources Development (EHRD), Eritrean Government for funding my studies and Mrs. Lula Gebreyesus for facilitating my study program.

I thank the Stellenbosch University for financial support and especially the International office for their support in every aspect of my stay in South Africa.

Thanks go to the South African Sugar Research Institute for funding to carry out this study.

Special thanks to Dr. Katherine Denby and Sally-Ann Walford for allowing me to complete this thesis while working with them in the Capar group (University of Cape Town). Their willingness and generosity in providing both time and academic support have greatly added to this work.

I observe special thanks for a dear colleague and a friend, Awot Kiflu, for her unlimited moral support, assistance and continuous encouragement throughout my study period. My special thanks are due to Million Girmay for his encouragement and critical proof reading of this thesis.

Lastly, I owe special thanks to Medahanie Zerabruk for allowing me unrestricted access to his computer facility during the preparation of the thesis.

CONTENTS

LIST OF FIGURES AND TABLES

Figures:

Tables:

LIST OF ABBREVATIONS

^{o}C	degrees centigrade
µg	microgram
µl	microlitre
µCi	microCurie
ATP	adenosine 5'-triphosphate
ATPase	adenosine 5'-triphosphatase
ACT1	rice actin-1 promoter
Adh1	maize alcohol dehydrogenase-1 promoter
BLAST	basic local alignment search tool
bp	nucleic acid base pairs
ca.	*'circa'* / approximately
CaMV	cauliflower mosaic virus
cDNA	complementary deoxyribonucleic acid
CHS15	bean chalcone synthase gene promoter
cm	centimetre
CTP	cytosine 5'-triposphate
DEPC	diethyl pyrocarbonate
DIG	digoxigenin
DNA	deoxyribonucleic acid
dNTP	deoxynucloside triphosphate
DPEs	downstream promoter elements
EDTA	ethylenediaminetetraacetic acid
eIFs	eukaryotic Initiation Factors
EST	expressed sequence tag
EtOH	ethanol
g	gram
g	gravitational acceleration (9.806 m/s)
GTP	Guanosine 5'-triposphate
GUS	β-glucuronidase

hr	hour
HATs	Histone Acetyltransferases
HCl	hydrochloric acid
HPR	hydroxypyruvate reductase
Inr	initiator
IPCR	inverse polymerase chain reaction
LiCl	lithium chloride
LR-iPCR	long-range inverse polymerase chain reaction
kb	kilobase
M	molar
m/v	molar per volume
mg/ml	milligram per millilitre
$MgCl_2$	magnesium chloride
Ml	mililitre
mM	milimolar
mm	millimetre
MS	Murashige and Skoog medium
MW	molecular weight
mas	mannopine synthase
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
NaOH	sodium hydroxide
NBT/BCIP	Nitro blue tetrazolium chloride/ 5-bromo-4-chloro-3-indolyl-phosphate
NTE	Sodium Tris- ethylenediaminetetraacetic acid buffer
nos	nopaline synthase
ocs	octapine synthase
PABP	poly(A)-binding protein
PBS	phosphate buffered saline
pH	acidity
PIC	pre-initiation complex
pol-II	ribonucleic acid polymerase II

RNA	ribonucleic acid
RNAse	ribonuclease
rbcS	riblose-1,5-bisphosphate carboxylase-oxygen gene
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate
SDHB	succinate dehydrogenase subunit B
Spp.	species
TBE	Tris(hydroxymethyl)-aminomethane borate ethylenediaminetetraacetic acid buffer
TAF	TBP associated factor
TBP	TATA binding protein
TFs	transcription factors
tRNA	transfer RNA
Tris	Tris(hydroxymethyl)-aminomethane
TSS	transcription start site
UsnRNA	uridylate rich small nuclear ribonucleic acid
UTR	untranslated region
Ubi1	maize polyubiquitin-1 promoter
UTP	uridine triphosphate
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
X-Gluc	5-bromo-4-chloro-3-indolyl- β-glucuronic acid

CHAPTER ONE

General Introduction

Sugarcane (*Saccharum* spp.) is a C_4 grass grown predominantly in tropical and subtropical regions for the production of sucrose. Cultivated sugarcane varieties are derived from complex interspecific hybridization between the species *S. officinarum* (2n = 80) and *S. spontaneum* (2n = 40-128) (Butterfield et al., 2001). It is being produced by both small-scale, traditional farmers, and modern, large-scale commercial farmers. Worldwide, more than 100 countries produce sugar (http://www.illovo.co.za). Sugarcane that is grown primarily in the tropical and subtropical zones of the Southern hemisphere is the source of 72% of world's sucrose while the rest is produced from sugar beet grown in the temperate zones of the northern hemisphere (http://www.illovo.co.za).

Sugarcane has been grown and milled in Southern Africa for centuries. Excellent growing condition, high yielding cane varieties and relatively low milling costs combine to boost the Southern African developing community to be one of the world's largest sugar producing regions. In this region an annual average of 5 million tons of sucrose is produced, of which South Africa has the largest share (http://www.illovo.co.za).

The South African sugar industry is one of the world's leading cost competitive producers of high quality sugar and contributes a great deal to the economy of the producing provinces (http:// www.sasa.org.za). It is a diverse industry that combines the agricultural activities of sugarcane cultivation with the industrial factory production of molasses, raw and refined sugar. The industry generates direct income and is an employment source in the regions within which it operates. It is estimated that the industry contributes R 2.38 billion to the country's foreign exchange earnings annually (http://www.sasa.org.za). Thus, the sugar industry has both social and economic significance to South Africa. Sucrose, as an agricultural product, is one of the main exported products by South Africa with a significant revenue generation every season (approximately 6 billion Rand). However, only two million tons of sugar is obtained from 20 million tons of harvested sugarcane (http:// www.sasa.org.za). Thus, it is one of the sugar

industry's objectives to increase the sucrose yield and thereby to increase its share of the internationally competitive market.

As with other crops, the modification of carbon partitioning by conventional breeding has made a major contribution to the increase of the sucrose content of sugarcane over the last century. Moore et al. (1997) attributes these increases in sugarcane yield and sucrose content to the success in overcoming the productivity barriers in both source and sink tissues. Under conditions that favour sucrose accumulation, the sugarcane stalk can store up to 25% of its fresh weight as sucrose, but the field performance of some varieties of sugarcane is less than 2% of their fresh weight (Moore et al., 1997). Hence, genetic manipulation of sugarcane aimed at increasing photoassimilates, directed towards stored sucrose in the culm, is a valuable and potentially profitable goal for both the breeders and environmentalists (Grof and Campbell, 2001).

Over the past decades, conventional breeding has mainly been used as an approach towards improving yield and resistance to the numerous pathogens and pests. Improvements in sugarcane using conventional plant breeding techniques are time-consuming and showed little success more recently. This has left genetic manipulation as an attractive alternative to introduce new desirable traits into sugarcane plants as a means of improving them (Yang et al., 2003; Schenk et al., 1999). The powerful combination of genetic engineering and conventional breeding techniques allows useful traits encoded by trangenes to be introduced into crops within an economically viable time frame (Hansen and Wright, 1999). Furthermore, this technique assists us to understand the role of existing sugarcane genes in complex traits such as yield or sugar content, and can introduce valuable 'novel' genes for new properties in sugarcane. Examples of these types of genes are those for insect or pest resistance (Birch, 1997).

Successful genetic manipulations of sugarcane would require appropriate genetic constructs, containing a promoter, the transgene and a terminating signal, to facilitate the integration and expression of foreign DNA in plants (Birch, 1997). An important part of the successful manipulation of a plant is getting the transgene to be expressed appropriately; that is, in the right organ and tissue, at the correct developmental stage and at the proper level (Stitt and Sonnewald, 1995). In plants and other organisms, control and regulation of the transgene expression can

occur at different stages, but particularly during transcription. The importance of this process is reflected in the percentage of genes in the genome that are implicated in transcription (Singh, 1998). For example, 15% of the genes encoded for on a 1.9 Mb fragment of chromosome 4 of *Arabidopsis thaliana* are involved in the transcription process (Bevan et al., 1998).

The promoter sequence ensures the control of transcription by interacting with other *trans*-acting, sequence specific, DNA binding proteins called transcription factors. These transcription factors are cell type specific and interact with promoter elements to drive gene expression in different cellular types, activating particular groups of genes in the different tissues of the particular organism (Meshi and Iwabuchi, 1995; Singh, 1998). The regulation of the differential expression of genes between different cell and tissue types is therefore dependent on the promoter elements of these genes. Successful genetic manipulation is therefore, to a large extent, dependent on the availability of appropriate promoters to achieve the desired expressions.

Promoters used for genetic manipulation are classified according to the way in which they control gene expression. They can be divided into two major classes, namely constitutive and tissue-specific or developmentally regulated promoters. Constitutive promoters direct expression in almost all tissues and are largely independent of environmental and developmental factors (Kuhlemeier et al., 1987). Tissue-specific promoters control gene expression in a tissue-dependent manner and according to the developmental stage of the plant and show a well-defined temporal expression pattern. The transgenes driven by these types of promoters will only be expressed in tissue where the transgene product is desired, leaving the rest of the tissue in the plant unmodified by the transgene expression (Stitt and Sonnewald, 1995).

Much emphasis is currently being placed on the identification of gene regulatory elements, which provide tissue and developmental specificity in sugarcane. There are currently very few promoters that have been shown to be stably active in transgenic sugarcane. The CaMV 35S promoter has been widely used for high-level constitutive expression in dicotyledonous, but shows lower levels of expression in monocotyledons (Wilmink et al., 1995; Grof and Campbell, 2001). Among the promoters tested in sugarcane, the *Emu* promoter and the maize ubiquitin promoter showed higher levels of expression than the CaMV 35S promoter (Gallo-Meagher and

Irvine, 1993; Rathus et al., 1993). Although some research groups have obtained successful transgene expression in sugarcane using promoters of other species, many groups have found aberrant or no expression with heterologous gene constructs following proven integration into the sugarcane genome (Birch, 1997). One explanation for silencing is that foreign or artificial promoters may lack some undefined features necessary for stable gene expression in sugarcane (Birch, 1997). Furthermore, promoters like maize ubiquitin and CaMV 35S are constitutively expressed which may increase the metabolic load in sugarcane plants. Therefore, successful transformation is dependent on the availability of tissue-specific promoters to achieve specific expression.

At present, no regulatory sequences are available that drive transgene expression in sugarcane in a tissue/organ-specific manner. In addition, patent limitation on already available promoters and genes, active in different plants, necessitates the isolation of novel regulatory elements from sugarcane. The overall aim of this project was therefore to isolate a mature culm-specific promoter from sugarcane that could be used to drive stable trangene expression in sugarcane.

The layout of this thesis is as follows: Chapter 2 will introduce some background knowledge on the origin, morphology and genetic make-up of sugarcane, regulation of gene expression in eukaryotes and our current knowledge of plant promoters. The identification, isolation and characterization of transcripts expressed in specific tissue and cell types are described in Chapter 3. From this work clone c22-a was identified as an appropriate target for promoter isolation. The successful isolation and computational sequence analysis of the corresponding promoter region is described in Chapter 4. Finally, in Chapter 5, the relevant results of the above mentioned chapters are discussed and conclusions are drawn from the characterisation of the culm-specific promoter.

CHAPTER TWO

Literature Review

2. 1. Introduction

Sugarcane is internationally regarded as one of most important crop species. Extensive plant breeding programs are in place in many countries around the world. However, it is becoming increasingly evident that little or no progress has been made to increase yield in sugarcane through conventional breeding. For this reason several research programmes in the world are focusing on the potential to genetically manipulating the crop through transgenic technology.

Successful implementation of transgene technology is dependent on three factors namely; an efficient transformation system, availability of suitable new genes, and promoter elements. Undoubtedly the lack of suitable promoter elements is one of the major stumbling blocks for progress in this programme.

The purpose of this chapter is three fold: firstly there will be a broad overview of the origin, morphology and genetic make up of sugarcane will be provided. Secondly, there will be a discussion of the regulation of gene expression in eukaryotes, and thirdly there will be a broad discussion of the current knowledge of plant promoters.

2. 2. Sugarcane

2. 2. 1. The origin of sugarcane

Sugarcane is one of the most important crops in the world, mainly cultivated for the high sucrose content in its stalk. Modern commercial cultivated sugarcane varieties are interspecific hybrids originating from several species of the genus *Saccharum*. *Saccharum* is a member of the Andropogoneae tribe of the grass (Poaceae) family and includes six species (Stevenson, 1965). The two wild Saccharum species are S*accharum spontaneum* and *Saccharum robustum* (2n = 60-170). The cultivated species *Saccharum officinarum* probably originated from *Saccharum*

robustum (Irvine, 1999). The other two cultivated species *Saccharum barberi* and *Saccharum sinense* are thought to be natural hybrids of S. *spontaneum* and *Saccharum officinarum*. The last one, *Saccharum edule,* has an intergenic origin between *Saccharum officinarum* or *Saccharum robustum* and the *Miscanthus* species (Daniels and Roach, 1987).

Current cultivated clones are essentially derived from interspecific hybridisation performed between *S. officinarum* (2n = 80) and *S. spontaneum* (2n = 40-128). *S. officinarum* is characterized by its high sucrose content in the stalk compared to other Saccharum species. *S. spontaneum* is used by sugarcane breeders due to its stress tolerance, vegetative vigor and disease resistance (Butterfield et al., 2001). The progenies from these hybridizations were backcrossed with *S. officinarum* (noble cane), a process called nobilisation, to recover the high sucrose phenotype (Bremer, 1961). During the process of nobilisation, there is asymmetric chromosome transmission (Bremer, 1961). In the hybridization process between *S. officinarum* (2n = 80) and S. *spontaneum* (2n = 40-128), the female parent (*S. officinarum*) transmits 2n gametes, giving rise to a 2n + n transmission. Crossing a F1 hybrid with *S. officinarum* again results in a 2n + n transmission, but in later generations the chromosomal transmission becomes normal (n + n) (Bremer, 1961; Lu et al., 1994; Butterfield et al., 2001). The consequence of this is that modern cultivated varieties have chromosome numbers between 2n = 100-130. Commercial sugarcane varieties therefore have complex polyploid genomes (Grivet and Arruda, 2002).

2. 2. 2. The Genetic complexity of sugarcane

In the plant kingdom, where polyploidy occurs much more frequently than in the animal kingdom, it is estimated that approximately 70% of Angiosperm spp. are polyploid (Pikaard, 2001). Many important crops, including banana, canola, coffee, maize, potato, oats, soybean, sugarcane and wheat are polyploids (Wendel, 2000; Osborn et al., 2003). Polyploidy could result from the duplication of a single genome (autoployploidy) or from combining two fully differentiated genomes into a common nucleus (allopolyploidy) in one of the parental cytoplasms. Both *S. officinarum* and S. *spontaneum* are thought to have complex autopolyploid genomes (D'Hont et al., 1996; Ming et al., 2001).

As stated above the current sugarcane cultivars originated from a cross between two or more than two polyploidy Saccharum spp, which makes the sugarcane cultivars to be genetically complex (Grivet and Arruda, 2002). *S. officinarum* and *S. spontaneum* have a basic chromosome number of x = 10 and x = 8 respectively. As a result of the difference in their basic chromosome number two distinct chromosome organizations co-exist in current varieties (Grivet and Arruda, 2002).

Polyploid organisms have multiple copies of a specific chromosome, which means that multiple copies of a gene are therefore present (Wendel, 2000). The fate of redundant genes resulting from genome duplication is poorly understood. When two different or related genomes are combined in a single cell they must respond to the consequence of genome duplication, especially multiple copies of genes with similar or redundant functions. One possible outcome of gene duplication might be the silencing of one of the duplicated copies, which is the loss or inactivation of the gene (Wendel, 2000). The silenced gene will remain in the genome as a pseudogene, accumulating mutations until it is no longer recognizable (Wendel, 2000). The most detailed description of the formation of a pseudogene from a duplicated gene is that of the PgiC2 gene in *Clarkia mildrediae* (Gottlieb and Ford, 1997). Of the 23 exons 18 were sequenced and 9 of them showed insertions or deletions, causing frame shift mutations and the insertion of stop codons. Some deletions also resulted in the loss of exon-intron splice junction.

Pseudogenes have also been characterized in the uridylate rich small nuclear RNA (UsnRNA) genes in plants. A pseudogene, truncated at the 3'-end and lacking the sequences necessary for the transcription at the 5' flanking region, was cloned from a tomato genomic library (Kiss et al., 1989). In potato, three out of ten UsnRNA genes were found to be pseudogenes with defective promoter or coding regions (Vaux et al., 1992). A non-functional promoter has also been isolated from sugarcane genomes. The peroxidase cDNA fragment (Spx42) was used to isolate the corresponding promoter sequence from a sugarcane genomic library. The functionality of the isolated promoter was tested by fusing it to a reporter gene in callus or in sugarcane plants, and it was found to be silenced (Birch et al., 1996).

After polyplodisation, genomic redundancy can occur at different levels; namely, duplicate chromosomes, duplicate genes and duplicate regulatory regions driving expressions. Each level

of redundancy might be subjected to the process of mutation (Force et al., 1999). Mutation in coding sequence or in regulatory regions may occur by several mechanisms including nucleotide substitutions, deletions or insertions of transposable elements (Force et al., 1999; Wendel, 2000). Gene silencing by insertion of transposable elements has been demonstrated in several plant species. In hexaploid wheat, a 8 kb insertion of a retrotransposon in the coding region leads to the loss of glutenin expression at the Glu-1 locus (Harberd et al., 1987). In tobacco, non-functional nitrate reductase genes were found to be a result of the insertion of a *copia*-like retrotransposon, *Tnt1* (Grandbastien, 1992). Alteration of expression can also be caused by insertion of transposable elements into regulatory regions as demonstrated in pea rbcS (White et al., 1994), maize R-s (May and Dellaporta, 1998) and the *nivea* chalcone synthase gene in *Antirrhinum* (Lister et al., 1993).

Polyploidy might also raise a problem for gene regulation. The expression of most genes is dependent on a network of regulators such as transcription factors (TFs). The numbers of TFs in a diploid network is high, but in a polyploid they can be expanded several fold and as a result the regulatory network may be modified. One way by which the organism solves this problem is by turning off or turning down the expression of some copies of some genes (Kellogg, 2003). Thus, a silencing strategy could balance the advantage and the disadvantage of having multiple copies of orthologous genes or gene products (example transcription factors) in polyploid cells.

As has been discussed above, recovery of functional promoters from sugarcane may therefore be complicated by the high polyploidy. Some or most copies of genes/promoters present in eight to ten copies in modern sugarcane cultivars may have accumulated mutations which rendered them inactive.

2. 3. Functional Morphology

2. 3. 1. Morphological features of the culm

Sugarcane is a perennial grass with tall culms bunched into stools, which are usually erect. Culms are divided into a number of joints, each consisting of a node and an internode

(Artschwager, 1925). The culm is a complex organ composed of epidermal, vascular, meristematic and storage parenchyma tissues (Moore, 1995). Numerous vascular bundles permeate the storage parenchyma tissue and are surrounded by a fiber sheath and two or more layers of thick walled, lignified sclerencyhma cells. Storage parenchyma cells become lignified at later stages of development (Artschwager, 1925).

Jacobsen et al. (1992) observed changes in morphological features and sucrose content of the sugarcane culm through histochemical and sugar assays. These include an increase in the number of vascular bundles and a concomitant decrease in size from the core to the peripheral tissues and an increase in lignification and suberisation down the culm in parallel with increased sucrose concentrations (Figure 2.1). In addition, Moore and Cosgrove (1991) observed a difference in the length and diameter of the storage parenchyma cells between young and mature internodes. It is therefore likely that the metabolic activity in these different cell types of the culm will be tailored to the specific function of the cell. Thus, to understand the significance of culm-specific gene expression at the cellular level, it is important to consider the specialized function of the different cell-types in the sugarcane culm.

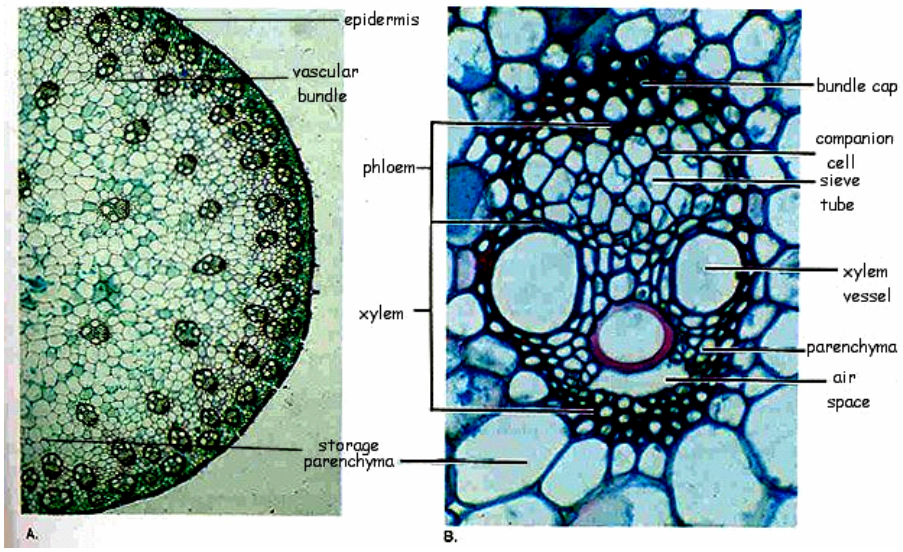


Figure 2.1: Cross section of a typical sugarcane stem (A) including the different cell types in the vascular bundle (B) (from Ellmore, 2000).

