CaMV Gene Expression:

The Analysis of two CaMV Promoters in Yeast and Higher Plants

Jennifer H Richardson BSc (York)

Leicester Biocentre and Department of Genetics,

University of Leicester

PhD 1988

UMI Number: U499751

UMI U499751

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgement

ABSTRACT

The aim of this study was to assess the feasibility of using the budding yeast Saccharomyces cerevisiae as a system in which to analyse plant promoters. The promoters chosen for study were the 19S and 35S promoters of cauliflower mosaic virus (CaMV) which, like cellular plant promoters, are transcribed in the plant nucleus by host cell RNA polymerase II. A complete CaMV genome was introduced into yeast on a 2 micron plasmid-based vector and using Northern blot analysis, several CaMV-hybridising transcripts were detected. More precise information on the activity of the promoters was obtained by constructing gene fusions in which the 19S and 35S promoters were linked to the bacterial lacZ gene. Biochemical assays for β-galactosidase showed that cells harbouring the 19S-lacZ gene expressed β-galactosidase but those harbouring the 35S-lacZ gene did not. The insertion of a yeast transcription termination signal upstream of the 19S promoter did not abolish or diminish expression of the 19S-lacZ gene. β-galactosidase was present at low levels in cells expressing 19S-lacZ, constituting less than 0.01% of total cell protein. The 5'ends of 19S-lacZ transcripts present in yeast were mapped by primer extension. The major RNA species initiated approximately 250bp upstream of the 19S-lacZ coding region, indicating the existence of a fortuitous promoter in this region of the CaMV DNA. Two less abundant RNA species initiated within the 19S-lacZ open reading frame at positions +9 and +25bp and may be produced from the genuine 19S promoter. There is evidence to suggest that one or both of these shorter transcripts is the functional mRNA for β-galactosidase. All three classes of RNA were polyadenylated. Coupling of the 19S-lacZ gene to a yeast enhancer (the GAL UAS) produced a 5-fold increase in β-galactosidase activity. At the transcriptional level, activation of the enhancer resulted in a massive increase in the level of the RNA initiating at -250bp but had a minor influence of the levels of the two RNA species initiating at +9 and +25. A series of deletion mutations within the 19S promoter was constructed using Bal31 nuclease. Analysis of these mutations in yeast revealed that sequences from -500 to -193bp and from -137 to -62bp were not required for 19S promoter function, but a deletion from -62 to -21bp (which removes the putative TATA box) severely reduced 19S-lacZ gene expression. Transgenic tobacco plants containing the 19S promoter deletions fused to a CAT gene were produced by A.tumefaciens-mediated gene transfer but the analysis of these plants was not completed.

CONTENTS

Abbreviations Used

BAP	benzyladenine 6-aminopurine
bp	base pair
BSA	bovine serum albumin
Ci	Curie
cpm	counts per minute
CTAB	cetyl triethylammonium bromide
DEPC	diethylpyrocarbonate
ddNTP	dideoxynucleotide triphosphate
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
EDTA	diaminoethanetetra-acetic acid
IMS	industrial methylated spirit
IPTG	isopropylthiogalactoside
kb	kilobase
MOPS	morpholinopropanesulphonic acid
NAA	α napthalene acetic acid
nt	nucleotide
NTP	nucleotide triphosphate
OD	optical density
ONPG	o-nitrophenol-β-D-galactopyranoside
PAGE	polyacrylamide gel electrophoresis
PMSF	phenylmethylsulphonylfluoride
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
TEMED	N,N,N',N',tetramethylethylenediamine
Tris	2-amino-2-(hydroxymethyl) propane-1,3-diol(tris)
TLC	thin layer chromatography
X gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CHAPTER 1

INTRODUCTION

The development of gene cloning and DNA transfer technology has
radically altered our approach to the study of gene expression and
provided valuable new insights into the mechanisms by which gene
expression is regulated.   An approach now frequently adopted is to
isolate a gene of interest and make specific mutations in the putative
regulatory region *in vitro*.   The gene is then re-introduced into the
natural host species or an alternative host, where its expression *in
vivo* can be assessed.

A variety of systems have been developed for stably introducing foreign
DNA into the plant nuclear genome.   The most widely used technique is
based on the natural DNA transfer system of <u>Agrobacterium tumefaciens</u>
(Caplan *et al* 1983) and has been successfully used to answer questions
about plant gene regulation (Herrera-Estrella *et al* 1984, Lamppa *et al*
1985).   Following DNA transfer by this or any other route, the
genetically modified cells must be amplified by tissue culture, and
ideally regenerated into mature plants, before expression of the
transforming DNA can be investigated.   This approach to genetic
analysis has met with a number of limitations.   For example, not all
plant species are susceptible to <u>A.tumefaciens</u>-mediated transformation.
The experiments are very time consuming as it takes many weeks of
tissue culture to generate sufficient callus tissue for analysis and
considerably longer if regeneration of mature plants is required.
Furthermore, studies of this type are restricted to those plants
species which can be propogated in tissue culture.

While there is much to be gained by replacing cloned plant genes in

their natural genetic background where sophisticated regulatory mechanisms such as tissue-specific or developmental control can be examined, there are clearly instances where an alternative expression system is necessary or desirable. Transient gene expression systems using plant protoplasts have been described for both monocot and dicot species (Fromm *et al* 1985, Werr and Lorz 1986). These provide a rapid system for gene expression studies, overcoming the need to select and propogate stably transformed tissue, and also permit the analysis of genes from species which are not amenable to tissue culture.

From a practical point of view, however, an equally rapid but more flexible and convenient way to analyse plant genes would be to express them in a eukaryotic micro-organism such as yeast. The aim of this project was to assess the feasibility of using the yeast S.cerevisiae as an alternative cell system in which to analyse plant promoters. Cauliflower mosaic virus is a DNA plant virus and its genome contains two promoters which interact with host cell RNA polymerase II and direct the synthesis of 19S and 35S transcripts. It is these two promoters which have been selected for study in yeast.

1.1   YEAST AS AN EXPERIMENTAL ORGANISM

The budding yeast <u>Saccharomyes</u> <u>cerevisiae</u> is an ascomycete fungus and has long been a popular research tool for genetic and biochemical studies. It is a unicellular eukaryote and can be cultured in much the same way as bacteria, making the growth and storage of genetic stocks relatively simple compared with other eukaryotic systems. Nevertheless, yeast is a typical eukaryote; its genetic material is confined to the nucleus in the form of chromatin, transcription and translation processes have characteristic eukaryotic features, nuclear division is by mitosis, and typically eukaryotic organelles, such as mitochondria, are found in the cytoplasm. Fundamental metabolic processes appear to be conserved among eukaryotes and, being comparatively simple, yeast is widely selected as a model organism for investigating basic problems in genetics and molecular biology.

1.11 Heterologous gene expression in yeast

The cloning in 1981 of a <u>Drosophila</u> gene by functional complementation of the yeast <u>ade</u>8 mutation (Henikoff *et al* 1981) was the first demonstration that genes from higher eukaryotes are able to work in yeast. Since that time, yeast has proved useful for both cloning and expressing genes from other eukaryotes. Several features make yeast a convenient organism for such studies. Firstly, it is genetically well defined and a large number of well characterised mutations are known. Secondly, yeast is very amenable to genetic manipulation. Efficient procedures for the transformation of sphaeroplasts (Beggs 1978) and more recently whole cells (Ito *et al* 1983) have been developed, together with a wide array of cloning vectors and selectable markers

(reviewed by Struhl 1983, Parent *et al* 1985). Thirdly, the secretory pathway of yeast can be manipulated to allow the export of heterologous proteins out of the cell.

A list of foreign eukaryotic genes which have been introduced into yeast is provided in Table 1.1, and can be split into two categories. The first and larger group consists of genes or cDNA clones encoding proteins of commercial interest such as starch-degrading enzymes (Filho *et al* 1986, Innis *et al* 1985), potential pharmaceutical products like human interferon and proinsulin (Hitzeman *et al* 1981, Stepien *et al* 1983), and viral antigens which may be used for vaccine production (Valenzuela *et al* 1983). To achieve maximum yields or inducible expression, these genes are invariably expressed using well characterised yeast promoters, and are frequently engineered so that the mature polypeptide is preceded by either its own (Edens *et al* 1984, Innis *et al* 1985) or a yeast (Filho *et al* 1986) signal peptide. The signal peptide is responsible for targeting the foreign polypeptide to the endoplasmic reticulum, from which it is exported to the cell surface via the secretory pathway.

The second category of foreign genes are those being expressed under the control of their natural 5' and 3' flanking sequences. These are genes which, for a variety of reasons, are more easily studied in a heterologous host, and it is with such genes that this project is primarily concerned. Some of the earlier attempts to express foreign genes in yeast using their natural promoters met with little success. Although, for example, a Drosophila gene corresponding to the yeast ADE8 locus had been identified by complementation, Drosophila genes complementing mutations at other yeast loci were not obtained (Henikoff *et al* 1981). Similarly, complementation of yeast mutations was rarely

4

Table 1.1   Eukaryotic genes expressed in yeast

SOURCE	GENE	REFERENCE
Schizosaccharomyces pombe	*triose phosphate isomerase	Russell 1985
	*alcohol dehydrogenase	Russell 1983
Saccharomyces diastaticus	*amyloglucosidase	Meaden et al 1985
	*glucoamylase	Yamashita and Fukui 1983
Saccharomycopsis fibuligera	*glucoamylase	Yamashita et al 1985
	*α-amylase	Yamashita et al 1985
Kluyveromyces lactis	*β-galactosidase	Dickson 1980
Kluyveromyces fragilis	*β-glucosidase	Raynal et al 1984
Rhizopus	*glucoamylase	Ashikari et al 1985
Aspergillus niger 1984	*β-glucosidase	Penttila et al
Aspergillus nidulans	*ornithine carba-moyl transferase	Berse et al 1983
Aspergillus awamori	glucoamylase	Innis et al 1985
Dictyostelium discoideum	*DIRS element	Cappelo et al 1984
	*dihydroorotate dehydrogenase	Jacquet et al 1982
	*orotate pyro-phosphorylase	Boy-Marcotte et al 1984
Triticum aestivum (wheat)	α-amylase	Rothstein et al 1984
	α-gliadin	Neill et al 1987
Phaseolus vulgaris (French bean)	*phaseolin	Harris-Cramer et al 1985
	phytohemagglutinin	Tague and Crispeels 1987
Thaumatococcus daniellii (West African shrub)	thaumatin	Edens et al 1984
Zea mays (maize)	*zein	Langridge et al 1984
Glycine max (soybean)	*leghaemoglobin	Jensen et al 1986
Drosophila melanogaster (fruit fly)	*ADE8	Henikoff et al 1981
	*alcohol	Watts et al 1983

	dehydrogenase	
Gallus domesticus (chicken)	ovalbumin	Mercereau et al 1980
	lysozyme	Oberto and Davison 1985
Bos prigmigenius (cow)	chymosin	Mellor et al 1983
	prochymosin	Goff et al 1984
	interleukin 2	Price et al 1987
pig (unspecified)	phospholipase A_2	van den Berg et al 1987
Mus musculus (mouse)	α-amylase	Filho et al 1986
	dihydrofolate-reductase	Zhu et al 1986
	GM-CSF	Price et al 1987
	interleukin 2	Miyajima et al 1985
Rat (unspecified)	cytochrome c	Scarpulla and Nye 1986
	carboxypeptidase A	Gardell et al 1985
Rabbit (unspecifed)	*β-globin	Beggs et al 1980
monkey (unspecified)	metallothionein	Thiele et al 1986
Homo sapiens (human)	proinsulin	Stepien et al 1983
	α1-antitrypsin	Rosenberg et al 1984
	interferon	Hitzeman et al 1981
	epidermal growth factor	Urdea et al 1983
	immunoglobulins	Wood et al 1985
	tissue plasminogen activator	Lemontt et al 1985
	trypsin inhibitor	Izumoto et al 1987
	GM-CSF	Miyajima et al 1986
	antithrombin III	Broker et al 1987
	gastric lipase	Bodmer et al 1987
	α-amylase	Nakamura et al 1986
	lysozyme	Hayakawa et al 1987
	superoxide dismutase	Hallewell et al 1987
Plasmodium knowlesi	surface antigen	Sharma et al 1985
Hepatitis B virus	surface antigen	Valenzuela et al 1982
Influenza virus	hemagglutinin	Jabbar et al 1985
Herpes simplex virus	*thymidine kinase	Kiss et al 1982
Semliki forest virus	E1 envelope protein	Keranen 1987

Epstein Barr virus	major envelope glycoprotein	Schultz et al 1987
Respiratory syncytial virus	G protein	Ding et al 1987
Adenovirus	Ela protein	Handa et al 1987
Human immunodeficiency virus	gag/pol	Kramer et al 1986

* expressed using natural promoter sequences

achieved using genomic DNA segments from other fungi such as Aspergillus (Penttila *et al* 1984, Innis *et al* 1985) and Neurospora (Vapnek and Case 1982). No expression of herpes simplex virus thymidine kinase gene could be detected when it was introduced into yeast cells (Kiss *et al* 1982) and the rabbit β-globin gene was shown to be transcribed in yeast but incorrect processing and premature termination of the precursor mRNA rendered the gene non-functional (Beggs *et al* 1980). The reasons for some of these problems have become apparent as our understanding of yeast gene expression has expanded, in particular with the finding that yeast is incapable of efficiently splicing the precursor mRNA of interrupted genes from other eukaryotes, (Watts *et al* 1983, Langford *et al* 1983) including other fungi (Innis *et al* 1985). Taking account of this knowledge, recent studies have been more successful. Langridge *et al* (1984) describe the accurate and efficient transcription in yeast of two uninterrupted seed storage protein genes from maize. cDNA sequences encoding the major seed storage glycoprotein of the French bean (Phaseolus) have been expressed in yeast under the control of a Phaseolus promoter (Harris-Cramer *et al* 1985). A stress-inducible promoter from Dictyostelium has also been shown to work in yeast, where its transcription is induced 10-fold by heat shock (Cappello *et al* 1984), and a soybean leghaemogoblin gene promoter has been analysed in yeast fused to the E.coli coding sequence for chloramphenicol acetyltransferase (Jensen *et al* 1986). The soybean promoter not only works efficiently in yeast but expression is regulated by haem at the level of translation, as may be the case in root nodules where translational control has been observed (Govers *et al* 1985). These findings lend support to the suggestion that yeast may be a suitable heterologous system in which to analyse the expression of some plant genes.

1.12 The 2 micron plasmid and yeast cloning vectors

A brief introduction to the 2 micron plasmid is pertinent as it forms the basis of many yeast cloning vectors, including those used in this study. This circular plasmid is present in most strains of S.cerevisiae and is so called because of its contour length of about 2μm. Present at 60 to 100 copies per cell, the plasmid comprises about 3% of the total yeast cellular DNA. It has a chromatin-like structure and replicates under nuclear control during the S phase of the cell cycle (Petes and Williamson 1975, Livingston and Kupfer 1977, Zakian *et al* 1979). A prominent structural feature of the molecule is the existence of two inverted repeats, 599bp in length, which separate two unique regions (Hartley and Donelson 1980). Site specific recombination takes place between these inverted repeat sequences, resulting in the formation of plasmid multimers as well as two interconvertible monomeric forms (see Figure 1.1)

While apparently cryptic for function, the plasmid has been extensively characterised at the molecular level. DNA sequence analysis has identified four open reading frames A to D (Hartley and Donelson 1980), each of which is transcribed into polyadenylated mRNA (Broach *et al* 1979, Sutton and Broach 1985). The A gene, also designated FLP, has been shown to encode a trans-acting product involved in the site specific recombination process (Broach and Hicks 1980) and is essential for plasmid copy number amplification (Volkert and Broach 1986). The B and C gene products are required for stable propagation of the plasmid, their primary role being to ensure efficient plasmid partitioning at cell division (Cashmore *et al* 1985). The D gene appears to be a fine tuning mechanism for regulating plasmid copy number, and acts by modulating the level of FLP gene expression (Murray *et al* 1987). Two
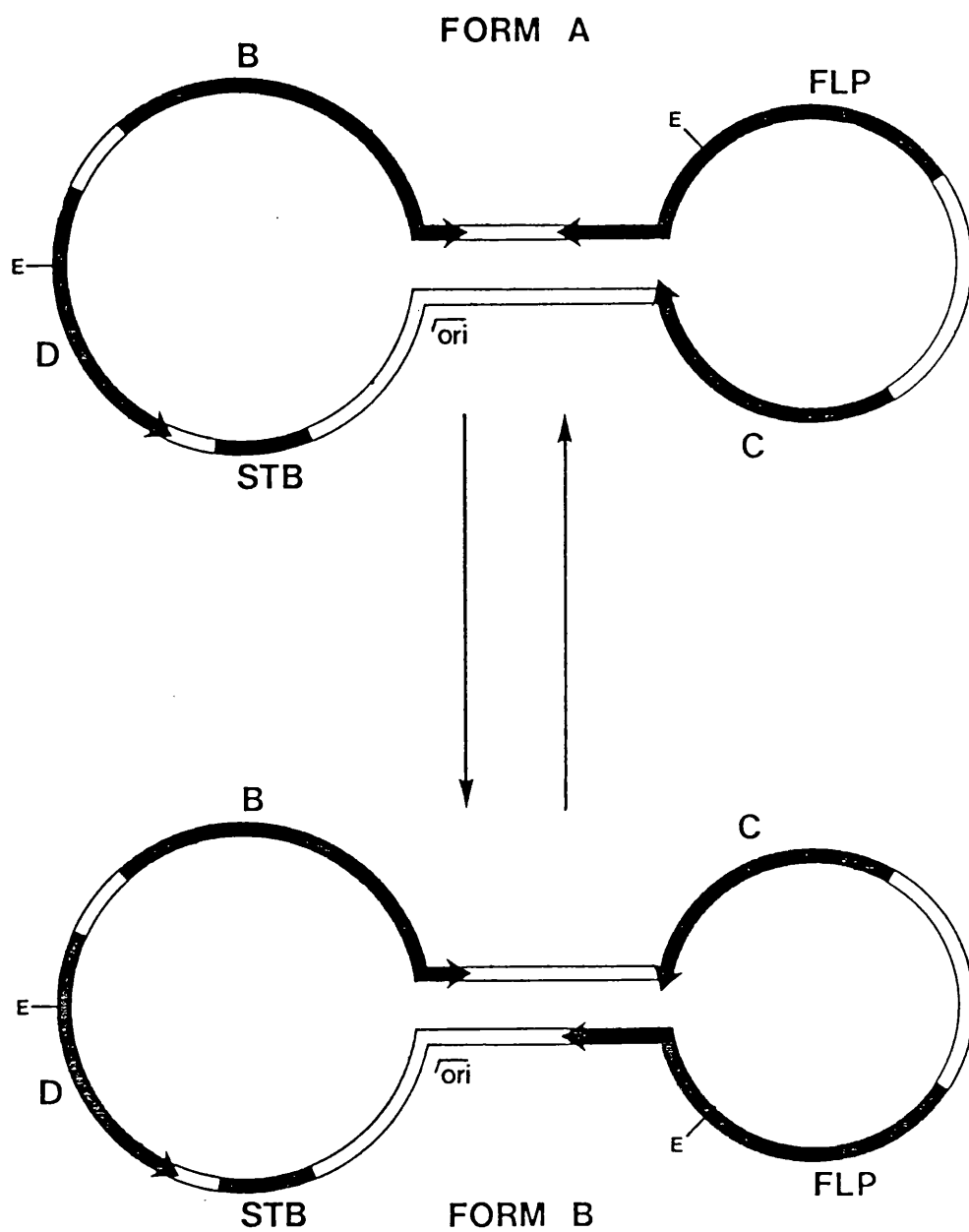
Figure 1.1. Diagramatic representation of the 2 micron plasmid A and B

forms with the inverted repeat sequences aligned. Open reading frames

FLP,B,C,D and the cis-acting STB locus are indicated. E = EcoRl

restriction site.

cis-acting loci required for stable plasmid maintenance have also been identified. These are the replication origin, defined by Broach *et al* (1983), and a set of direct repeats of a 62bp sequence collectively called the STB locus, with which the B and C gene products interact (Jayaram *et al* 1983).

The extensive characterisation, stable propagation and multi-copy status of the 2 micron pasmid make it a natural choice for the basis of a cloning vector. Indeed the first successful non-integrative yeast transformation was performed using a hybrid plasmid based on the 2 micron circle (Beggs 1978). Alternative types of vector have since been designed which permit chromosomal integration (YIp plasmids), autonomous replication (YRp plasmids), which are linear and bear telomeres (YLp plasmids), or which carry a yeast centromere and behave like mini-chromosomes (YCp plasmids). All the plasmids are yeast-<u>E.coli</u> 'shuttle vectors', containing DNA sequences for replication and selection in bacteria as well as yeast, so that the plasmids can be modified and amplified in <u>E.coli</u>. The following paragraphs will provide a brief description of the various classes of yeast transforming vector.

Yeast integrating (YIp) plasmids, first described by Hinnen *et al* (1978) contain a section of yeast chromosomal DNA and become incorporated into the yeast nuclear genome by homologous recombination. The plasmids are not capable of replicating in yeast and transforming efficiency is low, about 1-10 transformants per µg of DNA (Struhl 1983). Although very stable, the transforming event results in a gene duplication and the integration of bacterial DNA sequences, which may not be desirable.

Yeast episomal (YEp) plasmids contain a section of 2 micron DNA and are dependent on trans-acting factors specified by the native 2 micron circle for their efficient maintenance. The YEp plasmids described in this project contain the small EcoRl fragment of the B form (Figure 1.1) which carries the replication origin and STB locus, but not regions A to D. Plasmid copy number is influenced by a variety of factors including plasmid size, host strain and the selectable marker used (Erhart and Hollenberg 1983, Gerbaud and Guerineau 1980) but in general is between 10 and 50 copies per cell. The transformation frequency using YEp plasmids is in excess of 10^3 transformants per μg of DNA, and the plasmids are reasonably stable in strains containing endogenous 2 micron plasmid, hybrid plasmid-free cells being generated by uneven segregation at a rate of about 1% per generation in non-selective medium.

Yeast replicating (YRp) plasmids are able to replicate in yeast by virtue of an autonomously replicating sequence (ars) derived from the yeast nuclear genome (Struhl *et al* 1979). Like YEp plasmids they are present in multiple copies and transform efficiently, but are exceedingly unstable because of their deficient partitioning at cell division. Stable transformants are occasionally generated by integration into a host chromosome in a manner identical to that of Ylp vectors (Stinchcomb *et al* 1979).

Yeast centromere (YCp) plasmids carry both an ars and a functional centromere which confers on the plasmid some of the properties of a chromosome (Clarke and Carbon 1980, Hsiao and Carbon 1981). While not quite as stable as an authentic yeast chromosome, these are mitotically the most stable of all yeast vectors. They are inherited in a mendelian fashion, and are present at about one copy per cell.

Yeast linear (YLp) plasmids contain an ars and homologous or heterologous sequences which act as telomeres, preventing circularisation of the plasmid and protecting the DNA ends from host cell exonuclease action (Szostak and Blackburn 1982, Dani and Zakian 1983). Without a centromere they are present in multiple copies and behave like YRp plasmids. The inclusion of a centromere increases their mitotic and meiotic stability and reduces the copy number to about one per cell.

Other yeast plasmids have been specifically constructed for the selection or analysis of promoter sequences, for the expression of yeast or foreign genes, and for the secretion of foreign polypeptides (see Parent *et al* 1985). Clearly, the diversity of vectors available makes yeast a highly versatile system for the cloning and expression of foreign genes.

Cauliflower mosaic virus (CaMV) is an icosahedral virus with a double stranded DNA genome. Over recent years CaMV has been extensively studied as a potential plant gene cloning vector and consequently is one of the best characterised plant viruses at the molecular level (reviewed by Hohn *et al* 1982). The virus has a restricted host range, being mainly limited to species of the Cruciferae family although some strains are able to infect Solanaceae as well (Shepherd 1979), and is transmitted in nature by aphids. Usefully, for laboratory purposes, plants can also be infected mechanically by rubbing naked viral DNA onto the leaf surface. Once infection is established, the virus spreads systemically through the plant, causing chlorotic lesions which give a mosaic appearance to the leaves as well as retarding leaf growth.

Light and electron microscope studies reveal the existence of cytoplasmic inclusion bodies in CaMV-infected cells (Martelli and Castellano 1971). These are dense, proteinaceous structures of irregular size and shape, in which most of the virus particles (about 10^5 per cell) are located. Several lines of evidence point to the inclusion body as being the major site of viral DNA synthesis and virion assembly. Early electron micrograph studies showed the inclusion bodies to be associated with ribosomes (Martelli and Castellano 1971) and in an *in vitro* CaMV DNA replication system derived from infected leaves, inclusion bodies rather than nuclei were the principal site of active DNA synthesis (Modjtahedi *et al* 1984). Recent studies also suggest that viral DNA synthesis occurs concomitant with (or even subsequent to) virion assembly within the inclusion body (Marsh and Guilfoyle 1987). Current theories pertaining to the

intriguing replication cycle of CaMV will be presented below.

1.21 Genome Structure

CaMV has a circular genome of double-stranded DNA, 8kb in length. An unusual feature of the DNA, unique to caulimoviruses, is that it is not covalently closed but contains 2 or 3 (depending on strain) single-stranded discontinuities or 'gaps' at specific locations on each strand. These are not authentic nicks but regions of strand overlap resulting in the formation of triple stranded sections extending over 5 to 40nt (Franck *et al* 1980), see Figure 1.2. The 5' end of each discontinuity occurs at a fixed position and terminates with a ribonucleotide (Guilley *et al* 1983) whereas the 3' ends are heterogeneous (Richards *et al* 1981). The positions of the discontinuities are conserved in virtually all CaMV strains and it is likely that they are generated in the course of DNA replication. In this context, the ribonucleotide at the 5' extremity of each discontinuous strand may represent the vestige of an RNA primer that has not been completely excised. RNA molecules of about 9 residues are known to serve as primers for DNA synthesis in eukaryotes and are found covalently linked to the 5' end of nascent DNA chains (Okazaki fragments). Additional evidence that the gaps are of biological significance comes from the finding that CaMV DNA from which the gaps have been eliminated (by amplification in E.coli), is fully infectious when inoculated onto plants, however the virion DNA recovered has invariably reverted to the interrupted form (Howell *et al* 1980).

Most of the viral DNA in infected cells is encapsidated, but small amounts of free viral DNA have been detected, comprising the usual gapped form, a linear form and a supercoiled form in which the gaps

11
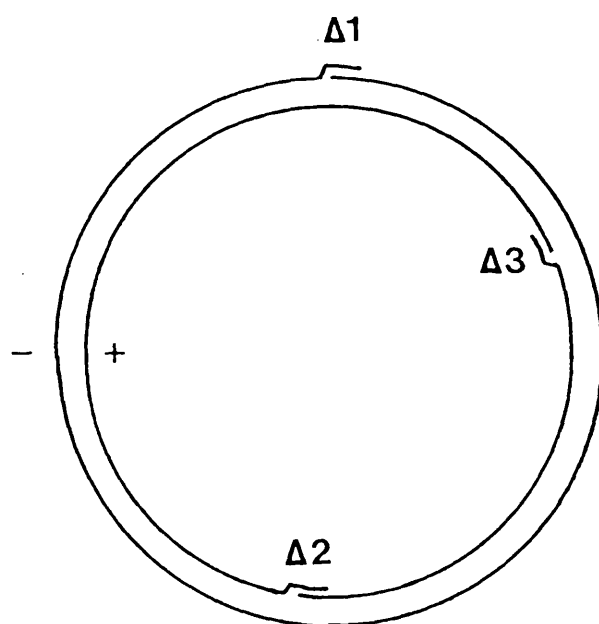
```
                      A T T T T T T T  rR-5'
                    A
  T C G G G G G C  G  OHA T T T T T T A A C C A T A G T C T C G G T A C
Δ1 A G C C C C G C T T A A A A A A T T G G T A T C A G A G C C A T G
                        OH A C C A U A G U C U C G G U
```

```
  T C C C C T C C T C C A A T A G T C T T C T T T T T G A G A A A T A
Δ2 A G G G G A G G A G G T T A T C A G A A G A A A A A  COH   C T T T A T
              5'-rR G G T T A T C A G A A G A A A A A C   T  C
```

```
  C A C C C C C C C A A C T A A T G A G C T C G G T T G A T T T C C T
Δ3 G T G G G G G G G T T G A T T A C T C G A  GOH   C A A C T A A A G G A
              5'-rR G G T T G A T T A C T C G A G   C  C
```

Figure 1.2. Genomic location and nucleotide sequence of the single stranded discontinuities in CaMV DNA.   + and − denote the plus and minus DNA strand.   Underscored sequence is the 3' terminus of plant tRNA[met].
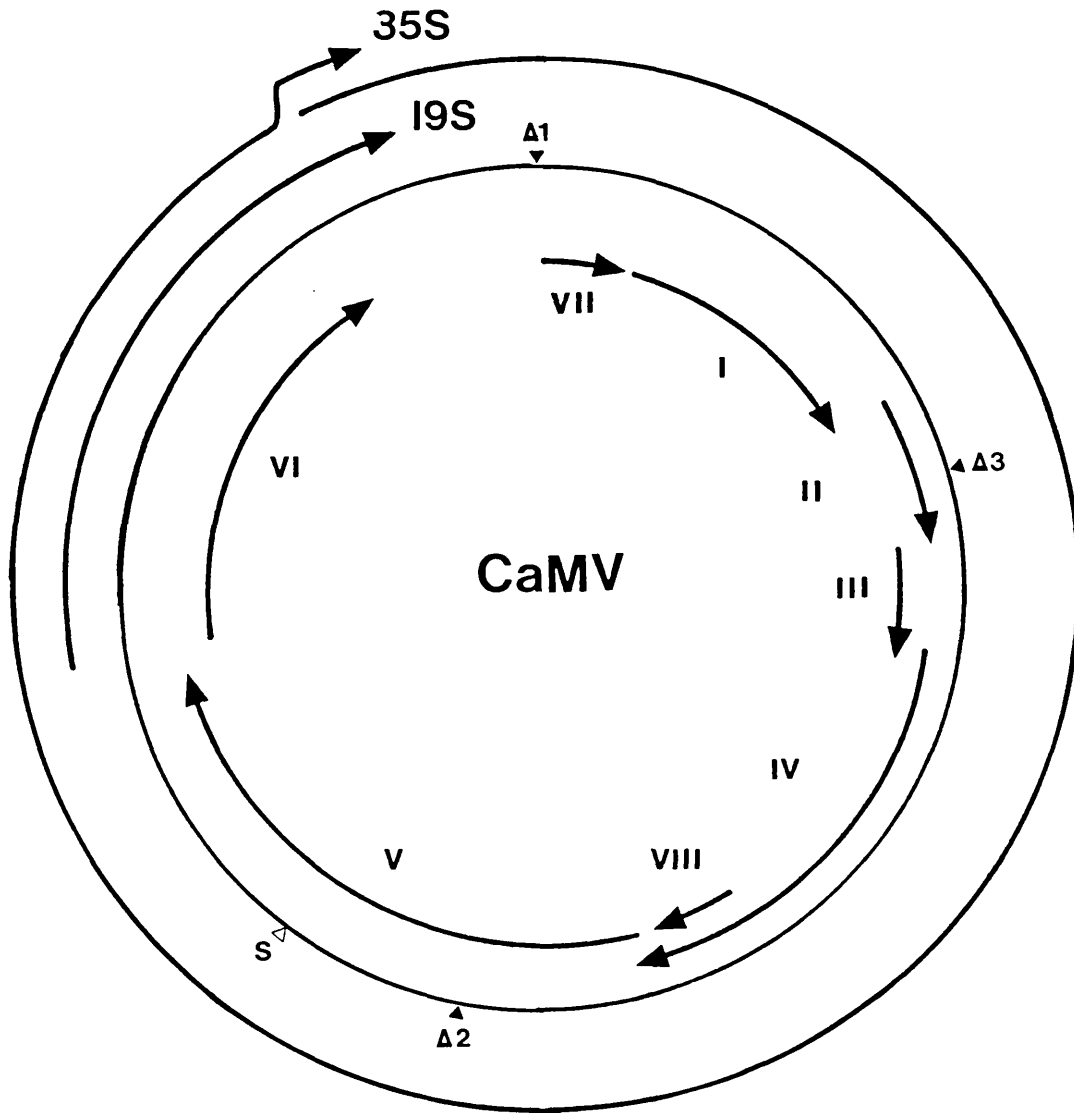
Figure 1.3. Diagram of CaMV showing the positions of open reading frames I to VIII (inner arrows), the two major RNA transcripts (outer arrows) and single strand gaps Δ1, Δ2 and Δ3. S = Sall restriction site.

have been covalently sealed (Menissier *et al* 1982). The supercoiled form, termed the CaMV mini-chromosome, is associated with the nuclei of infected cells and has been shown to possess a nucleosome-like structure based on its sensitivity to micrococcal nuclease digestion (Olszewski *et al* 1982). A subsequent electron microscope study by Menissier *et al* (1983) supports this, revealing the purified supercoiled molecule to have a beaded appearance typical of eukaryotic chromatin. These observations, together with the demonstration that it is a template for RNA polymerase II *in vitro* (Olszewski *et al* 1982), suggest that the mini-chromosome is the transcriptionally active form of CaMV DNA in infected cells.

1.22 Viral gene products

Nucleotide sequencing of CaMV DNA (Franck *et al* 1980) has revealed 6 tightly-packed long open reading frames (genes I to VI) and two short ones (VII and VIII), the significance of which is uncertain. These open reading frames occur exclusively on the plus strand and occupy 85% of the viral genome (Figure 1.3). With the exception of region VIII which lies wholly within gene IV, there is little or no overlap between adjacent genes, which alternate in reading frame and appear to be uninterrupted. Insertion mutagenesis indicates that the products of genes I, III, IV, V and VI are required for viral replication whereas those of genes II and VII are dispensible (Howell *et al* 1981, Daubert *et al* 1983). In accordance with this, characterisation of a naturally occuring deletion mutant of CaMV has revealed a 421bp deletion within gene II which does not render the virus inviable (Howarth *et al* 1981).

The protein products of all six genes have been detected in CaMV-infected plant cells, although functions have so far been assigned

to genes II, IV and V only. Using antiserum raised against a recombinant gene I product synthesized in E.coli, the native CaMV protein has been identified within the inclusion body but does not copurify with virions (Martinez-Izquierdo *et al* 1987). The gene III product has likewise been detected in infected plant tissues using immunological techniques (Xiong *et al* 1984). Data from a number of sources indicate that gene II encodes an 18,000 molecular weight product which is present in the inclusion body and is essential for insect transmission. These include studies of gene II deletion mutants (Armour *et al* 1983), the analysis of hybrid genomes constructed between aphid transmissible and non-transmissible CaMV strains (Woolston *et al* 1983) and insertion mutagenesis (Daubert *et al* 1983). The 57,000 dalton nucleocapsid protein precursor has been assigned to gene IV on the basis of its immunological detection when defined subgenomic fragments of CaMV were cloned and expressed in E.coli (Daubert *et al* 1982). In plant cells the primary translation product is subsequently cleaved to a mature 43,000 dalton polypetide (Hahn and Shepherd 1982) which is both phosphorylated and glycosylated. Cell-free translation studies have revealed that gene VI encodes a 62-66000 dalton protein identical to the major structural component of viral inclusion bodies (Odell and Howell 1980, Covey and Hull 1981). The N-terminal domain of this protein has more recently been shown to function as a host range determinant (Schoelz *et al* 1986). Finally, a reverse-transcriptase activity specific to CaMV-infected plant tissue (Volovitch *et al* 1984) has been assigned to gene V. This is the largest open reading frame and was postulated to encode a reverse transcriptase on the basis of amino acid sequence homology with retroviral reverse transcriptase (Toh *et al* 1983). Expression of the gene in yeast has since confirmed this suggestion (Takatsuji *et al* 1986). Interestingly, all attempts to to express gene V *in vitro* or in bacterial systems have generated a

protein without biological activity, implying that some post-translational modification of the protein is required. To date, no information is available regarding the roles of genes I and III, although both are absolutely required for viral infectivity. Conceivably they may be minor structural components of the inclusion body, function as part of the replication complex, or be part of a regulatory circuit controlling the expression or viral or cellular genes.

1.23 Transcription

Two major RNA transcripts of the CaMV genome have been identified, both synthesised from the minus DNA strand by RNA polymerase II (Guilfoyle 1980) (Figure 1.3). One is a 19S RNA, 1850nt in length, which hybridizes to gene VI (Odell and Howell 1980) and directs the synthesis of the major inclusion body protein *in vitro* (Xiong *et al* 1982). The RNA is capped and polyadenylated (Covey and Hull 1981) and clearly functions as a mRNA. A putative mRNA species corresponding to gene V has also been identified (Plant *et al* 1985).

The second major RNA species, also polyadenylated, has a sedimentation velocity of 35S and is a full length copy of the genome with a direct terminal repeat of 180nt (Covey *et al* 1981) (Figure 1.3). The role of this RNA has been the subject of much speculation. Since individual mRNAs have not been detected for genes I to IV, the full length transcript could function as a polycistronic mRNA and some evidence has been put forward in support of this theory (Dixon and Hohn 1984). Attempts to translate the 35S RNA *in vitro* have consistently been unsuccessful however (Guilley *et al* 1982), which perhaps argues against this role. A further possibility is that the 35S RNA is spliced to

create indiviual mRNAs for genes I to IV but neither splicing intermediates nor processed mRNAs for these genes have yet been identified.

While it is still unclear whether translation of genes I to IV is dependent upon the the 35S RNA, an alternative function for this molecule as a replicative intermediate has been proposed. According to the current model for CaMV DNA replication (see next section), the 35S transcript is a template for minus-strand DNA synthesis by virally-encoded reverse transcriptase.

The existence of a terminal repeat on the 35S RNA raises the question of how the RNA polymerase overrides a transcription termination signal on its first circuit of the genome. The problem has not been addressed experimentally, but secondary structure of the nascent RNA chain may be a factor influencing whether or not termination occurs. The other potential difficulty in producing a full length copy of the genome, that of traversing the minus-strand gap Δ1, is presumably circumvented by employing the covalently closed CaMV mini-chromosome as a template for transcription.

Two much less abundant RNA species, both polyadenylated, have been described: an 8S RNA of about 600nt which initiates at the same place as 35S RNA but terminates at Δ1, and a minor 35S RNA which begins and ends at Δ1 (Guilley *et al* 1982). The significance of these transcripts is uncertain and it is possible that they arise from aberrant transcription of "gapped" CaMV DNA prior to its conversion to a closed, supercoiled form.

The 5' and 3' ends these RNAs have been accurately determined by primer

extension and/or S1 nuclease analysis (Covey *et al* 1981, Guilley *et al* 1982). Such studies show that the 19S and 35S transcripts are exactly 3' co-terminal but several lines of evidence argue against the 19S RNA being a processed form of the 35S. Ultraviolet mapping of the CaMV transcripts by Howell (1981) first indicated that the 19S RNA is synthesised independently. More recently, sequences upstream of the gene VI coding region have been shown to contain an RNA polymerase II promoter which is active in an *in vitro* transcription system derived from Hela cells (Guilley *et al* 1982) and has been used to direct the expression of a bacterial neomycin phosphotransferase (NPT) II gene in transformed tobacco cells (Paszkowski *et al* 1984).

DNA sequences upstream of the 19S and 35S transcripts both contain features characteristic of a eukaryotic RNA polymerase II promoter, in particular an AT rich region (TATAAA) about 30bp from the start of transcription. The 35S promoter has been partially characterised by deletion analysis (Odell *et al* 1985). This revealed that the AT-rich region (TATA box) alone gives a basal level of transcription (roughly 5%) but additional promoter element(s) within the region -46 to -168bp are required for full activity. This upstream region contains a CCAAT box sequence, an inverted repeat region and a sequence resembling the consensus core for enhancers in animal systems (see section 1.32) but it remains to be demonstrated whether any of these features is relevant to 35S RNA transcription. A mutational analysis of the 19S promoter region has not yet been published.

1.24 Replication

It was the observation that CaMV has various features in common with retroviruses and hepatitis B viruses, both of which replicate using

reverse transcriptase (Varmus 1982, Summers and Mason 1982) which prompted speculation that CaMV might replicate by a similar mechansim. The most striking of these similarities include asymmetric RNA synthesis (plus strand only) and the formation of a full length transcript with terminally repeating sequences. A model for CaMV DNA replication that is directly analogous to the reverse transcription of animal retroviruses was put forward independently by Pfeiffer and Hohn (1983), Hull and Covey (1983) and Guilley *et al* (1983). The model, depicted in Figure 1.4, is as follows. 1) The 35S terminally redundant RNA is synthesized in the nucleus from the CaMV mini-chromosome and exported to the cytoplasm. 2) Here it becomes a template for minus-strand DNA synthesis by reverse transcriptase, primed by host cell tRNA^{met}. A 14 nucleotide sequence of perfect homology with the 3' end of plant tRNA^{met} is located adjacent to the minus strand gap Δ1 (Figure 1.2) some 600nt downstream of the 35S RNA 5' terminus. 3) Reverse transcription proceeds to the 5'end of the 35S RNA, producing a small minus strand DNA fragment covalently linked to the tRNA primer. Such a molecule has been identified in CaMV-infected cells (Turner and Covey 1984). 4) By a process of template strand switching (made possible because of the terminal repeat) the RNA template is circularised and reverse transcription continues until the minus strand is complete. 5) The completed minus strand is the template for plus-strand synthesis, which is primed by G-rich sequences located near the plus strand gaps Δ2 and Δ3 in virion DNA (Figure 1.2). These gaps presumably result from limited displacement synthesis by reverse transcriptase. An excellent summary of the research which has largely substantiated this model has been presented by Hohn *et al* (1985).
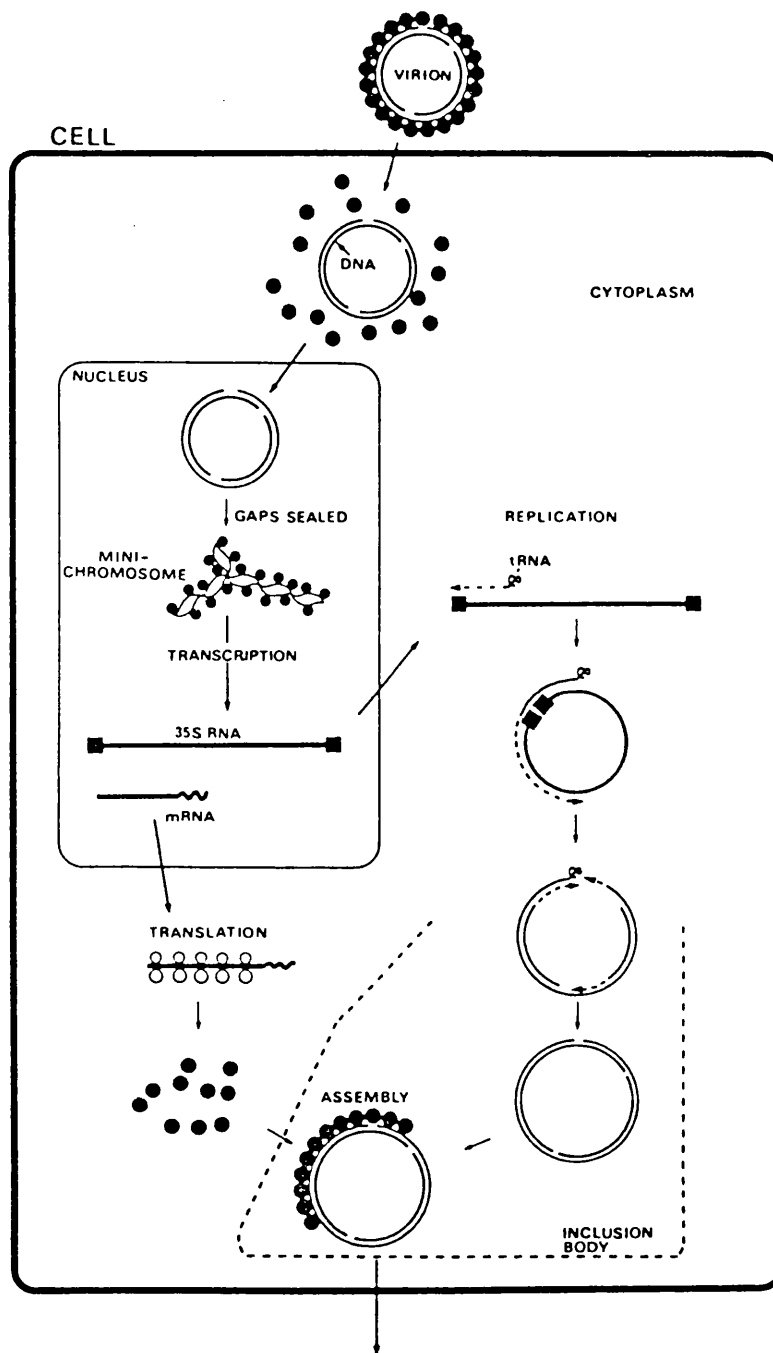
Figure 1.4.   The replication cycle of CaMV (from Grierson and Covey 1984).

1.25 CaMV as a research tool

At first sight CaMV looked like a promising vector for introducing foreign DNA into plant cells, however deletion and insertion mutagenesis experiments have revealed very few non-essential regions of the genome (Howell *et al* 1981, Daubert *et al* 1983). Furthermore those sites which are non-essential will not accommodate additional DNA sequences larger than 200-300bp (Gronenborn *et al* 1981, Daubert *et al* 1983), possibly reflecting DNA packaging constraints during virion assembly. Attempts to develop a helper virus system for the propagation of deleted viral genomes harbouring foreign DNA have proved problematical owing to a high rate of recombination between the two defective genomes, producing a wild-type virus (Walden and Howell 1982). Consequently, the development of CaMV as a cloning vector has largely been abandoned. Current interest in the virus stems firstly from its fascinating replication strategy and relationship with other members of the 'retrogene' family which includes the animal retroviruses, Ty transposable elements of yeast and the Copia family of Drosophila. Secondly, the two major promoters of CaMV (the 19S and 35S promoters) are widely used for the constitutive expression of cloned genes in the nuclei of transgenic plants, and as such merit extensive investigation.

GENE EXPRESSION

The cells of all eukaryotes that have been studied contain three RNA polymerases I, II and III which are required for the synthesis of ribosomal RNA, messenger RNA, and transfer and 5S RNA respectively. The purpose of this section is to review the regulation of protein-coding genes transcribed by RNA polymerase II, with particular reference to transcription initiation. The expression of protein-coding genes is a multi-step process involving transcription of the gene, the addition of a methyl guanine residue (cap) to the 5' end of the primary transcript and a poly(rA) tail to the 3' end, the removal of intervening sequences, transport of the mature mRNA into the cytoplasm and finally translation. The transcription, RNA splicing and translation processes in yeast and higher eukaryotes are essentially very similar, however areas of divergence which may have a bearing on the efficiency with which foreign genes are expressed in yeast will be discussed.

1.31   Chromatin Structure and Gene Expression

The fact that not all the genes in a cell are actively transcribed at the same time illustrates the importance of transcriptional control in eukaryotes. In any given differentiated cell type of higher eukaryotes only about 10% of the genes are actively expressed, the remainder being kept in an inactive state. While there are several well documented examples of *trans*-acting transcriptional repressors specific for single genes, it is likely that the dominant mechanism for transcriptional repression is more general, probably involving higher order chromatin structure. Repression can be very efficient since transcription of the rat growth hormone gene is repressed by a factor of at least 10^8 over

the fully active state (Ivarie *et al* 1983).

There is much evidence now that transcriptional derepression occurs in two distinct stages. The first of these involves a shift in chromatin organisation from a transcriptionally inaccessible to an accessible state. The second stage is the gene-specific induction of transcription brought about by the interaction of regulatory and transcription factors with the now accessible promoter.

Transcriptionally active, or potentially active, chromatin can be distinguished from the bulk inactive chromatin owing to its heightened susceptibility to a variety of endonucleases such as DNAase I. This susceptibility does not correlate with the absence of histone proteins as originally thought, since both electron microscope studies and carefully controlled experiments using micococcal nuclease (reviewed by Reeves 1984) indicate that most, if not all, actively transcribed genes are packaged into nucleosomes. Instead, nuclease sensitivity appears to reflect an altered chromatin conformation, in which the DNA wrapped round the outside of the histones is more accessible to external molecules such as nucleases and polymerases.

The high degree of DNA compaction in eukaryotic nuclei is achieved by several rounds of condensation. At the lowest level, DNA is coiled twice round each nucleosome particle. The string of nucleosomes is wound into a spiral or 'solenoid' visible in electron micrographs as a 30nm fibre, which is itself folded into loops or domains. Loops contain 35 to 100kbp of DNA and are anchored by specific non-histone proteins to a supporting nuclear structure termed the matrix or scaffold (Bodnar *et al* 1983, Lewis and Laemmli 1982).