2. 3. 2. Metabolic activities in the culm

Sucrose is the most important low-molecular weight carbohydrate produced in higher plants. In many plants, photosynthate, which is transported to storage tissues as sucrose, is converted to starch for long-term storage (Komor, 2000). The maturation of sugarcane is characterized by the accumulation of sucrose in the internodes (Moore and Maretzki, 1996). Sucrose is synthesized in the photosynthetic tissues (source tissues/leaves) and exported into the sink tissue of the sugarcane culm through the phloem as a long term-storage molecule (Moore, 1995; Zhu et al., 2000). The sink tissues, which may not have the ability to photosynthesize, rely on the supply of carbon in the form of sucrose for their different metabolic activities (Moore and Maretzki, 1996). Since different organs have different biological functions and biochemical requirements, sucrose is channeled according to these differences.

In sugarcane, sucrose unloaded from the phloem passes through three distinct cellular compartments, that is the apoplastic space (cell wall), the metabolic compartment (cytoplasm) and the storage compartment (vacuole) (Hongmei et al., 2000). Each compartment contains enzymes that contribute to sucrose degradation and synthesis. Physiological studies of key enzymes associated with sucrose metabolism such as sucrose synthase, sucrose phosphate synthase and the various isoenzymes of invertase (neutral invertase, soluble acid invertase and cell wall bound acid invertase) show that enzyme activity and expression vary depending on internode maturity (Moore, 1995; Zhu et al., 1997; Lingle, 1999; Vorster and Botha, 1999; Rose and Botha, 2000).

Based on the differences in morphological features and enzyme activities, it can be expected that genes are differentially expressed between the different tissue types in the culm and at the different developmental stages. Recently research approaches using expressed sequence tags (ESTs) have shown the differential expression of genes in sugarcane (Carson and Botha, 2000, 2002; Carson et al., 2002; Casu et al., 2003; Ma et al., 2004). Genes, which are differentially regulated are under the control of specific regulatory sequences (promoter regions) generally located immediately upstream of the gene (Birch, 1997). Hence, the identification and isolation of genes expressed under specific conditions will allow the isolation of their specific promoter

elements. These can then be used to regulate transgene expression in a temporal and spatial specific manner.

Numerous genes, and in some cases their associated promoters, which exhibit a wide range of tissue and/or developmental specificity have been characterized. Examples include genes that are specifically expressed in microspores (Custer et al., 1997), potato stolons (Trindade et al., 2003), embryos, cotyledons, endosperm, (Edwards and Coruzzi, 1990), leaves, roots, and fruits (Coupe and Deikman, 1997; Edward and Corruzzi, 1990).

Many factors need to be taken into account to improve strategies for genetic manipulation of sugarcane metabolism. Control of a particular metabolic pathway is shared by more than one regulatory enzyme and the activity of various enzymes in a pathway is highly coordinated and may require treatment as a quantitative trait (Stitt, 1995; Moore and Maretzki, 1997). The effects of the compartmentation of metabolism and metabolic channeling must be considered. Sucrose storage in sugarcane culm, parenchyma cells, is a highly regulated process, where anatomical features, metabolic reactions and transport through membranes are in close interaction (Jacobsen et al., 1992; Moore and Maretzki, 1997). All of these may require detailed information of relevant gene expression and the fate of the product of that expression. This information will aid decision-making in the choice of targets for manipulation. In addition to the identification of targets, effective manipulation of the complex metabolic process in sugarcane will require highly specific promoters able to regulate gene expression in a highly controlled manner, to ensure that spatial and temporal constraints are met.

2. 4. Regulation Of Gene Expression

The topic of gene regulation has received much attention because the key for the development of complex organisms does not lie as much in the entire number of genes but rather in their specific regulation and interaction. The characteristics of an organism are encoded in its DNA. In the eukaryotes nucleus, most of this information resides within thousands of genes and is organized into linear chromosomes. Plant mitochondria and plastids also contain circular DNA molecule (Lefebvre and Gellatly, 1997).

Genes consist of promoters (sequences of DNA that tell RNA polymerase where and usually when, to begin transcription), transcribed region and terminators (DNA sequences that tell RNA polymerase to halt transcription and release the RNA and DNAs) (Lefebvre and Gellatly, 1997). Genes encode information that specify functional products, either RNA molecules or proteins, used for various cellular functions. These proteins can be divided into a number of groups. Some examples include structural proteins that give form to a cell or an organism, proteins required for DNA synthesis, replication and cell division (DNA polymerase). In addition, proteins are involved in the synthesis of RNA (the proteins comprise RNA polymerase) and some others in regulating the activity of the organism, ensuring that metabolism is controlled and adapts to changes in the environment so that development occurs in the correct manner (Lefebvre and Gellatly, 1997).

Other genes encode specific RNA molecules, which are not translated into proteins but perform functions within the cell. Examples include the various RNAs that are major components of ribosomes (rRNA), and the transfer RNAs (tRNA), which are responsible for incorporating the correct amino acids during protein synthesis. By contrast, regulatory sequences do not encode a product. Yet without them a cell would be unable to express genes in an organised way. Nor could the cell coordinate the expression of the hundreds of thousands of genes in its nucleus, but select only certain genes for expression, and activate on them at precise moments in development. DNA on its own is not capable of catalysing any reaction, or building any structure. The information in DNA must be developed within the cell and translated into molecules that perform specific functions. This process of decoding and converting genetic information into molecules that perform the function of the organism is called gene expression. Gene expression is controlled by sequences (promoter regions) generally located immediately upstream of the coding gene, which determine the strength, developmental timing and tissue specificity of expression of the adjacent gene, and therefore, plays a crucial role in successful plant transformation (Birch, 1997). Figure 2.2 clearly indicates the very important role played by promoter elements during gene expression (Botha, 2000).

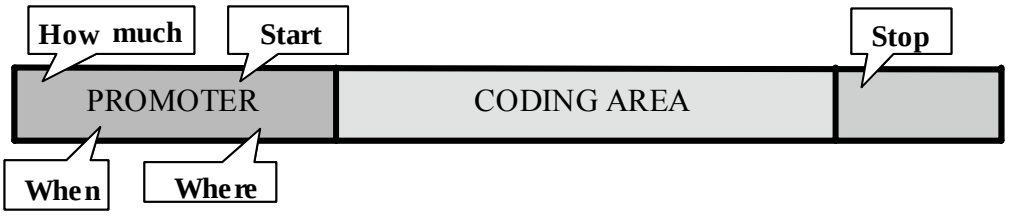


Figure 2.2: A schematic representation of the important regions in a plant promoter and other components of genetic construct which may be used for the genetic manipulation of plants (from Botha, 2000)

Even though every single cell of a multicellular organism contains all genetic information at all times, only a fraction of it is active in a given tissue (Tyagi, 2001). The temporal and spatial expression of specific genes is central to the process of development, differentiation and homeostasis in eukaryotes, and is regulated primarily at the level of transcription. Plants, during their development and differentiation, need to integrate different types of tissue, developmental and environmental signals to regulate complex patterns of gene expression (Singh, 1998). For example, seed-storage proteins are only expressed in the seed not in other parts of the plant, but they also are only expressed during a short period of time during the development of the seed (Singh, 1998). This variability is reflected in the organisation of promoters and regulatory elements as well as in genes characterised for the regulatory factors (Tyagi, 2001).

Gene expression involves a number of biochemical complex steps. Messenger RNA (mRNA) must be produced by RNA polymerase in a process called transcription. In cytoplasm the genetic code carried on the mRNA is translated into amino acid sequences by ribosomes. The net result is a polypeptide whose amino acid sequence is determined by the sequence of bases in the mRNA. Even though regulation occurs at all stages of protein synthesis, the control on the transcriptional level is an important regulation mechanism in gene expression. If this process fails in some respect, it will affect all the other steps that follow the production of the initial RNA transcript (for review of these stages, see Latchman, 1998).

2. 4. 1. Regulation at the level of transcription in the nucleus

Eukaryotes have three different RNA polymerases that are responsible for transcribing different subsets of genes: RNA polymerase I (pol I) transcribes genes encoding ribosomal RNA (rRNA),

RNA polymerase II (pol II) (which I will focus on in this part) is specific for the transcription of protein-coding genes (mRNA) and certain small nuclear RNAs, while RNA polymerase III (pol III) transcribes genes coding for transfer RNA (tRNA), and other small RNAs (Novina and Roy, 1996; Allison et al., 1985). In plants, there are also mitochondrial RNA polymerase and chloroplastic RNA polymerase. All of these enzymes are large complex enzymes composed of ten or more types of subunits (Allison et al., 1985). Although the three main polymerases (I, II and III) differ in overall subunit composition, they do contain some subunit in common. The existence of three different RNA polymerases acting on three different sets of genes (ribosomal genes, protein-coding genes and tRNA genes) implies that at least three different types of promoters exist to maintain specificity.

Transcription initiation by pol II is regulated by TFs interacting with transcription elements, and also with each other, and by an open chromatin structure that enables the factors to access the DNA (Nikolov and Burley, 1997). Transcriptional regulation may involve two levels of regulation: one involving chromatin unfolding and another involving the assembly of transcription machinery at pol II promoters. The first may be a prerequisite for the second, that is, chromatin unfolding may expose promoters for the assembly of the transcription machinery.

2. 4. 1. 1. Chromatin structure

The presence of large number of genes in eukaryote genomes would make it difficult should all of them compete for the components of the basal transcription machinery at the same time. Most genes are transcribed only inside a specific tissue or under rarely occurring conditions. The first level of transcriptional control in eukaryotes is at the level of chromatin structure. The DNA in eukaryotic cells is not naked, but packaged into a highly organised and compact nucleoprotein structure known as chromatin (Orphandies and Reinberg, 2002; Singh, 1998). The basic organisation unit of chromatin is the nucleosome, which consists of 146 bp of DNA wrapped almost twice around a protein core containing two copies of four histone proteins H2A, H2B, H3 and H4 (Luger et al., 1997). The physical structure of the DNA, as it exists compacted into chromatin, can affect the ability of transcriptional regulatory proteins (transcription factors) and RNA polymerase to find access to specific genes and to activate transcription from them

(Orphandies and Reinberg, 2002; Singh, 1998). However, not all chromatin are equal. Untranscribed regions of the genome are packaged into highly condensed "heterochromatin," while transcribed genes are present in more accessible "euchromatin" (Richards and Elgin, 2002).

Each cell type in eukaryotes packages its genes into a unique pattern of heterochromatin and euchromatin (Orphandies and Reinberg, 2002). This shows that each cell has found a way to shut down large regions of the genome that are not needed within a certain tissue effectively. This also guarantees that all cells of a tissue stay committed to expressing the same genes without losing parts of the genome.

Transcription is associated with structural changes of chromatin, called chromatin remodelling. Modifications of chromatin structure include acetylation, phosphorylation, methylation and ubiquination (Berger, 2002). The most characterised modification to date is acetylation, modification linked mainly to transcription activation (Berger, 2002; Kouzarides, 2002). The core histones in transcriptionally active regions are often acetylated by the action of Histone Acetyltransferases (HATs) (Berger, 2002; Singh, 1998). Addition of acetate to numerous lysine residues in the amino-terminal tails that protrude from the surface of the nucleosome reduces the affinity of the histone for DNA (Berger, 2002; Singh, 1998). There are deacetylases that remove the acetate from the histone which mediate transcriptional repression (Kouzarides, 2002). The yeast Gcn5 and mammalian p300/CBP proteins were initially identified as co-activators and were found to be required for histone acetylases. Histone acetylation has also been observed in plants (Belyaev et al., 1997) and a maize histone deacetylase has been identified (Lusser et al., 1997).

Chromatin remodelling also requires the involvement of protein complexes that actively displace nucleosomes, hydrolysing ATP in the process (Berger, 2002; Singh, 1998). A good example is the SWI/SNF complex, which creates hypersensitive sites in chromatin and stimulates the binding of the TFs to the regulatory sequences of the DNA (Berger, 2002; Singh, 1998).

2. 4. 1. 2. DNA methylation

In plants, mammals and some other organisms there is an additional level of gene regulation mediated through the modification of DNA by cytosine methylation (Finnegan, 2001). DNA

methylation can inhibit transcription by blocking the binding of basal transcriptional machinery or TFs through the modification of target sites (Bird and Wolffe, 1999; Finnegan, 2001). Inactive DNA contains nucleotides (especially cytosine) that have a methyl group attached to it. Most methylated DNA will remain inactive during differentiation and repeated cell divisions of that cell line (Finnegan, 1998). Methylation is probably important in preventing the transcription of the genes intended to be permanently turned off.

2. 4. 1. 3. The basal transcription machinery

Transcription initiation is directed by DNA sequences that lay upstream of the initiation site of a gene (Maniatis et al., 1987). This region of a gene, its promoter, can be seen to consist of a core promoter, a proximal promoter region, and distal enhancers, all of which contain transcription elements, short DNA sequence patterns that are targeted by specific auxilary proteins called TFs (Nikolov and Burley, 1997; Butler and Kadonaga, 2001; Tyagi, 2001). The core promoter is responsible for guiding the pol II to the correct transcription start site (TSS) (Maniatis et al., 1987).

Eukaryotic pol II does not recognise nor does it directly bind to the core promoter sequence, rather it recognises and binds to the TF that bind specifically to the promoter region (Vomt Endt et al., 2002; Orphanides et al., 1996). TFs play important roles in gene expression including chromatin remodelling and the recruitment of pol II transcription-initiation complex (Singh, 1998). TFs can be divided into a number of functional classes of which the major one is the activators and repressors (Singh, 1998). These proteins bind to specific DNA sequences that are present only in certain promoters and give rise to gene specific regulation (Singh, 1998). Co-activators or co-repressors are the second class of TFs. These proteins mediate the transcriptional effects of specific activator/repressors. Compared to the activators/repressors, this group of TFs are not able to bind directly to the DNA on their own, but they can still be promoter specific as a result of protein-protein interaction with specific activators and repressors (Singh, 1998). The third class of TFs are the general transcription factors, which are important components of the pol II transcription-initiation complex (Singh, 1998).

Accurate initiation of transcription depends on assembling a pre-initiation complex (PIC) containing pol II and at least six TFs, from the general initiation factors (TFIID, TFIIB, TFIIF, TFIIE, TFIIA and TFIIH) (Nikolov and Burley, 1997; Singh, 1998). The best characterised core promoter elements in eukaryote gene promoters are the TATA element located 25-30 bp upstream of the TSS and the less-well characterised a pyrimidine-rich initiator element located at the start site (Nikolov and Burley, 1997; Roeder, 1991; Yamaguchi et al., 1998). The TATA box is a target of TFIID, or more specifically, one component of TFIID, the TATA binding protein (TBP) (Stargell and Struhl, 1996). Assembly of the general transcription machinery is initiated by TBP binding with a variety of TBP associated factors (TAFs) in the form of TFIID. This is followed by the association of the remainder of the TFs and pol II (for review see Stargell and Struhl, 1996; Roeder, 1996). With the exception of TFIID and probably TFIIB, all TFs are recruited by protein-protein interaction.

There are however some promoters referred as TATA-less promoters which do not contain any TATA box (Smale, 1997; Zhu et al., 1995). In these promoters, the exact position of the transcriptional start point may instead be controlled by another element known as the initiator (Inr) (Smale, 1997; Zhu et al., 1995). Another type of promoter element, which was discovered in both humans and *Drosophila*, is present in some TATA-less, Inr-containing promoters about 30 bp downstream of the TSS (Butler and Kadonaga, 2001; Burke and Kadonaga, 1997). This element, which is known as the down stream promoter elements (DPEs), appears to be a down stream substitute of the TATA box in assisting the Inr in controlling precise transcriptional initiation (Burke and Kadonaga, 1997). TATA-less promoters are not well characterised in plants, though it was found that the tobacco eiF-4A promoter (Mandel et, al., 1995) and the maize ZMDJ1 promoter (Baszzynki et al., 1997) are TATA-less promoters.

2. 4. 1. 4. Activated transcription and regulatory sequences

In contrast to the basal transcription, activated transcription requires the entire promoter region that includes the core region plus proximal and distal enhancer regions (Nikolov and Burley, 1997). These elements are located at varying distances from the TSS. The proximal elements are adjacent to the core promoter while enhancers can be positioned several kilobases upstream or

downstream of the promoter depicting enhancers to be orientation independent (Nikolov and Burley, 1997). Both types of elements are binding sites for specific transcriptional regulatory proteins that increase the level of transcription from core promoters (Mitchell and Tjian, 1989; Maniatis et al., 1987). This phenomenon is referred as activated transcription. Activated transcription requires TBP and the remaining subunits of TFIID (TBP- associated factors), the other general initiation factors TFIIB, TFIIF, TFIIE, and TFIIH, plus transcriptional activators and coactivators (Nikolov and Burley, 1997). Enhancer elements located thousands of nucleotides away from the core promoter can still have an activating effect. This mechanism is puzzling, though it can be explained by the fact that the DNA strand is flexible and can form bends and loops that allow for activator proteins located at distant sites to interact with the core transcriptional machinery at the promoter (Adhya, 1989; Maniatis, 1987). This activated transcription is summarized in figure 2.3.

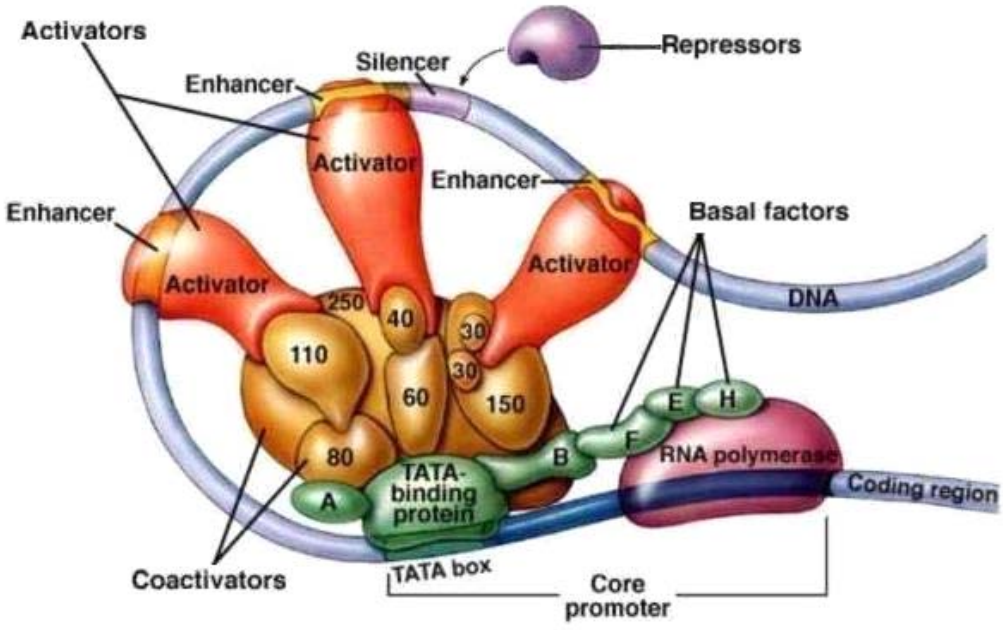


Figure 2.3: The structure of activated transcription complex (from Raven and Johnson, 1999).

Regulatory regions, controlling the transcription of eukaryotic genes, typically contain several transcription factor binding sites organized over a large region. Some of these individual binding

sites are able to bind several different members of a family of TFs. The SV40 enhancer best illustrates this characteristic (Maniatis et al., 1987). In the case of the SV40 enhancer at lease five different DNA-binding proteins are known to interact with certain parts of the enhancer sequence (Maniatis et al., 1987). If specific parts of the enhancer are disrupted, the other parts continue to function. Therefore, the SV40 enhancer consists of a collection of cis-acting DNA elements that bind to a variety of different trans-acting factors (Maniatis et al., 1987; Dynan, 1989). The particular factor binding to a given site, therefore, not only depends on the binding site, but also on what factors are available for binding in a given cell type at a given time (Zhang, 2003).

It is the modulation and combinatorial nature of the transcriptional regulatory regions that makes possible to precisely control the temporal and spatial expression patterns of the tens of thousands of genes present in higher eukaryotes (Dynan, 1989; Singh, 1998). Thus any gene will typically have its very own pattern of binding sites for transcriptional activators and repressors ensuring that the gene is only transcribed in the proper cell type(s) and at the proper developmental time. Other genes are expressed in response to environmental stimuli such as light, heat, cold and pathogen attack only (Vom Endt et al., 2002; Guilfoyle, 1997). TFS themselves are, of course, also subjected to similar transcriptional regulation. This can be achieved via a transcriptional cascade, as exemplified by VP1/C1 in seed maturation of maize (Vom Endt et al., 2002; Schwechheimer and Bevan, 1998). In addition, TF genes can be influenced by external stimuli such as light, hypoxia, salt stress, abscisic acid and gibbrelic acid (Liu et al., 1999).

2. 4. 2. mRNA processing

The RNA molecules produced by pol II are not the mature messenger RNAs until they have been processed and modified and leave the nucleus. The result of transcription is the pre-mRNA or primary transcript. The last transcriptional events, which lead from primary transcript to the final mRNA serving as a template for translation, starts with the capping of the pre-mRNA. A cap structure consists of 7-methyl guanosine residue linked in an unusual way (5'-5' link) to the 5' end of the RNA (Vasudevan and Peltz, 2003). Contrary to the modification at the 5' end which involves the adding of a single nucleotide, the 3' end is cleaved, a large RNA stretch is removed

and up to 200 adenosines are added. The cap serves as the binding for translation initiation factor eIF, and is also necessary to protect an RNA from degradation enzymes (Day and Tuite, 1998). Similar to the cap at the 5' end, the poly(A) tail serves as protection against degradation, and serves as the binding for the poly(A)-binding protein (PABP) (Day and Tuite, 1998).

The next step in RNA processing is splicing. Eukaryotic gene expression requires the removal of noncoding sequences (introns) from the precursor mRNA (pre-mRNA) (Will and Lührmann, 2001). The precise removal of pre-mRNA introns is the critical aspect of gene expression. Splicing occurs in a complex structure known as spliceosome, involving a number of RNA and protein factors and holds the upstream and downstream parts of the mRNA in the correct place while cutting out the intron (Albà and Pagès, 1998; Day and Tuite, 1998; Reed, 2000). If the introns are not removed, the mRNA can not be translated to give the complete protein. Not only do the splicing machinery recognize and remove introns to make correct messages for protein production but also for many genes, alternative splicing mechanism generate functionally diverse protein isoforms in a spatially and temporally regulated manner (Graveley, 2001).

Alternative splicing is regulated by tissue specific factors promoting a certain splice site as well as by the ratio balance of several proteins belonging to the spliceosome (Graveley, 2001; Lorković et al., 2000). Few examples of alternative splicing have been elucidated in plants. Alternative splicing of hydroxypyruvate reductase (HPR) gene in pumpkin produces two different proteins with different cellular localization (Lorković et al., 2000). In addition, a gene which is alternatively spliced to give two entirely different proteins, ribosomal protein S 14 (RPS 14) and succinate dehydrogenase subunit B (SDHB), is also observed in rice (Lorković et al., 2000). After the mRNA has been brought to the final shape it is transported from the nucleus through the nuclear membrane into the right place in the cytoplasm.

2. 4. 3. Translational regulation

Once a mature mRNA is produced in the nucleus, and transported to the cytoplasm, it must be translated by the protein biosynthetic machinery of the eukaryotic cell (ribosomes) (Orphanides and Reinberg, 2002). Protein synthesis in eukaryotes is a complex series of steps and involves a

number of protein translation factors that work in conjunction with ribosome and tRNA to decode a mRNA, thereby producing the encoded polypeptide chain (Day and Tuite, 1998; Gallie and Browning, 2001). The translation process can be divided into three different stages: initation, elongation and termination (Belly-Serres, 1999; Day and Tuite, 1998). One of the most important targets for translational control is at the initiation step. Control of translation initiation on individual transcript is modulated by structural properties of the mRNA namely the 5' untranslated region (5' UTR), 3' untranslated region (3' UTR), the coding region and the interaction between the 5' and 3' UTRs of the mRNA (Day and Tuite, 1998). Features of the 5' UTR that affect the translational efficiency include the 5' cap structure (7-methylGpppN), the leader length and sequence and the presence of secondary structures (reviewed in Kozak, 1991).

Translation is initiated by the binding of a ribosome at the cap structure on the 5' end of the mRNA. Cap recognition involves several eukaryotic Initiation Factors (eIFs) (Gallie and Browning, 2001; Bailey-Serres, 1999). A subunit of ribosome then migrates along the mRNA until it finds an appropriate start codon (AUG). General control at this stage is possible by inhibiting compontes of the ribosomes; specific mechanisms interact with patterns in 5' and 3' UTR of the mRNA that form characteristic secondary structures. Translation is influenced by the stability of the mRNA which determines the number of times that it is translated (Day and Tuite, 1998). RNA stability is an effective means to control the rate of protein synthesis, especially in cases where a rapid and transient change of a specific protein level is necessary, and is often accompanied by a change in transcription rate.

To summarise, gene expression involves which genes are used, which modifications are carried out to the transcript, and how efficiently the final product is synthesised.

2. 5. Plant Promoters

The regulation of gene expression in plants is controlled at both the transcriptional and post-transcriptional level. To turn a gene into a protein product, at least two general steps are required: the gene is transcribed, spliced and processed to form mRNA and then the mRNA is translated into a polypeptide. The complex pattern of gene regulation involves molecular signals

that act on DNA sequences encoding protein product. *Cis*-acting molecules act upon and modulate the expression of physically adjacent polypeptide encoding sequences. On the other hand, *trans*-acting factors affect the expression of genes. The expression of a particular gene may be regulated by the combined action of both types of *cis* and *trans*-acting elements. While multiple DNA regions are involved in the transcription of a gene, the promoter is the key *cis*-acting regulatory region that controls the transcription of the adjacent coding region into mRNA.

Promoters in eukaryotic organisms, for example plants and animals, comprise multiple *cis*-acting motifs some of which are found in nearly all promoters (Butler and Kadonaga, 2002). These include: TATA box (TATA(T/A)A(T/A), an AT-rich sequence usually located 20 to 35 nucleotides upstream of the transcription start site (Butler and Kadonaga, 2002; Guilfoyle, 1997). This region tends to be surrounded by GC-rich sequences which may play a role in its function (Lefebvre and Gellatly, 1997). The TATA box binds RNA ploymerase II through a series of transcription factors to form an initiation complex. In some genes the TATA box region is also a determinant of cell specific or organ specific expression (Butler and Kadonaga, 2002). However studies based on the deletion of the TATA box resulted in reduced promoter activity (An and Kim, 1993). In such cases, the initiator *cis* element, which is a less characterized consensus sequence, initiates transcription via binding of transcription factors for the placement of the start site (Smale and Baltimore 1989).

Another control motif, the CAAT box, is located close to -80 nucleotides from the TSS (+1) of several, but not all plant promoters (An and Kim, 1993). It plays an important role in promoter efficiency, by increasing its strength, and it seems to function in either orientation (An and Kim, 1993).

A hexamer sequence, TGACGT, occurs in most constitutive promoters within a few hundred nucleotides from the transcription start site (An and Kim, 1993). These hexamer motifs are often found as repeats separated by six to eight nucleotides, and deletion analysis has indicated that these motifs are essential for the transcription activity of the cauliflower mosaic virus (CaMV) 35S, octapine synthase (*ocs*) and nopaline synthase (*nos*) promoters (An and Kim, 1993). Genes for the transcription factors that specifically interact with hexamer motifs have been isolated from

both dicotyledonous and monocotyledonous plant species (Katagiri et al., 1989; Singh et al., 1990).

Motifs found in promoters of genes that are environmentally inducible include the G-box sequence (CCACGTGG) in various photosynthesis gene promoters (An and Kim, 1993) and the H-box sequence (CCTACC) which, together with the G-box sequence, are essential for the expression of the bean chalcone synthase gene (Arias et al., 1993). A sequence rich in guanidine (G) and cytidine (C) nucleotides is usually found in multiple copies in the promoter region, normally surrounding the TATA box. In addition a large number of less common types of elements have been found in specialized types of signal-dependent transcriptional regulation, such as in response to heat shock, hormones, and growth factors (Mitchell and Tjian, 1989). The importance of *cis* element can vary greatly in different cell types and in response to physiological signals, depending on which DNA binding factors are present in different tissues and under different circumstances (Mitchell and Tjian, 1989).

Promoters used and required for biotechnology are of different types according to the intended type of control of gene expression. They can be divided into two major classes namely constitutive and tissue-specifically or developmentally regulated.

2. 5. 1. Constitutive promoters

These types of promoters direct expression in almost all tissues and are largely independent of environmental and developmental factors (Kuhlemeier, et al., 1987). Little variation of mRNA abundance or translational product is observed during development, in different organs or upon application of various endogenous or environmental stimuli (Kuhlemeier, et al., 1987). Examples of constitutive plant promoters are the nuclear gene promoter of the β-subunit of the mitochondrial ATPase complex (Boutry and Chua, 1985) and the rice actin-1 (Act-1) promoter (Zhang et al., 1991). In addition, some constitutive promoters are even active across species and even across kingdoms. These include promoters of nopaline synthase (*nos*), octopine synthase (*ocs*), mannopine synthase (*mas*) from *Agrobacterium* (Ebert et al., 1987; Ellis et al., 1987) and the 35S promoter from cauliflower mosaic virus (CaMV) (Franck et al., 1980).

The CaMV 35S promoter has often been used as a strong and constitutive promoter for the expression of many foreign genes in transgenic plants of many plant species, especially dicotyledons. Although the widely used CaMV 35S promoter is active in monocotyledons cells, its relative strength is less than in dicotyledons cells (Vasil et al., 1994). Hence, in a strategy to obtain higher expression levels in monocotyledons genetic constructs were made by the duplicating of the CaMV 35S promoter sequences (Vasil et al., 1994). The search for other constitutive promoters has continued specially in relation to the expression of transgenes in monocotyledons. In some monocotyledons such as cereals, it has been found that introns (non-coding DNA region) in the upstream region of the structural genes transcribed, but not translated into proteins are essential for efficient gene expression (Callis et al., 1987). Higher expression of transgenes has been observed when introns are included in the genetic constructs (Vasil et al., 1994). Examples of such introns include intron 1 of the rice actin gene (McElroy et al., 1991), intron 1 of the maize ubiquitin gene (Christensen and Quail, 1996; Salgueiro et al., 2000) and intron 1 of the maize alcohol dehydrogenous gene (Rathus et al., 1993). The positive effect of introns on gene expression has also been observed with promoters which on their own have weak activity in monocot, such as CaMV 35S promoter (Rathus et al., 1993).

Promoters that work well in dicotyledonous plants and which lack introns generally do not work well on monocotyledonous species, especially the Poaceae (grass species, which include most of the important food crops). For this reason, promoters from monocotyledons have also been developed for control of transgene expression. Examples of the most widely used monocotyledon promoters are the maize ubiquitin 1 promoter (Ubi1) (Christensen et al., 1992), rice actin 1 promoter (Act-1) (McElroy et al., 1990) and maize alcohol dehydrogenase 1 promoter (Adh-1) (Dennis et al., 1984).

The number and types of promoter that drive strong and constitutive expression of trangenes in sugarcane are relatively few (Liu et al., 2003). Of promoters tested to date, maize ubiquitin 1 promoter (Gallo-Meagher and Irvine, 1993), the rice actin 1 promoter (Rathus et al., 1993), the artificial *Emu* promoter (Rathus et al., 1993) and CaMV 35S promoter (Elliott et al., 1998) have shown activity in sugarcane. However, expression levels in transgenic sugarcane are lower than

the level of transgenes expression driven by CaMV 35S promoter in dicotyledonous plants (Grof and Campbell, 2001).

2. 5. 2. Tissue – specific promoters

The life cycle of a plant includes both gametophytic and sporophytic phases of development, that require temporal and spatial regulation (Tyagi, 2001). Furthermore due to the sessile nature, plants have to respond rapidly to environmental signals (Tyagi, 2001). All these will require a complex array of regulatory elements responding to signals like heat, cold, hypoxia, light, dehydration, hormones, wounding or the cues that originated from the pattern of differentiation and development (Tyagi, 2001). Therefore, in any one cell at a particular time, only a subset of the total genetic information is expressed.

Tissue-specific promoters control gene expression in a tissue-dependent manner and according to the developmental stage of the plant and show a well defined temporal expression pattern. The transgenes driven by these types of promoters will only be expressed in tissue where the transgene product is desired, leaving the rest of the tissue in the plant unmodified by the transgene expression (Stitt et al., 1995). The lack of temporal and spatial regulation of constitutive and non-specific promoters such as CaMV 35S, rice actin 1 and maize Ub 1 promoters may be suitable for proof of concept experiment but they have a number of potential drawbacks for use in genetically modified crops (Gittins et al., 2000). For example, localization of transgene expression may be essential for the expression of certain antisense constructs, biosynthetic genes or transcription factors in order to obtain transgenic regenerants or to avoid phenotypic abnormalities (Kumar et al., 1996; Martin, 1996).

Sucrose cycling and compartmentalization are important factors in the sucrose accumulation process in sugarcane stem. The mechanisms by which these processes are regulated are not yet determined (Moore, 1995). But targeting the different key enzymes which are expressed in different cell types through specific promoters may provide clearer understanding of sucrose pathway. In addition, there is a need to understand the role of different type of tissues involved in sink-to-source interaction in sugarcane. For instance, targeting of yeast invertase to the

apoplast in transgenic potato provided clear evidence that sucrose required for growth and development in sink tissue is transported directly to phloem (Heineke et al., 1992).

Many tissue-specific genes and the corresponding promoters have been isolated. Examples include genes that are specifically expressed in microspore (Custers et al., 1997), flower (Faktor et al., 1996) and root (de Pater and Schilperoot, 1992). Furthermore, tissue-specific promoters have been isolated from genes for storage organs such as seeds, fruits, potato tubers and for vegetative organs like leaves (Edwards and Coruzzi, 1990). The most extensively studied leaf-specific gene is the gene that encodes the small subunit of Rubisco (rbcs) (Fluhr, et al., 1986). Sugarcane Rubisco small subunit (rbcs) genes have been isolated, sequenced and the corresponding promoter elements are isolated (Tang et al., 1996). The promoter showed expression specifically in leaves and preferentially in bundle sheath (Tang et al., 1996). The tomato Rubisco small subunit gene promoter was also found to be active in the green photosynthetic tissue of apple, specifically the mesophyll layer and the palisade cells (Gittins et al., 2000). Furthermore, the grape alcohol dehydrogenase promoter is expressed during ripening of the grape berry and hence characterised as a fruit-ripening promoter (Sarni-Manchado et al., 1997).

The specific regions of promoters which are necessary for tissue specific expression are identified by deletion analysis and reporter gene expression studies. It has been found that in most cases the proximal region of the promoter is able to confer specificity of expression (Tyagi, 2001). Deletion analysis was used to identify the location within the proximal 295 bp of the French bean β-phaselin *phas* promoter of sequence motifs that have been implicated in directing seed specific expression in transgenic tobacco plants (Van der Geest and Hall, 1996). These include legumin box (CATGCATG) and endosperm boxes (TGTAAAGT and RTGAGTCAT) (Van der Geest and Hall, 1996).

Deletion analysis of the helianthinin promoter (2.4 kp) of sunflower showed regionalized GUS-expression in transgenic tobacco seeds. Deletion to -74 resulted in β-glucuronidase (GUS) activity in the cotyledons extending through the shoot apical region but not in the mid zone. Additional sequences to -116 did not alter the shoot/cotyledon expression pattern. However,

sequences between -116 and -321 extended the tissue range of GUS expression to include most of the mid zone of the seed. Staining of cotyledons, shoot and entire root apical regions were obtained when the -739 deletion was examined (Nunberg, 1995).

Functional dissection of a bean chalcone synthase gene promoter (CHS15) revealed that maximal floral and root-specific expression required sequence elements located on both sides of the TATA-box. Two adjacent sequence motifs located near the TATA-box, the G-Box and H-box, were essential for floral expression (Faktor et al., 1996).

In monocots seed-specific promoters and regulatory elements have also been identified. Glutelin is a major seed storage protein of rice and it constitutes a small multigene family of about 10 members per haploid genome (Yoshihara et al., 1996). These genes are clearly classified into two subfamilies designated as GluA and GluB (Yoshihara et al., 1996). Promoters of the glutelin genes from rice have been found to contain two prolamin box [TG(t/a/c) AAA, GCN4 motif (5'-TGAGTCA-3'), ACGT motif and AACA motif (5'-AACAAACTCAATC-3'). A 45 bp (-104 / -60) GluA-3 promoter fragment contains GCN4 motif acting as positive regulatory element in tissues other than endosperm (Yoshihara et al., 1996). Deletion of these motifs in GluB-1 promoter results in loss of seed specificity and multimer of 21 bp containing GCN4 can confer endosperm-specific expression in transgenic rice on -46 bp 35S promoter (Wu et al., 1998).

Tissue specific promoters may be induced by endogenous or exogenous factors, so that they can be classified as inducible as well. For instance, the genes encoding the small subunit of Rubisco are strongly regulated by light (Edwards and Coruzzi, 1990), or, the tuber-specific patatin promoter B33 has been found to be inducible by metabolic signals derived from sucrose (Rocha-Sosa et al., 1989). Based on the nature of factors/compounds that trigger their expression, they can be divided into two groups, namely chemically and physically regulated promoters (Tyagi, 2001). In many cases, when precise temporal and spatial gene regulation is required to control transgene expression, a constitutive promoter like CaMV 35S promoter is unsuitable to deal with gene expression (Zuo and Chua, 2000). In contrast to constitutive promoters, inducible gene expression systems are inactive in the absence of inducers and therefore the plant development will not be affected by the foreign gene (Zuo and Chua, 2000). Inducible promoters can be

important tools in plant genetic engineering because the expression of genes linked to them can be regulated to function at certain stages of development of an organism or a tissue (Zuo and Chua, 2000).

The controlled expression of any transgene is primarily dependent on the choice of a suitable promoter. This promoter will determine the timing, location and amount of the transgene products. Much emphasis is currently being placed on the identification of gene regulatory elements which provide tissue and developmental specificity in sugarcane. Although some researcher groups have obtained successful transgene expression in sugarcane using promoters of other species, many groups have found aberrant or nil expression of heterologous gene constructs following proven integration into the host genome (Birch, 1997). One explanation for silencing is that foreign or artificial promoters may lack some undefined features necessary for stable gene expression in sugarcane. For this reason it makes sense to isolate desired promoter types from sugarcane (Birch, 1997).

CHAPTER THREE

Differential expression, copy number and putative identities of selected sugarcane gene sequences

3. 1. Introduction

During their development and differentiation plants need to integrate different types of tissue, developmental and environmental signals to regulate complex patterns of gene expression (Singh, 1998). Numerous genes, and in some cases their associated promoters, which exhibit a wide range of tissue and/or developmental specificity have been characterized. Examples include genes that are specifically expressed in microspores (Custer et al., 1997), potato stolons (Trindade et al., 2003), embryos, cotyledons, endosperm, (Edward and Corruzzi, 1990), leaves, roots, and fruits (Coupe and Deikman, 1997; Edward and Corruzzi, 1990).

In sugarcane one of the most important tissues is the mature culm which is the primary site for sucrose accumulation (for review see Moore, 1995). Sucrose cycling and compartmentalization are important factors in the sucrose accumulation process in sugarcane culm (Moore, 1995; Whittaker and Botha, 1997; Vorster and Botha 1999; Komor, 2000). Sucrose accumulation will involve different enzymes present in the distinct cellular compartments of the sugarcane culm. Physiological studies of key enzymes associated with sucrose metabolism such as sucrose synthase, sucrose phosphate synthase and the various isoenzymes of invertase (neutral invertase, soluble acid invertase and cell wall bound acid invertase) showed that enzyme activity and expression vary depending on internode maturity (Moore, 1995; Zhu et al., 1997; Lingle, 1999; Botha and Black, 2000).

Furthermore Jacobsen et al. (1992) observed changes in morphological features and sucrose content of the sugarcane culm through histochemical and sugar analyses. These include an increase in the number of vascular bundles and a concomitant decrease in size from the core to the peripheral tissues and an increase in lignification and suberisation down the culm in parallel with increased sucrose concentrations.

In addition, based on the differences in morphological features and enzyme activities, it can be expected that genes are differentially expressed between the different tissue types in the culm and at the different developmental stages. Recently research approaches using expressed sequence tags (ESTs) have shown the differential expression of genes in sugarcane (Carson and Botha, 2000, 2002; Carson et al., 2002; Casu et al., 2003; Ma et al., 2004). Genes, which are differentially regulated are under the control of specific regulatory sequences (promoter regions) generally located immediately upstream of the gene (Birch, 1997). Hence, the identification and isolation of genes expressed under specific conditions will allow the isolation of their specific promoter elements.

In most cases, promoters are identified by studying the expression pattern of different genes in the organism of interest in order to identify genes with a particular expression pattern (Kuhn, 2001). It is rare that the cellular distribution of the gene of interest is taken into account before the isolation of promoter elements. To identify differentially expressed genes many techniques have been used. These include, Northern hybridisation analysis, subtractive hybridisation, cDNA fragment fingerprinting, serial analysis of gene expression and array technology (Kuhn, 2001). These techniques usually focus on the differences between organs and developmental stages, ignoring differentiation in expression between different cell types (for example, xylem, phloem and parenchyma cells) within the same organ. A technique which can be used to characterise differential gene expression in different cell types is *in situ* hybridisation. This technique detects the presence of the differentially expressed genes on cellular level. *In situ* hybridisation, although routinely applied to determine the spatial patterns of gene expression at the cytological level in both plant and animals systems (Schwarzacher and Heslop-Harrison, 2000), has only recently been employed to investigate gene expression at the cellular level in sugarcane (Casu et al., 2003).