There is mounting evidence that the decondensation of the 30nm solenoid into an extended chain of nucleosomes is associated with the establishment of active chromatin (Kimura *et al* 1983). In agreement with this, various studies have implicated the non-core histone H1, which is essential for the formation of the solenoid, as a generalised repressor of gene activity (Weintraub 1984, Schlissel and Brown 1984). It appears, therefore, that in higher eukaryotes up to 90% of the genome is maintained in a transcriptionally repressed state as a result of being organised into an H1-dependent 30nm solenoid. Activation (often of entire domains) is associated with disruption of this structure, and while the mechanism(s) by which this is achieved are far from clear, roles have been suggested for the removal or structural modification of histone H1 (Caron and Thomas 1981), for high mobility group (HMG) proteins 14 and 17 which bind to nucleosomes (Weisbrod 1982) and for the ubiquitination or acetylation of core histones (Levinger and Varshavsky 1982, Reeves 1984). Factors such as these may affect DNA-histone interactions or else destabilise higher-order chromatin structures.

Interestingly, the chromatin structure of yeast differs from that of higher eukaryotes in that it does not appear to contain transcriptionally- repressed regions. The DNA is uniformly sensitive to DNAase I digestion (Lohr and Hereford 1979), perhaps due to the unusually high degree of histone acetylation reported by Davie *et al* (1981) which would inhibit the formation of a 30nm solenoid structure. This is entirely consistent with the small genome size of yeast in comparison with many higher eukaryotes, the fact that yeast is unicellular (so each cell must carry out all essential metabolic functions) and the observation that up to 40% of the yeast genome is transcribed during vegetative growth (Hereford and Rosbash 1977).

Nuclease digestion studies of chromatin from many species have also revealed the existence of short stretches of DNA up to 400bp in length that are hypersensitive to DNAase I (reviewed by Reeves 1984, Igo-Kemenes *et al* 1982). DNAase hypersensitive sites are found in the genomes of all eukaroytes including yeast, and are more sensitive by an order of magnitude than active or potentially active chromatin. The sites occur in a cell- and tissue- specific manner near the 5' and 3' ends of most actively expressed genes, and there is evidence to suggest that in contrast to protein-coding regions these regions are entirely free of nucleosomes. Hypersensitive sites may reflect the presence of DNA-binding proteins as they often coincide with mutagenically-defined *cis*-acting regulatory sequences (see for example Lohr and Hopper 1985).

DNA supercoiling and gene expression

Most of the DNA in eukaryotic chromosomes, despite being negatively supercoiled around nucleosomes, is not under superhelical tension (Sinden *et al* 1980). Recently, however, it has been found that most DNAase I hypersensitive regions contain specific sites that are sensitive to enzymes (eg S1 nuclease) and chemicals (eg bromoacetaldehyde) which indicate the presence of non B-form DNA (Larsen and Weintraub 1982, Kohwi-Shigematsu *et al* 1983). Non-B DNA forms such as Z (left-handed) DNA, single stranded DNA and cruciform structures are energetically favoured in supercoiled but not relaxed DNA *in vitro*, suggesting that DNA within the hypersensitive regions is indeed under torsional strain, perhaps due to the localised dissociation of a nucleosome and subsequent release of of superhelical turns.

There is circumstantial evidence to suggest that transcription *in vivo*

requires torsional stress within the template DNA. For example the transcription of a supercoiled DNA template micro-injected into frog oocytes ceases after relaxation of the template with novobiocin, an inhibitor of topoisomerase II (Ryoji and Worcel 1984). Other evidence that DNA topology is related to gene activity *in vivo* has come from studies in yeast where plasmids carrying transcriptionally repressible genes – acid phosphatase gene PHO5 and stored mating-type gene HMLα – were analysed (Bergman *et al* 1986, Abraham *et al* 1983). In both cases, when transcription of the gene was repressed (by the presence of inorganic phosphate in the case of PHO5 and by growth in a SIR$^+$ strain of yeast in the case of HMLα), the plasmid was found to have a pronounced increase in the number of superhelical twists. Furthermore, the sequences which mediate the topological change were mapped to a region coinciding with an upstream promoter element in PHO5 (Bergman *et al* 1986).

There are several ways in which torsional strain in DNA might promote transcription. Firstly, as indicated above, non-B DNA forms are energetically favoured in DNA which is under superhelical tension, and it might be that postive regulators of transcription have a higher affinity for their target sequences when they are in a non-B conformation. Specific Z-DNA binding proteins have been identified in a variety of eukaryotic species including Drosophila, wheat and man (Kolata 1983) and might represent such a class of regulators. Secondly, factors (such as H1) involved in repression may bind weakly or in a different manner to supercoiled DNA. Thirdly, torsional stress may be needed for long-range transmission of signals along the chromosome, for example to propagate localised DNA unwinding from an upstream promoter element to the mRNA start site.

1.32 Organisation of RNA polymerase II promoters

The major point of control for most eukaryotic genes is at the level of RNA synthesis. In transcriptionally active chromatin, the regulation of individual genes is achieved by means of *trans*-acting factors which bind to specific DNA sequences flanking the gene, thereby influencing the frequency of transcription initiation. These factors may be ubiquitous, tissue or stage specific and positively or negatively acting. Using such techniques as deletion and site-directed mutagenesis, gel retardation studies and DNA footprinting, these *cis*-acting regulatory sequences have been the subject of intense investigation in recent years, from which a fairly uniform pattern of promoter structure has emerged. Most regulated protein-coding genes are found to have a bi- or tripartite promoter, comprising a TATA box preceded by an upstream control region and in some cases an enhancer. The properties and significance of each of these elements will be discussed below. A different structural class of promoter (belonging to constitutively expressed mammalian 'housekeeping' genes) which has no TATA box and no requirement for upstream promoter elements has recently been reviewed by Dynan (1986) and will not be discussed here.

The TATA box

An AT rich sequence with the consensus TATA(T/A)A(T/A), referred to here as the TATA box, is an integral component of most eukaryotic promoters transcribed by RNA polymerase II. In higher eukaryotes the TATA box is located at a fixed distance of 25 to 30bp upstream of the mRNA start (Breathneach and Chambon 1981) and mutations within the sequence result in transcription initiation at novel sites, reduced mRNA levels or both (Grosschedl and Birnstiel 1980, Osborne *et al*

1982). From this it has been deduced that a major role of the TATA sequence is to accurately position the polymerase so that the initiation event occurs precisely.

Unlike prokaryotic RNA polymerase, purified RNA polymerase II lacks any inherent capacity to recognize a promoter and initiate transcription in vitro, implying that cellular transcription factors play a crucial role in the initiation event. Proteins which specifically bind to a region encompassing the TATA box have been purified from Drosophila (Parker and Topol 1984a) and human (Sawadogo and Roeder 1985) cell extracts. Clearly these represent strong candidates for such a transcription factor.

The TATA box is an essential component of most yeast promoters, although the role of this element in yeast appears to differ slightly from that in higher eukaryotes. This is illustrated by the fact that the distance between the TATA box and the transcription initiation site is not strictly preserved but varies between 40 and 140 nucleotides (see Hahn et al 1985). Additionally, multiple RNA initiation sites are common. Many yeast genes are preceded by several TATA-like sequences although mutational analyses have generally revealed that only one of these is functional (Nagawa and Fink 1985, Tajima et al 1986, Crabeel et al 1985). An interesting exception is the yeast iso-1-cytochrome oxidase (CYC1) gene which is preceded by at least 3 functional TATA boxes, each of which is required for a specific subset of the 20 or more mRNA starts (Hahn et al 1985, McNeil and Smith 1986).

A DNA footprinting study has revealed the presence of a factor bound at the yeast GAL1 TATA box (Selleck and Majors 1987). This factor may mediate the binding of RNA polymerase but a role in the accurate

determination of the RNA start site would seem unlikely in view of the variable location of the TATA box in yeast. The yeast TATA box does influence the position of transcription initiation in as much as a minimum distance of 40 to 60bp between this sequence and the initiation (cap) site appears to be required (Hahn *et al* 1985, Nagawa and Fink 1985, Healy *et al* 1987). The results of several investigations suggest that sequences surrounding the cap site are an important factor in the selection of that site, and the consensus sequences TC(G/A)A, YYRYY, and YAAR have been noted (Hahn *et al* 1985, Chen and Struhl 1985, Dobson *et al* 1982). A survey of 99 mRNA start sites occuring in 32 yeast genes provides support for the latter sequence, YAAR, as being the preferred site of initiation. The tendency for transcription to begin 40bp or more downstream of a TATA box may interfere with the expression of foreign genes from higher eukaryotes which have a short mRNA leader, since recognition of the usual TATA box may result in transcription initiation within the coding region.

Upstream elements and enhancers

Cis-acting promoter elements distal to the TATA box of protein-coding genes are diverse both in structure and function, and may be required i) for the general stimulation of transcription, ii) for tissue-specific gene expression and iii) for the induction (or repression) of transcription in response to specific environmental agents for example hormones, nutrients, light and temperature. In animal cells these regulatory regions have been classified as either upstream elements or enhancers depending on their location and properties, although the distinction between these types of element has become increasingly blurred as examples are found which combine the properties of both (Bienz and Pelham 1986). As originally defined,

upstream elements are located 40-100bp 5' of the mRNA cap site, whereas enhancers are able to activate transcription over longer distances (several kb), whether upstream or downstream of the cap site, and in a manner independent of their orientation. Plant promoters too contain enhancer-like elements (reviewed by Gelvin 1986, Schell 1987) which stimulate transcription in an orientation-independent manner although none of those so far studied functions as efficiently from a position downstream of the gene (Timko *et al* 1985). The upstream activator sequences (UASs) identified in the 5' flanking regions of all yeast genes so far examined (reviewed by Guarente 1984, Struhl 1987) have considerable positional flexibility and appear to fulfil the function of both the enhancer and upstream elements of higher eukaryotic promoters.

Certain upstream elements, such as the CCAAT and GC homologies described below, have been found in the promoters of many functionally unrelated genes and appear to be rather generalised activation elements. Originally identified as a conserved sequence upstream of virtually all vertebrate globin genes, the CCAAT box (consensus GGT/CCAATCT) has since been found in the promoter of many viral and eukaryotic genes. Mutational studies have shown it to be required for the efficient expression not only of globin genes (Dierks *et al* 1983, Mellon *et al* 1981) but of HSV thymidine kinase gene (reviewed by McKnight and Tijan 1986), chicken α-actin gene (Bergsma *et al* 1986), Xenopus hsp70 gene (Bienz and Pelham 1986) and a gene of prokaryotic origin, nopaline synthase, which is expressed in plant nuclei (Shaw *et al* 1984). A factor which binds specifically to the CCAAT box *in vitro* has recently been purified (Jones *et al* 1987) and appears to represent a ubiquitous nuclear protein.

Likewise, the cellular transcription factor Spl activates a range of viral and mammalian housekeeping gene promoters (McKnight and Tijan 1986, Dynan and Tijan 1985, Dynan 1986). This factor has been purified to homogeneity from Hela cells and DNAasel footprinting studies indicate that the hexanucleotide sequence GGGCGG (GC element) constitutes a consensus core for Spl binding sites (Kadonaga *et al* 1986).

Putative CCAAT boxes have been found upstream of several plant genes, including the 35S promoter of cauliflower mosaic virus (Odell *et al* 1985), but in most cases the functional significance of this homology remains to be assessed. While CCAAT and GC homologies are not observed in yeast promoters, two other 'generalised' promoter elements have been identified. Dobson *et al* (1982) noted a sequence common to the 5' flanking regions of several highly expressed yeast genes including the phosphoglycerate kinase gene PGK, the alcohol dehydrogenase gene ADH1, and the actin gene. The conserved sequence comprises a CT-rich block located 50-100bp upstream of the coding region followed about 10bp downstream by the sequence CAAG, and may be a requirement for high efficiency expression in yeast. Evidence has also been presented that the poly(dA-dT) stretches found in the 5' flanking regions of many yeast genes is a functional promoter element required for constitutive basal-level expression, as distinct from the UAS which confers high level inducible expression (Struhl 1985). These poly(dA-dT) regions may be targets for specific DNA binding proteins or alternatively affect the local chromatin structure. It has been shown that poly(dA-dT) regions inhibit nucleosome assembly *in vitro* (Kunkel and Martinson 1981) and may, therefore, present a permanently accessible stretch of DNA to the transcription machinery *in vivo*.

In addition to these general promoter elements, which are critical for efficient transcription but do not appear to have a regulatory function, a wide variety of gene-specific upstream elements have been identified. These usually confer regulated (ie inducible, tissue or stage-specific) gene expression and are too numerous to describe here but include, for example, i) the heat shock element found upstream of Drosophila heat shock genes, which is responsible for their stress-inducible transcription and binds a protein called the heat shock transcription factor (Parker and Topol 1984b) ii) the UAS located between the divergently transcribed GAL1 and GAL10 genes of yeast, which interacts with the GAL4 gene product to stimulate their transcription in the presence of galactose (West et al 1984) and iii) the light- responsive promoter element of the pea rbcS gene which is required for light-dependent stimulation of mRNA synthesis (Morelli et al 1985) and has recently been shown to interact with a protein factor designated GT-1 (Green et al 1987).

The heat shock element is particularly fascinating since the mechanism of stress-inducible transcription appears to be highly conserved in eukaryotes. Evidence for this came firstly from the observation that the promoters of heat shock genes from species as diverse as Xenopus, Dictyostelium, soybean and yeast contain a sequence similar to the Drosophila heat shock element and the general consensus CNNGAANNTTCNNG has been derived (Pelham 1985). Secondly, the Drosophila heat shock element confers heat-inducible expression upon a linked gene when introduced into both yeast (Wei et al 1986) and plant cells (Spena et al 1985), suggesting that these organisms possess a factor with the same function and specificity as the Drosophila heat shock transcription factor. This suggegstion has now been confirmed in yeast by the purification of a yeast heat shock transcription factor

identical in size and DNA binding properties to that of <u>Drosophila</u> (Wiederrecht *et al* 1987).

The widespread occurrence of upstream regulatory regions permits the co-ordinate expression of functionally related genes, which may be dispersed in the genome, by means of a common *trans*-acting regulator. This is a mechanism widely employed in yeast where, for example, the positive regulatory protein GCN4 (which is synthesized in response to amino acid starvation), binds specifically to a conserved sequence in the promoters of 30 to 50 different genes coding for amino acid biosynthetic enzymes, co-ordinately inducing their transcription (Hope and Struhl 1985). The co-ordinate expression of functionally related or stage-specific genes is achieved in a similar way in higher eukaryotes. The light-responsive element described above, for example, is found in the promoters of all known pea <u>rbcS</u> genes. Furthermore, the upstream region of two unrelated soybean genes (the lectin and Kunitz trypsin inhibitor genes) which are co-expressed during embryogenesis, both contain binding sites for an embryo-specific DNA-binding protein. This interaction may be responsible for the co-ordinate and tightly regulated transcription of these two (and possibly other) seed protein genes (Jofuku *et al* 1987).

Enhancer elements are functionally defined as *cis*-acting DNA sequences which can operate in either orientation and over considerable distances to activate transcription from a linked homologous or heterologous promoter. Most enhancers are located upstream of the cap site although enhancers have also been identified within the C-J intron of mammalian immunoglobulin heavy and light chain genes (Banerji *et al* 1983, Picard and Schaffner 1984), within the first intron of the human pro-α1 collagen gene (Rossouw *et al* 1987) and in the 3' portion of vertebrate

globin genes (Behringer *et al* 1987). Enhancers are generally 100–200bp in length and while no single sequence motif is common to all, a characteristic of many enhancers is the presence of one or more sequence motifs which are present in several copies. Individual copies of these repeats often have little stimulatory activity by themselves, but act co-operatively to create a strong enhancer (Sassone-Corsi and Borrelli 1986, Ondek *et al* 1987). Similarly, yeast UASs are often present in duplicate (Tajima *et al* 1986, Thiele and Hamer 1986, Wouldt *et al* 1986) or else contain multiply repeated elements (Sarokin and Carlson 1986).

The SV40 enhancer acts in a wide variety of tissues and hosts, but other viral enhancers and most animal cell enhancers will only activate transcription in certain cell-types or in response to an inducing signal. Likewise, most (if not all) yeast UASs are inducible rather than constitutive elements, and the best characterised plant enhancer, belonging to the pea rbcS gene, activates transcription in a light-dependent and organ-specific manner (Timko *et al* 1985, Fluhr *et al* 1986).

In vivo competition experiments first demonstrated that soluble cellular components interact with the enhancer to elicit stimulation (Scholer and Gruss 1984), and DNAaseI protection studies now suggest that some enhancers can bind several different factors simultaneously (Bohmann *et al* 1987). The existence of ubiquitous enhancer-binding factors has been suggested from both DNAase 1 footprinting and competition studies (Mercola *et al* 1985, Bohmann *et al* 1987). A core consensus sequence (TGTGA/TA/TAG) which is common to many enhancers (see Sassone-Corsi and Borrelli 1986) may be the target of one such general enhancer-binding factor. This interaction cannot be a

universal feature of enhancers, however, since not all enhancers contain the core sequence. Other interactions between *trans*-acting factors and enhancers appear to be more specific (Wasylyk and Wasylyk 1986) and probably contribute to the inducible or tissue-specific regulation of these enhancers.

Negative regulation of transcription

Rather less is known about the negative regulation of RNA synthesis, but it is becoming evident that many eukaryotic promoters are composite structures, contain negatively as well as positively acting elements. The inhibition of transcription in active chromatin by specific *trans*-acting repressors may be especially important in yeast, where blocks or domains of chromatin cannot be inactivated by any general repression mechanism (see section 1.31). The best characterised yeast repressor is the MATα2 gene product which co-ordinately represses the transcription of all genes specific to the 'a' mating type. A conserved element adjoining the a-specific genes has been shown to mediate the repression and binds MATα2 protein (Miller *et al* 1985, Johnson and Herskowitz 1985). The element functions to full effect only when situated between the TATA box and UAS although it represses weakly when placed distal to the UAS. This finding suggests a mechanism distinct from that in prokaryotes, where the operator-bound repressor sterically excludes RNA polymerase from binding to the promoter. The MATα2 repressor may instead function by antagonising the activity of the UAS.

The mechanism by which the two stored mating type genes of yeast, HMLα and HMRa, are repressed appears to be quite different. Brand *et al* (1985) have identified a *cis*-acting silencer element which is

responsible for the transcriptional repression and has properties akin to an enhancer, being able to function in either orientation, at a distance of 2.6kb from the gene, and when placed downstream of the cap site. The silencer is able to switch off transcription from heterologous promoters and recent studies have identified two silencer-binding protein factors (Shore *et al* 1987). Negative regulatory elements with properties similar to the HMR silencer are associated with the mouse c-myc gene (reviewed by Linzer 1986) and the chicken lysozyme gene (Baniahmad *et al* 1987).

Simpson *et al* (1986) have reported a 247bp element within the promoter of a pea cab gene which contains not only a light-inducible enhancer but a tissue-specific silencer, repressing the activity of a linked constitutive promoter in root tissue.

In their study of the human β-interferon gene enhancer, Goodbourn *et al* (1986) uncovered yet another mechanism of transcriptional repression when they found that the enhancer is actually composed of a positive (enhancer) element tightly coupled to a negative regulatory element. Using a combination of deletion analysis and DNA footprinting, they have demonstrated that prior to induction, a repressor molecule is bound to the negative element, apparently preventing the interaction of a constitutive transcription factor with the adjacent enhancer element. Upon induction (viral infection) the repressor dissociates, thereby allowing the transcription factor to bind and activate the enhancer (Goodbourn *et al* 1986, Zinn and Maniatis 1986). Physical proximity of the negative and postive elements appears to be a critical feature of this type of regulation, since the negative element is able to inhibit the activity of a heterologous (tk) promoter only when situated immediately adjacent to the tk upstream element.

Possible mechanisms of long-range gene regulation

Various models have been put forward to account for the stimulatory or inhibitory influence of DNA sequences situated hundreds and even thousands of base pairs from the transcription start. A so-called 'twisting' model proposes that regulatory proteins bind to DNA and induce a change in the DNA topology, for example localised unwinding or B$\rightleftharpoons$Z transition. This conformational change is transmitted to the TATA box or transcription start where it allows other proteins to bind and begin transcription. A second model, the 'sliding' model, proposes that upstream and enhancer sequences are regions at which components required for transcription enter the chromatin and thereupon slide along the DNA to the site of transcription initiation. In a third model it is suggested that the binding of a regulatory protein to its target sequence faciliates the binding of additional proteins to adjacent sequences, and this co-operative binding continues until a scaffold of proteins has built up extending as far as the initiation site. A fourth model proposes that regulation requires direct protein-protein contacts between factors bound to widely separated DNA sequences, and that this contact is achieved by looping out or bending of the intervening DNA. Recent studies of the phage λ repressor and the SV40 early promoter both provide strong precedents for the direct contact model. These and other experiments pertaining to the mechanism(s) of long-range gene regulation have been discussed in an excellent review by Ptashne (1986).

1.33   Transcription termination and 3' processing of mRNA

Analysis of pulse-labelled RNAs and *in vitro* nuclear run-on transcription studies have shown that in higher eukaryotes the 3'

terminus of the mature mRNA is generated by endonucleolytic cleavage of a longer RNA, rather than by transcription termination (see Birnstiel *et al* 1985). The primary transcript can extend for 1000 or more nucleotides beyond the processing site, and addition of the poly(A) tract (polyadenylation) appears to be mechanistically coupled to the cleavage event so that newly generated 3' ends are instantly protected by a poly(A) tail. The hexanucleotide sequence AAUAAA is highly conserved at a position 10 to 25 nucleotides upstream of the poly(A) addition site. This sequence is important but not sufficient for correct endonucleolytic cleavage and GU-rich sequences located downstream of the processing site are also required (Birnstiel *et al* 1985, Sadofsky *et al* 1985). A role for small nuclear RNA (snRNA) in 3' mRNA processing has been postulated although it appears that none of the major abundant snRNPs is involved (Berget and Robberson 1986). Antiserum to small nuclear ribonucleoproteins (snRNPs) does, however, inhibit cleavage and polyadenylation *in vitro* (see Birnstiel *et al* 1985). The sequence or structures directing termination of the primary transcript are so far ill-defined. Hybridisation and S1 mapping experiments have revealed extensive 3' heterogeneity of the primary transcripts (see Birnstiel *et al* 1985), consistent with the absence of a strong termination signal. Possibly the adoption of a 3' processing mechanism has removed the evolutionary pressure for an efficient and accurate termination event. Plant mRNAs contain the same consensus sequence AAUAAA shortly upstream of the polyadenylation site (Heidecker and Messing 1986, Joshi 1987a) and it is likely that these too are generated by endonucleolytic cleavage of a primary transcript. Most plant genes, as well as some animal genes, contain several poly(A) addition sites and it is unclear what factors, if any, control the choice of which site is used.

In contrast to the mRNAs in higher eukaryotes, few if any of those in yeast possess a AAUAAA sequence and polyadenylation appears to be tightly coupled to the termination of transcription (Zaret and Sherman 1982, Zaret and Sherman 1984). Virtually all mRNAs in yeast, including histone mRNAs, are polyadenylated, and a simple explanation not yet ruled out is that every mRNA 3'-OH is a substrate for poly(A) polymerase. Although transcription termination is a precise event, there appears to be little uniformity among yeast termination signals. From a mutational analysis of the CYC1 gene, Zaret and Sherman (1982) have proposed that the sequence TAG..(N)n..TAGT..(N)n..(ATrich)TTT constitutes a termination signal. Removal of part of this sequence resulted in a 90% drop in the steady-state mRNA level and the appearance of 3' heterogenous RNA species extending up to 1000 nucleotides beyond the normal 3' terminus. The sequence TTTTTATA is responsible for transcription termination in a Drosophila ADE8 gene cloned in yeast and has been proposed as an alternative termination signal (Henikoff and Cohen 1984). The 3' untranslated portions of many yeast mRNAs, however, contain neither of these proposed signals and furthermore, sequences resembling the tripartite consensus of Zaret and Sherman are found within the coding region or upstream of other genes. More recently, a third class of yeast terminator has been identified within the promoter region of the URA3 gene. This terminator shares no discernible homology with the two described above, but contains a region of dyad symmetry as well as a 15bp direct repeat (Yarger *et al* 1986). The absence of a common sequence or structural determinant for transcription termination in yeast is intriguing and might suggest that the binding of any one of several termination factors is sufficient to instigate termination.

The disparity between transcription termination mechanisms in yeast and

higher eukaryotes can create problems when genes from higher eukaryotes are expressed in yeast. On the whole, the termination and/or 3' processing signals of foreign genes are not recognised in yeast (Innis et al 1985, Watts et al 1983) although this does not necessarily impair gene expression providing an efficient yeast termination signal is subsequently encountered. The expression of both calf prochymosin and human interferon genes in yeast was found to be improved by placing a known yeast terminator sequence downstream of the gene (Mellor et al 1983, Hitzman et al 1983), indicating the importance of an efficient transcription termination step for high level gene expression. Another potential hazard to foreign gene expression is illustrated by the rabbit β-globin gene, in which 3' truncated transcripts were synthesised owing to premature termination at an AT-rich sequence within the gene (Beggs et al 1980). This appears to be an uncommon problem however, and to date no other examples of premature termination have been reported.

1.34 mRNA splicing

Splicing is the term applied to the precise removal of intervening sequences (introns) from precursor mRNA (pre-mRNA) to give a mature and translatable product. With the development of efficient in vitro splicing systems, the biochemical mechanism of the splicing reaction has been largely elucidated and found to be essentially the same in yeast and mammalian systems, although subtle differences in the mechanism (detailed below) have been at the root of several failures to demonstrate correct splicing of foreign introns in yeast (Watts et al 1983, Langford et al 1983).

The characterisation of RNA processing products synthesised in vivo, as
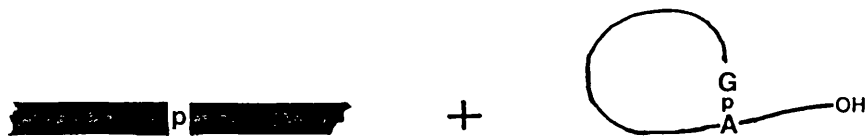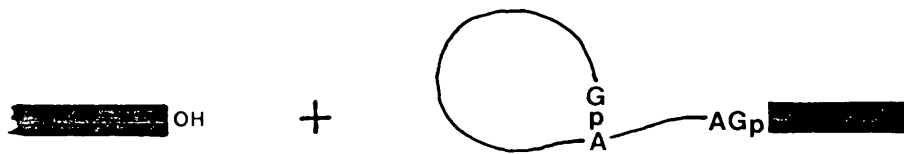
Figure 1.5   The eukaryotic mRNA splicing pathway.   Black boxes represent exons and the fine line is intervening DNA.

Table 1.2 Consensus sequences for eukaryotic splicing

	5' intron boundary	3' intron boundary	branch point
Animal	^C_AAG/GURAGU[1]	(Y)$_n$NYAG/G[1]	YNYURAY[2]
Yeast	/GUAYGU[3,1]	YAG/[3]	UACUAACA[3]
Plant	^C_AAG/GUAAGU[4]	(U)$_n$GCAG/G[4]	CURAY[4]

/ denotes cleavage site

1. Mount 1982

2. Green 1986

3. Langford et al 1984

4. Brown 1986

well as kinetic and structural data obtained from *in vitro* splicing reactions, (reviewed by Green 1986) have led to the formulation of a 2 stage pre-mRNA splicing pathway (Figure 1.5). The first step involves cleavage at the G residue of the 5' splice site and the formation of a lariat intermediate in which the free 5' end of the intron is joined by a 2'-5' phosphodiester bond to a specific A residue located near the 3' end of the intron. The second step consists of cleavage at the 3' splice site and ligation of the exons, resulting in excison of the intact intron in a lariat configuration.

Three *cis*-acting elements, identified by sequence comparison of introns and by site-directed mutagenesis, have been shown to influence the splicing reaction. These are the 5' splice junction, the 3' splice junction and the branch point. Consensus sequences for these elements in yeast and higher eukaryotic introns are compared in Table 1.2. *In vitro* splicing of both yeast and mammalian pre-mRNA requires the assembly of a dynamic ribonucleoprotein complex or 'spliceosome' composed of the RNA substrate, snRNPs and a variety of other factors (Maniatis and Reed 1987, Padgett *et al* 1986). snRNPs appear to be intimately involved in higher eukaryotic splicing and recent studies have shown that U1 snRNP binds to the 5' splice site, U2 snRNP binds to the branch point and an additional snRNP, possibly U5, interacts specifically with the 3' splice site (reviewed by Manaitis and Reed 1987). snRNAs also appear to be an integral component of the yeast splicing process (Pikielny and Rosbash 1986) but in neither system is it understood how these interactions might effect cleavage or ligation. In yeast, a genetic approach to the study of splicing is feasible and has been used to identify additional genes required for splicing. Analysis of a set of temperature sensisitive mutants which accumulate unspliced pre-mRNAs at the restrictive temperature has revealed at

least three genes, rna2, rna5 and rna11, that encode products directly involved in pre-mRNA splicing (Lustig *et al* 1986).

In spite of many similarities between yeast and higher eukaryotic splicing, a number of differences are apparent. Firstly, the yeast 5' splice site is highly conserved whereas the corresponding sequence in metazoan species may vary significantly. Secondly, the polypyrimidine tract which precedes the 3' splice site in higher eukaryotes and is essential for the first step in intron excision (Ruskin and Green 1985), is notably absent from most yeast 3' splice sites. Thirdly, mutations in the 5' splice site and branch point tend to abolish splicing in yeast, whereas similar mutations in higher eukaryotic introns lead to the activation of cryptic sites (see Green 1986). The fourth and most striking difference lies in the method of branch point selection. In higher eukaryotes there is no strict sequence requirement for branch formation, however there is a strong distance constraint: regardless of intron length (which can be several kilobases) all branch points are located 18 to 30nt upstream of the 3' splice site. In contrast, there is a rigid sequence requirement (the TACTAAC box) for branch formation in yeast and splicing tends to occur at the first AG dinucleotide downstream of this element (Langford *et al* 1984). Put another way, in higher eukaryotes the 3' splice site fixes the branch point, whereas in yeast it is the branch point which determines the 3' splice site.

The stringent sequence requirement for lariat formation in yeast may explain a) why cryptic branch points are not activated following mutation of the TACTAAC sequence and b) why higher eukaryotic pre-mRNAs which lack a good match to the TACTAAC box are not spliced in yeast. This is nicely illustrated by a study of the Drosophila ADH gene in

yeast, in which it was shown that while two introns remained unspliced, a third was inefficiently but accurately cleaved at the 3' end (Watts *et al* 1983). Unlike the unspliced introns, this intron contained a TACTAAC-like sequence (AACTAAC) 22 nucleotides from the 3' splicing site.

The differing sequence requirements for splicing in yeast and higher eukaryotes may reflect differences in the structural organisation and regulation of genes in these organisms. In yeast only 5-10% of structural genes are interrupted, introns are usually less than 300bp in length and there is only one per gene. Most genes of higher eukaryotes contain multiple introns, ranging in size from 31bp to several kb, and some genes encode transcripts that are differentially spliced. The less stringent sequence requirements for splicing in higher eukaryotes may therefore be a consequence of the need for greater splicing flexibility in these organisms.

1.35 Translation

The nuclear genetic code is universal so translation start and stop codons will be recognised in all species. Chain initiation is signalled by AUG, which codes for methionine, and chain termination by one of the nonsense codons UAA, G or UGA. There is no ribosome-binding sequence in eukaryotic messenger RNAs analogous to the Shine-Dalgarno consensus of prokaryotes, which interacts with the 3' end of 16S RNA, positioning the ribosome close to the initiating AUG codon. In at least 90% of the eukaryotic mRNAs examined, however, the initiator codon is the 5' proximal AUG. This led Kozak (1981) to propose a scanning model in which a 40S ribosomal subunit attaches at the capped 5' end of mRNA and migrates in a 3' direction until it

reaches the first AUG triplet, where translation is initiated providing the codon occurs in a favourable context (see below). If the sequence around the first AUG triplet is suboptimal, a proportion of the 40S sub-units bypass that site and initiate further downstream. Several observations are rationalised by the scanning model, for example eukaryotic cellular mRNAs are invariably monocistronic, a 5' cap structure is essential for efficient translation (Shatkin 1976), and ribosomes are unable to bind to circular mRNAs (Kozak 1979). Additionally, in both yeast and higher eukaryotes, the insertion of an AUG triplet into the 5' untranslated region of an mRNA severely inhibits translation initiation at the correct position (Zitomer *et al* 1984, Kozak 1984a).

The 5' non-coding region (or leader sequence) in eukaryotic mRNAs is typically 40-80nt in length. Leaders as long as 742nt and as short as 3nt have been characterised, however these are rare and generally confined to viral mRNAs. Leader sequences exhibit enormous variability of nucleotide sequence, but a survey of over 200 higher eukaryotic mRNA sequences has revealed that initiator AUG triplets tend to occur in a restricted sequence context with the consensus CC(A/G)CCAUGG (Kozak 1984b). The purines at positions -3 and +4 are especially highly conserved, and inspection of the few known naturally occuring non-functional AUG triplets located within mRNA leader regions revealed these to have a pyrimidine at one or both of these key positions (Kozak 1984b and see Table 1.3). The functional significance of the purines at positions -3 and +4 was first demonstrated by measuring the binding affinity of short synthetic oligonucleotides to wheat germ ribosomes *in vitro* (Kozak 1981). More recently, the optimal AUG context has been confirmed and refined to ACCAUGG by analysing the effects *in vivo* of single base substitutions around the initiator AUG triplet in a cloned

Table 1.3 Distribution of sequence contexts for functional

and non-functional AUG codons in eukaryotic genes

AUG context	Number of initiator AUG codons	Number of non-functional upstream AUG codons
ANNAUGG	74	1
ANNAUGA	38	0
ANNAUGY	29	0
GNNAUGG	18	2
GNNAUGA	4	0
GNNAUGY	4	4
YNNAUGG	11	4
YNNAUGA	1?	7
YNNAUGY	1	14

From Kozak (1983)

preproinsulin gene (Kozak 1986). Both this study and a report by Morlé (1985) again emphasise the dominant effect of a purine at position -3 for efficient translation. Morlé described an α-thalassaemia in which the deficiency in globin synthesis was attributed to a mutation altering CACC<u>AUG</u> to CCCC<u>AUG</u>. Inspections of plant and yeast translation initiation regions likewise reveal the conservation of A at position -3, although the bias in surrounding nucleotide distribution differs from that reported for animal genes (Kozak 1984b). The consensus sequence at plant initiator codons is UAAACA<u>AUG</u>GCU (Joshi 1987) and at yeast initiator codons is A/YAA/UAAUGUCU (Cigan and Donahue 1987).

The function of the untranslated leader sequence is largely unknown. In a few cases, for example the adenovirus late major transcript (Logan and Shenk 1984), the <u>GCN4</u> gene of yeast (Mueller and Hinnenbusch 1986) and <u>Drosophila</u> heat shock genes (Hultmark *et al* 1986), sequences in the leader have been implicated in translational control. In other mRNAs, however, deletion, insertion or recombination within the leader does not affect expression of the gene (see Kozak 1983 for references). Eukaryotic mRNAs are not translated with equal efficiency, however, and in the absence of an active translational control mechanism, it is probable that intrinsic structural features of the mRNA influence the rate of translation to some extent. It has been suggested from studies on the translational efficiencies of chimaeric mRNAs, that the leader region of well-translated messages might have either a diminished requirement or a higher affinity for limiting components of the translational machinery (Jobling and Gehrke 1987). It is not clear at present whether this property is associated with a specific nucleotide sequence or with a more general structural feature, for example the absence of secondary structure. Secondary structure in mRNA has

repeatedly been implicated in determining its translatability and numerous studies now show that excessive secondary structure in the 5' non-coding region of mRNAs impedes their translation (see Pelletier and Sonenberg 1987). A key (and rate-limiting) step in the translation initiation pathway is the binding of the 40S ribosome subunit (and associated factors) to the mRNA. It has been proposed that a 24kd cap-binding protein, in association with other factors, serves to destabilise local secondary structure in an ATP-requiring reaction, thereby facilitating attachment of the 40S ribosomal subunit (Pettelier and Sonenberg 1987, Rhoads 1988). If this is indeed the case, then energetic considerations alone would favour ribosome attachment to mRNAs having little secondary structure at the 5' end.

The generality of the ribosome scanning mechanism and the preferred AUG context in eukaryotes suggests that a given transcript will be translated with much the same efficiency in any system. A factor which may modulate the rate of chain elongation to a small extent during translation is the codon composition of the gene. There are 25 preferred codons in S.cerevisiae which correspond to the most abundant iso-accepting tRNAs (Bennetzen and Hall 1982). Certain very highly expressed yeast genes exhibit a marked bias for preferred codons (greater than 90% for glyceraldehyde-3-phosphate dehydrogenase gene) whereas genes encoding less abundant proteins show a more random codon usage. In extreme cases where foreign genes contain a high proportion of rare codons, the rate of protein synthesis may be impeded, but there is little direct evidence to support the idea that poor codon bias is a significant barrier to foreign gene expression in yeast (Kingsman *et al* 1985).

1.36 Other factors affecting gene expression

This comparative review of gene expression in eukaryotes has so far concentrated on the synthesis, splicing and translation of messenger RNA. A number of important areas not yet covered are the control of mRNA turnover and the role of DNA methylation in transcriptional activation. Higher eukaryotic genomes, in particular vertebrates and higher plants, are highly methylated, the modified base invariably being 5-methyl cytosine. 5-methyl cytosine occurs predominantly at the sequence CpG but has also been detected in the dinucleotides CpA, CpT and CpC. The biological significance of methylated DNA is not completely clear. In prokaryotes at least, methylation can affect DNA-protein interactions, a familiar example being restriction endonucleases which will not cleave a methylated substrate. In many vertebrate and viral genes, transcriptional activation is associated with hypomethylation of the upstream DNA, a topic reviewed by Doerfler (1983) and more recently by Bird (1987). An unresolved question is whether demethylation is a prerequisite for transcriptional activation or a consequence of it. The methylation state of genes in lower eukaryotes, including S.cerevisiae, is not generally a consideration as these organsims do not methylate their genomic DNA to any significant extent.

The rate of mRNA decay in the cytoplasm will have an important bearing on gene expression by influencing the steady-state levels of mRNAs. The metabolic stability of individual mRNA species differ widely, half lives varying between 3 and 100 minutes in yeast and from 15 minutes up to several hours or even days in higher eukaryotes. It is likely that intrinsic structural properties of mRNA molecules are the major factors determining their rates of decay, although in mammalian systems there

44

are examples of mRNAs whose rate of degradation can be modulated by endogenous or exogenous stimuli such as hormones (see Raghow 1987). The half life of vitellogenin mRNA, for example, is increased 30-fold (from 16 to 480 hours) upon exposure of cells to oestrogen. Likewise, histone mRNAs which are synthesised at a constant rate throughout the cell cycle are specifically stabilised during S phase when there is a high demand for new histones. The mechanisms by which mRNA turnover is altered in these circumstances is obscure.

Little is known about factors influencing mRNA stability in higher plants or yeast, although in yeast an inverse relationship between mRNA length and stability has been reported (Chinnapan-Santiago *et al* 1986). In mammalian systems, structural features of mRNA have been implicated in determining individual rates of turnover (see Raghow 1987, Brawerman 1986). In particular, a specific sequence promoting mRNA decay – a destabilising element – has been identified in the 3' non-coding region of a human gene encoding the lymphokine GM-CSF (Shaw and Kamen 1986). A similar (AU rich) sequence is present in a variety of other short-lived mRNAs including c-myc, c-fos and interferon. Again, the mechanism by which a specific sequence promotes decay is unknown. Overall, our knowledge of factors determining the differential stability of mRNAs in eukaryotes is very limited and the degradative enzyme(s) involved remain a mystery.

1.37 Outline of the project

The aim of this work was to ascertain to what extent the transcription machinery of <u>Saccharomyces</u> <u>cerevisiae</u> is able to utilise plant promoter signals. Using the two major promoters of cauliflower mosaic virus as a model system, the project examined the efficiency with which the CaMV promoters worked in yeast and the accuracy of the transcription initiation process. The 19S promoter was found to operate in yeast but RNA synthesis initiated at a site different from that used in plants. No evidence of 35S promoter function was obtained. A mutational analysis of the 19S promoter was carried out to compare DNA sequence requirements for recognition of the 19S promoter in higher plants and in yeast. The data suggests that some aspects of the transcription initiation process (*eg* TATA box recognition) are conserved whereas others (*eg* cap site selection) have diverged in the course of eukaryotic evolution.

CHAPTER 2

MATERIALS AND METHODS

2.1 Bacterial and yeast strains

All the strains used in this study are listed in Table 2.1. Strains were stored at 4°C on agar plates and restreaked at regular intervals. Stocks were also kept at -80°C in complex medium containing 15% glycerol.

2.2 Growth of liquid cultures

Single colonies were inoculated into sterile growth medium and incubated in a shaking water bath or cabinet air shaker at 28°C (A.tumefaciens), 30°C (S.cerevisiae) or 30°C (E.coli).

2.3 Bacterial growth media

L broth: (per litre) 10g Bacto-tryptone, 5g Bacto-yeast extract, 5g NaCl. pH was adjusted to 7.2 with sodium hydroxide.

L agar: L broth solidified with 1.5% Bacto-agar.

M9 agar: 40ml M9 salts, 4ml M9 additive and 8ml 20% glucose* was added to 360ml molten sterile 1.7% water agar. 0.4ml 2mg/ml thiamine** and 4ml 2mg/ml threonine** were added as required.

M9 salts (10X): (per litre) 70g Na_2HPO_4, 30g KH_2PO_4, 5g NaCl, 10g NH_4Cl.

M9 additive (100X): 0.1M $MgSO_4$, 0.01M $CaCl_2$.

Table 2.1 Bacterial and yeast strains

Organism	Strain	Genotype	Reference/source
E.coli	NM522	Δ(lac-proAB), hsdΔ5 (r_k^-, m_k^+), thi$^-$, supE, F' proAB, lacI^{Q}ZΔM15	Gough and Murray (1983)
E.coli	5K	thi-1, thr-1, leuB6, lacY1, hsdR, supE, tonA	A Mileham, Leicester Biocentre
E.coli	GM48	F$^-$, dam-3, dcm-6, gal, ara, lac, thr, leu, thi, tonA, tsx	Marinus (1973)
E.coli	HB101	F$^-$, hsdS20 (r_B^-, m_B^-) recA13, ara-14, proA2, lacY1, galK2, rpsL20, xyl-5, mtl-1, supE44, λ^-	Boyer and Roulland-Dussiox (1969)
E.coli	HB101/pRK2013	HB101 with plasmid pRK2013	Ditta et al (1980)
A.tumefaciens	LBA4404	Contains pAL4404, an extensively deleted Ti plasmid	Hoekema et al (1983)
S.cerevisiae	S150-2B	MATa, leu2-3, leu2-112, his3-Δ, trp1-284, ura3-52	Obtained from J Hicks, Cold Spring Harbor Laborarory, New York

BBL agar: (per litre) 10g Baltimore Biological Laboratories Tripticase, 5g NaCl, 10g Bacto-agar.

BBL soft agar: as for BBL agar except 6.5g Bacto-agar.

Nutrient broth: 13g nutrient broth (Oxoid) per litre.

T agar: 20ml T buffer, 20ml T salts and 10ml 20% glucose* were added to 350ml molten sterile 1.7% water agar.

T buffer (20X): (per litre) 210g K_2HPO_4, 90g KH_2PO_4.

T salts (20X): (per litre) 4g $MgSO_4$, 0.2g $CaCl_2$, 0.1g $FeSO_4$ 0.04g $MnCl_2$, 20g NH_4Cl.

Antibiotic stocks were stored at -20°C and used at the following concentrations:

Compound	Stock conc^n	Working conc^n
sodium ampicillin	25mg/ml in Q water	50μg/ml
kanamycin sulphate	50mg/ml in Q water	50μg/ml
chloramphenicol	30mg/ml in ethanol	5μg/ml
tetracycline hydrochloride	10mg/ml in 50% ethanol	10μg/ml
rifampicin	20mg/ml in methanol	50μg/ml

2.4 Yeast growth media

SD broth: (per litre) 6.7g yeast nitrogen base, 20g glucose.

SD agar: SD broth solidified with 2% Bacto-agar.

SG broth: (per litre) 6.7g yeast nitrogen base, 50g galactose.

YPD broth: (per litre) 10g Bacto-yeast extract, 20g Bacto-peptone, 20g glucose.

YPD agar: YPD broth solidified with 2% Bacto-agar.

Yeast media was routinely prepared without the sugar in 9/10 of the final volume. 1/9 volume of 20% glucose* or 50% galactose* was added after the media had been autoclaved.

Yeast minimal medium was supplemented as necessary with the following:

Compound	Stock conc^n	Storage temp	Final conc^n
uracil*	2mg/ml	room temp	20μg/ml
L-tryptophan*	2mg/ml	4°C	20μg/ml
L-histidine-HCl*	2mg/ml	4°C	20μg/ml
L-leucine*	3mg/ml	room temp	30μg/ml

Unless otherwise indicated media were sterilised by autoclaving for 15 minutes at 15 lbs/sq in.

* sterilised by autoclaving for 20 minutes at 10 lbs/sq in.

** filter sterilised.

2.5 Measurement of plasmid stability in yeast

To measure the rate at which plasmids carrying a selective marker (LEU2) were lost from a logarithmically-growing population of yeast cells, single transformant colonies were inoculated into liquid medium and incubated with good aeration at 30°C. The cultures were repeatedly diluted with fresh medium so that exponential growth was maintained for at least 30 generations. At intervals of approximately 5 generation,

samples of cells were removed from the cultures, diluted in sterile water, and plated on YPD agar (complex medium). The resulting colonies were replica plated onto SD agar (minimal medium) to screen for leucine-independent growth. The proportion of Leu$^+$ colonies is representative of the proportion of the viable cell population containing plasmid at the time of sampling. For each plasmid analysed, three independent transformants were tested, and the experiment was performed under both selective and non-selective growth conditions. Approximately 500 colonies were scored for Leu$^+$/Leu$^-$ phenotype at each sampling.

2.6 Construction of plasmids

Recombinant DNA techniques including restriction endonuclease digestion, ligation, filling in of recessed 3' DNA ends with Klenow fragment, 5'phosphorylation of DNA with T4 kinase, cloning linkers, dephosphorylation of 5'termini using calf intestinal alkaline phosphatase and the progressive digestion of double stranded DNA with Bal31 nuclease were all performed using standard buffers and conditions (Maniatis *et al*, 1982). Most enzymes were obtained from Bethesda Research Laboratories or New England Biolabs. Pharmacia PL ligase (6300u/ml) was used for blunt-ended ligations.

2.7 Gel electrophoresis of DNA

Preparation and electrophoresis of agarose and polyacrylamide gels was as described (Maniatis *et al* 1982) using 1X TBE buffer. Sample loading buffer was prepared as a 10X stock consisting of 0.4% bromophenol blue in 50% glycerol. Ethidium bromide (10mg/ml in water, stored at 4°C) was added to the reservoir buffer of agarose gels at 0.5μg/ml.

Polyacrylamide gels were stained subsequent to electrophoresis by soaking for 20 minutes in 1X TBE, 0.5μg/ml ethidium bromide then briefly destained in water. Gels were placed on a short-wave UV transilluminator to visualise the DNA and photographed using a Polaroid MP4 camera, type 55 or 57 film and an orange filter.

2.8 Recovery of DNA fragments from polyacrylamide gels

4% polyacrylamide gels were used for fragment purification, giving better yields than higher percentage gels. After electrophoresis, the gel was stained with ethidium bromide (section 2.7) and the DNA bands visualised using a long wave UV lamp. The desired band was excised and thoroughly squashed between two layers of Nescofilm. It was then transferred to an Eppendorf tube containing 0.2ml of elution buffer (0.5M ammonium acetate, 10mM magnesium acetate, 1mM Na_2EDTA, 0.1% SDS) and shaken at 37°C for 16 hours. The gel/buffer mixture was filtered to remove particles of acrylamide as follows. A plug of polyallomer wool was placed in a 0.5ml Eppendorf tube, the cap was removed and the bottom pierced with a hot wire. This tube was then placed inside a 1.5ml Eppendorf tube. The gel/ buffer mixture was pipetted into the smaller tube and spun for 5-10 seconds at low speed (2900g) in a microfuge. Buffer was recovered from the larger tube and the column was washed with a further 0.2ml of elution buffer. The filtrates were combined, vortexed with an equal volume of phenol then spun for 5 mins in a microfuge. The aqueous phase was transferred to a tube containing an equal volume of chloroform, vortexed and spun again. DNA was precipitated from the aqueous phase by the addition of 0.1 volume of 3M sodium acetate pH5.0 and 2.5 volumes of ethanol and chilling to -70°C for 30 minutes. The DNA was pelleted for 10 minutes in a microfuge, dried and resuspended in a small volume of TE buffer.

2.9 Preparation of phenol

Liquid phenol freshly distilled by Fisons was stored frozen at $-20^{\circ}C$. Batches were prepared for use by adding 8-hydroxyquinoline to a concentration of 0.1%. The phenol was equilibrated once with an equal volume of 1M Tris.HCl pH8.0 then twice with an equal volume of 0.1M Tris.HCl pH8.0 and stored at $4^{\circ}C$ under TE buffer.

2.10 Transformation of E.coli

Cells were made competent by treatment with calcium chloride as described by Morrison (1979), except that the (200ml) culture was concentrated 2-fold rather than 4-fold in $MgCl_2$ and $CaCl_2$. After the final resuspension in 75mM $CaCl_2$, 12.5% glycerol, cells were divided into 0.5ml aliquots and frozen in a dry ice/IMS bath for 10 minutes. Frozen cells were stored at $-80^{\circ}C$ for up to 18 months without significant loss of transformability. When required, an aliquot of frozen cells was thawed on ice for 15 minutes.

a) Transformation procedure when selecting for antibiotic resistance.
DNA was mixed with $100\mu l$ of 0.1M $CaCl_2$ in a 1.5ml Eppendorf tube and chilled on ice. $100\mu l$ of competent cells was added and the mixture left on ice for 20 minutes. Cells were given a $42^{\circ}C$ heat shock for 3 minutes then returned to ice for 10 minutes. 1ml of L broth was added and the tube incubated at $37^{\circ}C$ for 40 minutes. Cells were pelleted for 1 minute in a microfuge, resuspended in approximately $100\mu l$ of L broth and spread on L agar containing an appropriate antibiotic. The plates were incubated at $37^{\circ}C$ overnight.

b) Transformation procedure when selecting for complementation of an

auxotrophic marker.

The method was as above except expression time in L broth was increased to 1 hour and the cells were washed twice in 1X M9 salts before being spread on M9 agar plates. The plates were incubated at 37°C for 3 days.

2.11 Transfection of E.coli with M13 RF DNA

0.3ml of competent NM522 cells was mixed with the DNA in a chilled 15ml Falcon tube and left on ice for 30 minutes. Cells were heat shocked for 2 minutes at 42°C then the following were added sequentially: 100µl of a fresh overnight culture of NM522, 10µl of 0.1M IPTG (stored at -20°C), 50µl of Xgal (2% in dimethylformamide, stored at -20°C), and 3ml molten BBL soft agar cooled to 45°C. The tube was capped, gently inverted several times to ensure thorough mixing, then the contents poured evenly onto a BBL agar plate. The plates were left untouched for 15 minutes while the overlay solidified, then inverted and incubated at 37°C overnight.

2.12 Transformation of yeast cells

Cells from a 50ml YPD culture grown to a density of 10^7 cells/ml were harvested in a bench centrifuge (5mins at 5krpm) and washed twice with 10ml of TE pH7.6. The cell pellet was resuspended in 5ml of 0.1M lithium acetate in TE pH7.6 and shaken at 30°C for 1 hour. 150µl of competent cells was pipetted into an Eppendorf tube containing the transforming DNA, 350µl of 50% PEG 4000 was added and the tube contents mixed by inversion. Cells were incubated for 1 hour in a 30°C water bath, heat pulsed at 42°C for 5 minutes then plated directly onto plasmid-selective SD agar.

2.13 Small scale plasmid preparations from E.coli

Single colonies were inoculated into 2ml of L broth and shaken overnight at 37°C with antibiotic selection. 1.5ml of the culture was transferred to an Eppendorf tube and cells pelleted by 2 minutes centrifugation in a microfuge. The pellet was resuspended in 100μl of freshly prepared ice-cold lysis buffer (50mM glucose, 10mM Na$_2$EDTA, 25mM Tris.HCl pH8.0, 2mg/ml lysozyme) and left on ice for 5 minutes. 200μl of 0.2M NaOH, 1% SDS was added and gently mixed until the solution became clear and viscous. The tube was returned to ice for 5 minutes then 150μl of 3M sodium acetate pH5.0 was added and incubation on ice continued for 45 minutes. After 10 minutes centrifugation in a microfuge, the supernatant was poured into a fresh 1.5ml Eppendorf tube, 1ml of ethanol was added and the tube chilled at -20°C for 30 minutes. DNA was pelleted (5 mins in a microfuge) and resuspended in 100μl of 0.1M sodium acetate, 50mM Tris.HCl pH7.2. 0.5ml of ethanol was added and the DNA reprecipitated for 15 minutes at -20°C. After spinning for 5 minutes in a microfuge, the pellet was dried and resuspended in 50μl of TE pH8.0.

Preparations of this kind contain cellular RNA and traces of chromosomal DNA as well as plasmid DNA. The method typically yields 5-10μg of plasmid DNA.

2.14 Large scale plasmid preparations from E.coli

500ml of L broth was inoculated with 100μl of a stationary cell culture and incubated at 37°C overnight with antibiotic selection for the plasmid. Cells were harvested (5 mins, 5krpm in Sorvall GSA rotor) and resuspended in 25ml of 25mM Tris.HCl pH8.0, 5mM Na$_2$EDTA, 50mM NaCl. Cells were repelleted (5 mins, 5krpm in SS34 rotor) and suspended in

12.5ml of ice-cold 25% sucrose, 50mM Tris.HCl pH8.0. 1.5ml of freshly dissolved lysozyme (10mg/ml in 50mM Tris.HCl pH8.0) was added and the tube placed on ice for 8 minutes. 1.5ml of 0.25M Na_2EDTA was mixed in thoroughly and incubation on ice continued for 5 minutes. 15ml of lysis mix (o.2% Triton X-100, 50mM Tris.HCl pH8.0, 25mM Na_2EDTA) was added, the tube was shaken vigorously then placed on a 'multimix' shaker (Luckham) for 15 minutes at room temperature. The lysate was centrifuged (1 hour at 17krpm, 4°C in SS34) and the supernatant tipped into a 50ml measuring cylinder. 1g of solid caesium chloride and 50μl of 10mg/ml ethidium bromide was added for every ml and when the CsCl had dissolved, the solution was centrifuged for 10 minutes at 10krpm in the SS34 rotor. Avoiding the floating pellicle, the solution was transferred to a 36ml ultracentrifuge tube (Sorvall 03141), topped up with liquid paraffin and if necessary a balance tube was prepared. The samples were spun in a Sorvall OTD65B ultracentrifuge at 40krpm, 20°C for 20 hours in a TV865 vertical rotor. Chromosomal and plasmid DNA bands were visualised in the gradient using a long-wave UV lamp. The tube was pierced just below the (lower) plasmid band with a 1.1mm bore hypodermic needle and the DNA was carefully drawn off into a 5ml syringe. Ethidium bromide was removed from the DNA by 4 sequential extractions with 2 volumes of isopropanol equilibrated with CsCl-saturated TE buffer. The DNA solution was dialysed against 3 successive litre volumes of TE pH8.0 for at least 1 hour each, then precipitated with 0.1 volume of 3M sodiuim acetate pH5.0 and 2 volumes of ethanol at -20°C. DNA was recovered by centrifugation (15krpm, 10mins in SS34), rinsed with 70% ethanol, dried and redissolved in 0.5ml of TE pH8.0.

Plasmid DNA prepared by this method was completely free from RNA and chromosomal DNA and yields of around 1mg were obtained.

2.15 Preparation of total DNA from A.tumefaciens

Single colonies of strains carrying Bin19 (or derivatives of this plasmid) were inoculated into N broth containing 50μg/ml kanamycin and grown to stationary phase at 28°C. DNA was prepared from 1.5ml of culture according to a published method (Lichtenstein and Draper 1985) and resuspended in 100μl.

2.16 Small scale preparation of yeast DNA

10ml cultures were grown to late logarithmic phase (2×10^7 cells/ml) in SD medium with selection for plasmid where appropriate. Cells were harvested in a bench centrifuge (5 mins, 5krpm), washed with 1ml of 1M sorbitol and transferred to an Eppendorf tube. Cells were pelleted (1 min in microfuge) and resuspended in 1ml of sphaeroplasting solution (1M sorbitol, 50mM Tris.HCl pH7.5, 20mM Na_2EDTA, 30mM DTT, 1mg/ml Zymolyase 100T (Seikagaku Kogyo Co. Ltd.)). The reaction was incubated at 37°C until sphaeroplasting was complete (approx. 30 minutes). Sphaeroplasts were harvested by centrifugation for 30 seconds at low speed (6500rpm) in a microfuge, resuspended in 0.5ml lysis buffer (100mM NaCl, 10mM Na_2EDTA, 50mM Tris.HCl pH7.5, 1% sodium lauroyl sarkosine) and vortexed for 1 minute. The lysate was incubated for 10 minutes at 65°C then spun for 10 minutes in a microfuge. The supernatant was extracted once with an equal volume of phenol/chloroform (1:1) and spun for 5 minutes in a microfuge. The aqueous phase was extracted with an equal volume of chloroform, spun again and transferred to a clean tube. 0.1 volume of 3M sodium acetate and 2 volumes of ethanol were added and the DNA precipitated at -20°C for 1 hour. DNA was pelleted (10 mins in a microfuge), rinsed with 70% ethanol, dried and resuspended in 50μl of TE pH8.0. This is a total

DNA preparation enriched for small plasmid molecules and will also contain RNA.

2.17 Preparation of genomic DNA from plants

DNA was extracted from approximately 5g of fresh leaf tissue (mid ribs removed) according to a published procedure (Coen *et al* 1986) with the following modifications.   (i) Tissue frozen in liquid nitrogen was ground with a mortar and pestle, (ii) the phenol/chloroform extraction was omitted, (iii) CTAB/DNA complexes were pelleted by centrifugation rather than picked out of the solution, (iv) the DNA pellet was washed thoroughly with 70% ethanol before being briefly dried and resuspended in TE pH8.0.

Yields were typically 200μg–400μg.

2.18 Preparation of single stranded M13 DNA

A culture of NM522 was grown to mid log phase and diluted 20-fold with L broth.   Individual phage plaques were inoculated into 2ml of the diluted culture in sterile glass test tubes and shaken vigorously at 37°C overnight.   Cells were pelleted from 1.5ml of culture by 10 minutes centrifugation in a microfuge.   1350μl of the supernatant was added to 150μl of 20% PEG 6000, 2.5M NaCl and left for 15 to 30 minutes at room temperature.   Phage particles were pelleted (10 mins in a microfuge) and the supernatant discarded, taking care to remove all traces of the PEG solution.   Phage were resuspended in 100μl of TE pH8.0, vortexed well with 50μl of phenol/ chloroform/isoamyl alcohol (24:24:1) then spun for 5 minutes in a microfuge.   90μl of the aqueous phase was transferred to a tube containing 50μl of chloroform/iso-amyl alcohol (24:1), vortexed and respun.   The aqueous phase was mixed with

12μl of 2.5M sodium acetate pH5.0 and 200μl of ethanol and placed at −70°C for 1 hour. DNA was pelleted for 10 minutes in a microfuge, rinsed with ethanol, dried in air for 15 minutes and resuspended in 30μl of TE pH8.0.

2.19 DNA sequencing (dideoxy method)

a) ^{35}S sequencing of M13 single stranded templates

10μl primer (Biolabs 15mer diluted to 2ng/μl), 4μl of TM buffer (100mM Tris.HCl pH8.0, 100mM MgCl$_2$) and 6μl of Q water were combined to provide 'primer mix' for 4 templates. 5μl of template was mixed with 5μl of 'primer mix' in a small Eppendorf tube and incubated at 65°C for 1 hour. The mixture was allowed to cool to room temperature over 30 minutes. 1μl of ^{35}S-dATP (Amersham, 600Ci/mMol) was added to each annealed template and 2μl of the ^{35}S-dATP/primer/template mix was dispensed into each of 4 colour-coded Eppendorf tubes. 2μl of A mix was added to the first tube, 2μl of T mix to the second, 2μl of C mix to the third and 2μl of G mix to the fourth. The composition of these mixes is given below. 1μl of Klenow (diluted to 1 unit/μl in Q water) was added to each tube and the reactions allowed to proceed for 20 minutes at room temperature. 2μl of chase (a mixture of all 4 dNTPs, each at 5μM) was added and the reactions continued for a further 20 minutes. Reactions were terminated by the addition of 4μl of formamide dye (deionised formamide, 10mM Na$_2$EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue) and stored at −20°C overnight. Prior to electrophoresis the samples were thawed, boiled for 2 minutes then cooled rapidly on ice. Half of each reaction was loaded on a 6% sequencing gel (see below) with a drawn-out capillary. The gel was run at 25mA for about 2 hours until the bromophenol blue reached the bottom, then the gel was dried down using a Bio-Rad gel drier and

exposed to X-ray film (Fuji RX) for 48 hours.

b) ^{32}P sequencing of plasmid DNA

Supercoiled plasmid DNA purified from a CsCl gradient was sequenced according to a published procedure (Haltiner *et al* 1985).

To prepare the sequencing gel: two 20cm x 40cm glass plates (one of them notched) were scrupulously cleaned, first with water then with IMS and acetone. The notched plate was siliconised by wiping the inner surface with Sigmacote. The glass plates were sandwiched together with 0.4mm plasticard side spacers and firmly sealed at the side and bottom edges with adhesive tape and bulldog clips. 25μl of TEMED and 250μl of freshly prepared 10% ammonium persulphate were pipetted into a beaker containing 50ml of 6% gel stock (see below), the mixture was swirled vigorously and poured between the glass plates, taking care not to trap any pockets of air. The straight edge of a sharkstooth comb was inserted to a depth of 6mm to provide a flat upper surface to the gel, and the whole assembly was left in a horizontal position while the acrylamide polymerised (approximately 30 minutes). The tape was removed from the bottom of the glass plates and the sharkstooth comb was extricated before the gel was placed in a vertical tank containing 1X TBE buffer (see below). The comb was then inserted so that the teeth just pierced the surface of the gel, which was then ready for use.

6% gel stock: (per litre) 460g urea, 100ml 10X TBE, 150ml acrylamide stock

acrylamide stock: 38% acrylamide, 2% bis-acrylamide (both ultrapure grade), stored over 20g/l Amberlite MB1 resin.

10X TBE buffer: (per litre) 108g Trizma base, 55g boric acid, 9.3g Na$_2$EDTA.

100ml of formamide was deionised by stirring with 5g of AG501-X8 mixed bed resin (Bio-Rad) for 30 minutes at room temperature.   The formamide was filtered and stored at 4°C.

Composition of dideoxy NTP mixes for ^{35}S sequencing:

	T mix	C mix	G mix	A mix
0.5mM dTTP	5	500	500	500
0.5mM dCTP	500	25	500	500
0.5mM dGTP	500	500	10	500
10mM ddTTP	50			
10mM ddCTP		8		
10mM ddGTP			16	
10mM ddATP				5.8
TE	1020	1000	1015	500

2.20 E.coli colony hybridisation

A nitrocellulose disc was placed on the surface of an agar plate and using sterile wooden toothpicks, bacteria were spotted onto the nitrocellulose and onto a master plate in an identical grid array. Both plates were incubated overnight at 37°C until colonies were 2mm-3mm in diameter.   Bacteria were lysed by placing the nitrocellulose, colony-side up, on 4 thicknesses of Whatman 3MM paper saturated with 1.5M NaCl, 0.5M NaOH. After 5 minutes the nitrocellulose filter was transferred to a stack of 3MM paper saturated with 1.5M

NaCl, 0.5M Tris.HCl pH8.0. After another 5 minutes the filter was placed on a stack of 3MM paper soaked with 3X SSC for 5 to 10 minutes. The damp filter was then sandwiched between 2 dry sheets of 3MM paper and compressed for 20 seconds. Bacterial matter adhered to the paper as it was carefully peeled away, leaving the nitrocellulose relatively clean. The filter was air dried for 1 hour then baked for 2 hours at 80°C under vacuum. The baked filter was wetted with 6X SSC then prehybridised for 2 hours at 65°C with gentle agitation in a 10cm x 10cm plastic box containing 10ml of 6X SSC, 5X Denhardts solution, 100µg/ml denatured salmon sperm DNA. A radioactively-labelled DNA probe was then added and hybridisation continued for 16 hours at 65°C. The filter was washed with 2 changes of 4X SSC for 5 minutes each at room temperature then 4 changes of 4X SSC for 30 minutes each at 65°C. After blotting dry, the filter was encased in Saran Wrap and applied to X-ray film (Fuji RX) with an intensifying screen for 2 hours.

20X SSC: 3M NaCl, 0.3M sodium citrate

50X Denhardts solution: 1% Ficoll, 1% polyvinylpyrollidone, 1% bovine serum albumin. Stored at -20°C.

Double stranded DNA probes were labelled to a specific activity of 10^9 cpm/µg using random hexanucleotide primer extension with α^{32}P-dCTP (Amersham, 3000Ci/mMol) and Klenow (Feinberg and Vogelstein 1982), as described by Dalgleish (1987).

Salmon sperm DNA (Type III, Sigma) was dissolved in Q water at 10mg/ml, sheared by passing it 10 times through an 18-gauge hypodermic needle and stored at -20°C.

Radioactive probes and sheared salmon sperm DNA were denatured by boiling for 7 minutes, then cooled rapidly on ice before being added to the hybridisation solutions.

2.21 DNA dot blotting

A sheet of nitrocellulose of the required size was wetted in water for 5 minutes, then soaked in 10X SSC for 10 minutes and dried thoroughly. DNA samples dissolved in TE pH8.0 were applied to the nitrocellulose in 5μl aliquots and allowed to dry at room temperature. The DNA was denatured by placing the nitrocellulose filter, application-side up, on a series of pads (4 sheets thick) of Whatman 3MM paper saturated with the following solutions: 1.5M NaCl, 0.5M NaOH (for 5 minutes), 0.5M Tris.HCl pH7.5 (for 30 seconds), 1.5M NaCl, 0.5M Tris.HCl pH7.5 (for 5 minutes), 3X SSC (for 5 to 10 minutes). The filter was dried at room temperature for 1 hour then baked at 80°C for 2 hours under vacuum. The baked filter was rewetted for 5 minutes in 6X SSC and prehybridised for 2 to 4 hours at 65°C in a sealed plastic bag containing 5ml of 6X SSC, 5X Denhardts solution, 0.5% SDS, 100μg/ml salmon sperm DNA. Hybridisation was performed for 24 hours at 65°C in a sealed bag containing 3-4ml of 6X SSC, 5X Denhardts solution, 0.5% SDS, 10mM Na$_2$EDTA, 100μg/ml salmon sperm DNA and a radioactive probe (prepared as in section 2.20). The filter was washed with 3 changes of 2X SSC for 10 minutes each at room temperature then with 4 changes of 0.1X SSC, 0.1% SDS for 30 minutes each at 65°C. After blotting dry, the filter was wrapped in Saran Wrap and exposed to X-ray film (Fuji XR). A 48 hour exposure with an intensification screen at -70^oC was required to detect single copy sequences in 5μg of tobacco DNA.

2.22 Southern blotting

Southern blot analysis of bacterial plasmid and yeast DNA was performed as described (Maniatis *et al* 1982). For plant genomic DNA an alternative method was preferred (Dalgleish 1987).

2.23 Synthesis and purification of oligonucleotides

Oligonucleotide synthesis was performed by J Keyte according to the method of Matthes *et al* (1984) as modified by Brenner and Shaw (1985). The crude oligo preparation was brought to a volume of 90μl with TE pH8.0. 10μl of 2.5M sodium acetate pH5.0 and 200μl of ethanol were added and the oligos precipitated at -70°C for 3 hours. DNA was pelleted by centrifugation for 20 minutes in a microfuge, dried in air for 15 minutes and resuspended in 10μl of Q water. 2μl of the oligo solution was mixed with 2μl of formamide dye (see section 2.19), boiled for 2 minutes then chilled in iced water and loaded on a 20% sequencing gel. The gel was prepared as described in section 2.19 but used a 20% gel stock (460g urea, 100ml 10X TBE and 500ml acrylamide stock per litre). Electrophoresis was for 3 hours at 22 to 25W until the bromophenol blue reached the bottom of the gel. The glass plates were then removed and the gel sandwiched between 2 sheets of Saran Wrap. When a 20cm x 20cm TLC plate (Macherey-Nagel) was placed behind the gel, DNA bands could be visualised as dark areas or 'shadows' upon illumination with a short-wave UV lamp. Having marked the position of the major (longest) oligonucleotide species with a fibre-tip pen, that section of the gel was cut out and placed in a Eppendorf tube. The gel slice was covered with 100μl of Q water and left on the bench overnight. The eluted oligonucleotide DNA was recovered by spun column chromatography through 1ml of Sephadex G-25 (Pharmacia) equilibrated

with Q water (see Maniatis *et al* 1982 for the procedure) and stored at -20°C. The DNA concentration was determined by measuring the A_{260nm} of a 1/30 dilution. (An A_{260} of 1 corresponds to $20\mu g/ml$).

2.24 Preparation of RNA from yeast

Cells were harvested from a 200ml culture grown to late log phase (2×10^7 cells/ml) in minimal medium (5 mins, 5krpm in GSA rotor) and suspended in 20ml of Q water. After repelleting (5 mins, 5krpm in SS34 rotor) the cells were suspended in 5ml of guanidinium thiocyanate solution (4M guanidinium thiocyanate, 5mM sodium citrate pH7.0, 0.5% sodium lauroyl sarkosine, 0.1M β-mercaptoethanol) and sterile acid-washed glass beads (0.5mm diameter) were added to a level just below the meniscus. The tube was firmly capped and vortexed until 90% or more of the cells were disrupted. 6ml of phenol/choroform (1:1) was added, the tube was shaken for 15 minutes then centrifuged (10 mins at 12krpm in SS34). The aqueous phase was transferred to a sterile glass universal bottle and 1g of solid CsCl was added for every 2.5ml. When this had dissolved, the solution was carefully layered over a 3ml cushion of 5.7M CsCl, 0.1M Na_2EDTA pH7.5* in an ultracentrifuge tube (Sorvall 03954). The tube was topped up with water and a balance tube was prepared if necessary. Samples were spun in a TST41.14 swinging bucket rotor at 20°C, 25krpm for 24 hours in a Sorvall OTD65B ultracentrifuge. When the tube had been thoroughly drained of supernatant, the RNA pellet was dispersed in 1ml of Q water* and transferred to a siliconised 30ml Corex tube sterilised by baking. A further 4ml of Q water* was added and the RNA allowed to dissolve fully, warming to 42°C if necessary. 0.5ml of 3M sodium acetate pH5.0 and 12ml of ethanol were added and the RNA precipitated overnight at -20°C. RNA was pelleted (20mins, 10krpm in SS34), rinsed with 70%

ethanol, dried and redissolved in Q water*. Stocks were adjusted to 5mg/ml and stored at -20°C.

* Treated with 0.1% DEPC for 2 to 16 hours at 37°C then autoclaved.

2.25 Preparation of RNA from tobacco plants

10g of leaf tissue (mid ribs removed) was frozen in liquid nitrogen and ground to a fine powder in a mortar and pestle. While still frozen, the powder was tipped into a second mortar containing 6ml of guanidinium thiocyanate solution (see section 2.24) where it was allowed to thaw with continued grinding. Particulate material was removed by centrifugation (10 mins at 10krpm in SS34 rotor) and the soluble extract (9-10ml) transferred to a centrifuge tube containing 10ml of phenol/chloroform/iso amyl alcohol (24:24:1). After mixing for 15 minutes the tube was spun for 10minutes at 10krpm in an SS34 rotor. The aqueous phase was removed and stored in a sterile glass universal bottle for up to 3 hours before being layered over a 2ml cushion of 5.7M CsCl, 0.1M Na$_2$EDTA pH7.5 in an ultracentrifuge tube (Sorvall 03954). A balance tube was prepared if necessary and the samples were spun in a TST41.14 swinging bucket rotor at 20°C, 25krpm for 24 hours in a Sorvall OTD65B ultracentrifuge. The supernatant was carefully removed with a pasteur pipette and discarded. The RNA pellet was dispersed/dissolved in 0.5ml of DEPC-treated Q water and transferred to a 15ml Corex tube, where a further 1.5ml of Q water was added. When the pellet had dissolved fully, 0.2ml of sodium acetate pH5.0 and 6ml of ethanol were added and the tube placed at -20°C overnight. RNA was pelleted (20 mins, 10krpm in SS34 rotor), rinsed with 70% ethanol, dried and redissolved in DEPC-treated Q water. Stocks were adjusted to 5mg/ml and stored at -20°C.

2.26 Selection of poly(A)$^+$ RNA

Polyadenylated RNAs were separated from non-polyadenylated RNAs by one round of chromatography through oligo(dT)-cellulose (Pharmacia, Type7) using a standard procedure (Maniatis *et al* 1982).

2.27 Electrophoresis of RNA in formaldehyde-agarose gels

To prepare 100ml of agarose, 1.4g of agarose (Seakem ME) was dissolved in 65ml of water and allowed to cool slightly. 20ml of 5X MOPS buffer (0.1M MOPS pH7.0, 25mM sodium acetate, 5mM Na$_2$EDTA autoclaved with 0.1% DEPC) and 16.7ml formaldehyde were then mixed in thoroughly and the gel was poured in a fume hood. RNA samples (up to 10μg) were denatured for 10 minutes at 65°C in a buffer consisting of 1X MOPS, 50% deionized formamide (section 2.19), 16.7% formaldehyde. Samples were applied to the gel with 0.1 volume of sterile loading buffer (50% glycerol, 1mM Na$_2$EDTA, 0.4% bromophenol blue, 0.4% xylene cyanol). Reservoir buffer was 1X MOPS, 16.7% formaldehyde and electrophoresis was performed overnight at 30V in a fume hood.

RNA was stained with ethidium bromide by soaking the gel in several changes of dH20 for at least 2 hours, staining for 1 hour in 1mg/ml ethidium bromide, then destaining in dH20 for at least an hour before a photograph was taken. Better contrast was obtained with longer destaining *eg* overnight.

2.28 Northern blotting

RNA was fractionated in a formaldehyde gel. Additional tracks containing total RNA were usually included and these were cut away from

the rest of the gel, stained with ethidium bromide as described above, and photo graphed beside a ruler to provide rRNA size markers. RNA from the rest of the gel was transferred to GeneScreen nylon membrane (NEN) by capillary blotting as follows. The gel was placed on 2 layers of Whatman 3MM paper on top of 2 flat sponges (Spontex) soaking in a dish of transfer buffer (25mM Na_2HPO_4/NaH_2PO_4). A sheet of GeneScreen cut to the size of the gel was soaked in transfer buffer for 15 minutes then placed on the gel and covered with 2 sheets of 3MM paper and a stack of absorbant paper towels. A weight (eg a glass plate) was applied to the top and transfer was allowed to proceed for 16 hours or more. The membrane was rinsed in transfer buffer, baked for 2 hours at 80°C and prehybridised at 42°C for 16 hours in a sealed plastic bag containing 5ml of 50% deionized formamide (section 2.19), 2X Denhardts solution, 5X SSC, 1% SDS, 100μg/ml salmon sperm DNA (section 2.20). Hybridisation was for 48 hours at 42°C in a sealed bag containing 3-4ml of 50% deionized formamide, 1X Denhardts solution, 5X SSC, 1% SDS, 100 μg/ml salmon sperm DNA and a radioactive probe (section 2.20). The membrane was washed with 2 changes of 2X SSC for 5 minutes each at room temperature, 2 changes of 2X SSC, 1% SDS for 30 minutes each at 65°C then 2 changes of 0.1X SSC at room temperature for 30 minutes. The membrane was blotted dry, wrapped in Saran Wrap and exposed to X-ray film (Fuji RX) with an intensification screen at -70°C.

2.29 RNA dot blotting

A sheet of nitrocellulose was wetted with water, soaked in 10X SSC for 10 minutes and allowed to dry thoroughly. RNA samples were denatured in 20μl of 1X MOPS (section 2.27), 16.7% formaldehyde by heating to 65°C for 10 minutes. 5μl of 5M sodium acetate pH7.0 was added and 5μl of each sample was applied to the nitrocellulose. The remainder of the

samples was stored at -20°C and denatured again before use. The filter was dried in air for 1 hour then baked for 2 hours under vacuum. After rewetting in 5X SSC, the filter was prehybridised for 3 to 4 hours at 42°C in a sealed bag containing 5ml of 50% deionized formamide, 5X SSC, 5X Denhardts solution, $100\mu g/ml$ denatured salmon sperm DNA. Hybridisation was for 16 hours at 42°C in a sealed bag containing 3-4ml of 50% deionized formamide, 5X SSC, 1X Denhardts solution, 0.3% SDS, $100\mu g/ml$ salmon sperm DNA and a radioactive probe (section 2.20). The filter was washed with at least 4 changes of 3X SSC, 0.1% SDS at 65°C for 30 minutes each, then wrapped in Saran Wrap and exposed to X-ray film (Fuji RX) with an intensifying screen at -70°C.

2.30 Primer extension analysis

The oligonucleotide primer was 5'phosphorylated with ^{32}P in a $10\mu l$ reaction containing 10pmol of primer (53ng for a 16mer), $50\mu Ci$ of $\varsigma^{32}P$-ATP (Amersham, 3000Ci/mMol), 10 units of T4 kinase and $1\mu l$ of 10X kinase buffer (see below). The reaction was incubated at 37°C for 30 minutes then stopped by the addition of $2\mu l$ of 0.5M Na_2EDTA. The solution was diluted to $100\mu l$ with Q water, vortexed with an equal volume of phenol/chloroform (1:1) and spun for 5 minutes in a microfuge. The aqueous phase was applied to a 1ml spun column of Sephadex G-25 (Pharmacia) equilibrated with Q water (see Maniatis *et al* 1982 for the procedure). The void volume was transferred to an Eppendorf tube and evaporated to dryness in a freeze drier. The radioactively-labelled primer was resuspended in $10\mu l$ of Q water and stored overnight at -20°C.

Prior to the extension reaction with reverse transcriptase, the primer (1pmol) and RNA (up to $50\mu g$) were annealed for 1 hour at 42°C in $20\mu l$

68

of 1X RT buffer (see below), 25mM DTT containing 40 units of RNasin

(Promega Biotec). The solution was adjusted to 50μl of 1X RT buffer,

10mM DTT, 40 units RNasin, 1mM dATP, 1mM dTTP, 1mM dCTP, 1mM dGTP,

2.5μg/ml actinomycin D, 10 units of AMV reverse transcriptase

(Pharmacia) and incubation at 42°C continued for 2 hours. 1μl of 10%

SDS and 5μl 0.5M Na_2EDTA were added to terminate the reaction. RNA was

hydrolysed by boiling the reaction for 5 minutes with 30μl of Q water

and 20μl of 1M NaOH. When it had cooled to room temperature, the

solution was neutralised with 1M HCl. 5μg of carrier RNA and 250μl of

ethanol were added and cDNA products precipitated at -70°C for 30

minutes. After spinning for 5 minutes in a microfuge, the pellet was

dried and resuspended in 5μl of 50% formamide dye (section 2.19).

Samples were boiled for 2 minutes, chilled on ice and then analysed by

electrophoresis in 10% polyacrylamide/urea gels (see section 2.19 for

preparation of the gel; 10% gel stock consists of 460g urea, 100ml 10X

TBE and 250ml acrylamide stock per litre). Gels were run at 25W until

the bromophenol blue had run off the bottom, dried down and exposed to

X-ray film (Fuji RX) at -70°C with an intensifying screen. *nb*

unextended primer (16-mer) runs above bromophenol blue in 10% gels

10X RT buffer: 0.5M Tris.HCl pH8.3, 0.5M NaCl, 80mM $MgCl_2$

10X kinase buffer: 0.5M Tris.HCl pH7.6, 0.1M $MgCl_2$, 50mM DTT, 1mM

spermidine, 1mM Na_2EDTA.

2.31 Preparation of total cell extracts from yeast

Method 1. Cells were harvested (5 mins, 5krpm) from a 50ml culture

grown to late log phase in minimal medium, washed in 5ml of Q water and

resuspended in 350μl of TE pH7.5, 1mM PMSF in an Eppendorf tube. 0.75

volume of sterile acid-washed glass beads were added and the tube alternately vortexed for 30 seconds and placed on ice for 30 seconds until greater than 90% of the cells were disrupted. The extract was diluted with 0.5ml of ice-cold TE pH7.5, 1mM PMSF and centrifuged for 10 mins in a microfuge at 4°C. The supernatant was stored at -20°C and the protein content estimated using a Bio-Rad protein assay (based on the method of Bradford 1976) as follows. 1μl of the extract diluted in 0.8ml of Q water was mixed with 0.2ml of dye reagent concentrate (Bio-Rad). After 10 to 30 minutes, the OD_{595} was recorded against a reference of 0.8ml water mixed with 0.2ml dye reagent concentrate. A calibration curve of protein concentration (0-20μg/ml) versus OD_{595} was prepared using BSA.

Method 2.   Cells were harvested (5 mins, 5krpm) from 10ml cultures grown to late log phase in YPD, washed with 5ml of cold CAT buffer (50mM Tris.HCl pH7.5, 100mM NaCl, 0.1mM Na_2EDTA, 0.1mM chloramphenicol, 1mM PMSF) and resuspended in 150ml of ice-cold CAT buffer in an Eppendorf tube.   0.5 volume of sterile acid-washed glass beads was added and the tube alternately vortexed for 30 seconds and placed on ice for 30 seconds until at least 90% of the cells were disrupted. 50μl of ice-cold CAT buffer was added and gross cell debris removed by centrifugation (5 mins, 5krpm, 4°C). The supernatant was transferred to a clean Eppendorf tube and smaller debris was pelleted by a subsequent 5 minute spin in a microfuge at 4°C. The extract was stored at -20°C and protein concentration was estimated as described in method 1.

2.32 Polyacrylamide gel electrophoresis of proteins

The preparation, electrophoresis and Coomassie blue staining of protein gels was as described (Hames 1981). For dissociating gels, a discontinuous buffer system based on the method of Laemmli was used. Non-dissociating gels also used a Tris-glycine discontinuous buffer system.

2.33 Chloramphenicol acetyltransferase (CAT) assay

Yeast cell extracts prepared by method 2 (section 2.31) or tobacco cell extracts were added to an Eppendorf tube containing $17.5\mu l$ of Tris.HCl pH7.8, $1\mu l$ of ^{14}C-chloramphenicol (Amersham, 53mCi/mMol) and $10\mu l$ of 10mM acetyl coenzyme A (stored at -20^oC) and the volume brought to $150\mu l$ with Q water. Reactions were for 60 to 90 minutes at 28 to 30^oC. ^{14}C-chloramphenicol and its acetylated derivatives were extracted by vortexing the reaction with 1ml of ethyl acetate for 30 seconds. After spinning for 5 minutes in a microfuge, the organic phase was transferred to a fresh 1.5ml Eppendorf tube and evaporated to dryness in a freeze drier. The chloramphenicol was resuspended in $20\mu l$ of ethyl acetate and spotted onto a silica gel TLC plate (Macherey Nagel, obtained from Camlab). A chromatography tank was lined with filter paper, filled with a solution of chloroform/methanol (95:5) to a depth of 0.5cm and left to equilibrate for 15 minutes. The TLC plate was inserted vertically into the tank and subjected to ascending chromatography until the solvent front was 5mm from the top (10 to 15 minutes). The plate was dried in air for 20 minutes and exposed to X-ray film (Kodak AR) for 24 hours.

2.34 Measurement of β-galactosidase activity

a)   Plate assay.   Yeast strains were patch plated on SD agar buffered at pH7.0 containing 100μg/ml X gal.   To prepare the buffered medium 13.6g KH_2PO_4, 2g $(NH_4)_2SO_4$, 4.2g KOH, 0.2g $MgSO_4$, 0.5mg $FeCl_3$, 6.7g yeast nitrogen base and 20g glucose were dissolved in 500ml of Q water, filtered then filter sterilised.   100ml was mixed with an equal volume of molten sterile 4% water agar and 2ml of X gal (2% in dimethylformamide, stored at $-20^{\circ}C$) was added before pouring.

b)   ONPG assay of yeast cells (adapted from Miller 1972).   10ml cultures in minimal medium were grown to at least 10^7cells/ml and chilled on ice for 20 minutes while the cell density was counted accurately using a haemocytometer (Neubauer).   Cells were harvested in a bench centrifuge (5 mins, 5krpm) and resuspended in 1ml of 100mM KH_2PO_4 pH7.0.   Z buffer was prepared by adding 100μl of PMSF (100mM in ethanol, stored $-20^{\circ}C$), 25μl of 10% SDS and 27μl of β-mercaptoethanol to 10ml of Z buffer base (per liter: 1.6g Na_2HPO_4, 5.5g NaH_2PO_4, 0.75g KCl, 0.264g $MgSO_4$, pH7.0).   0.4ml of cells was vortexed for 5 seconds with 0.4ml of Z buffer and left at room temperature for 10 minutes. 30μl of chloroform was added, the tube vortexed for 5 seconds then placed in a $28^{\circ}C$ water bath for 10 minutes.   The reaction was started by adding 0.2ml of ONPG (4mg/ml in 100mM KH_2PO_4 pH7.0, warmed to $28^{\circ}C$). The tube was vortexed briefly and incubation at $28^{\circ}C$ continued until a yellow colour developed.   The reaction was terminated by the addition of 0.5ml of 1M Na_2CO_3, cells were removed by pelleting (2 mins in a microfuge) and OD_{420} of the supernatant was recorded in a spectrophotometer.   β-galactosidase activity was calculated using the following formula:

$$\text{activity} = \frac{OD_{420}}{t \times v \times d \times 0.0045} \quad \text{nmol ONPG converted/ml/min/}10^7 \text{cells}$$

where t = reaction time in minutes

v = volume of the (unconcentrated) culture sampled (ml)

d = density (cells/ml x 10^{-7}) of the unconcentrated culture

nb 1nmol/ml ONPG converted is equivalent to OD_{420} of 0.0045 with a 10mm light path (Miller 1972).

c) ONPG assay of Brassica protoplast extracts. 0.5 to 1×10^6 protoplasts were pelleted in a bench centrifuge (5 mins at 500rpm), transferred to an Eppendorf tube and 100μl of 2X extraction buffer (200mM KH_2PO_4 pH7.0, 10mM β-mercaptoethanol) was added. Cell extracts were prepared by vortexing the protoplast suspension for 10 seconds, either immediately or after storage of the protoplasts for up to 3 weeks at -20°C. Cell debris was removed by centrifuging (5 minutes in a microfuge) and the supernatant was transferred to a new tube and stored on ice until the assay was performed. The protein concentration of the extract was estimated as described in section 2.31.

Since some extracts were slightly yellow, enzyme assays were carried out in duplicate, one reaction being performed in the presence of ONPG and one in the absence of β-galactosidase substrate. For each reaction, 50μl of extract was added to 750μl of Z buffer (prepared as in section (b) but omitting the SDS) and placed in a 37°C water bath for 10 minutes to warm up. 0.2ml of ONPG (4mg/ml in 100mM KH_2PO_4 pH7.0, warmed to 37°C) was added to the test sample, vortexed briefly and incubation at 37°C was continued until sufficient yellow colour development had occurred. 0.2ml of 100mM KH_2PO_4 warmed to 37°C was

added to the control sample. The reactions were terminated by the addition of 0.5ml 1M Na_2CO_3 and OD_{420} of each sample was measured in a spectrophotometer. β-galactosidase activity was calculated using the following formula:

$$activity = \frac{OD_{420}(test) - OD_{420}(control)}{p \times t \times 0.0045} \ nmol \ ONPG \ converted/ml/mg/min$$

where p = mg total protein used per assay

t = reaction time in minutes

2.35 Neomycin phosphotransferase (NPT) II assay

100mg of tobacco leaf tissue was macerated at $0^{\circ}C$ with the tip of a glass capillary in an Eppendorf tube containing $20\mu l$ of extraction buffer (40mM Na_2EDTA, 150mM NaCl, 100mM NH_4Cl, 10mM Tris.HCl pH7.5, 15mM DTT, 1mM PMSF, 0.12mg/ml leupeptin). The tube was spun for 5 minutes and the supernatant ($40\mu l$ approx) transferred to a tube containing $2\mu l$ of 0.05% bromophenol blue. The extract was subjected to electrophoresis in a 10% polyacrylamide non-denaturing gel (section 2.32) at 150V, $4^{\circ}C$ until the bromophenol blue was 1cm from the bottom (approximately 5 hours). The resolving gel was equilibrated in 2X activity buffer (67mM Tris.malate pH7.1, 42mM $MgCl_2$, 0.4M NH_4Cl) for 15 minutes at $4^{\circ}C$, then drained and placed a glass plate at room temperature. The glass plate was edged with magic tape (Scotch) to contain any radioactive spillages. 20ml of molten $40^{\circ}C$ agarose (1% low gelling temperature agarose in 1X activity buffer) containing $100\mu Ci$ $\varsigma^{32}ATP$ (Amersham, 3000Ci/mMol) and $40\mu g/ml$ kanamycin sulphate (50mg/ml stock, stored at $-20^{\circ}C$) was very carefully pipetted over the protein gel until it was evenly covered. After 30 minutes the NPT II reaction

product, ^{32}P-phosphorylated kanamycin, was transferred to Whatman P81 ion exchange paper by capillary blotting as follows. A sheet of P81 paper was placed over the agarose, covered with a small stack of Whatman 3MM paper and a weight was applied to the top. After 3 hours transfer, the P81 paper was washed with 5 changes of water at 90°C for 5 minutes each, then with 10mM Na_2HPO_4/NaH_2PO_4 pH7.0 at 65°C for 30 minutes. The paper was wrapped in Saran Wrap and exposed to X-ray film overnight at -70°C with an intensifying screen.

2.36 Isolation and transformation of Brassica protoplasts

Protoplasts were isolated from turnip leaves (Brassica rapa L cv Just Right) by Dr C Woolston according to a published procedure (Ulrich et al 1980) with the following modifications. i) leaf pieces were plasmolysed prior to enzymic treatment by floating, lower epidermis downward, on 0.4M mannitol at 25°C for 1 hour with gentle agitation, ii) the enzyme solution was 2% cellulase, 0.5% macerozyme in 0.4M mannitol, iii) cells were washed with 0.4M mannitol, 1mM $CaCl_2$.

Freshly prepared protoplasts were transformed with DNA purified from a CsCl gradient (section 2.14) using a PEG-stimulated DNA uptake method (Maule 1983). The cells were cultured at 10^6 cells/ml in Km8p medium (Kao and Michayluk 1975) containing 500μg/ml carbenicillin and 6μg/ml nystatin at 23°C with continuous illumination.

2.37 Triparental mating

Overnight cultures of E.coli NM522 harbouring Bin19 or a related plasmid and E.coli HB101/pRK2013 (both kan^R, rif^S) were grown at 37°C in L broth, and a culture of A.tumefaciens LBA4404 (kan^S, rif^R) was

grown at 30°C in N broth. 200µl from each culture were combined in an Eppendorf tube, spread on a L agar plate and incubated overnight at 26°C. A loopful of cells from the resulting lawn was streaked on T agar containing kanamycin and rifampicin at 50µg/ml and incubated at 26°C for 3 days. Single colonies were restreaked twice.

2.38 A.tumefaciens-mediated transformation of tobacco leaf discs and regeneration of whole plants

Foreign DNA was introduced into Nicotiana tabacum cv Petit Havana clone SR1 (Maliga et al 1973) via A.tumefaciens, using a binary vector transformation system (Bevan 1984). To prepare leaf discs, a healthy young tobacco leaf was surface sterilised by immersing in 5% sodium hypochlorite for 5 minutes, then in 70% ethanol for 5 minutes and rinsed in 3 changes of sterile Q water. The leaf was placed on a ceramic tile that had been swabbed down with 70% ethanol and 0.75cm diameter discs were cut out with a sterile cork borer. These were incubated, lower epidermis down, on regeneration medium (see below) for 48 hours. All tissue culture material was kept at 23°C in a plant growth cabinet (Leec) continuously illuminated by warm white fluorescent tubes. A 10ml culture of A.tumefaciens LBA4404 carrying the foreign DNA inserted in plasmid Bin 19 was grown overnight at 28°C in N broth containing 50µg/ml kanamycin. Cells were pelleted, resuspended in 20ml of liquid regeneration medium and 10ml was dispensed into each of 2 petri dishes. 20 leaf discs were placed in each dish and floated on the Agrobacterium suspension for 48 hours with gentle agitation. After that time, the discs were transferred to regeneration medium containing kanamycin and carbenicillin. Discs were replated on fresh medium after 7 days. Callus formation was evident after another week and after a total of 4 weeks culture on regeneration

medium, callus tissue was detached from the leaf disc and transferred to shooting medium (see below). Shoots appeared after 3 to 6 weeks and when large enough to handle easily with forceps (1cm long) these were cut away from the callus and embedded in rooting medium (see below). Roots began to appear after 10 to 20 days. After a further 2 or 3 weeks the plantlets, now about 5cm tall, were potted up in compost and incubated for 1 week in a propagator inside an environmental growth cabinet (Fisons) maintained at 26°C and illuminated over a 16 hour photoperiod with warm white fluorescent tubes. Transformed plants were kept for another 2 weeks in the growth cabinet, then repotted in 10cm x 10cm disposable plant pots and removed to a temperature and humidity-controlled glasshouse where they were grown to maturity.

regeneration medium: (per litre) 4.71g MS salts (Flow Laboratories), 30g sucrose, 0.7% agarose (litex), 0.5mg NAA, 0.1mg kinetin, pH5.8

shooting medium: (per litre) 4.71g MS salts, 30g sucrose, 0.7% agarose, 0.1mg NAA, 1.0mg 6-BAP, pH5.8

rooting medium: (per litre) 4.71g MS salts, 30g sucrose, 0.7% agarose, pH5.8

Plant hormone and antibiotic stocks were as follows.

Compound	Stock conc^n, Storage temp.	Working conc^n
kanamycin sulphate	50mg/ml in water, -20°C	100μg/ml
carbenicillin	100mg/ml in water, -20°C	500μg/ml
NAA	1mg/ml in 50% ethanol, 4°C	see above
kinetin	1mg/ml in 0.1M HCl, 4°C	"
6-BAP	1mg/ml in 0.1M NAOH, 4°C	"

CHAPTER 3

TRANSCRIPTIONAL ANALYSIS OF THE CaMV GENOME IN YEAST

3.1 Introduction

As a preliminary measure in analysing the activity of CaMV promoters in yeast, a complete copy of the CaMV genome was introduced into yeast and northern blotting used to determine whether any portion of the viral DNA was transcribed. A YEp plasmid was chosen as the cloning vehicle since these are present in multiple copies which should facilitate the detection of CaMV-encoded RNAs.

3.2 Cloning of the CaMV genome into plasmid YEp213

The source of the CaMV DNA used throughout this study was plasmid pLW414, in which a complete genome of CaMV isolate CM4-184, linearised at a unique SalI restriction site within gene V (see Figure 1.3), is inserted at the SalI site of pBR322 (Howell *et al* 1981). This isolate contains a 421bp deletion within gene II which affects neither the pattern of transcription nor the viability of the virus in infected plants. To check that the viral DNA had not sustained any inactivating mutations as a consequence of being cloned, 10μg of plasmid pLW414 was restricted with SalI (liberating the viral DNA) and rubbed onto the leaf surface of a young turnip plant. Within 21 days the plant was exhibiting symptoms typical of systemic CaMV infection, confirming that the cloned viral DNA was fully infectious and therefore suitable material for the gene expression studies to be described in this and subsequent chapters. For preliminary analysis in yeast, the CaMV DNA was recovered from pLW414 and inserted into YEp213 as described in Figure 3.1. Resulting plasmids pJR1 and pJR2 contain the CaMV insert

in both of the possible orientations.

3.3 The stability of pJR plasmids in yeast

YEp plasmids are subject to instability for two reasons. Firstly, in the absence of selection, recombinant 2 micron plasmids are gradually lost from a dividing population of cells as a result of unequal segregation. Secondly, all YEp plasmids contain at least one of the 2 micron inverted repeat sequences (see Chapter 1, section 1.12 and Figure 1.1) and are therefore able to undergo site-specific recombination with the resident native 2 micron circle.

The segregational stability of pJR1 was monitored as described (Chapter 2 section 2.5) and the results are presented in Figure 3.2. Under non-selective growth conditions, pJR1 was found to be significantly less stable than the parent plasmid YEp213, being lost at a rate of 2% per generation. Since pJR1 and YEp213 are identical apart from the CaMV insert, the reduced stability of pJR1 in the absence of selection may be related to its larger size. Alternatively, specific sequences within the CaMV DNA might be deleterious to the yeast cell.

The impact of site-specific recombination on the stability of pJR plasmids was assessed by extracting the DNA from cultures which had been grown for 10 or more generations in selective medium and performing a Southern blot. Three independent pJR1 and pJR2 transformants were analysed and the result in Figure 3.3 indicates that in no case had the plasmid DNA undergone any discernible recombination, deletion or rearrangement. It should be noted, however, that the detection of rare structural variants represented at, say, one copy per cell, would require a longer exposure. While such molecules may well

be present, it is nevertheless clear that a vast majority of the plasmids are in their original configuration.