Preliminary work using expression arrays, has indicated that the periphery and core of mature sugarcane internodes contain differentially expressed genes (Rogbeer, 2002). Thirty cDNA fragments were identified by expression arrays from the core and periphery of internode 7 and 10 which apparently were differentially expressed (Rogbeer, 2002). Based on limited RNA blot data, four cDNA clones, c22, c23, cO and cS, were selected for further analyses. The two cDNA

clones (c22 and c23) were selected after confirmation by RNA blot analyses in internodes 7 and 10. In contrast, cDNA clones cO and cS were selected because they showed specific expression in stem by macroarray analysis and their sequences showed homology to known genes in public databases. However, this does not conclusively show the expression of the selected cDNA fragments throughout the mature culm tissue (all internodes) and in the different cell types within the culm.

Therefore, the purpose of this study was two fold: Firstly, to determine whether the previously identified expressed genes are expressed in all the mature internodes throughout the culm and secondly, to determine if the gene/s are expressed only in particular cell types.

3. 2. Materials and Methods

3. 2. 1. Plant material

Field grown sugarcane variety N19 was collected from Welgevallen, Stellenbosch South Africa. Tissues were sampled to represent different developmental stages of sugarcane including leaves, a young internode (4), mature internode (6, 8, 9 and 12), an old internode (18) and roots. Internodes were numbered from top to bottom, with the point of attachment of the first visible, dewlap leaf, taken as internode one. For the isolation of RNA, the mid-rib of the young leaves was removed, internodes 4, 6, 8, 9, 12 and 18 were excised from the sugarcane culm and the rind removed and cut into smaller pieces. The tissue was immediately frozen in liquid nitrogen and stored at -80^{o}C as a whole tissue.

3. 2. 2. RNA extraction

Total RNA was extracted from 4 g of ground, frozen sugarcane tissue materials using a modified phenol-based method (Bugose et al., 1995). The extraction buffer contained 0.1 M Tris HCl pH 7.5, 1 mM EDTA, 1% (w/v) sodium dodecyl sulphate (SDS), 0.1 M NaCl, and 100mM β-mecapoethanol. The samples were extracted once with phenol/chloroform/isoamyl alcohol (25:24:1). After the addition of 0.7 ml of 3 M sodium acetate (pH 5.2), the nucleic acids were

precipitated with an equal volume of isopropanol at -20^0C overnight. The nucleic acids were collected by centrifugation at 10,000 *g* for 10 min. The pellet was washed with 70% (v/v) ethanol. The pellet was resuspended in diethyl pyrocarbonate (DEPC)-treated water and the insoluble material was removed by centrifugation at 10,000 *g* for 5 min. LiCl was added to the supernatant to a final concentration of 2 M to selectively precipitate the RNA. Samples were incubated overnight and RNA was pelleted through centrifugation at 10,000 *g* for 15 min. The isolated RNA was quantified by UV spectrophotometry (Maniatis et al., 1989) and the quality was assessed by electrophoresis in 1.2% (m/v) agarose gels in 1 X TBE/running buffer (0.045 M Tris-borate, 0.001 M EDTA) (Maniatis et al., 1989). The RNA preparations were stored at -80^0C.

3. 2. 3. Preparation of probes for RNA and southern blots

From the preliminary work, four cDNA clones (c22, c23, cO and cS), which show differential expression in expression array analyses, were selected for further characterisation of their expression patterns. MC106l (*E.coli)* competent cells were transformed with plasmids of these selected cDNAs. Plasmids were extracted using Mini-prep kit (Promega) following the manufacturer instructions. The number of inserts present in each clone was assessed by digestion with EcoR I (Promega). Since cDNA clones c22, c23, cO and cS contained multiple inserts individual fragments were purified from the gel using a gel purification kit (QIAGEN) following the manufacturer's instruction. The purified fragments were used for probe preparation. For preparation of probes, the larger fragments of the previously identified cDNA clones (c22, c23, cO and cS) were radioactively labelled by Random Primer Labelling kit (Stratagene) using 25 μCi [α-^{32}P] dCTP (Amersham).

3. 2. 4. RNA blot analyses

RNA blots were prepared using total RNA from leaves, internodes 4, 6, 8, 9, 12 and 18 and roots. Ten μg per lane of RNA was separated on 1.2% (m/v) agarose gels in 1 X TBE/running buffer (0.045 M Tris-borate, 0.001 M EDTA) (Maniatis et al., 1989). RNA was transferred to positively charged nylon membrane (Roche) by upward capillary blotting using 10 X SSC (3 M NaCl,

Sodium citrate). RNA was UV cross-linked to the nylon membrane by exposing it to UV light for 2.5 min at 120 mJ/cm. Pre-hybridisation was done at 42°C in 50% (v/v) de-ionised formamide, 5 X Denhardt's reagent, 5 X SSC, 0.1% (w/v) SDS and 20 mg/ml Herring Sperm DNA. For preparation of probes, the larger fragments of the previously identified cDNA clones (c22, c23, cO and cS) were radioactively labelled by Random Primer Labelling kit (Stratagene) using 25 µCi [α-^{32}P] dCTP (Amersham).

Hybridisation was carried out in the same buffer overnight at 42°C in the presence of the denatured probe. The hybridised blots were washed twice for 15 min each at 50°C and 55°C in wash solution 1 (1 X SSC, 0.1% (w/v) SDS) and twice for 15 min each at 60°C and 65°C in wash solution 2 (0.5 X SSC, 0.1% (w/v) SDS). The blots were exposed to a supersensitive Cyclone Phosphor screen (Packard) for 16 to 18 hours. The hybridisation was visualised using a Cyclone imaging system (Packard).

3. 2. 5. Southern blot analyses

Genomic DNA was isolated from the leaves using the Dellaporta et al (1983) method. Eleven µg of genomic DNA was digested with EcoR I, Hind III, Xba I, BamH I, Pst I, and Sac I. Digestion reactions were incubated overnight at 37^{0}C. Digested DNA was precipitated with 1/20 volume of 3 M NaAc and 2 volumes of 95% (v/v) EtOH overnight 4^{0}C. DNA was collected and resuspended in distilled water. The restricted DNA was electrophoresed and visualised in 0.8% (m/v) ethidium bromide stained agarose gels, denatured with 1.5 M NaCl and 0.5 M NaOH for 1 hr, and transferred to neutralisation solution containing 0.5 M Tris-HCl pH 7.4, 1.5 M NaCl and for 1hr. DNA was then blotted overnight onto positively charged nylon membranes (Boehringer Mannheim, GmbH Mannheim, Germany) by upward capillary blotting (Sambrook et al., 1989) using 10 X SSC. The membrane was washed at room temperature for 5 min in 5 X SSC and the DNA was UV cross-linked to the nylon membrane by exposing it to UV light for 2.5 min at 120 mJ/cm.

Conditions for pre-hybridisation, hybridisation and visualising exposed screens were the same as described for northern analyses.

3. 2. 6. *In situ* hybridisation tissue preparation

For the purpose of this study, internodes 4, 8, 12 and 18 were sampled as representatives of young, maturing, mature and old tissues of the sugarcane stalks comprising of 18-20 internodes. Transverse sections of the sugarcane internodes spanning half the core, mid-internodal and peripheral regions were excised and sliced at a thickness of approximately 0.2 mm and a radius of 5.5 cm. These tissue slices were placed immediately into fixative containing 4% (w/v) paraformaldehyde in phosphate buffered saline (PBS) made with 130 mM NaCl, 7 mM Na_2HPO_4, 3 mM $NaH_2PO_4.2H_2O$, pH 7, with 0.1% (v/v) Triton X 100 and 0.1% (v/v) Tween 20) overnight. In order for the tissue to be more accessible to the RNA probes, and to reduce any non-specific binding of the probe to the slide, the tissue was washed twice for 5 min each under running Milli-Q water then in 0.2 M HCl for 10 min followed by Milli-Q water for 5 min and twice for 2 min each in PBS. The tissue de-proteinisation treatment involved exposure of the tissue for 10 min in 0.125 mg/ml pronase (Roche, Germany) in pronase buffer containing 50 mM Tris-HCl pH 7.5, 5 mM EDTA. This was followed by a 2 min wash in glycine (0.2% (v/v) in PBS). The tissue was again washed for 2 min in PBS. Restabilization of the tissue is crucial before hybridisation and thus the tissue was post-fixed for 10 min in the fixative. This was followed by two 2 min washes in PBS. In order to inactivate the proteases and reduce the electrostatic binding of the probe to sections, the tissue sections were washed for 10 min in 1% (v/v) acetic anhydride in 0.1 M triethanolamine (pH 8). The tissue was then washed for 2 min in PBS, before exposure to a dehydration series of 50%, 85% and 100% (v/v) ethanol for a duration of 1 min in each.

3. 2. 7. Probe preparation for *in Situ* hybridisation

In situ hybridisation requires the use of short probes (Drews, 1998). Digestion with BamH I of plasmid vectors (pBluescript (SK^+)) containing cDNA insert encoding dirigent-like protein was resulted in a 316 bp fragment and it was religated. Restriction endonucleases and T4 ligase were purchased from Promega and used as recommended by the manufacturer. The DNA template was linearised using Hind III and Xba I. The resulting fragments were purified using five volumes of 5 mM $MgCl_2$ in 70% (v/v) EtOH according to the protocol outlined by Sambrook et

al. (1989). The linearized and purified DNA was used in *in vitro* transcription reaction using T7 and T3 RNA polymerases from Roche molecular biochemicals. One microgram of linearized template DNA was incubated for 2 h at 37°C in a 20 µl mixture containing 40 U of the appropriate RNA polymerase (T3 or T7), NTP labeling mixture containing 1 mM of each ATP, CTP and GTP, 0.65 mM UTP, 0.35 mM digoxigenin-labeled UTP, transcription buffer containing 400 mM Tris-HCL (pH 8), 60 mM $MgCl_2$, 100 mM dithiothreitol and 20 mM spermidine, and 20 U RNAse inhibitor. The reaction was stopped by the addition of 1 µl 0.5 M EDTA (pH 8) and the labeled RNA was precipitated with 2.5 µl 4 M LiCl and 75 µl EtOH at –20°C for 2 hr. After pelleting and washing with cold 70% (v/v) EtOH, the transcripts were dissolved in 100 µl Milli-Q water. The riboprobes were then aliquoted in 10 µl volumes and stored at –80°C. Before storage, 5 µl of the transcript was run on a gel and thereafter a dilution series was set up in order to detect the riboprobe concentration using DIG quantification strips (Roche molecular biochemicals). The orientation of the gene fragment was determined by sequence analysis.

Generally, about 200 ng/ml of the probe is required for a hybridisation reaction. Before the probe was hybridized to the target RNA, it was diluted in the hybridisation buffer to the working concentration. For hybridisation the hybridisation buffer for *in situ* hybridisation of Sigma was used. Once the hybridisation mixture had been applied to the tissue (50 µl per tissue slice), the specimen was covered with RNAse-free hybrid-slips (Sigma). Hybridisation was carried out at 42°C and was done for 16-24 hours. Incubation was carried out in a sealed plastic container with a small amount of 50% (v/v) de-ionised formamide, 2 X SSC (1 X SSC contains 150 mM NaCl and 15 mM sodium citrate).

3. 2. 8. Washing

Hybridised tissue sections were placed into sterile 10 ml tubes and subjected to a series of stringency washes to remove the hybridisation mixture and unbound or weakly bound probes. The tissue was washed for 1 hr in fresh wash buffer (50% (v/v) de-ionised formamide, 2 X SSC) at 50°C. Subsequently the tissue was washed twice for 5 min each in NTE buffer containing 500 mM NaCl, 10 mM Tris-HCl (pH 8) and 1 mM EDTA at 37°C. The samples were then incubated

for 30 min in the same buffer with the addition of 25 μg/ml RNAse at 37^{o}C and then washed twice for 5 min each in NTE buffer at ambient temperature followed by wash buffer at 50^{o}C for 1 hr and PBS for 5 min at ambient temperature.

3. 2. 9. Detection

For the purpose of detecting RNA-RNA hybridisation sites in cells, a chromogenic, enzyme-mediated reporter system using the NBT/BCIP (Nitro blue tetrazolium chloride/ 5-bromo-4-chloro-3-indolyl-phosphate) and the anti-DIG antibody, coupled to alkaline phosphatase was used. The tissue was washed for 5 min in a detection buffer comprising of 100 mM Tris-HCl (pH 7.5) and 150 mM (w/v) NaCl. Thereafter the tissue was placed in blocking solution (1% (w/v) Roche molecular biochemicals blocking reagent in the detection buffer) overnight at 4^{o}C. The following day, the tissue was placed in fresh blocking solution containing the anti-DIG antibody coupled to alkaline phosphatase (Roche molecular biochemicals) at a concentration of 1:5000 for 90 min. The tissue was then washed twice for 40 min each in detection buffer with 0.3% (v/v) Triton X 100 and then with the detection buffer for 5 min and with 100 mM Tris-HCl (pH 9.5), 100 mM NaCl and 50 mM $MgCl_2$ for a further 5 min. The tissue was stained with a solution containing one NBT/BCIP tablet (Roche molecular biochemicals) per 10 ml of 10% (w/v) polyvinyl alcohol solution (Sigma) until significant colouration was observed. The detection reaction was stopped by rinsing the slides in running tap water for 10 min and subsequently rinsing in 0.1 M EDTA.

3. 2. 10. Microscopy

The sections were studied with a Nikon Eclipse E400 Microscope. Photographs were taken with a digital camera (Nikon coolpix 990).

3. 2. 11. Sequence analysis

The culm specific fragments of cDNA clones c22 and c23-a were subcloned into a pBluescript vector. Plasmids were prepared using the Mini-prep kit (Promega). Plasmid inserts were sequenced using T7 and T3 primers, and ABI Prism^{TM} Big Dye terminator Cycle Sequencing

Ready Reaction Kit. The cDNA sequences obtained were then compared to sequences in the GeneBank database provided by the National Center for Biological Information (NCBI).

3. 3. Results

3. 3. 1. cDNA Fragments

The first phase of this project was to characterise the cDNA inserts in the selected clones obtained from pervious study because many of the cDNA clones were found to contain multiple insert. The results obtained from EcoR I digestion, indicated that cDNA clones of c22, c23, cO and cS contained multiple inserts (Table 3.1, Figure 3.1).

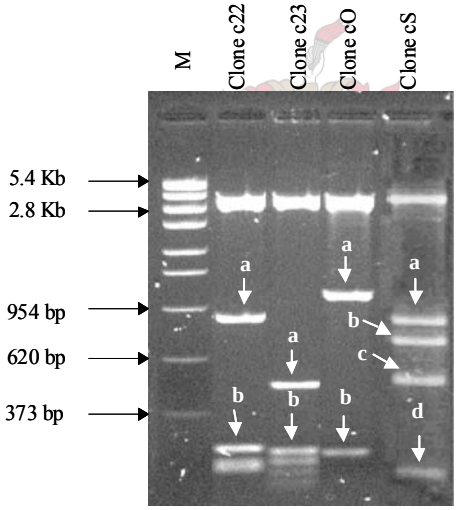


Figure 3.1: Agarose gel electrophoresis of EcoR I digested cDNA clones (c22, c23, cO and cS). Lane M indicates the molecular weight marker.

Table 3.1: Number of inserts present in the selected cDNA clones.

Clone	Number of cDNA inserts	cDNA fragments	Approximate Size (bp)	Characterised fragments by RNA Blot[a]
Clone c22	>2	c22-a	800	✓
		c22-b	200	x
Clone c23	>2	c23-a	450	✓
		c23-b	200	x
Clone cO	2	cO-a	950	✓
		cO-b	200	x
Clone cS	4	cS-a	800	✓
		cS-b	750	x
		cS-c	500	x
		cS-d	200	x

[a] ✓ = yes, x = no

3. 3. 2. RNA blot Analyses

Since the four cDNA clones (c22, c23, cO and cS) contained multiple inserts (Figure3.1, Table 3.1), a further confirmation of expression pattern was needed with individual fragments. In this study fragments c22-a, c23-a, cO-a and cS-a were selected based on the size of the fragments for RNA blot analyses. For RNA blot analyses, RNA was extracted from leaves, young internode (4), mature internodes (8, 9, 12), old internode (18) and roots.

Figure 3.2 A, C show expression patterns of cDNA clones c22-a and c23-a respectively. Results indicate that the cDNA inserts strongly hybridize to mRNA isolated from mature and old internodes (internodes 8, 9, 12 and 18 respectively). The expression level increases from internode 8 to old internode 18. The maximum level of expression was found in mature internode 12 and old internode 18 and intermediate level was detected in internodes 8 and 9. No expression was observed in leaves, young internode 4 and roots.

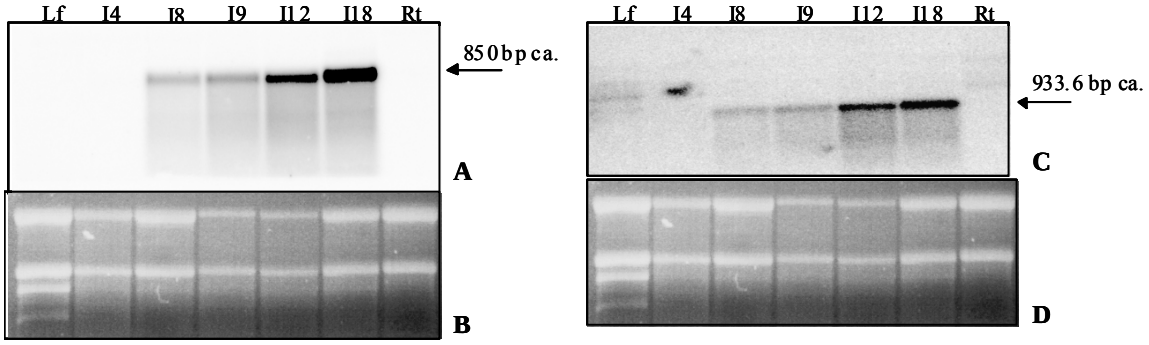


Figure 3.2: RNA blot analyses of selected fragments. Panel A and C show expression profiles of cDNA fragments c22-a, and c23-a respectively. Panels B and D show pictures of RNA gel used in the experiment. The blots contained 10 µg/lane of RNA from leaf (Lf), a young internode (I4), a mature internode (I8, I9, and I12), an old internode (I18) and roots (Rt).

Fragment cO-a showed very strong hybridization with mRNA isolated from leaves and an intermediate level of hybridization was also observed in internodal tissues and roots (Figure 3.3 A). Fragment cS-a hybridized to mRNA isolated from leaves and roots of sugarcane. This fragment also showed a weak of signal in internodal tissues upon longer exposure (Figure 3.3 B).

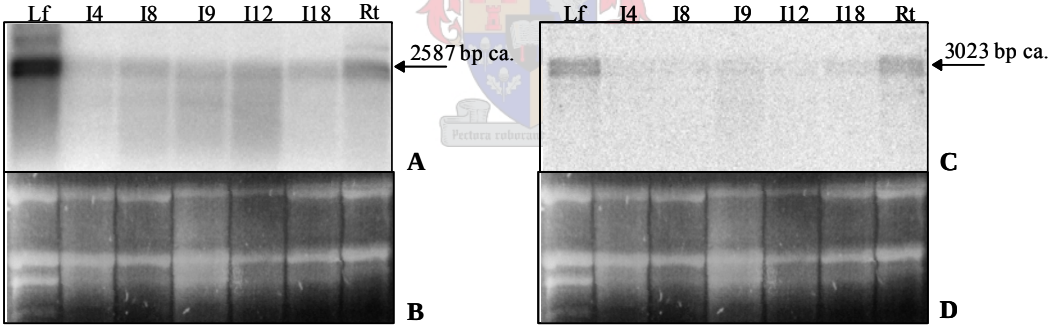


Figure 3.3: RNA blot analyses of selected fragments. Panel A and C show the expression profile of cDNA fragments cO-a and cS-a, respectively. Panel B and D contains pictures of ethidium bromide-stained RNA gel used in the experiment. The blots contain 10 µg/lane of RNA from leaf (Lf), a young internode (I4), a mature internode (I8, I9, I12), an old internode (I18) and roots (Rt).

3. 3. 3. Southern blot Analyses

The cDNAs that showed culm specific expression were used to probe sugarcane genomic DNA. Results obtained were used to get an indication of copy number of the respective genes in the sugarcane genome. The results are shown in figure 4 A, B. When cDNA clone 23-a was used to probe sugarcane genomic DNA, one to two bands were observed for each digest (Figure 3.4 A). cDNA clone c22-a also hybridised to one to two fragments in each digest (Figure 3.4 B).