3.4 Identification of CaMV-encoded transcripts in yeast

Total RNA extracted from yeast cultures harbouring pJRl and YEp213 was examined for the presence of specific transcripts by northern blotting. Control samples probed for transcripts of the 2 micron C gene gave the expected pattern of bands (Figure 3.4A), and confirm the integrity of messenger RNA in these preparations. Using full length CaMV DNA as a probe, at least two transcripts of approximately 2kb and 3.5kb are evident in pJRl RNA but absent from the YEp213 sample (Figure 3.4B). To gain further insight into the origin of these CaMV-hybridising RNAs, additional tracks were probed with the bacterially-derived DNA sequences adjoining the 5' and 3' extremities of the CaMV genome in pJRl (ie the disrupted tet gene) (Figure 3.4C). This probe hybridises to a set of transcripts indistinguishable in size from those detected with the CaMV probe, hence all of these RNAs must span one or other of the CaMV/tet gene junctions and none is fully internal to the viral DNA. The tet gene probe also hybridises to YEp213 RNA but this is not unexpected since the transcription of bacterial plasmid DNA sequences in yeast has been previously documented (Marczynski and Jaehning 1985).

If yeast RNA polymerase were to correctly recognize and utilise the two CaMV promoters and RNA processing signal, two viral transcripts would be predicted: i) a 1.85kb RNA corresponding to the 19S transcript identified in infected plants and ii) a longer transcript of at least 5kb, initiating at the same position as 35S RNA (Chapter 1, Figure 1.3) but terminating at an undefined position within the plasmid DNA. From the results in Figure 3.4 B and C it can be deduced that the

CaMV-specific RNAs detected in yeast cells harbouring pJR1 do not correspond in size and/or location to these predicted transcripts. Although a transcript of approximately the same size as 19S RNA is synthesised in yeast, this 2kb transcript is partially transcribed from bacterial DNA sequences. The genuine 19S RNA would be fully internal to the CaMV DNA, hence this 2kb RNA cannot initiate at the same position as 19S RNA. None of the transcripts observed in Figure 3.4B is long enough to represent the other predicted RNA, which would be at least 5kb in length.

The CaMV-specific transcripts synthesized in yeast have two possible origins. Firstly, they may initiate within yeast or bacterial DNA sequences in pJR1 and extend into the adjacent CaMV DNA. Conversely, they may arise within CaMV DNA owing to the presence of a fortuitous yeast promoter (which does not coincide with the 19S or 35S promoter) and extend into nearby plasmid DNA sequences. The position and polarity of these RNAs could be deduced by probing with a series of shorter, single stranded CaMV DNA fragments. This line of research was not pursued, however, since from the data presented it is clear that the CaMV transcripts originate either from outside the viral DNA or else from fortuitous yeast promoters within the CaMV genome that are unrelated to the known viral promoters.

These results do not conclusively demonstrate that the CaMV 19S and 35S promoters are inactive in yeast, since transcripts from the promoters may be synthesized at levels below the threshold of detection, or alternatively terminate heterogeneously such that a discrete transcript band cannot be produced by gel electrophoresis. Experiments designed specifically to test the activity of these CaMV promoters in yeast will be described in the following chapters.

3.5 Summary

A complete CaMV genome was introduced into yeast on a YEp plasmid, pJR1. Under growth conditions in which 70% of the cells harbour the plasmid, several CaMV-hybridising transcripts were detected. These were not extensively characterised but were shown to span one of both of the CaMV/tet gene junctions in pJR1.

Figure 3.1 Construction of pJR1 and pJR2.

CaMV DNA was excised from plasmid pLW414 as a 7.6kb SalI fragment and ligated into the unique SalI site of YEp213.   pJR1 and pJR2 differ only in the orientation of the CaMV DNA.   The positions of the viral promoters are indicated (p19S and p35S).   S = SalI restriction site.
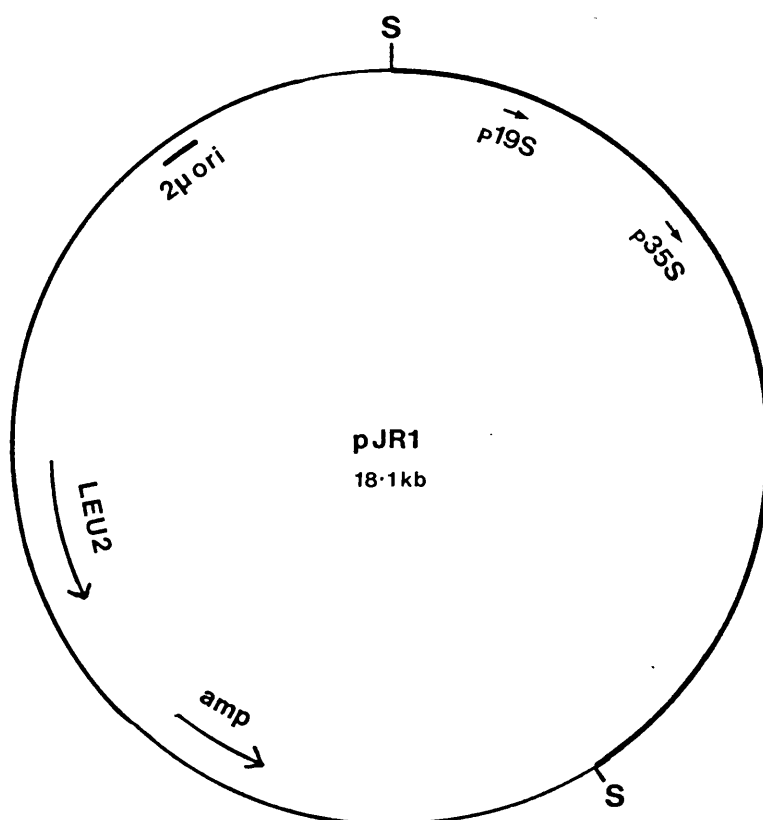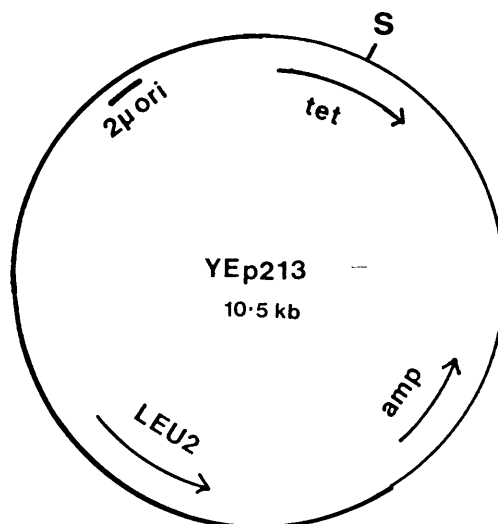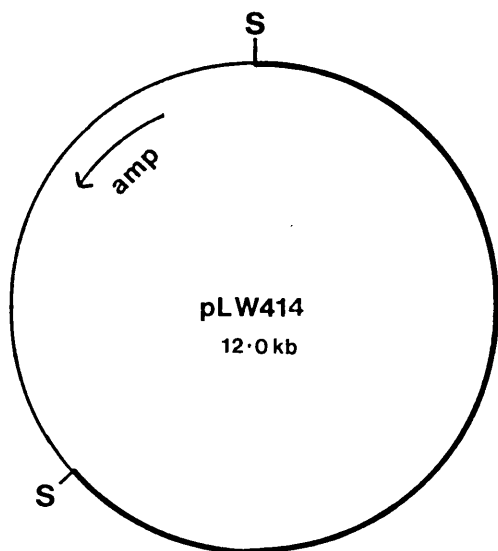
Figure 3.2   The stability of plasmids pJR1 (■) and YEP213 (□) in

yeast during continuous logarithmic culture in non-selective (A) and

plasmid- selective growth medium (B).
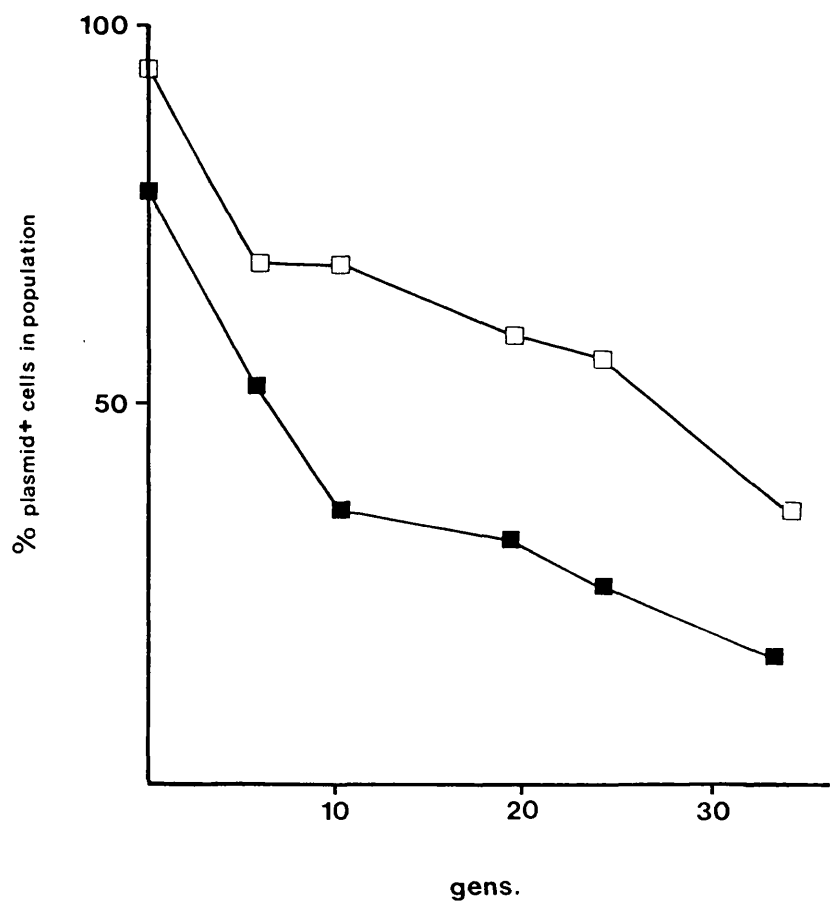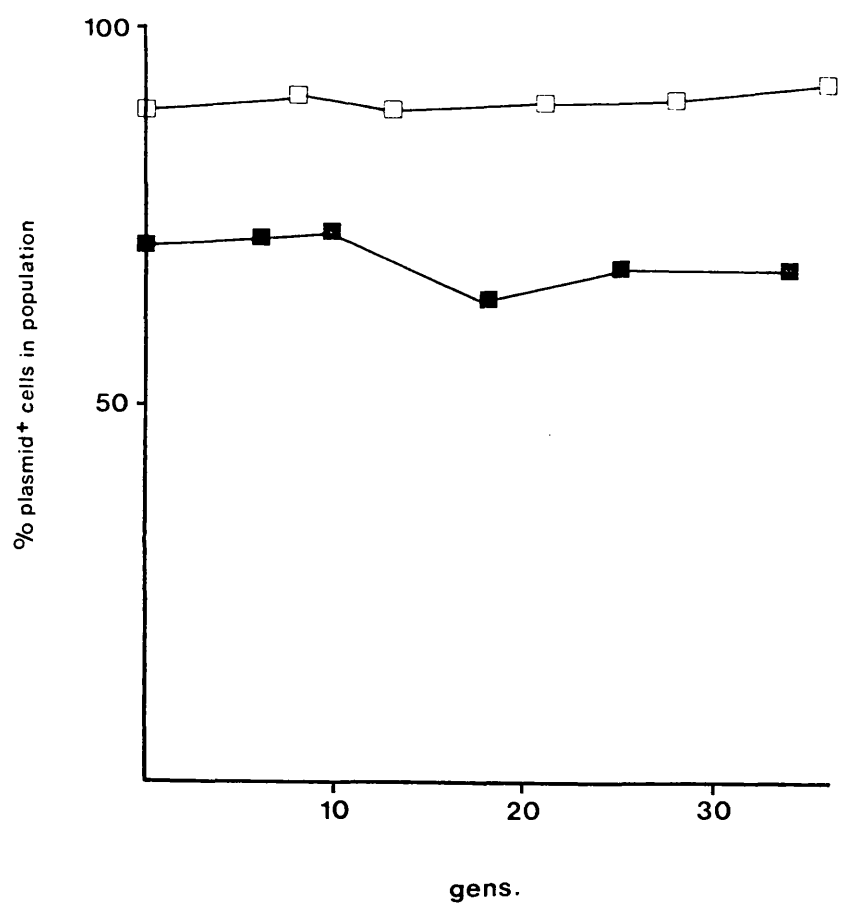
Figure 3.3 Southern blot analysis of yeast DNA.

10μl (1μg) of DNA was restricted with BamHl and probed with pLW414

DNA.

Lanes a and b, purified pJRl and pJR2 DNA restricted with BamHl

Lanes c-e, DNA from 3 independent S150-2B pJRl transformants

Lanes f-h, DNA from 3 independent S150-2B pJR2 transformants

Lane i, DNA from an S150-2B YEp213 transformant

Lane j, S150-2B DNA

Lanes a and b were exposed for 15 minutes and lanes c-j for 30

minutes, both with an intensifying screen.

a b c d e f g h i j
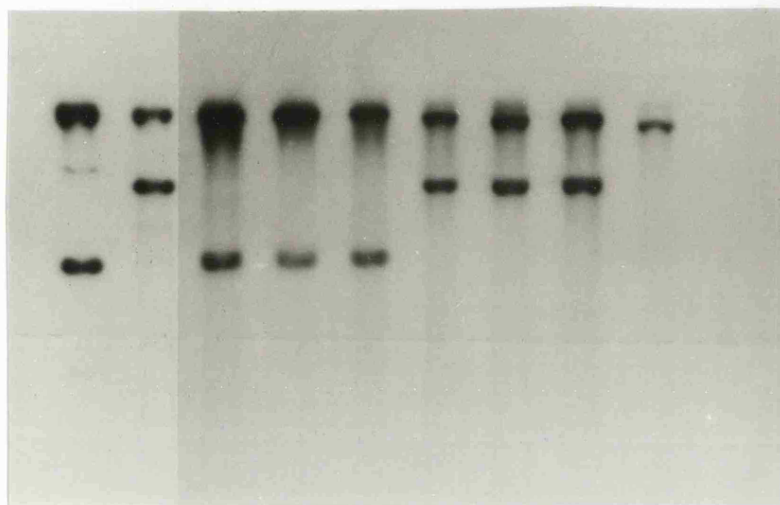
Figure 3.4 Northern blot analysis of yeast RNA.

10μg of total RNA from cells harbouring pJRl and YEp213 was probed with a restriction fragment of 2 micron DNA internal to the C gene (panel A), with CaMV DNA isolated as a 7.6kb SalI restriction fragment from pLW414 (panel B), or with the tet gene of pBR322 isolated as a HindIII–AvaI restriction fragment (panel C). Ethidium bromide-stained total RNA samples, showing the rRNA bands, are included for reference.

Autoradiographs were exposed for 3 days at -70°C with an intensifying screen.
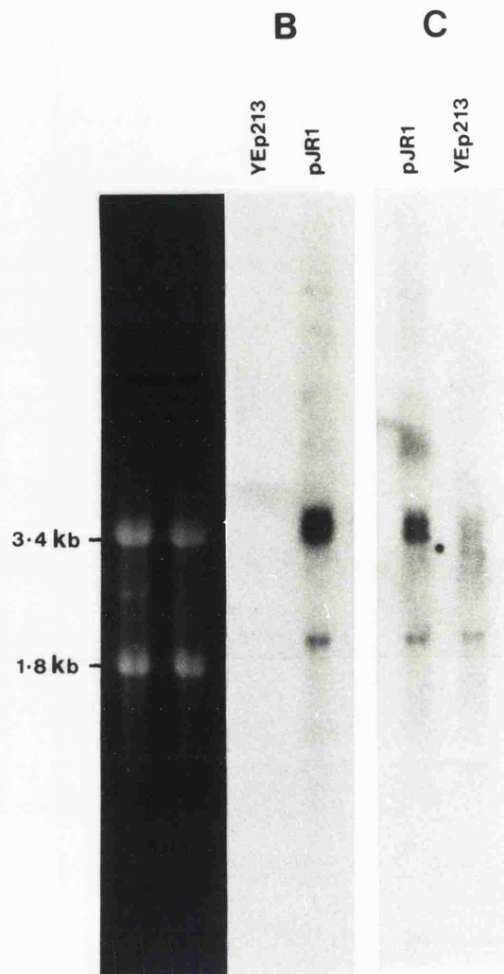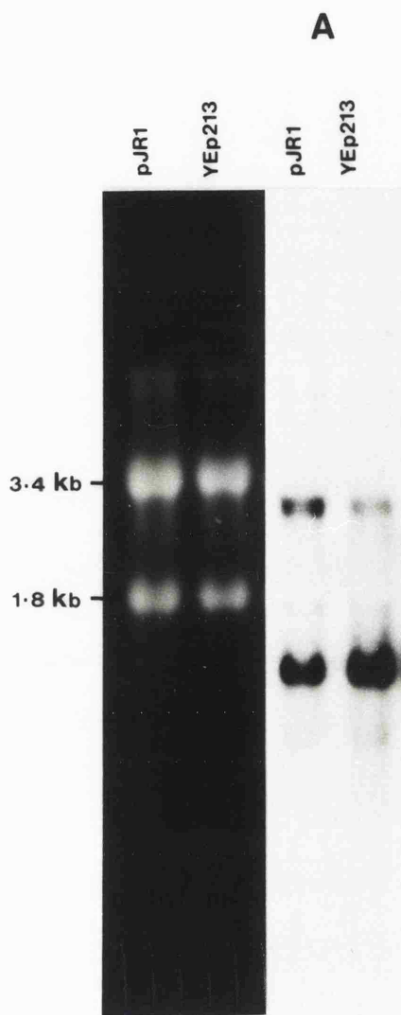
CHAPTER 4

ANALYSIS OF THE CaMV 19S PROMOTER IN YEAST USING GENE FUSION

4.1 Introduction

Gene fusion is a powerful tool that has been used to address a variety of biological problems, including protein localisation (Silhavy *et al* 1977) and the identification of functional protein domains (Brent and Ptashne 1985, Hope and Struhl 1986), but is most frequently used in studies of gene expression (West *et al* 1984, Timko *et al* 1985). In a gene fusion, the promoter (and possibly structural region) of one gene is fused to the structural part of a second easily assayed gene, the so-called reporter gene. The chief advantage of such an arrangement is that the reporter gene provides a readily detectable and quantifiable measure of gene activity, where the native gene product may confer no scorable or measurable phenotype upon the cell. There are several reporter genes in common usage including those encoding β-galactoside (Guarente and Ptashne 1981), chloramphenicol acetyltransferase (Herrera-Estrella *et al* 1984), neomycin phosphotransferase II (Werr and Lorz 1986) and galactokinase (Wright and Zitomer 1985). The lacZ gene of E.coli, which codes for β-galactosidase, is well suited to gene expression studies in yeast for a variety of reasons. Firstly, yeast has no endogenous β-galactosidase. Secondly, a range of β-galactosidase indicator media are available for both bacteria and yeast, hence lacZ provides a visible marker of gene expression. Thirdly, a biochemical assay for β-galactosidase has been devised which is one of the most sensitive enzyme assays available. Fourthly, the protein is tolerant of extensive modification at the N-terminus, making translational gene fusions feasible: up to 27 amino acids can be removed and replaced with another protein sequence without

substantially altering the specific activity of the enzyme (Brickman *et al* 1979). In this chapter I describe the construction and analysis of a gene fusion between lacZ and a subgenomic fragment of CaMV bearing the 19S promoter.

4.2 Construction of a plasmid for use in the investigation of CaMV promoter function in yeast

Figure 4.1 illustrates the construction of pPA, a YEp plasmid which contains a 5' truncated derivative of the lacZ gene (designated 'lacZ) from pMLB1034 (described in Silhavy *et al* 1984). The E.coli promoter and first eight codons are deleted in 'lacZ and replaced by a polylinker, into which putative promoter sequences can be inserted. The plasmid can then be introduced into yeast and β-galactosidase assays performed to monitor the level of gene expression. It should be noted that plasmid pPA is only suitable for generating translational gene fusions, since the 'lacZ gene is incomplete and cannot be translated unless it is first joined to the 5' end of another open reading frame.

4.3 Construction and *in vivo* analysis of a 19S-lacZ gene fusion

The restriction map of a 559bp Sau3A restriction fragment extending 54bp (18 codons) into CaMV gene VI and 505bp upstream is shown in Figure 4.2. A putative TATA box with the sequence TATATAAA is centred 30bp upstream of the cap site utilised in plants which is located 11bp from the start of the gene VI coding region (Covey *et al* 1981). It would seem reasonable, therefore, to assume that the 19S promoter is contained within this fragment. The Sau3A fragment was isolated from plasmid pLW414 and cloned into the BamHI site of pPA as described in

Figure 4.2b, creating plasmids pPA1 and pPA2. In pPA1 the 19S promoter fragment is inserted such that the first 18 triplets of CaMV gene VI are fused in frame to 'lacZ. Plasmid pPA2 contains the same fragment in the reverse orientation and the 'lacZ gene cannot be translated.

The activity of the 19S-lacZ gene in yeast was initially monitored by plating cells on medium containing the chromogenic indicator X-gal. X-gal is a lactose analogue which is hydrolysed by β-galactosidase to yield a blue compound. As shown in Figure 4.3, yeast cells harbouring pPA1 develop a blue colour in the presence of X-gal, demonstrating that the 19S-lacZ gene is expressed in yeast and encodes an active enzyme. Cells harbouring plasmids pPA and pPA2 remain white on this medium (not shown). S.cerevisiae was found to grow very poorly on the indicator medium, possibly because efficient colour development required that the agar be buffered at neutral pH whereas yeast grows optimally at a pH of about 6. For this reason, cells harbouring lacZ fusions were first grown on ordinary selective agar and then generously patch-plated onto the buffered agar containing X-gal to test for the presence of β-galactosidase.

It could not be deduced from the positive X-gal reaction whether expression of the 19S-lacZ gene was due to functional recognition of the CaMV promoter or whether transcription of the gene was dependent on yeast DNA sequences further upstream. A section of the yeast retrotransposon Ty-17 including one of the long terminal repeats (LTRs or δ sequences) is known to be situated downstream of the LEU2 gene; this element would thus be close to the 19S promoter in pPA1 - see footnote *** (Kingsman et al 1981, Andreadis et al 1982). As in the case of retroviral proviruses, the LTRs of active Ty elements contain a promoter which directs the synthesis of the major Ty transcript but is

also able to activate the transcription of cellular genes situated downstream (Scherer *et al* 1982, Roeder *et al* 1985).

To investigate whether expression of the 19S-lacZ gene was dependent upon, or strongly modulated by, adjacent vector sequences, two further plasmids were constructed. pPA4 is identical to pPA1 except that the EcoRI fragments bearing the LEU2 gene and 2 micron plasmid origin of replication are cloned in the reverse orientation with respect to the 19S promoter (Figure 4.4). The consequence of this is that the

19S-lacZ gene is now preceded by a section of 2 micron DNA and the Ty LTR is both displaced to a more distant position on the plasmid and reversed in orientation. In plasmid pPA5 the 3' end of the yeast phosphoglycerate kinase (PGK) gene has been inserted at a HindIII site within the CaMV Sau3A fragment (Figure 4.5). This section of the PGK gene contains an efficient transcription termination signal which should prevent any transcriptional readthrough from yeast sequences located upstream of the viral promoter.

When plated on medium containing X-gal, cells harbouring pPA4 and pPA5 develop a blue colour (Figure 4.3) demonstrating that the 19S-lacZ gene continues to be expressed in spite of the above modifications. This provides compelling evidence that expression of the gene occurs independently of the adjacent yeast DNA sequences and that mRNA synthesis initiates within the CaMV DNA.

*** This Ty element maps upstream of of the LEU2 gene in the original genomic clone, but in the course of constructing YEp213 a SalI-XhoI restriction fragment encompassing the gene has been inverted (see Broach 1983, where YEp213 is called YEp13S).

4.4 Quantitation of β-galactosidase levels

The level of β-galactosidase synthesis from pPA1 and other plasmids was quantified using a biochemical enzyme assay in which cells are exposed to the lactose analogue ONPG (o-nitrophenol-β-D-galactopyranoside). ONPG is hydrolysed by β-galactosidase yielding o-nitrophenol which is yellow in colour and can be measured with the aid of a spectrophotometer. In a preliminary experiment to ascertain the optimal growth conditions under which to test for β-galactosidase, a culture harbouring pPA1 was assayed at various time points during logarithmic and stationary phase growth. As shown in Figure 4.6, enzyme activity *per cell* was maximal during exponential growth and decreased gradually as the cells entered stationary phase but was still detectable 24 hours later. From this data, there appears to be no critical stage for sampling cells providing they have not entered stationary phase, and cultures were routinely harvested for analysis when they reached densities of 1 to 3 x 10^7 cells/ml.

The levels of β-galactosidase activity in cells containing pPA1 and other plasmids are given in Table 4.1. As expected, negligible enzyme activity is detected from the parent plasmid pPA which carries the untranslatable 'lacZ gene. pPA1, pPA4 and pPA5 all direct similar levels of β-galactosidase activity - around 0.7 enzyme units - from which the following facts can be inferred.

1. 19S-lacZ RNA synthesis must initiate within the CaMV DNA and not in adjacent vector sequences because the insertion of a yeast transcriptional terminator upstream of the gene does not profoundly affect its expression.

2. In the construction of pPA5, 110bp were removed from the 5' end of the CaMV Sau3A fragment. β-galactosidase synthesis was unaffected by

this, indicating that the promoter signals utilised in yeast are confined to the 400bp region immediately upstream of the 19S-lacZ gene.

3.   The activity of the 19S promoter in yeast appears not to be strongly modulated by enhancer-like elements in the adjoining plasmid DNA since in pPA1 and pPA4 the promoter is preceded by different DNA sequences, yet β-galactosidase is expressed at similar levels.

A previously documented hybrid gene, in which a native yeast promoter is fused to lacZ, was analysed for comparison.   Plasmid pLG669-Z contains a lacZ gene linked to the promoter of the iso-1-cytochrome oxidase (CYC1) gene, a moderately expressed yeast gene (Guarente and Ptashne 1981).   As shown in Table 4.1, the level of β-galactosidase derived from this plasmid was 30 times greater than that from the 19S-lacZ gene on pPA1.   This is consistent with the observation that blue colour development in the presence of X-gal was considerably slower in cells containing pPA1, pPA4 and pPA5 than in pLG669-Z transformed cells, and implies that the 19S-lacZ gene is poorly expressed in yeast.

Using the same ONPG assay but an alternative formula to calculate enzyme activity (from Miller 1972) β-galactosidase activities of 100 units and 73 units have been reported in glucose-grown cells harbouring pLG669-Z (Guarente and Ptashne 1981, Yocum *et al* 1984).   When the β-galactosidase activities in Table 4.1 are recalculated using the Miller formula, a similar value of 89 units is obtained for pLG669-Z, confirming that this plasmid is a reliable standard to use in β-galactosidase expression studies.   It has been calculated that 22 units (100 Miller units) of β-galactosidase corresponds to about 0.2% of total cell protein (Guarente and Ptashne 1981).   Assuming that the specific activity of the 19S-lacZ hybrid protein is comparable to that

of the wild-type enzyme, an activity of 0.7 β-galactosidase units would correspond to less than 0.01% of total cell protein, or 50-100 enzyme molecules per cell.

Cell extracts were prepared from yeast cells carrying various plasmids and analysed by SDS-PAGE in an attempt to identify the lacZ gene product (Figure 4.7). This task was complicated by the finding that in all extracts (including those of cells not expressing β-galactosidase, lane b) a doublet band was found to comigrate with the purified enzyme (lane a). In extracts of cells known to contain β-galactosidase (lanes c-f) no additional bands were discernible in this region of the gel, again emphasising the low abundance of the hybrid protein in transformed cells.

4.5 Discussion

Both the low β-galactosidase activity as measured biochemically, and the apparent absence of a novel polypeptide of the expected size (116kd) in transformed cells, suggest that the 19S-lacZ gene is inefficiently expressed in yeast. This could be due to rapid turnover of the gene product, as appeared to be the case with human insulin in yeast (Stepien et al 1983), but the persistence of the enzyme activity in stationary phase cells (Figure 4.6) would suggest just the opposite: that the protein is relatively stable. The "N-end rule" proposed by Bachmair et al (1986), which states that the in vivo half life of a protein is a function of its amino-terminal residue, predicts a half life of 20 hours or more for the 19S-lacZ gene product if the N-terminal methion ine is retained and 30 minutes or less if it is removed, exposing a glutamic acid residue.

In estimating the level of 19S-lacZ gene expression it has been assumed that the hybrid β-galactosidase protein has the same specific activity as the native enzyme. From analysis of many lacZ gene fusions it has been concluded that the substitution of amino-terminal sequences rarely has a significant impact on the activity of the enzyme (Brickan *et al* 1979, Fowler and Zabin 1983). Proteins with reduced enzyme activity tend to result from fusions between lacZ and genes encoding membrane proteins (see Silhavy and Beckwith 1985 for references). In these instances the hybrid protein becomes lodged in the cell membrane, presumably rendering the active site inaccessible. The 18 amino acids of CaMV inclusion body protein which have been fused to β-galactosidase in this study are not hydrophobic and there is no reason to suppose that the hybrid protein is localised in a membrane. Enzyme activity probably is, therefore, a realistic guide to the amount of 19S-lacZ product in the cell.

Excessive instability of the plasmids bearing 19S-lacZ gene fusions is a factor which might contribute to their inefficient expression. The segregational stability of pPA and pPA1 was measured and found to be very similar to that of plasmid YEp213 (Figure 4.8). Under the selective growth conditions typically used in these experiments, approximately 90% of the cells would contain plasmid at the time of harvesting, hence plasmid instability can be discounted as a cause of low level gene expression.

Maintenance of a low steady-state mRNA level and/or inefficient translation of the 19S-lacZ mRNA would seem to be the most plausible explanation for the small amounts of β-galactosidase in transformed cells. Low levels of 19S-lacZ mRNA might arise due to inefficient functioning of the 19S promoter in yeast. This promoter has evolved to

work optimally in a plant cell nucleus and may not interact as effectively with the yeast RNA polymerase and transcription factors. Instability of the mRNA may also contribute to a low steady state level of 19S-lacZ mRNA. Little is known about factors affecting the half life of mRNA in yeast, which can vary between 2 and 100 minutes (Chinnapan-Santiago *et al* 1986). No attempt has been made here to quantitate the relative rates of 19S-lacZ mRNA synthesis and turnover. Consequently it is not possible to unravel the way in which these factors interact to dictate the final steady-state mRNA level.

The efficiency of 19S-lacZ mRNA translation will depend on the position of the mRNA cap site. The initiator codon of CaMV gene VI occurs in a favourable context (AGCAUGA) but translation would be seriously impaired if transcription were to initiate a significant distance up or downstream of the natural cap site. Initiation downstream might result in the RNA starting within the coding sequence while initiation far upstream would result in the presence of AUG triplets within the mRNA leader. In yeast, as in higher eukaryotes, initiator codons are usually the first AUGs to be encountered by a scanning 40S ribosomal subunit. Zitomer *et al* (1984) demonstrated that the presence of additional AUG triplets within a yeast mRNA leader region severely inhibited translation initiation at the correct position. The extent of 19S-lacZ gene expression in yeast will clearly be determined by a variety of factors including the inherent strength of the 19S promoter, the stability of the mRNA and the choice of mRNA start site.

4.6 Summary

A gene fusion was constructed in which the CaMV 19S promoter and 18 codons of CaMV gene VI were fused to the E.coli lacZ coding region. Expression of this gene in yeast was demonstrated by plating cells on indicator medium containing X-gal and also by performing a biochemical assay for β-galactosidase. The insertion of a transcription termination signal upstream of the 19S promoter did not significantly affect the expression of the gene, indicating that transcription of 19S-lacZ is not dependent upon adjacent vector DNA sequences and confirming that the CaMV DNA contains a functional promoter element. The 19S-lacZ gene product was found to be present at very low levels in transformed cells, constituting less than 0.01% of total cell protein, and was not detectable in Coomassie-blue stained protein gels of yeast cell extracts.

Table 4.1 β-galacotosidase activity specified by

various plasmids in yeast

Plasmid	Gene fusion	β-galactosidase activity*	Relative activity
pPA	(no promoter)	0.017 (0.08)	0.09
pPA1	19S-lacZ	0.63 (2.8)	3.2
pPA4	19S-lacZ (upstream DNA inverted)	0.37 (1.7)	1.9
pPA5	19S-lacZ (PGK terminator upstream)	0.78 (3.5)	3.9
pLG669Z	CYC1-lacZ	20 (90)	100

* Activity is expressed in nmol ONPG converted/min/10^7 cells at 28°C. Figures in brackets are activities expressed in Miller units.

Figure 4.1 Construction of plasmids pPA and pPAX.

YEp213-12 was constructed by Dr A Boyd and is identical to YEp213 except that a HindIII site in the 2 micron DNA has been destroyed by filling in of the 3' recessed end with klenow fragment (the large fragment of E.coli DNA polymerase I). This plasmid was partially digested with EcoRl and the fragments ligated into the EcoRl site of pMBL1034. The ligated mixture was restricted with HindIII (to enrich for molecules based on pMBL1034 replicon) and then used to transform a leu⁻ strain of E.coli. Plasmid pPA was identified by restriction mapping of plasmids recovered from leu⁺ amp^R transformants. A plasmid named pPAX was also obtained, in which the same two EcoRl fragments from YEp213-12 were inserted in the reverse orientation with respect to the polylinker. The nucleotide sequence of the polylinker is phased to indicate the lacZ translational reading frame. E = EcoRl, S = SmaI, B = BamHI

EH

2μ ori

tet

YEp213-12
10·5 kb

LEU2

amp

E

E

ESB

amp

pMLB1034
6·3 kb

'lacZ

EcoR1   BamH1
GA ATT CCC GGG GAT CCC
            Sma1

E

LEU2

'lacZ

pPA
12·4 kb

2μ ori   amp

E
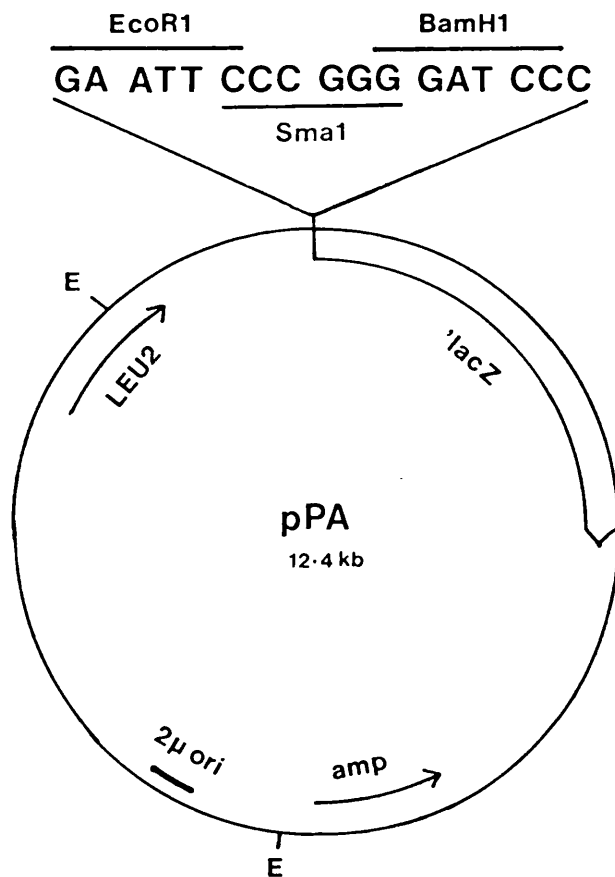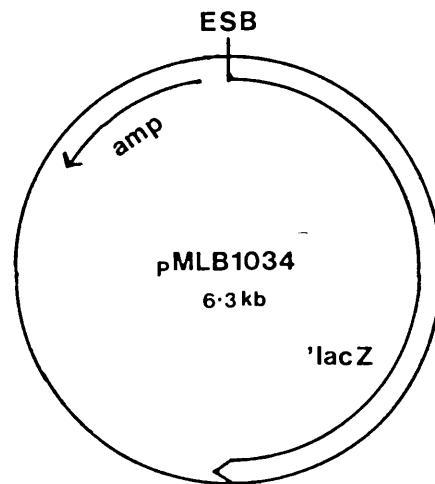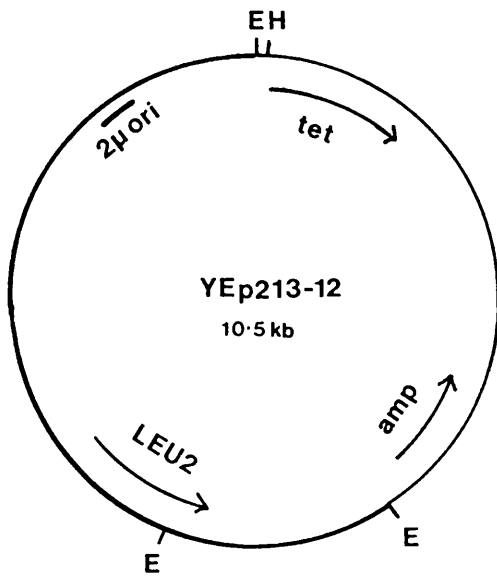
Figure 4.2

a) Restriction map of the Sau3A fragment isolated from pLW414. Filled box represents the coding region of gene VI and in the scale below, position +1 corresponds to the first base of the initiator ATG codon. S = Sau3A, H = HindIII, P = PstI, Hp = HpaI, E = EcoRl, V = EcoRV, Sc = SacI.

b) Construction of pPA1 and pPA2.

pLW414 was digested with Sau3A and the fragments separated by PAGE. A 560bp fragment carrying the 19S promoter of CaMV (p19S) was recovered from the gel and ligated into the BamHI site of pPA. The BamHI restricted vector (pPA) had been treated with calf intestinal alkaline phosphatase to reduce the frequency of vector-vector ligations. Plasmids pPA1 and pPA2 differ in the orientation of the CaMV insert, which was ascertained by restriction mapping of the asymmetric PstI site. B = BamHI.

A

S · · H P · · Hp · · E · V · · Sc S · · gene VI

-500 · -400 · -300 · -200 · -100 · 1 · bp

B

B/S · S/B

pI9S

LEU2

lac Z

pPA1

13·0 kb

2μ ori · amp
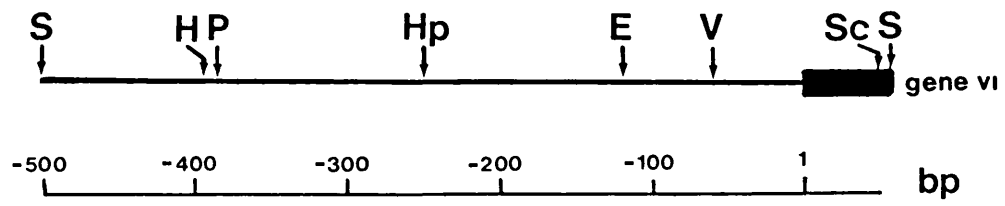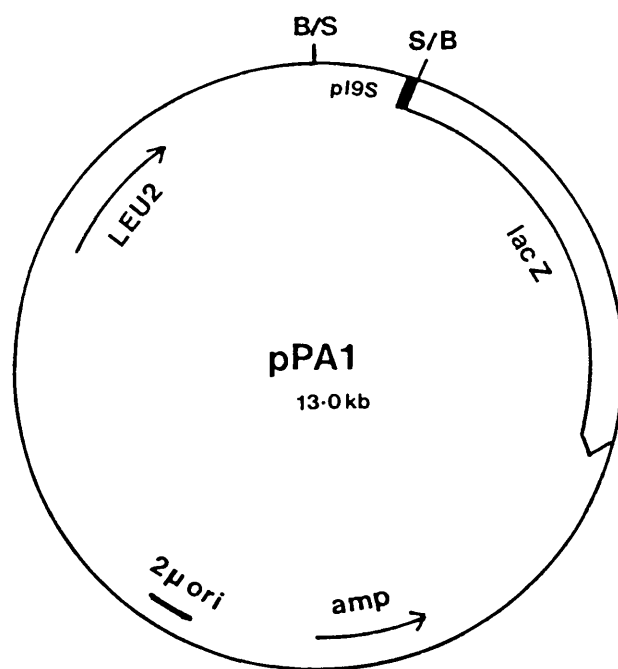
Figure 4.3   Yeast (A) and E.coli strain NM522 (B) harbouring various

plasmids were grown in the presence of 100μg/ml X-gal.

a: pPA6 (35S-lacZ fusion, -168 form)

b: pPA7 (35S-lacZ fusion, -343 form)

c: pPA1 (19S-lacZ fusion)

d: pPA4 (19S-lacZ fusion, upstream DNA inverted)

e: pPA5 (19S-lacZ fusion, PGK terminator upstream)

f: pLG669-Z (CYC1-lacZ fusion)

g: pLGSD5 (CYC1-lacZ fusion incorporating GAL UAS)
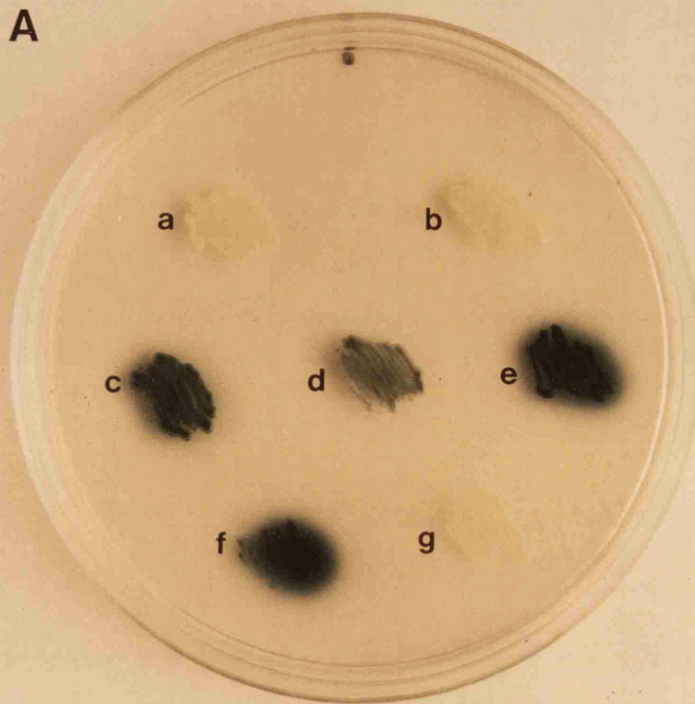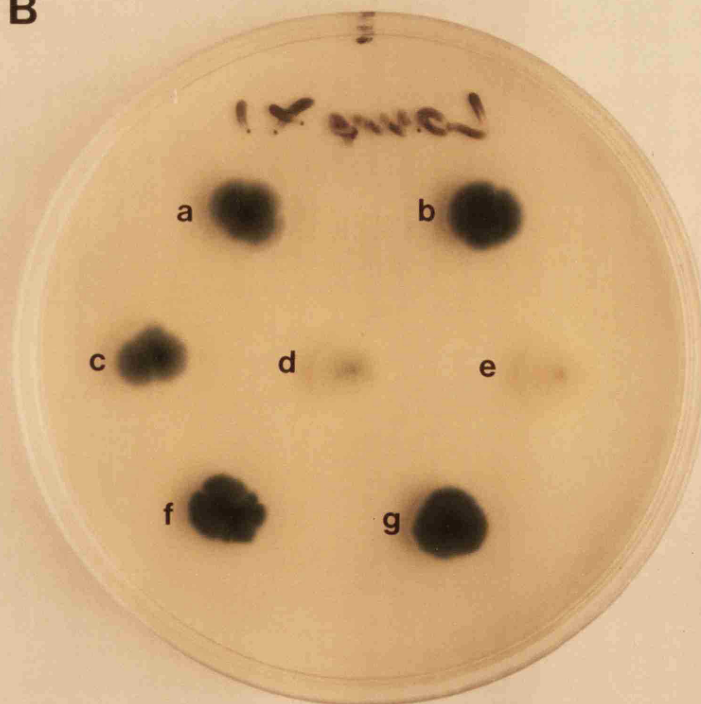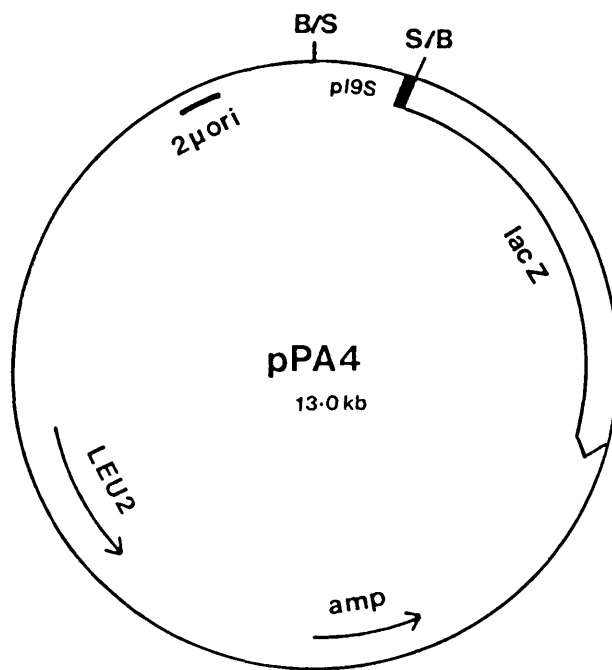
Figure 4.4 Plasmid pPA4. This plasmid was constructed in an identical manner to pPA1 (Figure 4.2) except that the Sau3A promoter fragment was cloned into pPAX rather than pPA. B = BamHI, S = Sau3A. See legend to Figure 4.1 for a description of pPAX.

<u>Figure 4.5</u> Construction of pPA5.

Plasmid pUN121-PGK was obtained from Dr C Hadfield and contains a
2.9kb HindIII fragment of the yeast genome encompassing the PGK gene.
The plasmid was digested with HindIII and the resulting mixture of
two fragments was ligated together with pPA1 which had been
linearised at the unique HindIII site lying within the CaMV promoter
fragment.   Molecules containing the larger HindIII fragment of
pUN121-PGK were eliminated prior to <u>E.coli</u> transformation by
digestion with BamHI and the remaining plasmids screened for the
desired orientation of the PGK gene insert by restriction mapping.
The 5' portion of the PGK gene was removed by restricting the 15.9kb
plasmid with SmaI and SalI, filling in the SalI cohesive end using
the klenow fragment of DNA polymerase I and recircularising the
molecule by blunt end ligation. H = HindIII, S = SalI, Sm = SmaI.
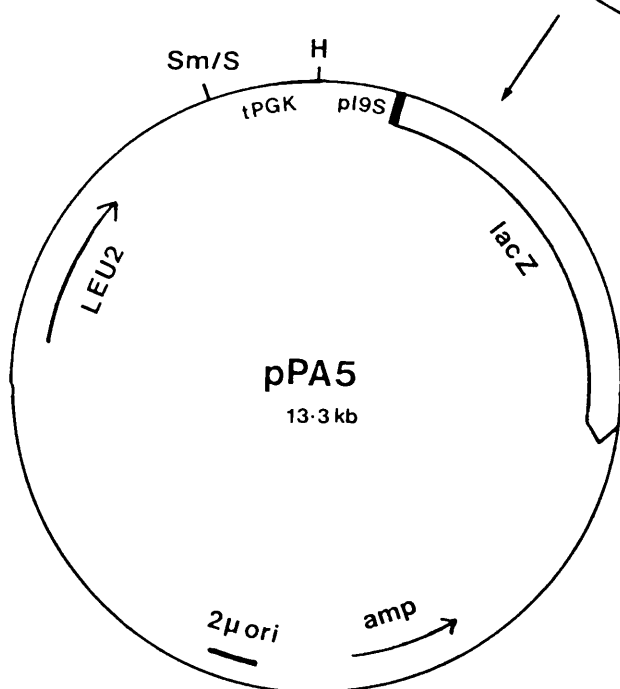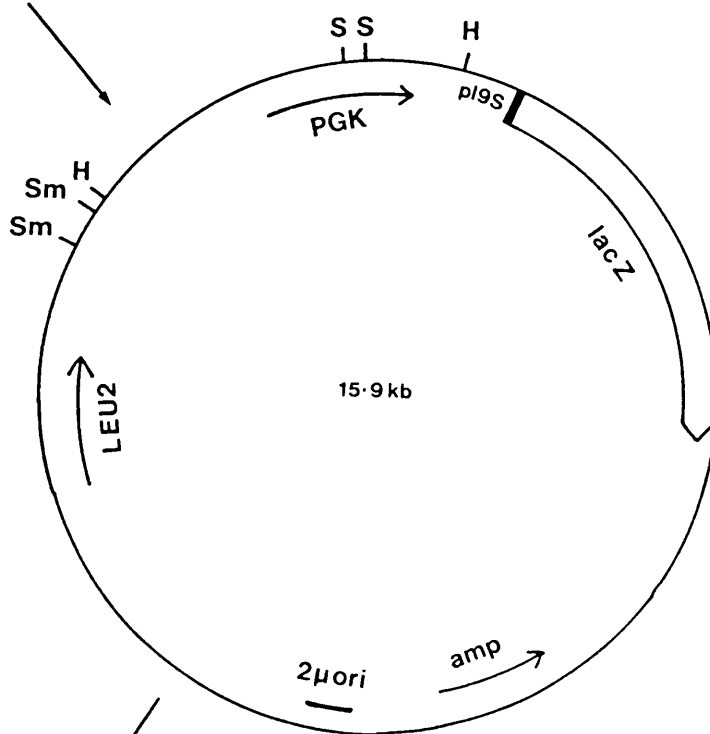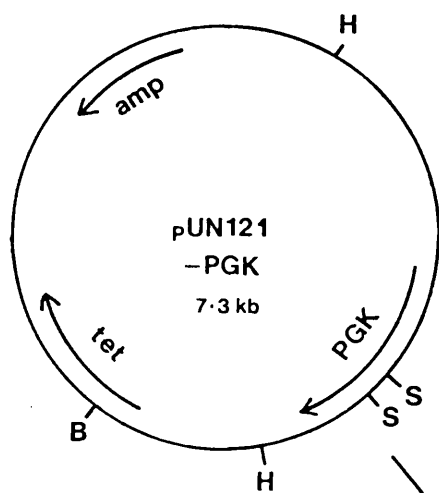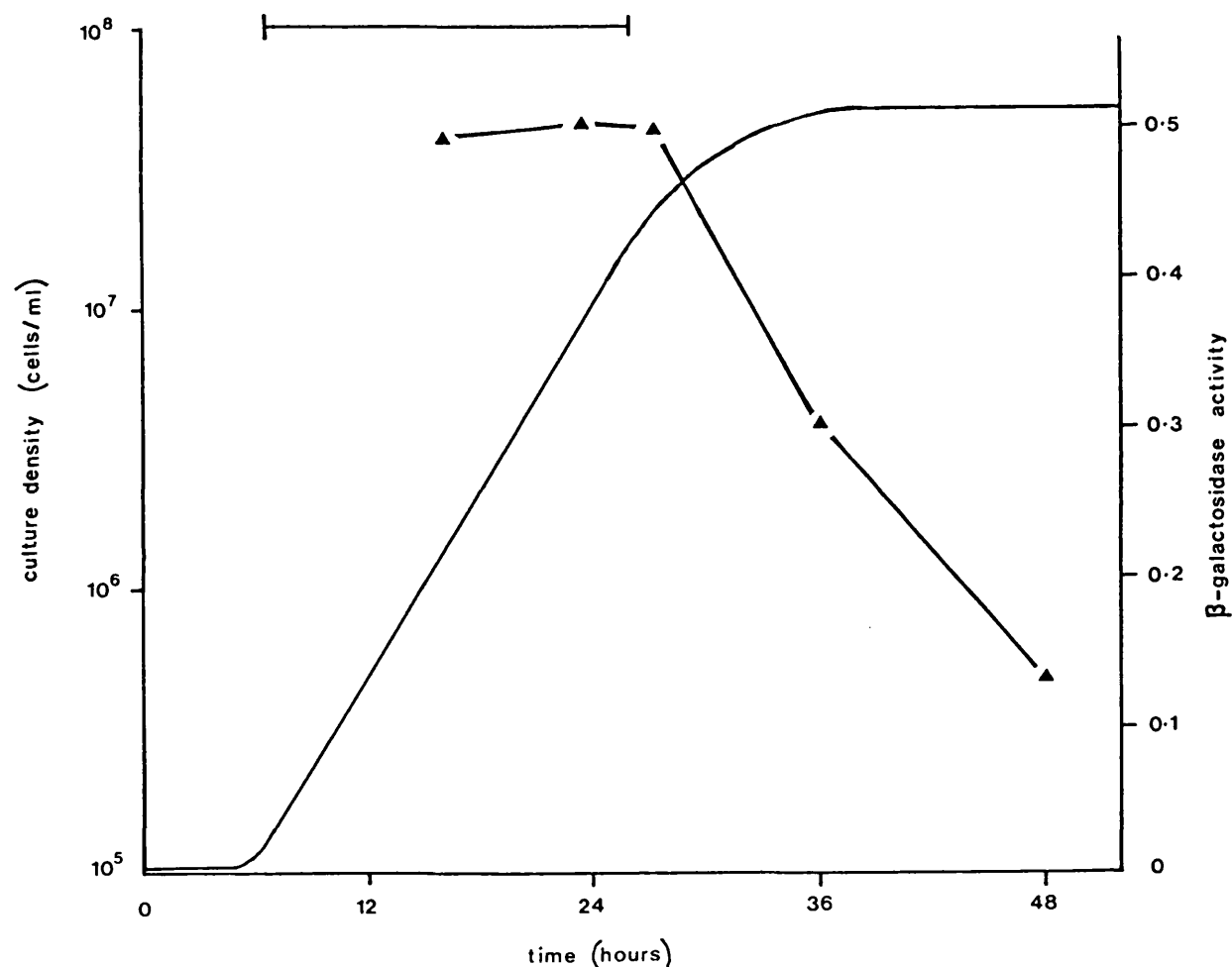
Figure 4.6 A culture of yeast harbouring pPA1 was grown to stationary phase at 30°C in minimal medium with selection for the plasmid. At various stages of growth, cells were removed from the culture and β-galactosidase activity was measured using the ONPG enzyme assay. Culture density (continuous line) and enzyme activity (interrupted line) are plotted against time. Units of β-galactosidase activity are nanomoles of ONPG metabolised / minute / 10^7 cells. The bar above these curves represents the period of logarithmic growth.

Figure 4.7   Yeast cell extracts containing 50μg of protein were subject to SDS-PAGE analysis using a 5-15% linear gradient gel.

lane a: 1μg of purified β-galactosidase (Sigma)

lane b: 10μg of purified β-galactosidase

lane c: S150-2B/pPA extract

lane d: S150-2B/pPA5 extract

lane e: S150-2B/pPA4 extract

lane f: S150-2B/pPA1 extract

The doublet band comigrating with β-galactosidase is arrowed.
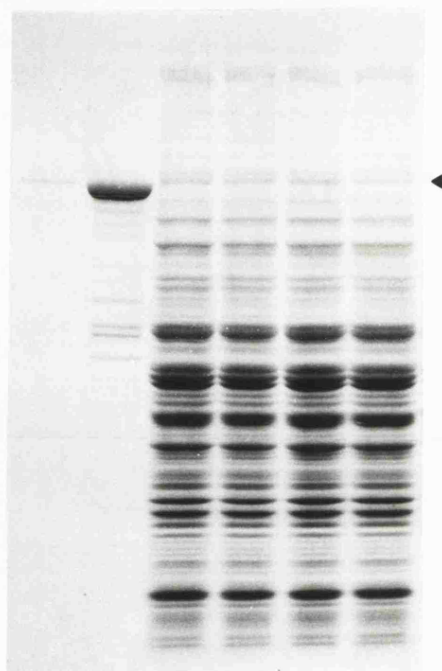
Figure 4.8   The stability of plasmids YEp213 (□), pPA (△) and pPA1
(▲) in yeast during continuous logarithmic culture in
plasmid-selective (B) and non-selective growth medium (A).
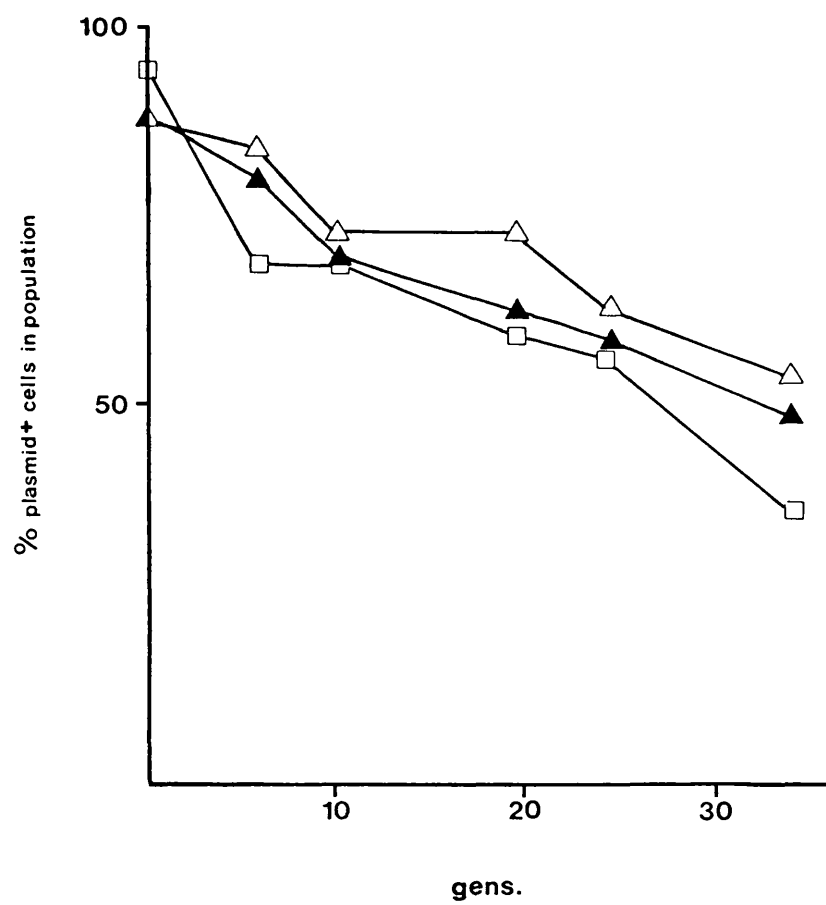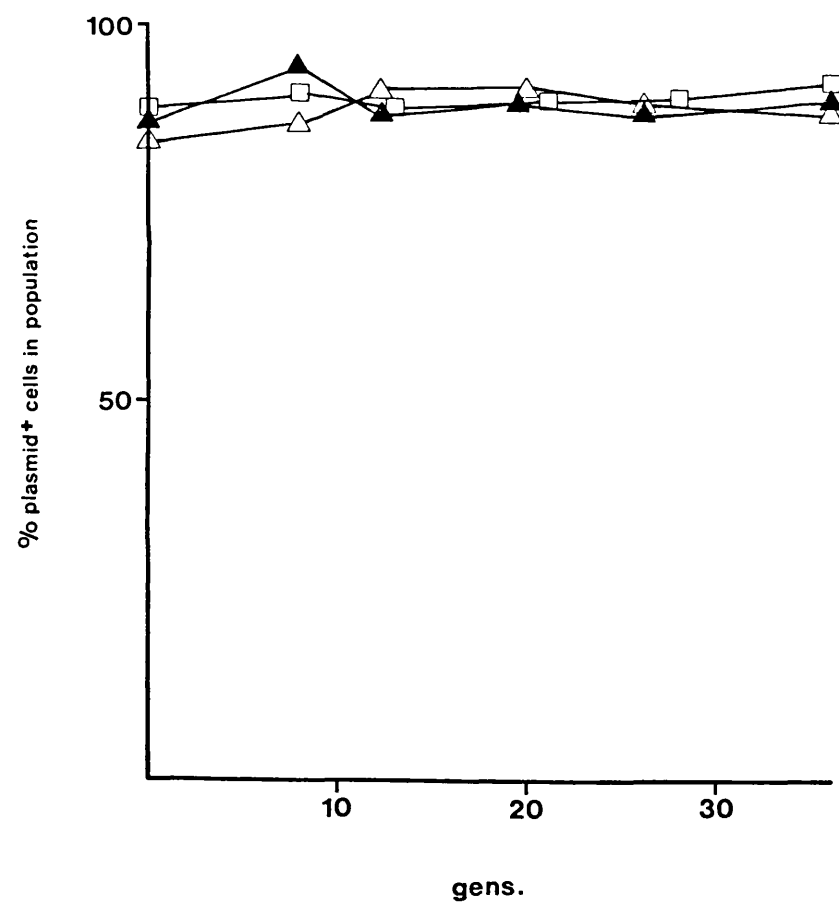
CHAPTER 5

INVESTIGATION OF 35S PROMOTER FUNCTION IN YEAST USING GENE FUSION

5.1 Introduction


The 35S promoter of CaMV is a constitutive promoter gaining widespread
use in plant molecular biology, both as a control (constitutive)
promoter in studies of regulated gene expression (Morelli *et al* 1985,
Ellis *et al* 1987) and for the expression of foreign genes in transgenic
plants (Bevan *et al* 1985, De Block *et al* 1987, Hilder *et al* 1987). To
this end, the 35S promoter has been built into a number of general-
purpose plant expression vectors (Sanders *et al* 1987, Rothstein *et al*
1987). The promoter has been partially characterised by Odell *et al*
(1985), who examined the effects of various deletions on the activity
of the promoter in transgenic tobacco plants. The study revealed that
the 46 base pairs immediately upstream of the RNA cap site (which
contain a typic al TATA box) were sufficient for accurate initiation of
transcription, but that additional promoter element(s) between -46bp
and -168bp were required for maximal transcription in tobacco cells.
Two fully active promoter clones described by Odell extend from +8bp to
-168bp and -343bp (Figure 5.2a) and were generously provided for use in
this project. The promoter fragments were analysed in yeast after
fusion to the E.coli lacZ coding region.

5.2 Construction of two 35S-lacZ gene fusions

Plasmid pPA, described in the previous chapter, was used to construct
the 35S-lacZ gene fusions, but the CaMV promoter fragments could not be
inserted directly into pPA for two reasons. Firstly, the fragments
were not flanked by suitable restriction sites. Secondly, and more

importantly, direct insertion of the promoter fragments would not create a translatable gene. As mentioned previously, the first 8 codons are lacking from the 'lacZ gene in pPA and the gene cannot be expressed unless the coding region is supplied with a translational start codon (ATG). The 35S promoter fragments do not meet this requirement as no ATG triplet occurs in the 8bp downstream of the cap site.

Construction of a translatable 35S-lacZ gene fusion was achieved by synthesizing an oligonucleotide "adaptor" sequence (Figure 5.1) with the following features: 1) SmaI and BamHI termini to enable cloning of the adaptor into the polylinker of pPA. 2) Internal restriction sites for ClaI and HindIII, into which the 35S promoter fragments can be inserted. 3) an ATG codon (context AGCATGG) located downstream of the promoter insertion site so as to be in phase with the 'lacZ coding sequence when the adaptor is cloned into pPA.

Plasmids pPA6 and pPA7, which contain 35S-lacZ fusions using the -168bp and -343bp promoter fragments respectively, were constructed in three stages. Firstly, the adaptor molecule was synthesised from synthetic oligonucleotides and ligated into bacterial cloning vector pTTQ9 (Stark 1987) as described in the legend to Figure 5.1. Plasmid pTTQ9 was used because it has no restriction sites for ClaI or HindIII, hence those introduced in the adaptor will be unique. The structure of the cloned adaptor was verified by DNA sequencing (Figure 5.1). Secondly, the two 35S promoter clones were isolated as ClaI-HindIII restriction fragments and inserted into the plasmid-borne adaptor sequence. Thirdly, the promoters were recovered as SmaI-BamHI restriction fragments and ligated into pPA, where the ATG codon of the adaptor provides an initiator codon for lacZ translation (Figure 5.2).

5.3 Analysis of the 35S-lacZ genes *in vivo*

Expression of the two 35S-lacZ genes on pPA6 and pPA7 was initially monitored by plating yeast cells harbouring these plasmids on agar medium containing the chromogenic indicator X-gal. After 4 weeks incubation at 30°C in the presence of X-gal, the strains showed no blue colouration whatsoever (Figure 4.3), suggesting that the genes are expressed at an extremely low level, if at all, in yeast. The biochemical enzyme assay based on the lactose analogue ONPG likewise revealed no evidence of β-galactosidase activity (Table 5.1). The absence of detectable β-galactosidase cannot be attributed to the synthesis of a non-functional hybrid polypeptide, since lacZ⁻ E.coli colonies harbouring pPA6 or pPA7 are blue in the presence of X-gal (Figure 4.3), clearly demonstrating that a functional enzyme is produced. Neither can the result be attributed to loss or rearrangement of the plasmids, as the presence and integrity of pPA6 and pPA7 in yeast was confirmed by Southern blot analysis (Figure 5.3). *nb* the largest PstI fragment of pPA6 and pPA7 is slightly longer than that of pPA: this is caused by insertion of the 35S promoter and is not due to unscheduled plasmid rearrangement.

5.4 Discussion

The biochemical assay for β-galactosidase is one of the most sensitive enzyme assays available; it will detect fewer than 10 tetramers per cell (β-galactosidase is active as a homotetramer) and the activity of a HIS4-lacZ gene, transcribed at a level of one mRNA molecule per cell, was easily measured (Ruby *et al* 1983). The inability to detect expression of the 35S-lacZ genes is all the more striking, therefore, and contrasts sharply with the 19S-lacZ gene, the expression of which

was readily detected in yeast (see chapter 4).

Possible explanations for the absence of detectable β-galactosidase in pPA6 and pPA7-transformed cells include   i) inefficiency of the transcription and translation processes   ii) plasmid instability   iii) production of a protein with very low specific activity or short half life.   For reasons discussed in the previous chapter, the specific activity of the 35S-lacZ gene product is unlikely to differ significantly from that of the wild-type enzyme and the hybrid protein quite clearly has some activity in E.coli.   Unless the protein is very rapidly degraded *in vivo*, the lack of β-galactosidase in yeast would suggest that little or no 35S-lacZ gene product is synthesised. According to the N-end rule (Bachmair *et al* 1986) the stability of the 35S-lacZ product should be similar to that of the 19S-lacZ gene, *ie* relatively stable, as they initiate with the same two amino acids (met-glu).   Plasmid instability has been excluded as a causative factor, therefore inefficient expression of the gene would seem to be the primary reason for failure to detect β-galactosidase in transformed cells.

Barriers to 35S-lacZ expression may be either transcriptional or translational.   In the former instance, the viral promoter may not function in yeast, *ie* those features of the DNA which signal a promoter in the plant nucleus may not be recognised by the yeast transcription machinery.   Previously documented examples of this kind are the HSV thymidine kinase gene promoter (Kiss *et al* 1982) and an Aspergillus glucoamylase gene promoter (Innis *et al* 1985), neither of which was functional in S.cerevisiae.   There are at least two components to the 35S promoter of CaMV: a TATA box located 30bp 5' of the cap site and additional element(s) further upstream.   Upstream promoter elements are

typically gene specific and interact with specialised regulatory molecules (see chapter 1). The absence of a yeast nuclear factor equivalent in function to the putative 35S regulator may partially or wholly account for the lack of 35S-lacZ gene expression in yeast. Choice of a mRNA initiation site either up- or downstream of the natural 35S RNA cap site may also lead to translational difficulties, as discussed in chapter 4. Without further experimentation it is not possible to deduce which of these factors (protein instability, inefficient transcription initiation or translation) account for the observed lack of 35S-lacZ gene expression in yeast.

5.5 Summary

Gene fusions were constructed between the CaMV 35S promoter and the 'lacZ gene of E.coli, the translation initiation codon being provided by a synthetic adaptor molecule. Two 35S-lacZ genes, bearing promoter fragments 168bp and 343bp in length, were introduced into yeast on a YEp plasmid derived from pPA. Using indicator medium containing X-gal and also the more sensitive ONPG assay for enzyme activity, no β-galactosidase could be detected in the transformed cells. This result, which could not be attributed to plasmid instability or the synthesis of a non-functional protein, suggests that expression of the 35S-lacZ gene is blocked at the level of transcription or translation.

Table 5.1 β-galactosidase activity specified by various

plasmids in yeast

Plasmid	Gene fusion	β-galactosidase activity*
pPA1	19S-lacZ	0.96
pPA6	35S-lacZ (−168 clone)	0.014
pPA7	35S-lacZ (−343 clone)	0.019
pPA	(no promoter)	0.017

* nmol ONPG converted/min/10^7 cells at 28°C.

Figure 5.1 Synthesis and cloning of the adaptor molecule.

The nucleotide sequence of the adaptor is shown in (A). The translational start codon is emboldened. The adaptor was synthesised chemically in four parts, each strand being composed of two oligonucleotides joined at the position marked by the asterisk. The individual oligos were phosphorylated at the 5' end with T4 kinase, then annealed together with the plasmid vector (pTTQ9) as follows: a 10μl reaction containing 1x linker-kinase buffer (Maniatis et al 1982) 1μg of each of the phosphorylated oligos and 2μg of pTTQ9 restricted with BamHI and SmaI was incubated at 80°C for 5 minutes then gradually cooled to 0°C over 4 hours. The annealed mixture was then ligated, used to transform E.coli, and recombinant molecules containing the adaptor were identified by virtue of the newly introduced ClaI and HindIII restriction sites. The resulting plasmid was called pX and the structure of the cloned adaptor in pX was verified by direct sequence analysis of the plasmid DNA (B).

A

5′GGGATCGATAAGCTTAGC**ATG**GAG
3′CCCTAGCTATTCGAATCGTACCTCCTAG

Smal   ClaI   HindIII   BamHI

amp

pX

B

BamHI

HindIII

ClaI

Smal

A G C T
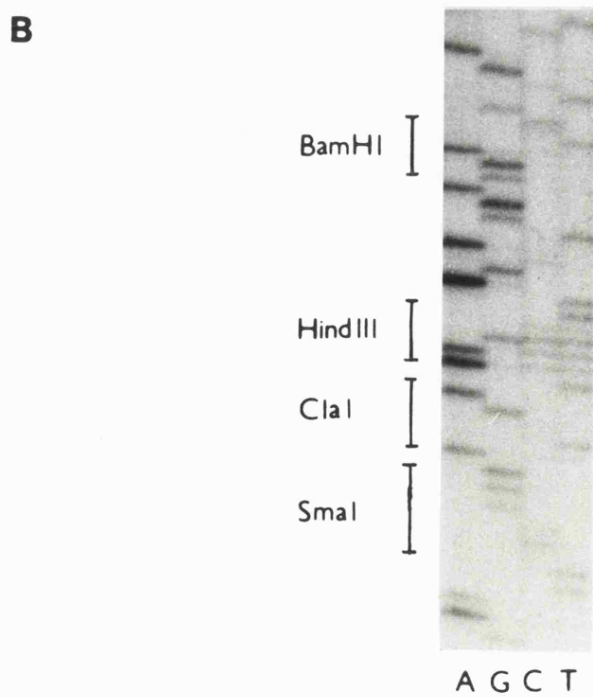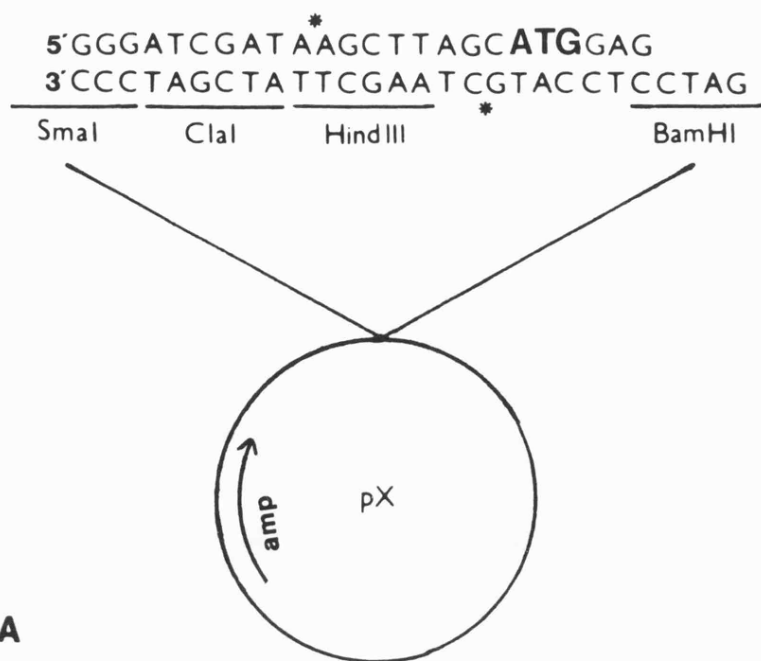
Figure 5.2 Construction of pPA6 and pPA7.

Two 35S promoter fragments (shown in A) were isolated as ClaI-HindIII
restriction fragments from the plasmids in which they were supplied
and ligated into plasmid pX giving pX168 and pX343.   The promoters
were   recovered   from   these   plasmids   as   SmaI-BamHI   restriction
fragments and inserted into the polylinker of pPA (Figure 4.1) to
produce pPA6 and pPA7.   These contain the -168bp and -343bp 35S
promoter fragments respectively.   C = ClaI, H = HindIII, Sm = SmaI, B
= BamHI.

A

C ——————————————————————— H
-343   +8

                    C ——————————————— H
                  -168   +8
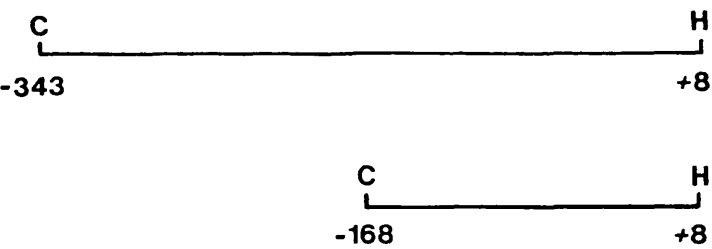
B
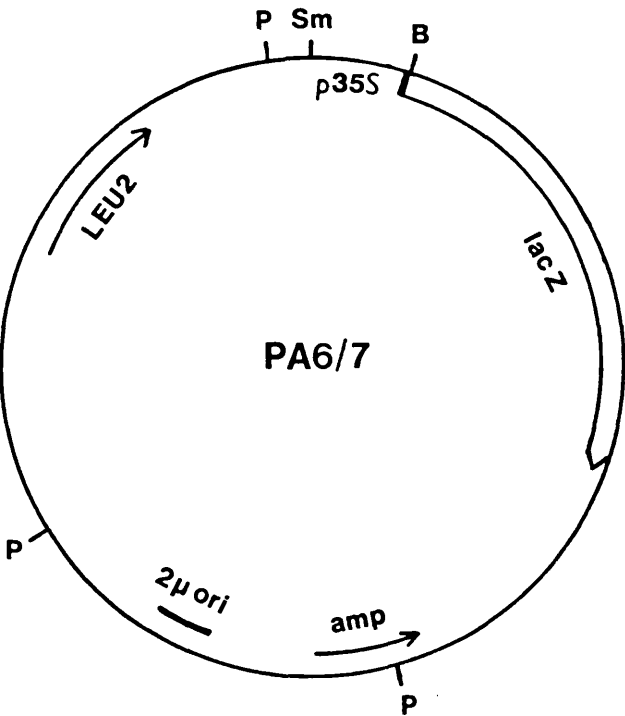
Figure 5.3   Southern blot analysis of DNA from yeast cultures
harbouring pPA6 and pPA7.   DNA was extracted from cells of the same
culture as was used for β-galactosidase assays.   5μl of DNA
(approximately 1μg) was digested with PstI and probed with plasmid
CV9, which contains the LEU2 gene of S.cerevisiae in pBR322, and
shares homology with each of the PstI fragments of pPA but not with
the native 2μ plasmid.

lane a: 0.1μg of purified pPA DNA restricted with PstI

lanes b-d: DNA from 3 independent pPA7 transformants

lanes e-f: DNA from 3 independent pPA6 transformants

lane a was exposed for 3 hours and lanes b-g for 24 hours without an
intensifying screen.
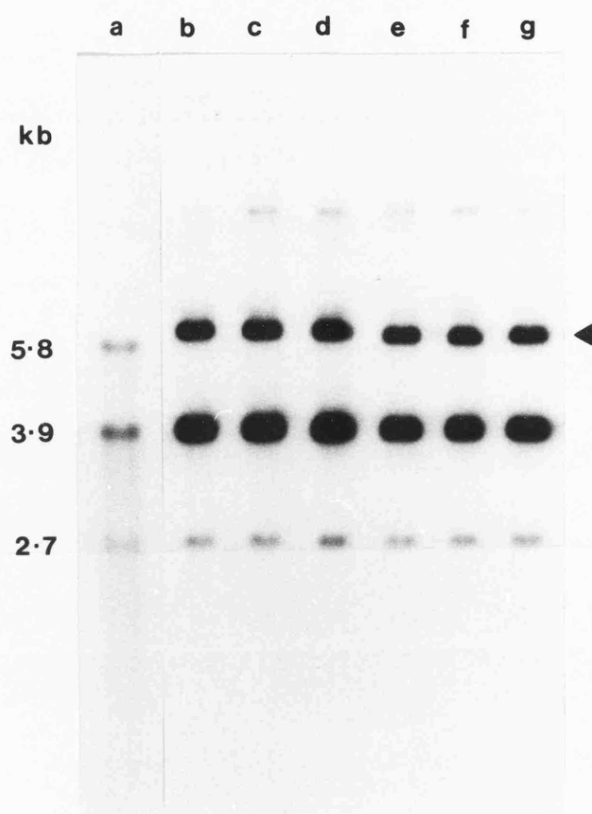
CHAPTER 6

MAPPING THE 5'ENDS OF THE 19S AND 19S-lacZ mRNAs

USING PRIMER EXTENSION

6.1 Introduction

The utility of yeast as a host organism for analysing plant promoters will be highly dependent upon such factors as promoter strength, regulation and choice of mRNA cap site being faithfully reproduced in the yeast cell. The third factor, accuracy of transcription initiation, is addressed in this chapter, in which the 5' terminus of the 19S-lacZ mRNA has been precisely mapped by primer extension (Figure 6.1).

6.2 Optimisation of reaction conditions

i) Choice of DNA primer.

A synthetic oligodeoxynucleotide (oligo) was used to prime cDNA synthesis by reverse transcriptase. The oligo, of minus strand polarity, had the sequence 5'GGGATCGAGCTCTAGC3' and was complementary to a 16 nucleotide sequence near the 5' end of the 19S-lacZ coding region (Figure 6.2A). Because the oligo is complementary to the gene VI portion of the 19S-lacZ RNA, it could also be used to prime cDNA synthesis from the native 19S transcript. The former is a relatively abundant RNA species in CaMV-infected plant cells and was used as a control in the 19S-lacZ RNA mapping studies.

ii) Comparison of various primer/RNA annealing regimes.

According to literature reports, temperatures ranging from 25°C to 95°C have been employed for the primer/RNA annealing step. To determine

which was the most efficient procedure for annealing the oligonucleotide with the 19S transcript, total RNA extracted from CaMV-infected turnip leaves was incubated with the oligo under three different conditions. RNA plus primer were either boiled for 5 minutes then gradually cooled to 70°C over a 60 minute period (track 1, Figure 6.2B), or incubated at 65°C for 90 minutes (track 2) or incubated at 42°C for 90 minutes in the presence of placental ribonuclease inhibitor (track 3). The annealed primers were then extended using reverse transcriptase and analysed by electrophoresis and autoradiography. At all stages subsequent to the annealing, the three samples were treated identically.

The conditions of primer annealing were found to have a profound effect on the range of cDNA products obtained (Figure 6.2B). The most stringent annealing (5 minutes boiling) resulted in low molecular weight products only with no discrete bands. Annealing at 65°C produced a wide array of cDNAs whereas annealing at 42°C gave predominantly a single cDNA species plus a small amount of very high molecular weight material. This major band appears to be the genuine primer extension product of the 19S RNA since subsequent experiments (not shown) revealed the cDNA to be of exactly the predicted length and to be absent when the reaction was performed using RNA from uninfected turnip leaves. The high molecular weight RNA was invariably present to a greater or lesser extent when the 42°C annealing conditions were used and may represent cDNAs synthesised using the CaMV 35S transcript as a template. This is to be expected since any primer complementary to the 19S RNA will hybridise to the 35S RNA as well, producing cDNAs of up to 5kb in length. The lowest stringency annealing conditions (42°C) were unequivocally the most efficient in this system and were used routinely in subsequent reactions.

6.3 Analysis of the 19S-lacZ RNAs produced in yeast

Having ascertained the experimental conditions necessary for efficient cDNA synthesis using the 19S transcript as a template, the primer exter ion method was used to examine the hybrid 19S-lacZ RNAs present in yeast cells. RNA samples from yeast harbouring plasmids pPA5 (which carries the 19S-lacZ gene preceded by a transcription termination signal), pPA9 (which contains the 19S-lacZ gene coupled to an inducible enhancer - see chapter 7 for details) and YEp213 (which has no 19S-lacZ gene) were analysed. The data, presented in Figure 6.3, reveals a number of interesting points, listed below.

1. The 19S-lacZ transcripts are heterogeneous at their 5' ends: at least three classes of cDNA are synthesised from yeast/pPA5 RNA (Figure 6.3, lanes c,d, bands 1,2,3) which are absent from the control reaction using YEp213 RNA (lane e). *nb* an additional product (band 4) is common to all reactions performed with yeast RNA, including the control reaction, and cannot therefore be specific to the 19S-lacZ gene.

2. cDNA bands 1,2 and 3 are synthesised from polyadenylated RNAs because a far stronger signal is obtained using pPA5 RNA enriched for polyA$^+$ transcripts than with total RNA (lanes c,d).

3. None of the 19S-lacZ cDNAs co-migrates with the CaMV 19S RNA primer exten ion product (band 5) hence none of the three mRNA cap sites in yeast is coincident with that of the native 19S transcript. The major 19S-lacZ primer extension product (band 1) is extremely long and runs near the top of the gel where it is difficult to size accurately, however I estimate the cap site to be approximately 250bp upstream from the gene VI coding region. From close inspection of the autoradiograph

and also from a sequencing ladder run on the same gel, it is clear that the two other cDNAs (bands 2 and 3) are 20 and 36 nucleotides respectively shorter than the native 19S product. The 5'non coding region of the 19S RNA is only 11nt in length, from which it can be deduced that the 19S-lacZ RNAs initiate within the coding sequence of the 19S-lacZ gene at positions +9 and +25 (Figure 6.4).

4.   When the inducible yeast enhancer, the GAL UAS, on pPA9 (this plasmid is described in chapter 7) is activated by growing cells in the presence of galactose, it is primarily the longer 19S-lacZ RNA (giving rise to band 1) which is elevated (Figure 6.3, lane b). Hence the yeast UAS displays preferential stimulation of the nearer of two (or more) tandem promoters, a property shared with various viral and animal cell enhancer elements.

The migration of several cDNA bands (notably the 19S primer extension product) as a doublet was a recurring feature of these experiments. A possible explanation for this is that the doublets represent genuine heterogeneity of the RNA template, *ie* the existance of two cap sites staggered by a single nucleotide. Both Covey *et al* (1981) and Guilley *et al* (1982) have demonstrated a unique 19S RNA initiation site in infected plant cells, however, arguing against this explanation. Heterogeneity of primer length could produce a similar effect and fractionation of the labelled primer in a 20% acrylamide/urea gel did indeed reveal the sample, theoretically a 16mer, to contain a trace of 15mer (Figure 6.2C). If primer breakdown were the sole reason for the doublet phenomenon then one might expect every cDNA band to run as a doublet. This is not the case, invoking a third explanation, which is that the doublets are an artefact arising from "cap effects". Secondary cDNAs with an apparent length 1-2 nucleotides shorter than

the full length version may be generated in primer extension reactions and are thought to be due to premature termination of reverse transcription at the capped residue (Williams and Mason 1985). The doublets observed here are probably due to a combination of variable primer length and cap effects.

6.4 Discussion

The data presented in this chapter raise interesting questions concerning expression of the 19S-lacZ gene and also highlight the structural disparity between yeast and higher eukaryotic promoters. The existence of an RNA species initiating about 250bp 5' of the 19S-lacZ coding region is indicative of a promoter signal shortly upstream, although this region does not act as a promoter in plant cells. Inspection of the DNA sequence does indeed reveal promoter-like elements in this region. For example, the section -300 to -350 has a very high AT content (90%) and contains a precise match to the canonical TATA box, which appears to direct RNA polymerase II binding in many eukaryotic cells. In addition, the AT rich nature of the region may facilitate the assembly of a transcription complex as proposed by Struhl (1985).

The 19S-lacZ RNAs initiating at +9 and +25bp (with respect to the start of the coding region) may well reflect functional recognition of the 19S promoter TATA box centred around -40bp. This is because transcription initiation in yeast tends to occur not 30bp (as in higher eukaryotes) but 40bp or more downstream of this element (see chapter 1). From the relative steady-state levels of the 19S-lacZ mRNAs, this 19S promoter signal would seem to be considerably less effective than the fortuitous promoter further upstream.

In the absence of a mRNA initiating shortly upstream of the 19S-lacZ coding region, one of the observed RNAs must be translated to produce β-galactosidase. Translation of the major transcript is probably exceedingly inefficient because the 250 nucleotide leader region contains several AUG triplets (see Figure 6.4). Studies in yeast by Zitomer *et al* (1984) have shown that, in accordance with the ribosome scanning model of translation initiation proposed by Kozak (see chapter 1), the presence of AUG triplets in the 5' untranslated portion of a mRNA severely inhibits translation initiation at the correct site. It is therefore highly unlikely that this molecule is a functional mRNA.

The two shorter 19S-lacZ transcripts both initiate within the coding region of the gene, which at first sight would appear to preclude their functioning as mRNAs. Inspection of the CaMV gene VI sequence, however, reveals two internal methionine codons (codons 9 and 15) which are situated downstream of the observed cap sites (Figure 6.4). In the absence of the usual translational start signal, it is possible that one of these "internal" AUGs may serve as an initiation codon. Data presented in chapter 8 lends support to this suggestion.

6.5 Summary

Analysis of yeast 19S-lacZ transcripts by primer extension revealed three classes of RNA. The major RNA species initiates about 250bp upstream of the 19S-lacZ coding region and two less abundant species initiate within the open reading frame at nucleotides +9 and +25. All three species are polyadenylated and none has a 5' end that is coincident with that of the native 19S transcript synthesised in CaMV-infected plant cells. The effect of placing a galactose-inducible enhancer (the GAL UAS) upstream of the 19S-lacZ gene was to elevate the

steady-state level of RNA initiating at the most UAS-proximal cap site and had a relatively minor influence on the level of transcripts initiating further downstream.

<u>Figure 6.1</u>   The   principle   of   transcript   mapping   by   primer

extension.

A single-stranded DNA primer is phosphorylated with ^{32}P at its 5' end

using T4 kinase (asterisk denotes ^{32}P).   The labelled primer is

incubated with total or poly(A)$^+$ RNA and anneals specifically to the

complementary sequence in the RNA molecule of interest.   On addition

of dNTPs and AMV reverse transcriptase, the primer is extended to

produce a ^{32}P-labelled cDNA molecule of defined length.   The products

of the extention reaction are fractionated in a urea-acrylamide gel

and detected by autoradiography.   From the length of the cDNA

synthesized (which is determined using appropriate size markers eg a

sequencing ladder), the position of the 3' terminus (and therefore

the mRNA 5' terminus) is deduced.

mixed population of
RNA molecules

+

—∗

radio-labelled DNA primer

specific annealling

5′————————————∗————————————3′

reverse transcription

gel electrophoresis and
autoradiography

◄ cDNA product

◄ unextended primer
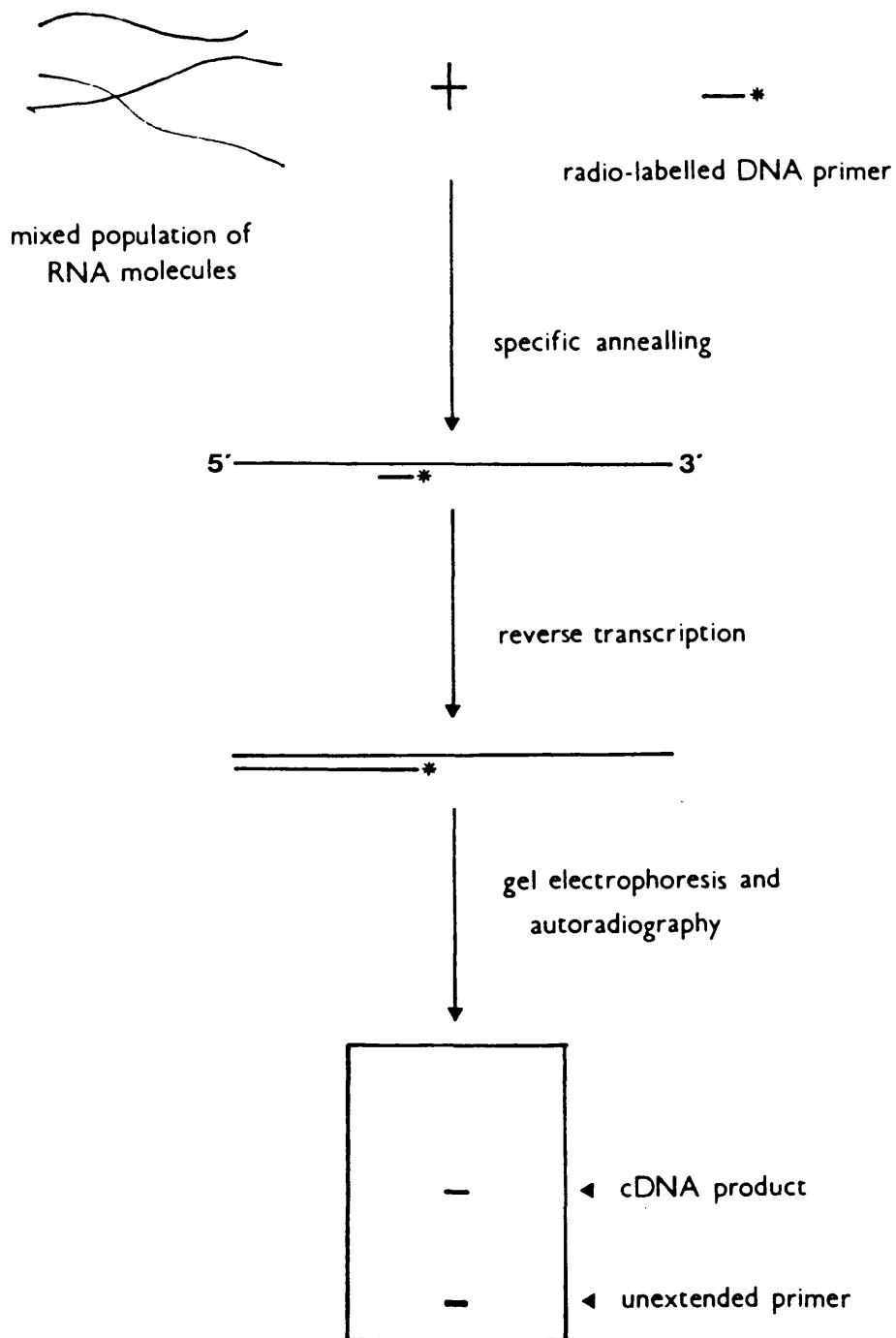
Figure 6.2

A. Regions of the 19S-lacZ gene and CaMV gene VI to which the oligonucleotide primer corresponds is indicated by the central bar. Open box = gene VI coding region, hatched box = lacZ coding region.

B. Primer extension: various annealing temperatures compared. 10 pmoles of kinase-labelled primer were mixed with 10µg of total RNA from CaMV-infected turnip leaves and annealed at 95°C for 5 minutes then cooled to 70°C over 60 minutes (track 1), at 65°C for 90 minutes (track 2) or at 42°C for 90 minutes in the presence of RNasin (track 3). Extension reactions were performed as described in chapter 2 and the reaction products were separated in a 20cm x 17cm 10% urea-acrylamide gel, which was then covered in Saran wrap and exposed to X ray film for 5 days at -70°C with an intensifying screen.

C. Electrophoretic analysis of the primer oligonucleotide.

The oligonucleotide was labelled with ^{32}P using T4 kinase and 0.25ng (0.05 pmole) was electrophoresed in a 20% urea-acrylamide gel until the bromophenol blue had reached the bottom. The gel was covered with Saran wrap and exposed directly to X ray film overnight at -70°C with an intensifying screen.
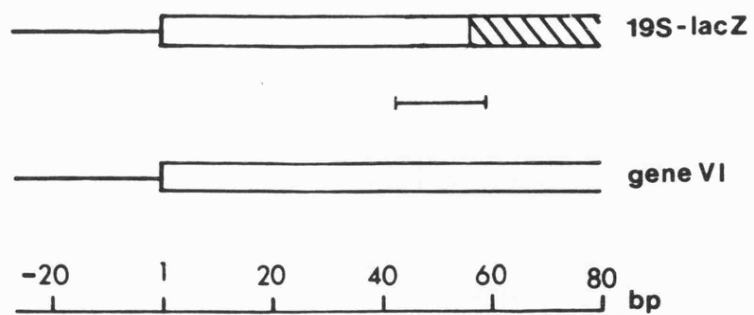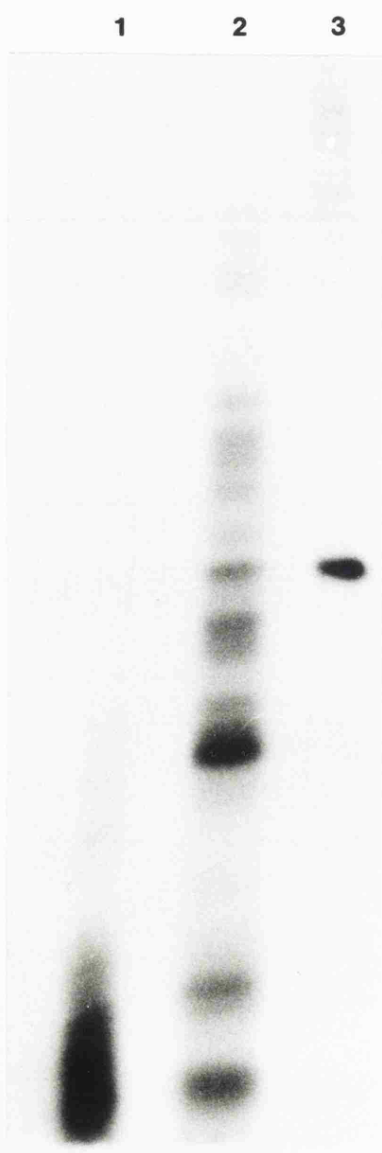
Figure 6.3   Primer extension analysis of 19S and 19S-lacZ
transcripts.

A and B are separate experiments in which primer extension reactions
were performed using the following RNA samples:

A. a,f) 2μg total RNA from CaMV-infected turnip leaves.

    b)   50μg total RNA from yeast transformed with pPA9 (galactose
induced)

    c)   16μg poly(A)$^+$ RNA from yeast transformed with pPA5

    d)   50μg total RNA from yeast transformed with pPA5

    e)   50μg total RNA from yeast transformed with YEp213

B. g) 2.5μg total RNA from CaMV-infected turnip leaves

    h)   50μg total RNA from yeast transformed with pPA9 (galactose
induced)

1 to 5 indicate specific cDNA bands refered to in the text.