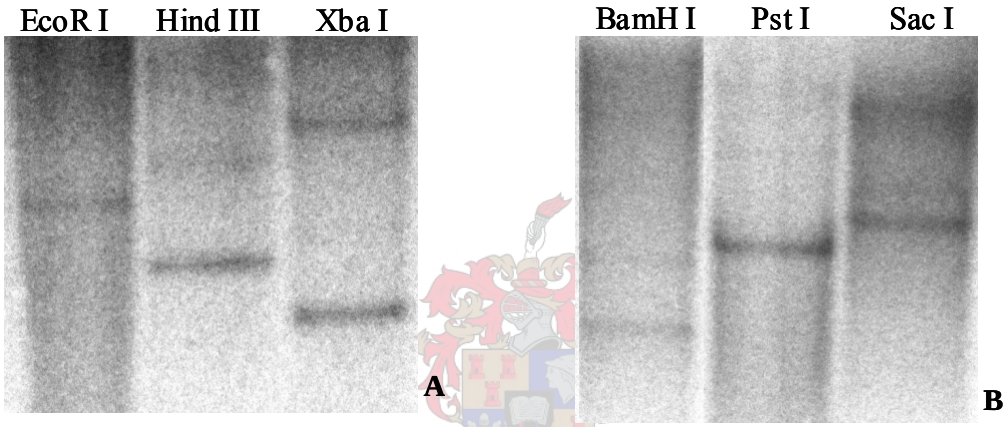


Figure 3.4: Southern blot analyses of sugarcane genomic DNA isolated from leaves. Eleven µg of genomic DNA was restricted with EcoR I, Hind III and Xba I (panel A) and with BamH I, Pst I and Sac I (panel B) and fractionated on 0.8% agarose gels. Blots were probed with c23-a (A) and c22-a (B).

Table 3.2: Summary of results obtained from the RNA blot analyses and Southern blot analyses

CDNA Clones	Approximate Size (bp)	Tissue specificity [a]	Selected for Promoter recovery YES/NO
c22-a	800	C	YES
c23-a	450	C	YES
cO-a	950	L	NO
CS-a	800	L+R	NO

[a] L= Leaves, C= Culm, R= Roots

3. 3. 4. Identification of differentially expressed cDNA clones

Nucleotide sequences of cDNA 22-a and 23-a, which showed culm specific expression, were determined. Sequence comparison was carried out by searching in the GeneBank using the BLAST program at the National Center for Biological Information (NCBI). The search revealed that both c22-a and c23-a show homology to known gene sequences (Table 3.3).

Table 3.3: Summary of results obtained from sequence comparison in BLAST search, of the culm -specific fragments.

	c22-a			c23-a		
	Sequence	Homology (%)	Length	Sequence	Homology (%)	Length
BLASTN NR	1. Saccharum hybrid cultivar dirigent mRNA AY421731	96	566bp	1. Zeamays CL17814_1mRNA sequence	90	377bp
	2. Zeamays CL10323_1 mRNA sequence AY111001	90	41bp			
BLASTN EST	1. Saccharum sp. cDNA clone MCSA100F06 5' similar to dirigent CF573900	95	698bp	1. *Saccharum officinarum* cDNA SCRUFL1112E09	98	371bp
	2. *Saccharum officinarum* cDNA clone SCBGHR1060E05 5', mRNA sequence. CA102512	96	557bp	2. Sorghum propinquum cDNA RHIZ2	95	369bp
				3. Zeamays cDNA tassel primordium	89	377bp
BLASTX	1. Saccharum hybrid cultivar dirigent AAR00251	91	161aa	1. *T.aestivum* serine/threonine protein kinase	82	28aa
	2. Dirigent protein (*Thuja plicata*) AAF25363	36	88aa	2. *Arabidopsis thaliana* putative protein kinase	67	28aa
	3. Dirigent protein *Forsythia intermedia*. AAF25357	33	92aa			
	4. Jacalin homolog *Hordeum vulgare*].AAB72097	36	146aa			

Homology (%): The percentage identity with the specific sequence in the search.

Sequence: The sequences in the database to which the highest homology were recorded.

Length: The number of base pairs (bp) or amino acids (aa) which was compared in the search.

BLASTN: Sequence comparison nucleotide to nucleotide level.

BLASTX: Sequence comparison nucleotide to amino acid level.

NR: Non-redundant database.

EST: Expressed Sequence Tags.

3. 3. 5. *In situ* hybridisation

In situ hybridisation was used to localise the expression of selected gene in sugarcane culm. Clone 22-a (a putative dirigent-like protein) was selected for cellular localization based on its expression pattern in RNA blot and the length of the cloned cDNA fragment. This gene fragment used in RNA blot, was used to probe the tissue sections. The *in situ* hybridisation was performed with DIG-labeled sense and anti-sense transcript of the gene fragment. The controls (sense probes) showed no hybridisation signal whereas the anti-sense probes hybridised to the transcribed sequences in the tissue, resulting in a blue colouring that could be clearly observed under the microscope (Figure 3.5). Figure 2.1 in chapter 2 shows the various tissue and cell types that comprise the sugarcane culm. The *in situ* hybridisation showed that the dirigent-like protein mRNA in the culm is preferentially localised in the parenchyma cells surrounding the vascular tissue and in the parenchyma cells of the vascular complex of internodes 8, 12 and 14 (Figure 3.5). The blue staining was restricted to certain parts of the cell, that is, the cytoplasm. *In situ* hybridisation revealed no expression of the gene in the young internode (internode 4) (Figure 3.5 A), once again confirming the results obtained with the RNA blot experiment.

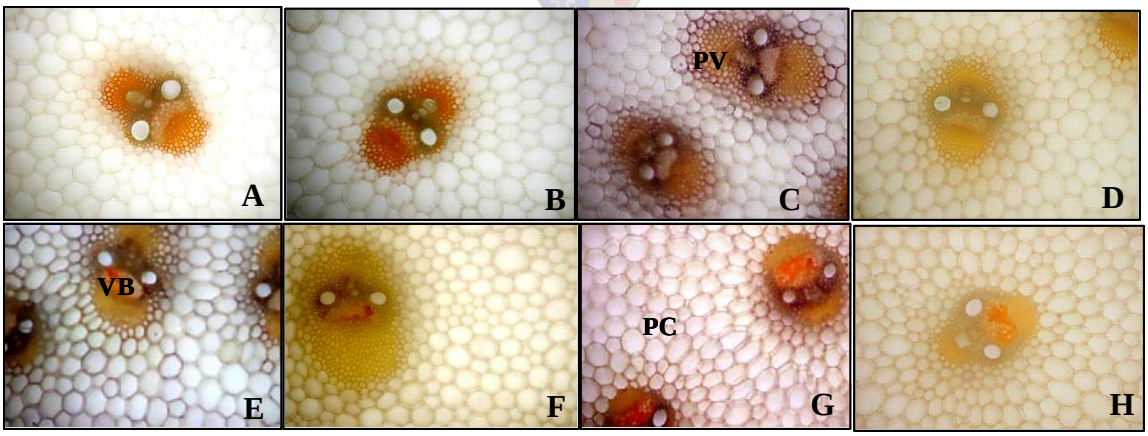


Figure 3.5: Detection of a non-radioactive DIG labelled dirigent like protein transcript in sugarcane internode 4 (A: antisense probe; B: sense probe), internode 8 (C: antisense probe; D: sense probe), internode 12 (E: antisense probe; F: sense probe) and internode 18 (G: antisense probe; H: sense probe) by employing the technique of *in situ* hybridisation. PC, parenchyma contiguous to the vascular bundle; PV, parenchyma associated with the vascular bundle; VB, vascular bundle.

3. 4. Discussion

In the previous work, thirty cDNA putative differentially expressed fragments were identified by expression arrays from the core and periphery of internode 7 and 10. Based on limited RNA blot data 4 cDNA clones were selected (c22, c23, cO and cS) for further analyses. The first objective of this study was to confirm that the previously identified culm-specific sequences are truly differentially expressed throughout the mature culm. For this purpose RNA was extracted from leaves, a young internode, a mature internode, an old internode and roots and used in RNA blot analyses. Secondly, to determine if the gene/s are expressed only in particular cell types of the culm. The result obtained from RNA blot shows that two cDNA clones (c22-a and c23-a) are culm-specific in sugarcane N19 variety. The results also indicate that within the stalk the transcripts of c22-a and c23-a are highly accumulated in mature and older internodes.

For the other two clones, cO-a and cS-a, the RNA analyses showed that these clones are not culm specific. The transcripts of cDNA clone, cO-a, were predominantly detected in leaves, at an intermediate level in roots and less in internodes. cDNA cS-a has been confirmed to be leaf and root specific. However, it remains to be determined whether the other fragments of cO and cS represent culm specific expression.

The culm-specific cDNA c22-a has been sequenced and putatively identified as a dirigent-like protein. This family contains a number of proteins which are induced during disease response in plants. Members of this family are involved in lignification. Lignin is important for mechanical support; it gives rigidity to cell wall, facilitates water transport, and hinders the degradation of wall polysaccharides, thus acting as a major line of defence against pathogens, insects, and other herbivores (Burlat, et al., 2001; Önnerud et al., 2002).

The increase of tissue maturation down the length of the sugarcane culm involves an increase in lignification and suberisation of cells in the culm (Jacobsen et al., 1992). In young internodes the protoxylem are the only cells which gets lignified. With maturity, lignification spreads to metaxylem, then to sclerenchymatic bundle sheaths and a few parenchyma cells. In mature and old internodes, liginifcation carries on from the vascular bundle to storage parenchyma cells

(Jacobsen et al., 1992). The presence of dirigent-like protein in the parenchyma tissue surrounding the vascular bundles and in the parenchyma cells of the vascular complex of the mature sugarcane culm (Figure 3.5) was therefore a probable consequence of the requirement for lignification of cells in order to maintain cell structure and function. A similar tissue specific expression was observed for *Forsythia intermedia* (Burlat et al., 2001) and *Thuja plicata* (Kim et al., 2002a) where dirigent proteins gene was localized in different types of lignified cell walls of stems. In addition, the results suggested that the gene is under tissue-specific control since its expression differed between young and mature culm tissue of sugarcane. Dirigent protein genes are expressed in cells during lignification (Önnerud et al., 2002).

cDNA c23-a has been putatively identified as a signalling molecule, exhibiting sequence similarity with a serine/threonine protein kinase. Protein kinases can be involved in hormone, defence and environmental stress responses (Hardie, 2000). In addition from the EST, database this cDNA showed similarity in sequences with cDNA clones of sorghum and maize which confirms that it is an expressed sequence. It is not yet known what the exact function of the gene is or why it is expressed only in stalk tissues.

The Southern blot data of clones c22-a and c23-a, suggests that the corresponding genes are present at a low copy number in the sugarcane genome. If a gene is present in multiple copies, there is a possibility that the duplicate copies will have accumulated sequence changes to incapacitate transcription. Because of these phenomena, most cloned copies of these genes may have non-functional promoters. From Southern blot, RNA blot and *in situ* hybridisation results, dirigent-like protein gene will be ideal candidate for functional promoter recovery. A promoter is a DNA sequence generally located immediately upstream of a gene, and determines the developmental timing, location and the levels of gene expression. After tissue specificity of potential clones had been confirmed, the dirigent-like protein gene (c22-a) was selected for promoter isolation based on its expression pattern and the length of the cloned cDNA fragment. This promoter could be valuable since it could act as powerful genetic "switch" that is tissue and developmental stage specific. It could possibly be used to drive expression of other important genes, particularly those that need to be expressed in the mature internodes of sugarcane to manipulate sucrose accumulation.

CHAPTER FOUR

Isolation of the promoter of a low copy number gene (Dirigent-like protein), differentially expressed at high levels in the mature culm

4. 1. Introduction

Conventional plant breeding techniques have been widely used to enhance important economic traits in agronomic crops, but this approach is laborious and time-consuming in crops like sugarcane (Liu et al., 2003). In addition, several important traits such as resistance to viruses, insects and herbicides are sometimes absent from the normal sugarcane germplasm (Liu et al., 2003). Genetic engineering can serve as an alternative way to introduce useful genes into sugarcane (Grof and Campbell, 2001).

Appropriate genetic constructs, containing a promoter, the transgene and a terminating signal, are required to facilitate the integration and expression of foreign DNA in plants (Birch, 1997). The effective manipulation of a phenotypic trait depends on the effective and predictable expression of the transgene (Stitt and Sonnewald, 1995). Successful genetic manipulation is therefore, to a large extent, dependent on the availability of appropriate promoters to achieve the desired expressions.

There are currently very few constitutive promoters that have been shown to be active in sugarcane (Grof and Campbell, 2001). These include maize ubiquitin 1 promoter (Gallo-Meagher and Irvine, 1993), the rice actin 1 promoter (Rathus et al., 1993), the artificial Emu promoter (Rathus et al., 1993) and the CaMV 35S promoter (Elliott et al., 1998). However, expression levels in transgenic sugarcane are low relative to reported levels from strong constitutive promoters, like CaMV 35S, in transgenic dicotyledonous plants (Grof and Campbell, 2001). In addition, from those studies it became apparent that gene silencing is a real problem. One explanation for silencing is that foreign or artificial promoters may lack some undefined features necessary for stable gene expression in sugarcane (Birch, 1997). Although constitutive expression of a transgene may sometimes be required, targeting expression to a specific tissue

type or developmental stage would decrease the metabolic load on the plant, in comparison to a constitutively expressed transgene.

Tissue-specific promoters control gene expression in a tissue-dependent manner and according to the developmental stage of the plant. The transgenes driven by these types of promoters will only be expressed in tissue where the transgene product is desired, leaving the rest of the tissue in the plant unmodified by the transgene expression (Stitt and Sonnewald, 1995). Unlike constitutive expression of genes, tissue-specific expression is the result of several interacting levels of gene regulation. It is therefore preferable to use promoters isolated from sugarcane itself to achieve efficient and reliable transgene expression in a particular tissue type (Birch, 1997).

At present, no regulatory sequences are available that drive transgene expression in sugarcane in a tissue/organ-specific manner. Patent limitation on already available promoters and genes, active in different plants, necessitate the isolation of novel regulatory elements from sugarcane. The aim of this project was therefore to isolate a mature culm specific promoter from sugarcane that could be used to drive stable trangene expression in sugarcane.

One possible way to obtain promoters directing the desired pattern of expression is to look for genes already expressed in the desired pattern in sugarcane, and then to isolate the corresponding promoters from sugarcane genome. The c22-a cDNA was selected as a target for isolation of the corresponding regulatory sequence, based on the expression pattern it showed in RNA blot and *in situ* hybridisation analyses (chapter 3). This apparently encodes a dirigent-like protein. This family contains a number of proteins which are induced during disease response in plants and members of this gene family are involved in lignification (Davin and Lewis, 2000).

Several methods have been used to isolate promoter elements from different plant genes. These include genomic library screening by cDNA or heterologous probes, promoter tagging (An and Kim, 1993) and inverse polymerase chain reaction (IPCR) (Triglia et al., 2000). IPCR is a convenient and versatile method of cloning unknown sequences upstream or downstream of known sequences (Triglia et al., 2000). IPCR avoids the laborious procedures of producing and

screening genomic libraries. This method has, for example, been successfully used to isolate the promoter sequences of two invertase genes from potato (Hedely et al., 2000) and seed lipoxgenase promoter from pea (Forster et al., 1994). Since the sequence of the targeted dirigent-like protein is known (refer chapter 3), it was decided to use IPCR to isolate the promoter sequence.

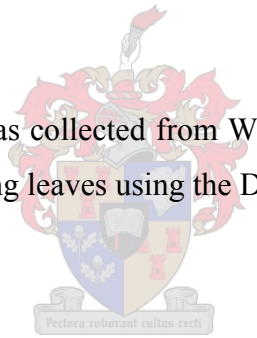
In this chapter the successful isolation of the promoter is described. In addition a detailed sequence analysis and comparison with other homologues is presented.

4. 2. Materials and Methods

4. 2. 1. Isolation, cloning and sequencing of the dirigent-like protein promoter

Isolation of Genomic DNA

Field grown sugarcane variety N19 was collected from Welgevallen, Stellenbosch South Africa. Genomic DNA was isolated from young leaves using the Dellaporta et al., (1983) method.



Inverse-PCR

To clone the genomic sequences immediately upstream of the coding region contained in the putative dirigent-like cDNA clone, a long-range inverse PCR amplification (LR-iPCR) was conducted. Based on the dirigent-like gene sequence, a pair of primers was designed at the 5' and 3' ends of the sequence, extending in an outward direction for the LR-iPCR. These are Dirfrw2-forward (5'-CTCATCCAGTGCGAACGATGCAG-3' towards the 3'-UTR) and Dirrev2-reverse (5'-CTACCTGGGGAGTTCGGCAAC-3' towards the 5'-UTR).

Three µg of genomic DNA isolated from the leaves of sugarcane variety N19 was digested with 6 bp-cutter restriction enzymes in a total volume of 30 µl at 37^{o}C overnight. Digested DNA (600 ng) was circularized by self-ligation in a ligation mix containing 9 U T4 DNA ligase, 40 µl ligase buffer (Promega Corporation, Madison, USA) and made up in a final volume of 400 µl at 16 ^{o}C overnight. Circularised DNA fragments were phenol/chloroform (1:1) extracted, precipitated and

resuspended in 40 µl distilled water. Digestion and recircularisation of genomic DNA were visualised in ethidium bromide-stained 0.8% (m/v) agarose gels. The re-ligated DNA was quantified and 150 ng circularised DNA was taken as template for LR-iPCR reactions.

The PCR was carried out in a 50 µl mixture containing 1 µl Elongase®enzyme mix, 10 µl total volume of Buffer A and B (Invitrogen Corporation), 0.2 mM of each dNTP and 0.2 mM of each primer (Dirfrw2-forward and Dirrev2-reverse). The cycling conditions were as follows: an initial denaturing for 30 sec at 94 ^{o}C was followed by 35 cycles of amplification (94 °C denaturation, 1 min; 55 °C annealing, 1 min; 68 °C polymerization, 10 min). The PCR products were run on ethidium bromide-stained 1.2% (m/v) agarose gel to check for fragments. Following electrophoresis of the PCR product and elution from agarose gel, fragments were cloned into the pGEM T-easy vector (Promega Corporation, Madison, USA) and sequenced (ABI PRISM® 3100 genetic analyser) using the ready reaction kit with AmpliTaq® DNA polymerase (The Perkin Elmer Corporation, Norwalk, USA).

4. 2. 2. Sequence analyses

To identify preliminary proximal promoter regions (TATA-box and putative transcriptional start site), the genomic clone sequence was analysed by promoter prediction database software, Neural Network Promoter Prediction (NNPP, http://www.fruitfly.org/seq tools/promoter.html) (Rees and Eeckman, 1995). Putative cis-acting elements were found using the PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE) (Higo et al., 1998) and PlantCARE (http://intra.psb.ugent.be:8080/PlantCARE) (Rombauts et al., 1999) databases, which describe elements from vascular plants. All sequence similarity search and alignment analyses were conducted using BLAST: http://www.ncbi.nlm.nih.gov/blast/ (Altschul et al., 1990) and DNASIS®MAX software (Hitachi Software Engineering Co., Ltd., Japan).

4. 2. 3. Plasmid construction

To construct the plasmid pBSGUS DPB and pBSGUS DPS, the vector pAHC27 (Christensen and Quail, 1996) was digested with BamH I and EcoR I to isolate the *uid*A reporter gene with the nopaline synthase (nos) terminator sequence. This 2114 bp BamH I-EcoR I fragment from

pAHC27 was then cloned into pBlueScript SK^+ to make a promoter less construct and named as pBScr-GUS000. To build gene constructs including the putative dirigent promoter fused to the GUS reporter gene, BamH I and Xba I sites were added to the 5' and 3' ends of the promoter sequences. This was done by PCR amplification of the cloned promoter fragments with primers containing the respective restriction sites (DPBfrw-forward 5'-AGGC**TCTAGA**GGTGTGGAAGACTA-3', DPSfrw-forward 5'-AGG**TCTAGA**GAGCTCTCAATGGGA-3' and DirPrev-reverse 5'-ACC**GGATCC**TGTATTAGTTATGGCAGC-3'**).** This resulted in a 1151 bp DNA fragment containing the Bgl I-derived promoter and a 985 bp fragment containing the Sac I-derived promoter. These two promoter fragments were fused transcriptionaly into the Xba I and BamH I sites of the expression vector pBScr-GUS000, to obtain the construct pBSGUS DPB and pBSGUS DPS. These constructs were subjected to restriction analyses using the enzymes BamH I, Xba I and EcoR I where after they were sequenced.

4. 2. 4. Preparation of target tissue

Five types of target tissues were used for functional analysis of putative DPB and DBS promoters. Sugarcane callus, type III and IV, slices of culm tissue, young internode 5 and matured internode 15, and maize hypocotyls. The maize hypocotyl was used because of its taxonomic relatedness to sugarcane and the convenience of this material for GUS expression. These tissues were placed in the centre of petri dishes containing wet filter paper.

4. 2. 5. Particle bombardments

Plamids (GUS constructs) were isolated by rapid alkaline extraction, dissolved in MQ water and checked for quality by gel electrophoresis, quantified by flourometry. Bombardment was carried out using a particle bombardment device, constructed according to the Finer Particle Inflow Gun design (Finer et al., 1992) and tungsten particles (0.7 µm, Grade M10, Bio-Rad) coated with the plasmid DNA. For bombardment, 10 µl of plasmid DNA (10 µg) was co-precipitated onto 50 µl of sterile tungsten particles (5 mg) in the presence of 50 µl of 2.5 M $CaCl_2$ and 20 µl of 100 mM spermidine, free base. The mixture was incubated on ice for 5 min where upon 100 µl of the supernatant was removed and discarded. For bombardment, 5µl of the dispersed tungsten-DNA

was pipetted onto the centre of the support screen of the disassembled 13 mm stainless steel syringe filter unit. Petri dishes containing the target tissue was placed 12.5 cm from the support screen in the syringe filter unit. The tissue was covered with a baffle made of 2 mm stainless steal screen, mounted 9 cm above the tissue (Finer et al., 1992). Particle bombardment was performed with a chamber vacuum of 85 kPa and a helium delivery pressure of 1000 kPa. Target tissues were bombarded once. All bombarded tissues were incubated at 37^0C in the dark for 48-72 hrs.