The three panels in A represent a 16h exposure (left), a 24h exposure
(centre) and a 24h exposure with intensifying screens (right).

bp

-500   +1   +100

AT-rich

TATAAA   TATATAAA
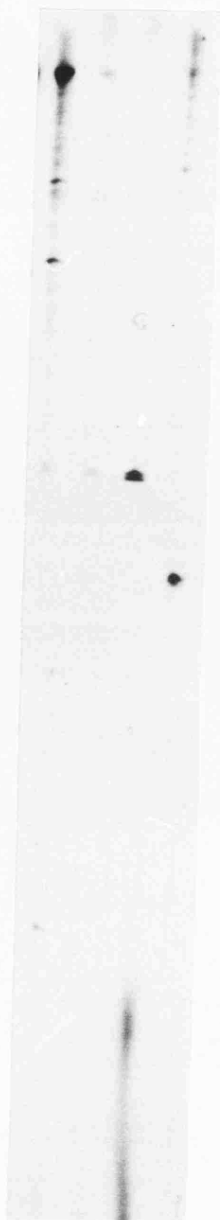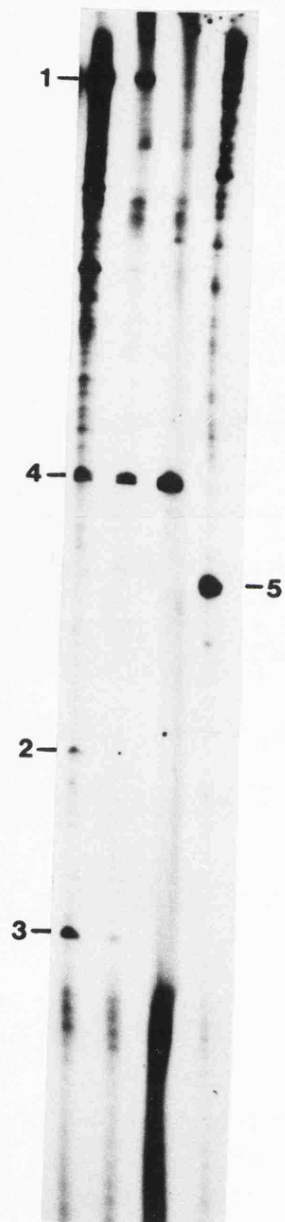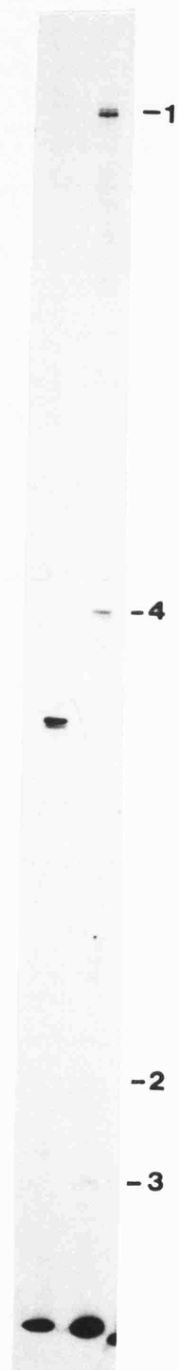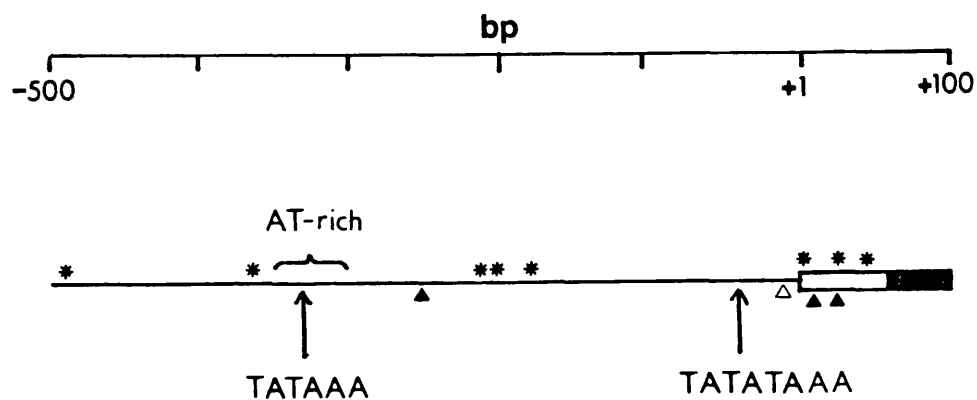
Figure 6.4   Cap sites and structural features of the 19S-lacZ promoter region.

The diagram shows the 5' end of the 19S-lacZ gene. Open box represents CaMV gene VI sequences and filled box, lacZ sequences. The position of ATG triplets is indicated by (*). (△) shows the cap site of the native 19S transcript and (▲) denote initiation sites of 19S-lacZ RNAs synthesized in yeast.

CHAPTER 7

THE EFFECTS OF A YEAST ENHANCER ON EXPRESSION OF THE 19S-lacZ GENE

7.1 Introduction

In chapter 4, the level of β-galactosidase synthesised by the 19S-lacZ

gene was found to be low, of the order of 50 to 100 enzyme molecules

per cell. This chapter describes an attempt to boost expression of the

gene by coupling it to a yeast upstream activator sequence (UAS). The

best characterised yeast UAS is the galactose-inducible GAL UAS,

located between the divergently transcribed GAL1 and GAL10 genes on

chromosome 2 (Figure 7.1A). The GAL UAS has been shown to elevate the

transcription of nearby yeast genes up to 1000-fold in the presence of

galactose and to depress their expression in glucose (St John and

Davies 1981, Guarente et al 1982). Whilst the mechanism of glucose

repression is poorly defined, molecular events leading to gene

activation have been intensively studied. From genetic analyses of the

GAL system (Oshima 1982), it was deduced that the GAL4 gene product was

required for galactose specific induction of GAL1 and GAL10

transcription. More recent molecular studies have confirmed that in

the presence of inducer (galactose or one of its metabolites), the GAL4

product binds to a conserved sequence which occurs upstream of a

variety of galactose-inducible genes (Giniger et al 1985, Bram et al

1986). In the case of the GAL1-GAL10 promoter region, the GAL4 binding

sites coincide precisely with the UAS previously defined by deletion

analyses (West et al 1984, Johnston and Davies 1984). This chapter

describes the effects on 19S-lacZ gene expression of · placing the GAL

UAS directly upstream.

7.2 Subcloning of the GAL UAS into pPA1

Plasmid pBM150, which contains the entire GAL1-GAL10 intergenic region, was used as a source of GAL UAS (Johnston and Davies 1984). A 143bp RsaI-AluI restriction fragment carrying the UAS was recovered from pBM150 and ligated into the SmaI site of pPA1, as described in figure 7.1. The presence of the RsaI-AluI fragment in the resulting plasmid pPA9 was confirmed by Southern hybridisation (not shown). The orientation of the inserted UAS was not determined.

7.3 Analysis of the UAS-coupled 19S-lacZ gene in vivo

The influence of the UAS on 19S-lacZ gene expression was assessed *in vivo* by comparing β-galactosidase enzyme levels in cells harbouring pPA1 and those harbouring pPA9. lacZ RNA levels were also compared.

a) Measurement of enzyme activity. β-galactosidase activity was measured using the ONPG biochemical assay and found to be elevated approximately 5-fold in pPA9-transformed cells grown in the presence of galactose (Table 7.1). When the UAS was uninduced (by growth in glucose), β-galactosidase activity was significantly lower in pPA9-containing cells than those containing pPA1. This finding is not unexpected, since sequences mediating catabolite repression of the yeast GAL1 and GAL10 genes are known to lie close to (and possibly overlap with) the UAS (West *et al* 1984, Johnston and Davies 1984). The UAS region has been shown to reduce transcription from a heterologous yeast promoter HIS3 in the presence of glucose (Struhl 1985a). It is likely, therefore, that in constructing pPA9, sequences responsible for glucose repression have been introduced in addition to the UAS.

b) RNA analysis. RNA dot blotting revealed a corresponding (12-fold) increase in the steady-state level of lacZ RNA in galactose-grown cells harbouring pPA9 when compared with cells harbouring pPA1 (Figure 7.2). Thus the galactose induction of 19S-lacZ gene expression appears to operate at the level of transcription. An identical set of RNA samples was probed for 2 micron plasmid B gene transcripts, the levels of which were fairly constant in all strains.

An unexpected finding from the data presented in Figure 7.2 was that the truncated and promoterless 'lacZ gene in plasmid pPA is transcribed in vivo. Experiments described in chapter 4 have ruled out transcriptional readthrough from vector sequences as a factor contributing to 19S-lacZ expression (the insertion of a known yeast transcription termination signal upstream of the 19S promoter did not significantly alter β-galactosidase expression in vivo). The existence of the pPA transcripts is intriguing nonetheless, particularly as lacZ RNA is more abundant in cells containing pPA than in cells containing pPA1. It is possible that the presence of the CaMV DNA blocks the synthesis of transcripts originating further upstream, which would otherwise continue through the lacZ gene. The cloned CaMV DNA is exceedingly AT-rich and certain AT-rich sequences are known to cause transcription termination in yeast (Beggs et al 1980, Henikoff and Cohen 1984).

7.4 Discussion

The data presented here indicate that a yeast UAS can be coupled to a foreign promoter to achieve enhanced and inducible expression of the accompanying gene. Dot blot experiments have revealed the level of lacZ RNA to be increased 10-fold or more when the 19S-lacZ gene is

linked to the GAL UAS. This is a modest elevation however in comparison with other yeast genes coupled to the same UAS, for which 1000-fold enhancement of RNA levels have been reported (St John and Davies 1981, Guarente *et al* 1982). The primer extention analysis of 19S-lacZ transcripts described in chapter 6 did, however, reveal a massive increase in the level of one specific lacZ RNA in galactose-induced cells harbouring pPA9 (band 1 in Figure 6.3). This transcript appeared to arise from a fortuitous promoter within the CaMV DNA, distinct from the 19S promoter and located approximately 300bp upstream of the 19S-lacZ coding sequence. These results present a paradox in that we detect a greater than 100-fold increase in the level of one specific 19S-lacZ transcript but only a 10-fold increase in the overall amount of lacZ-hybridising RNA. The data can be rationalised by postulating the existence of a body of (largely non-functional) lacZ encoded transcripts which are not inducible by the UAS (Figure 7.3). Enhancer-like elements generally exert their greatest effect on the nearest promoter/cap site. The non- or weakly-inducible transcripts may therefore be negative strand RNAs or positive strand RNAs which arise from promoters located further downstream. If the galactose inducible RNA species (α in Figure 7.3) arising from the fortuitous promoter at -300bp formed, say, 1% of the total lacZ hybridising transcripts under non-induced conditions, then a 1000-fold increase in the abundance of this RNA would produce only a 10-fold elevation in the overall amount of lacZ RNA. The fact that β-galactosidase activity is enhanced only 5-fold upon galactose induction of cells harbouring pPA9 suggests that the major induced transcript is not a functional mRNA for β-galactosidase. This is probably due to the translational difficulties associated with the very long leader sequence, as discussed in chapter 6. Those RNAs which are translated appear, therefore, to be among the relatively non-inducible transcripts

initiating further downstream (group β transcripts according to the model in figure 7.3). This group would include the two 19S-lacZ transcripts detected by primer extension which initiate at +9 and + 25bp (bands 2 and 3 in Figure 6.3) and may result from recognition of the genuine 19S promoter. While not having a great impact on the amount of β-galactosidase produced, the exercise of introducing a UAS has shown that the stimulatory action of yeast UASs is not confined to native yeast promoters and UASs can be effectively used to elevate transcription from foreign promoters.

7.5 Summary

A galactose-inducible yeast upstream activator sequence (UAS) was placed in front of the 19S-lacZ gene. In the presence of galactose, the steady-state level of 19S-lacZ RNA was enhanced 12-fold and β-galactosidase activity was increased 5-fold. The UAS also conferred glucose repression on 19S-lacZ expression.

Footnote: Northern blot analyses of lacZ-hybridising RNAs in yeast were difficult to interpret and are therefore not included in this thesis. The pattern obtained (a smear with few discrete bands) appeared to indicate the presence of multiple "aberrant" transcripts, possibly initiating within the gene and terminating heterogeneously.

7.6 Resumé

Information from chapters 4,6 and 7 concerning expression of the 19S-lacZ gene can be summarised as follows:

There is much background transcription of the lacZ gene in yeast; the origin and polarity of these transcripts is not known. Transcripts arising in plasmid vector sequences which continue into the 'lacZ gene in pPA may terminate within the CaMV DNA in pPA1. 19S-lacZ expression does not require promoter signals in the adjacent plasmid DNA nor transcriptional readthrough from this region, ie the CaMV DNA contains one or more bona fide promoters. Primer extension experiments revealed a major RNA species initiating about 250bp upstream of the 19S-lacZ coding region, suggestive of a fortuitous promoter shortly upstream. Two further transcripts initiating at +9 and +25bp were detected. The initiation sites of these RNAs is consistent with their being the result of 19S promoter TATA box recognition. Two facts suggest that the more abundant transcript initiating at -250bp is not a functional mRNA for β-galactosidase. The first is that the 5' non-coding region of this RNA contains at least 3 AUG codons, which would interfere with its translation. The second is that activation of the yeast enhancer in pPA9 resulted in massive overproduction of this RNA species but was not accompanied by an equivalent increase in β-galactosidase synthesis. This suggests that one or both of the 19S-lacZ transcripts initiating at +9 and +25bp are mRNAs for β-galactosidase. On the basis of this information, a mutational analysis of the 19S promoter region was undertaken to define specific sequences important for 19S-lacZ expression in yeast. This work is described in chapter 9.

Table 7.1 β-galactosidase activity specified by various

plasmids in yeast

| Plasmid	Gene fusion	Carbon source	β-galactosidase activity*
pPA1	19S–lacZ	glucose	0.74
pPA9	UAS–19S–lacZ	glucose	0.095
pPA9	UAS–19S–lacZ	galactose	3.48

* nmoles ONPG converted / min / 10^7 cells at 28°C

Figure 7.1   A.   The GAL1-GAL10 intergenic region.

Sites of mRNA initiation are indicated by (▲).   Bars represent the

GAL4 protein binding sites and are marked L or H for low or high

affinity.   Bracket corresponds to the core UAS delimited by deletion

analysis which retains greater than 60% UAS activity.   The positions

of RsaI and AluI restriction sites used for subcloning of the UAS are

shown.

B.   Cloning the GAL UAS into pPA1.

Plasmid pBM150 was restricted with BamHI and EcoRI.   A 760bp fragment

carrying the GAL1-GAL10 promoter region was recovered from a

polyacrylamide gel and further restricted with RsaI and AluI.   This

yielded a 143bp blunt ended fragment containing the UAS, which was

purified from a polyacrylamide gel and ligated into the SmaI site of

plasmid pPA1 (after SmaI digestion the recipient plasmid had been

treated with alkaline phosphatase to reduce the frequency of

self-ligation).   Replacement of a 360bp SmaI fragment with the

RsaI-AluI fragment was demonstrated by restriction analysis and

confirmed by Southern blotting.

A

GAL 10

UAS$_G$

L H L   H

Rsal   Alul

GAL 1

B

(Sm)

p19S

UAS$_G$

LEU2

lac Z

pPA9
12·7 kb

2μori   amp
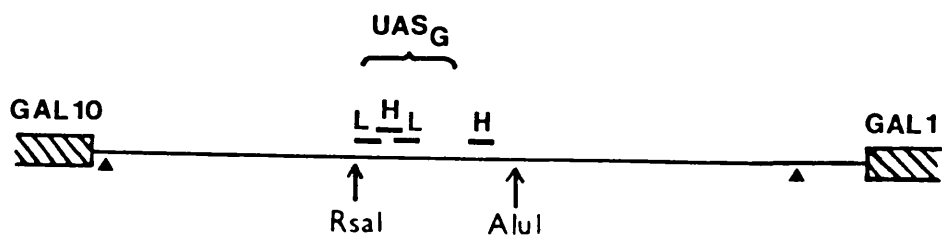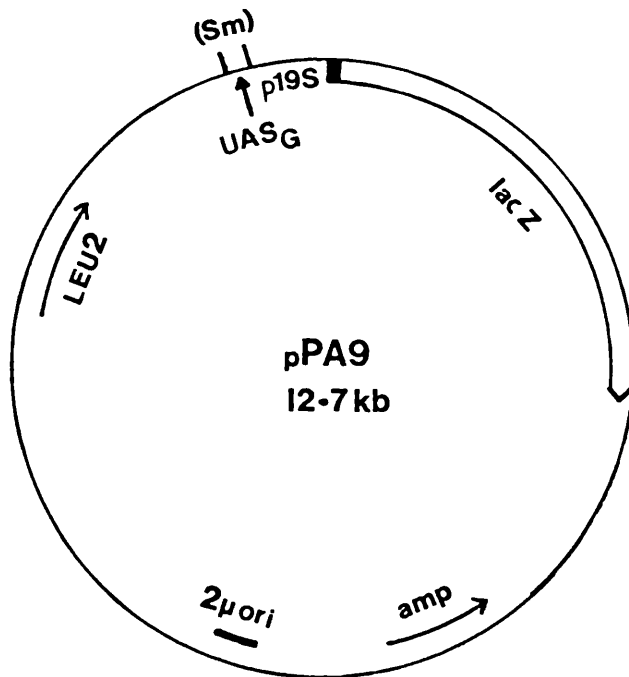
Figure 7.2 Dot blot analysis of yeast RNAs.

RNA was extracted from log-phase yeast cultures harbouring plasmids
and growing in medium containing 2% glucose (YEp213, pPA, pPA1, pPA9)
or 5% galactose (pPA9 only). 5 spots, each containing 4μg total RNA,
were applied to the nitrocellulose filter and hybridised with either
a 2 micron plasmid B gene probe or a lacZ gene probe. The figure
shows an overnight exposure with intensifying screens. After
autoradiography, the spots were cut out and the amount of probe
hybridised was measured in a scintillation counter. From this data
the relative amounts of lacZ-hybridising RNA were calculated.
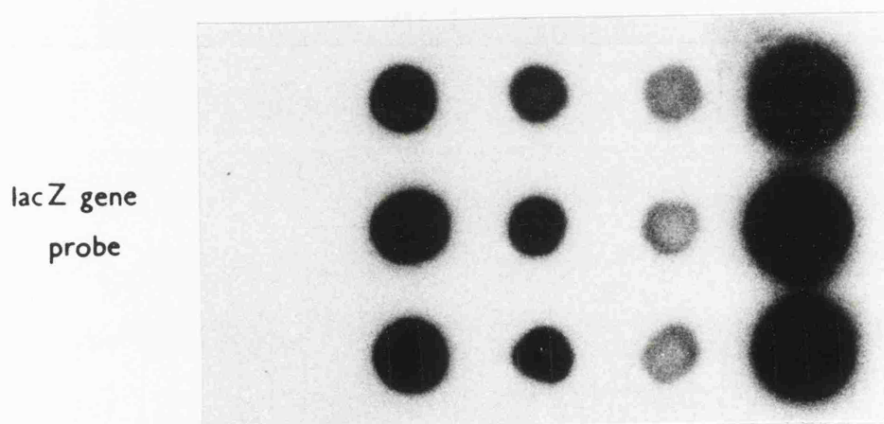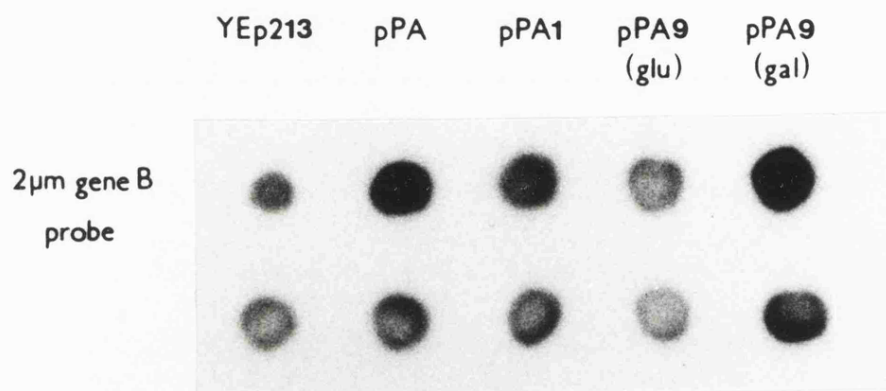
|	YEp213	pPA	pPA1	pPA9 (glu)	pPA9 (gal)
2 μm gene B probe					
lac Z gene probe					
relative lacZ RNA levels:	0·06	2·82	1·00	0·63	12·52

A
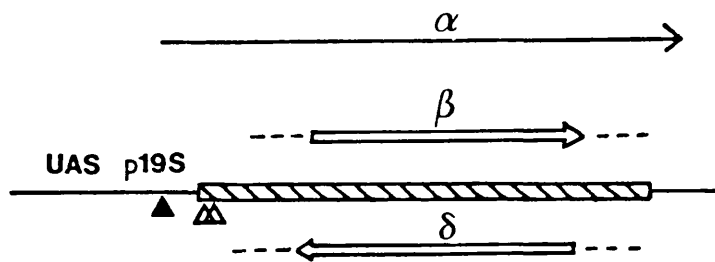
B
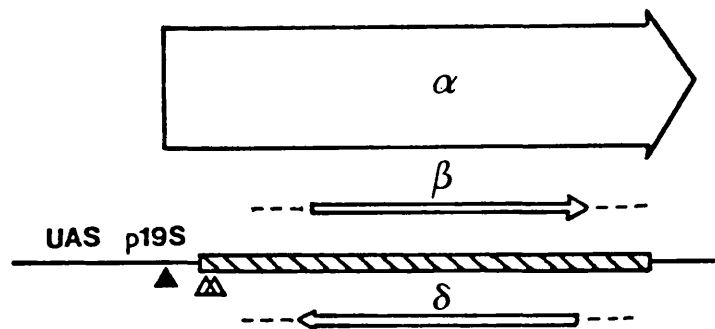
Figure 7.3 Model of 19S-lacZ gene transcription from pPA9 before (A) and after (B) galactose induction of the GAL UAS.   Arrows β and δ represent theoretical classes of RNA, the synthesis of which is not affected by activation of the UAS.   Arrow α represents a specific galactose-inducible RNA species arising from a fortuitous promoter near the UAS.   The relative abundances of these transcripts (or collections of transcripts) is indicated by the depth of the arrows.   The initiation points of transcript α (▲) and of two other 19S-lacZ transcripts detected by primer extension (△) are indicated.

CHAPTER 8

CONSTRUCTION OF DELETIONS WITHIN THE 19S PROMOTER REGION

AND THEIR ANALYSIS IN YEAST

8.1 Introduction

The 19S promoter of CaMV, responsible for the transcription of gene VI, is less well characterised than the 35S promoter. At present it is not known which specific sequences upstream of gene VI are required for its efficient transcription in plant (or any other) cells. In the remaining part of this thesis I describe attempts to identify and compare the DNA sequences required for 19S promoter activity in plant cells and in yeast. This chapter describes the construction of a series of deletion mutations and their effects on 19S promoter function in yeast. Analysis of the promoter mutations in transgenic plants is presented in chapter 9.

8.2 Choice of a reporter gene

To facilitate their analysis *in vivo*, the 19S promoter and its deleted derivatives were again fused to a reporter gene *ie* a structural gene with an easily assayed product. It would have been convenient to use the already existing 19S-lacZ gene fusion as a basis for the study but measurement of β-galactosidase in plant tissue is complicated by the existence of an endogenous β-galactosidase which appears to be ubiquitous among plant species (see Harding 1984 for references). Preliminary experiments revealed the endogenous β-galactosidase to be easily measurable in protoplasts isolated from radish, turnip, tobacco and petunia species. The endogenous activity was sufficient to mask expression of the 19S-lacZ gene in turnip protoplasts transfected with

pPAl DNA (Figure 8.1) indicating that a transient assay system based on inoculation of plant protoplasts would not be suitable for the analysis of 19S-lacZ gene fusions. The problem of a high background activity might be overcome by employing a more efficient transformation system so that a greater proportion of the target cells receive and express the input DNA (in a transient assay such as that used here, the proportion is typically fewer than 1%). A.tumefaciens-mediated DNA transfer gives rise to callus tissue in which all cells, theoretically, are stably transformed. Using this technique, Helmer *et al* (1984) found only a two-fold elevation of β-galactosidase activity in tobacco callus transformed with a chimeric nopaline synthase-lacZ gene. The same study revealed the native β-galactosidase activity to be exceedingly variable in cultured tissue, further reinforcing the unsuitability of this enzyme as an indicator of gene expression in plants.

In view of these facts, an alternative reporter gene was deemed necessary for use in plants. The chloramphenicol acetyltransferase (CAT) gene derived from the bacterial transposon Tn9 was chosen because a variety of biochemical enzyme assays are available and no endogenous CAT activity is found in those plant species most commonly used for genetic analysis.

8.3 Construction of a 19S-CAT gene fusion

A translational gene fusion was produced between CaMV gene VI and the CAT gene, the construction of which is described in Figure 8.2. The open reading frame comprises the first 19 codons of CaMV gene VI, 30bp of DNA from Tn9 containing no stop codons, followed by the entire CAT coding sequence. The 3' end of the A.tumefaciens nopaline synthase

gene, which contains a transcription termination signal, was placed downstream of this construct to ensure efficient termination of transcription in plants. The resulting plasmid was called p19S-CAT.

8.4 Functional analysis of the 19S-CAT gene in yeast

The 19S-CAT gene was introduced into yeast to ascertain whether the translational fusion was functional *ie* whether the hybrid CAT protein retained its enzyme activity despite possessing an N-terminal extension of 33 amino-acids. Plasmid YEp19SCAT was constructed for this purpose and contains the 19S-CAT gene followed by transcription termination signals from the yeast CYC1 gene (Figure 8.3).

Extracts were prepared from yeast cells harbouring the 19S-CAT gene and assayed for CAT enzyme activity using a ^{14}C-chloramphenicol TLC method. Briefly, cell extracts are incubated with ^{14}C-chloramphenicol and acetyl coenzyme A. In the presence of CAT, acetylation of the chloramphenicol occurs, giving two mono- and a rarer di-acetoxy form. The reaction products are separated by thin layer chromatography and detected by autoradiography. Figure 8.4a shows a typical assay comparing CAT activity in extracts of yeast cells harbouring YEp19SCAT, YEp213-12, which has no CAT gene, and pCH100, which carries a highly expressed ADC1-CAT gene fusion (Hadfield *et al* 1986). The assay shows no evidence of CAT activity in cells containing YEp19SCAT or YEp213-12. SDS-PAGE analysis of the same cell extracts reveals a typical array of proteins, suggesting that the negative result cannot be attributed to generalised proteolysis and also shows no discernible polypeptide of the expected molecular weight for CAT (26kd) in the YEp19SCAT extract (Figure 8.4b). The absence of CAT activity in yeast suggests that the 19S-CAT gene product is non-functional, although it is also possible

that this assay is not sensitive enough to detect extremely low levels of CAT activity. In either instance, the gene fusion is clearly not suitable for monitoring the effects pf 19S promoter deletions in yeast. The 19S-CAT gene was retained as the marker gene in plants, where CAT activity might prove detectable (due to more efficient functioning of the promoter?) and the CAT gene should in any case be suitable for transcriptional analysis.

8.5 Construction of 19S promoter deletions

Two sets of deletions were generated by the action of Bal31 nuclease on p19S-CAT DNA. The first is a series of deletions in which DNA is progressively deleted from the 5' end of the promoter (plasmids 2-6, Figure 8.5). The deletions terminate 256, 185, 131, 62 and 21bp from the start of the coding region. The second series (plasmids 11-14) consists of a group of internal deletions ranging in length from 41bp to 71bp. The construction of these mutations is described in detail in Figure 8.5. The precise end-points of the deletions were determined by sequence analysis, for which purpose the HindIII-SacI fragments encompassing the deleted promoters in plasmids 11 to 14 were isolated and cloned into M13 vectors mp18 or mp19. DNA sequences spanning the deletions are presented in Figure 8.6.

8.6 Analysis of the 19S promoter deletions in yeast

The original (undeleted) 19S-CAT gene produced no detectable CAT activity in yeast, so the mutated 19S promoters were fused again to the lacZ coding region in order to compare and quantify their activities in yeast. This was achieved by substituting the deleted promoters for the full length promoter in pPA5. *nb* this plasmid contains a transcription

termination signal between the 19S promoter and vector sequences upstream. Four deletions (numbers 2, 11, 12, and 14), between them spanning greater than 80% of the cloned CaMV DNA, were fused to lacZ by this means (see Figure 8.7 for details). β-galactosidase assays performed on cells containing pPA5 and its derivatives showed that deletions 2, 12 and 14 had no significant effect on 19S-lacZ expression (Table 8.1). Deletion 11, on the contrary, caused a marked reduction in β-galactosidase activity implying that sequences between -22bp and -62bp are involved in (and possibly sufficient for) expression of the 19S-lacZ gene in yeast. Interestingly, this region encompasses a putative TATA box centred at -41bp.

8.7 Discussion

The first point of interest emerge from this work concerns chloramphenicol acetyltransferase and its limitations as a reporter enzyme. The CAT gene is widely used as a reporter in molecular studies of gene regulation but in virtually all instances the constructs analysed are transcriptional gene fusions, giving rise to wild type CAT enzyme. In this study, for reasons of speed and simplicity, a translational gene fusion was constructed between CaMV gene VI and the CAT gene of Tn9. The product of the gene is a novel CAT protein bearing a 33 amino-acid extension at the N-terminus (Figure 8.2). In yeast cells harbouring this gene, CAT activity was completely undetectable. Some possible explanations for this include i) limited sensitivity of the biochemical assay, ii) rapid turnover of the 19S-CAT mRNA or translation product, iii) reduced activity of the novel CAT protein. None of these factors has been addressed experimentally but the latter is perhaps the more likely, *ie* the additional amino-acids at the N-terminus of the CAT protein interfere

with its catalytic activity. Very few translational gene fusions using CAT have been described in the literature. In one instance CAT activity was retained (Hentze *et al* 1987) but in others no enzyme activity was recorded (Stark 1987). E.coli cells containing a fusion between the lacZ α-peptide and CAT were not selectable on agar plates containing chloramphenicol despite the fact that the hybrid CAT protein formed 10% of the total cellular protein (M. Stark pers. comm.). The available data suggest that some modifications of the CAT enzyme destroy its catalytic activity whereas others do not. Those modifications which abolish activity might influence the tertiary structure of the enzyme, rendering the active site (centred around amino-acid 193, HIS) less readily accessible. Alternatively, the additional amino-acids might affect the quaternary structure of the enzyme, for example by inhibiting formation of the trimeric structure in which native CAT enzyme exists.

Analysis of the 19S promoter mutations in yeast revealed that sequences removed by deletions 2(-500 to-256bp), 14(-256 to -186bp) and 12(-136 to -63bp) are not required for expression of the 19S-lacZ gene, whereas sequences between -62 and -21bp appear to be critical. The data provides direct evidence that the fortuitous promoter located at about -350bp (see chapter 6), which gives rise to the longest and most abundant 19S-lacZ RNA, does not contribute to β-galactosidase production. Synthesis of this long RNA is presumably abolished by deletion 2, which would remove the promoter, yet β-galactosidase activity is unaffected. As postulated in chapters 6 and 7, therefore, the long RNA is almost certainly not a functional .mRNA and it is probably one or both of the shorter RNAs initiating at +9bp and +25bp which is translated to give β-galactosidase. The production of β-galactosidase appears to be connected with a sequence element

eliminated in deletion 11. This mutation removes the 19S promoter TATA box, hence it is tempting to speculate that a) β-galactosidase production is related to the presence of the two shorter 19S-lacZ RNAs and b) the viral TATA box at -41bp is required for synthesis of these transcripts. While direct evidence is lacking, the data is highly suggestive of a role for the CaMV TATA box in 19S-lacZ gene expression, implying the functional recognition of a viral promoter element in yeast.

8.8 Summary

A translational gene fusion, in which the 19S promoter and 5' end of CaMV gene VI were fused to the bacterial CAT gene, was found to be non-functional in yeast. Using Bal31 nuclease, a series of mutations within the 19S promoter region was produced. Analysis of some of these mutations in yeast* revealed that sequences between -500 and -193bp and also between -137 and -62bp are not required for 19S promoter function. A deletion spanning -62 to -21bp (which removes the gene VI TATA box) severely reduced 19S-lacZ expression.

* using lacZ as a reporter gene

Table 8.1 β-galactose activity in yeast strains harbouring

19S-lacZ genes with promoter deletions

| Plasmid	Region deleted[+]	β-galactosidase activity*
pPA5	none	1.1
pPA5Δ2	−500 to −256	1.2
pPA5Δ14	−257 to −185	1.1
pPA5Δ12	−137 to −62	1.2
pPA5Δ11	−63 to −21	0.12

* nmoles ONPG converted / min / 10^7 cells at 28°C.

+ position (bp from the start of CaMV gene VI coding

region) of the nucleotides flanking the deletion.

2·0

relative enzyme activity

1·0

1   2   3

time (days) after DNA inoculation
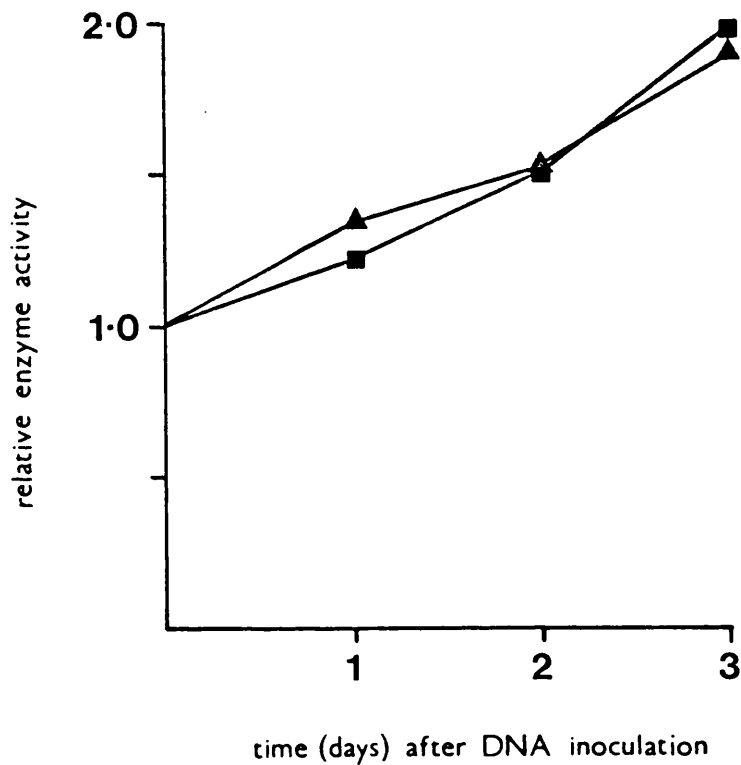
Figure 8.1   β-galactosidase activity in pPA1 transfected protoplasts.

Freshly prepared protoplasts were inoculated with plasmid pPA1 using PEG-stimulated DNA uptake. A control batch of protoplasts were mock transfected ie similarly treated with PEG but omitting the DNA. At 0, 1, 2 and 3 days after inoculation, samples containing 0.5×10^6 cells were removed from culture and stored at $-20^\circ C$. β-galactosidase assays were performed simultaneously on the thawed samples. The figure shows enzyme activity relative to that immediately after inoculation (arbitrary units). (■) = pPA1 transfected cells, (▲) = control cells.

Figure 8.2

A. Structure of the 19S-CAT gene.

The coding region of the hybrid CAT gene is represented by the boxed
area. The filled portion consists of the first 19 codons of CaMV
gene VI and is preceded by the 19S promoter. The hatched portion is
derived from pUC18 polylinker and a 30bp sequence which precedes the
CAT open reading frame in Tn9 (14 codons). The open box consists of
the complete CAT coding region (220 codons). The bold line is DNA
from A.tumefaciens comprising the last 150bp of the nopaline synthase
coding region and 695bp of downstream DNA which includes a plant
transcription termination signal (nos$_t$). The arrow beneath the gene
is the 1200nt transcript predicted in plants.

B. Construction of p19S-CAT.

Cloning vector pUC18 was digested with BamHI and ligated with a 559bp
Sau3A fragment of CaMV (containing the 19S promoter) which had been
isolated from plasmid pLW414. Ligation products were restricted with
BamHI prior to E.coli transformation and recombinant plasmids having
the desired orientation of the CaMV insert were identified by
restriction mapping. The CAT gene, with 30bp of upstream DNA was gel
purified from pCM1 (Close and Rodriguez 1982) as a 779bp SalI
restriction fragment and ligated into pUC18/CaMV which had been
digested with SalI and treated with alkaline phosphatase. A plasmid
having the desired orientation of CAT gene was identified and further
restricted with HindIII and SphI. The 3' end of the nopaline
synthase gene was obtained from plasmid pGV601 as an 0.84kb SphI-
HindIII fragment and ligated with the restricted plasmid. Cm^r Ap^r
transformants were selected and screened for the presence of nopaline
synthase DNA. (B) = BamHI/Sau3A junction, S = SalI, Sp = SphI, H =
HindIII, X = XmnI.

A

0   1·0   2·0 kb

p19S   CAT   nos$_t$

B

(B) H

p19S

(B)

S

CAT

p19SCAT
4·86 kb

amp

X

S
Sp

nos$_t$

H
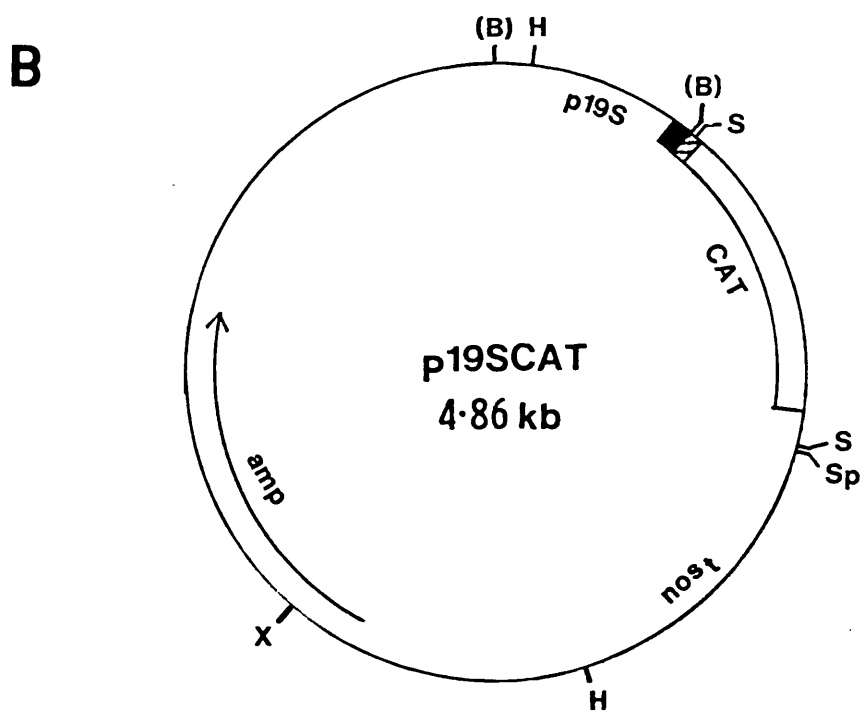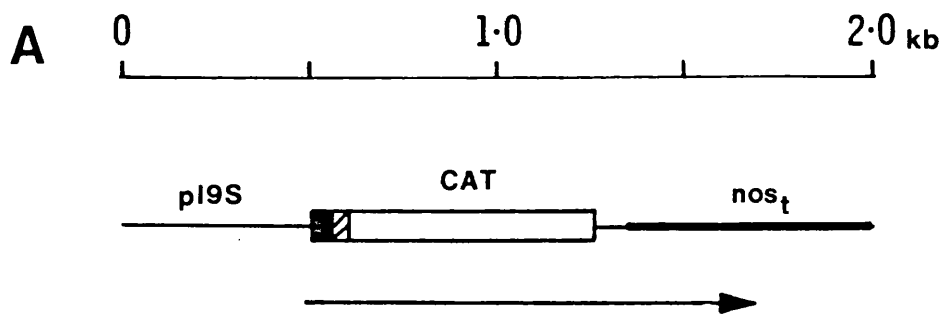
Figure 8.3 Construction of YEp19SCAT.

The plasmid was constructed in two stages. A 300bp SalI-XhoI fragment bearing the transcription termination signals from the yeast CYC1 gene was purified from pC78/28 and cloned into the SalI site of YEp213-12. A plasmid having the desired orientation of the CYC1 terminator was designated YEp20. The 19S-CAT gene was recovered from p19S-CAT as a HindIII-SphI restriction fragment and ligated into YEp20 which had been digested with HindIII and SphI. H = HindIII, Sp = SphI, S = SalI, (X) = SalI/Xho joint.

H

p19S

CAT

Sp
S

CYCt

(X)

LEU2

YEp19S-CAT
11·7 kb

2μ ori

amp
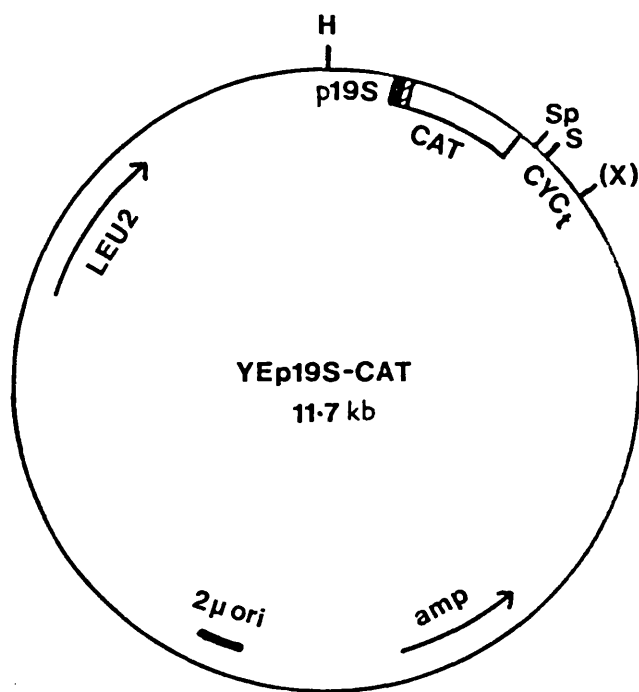
Figure 8.4   A.   CAT activity in various yeast cell extracts.

a)   extract of S150-2B harbouring YEp213 (300μg protein)

b)   extract of S150-2B harbouring YEp19SCAT (300μg protein)

c)   as for b except cells were grown in the presence of 500μg/ml
chloramphenicol.

d)   extract of a distillery yeast strain harbouring pCH100 (less than
10μg protein)

e)   no cell extract

reaction times were 90 minutes (a,b,c e) or 10 minutes (d).

B.   SDS-PAGE analysis of yeast cell extracts.

Extracts a, b and c (as above) containing 70μg protein were
fractionated in an 11% polyacrylamide gel and stained with Coomassie
blue.
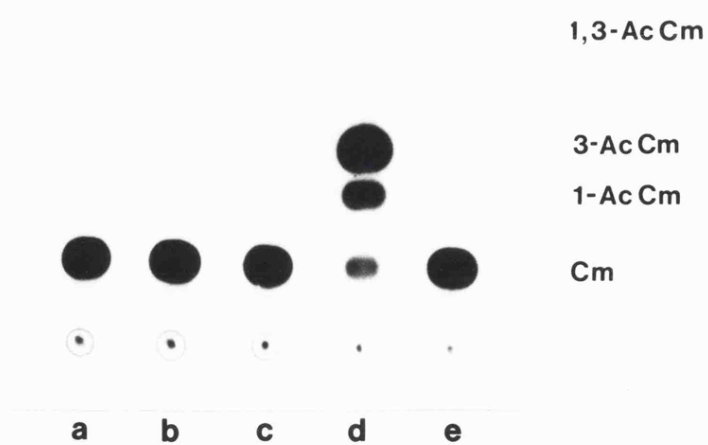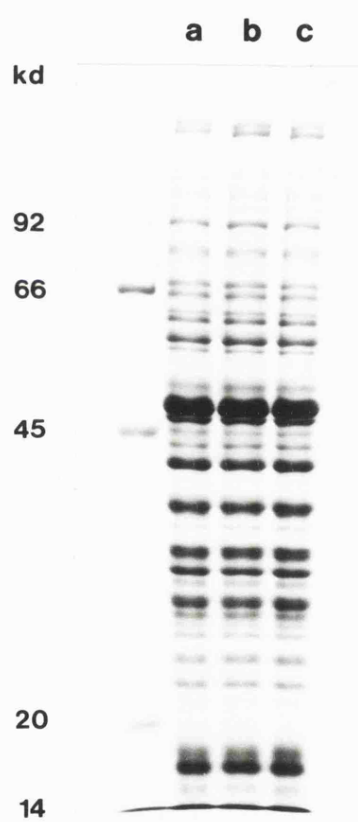
<u>Figure 8.5</u> Construction of 19S promoter deletions.

p19S-CAT was digested with HpaI or EcoRV and treated with Bal31 nuclease to produce a range of deletions. DNA termini were rendered flush by treatment with klenow, linkers were added and the plasmids recircularised by ligation. A large number of plasmids were screened by restriction analysis to determine the approximate end-point of the deletion and 5 were selected for use. The full length CaMV promoter and part of the hybrid CAT gene is shown at the top of the figure. Sm = SmaI, H = HindIII, Hp = HpaI, E = EcoRI, V = EcoRV, S = SacI. T = TATA box, ▲ cap site (plants), △ cap site (yeast). Beneath are deleted plasmids 2 to 14 with deletion end-points (relative to the translation start) shown to the right. B = BamHI linker, Bg = BglII linker. The construction of individual plasmids was as follows.

<u>SERIES I</u>

2.   p19S-CAT was restricted with HpaI and SmaI then recircularised with BamHI linkers.

3,4,6.   p19S-CAT was restricted with HpaI, treated with Bal31 nuclease then klenow, digested with SmaI and recircularised with BamHI linkers.

5.   p19S-CAT was digested with EcoRV and SmaI then recircularised with BamHI linkers.

<u>SERIES II</u> (used in the construction of plasmids 11-14)

7.   p19S-CAT was digested with EcoRV, BglII linkers were added and the plasmid religated.

8,9.   p19S-CAT was digested with EcoRV, treated with Bal31 nuclease and klenow, BglII linkers were added and the plasmids religated.

10.   p19S-CAT was restricted with HpaI, BglII linkers were added and the plasmid religated.

<u>SERIES III</u>

Internal deletions 11 to 14 were constructed by coupling series I and II plasmids so that 11 is a combination of 6 and 7; 12 of 5 and 8; 13 of 4 and 9; 14 of 3 and 10. Plasmids were cleaved into two parts by restriction with XmnI (which cuts at a single site within the amp gene - see figure 8.2) and either BamHI (series I plasmids) or BglII (series II plasmids). Relevant pairs were combined, ligated and subsequently restricted with BamHI and BglII to enrich for seriesI/seriesII recombinant molecules.

11   5'TGTGGCTGAT | CAGATCCCG | GAGAAAATC3'

12   5' GCGGACTTC | CAGATCCCG | ATCAAAAAG3'

13   5'TGGCTTAGCC | CAGATCCCG | AAGAGAATT3'

14   5'GAGTTTTGTT | CAGATCCCG | GTTTGATGT3'
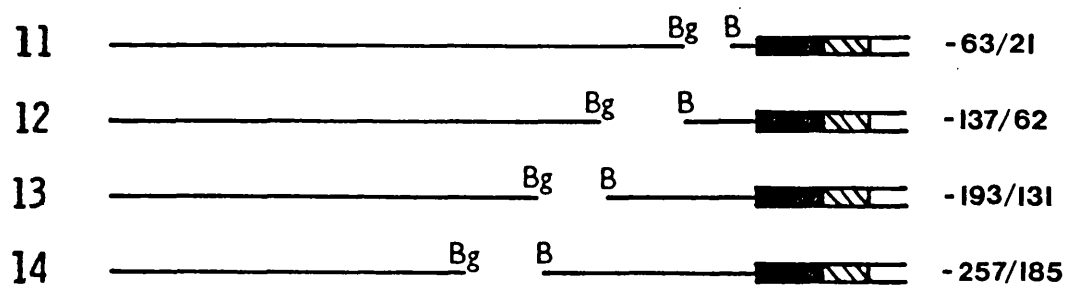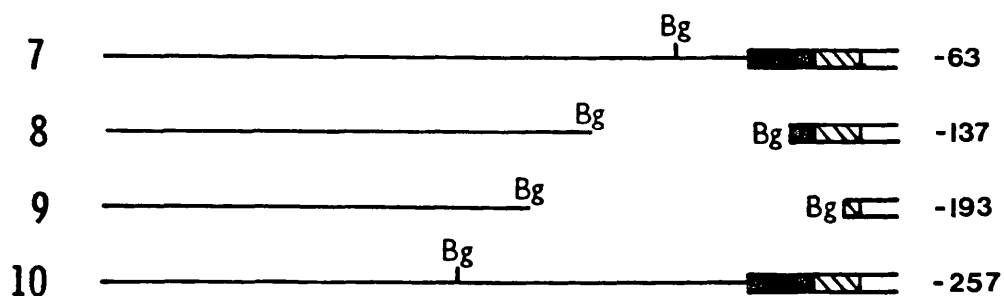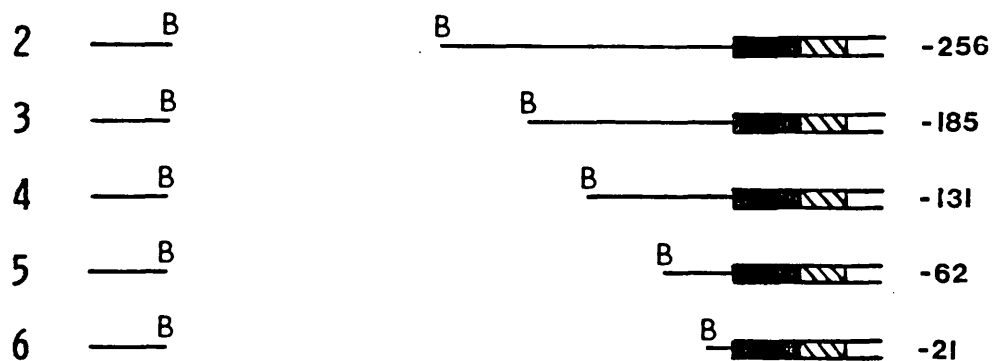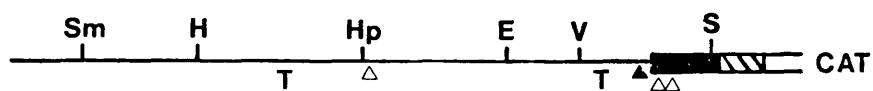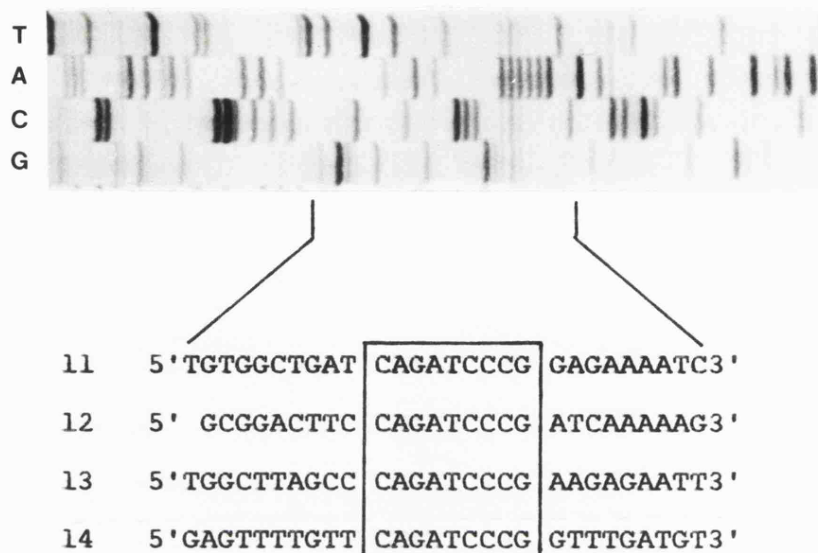
Figure 8.6   DNA sequence at the BamHl/BglII juncture in plasmids 11 to 14.   Sequences of the coding strands are given. Boxed area is the recombinant BamHI/BglII linker. The junction sequence of plasmid 11 is shown as an illustration.
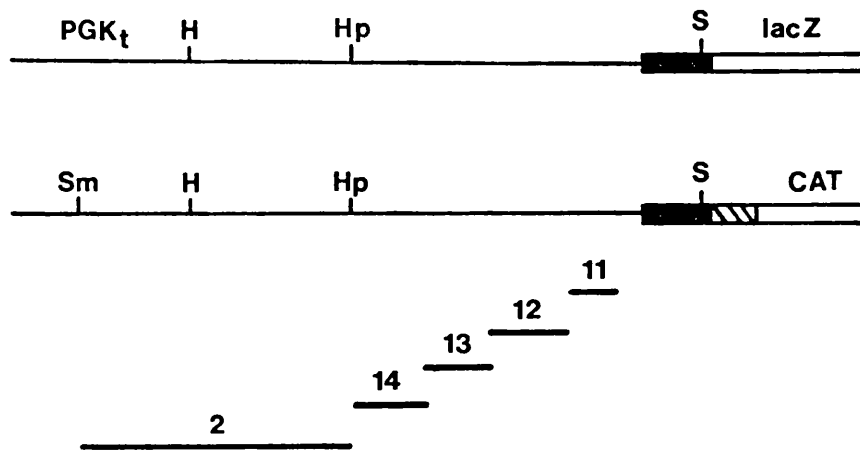
Figure 8.7 Construction of mutant 19S-lacZ genes.