4. 2. 6. GUS Assay/ Transient expression

Histochemical detection of GUS activity was done using the substrate 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) according to the method of Jefferson et al. (1987). Sugarcane tissues (callus and culms slices) and maize hypocotyls were stained for GUS in a medium containing 2 mM X-Gluc (company), 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 100 mM sodium phosphate buffer (pH 7.0) and 0.3% (v/v) Triton-X100, vacuum infiltrated for 15 min and incubated for 24–48 hrs at 37^0C. Stained sections were observed using a light microscope.

4. 2. 7. Stress induction experiments

Sugarcane calli of variety NCo310 were placed in MSC_3 media containing MS salts and vitamins, 20 g/l sucrose, 0.5 g/l casein hydrolysate, 3 mg/l 2,4-D, solidified with 8 g/l agar, pH 7.5 for bombardment. After bombardment the callus was transferred to MSC_3 supplemented with 50, 100 and 150 mM of NaCl and 0.4, 0.8, 1 M of mannitol for 48 hrs at 25 ± 1 ^{0}C. Control treatments were callus placed in MSC_3 medium and bombarded with pAHC27 and the promoter-less construct pBScr-GUS000.

4. 2. 8. RNA Blot Analyses

Refer to Chapter 3 (sections on RNA extraction, preparation of probes sequence analyses and RNA blot analyses) for details of the technique used on RNA extraction, RNA blot preparation, manipulation of nucleic acids, sequence analyses and on how to prepare the probe.

4. 2. 9. Nucleotide sequence accession number

The dirigent promoter sequence (DPB) described in this work has been deposited in the DDBJ/EMBL/GenBank nucleotide sequence database under accession no. AJ626722.

4. 3. Results and Discussion

4. 3. 1. Cloning of two putative dirigent promoters sequences

IPCR was performed based on the available, partial dirigent-like cDNA sequence, in order to isolate the corresponding promoter region from the sugarcane genome. Different enzymes were used to digest sugarcane genomic DNA, but subsequent IPCR amplification was only successful for two of the enzymes used, that is Bgl I and Sac I (Figure 4.1A). There was also a need to design the gene specific primers twice because the first set of primers gave false positive IPCR products (Figure 4.1B). One possible explanation for this might be that one of the primers, based on the cDNA sequence, spanned an intron/exon junction in the genomic DNA. Therefore, the presence of introns in the target gene must also be considered when carrying out IPCR from genomic DNA.

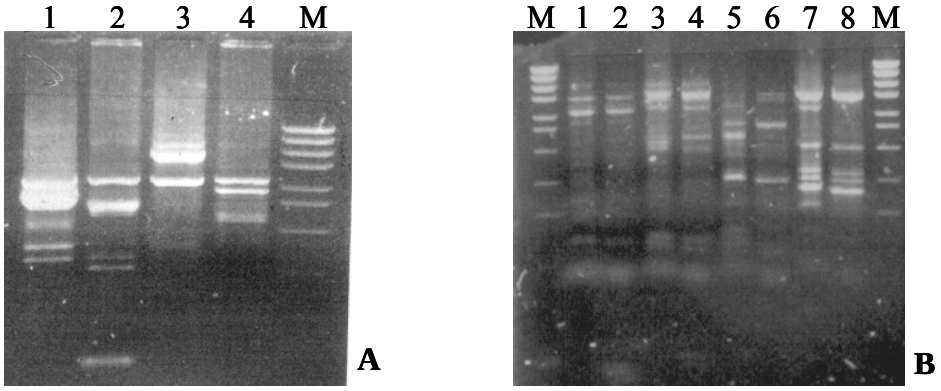


Figure 4.1: LR-iPCR analysis of sugarcane genomic DNA. Panel A IPCR analysis Bgl II- and Sac I-prepared genomic DNA using Dirfrw2-forward and Dirrev2-reverse primers. Lane 1 and 3 are PCR reactions with Bgl II-and Sac I-prepared genomic DNA respectively; lane 2 are 4 are IPCR products digest with Bgl II a d Sac I respectively. Panel B IPCR analysis of using the first set of primers. Lane 2, 4, 6 and 8 are PCR reactions with EcoR V, Hind III, Pst I and Xba I respectively; Lane 1, 3, 5 and 7 are IPCR products digested with EcoR V, Hind III, Pst I and Xba I respectively. Lane M indicates the molecular weight marker.

Two IPCR experiments respectively using Bgl I and Sac I digested, circularized N19 gDNA and using the primer set, Dirfrw2 5'-and Dirrev2, allowed the amplification of 1151 and 985 bp fragments upstream of the putative ATG codon (Figure 4.5 and 4.6). The original dirigent-like cDNA was a partial sequence, but IPCR allowed the amplification of an additional 25 bp of the coding sequence that included a putative ATG (Figure 4.2 and 4.3). BLASTN analyses also confirmed this to be the initiation codon. The BLASTN results using the complete sequence showed a 96% homology with a *Saccharum* hybrid dirigent mRNA (complete coding sequence, accession no. AY421731). The putative promoters isolated were named **D**irigent **P**romoter from **B**gl I digest (DPB) and **D**irigent **P**romoter from **S**ac I digest (DPS) respectively. To confirm that the promoter sequences isolated were from the original dirigent-like cDNA, the sequences were aligned with original cDNA. The 3' regions of these promoter sequences were aligned with the 5' coding region of cDNA and were found to be identical to dirigent-like cDNA sequence, confirming that the correct fragments have been amplified and cloned (Figure 4.2 and 4.3).

Figure 4.2: Alignment of nucleotide sequence of 3' region of DPB and 5' dirigent-like cDNA fragment. The sequence corresponding to the primers used for IPCR are underlined and marked by arrows. The coding sequence of dirigen gene is represented in lowercase. The translational start codon ATG is in bold type.

Figure 4.3: Alignment of nucleotide sequence of 3' region of DPS and 5' dirigent-like cDNA fragment. The sequence corresponding to the primers used for IPCR are underlined and marked by arrows. The coding sequence of dirigen gene is represented in lowercase. The translational start codon ATG is in bold type.

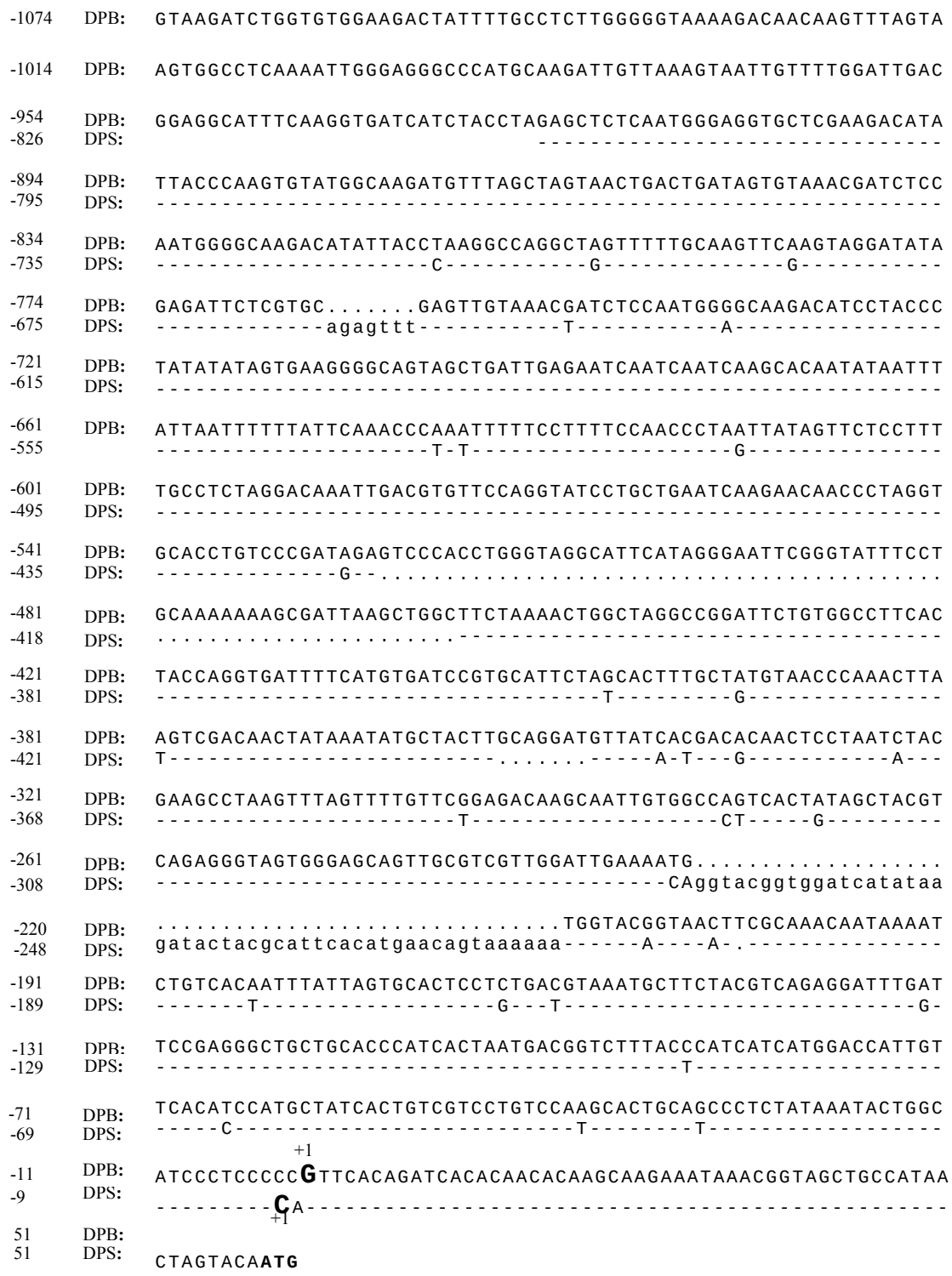


Figure 4.4: Nucleotide sequences of the 5' regulatory regions of the two putative dirigent like genes. Homologous nucleotides are represented by a dash (-), deletions by a dot (.) and additions by lowercase letters. Putative transcriptional start sites are indicated by +1 and are in bold. The translation initiation codon is also indicated in bold.

From the BLAST 2 sequences alignments it was evident that the DPB and DPS sequences differ only slightly (95% homologous). There are also deletions and substitutions in these two promoters (Figure 4.4). In the DPB gene promoter sequence, 50 bp corresponding to nucleotide 659 to 709 in the DPS gene promoter were deleted and in DPS gene promoter sequence 60bp corresponding to nucleotide 557 to 623 were also deleted (Figure 4.4). This could suggest that they represent the promoter regions of two forms of dirigent protein in sugarcane. Isolating two highly homologous promoter sequences from the sugarcane genome is not surprising because of the polyploidy nature of sugarcane. In addition, Southern blot analyses have already suggested that more than one copy of the dirigent-like gene is present in the sugarcane genome (Figure 3.4B in chapter 3).

4. 3. 2. Transcription factor and regulatory region mapping of the putative dirigent promoters

A BLASTN search of a DPB and DPS dirigent promoter sequences did not reveal any significant homology to any known nucleotide sequences in the GenBank database, but it showed 100% homology with a 35 bp 5' UTR sequence of a *Saccharum* hybrid dirigent mRNA coding sequence (variety Q117 accession no. AY421731). This is the same sequence which aligned to the original cDNA fragment. A detailed analysis of these promoter sequences revealed several promoter elements that are common to other plant promoters. The Neural Network Promoter Prediction (NNP2.1, Rees and Eeckman, 1995) tool was used in order to find putative TATA-box elements and transcription starting sites (TSS) for the putative DPB and DPS promoters (Table 4.1). The promoter sequences of both DPB and DPS have identical TATA boxes (TATAAATA) at position -24 bp and -22 bp respectively, in relation to the TSS (Figure 4.5 and 4.6). The presence of a putative TATA box sequence categorises these promoters as a TATA box-containing promoters as classified by Joshi (1987). The localisation TATA-box, TSS and translation-start site elements in the DPB and DPS promoters are in agreement with the results previously obtained by Joshi (1987) for 79 higher plant promoters. The distance between putative TATA-box and the TSS for both promoters falls in the range of 32 ± 7 bp as determined by Joshi (1987) this distance is 24 bp and 22 bp for DPB and DPS promoters respectively. Furthermore, according to Joshi (1987) the distance between the TSS and the translation-starting site is

approximately between 9 and 193 bp, this distance is 56 bp and 58 bp for the DPB and DPS promoters respectively.

Table 4.1: Putative core promoter elements in the upstream sequences of the putative dirigent-like gene as predicted by the NNP2.1 tool[a]

Promoter sequence	Length of the 5' end flanking sequence	Position of core promoter	Score	Promoter sequence[b]
DPB	1151	1061-1111	0.99	AAGCACTGCAGCCCTCTATAAATACTGGCATCCCTCCCCC**G**TTCA CAGAT
DPS	985	893-943	0.99	CCATGCACTGCATCCCTCTATAAATACTGGCATCCCTCCC**C**AGTT CACAG

[a] Neural network promoter prediction (see Section 4.2.2).

[b] Predicted transcriptional start sites are indicated in bold type, predicted TATA box sequences are underlined.

The consensus sequences of the CAAT-boxes identified in the two putative promoter sequences were different. In DPB the putative CAAT promoter element (CACCAAT) is positioned 163 bp upstream from the putative TATA-box whereas in DPS the putative CAAT-box (CAAT) is located 186 bp upstream of the TATA-box (Figure 4.5 and 4.6). In these promoter sequences, the distance between the putative CAAT-boxes and the TSS is greater than the average of 80 bp as reported for several plant promoters by An and Kim (1993).

In addition, PLACE and PLANTCARE based searches for regulatory elements in the putative dirigent promoters revealed six and seven MYB-like recognition sites for DPB and DPS respectively (Figure 4.5, 4.6 and Table 4.2). The MYB gene family represents one of the largest regulatory factor families in plants, and one of the important functions for MYB factor is to control development and determine cell fate and identity (Stracke et al., 2001). MYB recognition sites were also identified in the promoter region of LpLAC5-4 laccase from ryegrass, which is involved in lignin polymerization (Gavnholt et al., 2002). In addition, MYB transcription factors are involved in regulation of signal transduction pathways responding to plant growth regulators as well as in the regulation of the branch of phenylpropanoid metabolism involved in lignin formation (Martin and Paz-Ares, 1997). MYB genes are highly expressed in tissues such as differentiating xylem, supporting the idea that they play a role in the lignification process (Gavnholt et al., 2002).

Furthermore, various other putative *cis*-acting elements were identified in the putative DPB and DPS promoter regions. These include various signal-responsive elements such as light- and phytohormones–responsive elements. Several light responsive elements, namely three GATA-boxes, two GT-1 motifs and an I-box, have been identified which are characteristics *cis*-regulatory elements of light-regulated promoters (Guilfoyle, 1997). Phytohormone–responsive elements identified include auxin, abscissic acid and gibberelin responsive elements.

The processes of lignification and secondary vascular tissue formation are controlled by a wide variety of factors, both exogenous (including photoperiod and temperature) and endogenous (including phytohormones) and by the interaction between them (Plomion et al., 2001). Therefore, the presence of putative light and hormone responsive elements on the promoter of the dirigent-like gene support their putative role in lignification. For example, auxin has shown to enhance lignification in the primary phloem fibres of *Coleus blumei* stems (Aloni et al., 1990) and in secondary xylem of tobacco (Sitbon et al., 1999).

Figure 4.5: Nucleotide sequence of the putative DPB promoter, including 115bp of the coding sequence (GeneBank accession no. AJ626722). The initiation codon (ATG) is in bold and the transcriptional start site (+1) is marked by an arrow. The putative *cis*-acting elements are indicated by boxes, underlining and shading. Names are given directly above or below the elements.

The AAAG-motif ((A/T)AAAG) is a potential target site of the Dof class of DNA binding proteins. Dof proteins are TFs that are unique to plants and they have been found to interact with a variety of promoters from photosynthetic genes, seed storage protein genes and genes responsive to plant hormones and/or stress signals (Yanagisawa, 2002; Yanagisawa and Schmidt, 1999). Three AAAG motifs were identified in the putative DPB promoter. In contrast to DPB, the putative DPS promoter sequence does not contain any AAAG motifs (Figure 4.5, 4.6 and Table 4.2). The presence of MYB recognition sites, the plant-specific Dof motif, light responsive

elements and phytohormones–responsive elements in the dirigent promoter sequence could therefore play a cooperative role with other factors during the lignification process of mature sugarcane culm tissues.

Figure 4.6: Nucleotide sequence of the putative DPS, including 115 bp of the coding sequence. The start codon (ATG) is in bold and the transcriptional start site (+1) is marked by an arrow. The putative *cis*-acting elements are in a box, underlined. Names are given above and below the elements.

Table 4.2: Putative regulatory elements present in the putative DPB and DPS promoters.

Regulatory element	Consensus sequence	Position[a]		Proposed function	References
		DPB[b]	DPS[c]		
ABRE	ACGTG	-492	-350	Abscisic acid responsive element	Simpson et al., 2003
AuxRE	TGACGTAA	-931	NF	Auxin responsive element	Guilfoyle et al., 1998
Dof	AAAG	-42	NF	Dof core recognition sequence, unique to plants	Yanagisawa and Schmidt, 1999
		-92			
		-600			
GATA-box	GATAA	-222	-73	Cis-element for light-and tissue specific responsiveness	Teakle et al. 2002
		-295	-146		
		-545	-679		
GARE (P-box)	CCTTttc	-443	-300	Gibberellin-responsive element	Kim et al., 1992
		-469	-326		
HSE	aAAAAatgg	-867	-703	Heat shock element	Pastuglia et al., 1997
		-897	-730		
	aTAAAatct				
MRE	AACCcaa	-430	-493	MYB binding site involved in light responsiveness	Feldbrugge et al., 1997
		-702			
MYB2AT	TAACTG	-213	-64	An Arabidopsis myb homolog and its gene products are involved in regulation of genes that are responsive to water stress in Arabidopsis	Urao et al., 1993
MYBCORE	CAGTTG	-850	-635	Binding site for all animal MYB and at least two plant MYB proteins ATMYB1 and ATMYB2, ATMYB2 is involved in regulation of genes that are responsive to water stress in Arabidopsis	Urao et al., 1993
MYBPZM	CCT(A)ACC	-347	-205	Maize myb-homologous P gene, binds to and activates transcription of the A1 gene required for 3-deoxy flavonoid and phlobaphene biosynthesis	Grotewold et al., 1994
		-449	-307		
MYBST1	GGATA	-294	-145	Myb-related protein isolated form potato cDNA functioning as a transcriptional activator	Baranowskij et al., 1994
MYCATERD1	CATGTG	-668	-460	MYC recognition sequence (from -466 to -461) necessary for expression of erd1 (early responsive to dehydration) in dehydrated Arabidopsis	Simpson et al., 2003
MYCATRD22	CACATG	NF	-692	Binding site for MYC (rd22BP1) in Arabidopsis dehydration-responsive gene	Abe et al., 1997

NF= not found

[a] Positions are relative to the TSS in the DPB and DPS sequences and the 5' most nucleotide of the sequence is numbered.

B **D**irigent **P**romoter from **B**gl I digest

c **D**irigent **P**romoter from **S**ac I digest

4. 3. 3. Transient expression analyses

Transient expression and stable transformation systems are used to study plant regulatory elements. In this study the activity of the putative promoter sequences (DPB and DPS) was assayed in a transient expression system using the GUS reporter gene. The expression of the GUS gene was analysed by β-glucronidase histochemical assays in various bombarded tissue types. No GUS activity was detected in any of the callus types or in the hypocotyls of maize with the expression vectors containing the putative DPB and DPS promoter sequence. The efficiency of the assay system was verified using an expression vector in which the constitutive UBI promoter drives GUS expression. A great number of GUS-expressing blue cells were detected in sugarcane callus and hypocotyls of maize bombarded with this construct.

The negative results of the functional analyses of the putative DPB and DPS promoters in sugarcane callus and maize hypocotyls probably support the tissue specificity of the promoters. Moreover the lack of activity of these promoters might be due to the absence of the necessary transcription factors, occurring in the culm tissue. Consequently, mature tissues (internode 15) where the dirigent-like gene is highly expressed were used to evaluate the activity of the DPB and DPS promoters. GUS activity was also undetectable in these tissues where the endogenous gene is highly expressed (refer chapter 3, Figure 3.2A and 3.5). In addition, no GUS-expressing cell could be detected in these mature tissues when bombarded with the UBI promoter construct. This clearly indicates that mature culm tissues cannot be used in transient assay for functional analyses of putative promoters.