CaMV DNA encompassing deletions 11, 12 and 14 were purified as HindIII-SacI restriction fragments and substituted for the full length HindIII-SacI fragment of pPA5. A 19S-lacZ gene bearing promoter deletion 2 was generated by recovering a HpaI-SacI CaMV fragment from p19SCAT and inserting this in place of the HindIII-SacI fragment of pPA5 (nb the HindIII 4bp overhang of pPA5 was first made flush by treatment with Klenow fragment). The figure shows the relevant sections of pPA5 and p19SCAT. The extent of deletions 2, 11, 12 and 14 are indicated beneath. H = HindIII, Hp = HpaI, S = SacI, PGK$_t$ = transcription termination signal of the yeast PGK gene.

CHAPTER 9

INVESTIGATION OF THE EFFECTS OF MUTATIONS WITHIN THE 19S PROMOTER

IN TRANSGENIC PLANTS

9.1   Introduction

Transgenic plants, that is plants in which a segment of foreign DNA is
stably integrated into the nuclear genome, provide a vital tool for
dissecting the molecular basis of plant gene regulation.   In this
chapter I describe the production and analysis of transgenic tobacco
plants harbouring the 19S-CAT gene and its deleted derivatives.   The
production of such genetically engineered plants requires the genetic
modification of individual cells, which are then cultured *in vitro* and
ultimately regenerated into mature plants.   The initial transforming
event has been achieved by micro-injection of plant protoplasts (Reich
1986, Crossway *et al* 1986), liposome fusion with protoplasts (Deshayes
et al 1985) and by PEG-stimulated DNA uptake (Paszkowski *et al* 1984).
The most widely used DNA delivery system, and that used in this study,
exploits   the   natural   DNA   transfer   mechanism   of   Agrobacterium
tumefaciens.   Nicotiana tabacum (tobacco) was chosen as the recipient
plant   species,   being   easier   to   propagate   in   tissue   culture   than
Brassica species, which are the natural hosts of CaMV.   Both the 19S
and 35S promoters of CaMV have previously been shown to function
efficiently in tobacco cells (Paszkowski *et al* 1984, Odell *et al* 1985,
Shewmaker *et al* 1985).

Plant transformation by A.tumefaciens

A.tumefaciens is a soil bacterium and phytopathogen which invades plant
tissue at the site of a wound, causing the formation of a tumour.   The

genetic basis of tumour formation has been elucidated and hinges upon the transfer of a specific DNA segment (called T-DNA) from the bacterium into plant cells, where it becomes incorporated into the nuclear genome. Subsequent expression of genes within the T-DNA has a profound effect on the metabolism of the transformed cell, which begins to synthesise novel compounds (opines) that are catabolised by the infecting bacteria. More importantly, the T-DNA specifies the production of phytohormone biosynthetic enzymes, which release the cells from the natural hormonal control of growth and differentiation and lead to tumour formation.

The T-DNA is derived from an extra-chromosomal element in A.tumefaciens, the Ti (tumour inducing) plasmid, and is flanked by a 25bp direct repeat. These repeats or "border sequences" are cis-acting elements essential for efficient transfer of the T-DNA into plant cells. A number of trans - acting factors also required for T-DNA mobilisation are encoded by a separate region of the Ti plasmid, the vir (virulence) region. vir genes are activated when the bacterium encounters specific phenolic compounds released by plant cells in response to wounding.

This highly specialised mechanism of gene transfer has been exploited by molecular biologists following the observations that i) any genetic material inserted between the T-DNA border sequences is transferred to the plant nucleus ii) deletion of the oncogenic T-DNA genes does not impair T-DNA transfer and permits normal rather than neoplastic growth of the transformed tissue.

<u>Binary Agrobacterium vectors</u>

A variety of disarmed (non-oncogenic) Ti plasmid-based vectors have
been developed for use in transforming plant tissue, the most
sophisticated to date being the binary vectors. Binary systems make
use of the fact that <u>vir</u> gene products act in <u>trans</u>, so that the T-DNA
and <u>vir</u> genes can exist on separate replicons. The binary system used
in this study is that developed by Bevan (1984) and has two essential
components:

1) <u>A. tumefaciens</u> strain LBA4404, which contains a deleted Ti plasmid
with an intact <u>vir</u> region but no T-DNA.

2) plasmid Bin19, which contains T-DNA left and right border sequences
flanking a site for the insertion of foreign DNA (the pUC19 polylinker)
and a selectable marker conferring kanamycin resistance on transformed
plant cells. Bin19 is derived from the wide host range plasmid RK2 and
consequently is able to replicate in both <u>E.coli</u> and <u>A.tumefaciens</u>.

The introduction of foreign DNA into plant cells using Bin19 is
achieved in the following steps. The gene of interest is first
incorporated into the T-region of Bin19 by insertion into the
polylinker sequence. Recombinant binary plasmids are subsequently
transferred from <u>E.coli</u> to <u>A.tumefaciens</u> by conjugation. Helper
plasmid pRK2013 is required for the transfer of binary plasmids between
bacteria (Ditta *et al* 1981) so a tri-parental mating is performed
involving the donor strain (<u>E.coli</u>/ binary plasmid), the recipient
strain (<u>A.tumefaciens</u> LBA4404) and the mobilising strain (<u>E.coli</u>/
pRK2013). Progeny <u>A.tumefaciens</u> cells harbouring both the Ti plasmid
and binary plasmid are recovered from the mating and inoculated onto
tobacco leaf sections. Kanamycin selection is imposed after 48 hours

to restrict the growth of untransformed plant cells and is maintained throughout the process of plant regeneration.

9.2   Construction of A.tumefaciens strains carrying 19S-CAT genes inserted in Bin19

The complete 19S-CAT gene and deleted forms 2-6 and 11-14 (see figure 8.5) were cloned into the polylinker of Bin19 between the T-DNA borders (Figure 9.1). Care was taken to obtain each of the constructs in the same orientation.   The resulting binary plasmids were conjugally transferred to A.tumefaciens via a triparental mating and progeny Agrobacterium cells which had acquired the binary plasmid were selected on medium containing kanamycin and rifampicin.   None of the three parental strains showed any growth on this medium, E.coli being sensitive to rifampicin and A.tumefaciens to kanamycin.   The presence of binary plasmids in the kanamycin resistant A.tumefaciens strains was confirmed by Southern blotting (Figure 9.2).

9.3   Tobacco leaf disc transformation and regeneration of transgenic plants

Tobacco leaf discs were inoculated with binary plasmid - containing A.tumefaciens strains as described in chapter 2.   After 48 hours, the discs were placed on regeneration medium containing carbenicillin (to kill the bacteria) and kanamycin. Callus tissue began to appear at the periphery of A.tumefaciens-treated leaf discs after 12 days (Figure 9.3c) and after a further week the callus was detached and placed on shooting medium (Fig 9.3d).   Leaf discs which were not exposed to A.tumefaciens showed little or no callus formation after 21 days on regeneration medium (Fig 9.3b), indicating that kanamycin does indeed

restrict the growth of untransformed tissue. Shoots were formed from the calli after 4 to 8 weeks (Fig 9.3e) and when large enough to handle these were excised and embedded in rooting medium. Roots began to appear in 10 to 20 days and the rooted plantlets were removed from the sterile environment and potted up in compost when about 5cm tall (Fig 9.3f,g). After a period of acclimatisation in an environmental growth cabinet, plants were repotted and transported to a temperature and humidity controlled glasshouse where they were grown to maturity (Fig 9.3i). Self pollination of individual plants was encouraged by loosely covering the flower heads with plastic bags and seeds were later collected from all specimens. Between 2 and 10 transgenic plants were obtained for each gene construct.

9.4 Analysis of DNA from regenerated plants

Genomic DNA extracted from each of the regenerated plants was initially screened for the presence of 19S-CAT sequences by dot blot hybridisation (Figure 9.4). Of the 54 plants analysed, only one (19S-CAT plant 7) proved to be negative, attesting to the efficiency of kanamycin for selecting transformant tissue. The integrity of the 19S-CAT gene in these DNA samples was confirmed by Southern blotting (Figure 9.5). A CAT-hybridising restriction fragment of the predicted length was detected in all cases when genomic DNA was digested with BamHI or BamHI and HindIII, indicating that no major rearrangement had occured. The number of integrated copies of the 19S-CAT genes varied between 1 and 12 as estimated by densitometry scanning of the autoradiographs. Figure 9.4B shows the distribution of estimated copy numbers in the transgenic plants.

122

9.5 Transcriptional analysis of 19S-CAT gene expression in transgenic plants

Total RNA was prepared from each of the transgenic plants and also from the original (untransformed) tobacco strain SR1. RNA from three independently regenerated plants containing the intact 19S-CAT gene (plants 1, 3 and 8) was analysed by northern blotting but no CAT-hybridising transcripts were detectable (Figure 9.6, one plant only shown). This cannot be attributed to breakdown of the RNA because ethidium bromide staining of the samples after electrophoresis revealed the cytoplasmic and chloroplast ribosomal RNAs to be intact. Furthermore, the chlorophyll a/b binding protein (cab) gene transcript was detectable in the same samples and showed no signs of degradation (Figure 9.6). These findings suggest that the 19S-CAT gene transcript is not an abundant RNA species and that its detection will require the preparation of polyadenylated RNA. Unfortunately no further time was available in which to do this.

9.6 Discussion

To examine the effects of the 19S promoter mutations *in vivo*, a large number of transformed plants were produced, deemed to be transgenic by the following criteria: 1. kanamycin resistance during tissue culture, 2. positivity in neomycin phosphotransferase (NPT) II assays (some plants only), 3. southern blot analysis. CAT enzyme assays performed using leaf extracts from transgenic plants were negative (not shown), implying once again that the product of the 19S-CAT gene fusion has no catalytic activity (see chapter 8). Assessment of the relative efficiencies of the intact and deleted forms of the 19S promoter must therefore entail a comparison of 19S-CAT mRNA levels. The 19S-CAT gene

transcript proved to be undetectable when total RNA from 3 plants transformed with the intact 19S-CAT gene was examined by northern and dot blotting. Identification and quantitation of the 19S-CAT mRNA levels consequently requires the preparation and analysis of polyadenylated RNA fractions but owing to lack of time this has not been done. Thus, the effects of the mutations within the 19S promoter are not yet known and no conclusions can be drawn regarding the location of functional sequence elements within the 19S promoter.

9.7 Summary

Using an A.tumefaciens binary transformation system, transgenic tobacco plants have been generated which possess intact or deleted forms of the 19S-CAT gene. 19S-CAT mRNA was not detectable in total RNA extracted from plants containing the intact gene, indicating that analysis of 19S-CAT gene transcription will require the preparation of polyadenylated RNA.

<u>Figure 9.1</u> Cloning of <u>19S-CAT</u> genes into Bin19

p19S-CAT and plasmids with deletions 11-14 were cut with HindIII and

plasmids with deletions 2-6 were cut with BamHI and HindIII,

releasing the <u>19S-CAT</u> genes. Restricted plasmids were ligated with

Bin19 which had been digested with HindIII or BamHI and HindIII and

treated with alkaline phosphatase. White colonies (on X-gal/IPTG)

which were Km^r, Ap^r and Cm^r were screened by restriction mapping to

determine the orientation of the <u>19S-CAT</u> genes. The resulting series

of plasmids was designated pB19S-CAT, pB19S-CATΔ2 etc. H = HindIII,

RB, LB are right and left T-DNA borders.

H

pl9S

CAT

nost

H

Kan^r

NPT II

RB

LB

pBI9SCAT

12·1 kb

PRK252 replicon (Kan^r)
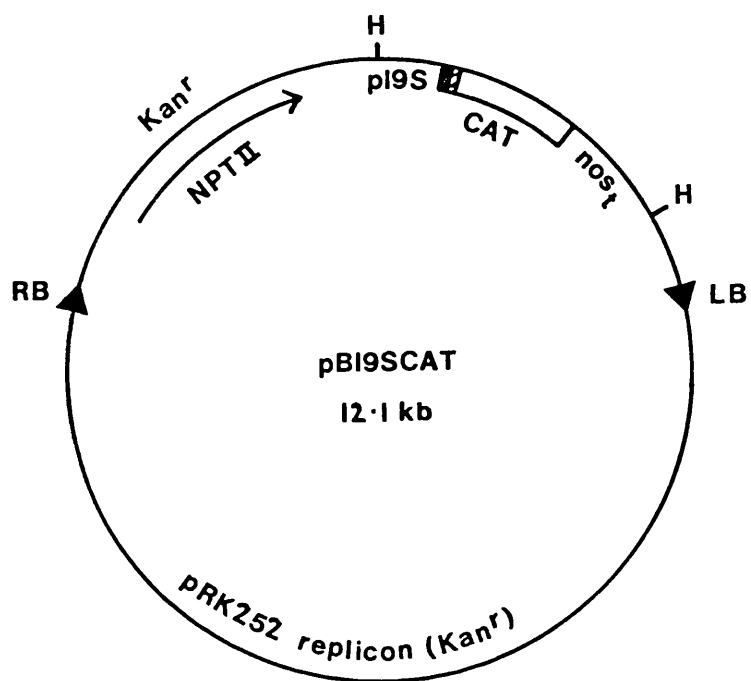
Figure 9.2 Southern blot analysis of A.tumefaciens DNA

10μl of DNA from kanamycin resistant A.tumefaciens clones generated during triparental matings, was digested with HindIII (pB19S-CAT, Δ11 to Δ14, Bin19) or BamHI and HindIII (Δ2 to Δ6) and fractionated in a 0.9% agarose gel. DNA was transferred to nitrocellulose and probed with p19S-CAT. A shows the ethidium bromide-stained gel prior to transfer. B shows a 2 hour autoradiographic exposure of the hybridised filter. Conjugally transferred plasmids were as follows. lane a: pB19S-CAT, lanes b-f: Δ2 to Δ6, lanes g-j: Δ11 to Δ14, lane k: Bin19.
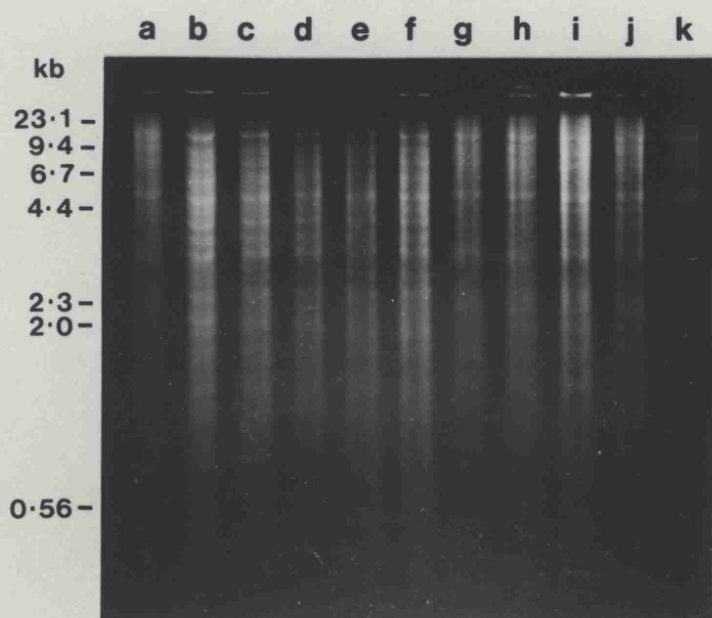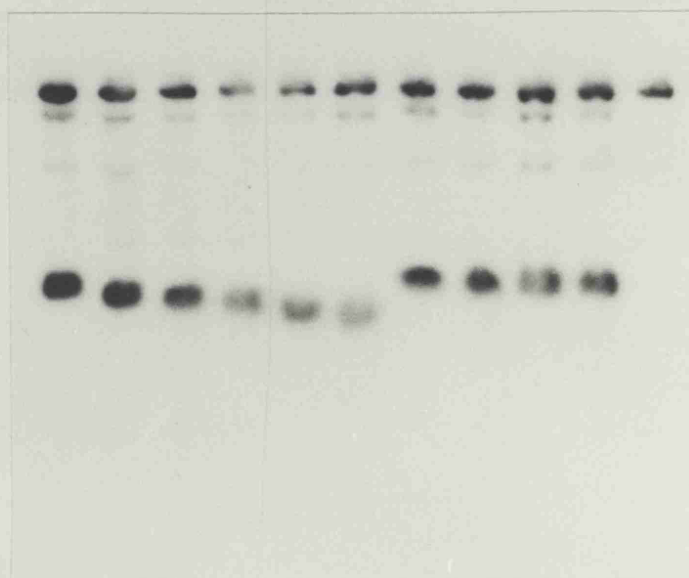
Figure 9.3 Regeneration of transgenic plants.

a - i show various stages of the tissue culture and plant
regeneration process.   time (t) = 0 on the day of A.tumefaciens
inoculation.

a) t = -2 days
freshly cut disks from a surface sterilised tobacco leaf

b) t = 3 weeks
control (uninfected) leaf disks after 3 weeks growth on regeneration
medium

c) t = 3 weeks
leaf disks previously inoculated with A.tumefaciens showing formation
of callus tissue

d) t = 4 weeks
calli continue to grow on shooting medium

e) t = 10 weeks
shooting calli

f) t = 14 weeks
rooted plantlet

g) t = 14 weeks
newly potted up plantlet

h) t = 16 weeks
established plant ready for transfer to glasshouse

i) t = 20 weeks
mature, flowering plant

a

b

c

d

e

f

g
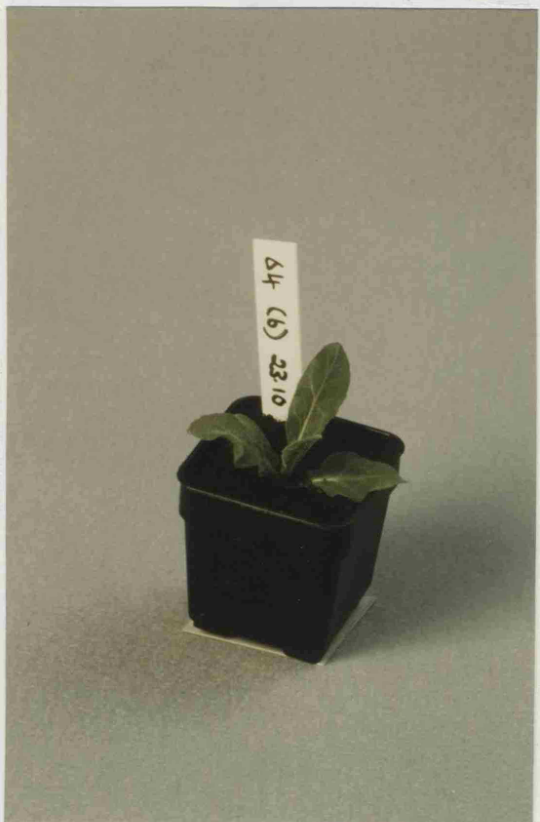
h

i

Figure 9.4

A. Dot blot analysis of DNA from regenerated plants

DNA samples (5μg) were spotted onto a nitrocellulose filter and hybridised with a CAT gene probe. DNA from the parental tobacco strain SR1 was included as a negative control (C). Standards (S) were 60, 30, 15, 7.5, and 3.75 picograms of plasmid p19S-CAT, which correspond to 16, 8, 4, 2, and 1 19S-CAT gene equivalents. The figure shows a 4 day exposure at -70°C with an intensifying screen.

B. Distribution of 19S-CAT gene copy numbers in transgenic plants

A gene construct
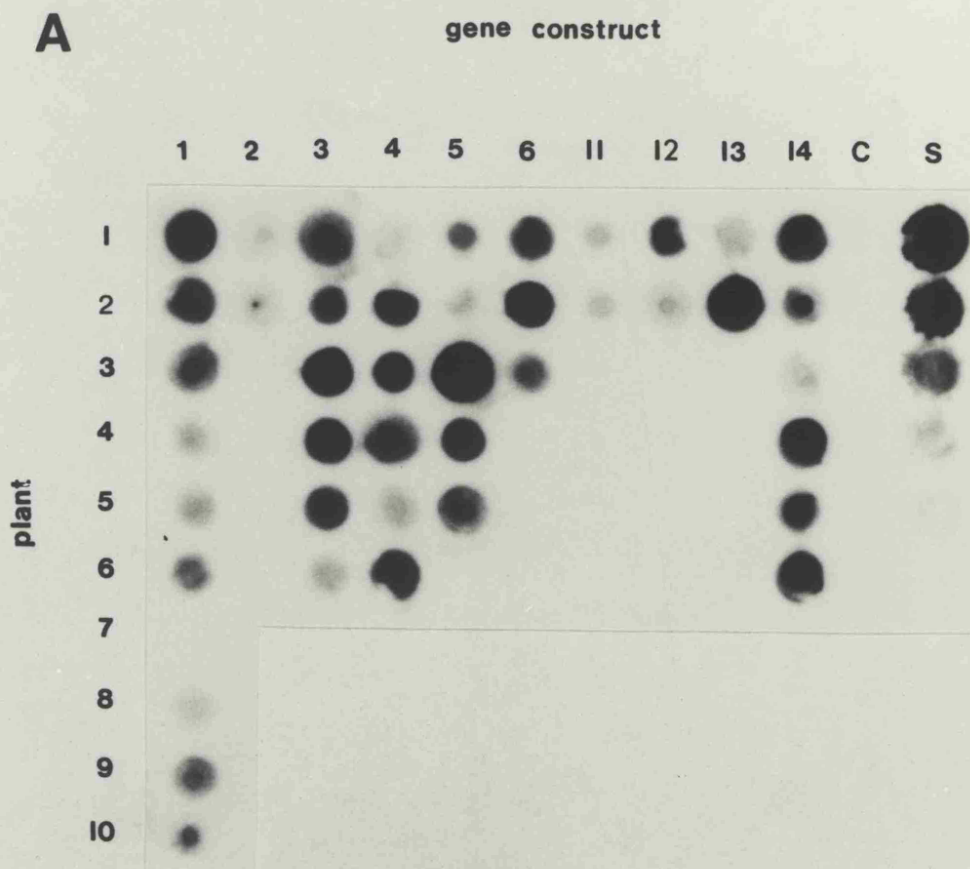
B

gene dose	number of plants
1	16
2–5	22
6–10	7
>.10	2

Figure 9.5 Southern blot analysis of DNA from regenerated plants

8µg of DNA was digested with HindIII (for intact 19S-CAT and Δ11 to

Δ14) or BamHI and HindIII (for Δ2 to Δ6) and fractionated in 1%

agarose gels. lanes a and b in each gel contain 120 and 12 picograms

respectively of p19S-CAT restricted with HindIII. These amounts are

equivalent to 16 and 1.6 19S-CAT genes per genome. DNA was

transferred to Hybond N membrane and hybridised with a CAT gene

probe. Figures a, b, and c show ethidium bromide-stained gels prior

to transfer and the corresponding autoradiographs developed after 4

days exposure at -70°C with intensifying screens.

a
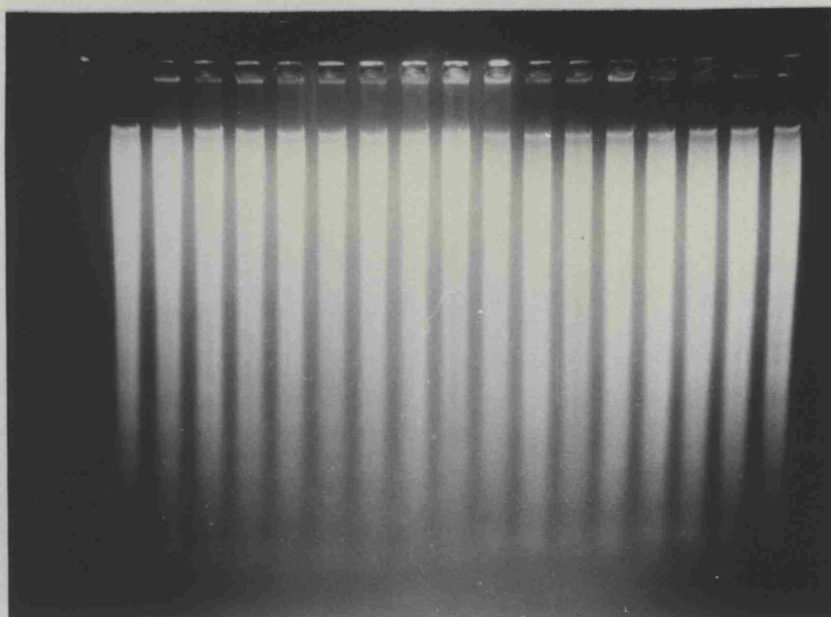
gene   I9SCAT   Δ2   Δ3
            ─────────────────────   ──────   ──────────────

plant   a b 1 2 3 4 5 6 8 9 10 1 2 1 2 3 4 5 6

b
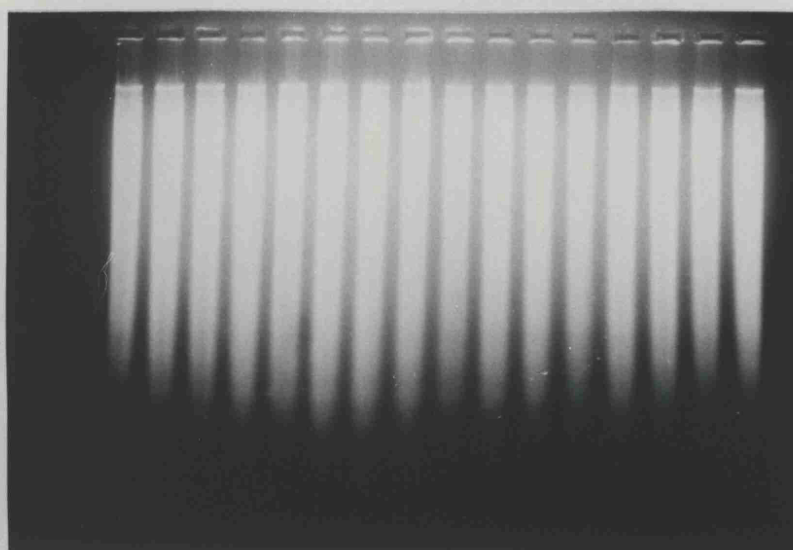
gene   Δ4   Δ5   Δ6

plant   a   b   1   2   3   4   5   6   8   9   1   2   3   4   5   1   2   3

C
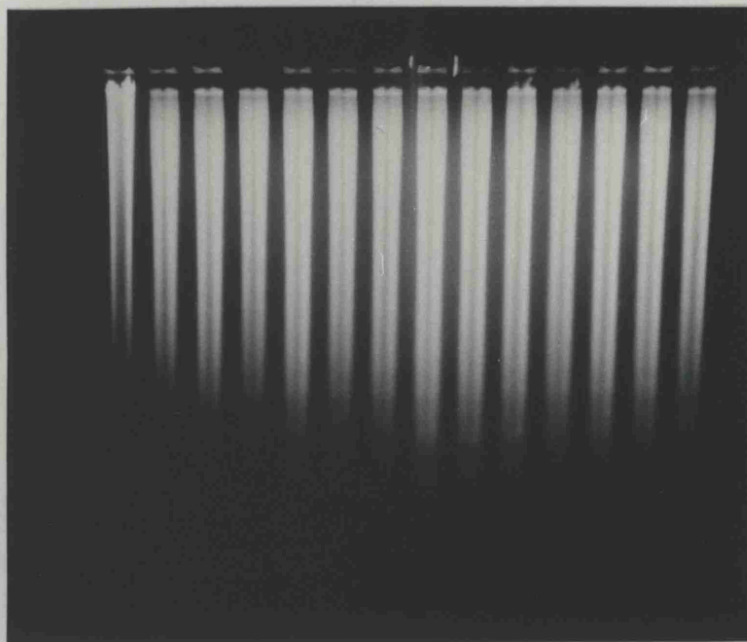
gene				Δ11			Δ12		Δ13			Δ14					
plant			b	1	2	4	1	2	1	2	3	1	2	3	4	7	8
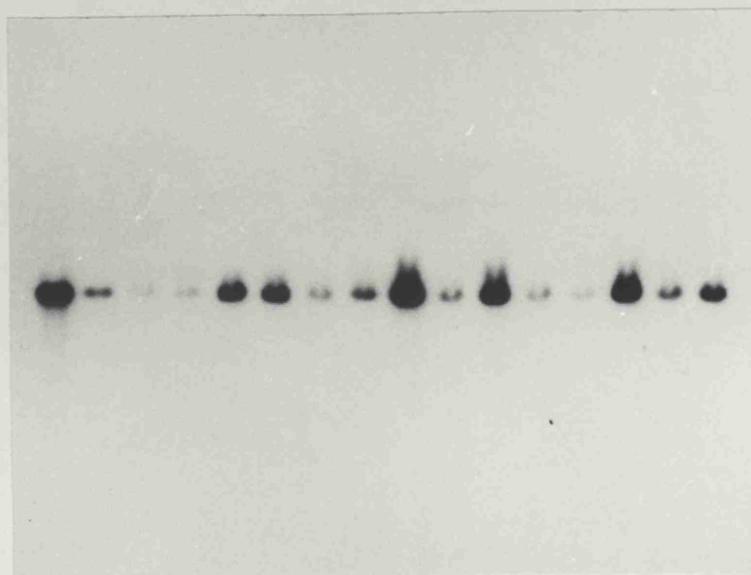
Figure 9.6 Northern blot analysis of tobacco RNA.

Triplicate samples, each containing 5µg of total RNA from N.tabacum

SR1 (tracks A) or N.tabacum SR1/19S-CAT plant 1 (tracks B), were

fractionated in a formaldehyde agarose gel. One pair of samples was

stained with ethidium bromide, the second was probed with a cDNA

clone from a cucumber cab gene and the third pair was probed with the

CAT gene. The figure shows a 3 day autoradiograpic exposure.

probe: Cab   CAT
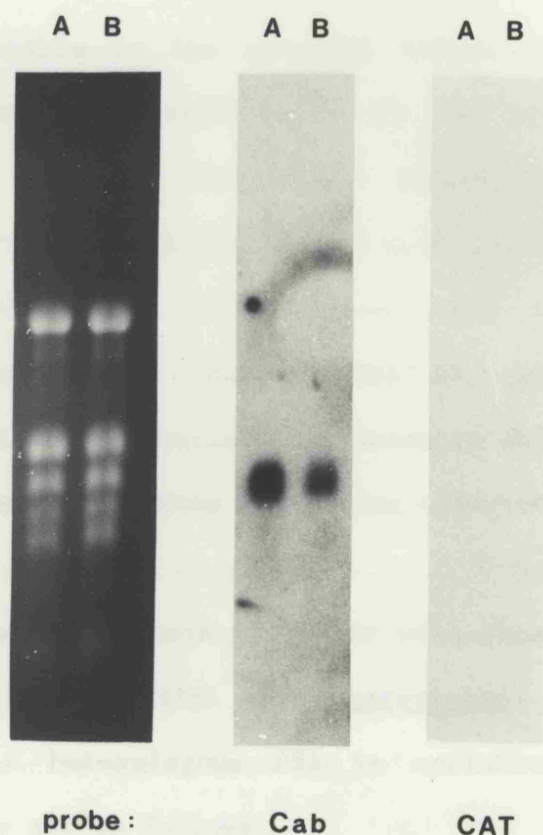
CHAPTER 10

GENERAL DISCUSSION

Yeast is a remarkable experimental organism, combining the convenience of a microbial system (ease of handling, rapid growth *etc*) with the ability to perform both classical and molecular genetic analyses. Yeast is especially amenable to the powerful techniques of genetic manipulation, ranging from gene replacement to the introduction of artificial chromosomes and has contributed significantly to our understanding of such areas as eukaryotic DNA replication, chromosome structure and gene regulation. The convenience and versatility of yeast as an experimental system has prompted us, and others, to investigate whether it might be a suitable organism in which to analyse the expression and regulation of genes from other eukaryotes.

The findings described in this thesis, together with those reported by others, suggest that the ability of S.cerevisiae to recognize transcription signals in heterologous DNA is variable. The main conclusions of this study are as follows.

1. A 500bp segment of DNA encompassing the 19S promoter of CaMV has some transcriptional activity in yeast. When fused to the E.coli lacZ coding region, the promoter gave rise to measurable amounts of β-galactosidase.

2. The putative TATA box of the 19S promoter may be required for expression of the 19S-lacZ gene in yeast as a 41bp deletion which specifically removed this element caused a 10-fold reduction in the level of β-galactosidase.

3. Primer extension analyis revealed the existence of two 19S-lacZ transcripts in yeast, both of which initiated downstream of the cap

125

site used in CaMV-infected plant cells. No transcript initiating at the correct position could be detected.

4. Primer extention studies also demonstrated the presence of a transcript originating 200-250bp upstream of the 19S-lacZ coding region, indicating the existence of a fortuitous promoter in this vicinity.

5. A fusion between the 35S promoter of CaMV and the lacZ gene was not expressed in yeast.

Some of these pitfalls to foreign gene expression in yeast (inaccurate or inefficient transcription, the existence of fortuitous promoters etc) have been noted by others. The native promoters of the rabbit β-globin, Drosophila ADE8, and french bean phaseolin genes all showed some transcriptional activity in S.cerevisiae but in each instance RNA initiation was at an incorrect site (Beggs et al 1980, Henikoff and Furlong 1983, Harris Cramer et al 1985). Inaccurate transcription initiation in S.cerevisiae was also observed for some fungal genes including an Aspergillus glucoamylase gene (Innis et al 1985) and two genes from Schizosaccharomyces pombe (Russell 1983, Russell 1985). In other instances no promoter activity could be detected, for example the thymidine kinase gene of herpes simplex virus (HSV) did not appear to be transcribed in yeast (Kiss et al 1982). On the whole, it would appear that S.cerevisiae is not a satisfactory host species in which to examine the normal expression pattern of genes from other eukaryotes, although a degree of transcription may be obtained and in exceptional cases foreign genes may be accurately transcribed. Two zein genes from maize, for example, were efficiently transcribed in yeast, starting from the usual cap site (Langridge et al 1984) and, incidentally, in animal cells.

Inaccurate and inefficient initiation of transcription appears to be the major obstacle to foreign gene expression in yeast, although other complications such as incorrect RNA splicing and premature termination of transcription have been reported (Beggs *et al* 1980, Watts *et al* 1983). While yeast may be an excellent model organism for the study of more general biochemical processes, there has clearly been divergence of the mRNA transcription initiation mechanism, such that accurate recognition of foreign promoters and cap sites is rarely achieved. The following paragraphs will examine the evidence for divergence of the transcription apparatus in eukaryotes.

The failure of eukaryotic genes to be efficiently transcribed in heterologous species may be due to the absence of gene-specific activator proteins as well as a more generalised inability of the transcription machinery to recognize foreign promoter and initiation signals. *Trans*-acting factors which interact with upstream promoter elements to regulate transcription are not necessarily present in all cell types of the native host species and an equivalent molecule may be completely lacking in yeast. Both the rabbit β-globin and HSV thymidine kinase genes contain upstream promoter elements which are crucial for their efficient expression in mammalian cells. The failure of these (and other) genes to be efficiently transcribed in yeast may well be partly attributable to the absence of appropriate stimulatory trancription factors.

Other reasons might include subtle differences in the structure or specificity of RNA polymerase II (and its associated factors) in different species. RNA polymerase II is a complex enzyme composed of two large non-identical subunits (Mr 220kd and 140kd) and between 6 and 9 smaller polypeptides (Mr less than 45kd). Little is known about the

function of the smaller subunits but collectively the proteins must have sites for interacting with the DNA template, the nascent RNA chain, substrate nucleotide triphosphates, neighbouring enzyme subunits and probably transcription factors. DNA protection experiments indicate that the largest subunit as well as a 23kd subunit appear to be involved in DNA binding (see Sentenac 1985). The finding has been supported by the generation of a monoclonal antibody directed against the largest subunit of calf RNA polymerase II which blocks the formation of enzyme-DNA complexes (Carroll and Stollar 1983). DNA sequence analysis of the yeast 140kd subunit gene revealed a putative purine-binding site (Sweetser *et al* 1987) and a chemical affinity probe has been used to show that this subunit does indeed interact with substrate NTP and contains the active site of the polymerase (Riva *et al* 1987). Affinity labelling studies also show that the nascent RNA is in close contact with the largest subunit and to a lesser extent with the 140kd subunit, leading to the proposal that these large subunits form a channel through which the growing RNA chain passes (Riva *et al* 1987, Bartholomew *et al* 1987).

Immunological studies using polyclonal or monoclonal antibodies raised against purified RNA polymerase II subunits have been used to compare the relatedness of mRNA polymerases from different species (reviewed by Sentenac 1985). Overall immunological cross reactivity of the enzymes decreases steadily with increasing evolutionary distance although certain antigenic determinants are highly conserved (Carroll and Stollar 1983). When antibodies raised against each of the yeast RNA polymerase II subunits were tested against RNA polymerase II from a range of eukaryotic species, it was found that enzymes from all species tested cross react with the two large subunits and a 23kd subunit, whereas limited or no cross reactivity was observed with other subunits

(Huet *et al* 1982)   The data indicates that the evolution of the two large and certain smaller enzyme subunits is subject to strong structural and functional constraints, while other subunits are more free to evolve.   Possibly the more conserved subunits are those involved in the catalytic (or DNA binding) process whereas the divergent components interact with cellular transcription factors.

Evidence of functional as well as structural divergence of eukaryotic mRNA polymerases has come from experiments using cell-free transcription systems.   In these it was shown that invertebrate RNA polymerase II was unable to complement reconstituted *in vitro* transcription systems derived from mammalian cells.   In one report, the adenovirus 2 major late transcription unit could be transcribed by human, mouse and Xenopus RNA polymerase II but the same enzyme from wheat germ was ineffective (Weil *et al* 1979).   Similarly, the chick conalbumin and ovalbumin genes were transcribed with equal efficiency by calf thymus or hen oviduct polymerase but RNA polymerase II isolated from Drosophila, yeast or wheat germ was unable to initiate transcription (Wasylyk *et al* 1980).   It is known that purified RNA polymerase II is not able to recognize and bind to promoters by itself and that protein factors, supplied in these *in vitro* transcription systems by a mammalian (Hela) cell extract, are essential for efficient initiation of RNA synthesis.   A likely explanation for the above findings is that the invertebrate enzymes are not compatible with mammalian transcription factors, possibly due to evolution of the polymerase subunits which interact with the transcription factors and/or divergence of the factors themselves.

Transcription factors are defined as protein components, distinct from polymerase subunits, which are required as essential components of the

transcription complex at various stages of RNA synthesis (initiation, elongation, termination). They may bind to promoter sequences on the DNA or alternatively interact directly with the RNA polymerase and can also be subdivided into gene-specific transcription factors (such as mammalian factor Sp1 or the yeast GAL4 protein) and general transcription factors required for the synthesis of all mRNA. The process of transcription initiation can be divided into the following three functional stages and it is probable that transcription factors are required at each. 1. one or more transcription factor binds to the promoter 2. RNA polymerase II becomes associated, forming a stable initiation complex 3. The RNA polymerase initiates RNA chain elongation by catalysing phosphodiester bond formation between the first two nucleotide triphosphates.

Biochemical fractionation of nuclear extracts has led to the identification and partial purification of several general transcription factors from <u>Drosophila</u> and human cells (Price *et al* 1987, Reinberg and Roeder 1987). The precise function of most of these factors is not clear although some are known to be required for initiation while others affect chain elongation *in vitro*. One of the better characterised mammalian transcription factors, TFIID, binds to DNA at a site overlapping the TATA box and cap site. This is the only known transcription factor to interact directly with promoter sequences and is almost certainly responsible for primary recognition of the promoter (Workman and Roeder 1987). The mechanism by which the polymerase then binds to the complex is not known.

In view of their complex functional interactions with the polymerase, it is almost inevitable that transcription factors have diversified alongside polymerases during eukaryotic evolution. This point was

nicely demonstrated by Burton *et al* (1986) in the course of isolating putative transcription factors by passing Hela cell nuclei through an affinity column containing immobilised calf thymus RNA polymerase II. The polymerase-binding proteins were shown to be *bona fide* transcription factors but none of the factors was required for formation of the promoter-associated initiation complex, suggesting that their primary interaction is with the polymerase rather than with promoter DNA. A significant finding, however, was that none of these polymerase- associating transcription factors was retained on an affinity column of yeast RNA polymerase II, emphasising that polymerase specificity of transcription factors does exist. The failure of yeast RNA polymerase II to interact with mammalian transcription factors may well account for the inactivity of this enzyme in Hela cell-derived *in vitro* transcription systems but does not, of course, explain the inefficient transcription of foreign genes in yeast, where the enzyme is equipped with its usual set of transcription factors.

Because eukaryotic RNA polymerases have no intrinsic affinity for promoter sequences, the transcription of a heterologous DNA sequence in yeast will depend to a large extent on its affinity for the general transcription factors involved in promoter recognition. One such factor may be the protein which interacts with a yeast TATA box and may, by analogy with the mammalian factor TFIID, mediate the attatchment of RNA polymerase (Selleck and Majors 1987). Such a role for the TATA box-binding protein would seem probable in view of the fact that the TATA box is the only promoter element common to virtually all protein-coding genes in yeast. One might therefore expect (of foreign DNA sequences) that possession of a TATA-like element would guarantee a basal level of transcription in yeast.

Of those foreign genes which are transcribed in yeast, none has previously been characterised with a view to defining DNA sequences required for RNA synthesis or identifying sites of protein-DNA interactions. The mutational analysis of the CaMV 19S promoter described in this thesis is therefore the first to address this question. Data presented in chapter 8 reveals expression of the downstream gene to be dependent on a 41bp sequence encompassing the viral TATA box, whereas deletion of other segments of the promoter had a negligible effect on expression. These findings not only confirm the importance of the TATA box but demonstrate that the yeast transcription machinery is able to recognise promoter elements in foreign DNA. An exception to this might be the thymidine kinase gene of HSV, which is preceded by a conventional TATA sequence but did not give rise to detectable transcripts in yeast (Kiss *et al* 1982). The tk transcript may however be very unstable or the experimental system not sensitive enough to detect extremely low levels of tk RNA.

Another important point to emerge from this work is that recognition of the heterologous TATA box does not guarantee transcription initiation at the usual position, probably due to divergence in the mechanism of cap site selection in eukaryotes. Unlike higher eukaryotes where transcription initiates a fixed distance of 25 to 30bp 3' of the TATA box, mRNA synthesis in yeast occurs 40bp or more downstream of the TATA box at a site largely dictated by the surrounding nucleotide sequence (see introduction). In accordance with this, apparent utilisation of the 19S promoter TATA box in yeast resulted in two transcripts initiating 45bp and 61bp downstream. The 19S transcript synthesized in CaMV-infected plant cells begins only 25bp downstream of the same TATA box (Guilley *et al* 1982). Transcriptional analysis of several other higher eukaryotic genes expressed in yeast again reveals RNA synthesis

to start 10 to 40bp downstream of the expected position, at a site consistent with recognition of the foreign TATA box (Beggs *et al* 1980, Henikoff and Furlong 1983, Harris-Cramer *et al* 1985, Watts *et al* 1983). Paradoxically, then, it appears that recognition of a foreign promoter (TATA element) is rarely compatible with transcription initiation at the normal site.

The above discussion underscores the necessity to perform plant gene expression studies in a less distantly related genetic background, ideally in a plant cell where the availability of regulatory and general transcription factors is not in question. Two complementary plant-based experimental systems are currently available for the analysis of cloned plant genes. These are transient expression systems using plant protoplasts and regenerated plants in which the foreign DNA is stably integrated. The following paragraphs will provide an overview of the relative merits and limitations of these systems.

Transgenic plants represent the most natural, unperturbed system for gene expression studies and offer scope for investigating more complex aspects of gene regulation such as tissue specificity and environmental control of transcription. The production of transgenic plants is, however, a painstaking and above all lengthy process, requiring several months of tissue culture and growth. The system is obviously not appropriate for rapid analyses of gene expression. Moreover, since the introduction of foreign DNA must take place either via A.tumefaciens infection of tissue sections or via DNA uptake by isolated protoplasts, the production of transgenic plants is restricted to those species which are fully regeneratable from the transformed callus or protoplast. For many plant species, including most monocotyledonous species, tissue culture conditions have not been defined which allow

this complex developmental process to occur. Genes isolated from non-regeneratable species such as wheat and maize must therefore be reintroduced into a heterologous but regeneratable species such as tobacco or petunia for *in vivo* analysis. Fortunately there appears to have been relatively little divergence of the transcription apparatus among higher plants, as illustrated by the fact that genes isolated from dicotyledonous species are not only accurately transcribed but display appropriate regulation in a transgenic dicotyledonous host such as tobacco or petunia (Willmitzer 1988). Results with monocotyledonous genes are more variable, probably reflecting the greater evolutionary distance between the native and transgenic species. Another problem associated with the analysis of (randomly) integrated DNA sequences is that their level of expression can vary enormously depending on the site of chromosomal integration, a phenomenon known as "position effect". Position effects cannot be avoided, so an accurate comparison of gene expression levels can only be obtained by analysing a large number of independently derived transgenic plants.

An alternative approach for measuring gene expression is the transient assay. In transient assay systems, foreign DNA is introduced into plant protoplasts (that is cells from which the cell wall has been enzymically removed) usually by treatment with polyethylene glycol or electroporation. The protoplasts are maintained in culture and expression of the input DNA is monitored after a fixed time period. Compared with transgenic plants, the method is very rapid, results being obtainable within a matter of days, and expression levels are not subject to position effect as the input DNA appears to remain extrachromosomal (Werr and Lorz 1986). Furthermore, transient assays can be performed using protoplasts prepared from monocot as well as dicot species. Most protoplast-based transient assays have so far been

134

monitored by recording the activity of reporter enzymes fused to the promoters under study. A recent report by Ellis *et al* (1987) showed, however, that RNA analysis (in this case S1 nuclease mapping) is also feasible. Transient expression systems provide a rapid means of measuring gene activity in a specified cell type and obviate the need for tissue culture and regeneration. It should be borne in mind, however, that these cells have had their cellulose cell wall stripped away, have lost their contact with neighbouring cells and are being maintained in an artificial growth medium; as such they represent a highly perturbed system.