4. 3. 4. Stress inducablity of the putative dirigent-like gene promoter in sugarcane callus

After evaluating different tissue types in transient expression experiments to verify promoter activity, it was decided to conduct transient expression assays in stressed callus. The c22-a cDNA has showed homology to dirigent and jacalin proteins (refer chapter 3, Table 3.3). The dirigent-like proteins are induced during disease response in plants. Similarly, the jacalin gene

has been isolated and characterised from salt stressed rice plants (Zhang et al., 2000). In the same report the authors claimed that the rice jacalin act as a stress responsive protein. So, based on the homology that the c22-a cDNA showed to these stress responsive proteins, salt and drought stress experiments were conducted to evaluate the promoters' activities in stressed callus. If gene expression was induced in stressed callus it could be used as a transient system under these conditions. No reporter gene expression could be detected for the DPB and DPS promoter in the stressed callus.

So the next step was to confirm if the endogenous gene (the c22-a mRNA) was expressed in the untreated or stress treated callus using RNA blot analysis. These results confirmed that the c22-a cDNA was not expressed in either the untreated or stressed callus. Although the c22-a probe hybridised to a homologous message under stringent conditions this was confirmed not to be c22-a message (Figure 4.7A and B). The observed size difference between c22-a transcript and the one expressed in the callus justifies the conclusion (Figure 4.7A and B). The complete c22-a cDNA has been sequenced and its size is 850 bp (refer chapter 3). In comparison the size of the homologous transcript, which is expressed in callus, is approximately 2100 bp (Figure 4.7A). The size of the homologous transcript was confirmed by its relative position to the small rRNA, which is 1926 bp (Dyer, 1982) (Figure 4.7A). The strong signal observed in the callus RNA blot indicates that both these sequences are highly homologous (Figure 4.7A). The RNA blot result from the callus tissues further confirms the highly mature culm specificity of the c22-a gene (Figure 4.7A and B). The observed expression pattern, the fact that callus tissue is not lignified and the putative identity of the c22-a sequence therefore all ties in.

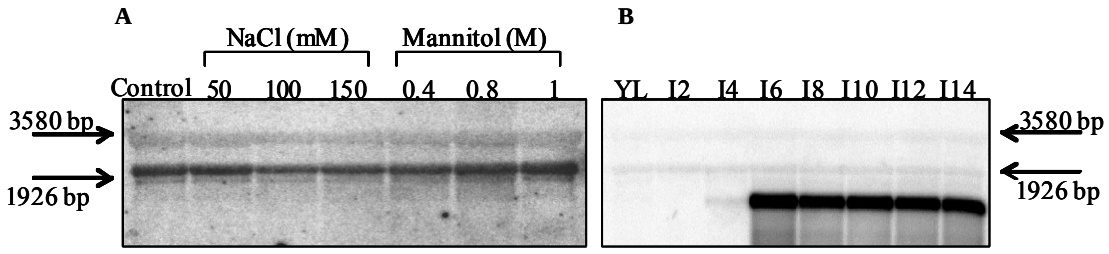


Figure 4.7: RNA blot analyses of c22-a. Panel A shows an RNA blot analysis of c22-a expression in untreated (control) and stress induced callus, salt and drought stress was simulated using NaCl and mannitol at different concentrations. Panels B show the expression profile of c22-a , young leaf (YL), young internodes (I2 and I4), mature internodes (I6, I8, I10, and I12) and an old internode (I14) . Arrows indicate the relative size of two ribosomal RNA, large subunit (3580 bp) and small subunit (1926 bp). The blots contain 10 µg/lane of total RNA.

4. 4. Conclusion

Using the sequence information of the dirigent-like gene, two 5'- flanking sequences of the putative sugarcane dirigent-like gene were cloned. Promoters such as these are of increasing importance because gene silencing is often induced when constitutive promoters, isolated from foreign environments, are used.

The ability of the isolated sequences to function as promoters was evaluated using transient expression assays in different tissues and under various conditions. Although the available system apparently works well for constitutive and some young tissue specific promoters this study indicates that it might be unsuitable for the evaluation of mature culm specific promoters. These results emphasised the necessity to develop molecular genetic tools for transient expression assays for mature culm specific promoters in sugarcane, which would facilitate the *in vivo* evaluation of these elements before transgenic plants are generated.

Final confirmation of the promoter activity or lack thereof of these putative promoter elements will only be possible once mature stably transformed sugarcane plants are available.

CHAPTER FIVE

General Discussion and Conclusions

Sugarcane is one of the world's most important crops, with a significant sucrose storage capacity. During the past two decades, little progress has been made in conventional breeding to increase sucrose content of the commercial sugarcane varieties. Thus considerable efforts are being undertaken in many research programmes to alter carbon partitioning in sugarcane through genetic manipulation. One of the major obstacles in such endeavours, however, is the lack of suitable promoter elements to control transgene expression in the culm. The controlled expression of any transgene is mainly dependent on the choice of a suitable promoter. This promoter will determine the timing, location and amount of the transgene products. It was therefore a primary objective of this project to isolate a mature tissue specific promoter for use in transgenic sugarcane plants. This project had three goals. Firstly, to determine whether the previously identified cDNA clones are expressed in all the mature internodes throughout the culm. Secondly, to determine if the gene/s are expressed in particular cell types only. Thirdly, to isolate the promoter region associated with one of the most appropriate genes

Different physiological, metabolic and genetic studies have been conducted to characterize the sucrose accumulation process in the sugarcane culm. Results obtained from such studies have revealed differences in morphological features and enzyme activities between the different developmental stages (Chapter 2). Each stage during sugarcane maturation is likely to be regulated by a set of interacting genes that are directly involved in the different processes. Genes, which are differentially regulated are under the control of specific regulatory sequences (promoter regions) generally located immediately upstream of the gene (Birch, 1997). Hence, the identification and isolation of genes expressed under specific conditions will allow the isolation of their specific promoter elements.

Preliminary work using cDNA macroarrays has indicated that sugarcane internodes contain a number of cDNA clones, which appear to be expressed only in mature culm specific (Rogbeer, 2002). Based on limited RNA blot data, four cDNA clones; namely, c22, c23, cO and cS were

selected for further analyses. The first phase of this study, described in chapter three, was to analyse the expression pattern of the selected fragments along the length of the culm and in other non-storage tissues, that is immature internodes, leaf roll and roots. The cDNA clones obtained from the previous study were found to contain more than one insert, after EcoR I restericiton analysis c22-a, c23-a, cO-a and cS-a were seleted for the further analysis (Chapter 3, Figure 3.1 and Table 3.1). Further characterisation by RNA blot analysis confirmed two clones, c22-a and c23-a, to be mature culm-specific. The selection of these clones was based on their high degree of mature tissue specificity and homology to relevant genes. Consequently the representation of these genes in the whole genome was determined by Southern blot analysis. The Southern blot data suggested that the corresponding genes are present at a low copy number in the sugarcane genome.

The culm-specific cDNA clone, c22-a, was sequenced and putatively identified as a dirigent-like protein. The transcript is not only specific to mature culm tissues but also to certain cell types. *In situ* hybridization demonstrated that the transcript is highly expressed in parenchyma cells surrounding vascular bundles and storage parenchyma (Chapter 3, Figure 4.5). Similar tissue specific expression patterns were observed for the dirigent gene in *Forsythia intermedia* (Burlat et al., 2001) and *Thuja plicata* (Kim et al., 2002) where it was localized in different types of lignified cell walls of stems. Similarly, the increase in tissue maturity down the length of the sugarcane culm is characterized by an increase in the lignification and suberisation of the cells (Jacobsen et al., 1992). The expression pattern of this gene therefore corresponds with its identity. Furthermore, the results suggest that the gene is under tissue-specific control since its expression differed between young and mature culm tissue of sugarcane.

In addition, in order to determine the number of variants of dirigent-like protein gene in sugarcane, the cDNA fragment of this gene was compared to the sugarcane EST database (www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=s_officinarum). From this comparison eight variants showing homology to each other were obtained. In this study the tissue specificity and expression level of this gene was determined based on hybridisation techniques (RNA blot and *in situ* hybridisation). However, these techniques do not always discriminate between closely related genes with in a gene family. If a high level of transgene expression driven by a strong

promoter is required it might be necessary check the expression level of each dirigent multigene member. This can be done by reverse transcription polymerise chain reaction (RT-PCR) uisng their 5'-UTR or 3'-UTR sequences. These sequences are ideal for RT-PCR, because they are more variable, both in length and content, than the coding sequence for genes and would thus allow discrimination between the different members of the gene family. Furthermore, the 5'-UTR sequence can be used in the design of gene specific primers for inverse-PCR from genome to obtain specific promoter sequences for genes which show high level of expression. A nine member dirigent gene family has been previously described from western red cedar and the corresponding promoter regions for each dirigent multigene family were obtained (Kim et al., 2002a; Kim et al., 2002b).

The differential expression pattern, copy number and the proximity to the 5' end of the cloned cDNA fragment c22-a made it ideal for functional promoter recovery. Consequently, the next phase of this study, described in chapter four, involved the isolation and computational sequence analysis of the dirigent promoter. The isolation of the promoter was performed using the long-range inverse PCR (LR-iPCR) method. LR-iPCR proved to be a convenient and versatile method for cloning unknown sequences upstream or downstream of known sequences. In this study, as described in chapter four, we have successfully isolated two highly homologous promoter regions of the dirigent like gene of respectively 1151 and 985 base pairs. The putative promoters isolated were designated as **D**irigent **P**romoter from **B**gl I digest (DPB) and **D**irigent **P**romoter from **S**ac I digest (DPS). Promoter regions are organised in a series of *cis*-elements where transcription factors binds in order to regulate transcription. These *cis*-elements are highly conserved among different eukaryotic organisms (Butler and Kadonaga, 2002). Therefore, the isolated promoter sequences have been sequenced and the region was analysed for the presence of consensus sequences where transcription factors can potentially bind.

Various websites offer search engines that assess sequences for potential transcription factor binding sites. For this study the PLACE , PlantCARE and NNPP databases were used to search for potential transcription factor binding sites and putative TATA-box elements and TSS (Chapter 4). Based on sequence similarity to well characterised plant *cis*-elements different potential transcription factor binding sites were identified: namely TATA-box, CAAT-box, TSS

and specific boxes involved in response to light and hormones (Chapter 4, Figure 4.5 and 4.6, Table 4.1 and 4.2). The computational analyses result shows which transcription factors are potentially involved in the control of a particular gene and that information in turn suggests the possible signal transduction pathways that may regulate gene expression.

For this study LR-iPCR proved to be simple and much faster to use if compared to other promoter isolation techniques like genomic DNA screening. However like other methods LR-iPCR has also its own drawbacks; one of which is the need for restriction enzyme sites that flank the priming region. The lack of data on genomic sequences of sugarcane and restriction sites greatly reduces the successful cloning rates of promoters.

However, experimental verification is needed to fully assess the functionality of these promoter regions. Verifying the activity of the isolated promoters through transient expression analysis proved to be problematic. Different tissue types have been used in order to evaluate activity of the isolated promoters these includes, sugarcane callus, type III and IV, slices of culm tissue, young internode 5 and matured internode 15, and maize hypocotyls. No GUS activity was detected in any of the tissue used. The negative results of the functional analyses of the putative DPB and DPS promoters in sugarcane callus and maize hypocotyls probably support the tissue specificity of the promoters. Moreover the lack of activity of these promoters might be due to the absence of the necessary transcription factors, occurring in the culm tissue.

Both constructs (pBSGUS DPB and pBSGUS DPS) are currently being used to generate stable transformants in which promoter activity can be evaluated. Future work will focus on the analysis of these transgenic sugarcane plants. If these promoters prove to be functional, tissue specific and stable they will be important inputs for the sugarcane transgenic programme because it will for the first time allow the targeting of transgene activity to mature culm tissue and in doing so reduce the impact on non-target tissues. In addition, the isolated promoters may have applications to manipulate certain aspects of sucrose metabolism and to drive pathogen resistance transgene in mature culm tissue and this in turn will enhance the productivity and economic well-being of the South African sugar industry.

REFERENCES

Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K (1997) Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell **9:** 1859-1868

Adhya S (1989) Multipartite genetic control elements communication by DNA looping. Annual Review of Genetics **23:** 227-250

Alba MM, Pages M (1998) Plant proteins containing the RNA-recognition motif. Trends in Plant Science **3:** 15-21

Allison B, Moyle M, Shales M, Ingles C (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell **42:** 599-610

Aloni R, Tollier M, Monties B (1990) The role of auxin and gibberellin in controlling lignin formation in primary phloem fibers and in xylem of *Coleus Blumei* stems. Plant Physiol. **94:** 1743-1747

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. Journal of Molecular Biology **215:** 403-410

An G, Kim YS (1993) Techniques for isolating and characterising plant transcription promoters, enhancers, and terminators. *In* B Glick, J Thompson, eds, Methods in plant molecular biology and biotechnology. C.R.C Press, London, pp 155-166

Annonymous International sugar statistics. *In* World of sugar, Vol 2004. http://www.illovo.co.za/, Durban

Annonymous SASA Crop Estimates. *In* Sugar business, Vol 2004. http://www.sasa.org.za, Durban

Annonymous Southern African Development Community (SADC) sugar production statistics. *In* World of sugar, Vol 2004. http://www.illovo.co.za/, Durban

Annonymous Sugar Industry. *In* Sugar business, Vol 2004. http://www.sasa.org.za, Durban

Annonymous (2004) TIGR S. officinarum Gene Index. *In* TIGR Gene Indices, Vol 2005. http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=s_officinarum, Rockville

Arias JA, Dixon RA, Lamb CJ (1993) Dissection of the functional architecture of a plant defense gene promoter using a homologous *in vitro* transcription initiation system. Plant Cell **5:** 485-496

Artschwager E (1925) Anatomy of the vegetative organs of sugarcane. Journal of Agricultural Research **30:** 197-221

Baranowskij N, Frohberg C, Prat S, Willmitzer L (1994) A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. EMBO J **13:** 5383-5392

Baszczynski C, Barbour E, Zeka B, Maddock S, Swenson J (1997) Characterisation of a genomic clone for a maize DnaJ-related gene, *ZMDJ1*, and expression analysis of its promoter in transgenic plants. Maydica **42:** 189-201

Belly-Serres J (1999) Selective translation of cytoplasmic mRNA in plants. Trends in Plant Science **4:** 142-148

Belyaev N, Houben A, Baranczewski P, Schubert I (1997) Histone H4 acetylation in plant heterochromatin is altered during the cell cycle. Chromosoma **106:** 641-643

Berger SL (2002) Histone modifications in transcriptional regulation. Current Opinion in Genetics & Development **12:** 142-148

Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, Van Staveren M, Stiekema Wa, others (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature **391:** 485-488

Birch R (1997) Transgenic sugarcane: opportunities and limitations. *In* B Keating, J Wilson, eds, Intensive sugarcane production: meeting the challenges beyond 2000. CAB International, Wallingford, UK, pp 125-140

Birch R, Bower R, Elliott A, Potier B, Franks T, Cordeiro G (1996) Expression of foreign genes in sugarcane. *In* J Cock, T brekelbaum, eds, Proceedings of the International Society of Sugarcane Technologists XXII Congress, Cartegena, September 1995, Vol 2, Tecnicana, Cali, Colombia

Bird A, Wolffe A (1999) Methylation-induced repression-belts, braces, and chromatin. Cell **99:** 451-454

Botha FC (2000) Obstacles impeding the use of genetic manipulation for the improvement of grapes. *In* C du Plessis, ed, Wineland, Vol 2002. http://www.wynboer.co.za/recentarticles/0899obstacles.php3, Capetown

Botha FC, Black K (2000) Sucrose phosphate synthase and sucrose synthase activity during maturation of internodal tissue in sugarcane. Australian Journal of Plant Physiology **27:** 81-85

Boutry M, Chua N-H (1985) A nuclear gene encoding the beta subunit of the mitochondrial ATPase in *N.plumbaginifolia*. EMBO J **4:** 2159-2165

Bremer G (1961) Problems in breeding and cytology of sugarcane. Euphytica **10:** 59-78

Bugos R, Chiang V, Zhang X, Campbell E, Podila G, campbell W (1995) RNA isolation from plant tissues recalcitrant to extraction in guanidine. Biotechniques **19:** 734-737

Burke T, Kadonaga JT (1997) The downstream promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. Genes Dev **11:** 3020-3031

Burlat V, Kwon M, Davin LB, Lewis NG (2001) Dirigent proteins and dirigent sites in lignifying tissues. Phytochemistry **57:** 883-897

Butler JEF, Kadonaga JT (2001) Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev. **15:** 2515-2519

Butler JEF, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev. **16:** 2583-2592

Butterfield M, D'hont A, Berding N (2001) The sugarcane genome: A synthesis of current understanding, and lessons for breeding and biotechnology. Proc S Afr Sug Technol Ass **75:** 1-5

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

Carson DL, Botha FC (2000) Preliminary Analysis of Expressed Sequence Tags for Sugarcane. Crop Sci **40:** 1769-1779

Carson DL, Botha FC (2002) Genes expressed in sugarcane maturing internodal tissue. Plant Cell Reports **20:** 1075-1081

Carson DL, Huckett BI, Botha FC (2002) Sugarcane ESTs differentially expressed in immature and maturing internodal tissue. Plant Science **162:** 289-300

Casu R, Grof C, Rae A, McIntyre C, Dimmock C, Manners J (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol **52:** 371-386

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

Christensen A, Sharrock R, Quail P (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Molecular Biology **18:** 675-689

Coupe S, Deikman J (1997) Characterisation of a DNA-binding protein that interacts with 5' flanking regions of two fruit-ripening genes. The Plant Journal **11:** 1207-1218

Custers JBM, Oldenhof MT, Schrauwen JAM, Cordewener JHG, Wullems GJ, Campagne vL, Michiel M (1997) Analysis of microspore-specific promoters in transgenic tobacco. Plant Mol Biol **35:** 689-699

Daniels J, Roach B (1987) Taxonomy and evolution. *In* D Heinz, ed, Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 7-84

Davin LB, Lewis NG (2000) Dirigent proteins and dirigent sites explain the mystery of specificty of radical precursor coupling in lignin and lignin biosynthesis. Physiol Plant **123:** 453-461

Day D, Tuite M (1998) Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. Journal of Endocrinology **157:** 361-371

de Pater B, Schilperoort R (1992) Structure and expression of a root-specific rice gene. Plant Mol Biol **18:** 161-164

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation. Plant Molecular Biology Reporter **1:** 19-21

Dennis E, Gerlach W, Pryor A, Bennetzen J, Inglis A, Llewellyn D, Sachs M, Ferl R, Peacock W (1984) Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. Nucl. Acids Res. **12:** 3983-4000

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. Molecular Genetics and Genomics **250:** 405-413

Drews G (1998) *In situ* hybridization. *In* J Martinez-Zapater, J Salinas, eds, Methods in molecular biology, Arabidopsis protocols. Humana Press Inc, Totowa, pp 353-371

Dynan W (1989) Modularity in promoters and enhancers. Cell **58:** 1-4

Ebert P, Ha S, An G (1987) Identification of an essential upstream element in the nopaline synthase promoter by stable and transient expression. PNAS **83:** 5372-5376

Edwards JW, Coruzzi GM (1990) Cell-Specific Gene Expression in Plants. Annual Review of Genetics **24:** 275-303

Elliott A, Campbell J, Brettell R, Grof CP (1998) *Agrobacterium*-mediated transformation of sugarcane using GFP as a screenable marker. Australian Journal of Plant Physiology **25:** 739-743

Ellis J, Llewellyn D, Walker J, Dennis E, Peacock W (1987) The OCS element : a 16 base pair palindrome essential for activity of the octopine synthase enhancer. EMBO J **6:** 3203-3208

Ellmore G (2000) Biology 202. *In* Anatomy Figures, Vol 2003. http://www.science.smith.edu/.../overheads/monocotstemxs.jpg.

Faktor O, Kooter JM, Dixon RA, Lamb CJ (1996) Functional dissection of bean chalcone synthase gene promoter in transgenic tobacco plants reveals sequence motifs essential for floral expression. Plant Mol Biol **32:** 849-859

Feldbrugge M, Sprenger M, Hahlbrock K, Weisshaar B (1997) PcMYB1, a novel plant protein containing a DNA-binding domain with one MYB repeat, interacts in vivo with a light-regulatory promoter unit. Plant Journal **11:** 1079-7093

Finer J, Vain P, Jones M, McMullen M (1992) Development of the particle inflow gun for DNA delivery to plant cells. Plant Cell Reports **11:** 323-328

Finnegan EJ (2001) Is plant gene expression regulated globally? Trends in Genetics **17:** 361-365

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (1998) DNA Methylation in plants. Annual Review of Plant Physiology and Plant Molecular Biology **49:** 223-247

Fluhr R, Kuhlemeier C, Nagy F, Chua N-H (1986) Organic-specific and light-induced expression of plant genes. Science **232:** 1106-1112

Force A, Lynch M, Pickett FB, Amores A, Yan Y-l, Postlethwait J (1999) Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics **151:** 1531-1545

Forster C, Arthur E, Crespi S, Hobbs S, Mullineaux P, Casey R (1994) Isolation of a pea (*Pisum sativum*) seed lipoxygenase promoter by inverse polymerase chain reaction and characterization of its expression in transgenic tobacco. Plant Mol Biol **26:** 235-248

Franck A, Guilley H, Jonard G, Richards K, Hirth L (1980) Nucleotide sequence of cauliflower mosaic virus DNA. Cell **21:** 285-294

Gallie D, Browning K (2001) eIF4G functionally differs from eIFiso4G in promoting internal initiation, cap-independent translation, and translation of structured mRNAs. The Journal of Biological Chemistry **276:** 36951-36960

Gallo-Meagher M, Irvine J (1993) Effects of tissue type and promoter strength on transient GUS expression in sugarcane following particle bombardment. Plant Cell Reports **12:** 666-670

Gavnholt B, Larsen K (2002) Molecular biology of plant laccases in relation to lignin formation. Physiologia Plantarum **116:** 273-280

Gittins JR, Pellny TK, Hiles ER, Rosa C, Biricolti S, James DJ (2000) Transgene expression driven by heterologous ribulose-1,5-bisphosphate carboxylase/oxygenase small-subunit gene promoters in the vegetative tissues of apple (Malus pumila Mill.). Planta **210:** 232-240

Gottlieb L, Ford V (1997) A recently silenced, duplicate PgiC locus in Clarkia. Mol Biol Evol **14:** 125-132

Grandbastien M-A (1992) Retroelements in higher plants. Trends in Genetics **8:** 103-108

Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. Trends in Genetics **17:** 100-107

Grivet L, Arruda P (2002) Sugarcane genomics: depicting the complex genome of an important tropical crop. Current Opinion in Plant Biology **5:** 122-127

Grof C, Campbell J (2001) Sugarcane sucrose metabolism: scope for molecular manipulation. Australian Journal of Plant Physiology **28:** 1-12

Grotewold E, Drummond B, Bowen B, Peterson T (1994) The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell **76:** 543-553

Guilfoyle T (1997) The structure of plant gene promters. *In* J Setlow, ed, Genetic Engineering, Vol 19. New York, Plenum Press, pp 15-47

Guilfoyle T, Hagen G, Ulmasov T, Murfett J (1998) How does auxin turn on genes? Plant Physiol. **118:** 341-347

Harberd N, Flavell R, Thompson R (1987) Identification of a transposon-like insertion in a *Glu-1* allele of wheat. Molecular Genetics and Genomics **209:** 326-332

Hardie D (2000) Plant Protein-Serine/Threonine Kinases: classification into subfamilies and overview of function. *In* M Kreis, J Walker, eds, Advances in Botanical Research incorporating advances in plant Pathology, Vol 32. Acadamic Press, London, pp 1-33

Hedley PE, Maddison AL, Davidson D, Machray GC (2000) Differential expression of invertase genes in internal and external phloem tissues of potato (*Solanum tuberosum* L.). J. Exp. Bot. **51:** 817-821

Heineke D, Sonnewald U, Bussis D, Gunter G, Leidreiter K, Wilke I, Raschke k, Willmitzer L, Heldt H (1992) Apoplastic expression of yeast derived invertase in potato. Effects on photosynthesis, leaf solute composition, water relations and tuber composition. Plant Physiol. **100:** 301-308

Higo K, Ugawa Y, Iwamoto M, Higo H (1998) PLACE: a database of plant cis-acting regulatory DNA elements. Nucl. Acids Res. **26:** 358-359

Hongmei M, Henrik H, Robert P, Moore P (2000) Metabolic engineering of invertase activities in different subcellular compartments affects sucrose accumulation in sugarcane cells. Australian Journal of Plant Physiology **27:** 1021-1030

Irvine JE (1999) *Saccharum* species as horticultural classes. Theor Appl Genet **98:** 186-194

Jacobsen K, Fisher D, Maretzki A, Moore P (1992) Developmental changes in the anatomy of the sugarcane stem in relation to phloem unloading and sucrose storage. Bot. Acta **105:** 70-80

Jefferson R (1987) Assaying Chimeric genes in plants: the GUS-fusion system. Plant Mol Biol **5:** 387-405

Joshi C (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl. Acids Res. **15:** 6643-6653

Katagiri F, Lam E, Chua N-H (1989) Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature **340:** 727-730

Kellogg EA (2003) What happens to genes in duplicated genomes. PNAS **100:** 4369-4371

Kim J, Cao J, Wu R (1992) Regulation and interaction of multiple protein factors with the proximal promoter regions of a rice high pI alpha-amylase gene. Mol Gen Genet **232:** 383-393

Kim MK, Jeon J-H, Fujita M, Davin LB, Lewis NG (2002a) The western red cedar (*Thuja plicata*) 8-8' DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. Plant Mol Biol **49:** 199-214

Kim MK, Jeon J-H, Davin LB, Lewis NG (2002b) Monolignol radical-radical coupling networks in western red cedar and Arabidopsis and their evolutionary implications. Phytochemistry **61:** 311-322

Kiss T, Abel S, Solymosy F (1989) A plant pseudogene for U1 RNA. Plant Mol Biol **12:** 709-711

Komor E (2000) The physiology of sucrose storage in sugarcane. *In* A Gupta, N Kaur, eds, Carbohydrate Reserves in Plants, pp 35-54

Kouzarides T (2002) Histone methylation in transcriptional control. Current Opinion in Genetics and Development **12:** 198-209

Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. **266:** 19867-19870

Kuhlemeier C, Green P, Chua N-H (1987) Regulation of gene expression in higher plants. Annual Review of Plant Physiology **38:** 221-257

Kuhn E (2001) From Library Screening to Microarray Technology: Strategies to Determine Gene Expression Profiles and to Identify Differentially Regulated Genes in Plants. Ann Bot **87:** 139-155

Kumar A, Taylor MA, Arif SAM, Davies HV (1996) Potato plants expressing antisense and sense S-adenosylmethionine decarboxylase (SAMDC) transgenes show altered levels of polyamines and ethylene: antisense plants display abnormal phenotypes. Plant J **9:** 147-158

Latchman D (1998) Gene regulation- A eukaryotic perspective. Stanley thorens Ltd, 3rd edition

Lefebvre D, Gellatly K (1997) Fundamentals of gene structure and control. *In* D Dennis, D Turpin, D Lefebvre, D Layzell, eds, Plant Metabolism (2nd Ed.). Longman Singapore publishers, Singapore, pp 3-16

Lingle S (1999) Sugar metabolism during growth and development in sugarcane internodes. Crop Sci **39:** 480-486

Lister C, Jackson D, Martin C (1993) Transposon-induced inversion in antirrhinum modifies nivea gene expression to give a novel flower color pattern under the control of cycloidearadialis. Plant Cell **5:** 1541-1553

Liu D, Oard SV, Oard JH (2003) High transgene expression levels in sugarcane (*Saccharum officinarum* L.) driven by the rice ubiquitin promoter RUBQ2. Plant Science **165:** 743-750

Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants: Functional domains, evolution and regulation. Eur J Biochem **262:** 247-257

Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MHL, Filipowicz W (2000) Pre-mRNA splicing in higher plants. Trends in Plant Science **5:** 160-167

Lu Y, D'hont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC (1994) Molecular diversity and genome structure in modern sugarcane varieties. Euphytica **78:** 217-226

Luger K, Mader A, Richmond R, Sargent D, Richmond T (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature **389:** 251-260

Lusser J, Brosch G, Loidl A, Haas H, Loidl P (1997) Identification of maize histone deacetylase HD2 as an acidic nucleolar phosphoprotein. Science **277:** 88-91

Ma H, Schulze S, Lee S, Yang M, Mirkov E, Irvine J, Moore P, Paterson A (2004) An EST survey of the sugarcane transcriptome. Theor Appl Genet **108:** 851-863

Mandel T, Fleming A, Krahenbuhl R, Kuhlemeier C (1995) Definition of constitutive gene expression in plants: the translation initiation factor 4A gene as a model. Plant Mol Biol **29:** 995-1004

Maniatis T, Fritsch E, Sambrook J (1989) Molecular Cloning. A Laboratory Manual, 2nd edition. Cold Spring Habour Laboratory Press, New York

Maniatis T, Goodbourn S, Fischer J (1987) Regulation of inducible and tissue-specific gene expression. **236:** 1237-1245

Martin C (1996) Transcription factors and the manipulation of plant traits. Current Opinion in Biotechnology **7:** 130-138

Martin C, Paz-Ares J (1997) MYB transcription factors in plants. Trends in Genetics **13:** 67-73

May BP, Dellaporta SL (1998) Transposon sequences drive tissuespecific expression of the maize regulatory gene R-*s*. Plant J **13:** 241-247

McElroy D, Blowers A, Jenes B, Wu R (1991) Construction of expression vectors based on the rice actin 1 (*Act 1*) 5' region for use in monocot transformation. Molecular Genetics and Genomics **231:** 150-160

McElroy D, Zhang W, Cao J, Wu R (1990) Isolation of an efficient actin promoter for use in rice transformation. Plant Cell **2:** 163-171

Meshi T, Iwabuchi M (1995) Plant transcription factors. Plant Cell Physiology **36:** 1405-1420

Ming R, Liu S-C, Moore PH, Irvine JE, Paterson AH (2001) QTL Analysis in a Complex Autopolyploid: Genetic Control of Sugar Content in Sugarcane. Genome Res. **11:** 2075-2084

Mitchell P, Tjian R (1989) Transcriptional regulation in mamalian cells by sequence-specific DNA binding proteins. Science **245:** 371-378

Moore P, Botha FC, Furbank R, Grof CPL (1997) Potential for overcoming physico-biochemical limits to sucrose accumulation. *In* B Keating, J Wilson, eds, Intensive sugarcane production: meeting the challenges beyond 2000. CAB International, Wallingford, UK, pp 141-155

Moore P, Cosgrove D (1991) Developmental changes in cell and tissue water relations parameters in storage parenchyma of sugarcane. Plant Physiol. **96:** 794-801

Moore P, Maretzki A (1996) Sugarcane. *In* E Zamski, M Schaffer, eds, Photoassimilate distribution in plants and crops. Source and sink relationship. Marcel Dekker, New York, pp 643-669

Moore PH (1995) Temporal and spatial of sucrose accumulation in the sugarcane stem. Australian Journal of Plant Physiology **22:** 661-679

Nikolov D, Burley S (1997) RNA polymerase II transcription initiation: A structural view. PNAS **94:** 15-22

Novina C, Roy A (1996) Core promoters and transcriptional control. Trends Genet **12:** 351-355

Nunberg A, Li Z, Chung H-J, Reddy A, Thomas T (1995) Proximal promoter sequences of sunflower helianthinin genes confer regionalized seed-specific expression. J Plant Physiol **145:** 600-605

Önnerud H, Zhang L, Gellerstedt G, G H (2002) Polymerization of Monolignos by redox shuttle-mediated enzaymatic oxidation: A new model in lignin biosynthesis I. The Plant Cell **14:** 1953-1962

Orphanides G, Lagrange T, Reinberg D (1996) The general transcription factors of RNA polymerase II. Genes Dev **10:** 2657-2683

Orphanides G, Reinberg D (2002) A unified theory of gene expression. Cell **108:** 439-451

Osborn TC, Chris PJ, Birchler JA, Auger DL, Jeffery CZ, Lee H-S, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA (2003) Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics **19:** 141-147

Pastuglia M, Roby D, Dumas C, Cock J (1997) Rapid induction by wounding and bacterial infection of an S gene family receptor-like kinase gene in Brassica oleracea. Plant Cell **9:** 49-60

Peumans W (2000) Isolation and characterization of a jacalin-related mannose-binding lectin from salt-stressed from rice (*Oryza sativa*) plants. Planta **210:** 970-978

Pikaard CS (2001) Genomic change and gene silencing in polyploids. Trends in Genetics **17:** 675-677

Plomion C, Leprovost G, Stokes A (2001) Wood formation in trees. Plant Physiol. **127:** 1513-1523

Ptashne M, Gann A (1997) Transcriptional activation by recruitment. Nature **386:** 569-577

Rathus C, Bower R, Birch R (1993) Effects of promoter, intron and enhancer elements on transient gene expression in sugar-cane and carrot protoplasts. Plant Mol Biol **23:** 613-618

Raven P, Johnson G (1999) Biology, Ed 5. The Mcgraw-Hill Companies, Inc., Boston

Reed R (2000) Mechanisms of fidelity in pre-mRNA splicing. Current Opinion in Cell Biology **12:** 340-345

Rees M, Eeckman F (1995) New neural network algorithms for improved eukaryotic promoter site recognition. Genome Sci. Technol. **1:** 45

Richards E, Elgin S (2002) Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell **108:** 489-500

Rocha-Sosa M, Sonnewald U, Frommer W, Stratman M, Schell J, Willmitzer L (1989) Both development and Metabolic signals activate the promoter of a class I patatin gene. EMBO J **8:** 23-29

Roeder R (1996) The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci **21:** 327-335

Roeder RG (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. Trends in Biochemical Sciences **16:** 402-408

Rogbeer O (2002) Differential gene expression in the culm of sugarcane during development, with special emphasis on the storage parenchyma cells. Msc Dissertation. Stellenbosch

Rombauts S, Déhais P, Van Montagu M, and Rouzé P (1999) PlantCARE, a plant cis-acting regulatory element database. Nucleic Acids Res **27:** 295-296

Rose S, Botha FC (2000) Distribution patterns of neutral invertase and sugar content in sugarcane internodal tissues. Plant Physiology and Biochemistry **38:** 819-824

Salgueiro S, Pignocchi C, Parry MJ (2000) Intron-mediated gusA expression in tritordeum and wheat resulting from particle bombardment. Plant Mol Biol **42:** 615-622

Sambrook J, Fritsch E, Maniatis T (1989) Molecular Cloning. A Laboratory Manual, 2nd edition. Cold Spring Habour Laboratory Press, New York

Sargent T (1987) Isolation of differentially expressed genes. Methods in Enzymology **152:** 423-432

Sarni-Manchado P (1997) Molecular characterization and structural analysis of one alcohol dehydrogenase gene (GV-*Adh1*) expressed during ripening of grapevine (*Vitis Vinifera*) berry. Plant Science **125:** 177-187

Schafer WE, Rohwer JM, Botha FC (2004) Protein-level expression and localization of sucrose synthase in the sugarcane culm. Physiol Plant **121:** 187-195

Schwarzacher T, Heslop-Harrison P (2000) Practical *in situ* hybridization. BIOS Scientific Publishers Ltd, Oxford

Schwechheimer C, Bevan M (1998) The regulation of transcription factor activity in plants. Trends in Plant Science **3:** 378-383

Simpson S, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant Journal **33:** 259-270

Singh K, Dennis ES, Ellis JG, Llewellyn DJ, Tokuhisa JG, Wahleithner JA, Peacock WJ (1990) OCSBF-1, a Maize Ocs enhancer binding factor: isolation and expression during development. Plant Cell **2:** 891-903

Singh KB (1998) Transcriptional regulation in plants: The Importance of Combinatorial Control. Plant Physiol. **118:** 1111-1120

Sitbon F, Hennion S, Little C, Sundberg B (1999) Enhanced ethylene production and peroxidase activity in IAA-overproducing transgenic tobacco plants in associated with increase lignin content and altered lignin composition. Plant Science **141:** 165-173

Smale S (1997) Transcription initiation from TATA-less Promoters within eukaryotic protein-coding genes. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1351:** 73-88

Smale S, Baltimore D (1989) The "initiator" as a transcription control element. Cell **57:** 103-113

Stargell LA, Struhl K (1996) Mechanisms of transcriptional activation in vivo: two steps forward. Trends in Genetics **12:** 311-315

Stevenson G (1965) Genetics and breeding of sugarcane. *In*. Longman, Green and Co. Ltd., London, pp 1-38

Stitt M (1995) The use of transgenic plants to study the regulation of plant carbohydrate metabolism. Australian Journal of Plant Physiology **22:** 635-646

Stitt M, Sonnewald U (1995) Regulation of metabolism in transgenic plants. Annual Review of Plant Physiology and Plant Molecular Biology **46:** 341-368

Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. Current Opinion in Plant Biology **4:** 447-456

Tang W, Sun S, Nagai C, Moore P (1996) Regulation of expression of ß-glucuronidase in transgenic sugarcane by promoter of rubisco small subunit genes. *In* J Wilson, D Hogarth, J Campbell, A Garside, eds, Sugarcane research towards efficient and sustainable production, CSIRO Division of Tropical Crops and Pastures, Birsbane Australia, pp 117-119

Teakle G, Manfield I, Graham J, Gilmartin P (2002) Arabidopsis thaliana GATA factors: organisation, expression and DNA-binding characteristics. Plant Mol Biol **50:** 43-57

Triglia T (2000) Inverse PCR (IPCR) for obtaining promoter sequence. *In* M Tymms, ed, Methods in Molecular Biology: Transcription factor protocols. Humana Press Inc, Totowa, NJ

Trindade LM, Horvath B, Bachem C, Jacobsen E, Visser RGF (2003) Isolation and functional characterization of a stolon specific promoter from potato (*Solanum tuberosum* L.). Gene **303:** 77-87

Tyagi A (2001) Plant gene and their expression. Current science **80:** 161-169

Urao T, Yamaguchi-Shinozaki K, Urao S, Shinozaki K (1993) An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell **5:** 1529-1539

Van der Geest A, Hall T (1996) A 68 bp element of the ß-phaseeolin promoter functions as a seed-specific enhancer. Plant Mol Biol **32:** 579-588

Vasil IK (1994) Molecular improvement of cereals. Plant Mol Biol **25:** 925-937

Vasudevan S, Peltz S (2003) Nuclear mRNA surveillance. Current Opinion in cell Biology **15:** 332-337

Vaux P, Guerineau F, Waugh R, Brown J (1992) Characterization and expression of U1snRNA genes from potato. Plant Mol Biol **19:** 959-971

Vom Endt D, Kijne JW, Memelink J (2002) Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry **61:** 107-114

Vorster D, Botha F (1999) Sugarcane internodal invertases and tissue maturity. J Plant Physiol **155:** 470-476

Wendel J (2000) Genome evolution in polyploids. Plant Mol Biol **42:** 225-249

Wendler R, Veith R, Dancer J, Stitt M, Komor E (1990) Sucrose storage in cell suspension cultures of *Saccharum* sp. (sugarcane) is regulated by a cycle of synthesis and degradation. Planta **183:** 31-39

White S, Habera L, Wessler S (1994) Retrotransposons in the flanking region of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. PNAS **91:** 11792-11796

Whittaker A, Botha FC (1997) Carbon partitioning during sucrose accumulation in sugarcane internodal tissue. Plant Physiol. **115:** 1651-1659

Will CL, Luhrmann R (1997) Protein functions in pre-mRNA splicing. Current Opinion in Cell Biology **9:** 320-328

Will CL, Luhrmann R (2001) Spliceosomal UsnRNP biogenesis, structure and function. Current Opinion in Cell Biology **13:** 290-301

Wilmink A, van de Ven B, Dons J (1995) Activity of constitutive promoters in various species from the Liliaceae. Plant Mol Biol. **28:** 949-955

Wu C-Y, Suzuki A, Washida H, Takaiwa F (1998) The GCN4 motif in a rice glutelin gene is essential for endosperm specific gene expression and is activated by Opaque2 in transgenic rice plants. Plant J **14:** 673-683

Wu C-Y, Washida H, Ondera Y, Harada Y, Takaiwa F (2000) Quantitative nature of the prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal cis-element requirements for endosperm- specific gene expression. The Plant Journal **23:** 415-421

Yamaguchi Y, Itoh Y, Takeda Y, Yamazaki K-i (1998) TATA sequence requirements for the initiation of transcription for an RNA polymerase II *in vitro* transcription system from *Nicotiana tabacum*. Plant Mol Biol **38:** 1247-1252

Yanagisawa S (2002) The Dof family of plant transcription factors. Trends Plant Sci **7:** 555-560

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

Yang M, Bower R, Burow MD, Paterson AH, Mirkov TE (2003) A Rapid and direct approach to identify promoters that confer high levels of gene expression in monocots. Crop Sci **43:** 1805-1813

Yoshihara T, Washida H, Takaiwa F (1996) A 45-bp proximal region containing AACA and GCN4 motif is sufficient to confer endosperm-specific expression of the rice storage protein glutelin gene, *GluA*-3. FEBS **383:** 213-218

Zhang JZ (2003) Overexpression analysis of plant transcription factors. Current Opinion in Plant Biology **6:** 430-440

Zhang W, McElroy D, Wu R (1991) Analysis of rice Act 5' region activity in transgenic rice plants. . The Plant Cell **3:** 1155-1165

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

Zhu Y, Henrik H, Moore P (2000) Differential expression of soluble acid invertase genes in the shoots of high-sucrose and low-sucrose species of *Saccharum* and their hybrids. Australian Journal of Plant Physiology **27:** 193-199

Zhu YJ, Komor E, Moore PH (1997) Sucrose accumulation in the sugarcane stem is regulated by the difference between the activities of soluble acid invertase and sucrose phosphate synthase. Plant Physiol. **115:** 609-616

Zuo J, Chua N-H (2000) Chemical-inducible systems for regulated expression of plant genes. Current Opinion in Biotechnology **11:** 146-151