A novel mechanism for delivering DNA into intact plant cells was recently reported by Klein *et al* (1987), who showed that DNA-coated tungsten particles (4μm spheres) "fired" at sections of onion epidermis would pierce the cell wall and penetrate cells without killing them. Moreover, expression of the incoming DNA could be demonstrated. As a transient assay system, the particle bombardment technique has some advantages over protoplast-based assays. For example, the target cells are in a more natural state and, unlike protoplasts, require no prior preparation.

The choice between these gene expression systems will clearly depend upon the question being posed and on the facilities and time available to the experimentor. Some studies will require the production of transgenic plants while others may be adequately answered by short-term expression systems. Data presented in this thesis, as well as elsewhere, would caution against the use of more distantly related eukaryotic species as host organisms for plant gene expression studies.

CHAPTER 11

LITERATURE CITED

Abraham J, Feldman J, Nasmyth K, Strathern JN, Klar AJS, Broach JR, Hicks JB (1983). Sites required for position effect regulation of mating-type information in yeast. Cold Spring Harbor Symp Quant Biol 47 989-998.

Andreadis A, Hsu Y, Kohlaw GB, Schimmel P (1982). Nucleotide sequence of yeast LEU2 shows the 5' non-coding region has sequences cognate to leucine. Cell 31 319-325.

Armour SL, Melcher U, Pirone TP, Lyttle DI, Essenberg RC (1983). A helper component for aphid transmission encoded by region II of cauliflower mosaic virus DNA. Virology 129 25-30.

Ashikari T, Nakamura N, Tanaka Y, Kiuchi N, Shibano Y, Tanaka T, Amachi T, Yoshuzumi H (1985). Cloning and expression of the Rhizopus glucoamylase gene in yeast. Agric Biol Chem 49 2521-2523.

Bachmair A, Finley D, Varshavsky A (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science 234 179-186.

Baim SB, Pietras DF, Eustice DC, Sherman F (1985). A mutation allowing an mRNA secondary structure diminishes translation of Saccharomyces cerevisiae iso-1-cytochrome C. Mol Cell Biol 5 1839-1846.

Banerji J, Olson L and Schaffner W (1983). A lymphocyte- specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33 729-740.

Baniahmad A, Muller M, Steiner C, Renkawitz R (1987). Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. EMBO J 6 2297-2303

Bartholomew B, Dahmus ME, Meares CF (1986). RNA contacts subunits II_O and II_C in Hela RNA polymerase II transcription complexes. J Biol Chem 261 14226-14231.

Beggs JD, van den Berg J, van Ooyen A, Weissman C (1980). Abnormal expression of chromosomal rabbit β-globin gene in <u>Saccharomyces cerevisiae</u>. Nature <u>283</u> 835-840.

Beggs JD (1978). Transformation of yeast by a replicating hybrid plasmid. Nature <u>275</u> 104-109.

Behringer RR, Hammer RE, Brinster RL, Palmiter RD, Townes TM (1987). Two 3' sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. Proc Natl Acad Sci USA <u>84</u> 7056-7060.

Bennetzen JL and Hall BD (1982). Codon selection in yeast. J Biol Chem <u>257</u> 3026-3031.

Berget SM and Robberson BL (1986). U1, U2 and U4/U6 small nuclear ribonucleoproteins are required for *in vitro* splicing but not polyadenylation. Cell <u>46</u> 691-696.

van den Bergh CJ, Bekkers AC, de Geus P, Verheij HM, de Haas GH (1987). Secretion of biologically active porcine phospholipase A_2 by <u>Saccharomyces cerevisiae</u>. Eur J Biochem <u>170</u> 241-246.

Bergman LW, Stranathan MC, Preis LH (1986). Structure of the transcriptionally repressed phosphate-repressible acid phosphatase gene (PHO5) of <u>Saccharomyces cerevisiae</u>. Mol Cell Biol <u>6</u> 38-46.

Bergsma DJ, Grichnik JM, Gossett LMA, Schwartz RJ (1986). Delimitation and characterisation of *cis*-acting DNA sequences required for the regulated expression and transcriptional control of the chicken skeletal α-actin gene. Mol Cell Biol <u>6</u> 2462-2475.

Berse B, Dmochowska A, Skrzypek M, Weglenski P, Bates MA, Weiss RL (1983). Cloning and characterisation of the ornithine carbamoyl transferase (E.C.2.1.3.3.) gene from <u>Aspergillus nidulans</u>. Gene <u>25</u> 109-118.

Bevan M (1984). Binary <u>Agrobacterium</u> vectors for plant transformation. Nucl Acids Res <u>12</u> 8711-8721.

Bevan MW, Mason SE and Goelet P (1985). Expression of tobacco mosaic virus coat protein by a cauliflower mosaic virus promoter in plants transformed by Agrobacterium. EMBO J 4 1921-1926.

Bienz M and Pelham HRB (1986). Heat shock regulatory elements function as an inducible enhancer in the Xenopus hsp70 gene and when linked to a heterologous promoter. Cell 45 753-760.

Bird AP (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3 342-347.

Birnstiel ML, Busslinger M, Strub K (1985). Transcription termination and 3' processing: the end is in site! Cell 41 349-359.

Bodmer MW, Angal S, Yarranton GT, Harris TRS, Lyons A, King DJ, Pieroni G, Riviere C, Verger R, Lowe PA (1987). Molecular cloning of a human gastric lipase and expression of the enzyme in yeast. Biochim Biophys Acta 909 237-244.

Bodnar JW, Jones CJ, Coombs DH, Pearson GD, Ward DC (1983). Proteins tightly bound to Hela cell DNA at nuclear matrix attachment sites. Mol Cell Biol 3 1567-1579.

Bohmann D, Keller W, Dale T, Schöler HR, Tebb G, Mattaj IW (1987). A transcription factor which binds to the enhancers of SV40, immunoglobulin heavy chain and U2 snRNA genes. Nature 325 268-272.

Boy-Marcotte E, Vilaine F, Camonis J, Jacquet M (1984). A DNA sequence from Dictyostelium discoideum complements ura3 and ura5 mutations of Saccharomyces cerevisiae. Mol Gen Genet 193 406-413.

Boyer HW and Roulland-Dussoix D (1969). A complementation analysis of the restriction and modification of DNA in Escherichia coli. J Mol Biol 41 459-472.

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

Bram RJ, Lue NF and Kornberg RD (1986). A GAL family of upstream

activating sequences in yeast: roles in both induction and repression of transcription. EMBO J <u>5</u> 603-608.

Brand AH, Breeden L, Abraham J, Sternglanz R, Nasmyth K (1985). Characterisation of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. Cell <u>41</u> 41-48.

Brawerman G (1987). Determinants of messenger RNA stability. Cell <u>48</u> 5-6.

Breathneach R and Chambon PA (1981). Organisation and expression of eukaryotic split genes coding for proteins. Ann Rev Biochem <u>50</u> 349-383.

Brenner DG and Shaw WV (1985). The use of synthetic oligonucleotides with universal templates for rapid DNA sequencing: results with staphylococcal replicon pC221. EMBO J <u>4</u> 561-568.

Brent R and Ptashne M (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell <u>43</u> 729-736.

Brickman E, Silhavy TJ, Bassford PJ, Shuman HA, Beckwith JR (1979). Sites within gene <u>lacZ</u> of <u>Escherichia</u> <u>coli</u> for formation of active hybrid β-galactosidase molecules. J Bacteriol <u>139</u> 13-18.

Broach JR (1983). Construction of high copy yeast vectors using 2μm circle sequences. Methods Enzymol <u>101</u> 307-325.

Broach JR, Atkins JF, McGill C, Chow L (1979). Identification and mapping of the transcriptional and translational products of the yeast 2μm circle. Cell <u>16</u> 827-839.

Broach JR, Guarascio VR, Jayaram M (1982). Recombination within the yeast 2μm circle is site-specific. Cell <u>29</u> 227-234.

Broach JR and Hicks JB (1980). Replication and recombination functions associated with the yeast 2μm circle. Cell <u>21</u> 501-508.

Broach JR, Li YY, Feldman J, Jayaram M, Abraham J, Nasmyth KA, Hicks JB (1983). Localisation and sequence analysis of yeast origins of DNA replication. Cold Spring Harbor Symp Quant Biol 47 1165-1173.

Bröker M, Ragg H, Karges HE (1987). Expression of human antithrombin III in Saccharomyces cerevisiae and Schizosaccaromyces pombe. Biochim Biophys Acta 908 203-213.

Brown JWS (1986). A catalogue of splice junction and putative branch point sequences from plant introns. Nucl Acids Res 14 9549-9559.

Burton ZF, Ortolan LG, Greenblatt J (1986). Proteins that bind RNA polymerase II are required for accurate initiation of transcription at the adenovirus 2 major late promoter. EMBO J 5 2923-2930.

Caplan A, Herrera-Estrella L, Inze D, van Haute E, van Montagu M, Schell J, Zambryski P (1983). Introduction of genetic material into plant cells. Science 222 815-821.

Cappello J, Zuker C and Lodish HF (1984). Repetetive Dictyostelium heat shock promoter functions in Saccharomyces cerevisiae. Mol Cell Biol 4 591-598.

Caron F and Thomas JO (1981). Exchange of histone H1 between segments of chromatin. J Mol Biol 146 513-537.

Carroll SB and Stollar BD (1983). Conservation of a DNA binding site in the largest subunit of eukaryotic RNA polymerase II. J Mol Biol 170 777-790.

Cashmore AM, Albury MS, Hadfield C and Meacock PA (1985). Genetic analysis of partitioning functions encoded by the 2μm circle of Saccharomyces cerevisiae. Mol Gen Genet 203 154-162.

Chen W and Struhl K (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J 4 3273-3280.

Chinnapan-Santiago T, Purvis IJ, Bettany JE, Brown AJP (1986). The relationship between mRNA stability and length in Saccharomyces

cerevisiae. Nucl Acids Res 14 8347-8360.

Cigan AM and Donahue TF (1987). Sequence and structural features associated with translational initiator regions in yeast - a review. Gene 59 1-18.

Clarke L and Carbon J (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature 287 504-509.

Close TJ and Rodriguez RL (1982). Construction and characterisation of the chloramphenicol-resistance gene cartridge: A new approach to the transcriptional mapping of extrachromosomal elements. Gene 20 305-316.

Coen ES, Carpenter R and Martin C (1986). Transposable elements generate novel spatial patterns of gene expression in Antirrhinum majus. Cell 47 285-296.

Cohen RB, Sheffery M and Kim CG (1986). Partial purification of a nuclear protein that binds to the CCAAT box of the mouse α_1-globin gene. Mol Cell Biol 6 821-832.

Covey S and Hull R (1981). Transcription of cauliflower mosaic virus DNA. Detection of transcripts, properties and location of the gene encoding the virus inclusion body protein. Virology 111 463-474.

Covey SN, Lomonossoff P and Hull R (1981). Characterisation of cauliflower mosaic virus DNA sequences which encode major polyadenylated transcripts. Nucl Acids Res 9 6735-6747.

Crabeel M, Huygen R, Verschueren K, Messenguy F, Tinel K, Cunin R and Glansdorff N (1985). General amino acid control and specific arginine repression in Saccharomyces cerevisiae: Physical study of the bifunctional regulatory region of the ARG3 gene. Mol Cell Biol 5 3139-3148.

Crossway A, Oakes JV, Irvine JM, Ward B, Knauf VC, Shewmaker LK (1986). Integration of foreign DNA following microinjection of tobacco mesophyll protoplasts. Mol Gen Genet 202 179-185.

Dalgleish R (1987). In Gene Cloning and Analysis: a Laboratory Guide. Ed GJ Boulnois. Blackwell Scientific.

Dani GM and Zakian VA (1983). Mitotic and meiotic stability of linear plasmids in yeast. Proc Natl Acad Sci USA 80 3406-3410.

Daubert S, Richins R, Shepherd RJ and Gardner RC (1982). Mapping of the coat protein of cauliflower mosaic virus by its expression in a prokaryotic system. Virology 122 444-449.

Daubert S, Shepherd RJ and Gardner RC (1983). Insertional mutagenesis in the cauliflower mosaic virus genome. Gene 25 201-208.

Davie JR, Saunders CA, Walsh JM and Weber SC (1981). Histone modifications in the yeast Saccharomyces cerevisiae. Nucl Acids Res 9 3205-3216.

De Block M, Botterman J, Vandewiele M, Dockx J, Thoen C, Gosselé V, Rao Movva N, Thompson C, Van Montagu M, Leemans J (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J 6 2513-2518.

Deshayes A, Herrera-Estrella L, Caboche M (1983). Liposome- mediated transformation of tobacco mesophyll protoplasts by an Escherichia coli plasmid. EMBO J 4 2731-2737.

Dickson RC (1980). Expression of a foreign eukaryotic gene in Saccharomyces cerevisiae: β-galactosidase from Kluyvromyces lactis. Gene 10 347-356.

Dierks PA, van Ooyen A, Cochran MD, Dobkin C, Reiser J and Weissmann C (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32 695-706.

Ding M, Wen D, Schlesinger MJ, Wertz GW, Ball LA (1987). Expression and glycosylation of the respiratory syncytial virus G protein in Saccharomyces cerevisiae. Virology 159 450-453.

Ditta G, Stanfield S, Corbin D, Helsinki D (1980). Broad host range

DNA cloning system for gram negative bacteria: construction of a gene bank of Rhizobium meliloti. Proc Natl Acad Sci USA 77 7347-7351.

Dixon LK and Hohn T (1984). Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. EMBO J 3 2731-2736.

Dobson MJ, Tuite MF, Roberts NA, Kingsman AJ, Kingsman SM (1982). Conservation of high efficiency promoter sequences in Saccharomyces cerevisiae. Nucl Acids Res 10 2625-2637.

Doerfler W (1983). DNA methylation in gene expression. Ann Rev Biochem 52 93-124.

Dynan WS (1986). Promoters for housekeeping genes. Trends Genet 2 196-197.

Dynan WS and Tijan R (1985). Control of eukaryotic messenger RNA systhesis by sequence-specific DNA-binding proteins. Nature 316 774-778.

Edens L, Bom I, Ledeboer HM, Maat J, Toonen Y, Visser C, Verrips CT (1984). Synthesis and processing of the plant protein thaumatin in yeast. Cell 37 629-633.

Ellis JG, Llewellyn DJ, Walker JC, Dennis ES and Peacock WJ (1987). The ocs element: a 16 base pair palindrome essential for activity of the octopine synthase enhancer. EMBO J 6 3203-3208.

Erhart E and Hollenberg CP (1983). The presence of a defective LEU2 gene on 2μm DNA recombinant plasmids of Saccharomyces cerevisiae is responsible for curing and high copy number. J Bacteriol 156 625-635.

Favali MA, Bassi M and Conti GG (1973). A quantitative autoradiographic study of intracellular sites for replication of cauliflower mosaic virus. Virology 53 115-119.

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

Filho SA, Galembeck EV, Faria JB, Frascino ACS (1986). Stable yeast transformants that secrete functional α-amylase encoded by cloned mouse pancreatic cDNA. Bio/Technology 4 311-315.

Fluhr R, Kuhlemeier C, Nagy F and Chua NH (1986). Organ- specific and light-induced expression of plant genes. Science 232 1106-1112.

Fowler AV and Zabin I (1983). Purification, structure and properties of hybrid β-galactosidase proteins. J Biol Chem 258 14354-14358.

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

Fromm M, Taylor LP and Walbot V (1985). Expression of genes transferred into monocot and dicot plant cells by electro- poration. Proc Natl Acad Sci USA 82 5824-5828.

Gardell SJ, Craik CS, Hilvert D, Urdea S, Rutter WJ (1985). Site directed mutagenesis shows that tyrosine 248 of carboxy- peptidase A does not play a role in catalysis. Nature 317 551-553.

Gelvin SG (1986). Transcription "enhancer-like" sequences in plant genes. Plant Mol Biol 7 229-232.

Gerbaud C and Guerineau M (1980). 2μm plasmid copy number in different yeast strains and repartition of endogenous and 2μm chimeric plasmids in transformed strains. Curr Genet 1 219-228.

Giniger E, Varnum SM and Ptashne M (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell 40 767-774.

Goff CG, Moir DT, Kohno T, Gravius TC, Smith RA, Yamasaki E and Taunton-Rigby A (1984). Expression of calf prochymosin in Saccharomyces cerevisiae. Gene 27 35-46.

Goodbourne S, Burnstein H and Maniatis T (1986). The human β-interferon gene enhancer is under negative control. Cell 45 601-610.

Gough JA and Murray NE (1983). Sequence diversity among related genes

for recognition of specific targets in DNA molecules. J Mol Biol 166 1-19.

Govers F, Gloudemans T, Moerman M, van Kammen A, Bisseling T (1985). Expression of plant genes during development of pea root nodules. EMBO J 4 861-867.

Green MR (1986). Pre-mRNA splicing. Ann Rev Genet 20 671-708.

Green PA, Kay SA, Chua NH (1987). Sequence-specific interactions of a pea nuclear factor with light responsive elements upstream of the rbs-3A gene. EMBO J 6 2543-2549.

Grierson D and Covey S (1984). Plant Molecular Biology. Blackie.

Gronenborn B, Gardner RC, Schaefer S, Shepherd RJ (1981). Propagation of foreign DNA in plants using cauliflower mosaic virus as a vector. Nature 294 773-776.

Grosschedl R and Birnstiel ML (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc Natl Acad Sci USA 77 1432-1436.

Guarente L (1984). Yeast promoters: positive and negative elements. Cell 36 799-800.

Guarente L and Ptashne M (1981). Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. Proc Natl Acad Sci USA 78 2199-2203.

Guarente L, Yocum RR and Gifford P (1982). A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. Proc Natl Acad Sci USA 79 7410-7414.

Guilfoyle TJ (1980). Transcription of the cauliflower mosaic virus genome in isolated nuclei from turnip leaves. Virology 107 71-80.

Guilley H, Dudley RK, Jonard G, Balàzs E, Richards K (1982). Transcription of cauliflower mosaic virus DNA: Detection of promoter

sequences and characterisation of transcripts. Cell <u>30</u> 763-773.

Guilley H, Richards K and Jonard G (1983). Observations concerning the discontinuous DNAs of cauliflower mosaic virus. EMBO J <u>2</u> 277-282.

Hadfield C, Cashmore AM, Meacock PA (1986). An efficient chloramphenicol-resistance marker for <u>Saccharomyces</u> <u>cerevisiae</u> and <u>Escherichia</u> <u>coli</u>. Gene <u>45</u> 149-158.

Hahn S, Hoar ET and Guarente L (1985). Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC1 promoter of <u>Saccharomyces</u> <u>cerevisiae</u>. Proc Natl Acad Sci USA <u>82</u> 8562-8566.

Hahn P and Shepherd RJ (1982). Evidence for a 58 kilodalton polypeptide as precusor of the coat protein of cauliflower mosaic virus. Virology <u>116</u> 480-488.

Hallewell RA, Mills R, Tekamp-Olson P, Blacher R, Rosenberg S, Ötting F, Masiarz FR, Scandella CJ (1987). Amino terminal acetylation of authentic human Cu, Zn superoxide dismutase produced in yeast. Bio/Technology <u>5</u> 363-367.

Hames BD and Rickwood D (1981). Gel electrophoresis of proteins, a practical approach. IRL press.

Handa H, Toda T, Tajima M, Wada T, Iida H, Fukasawa T (1987). Expression of the human adenovirus Ela product in yeast. Gene <u>58</u> 127-136.

Harding K (1984). Studies on plant genetic transformation using tobacco protoplasts. PhD thesis, University of Nottingham, 37-39.

Harris Cramer J, Lea K, Slightom J (1985). Expression of phaseolin cDNA genes in yeast under the control of natural plant DNA sequences. Proc Natl Acad Sci USA <u>82</u> 334-338.

Hartley JL and Donelson JE (1980). Nucleotide sequence of the yeast plasmid. Nature <u>286</u> 860-865.

Hayakawa T, Toibana A, Marumoto R, Nakahama K, Kikuchi M, Fujimoto K, Ikehara M (1987). Expression of human lysozyme in an insoluble form in yeast. Gene <u>56</u> 53-59.

Healy AM, Helser TL and Zitomer RS (1987). Sequences required for transcription initiation of the <u>Saccharomyces cerevisiae</u> CYC7 gene. Mol Cell Biol <u>7</u> 3785-3791.

Heidecker G and Messing J (1986). Structural analysis of plant genes. Ann Rev Plant Physiol <u>37</u> 439-466.

Helmer G, Casadaban M, Bevan M, Kayes L, Dell Chilton M (1984). A new chimeric gene as a marker for plant transformation: the expression of <u>Escherichia coli</u> β-galactosidase in sunflower and tobacco cells. Bio/Technology <u>2</u> 520-527.

Henikoff S and Cohen EH (1984). Sequences responsible for transcription termination on a gene segment in <u>Saccharomyces cerevisiae</u>. Mol Cell Biol <u>4</u> 1515-1520.

Henikoff S and Furlong C (1983). Sequence of a <u>Drosophila</u> DNA segment that functions in <u>Saccharomyces cerevisiae</u> and its regulation by a yeast promoter. Nucl Acids Res <u>11</u> 789-800.

Henikoff S, Tatchell K, Hall BD and Nasmyth KA (1981). Isolation of a gene from <u>Drosophila</u> by complementation in yeast. Nature <u>289</u> 33-37.

Hentze MW, Rouault TA, Wright Caughman S, Dancis A, Harford JB, Klausner RD (1987). A <u>cis</u>-acting element is necessary and sufficient for translational regulation of human ferritin in response to iron. Proc Natl Acad Sci USA <u>84</u> 6730-6734.

Hereford L and Rosbash M (1977). Number and distribution of polyadenylated RNA sequences in yeast. Cell <u>10</u> 453-462.

Herrera-Estrella L, Van den Broek G, Maenhaut R, Van Montagu M, Schell J, Timko M, Cashmore A (1984). Light inducible and chloroplast associated expression of a chimaeric gene introduced into <u>Nicotiana tabacum</u> using a Ti plasmid vector. Nature <u>310</u> 115-120.

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

Himmelfarb HJ, Simpson EM, Friesen JD (1987). Isolation and characterisation of temperature-sensitive RNA polymerase II mutants of Saccharomyces cerevisiae. Mol Cell Biol 7 2155-2164.

Hinnen A, Hicks JB and Fink GR (1978). Transformation of yeast. Proc Natl Acad Sci USA 75 1929-1933.

Hitzeman RA, Hagie FE, Levine HL, Goeddel DV, Ammerer G, Hall BD (1981). Expression of a human gene for interferon in yeast. Nature 293 717-722.

Hitzeman RA, Leung DW, Perry LJ, Kohr WJ, Levine HL, Goeddel DV (1983). Secretion of human interferons by yeast. Science 219 620-625.

Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA (1983). A binary plant vector strategy based on separation of vir and T-regions of the Agrobacterium tumefaciens Ti plasmid. Nature 303 179-180.

Hohn T, Hohn B and Pfeiffer P (1985). Reverse transcription in CaMV. Trends Biochem Sci 10 205-209.

Hohn T, Richards K, Lebeurier G (1982). Cauliflower mosaic virus on its way to becoming a useful plant vector. Curr Top Microbiol Immunol 96 193-236.

Hope IA and Struhl K (1985). GCN4 protein synthesised *in vitro* binds to HIS3 regulatory sequences: implications for the general control of amino acid biosynthetic genes in yeast. Cell 43 177-188.

Hope IA and Struhl K (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46 885-894.

Howarth AJ, Gardner RC, Messing J, Shepherd RJ (1981). Nucleotide sequence of naturally occuring deletion mutants of cauliflower mosaic virus. Virology 112 678-685.

Howell SH (1981). Ultraviolet mapping of RNA transcripts encoded by the cauliflower mosaic virus genome. Virology 112 488-495.

Howell SH, Walker LL and Dudley RK (1980). Cloned cauliflower mosaic virus DNA infects turnips (Brassica rapa). Science 208 1265-1267.

Howell SH, Walker LL and Walden RM (1981). Rescue of *in vitro* generated mutants of cloned cauliflower mosaic virus genome in infected plants. Nature 293 483-486.

Hsiao CL and Carbon J (1981). Characterisation of a yeast replication origin (ars2) and construction of stable mini- chromosomes containing cloned yeast centromere DNA (CEN3). Gene 15 157-166.

Huet J, Sentenac A, Fromageot P (1982). Spot immunodetection of conserved determinants in eukaryotic RNA polymerases. J Biol Chem 257 2613-2618.

Hull R (1980). Structure of the cauliflower mosaic virus genome III. Restriction endonuclease mapping of thirty three isolates. Virology 100 76-90.

Hull R and Covey SN (1983). Does CaMV replicate by reverse transcription? Trends Biochem Sci 8 119-121.

Hultmark D, Klemenz R and Gehring WJ (1986). Translational and transcriptional control elements in the untranslated leader of the heat shock gene hsp22. Cell 44 429-438.

Igo-Kemenes T, Horz W, Zachau HG (1982). Chromatin. Ann Rev Biochem 51 89-121.

Innis MA, Holland MJ, McCabe PC, Cole GE, Wittman VP, Tal R, Watt KWK, Gelfand DH, Holland JP, Meade JH (1985). Expression, glycosylation and secretion of an Aspergillus glucoamylase by Saccharomyces cerevisiae. Science 228 21-26.

Ito H, Fukuda Y, Murata K, Kimura A (1983). Transformation of intact yeast cells treated with alkali cations. J Bacteriol 153 163-168.

Ivarie RD, Schacter BS, O'Farrell PH (1983). The level of expression of the rat growth hormone gene in liver tumour cells is at least eight orders of magnitude less than that in anterior pituitary cells. Mol Cell Biol <u>3</u> 1460-1467.

Izumoto Y, Sato T, Yamamoto T, Yoshida N, Kikuchi N, Ogawa M, Matsubara K (1987). Expression of human pancreatic secretory trypsin inhibitor in <u>Saccharomyces</u> <u>cerevisiae</u>. Gene <u>59</u> 151-159.

Jabbar MA, Sivasubramanian N, Nayak DP (1985). Influenza viral (A/WSN/33) hemagglutinin is expressed and glycosidated in the yeast <u>Saccharomyces</u> <u>cerevisiae</u>. Proc Natl Acad Sci USA <u>82</u> 2019-2023.

Jacquet M, Boy-Marcotte E, Rossier C, Kessin R (1982). A fragment of <u>Dictyostelium</u> <u>discoideum</u> genomic DNA which complements the ural mutation of <u>Saccharomyces</u> <u>cerevisiae</u>. J Mol Appl Genet <u>1</u> 513-525.

Jayaram M, Li YY, Broach JR (1983). The yeast plasmid 2μm circle encodes components required for its high copy propagation. Cell <u>34</u> 95-104.

Jensen EO, Marcker KA, Villadsen IS (1986). Heme regulates the expression in <u>Saccharomyces</u> <u>cerevisiae</u> of chimaeric genes containing 5' flanking soybean leghemoglobin sequences. EMBO J <u>5</u> 843-847.

Jobling SA and Gehrke (1987). Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. Nature <u>325</u> 622-625.

Jofuku KD, Okamuro JK and Goldberg RB (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature <u>328</u> 374-377.

Johnson AD and Herskowitz I (1985). A repressor (MATα2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell <u>42</u> 237-247.

Johnston M and Davies RW (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in <u>Saccharomyces</u> <u>cerevisiae</u>. Mol Cell Biol <u>4</u> 1440-1448.

Jones KA, Kadonaga JT, Rosenfeld PJ, Kelly TJ, Tijan R (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48 79-89.

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

Joshi CP (1987a). Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nucl Acids Res 15 9627-9640.

Kadonaga JT, Jones KA, Tijan R (1986). Promoter-specific activation of RNA polymerase II transcription by Spl. Trends Biochem Sci 11 20-23.

Kao KN and Michayluk MR (1975). Nutritional requirements for growth of Vicia hajastana cells and protoplasts at a very low population density in liquid media. Planta 126 105-110.

Keränen S (1986). Synthesis and processing of semliki forest virus polyprotein in Saccharomyces cerevisiae: a yeast-type glycosylation of El envelope protein. Gene 48 267-275.

Kimura T, Mills FC, Allan J, Gould H (1983). Selective unfolding of erythroid chromatin in the region of the active β-globin gene. Nature 306 709-712.

Kingsman AJ, Gimlich RL, Clarke L, Chinault AC and Carbon J (1981). Sequence variation in dispersed repetetive sequences in Saccharomyces cerevisiae. J Mol Biol 145 619-632.

Kingsman SM, Kingsman AJ, Dobson MJ, Mellor J, Roberts NA (1985). Heterologous gene expression in Saccharomyces cerevisiae. Biotechnology and Genetic Engineering Reviews, volume 3, 377-416. Ed Russell GE, Intercept.

Kiss GB, Pearlman RE, Cornish KV, Friesen JD, Chan VL (1982). The herpes simplex virus thymidine kinase gene is not transcribed in Saccharomyces cerevisiae. J Bacteriol 149 542-547.

Klein TM, Wolf ED, Wu R, Sanford JC (1987). High velocity microprojectiles for delivering nucleic acids into living cells. Nature 327 70-73.

Kohwi-Shigematsu T, Gelinas R, Weintraub H (1983). Detection of an altered chromatin DNA conformation at specific sites in chromatin and supercoiled DNA. Proc Natl Acad Sci USA 80 4389-4393.

Kolata G (1983). Z-DNA moves towards "real biology". Science 222 495-496.

Kozak M (1979). Inability of circular mRNAs to attach to eukaryotic ribosomes. Nature 280 82-85.

Kozak M (1981). Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucl Acids Res 9 5233-5252.

Kozak M (1983). Comparison of initiation of protein synthesis in prokaryotes, eukaryotes and organelles. Microbiol Rev 47 1-45.

Kozak M (1984a). Selection of initiation sites by eukaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. Nucl Acids Res 12 3873-3893.

Kozak M (1984b). Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. Nucl Acids Res 12 857-872.

Kozak M (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44 283-292.

Kramer KA, Sehaber MD, Skalka AM, Ganguly K, Wong-Staal F, Reddy EP (1986). HTLV-III gag protein is processed in yeast by the virus pol protease. Science 231 1580-1584.

Kunkel GR and Martinson HG (1981). Nucleosomes will not form on double stranded RNA or over poly(dA).poly(dT) tracts in recombinant DNA. Nucl

Acids Res 9 6869-6888.

Lamppa G, Nagy F and Chua NH (1985). Light-regulated and organ-specific expression of a wheat cab gene in transgenic tobacco. Nature 316 750-752.

Langford CJ, Klinz FJ, Donath C, Gallwitz D (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. Cell 36 645-653.

Langford C, Nellen W, Niessing J, Gallwitz D (1983). Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. Proc Natl Acad Sci USA 80 1496-1500.

Langridge P, Eibel H, Brown JWS, Feix G (1984). Transcription from maize storage protein gene promoters in yeast. EMBO J 3 2467-2471.

Larsen A and Weintraub H (1982). An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. Cell 29 609-622.

Lemontt JF, Wei CM, Dackowski WR (1985). Expression of active human uterine tissue plasminogen activator in yeast. DNA 4 419-428.

Levinger L and Varshavsky A (1982). Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. Cell 28 375-385.

Lewis CD and Laemmli UK (1982). Higher order metaphase chromosome structure: evidence for metalloprotein interactions. Cell 29 171-181.

Lichtenstein C and Draper J (1985). Genetic engineering of plants. DNA cloning volume II: a practical approach, 67-119. Ed DM glover, IRL press.

Linzer D (1986). Negative transcriptional regulation of c-myc. Trends Genet 2 195.

Livingston DM and Kupfer DM (1977). Control of Saccharomyces cerevisiae 2μm DNA replication by cell division cycle genes that

control nuclear DNA replication. J Mol Biol 116 249-260.

Logan J and Shenk T (1984). Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. Proc Natl Acad Sci USA 81 6355-6359.

Lohr D and Hereford L (1979). Yeast chromatin is uniformly digested by DNaseI. Proc Natl Acad Sci USA 76 4285-4288.

Lohr D and Hopper JE (1985). The relationship of regulatory proteins and DNAaseI hypersensitive sites in the yeast GAL1-10 genes. Nucl Acids Res 13 8409-8423.

Lustig AJ, Lin RJ and Abelson J (1986). The yeast RNA gene products are essential for mRNA splicing *in vitro*. Cell 47 953-963.

Maliga P, Breznovitz A, Marton L (1973). Streptomycin- resistant plants from callus culture of haploid tobacco. Nature New Biol 244 29-30.

Maniatis T, Fritsch EF and Sambrook J (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, New York.

Maniatis T and Reed R (1987). The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature 325 673-678.

Marczynski GT and Jaehning JA (1985). A transcriptional map of a yeast centromere plasmid: unexpected transcripts and altered gene expression. Nucl Acids Res 13 8487-8506.

Marinus MG (1973). Location of DNA methylation genes on the Escherichia coli K-12 genetic map. Mol Gen Genet 127 47-55.

Marsh LE and Guilfoyle TJ (1987). Cauliflower mosaic virus replication intermediates are encapsidated into virion-like particles. Virology 161 129-137.

Martelli GP and Castellano MA (1971). Light and electron microscopy of the intracellular inclusions of cauliflower mosaic virus. J Gen Virol 13 133-140.

Martinez-Izquierdo JA, Fütterer J, Hohn T (1987). Protein encoded by orfI of cauliflower mosaic virus is part of the viral inclusion body. Virology 160 527-530.

Matthes HWD, Zenke WM, Grundstrom T, Staub A, Wintzerith M, Chambon P (1984). Simultaneous rapid synthesis of over one hundred oligonucleotides on a microscale. EMBO J 3 801-805.

Maule AJ (1983). Infection of protoplasts from several Brassica species with cauliflower mosaic virus following inoculation using polyethylene glycol. J Gen Virol 64 2655-2660.

McKnight S and Tijan R (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell 46 795-805.

McNeil JB and Little JG (1985). Yeast/herpes simplex virus thymidine kinase gene fusions yield fusion proteins with thymidine kinase activity. Curr Genet 9 567-572.

McNeil JB and Smith M (1986). Transcription initiation of the Saccharomyces cerevisiae iso-1-cytochrome c gene: multiple independent T-A-T-A sequences. J Mol Biol 187 363.

Meaden P, Ogden K, Bussey H, Tubb RS (1985). A DEX gene conferring production of extracellular amyloglucosidase on yeast. Gene 34 325-334.

Mellon P, Parker V, Gluzman Y, Maniatis T (1981). Identification of DNA sequences required for transcription of the human α_1-globin gene in a new SV-40 host-vector system. Cell 27 279-288.

Mellor J, Dobson M, Roberts NA, Tuite MF, Emtage JS, White S, Lowe PA, Patel T, Kingsman AJ, Kingsman SM (1983). Efficient synthesis of enzymically active calf chymosin in Saccharomyces cerevisiae. Gene 24 1-14.

Menissier J, Lebeurier G and Hirth L (1982). Free cauliflower mosaic virus supercoiled DNA in infected plants. Virology 117 322-328.

Menissier J, Murcia G, Lebeurier G, Hirth L (1983). Electron microscopic studies of the different topological forms of the cauliflower mosaic virus DNA: knotted encapsidated DNA and nuclear minichromosome. EMBO J 2 1067-1071.

Mercereau-Puijalon O, Lacroute F, Kourilsky P (1980). Synthesis of a chicken ovalbumin-like protein in the yeast Saccharomyces cerevisiae. Gene 11 163-167.

Mercola M, Goverman J, Mirell C, Calame K (1985). Immunoglobulin heavy-chain enhancer requires one or more tissue specific factors. Science 227 266-270.

Miller JH (1972). Assay of β-galactosidase. Experiments in Molecular Genetics, 352-355. Cold Spring Harbor Laboratory, New York.

Miller AM, Mackay VL, Nasmyth KA (1985). Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast. Nature 314 598-603.

Miyajima A, Bond MW, Otsu K, Arai K, Arai N (1985). Secretion of mature mouse interleukin-2 by Saccharomyces cerevisiae: use of a general secretion vector containing promoter and leader sequences of the mating pheromone α-factor. Gene 37 155-161.

Miyajima A, Otsu K, Schreurs J, Bond MW, Abrams J, Arai K (1986). Expression of murine and human granulocyte-macrophage colony stimulating factors in Saccharomyces cerevisiae; mutagenesis of the potential N-linked glycosylation sites. EMBO J 5 1193-1197.

Modjtahedi N, Volovitch M, Sossountzov L, Habricot Y, Bonneville JM, Yot P (1984). Cauliflower mosaic virus-induced viroplasms support viral DNA synthesis in a cell-free system. Virology 133 289-300.

Morelli G, Nagy F, Fraley RT, Rogers SG, Chua NH (1985). A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose 1,5-bisphosphate carboxylase small subunit of pea. Nature 315 200-204.

Morlé F, Lopez B, Henni T, Godet J (1985). α-thalassaemia associated

with the deletion of two nucleotides at positions -2 and -3 preceding the AUG codon. EMBO J <u>4</u> 1245-1250.

Morrison DA (1979). Transformation and preservation of competent bacterial cells by freezing. Methods Enzymol <u>68</u> 326-331.

Mount SM (1982). A catalogue of splice junction sequences. Nucl Acids Res <u>10</u> 459-472.

Mueller PP and Hinnenbusch AG (1986). Multiple upstream AUG codons mediate translational control of GCN4. Cell <u>45</u> 201-207.

Murray JAH, Scarpa M, Rossi N, Cesareni G (1987). Antagonistic controls regulate copy number of the yeast 2μm plasmid. EMBO J <u>6</u> 4205-4212.

Nagawa F and Fink GR (1985). The relationship between the "TATA" sequence and transcription initiation sites at the HIS4 gene of <u>Saccharomyces cerevisiae</u>. Proc Natl Acad Sci USA <u>82</u> 8557-8561.

Nakamura Y, Sato T, Emi M, Miyanohara A, Nishide T, Matsubara K (1986). Expression of human salivary α-amylase gene in <u>Saccharomyces cerevisiae</u> and its secretion using the mammalian signal sequence. Gene <u>50</u> 239-245.

Neill JD, Litts JC, Anderson OC, Greene FC, Stiles JI (1987). Expression of wheat α-gliadin in <u>Saccharomyces cerevisiae</u>. Gene <u>55</u> 303-317.

Oberto J and Davison J (1985). Expression of chicken egg white lysozyme by <u>Saccharomyces cerevisiae</u>. Gene <u>40</u> 57-65.

Odell JT and Howell SH (1980). The identification, mapping and characterisation of mRNA for p66, a cauliflower mosaic virus-coded protein. Virology <u>102</u> 349-359.

Odell JT, Nagy F and Chua NH (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature <u>313</u> 810-812.

Olszewski N, Hagen G and Guilfoyle TJ (1982). A transcriptionally active, covalently closed mini-chromosome of cauliflower mosaic virus DNA isolated from infected turnip leaves. Cell 29 395-402.

Ondek B, Shephard A and Herr W (1987). Discreet elements within the SV40 enhancer region display different cell- specific enhancer activities. EMBO J 6 1017-1025.

Osborne TF, Gaynor RB and Berk AJ (1982). The TATA homology and the mRNA 5' untranslated sequence are not required for expression of essential adenovirus Ela functions. Cell 29 139-148.

Oshima Y (1982). Regulatory circuits for gene expression: the metabolism of galactose and phosphate. Molecular Biology of the yeast Saccharomyces cerevisiae, 159-180. Eds Strathern JN, Jones EW, Broach JR. Cold Spring Harbor Laboratory, New York.

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA (1986). Splicing of messenger RNA precursors. Ann Rev Biochem 55 1119-1150.

Parent SA, Fenimore CM and Bostian KA (1985). Vector systems for the expression, analysis and cloning of DNA sequences in Saccharomyces cerevisiae. Yeast 1 83-138.

Parker CS and Topol J (1984a). A Drosophila RNA polymerase II transcription factor contains a promoter region-specific DNA-binding activity. Cell 36 357-369.

Parker CS and Topol J (1984b). A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp70 gene. Cell 37 273-283.

Paszkowski J, Shillito RD, Saul M, Mandák V, Hohn T, Hohn B, Potrykus I (1984). Direct gene transfer to plants. EMBO J 3 2717-2722.

Pelham HRB (1985). Activation of heat-shock genes in eukaryotes. Trends Genet 1 31-35.

Pelletier J and Sonenberg N (1987). The involvement of mRNA secondary structure in protein synthesis. Biochem Cell Biol 65 576-581.

Penttilä ME, Nevalainen KMH, Raynal A, Knowles JKC (1984). Cloning of *Aspergillus* *niger* genes in yeast. Expression of the gene encoding *Aspergillus* β-glucosidase. Mol Gen Genet 194 494-499.

Petes TD and Williamson DH (1975). Replicating circular DNA molecules in yeast. Cell 4 249-253.

Pfeiffer P and Hohn T (1983). Involvement of reverse transcriptase in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. Cell 33 781-789.

Picard D and Schaffner W (1984). A lymphocyte-specific enhancer in the mouse immunoglobulin κ gene. Nature 307 80-82.

Pikielny CW and Rosbash M (1986). Specific small nuclear RNAs associated with yeast spliceosomes. Cell 45 869-877.

Plant AI, Covey SN and Grierson G (1985). Detection of a subgenomic mRNA for gene V, the putative reverse transcriptase gene of cauliflower mosaic virus. Nucl Acids Res 13 8305-8321.

Price DH, Sluder AE and Greenleaf AL (1987). Fractionation of transcription factors for RNA polymerase II from *Drosophila* Kc cell nuclear extracts. J Biol Chem 262 3244-3255.

Price V, Mochizuki D, March CJ, Cosman D, Deeley MC, Klinke R, Clevenger W, Gillis S, Baker P, Urdal D (1987). Expression, purification and characterisation of recombinant murine granulocyte-macrophage colony stimulating factor and bovine interleukin-2 from yeast. Gene 55 287-293.

Ptashne M (1986). Gene regulation by proteins acting nearby and at a distance. Nature 322 697-701.

Raghow R (1987). Regulation of messenger RNA turnover in eukaryotes. Trends Biochem Sci 12 358-360.

Raynal A and Guerineau M (1984). Cloning and expression of the structural gene for β-glucosidase of *Kluyvromyces* *fragilis* in *Escherichia* *coli* and *Saccharomyces* *cerevisiae*. Mol Gen Genet 195

108-115.

Reeves R (1984). Transcriptionally active chromatin. Biochim Biophys Acta 782 343-393.

Reich TJ, Iyer VN, Miki BL (1986). Efficient transformation of alfalfa protoplasts by the intranuclear microinjection of Ti plasmids. Bio/Technology 4 1001-1004.

Reinberg D, Horikoshi M and Roeder RG (1987). Factors involved in specific transcription in mammalian RNA polymerase II. J Biol Chem 262 3322-3330.

Rhoads RE (1988). Cap recognition and the entry of mRNA into the protein initiation cycle. Trends Biochem Sci 13 52-56.

Richards KE, Guilley H and Jonard G (1981). Further characterisation of the discontinuities in cauliflower mosaic virus. FEBS Lett 134 67-70.

Riva M, Schäffner AR, Sentenac A, Hartmann GR, Mustaev AA, Zaychikov EF, Grachev MA (1987). Active site labelling of the RNA polymerases A, B and C from yeast. J Biol Chem 262 14377-14380.

Roeder GS, Rose AB and Pearlman RE (1985). Transposable element sequences involved in the enhancement of yeast gene expression. Proc Natl Acad Sci USA 82 5428-5432.

Rosenberg S, Barr PJ, Najarian RC, Hallewell RA (1984). Synthesis in yeast of a functional oxidation-resistant mutant of human α_1-antitrypsin. Nature 312 77-80.

Rossouw CMS, Vergeer WP, du Plooy SJ, Bernard MP, Ramirez F, de Wet WJ (1987). DNA sequences in the first exon of the human pro-α1(I) collagen gene enhance transcription. J Biol Chem 262 15151-15157.

Rothstein SJ, Lahners KN, Lotstein RJ, Carozzi NB, Jayne SM, Rice DR (1987). Promoter cassettes, antibiotic resistance genes and vectors for plant transformation. Gene 53 153-161.

Rothstein SJ, Lazarus CM, Smith WE, Baulcombe DC, Gatenby AA (1984). Secretion of a wheat α-amylase expressed in yeast. Nature 308 662-665.

Ruby SW, Szostak JW, Murray AW (1983). Cloning regulated yeast genes from a pool of lacZ fusions. Methods Enzymol 101 253-269.

Ruskin B and Green MR (1985). Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. Nature 317 732-734.

Russell PR (1983). Evolutionary divergence of the mRNA transcription initiation mechanism in yeast. Nature 301 167-169.

Russell PR (1985). Transcription of the triose-phosphate- isomerase gene of Schizosaccharomyces pombe initiates at a start point different from that in Saccharomyces cerevisiae. Gene 40 125-130.

Ryoji M and Worcel A (1984). Chromatin assembly in Xenopus oocytes: in vivo studies. Cell 37 21-32.

Sadofsky M, Connelly S, Manley JL, Alwine JC (1985). Identification of a sequence element on the 3' side of AAUAAA which is necessary for simian virus 40 late mRNA 3' end processing. Mol Cell Biol 5 2713-2719.

Sanders PR, Winter JA, Barnason AR, Rogers SG, Fraley RT (1987). Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. Nucl Acids Res 15 1543-1558.

Sarokin L and Carlson M (1986). Short repeated elements in the upstream regulatory region of the SUC2 gene of Saccharomyces cerevisiae. Mol Cell Biol 6 2324-2333.

Sassone-Corsi P and Borrelli E (1986). Transcriptional regulation by trans-acting factors. Trends Genet 2 215-219.

Sawadogo M and Roeder RG (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box. Cell 43 165-175.

Scarpulla RC and Nye SH (1986). Functional expression of rat

cytochrome c in <u>Saccharomyces</u> <u>cerevisiae</u>. Proc Natl Acad Sci USA <u>83</u> 6352-6356.

Schell J (1987). Transgenic plants as tools to study the molecular organisation of plant genes. Science <u>237</u> 1176-1183.

Scherer S, Mann C and Davis RW (1982). Reversion of a promoter deletion in yeast. Nature <u>298</u> 815-819.

Schlissel MS and Brown DD (1984). The transcriptional regulation of <u>Xenopus</u> 5S RNA genes in chromatin: the roles of active stable transcription complexes and histone Hl. Cell <u>37</u> 903-913.

Schoelz J, Sheperd RJ, Daubert S (1986). Region VI of cauliflower mosaic virus encodes a host range determinant. Mol Cell Biol <u>6</u> 2632-2637.

Schöler HR and Gruss P (1984). Specific interaction between enhancer-containing molecules and cellular components. Cell <u>36</u> 403-411.

Schultz LD, Tanner J, Hofmann KJ, Emini EA, Condra JH, Jones RE, Kieff E, Ellis RW (1987). Expression and secretion in yeast of a 400kDa envelope glycoprotein derived from Epstein- Barr virus. Gene <u>54</u> 113-123.

Selleck SB and Majors J (1987). Photofootprinting *in vivo* detects transcription-dependent changes in yeast TATA boxes. Nature <u>325</u> 173-177.

Sentenac A (1985). Eukaryotic RNA polymerases. CRC Crit Rev Biochem <u>18</u> 31-90.

Sharma S and Godson GN (1985). Expression of the major surface antigen of <u>Plasmodium</u> <u>knowlesi</u> sporozoites in yeast. Science <u>228</u> 879-882.

Shatkin AJ (1976). Capping of eukaryotic mRNAs. Cell <u>9</u> 645-653.

Shaw CH, Carter GII, Watson MD, Shaw CH (1984). A functional map of the nopaline synthase promoter. Nucl Acids Res <u>12</u> 7831-7846.

Shaw G and Kamen R (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA that mediates selective mRNA degradation. Cell 46 659-667.

Shepherd RJ (1979). DNA plant viruses. Ann Rev Plant Physiol 30 405-423.

Shewmaker CK, Caton JR, Houck CM, Gardner RC (1985). Transcription of cauliflower mosaic virus integrated into plant genomes. Virology 140 281-288.

Shore D, Stillman DJ, Brand AH, Nasmyth KA (1987). Identification of silencer-binding proteins from yeast: possible roles in SIR control and DNA replication. EMBO J 6 461-467.

Silhavy T and Beckwith JR (1985). Uses of lac fusions for the study of biological problems. Microbiol Rev 49 398-418.

Silhavy T, Berman ML and Enquist LW (1984). Experiments with gene fusions, Appendix I. Cold Spring Harbor Laboratory, New York.

Silhavy T, Shuman H, Beckwith J, Schwartz M (1977). Use of gene fusions to study outer membrane protein localisation in Escherichia coli. Proc Natl Acad Sci USA 74 5411-5415.

Simpson J, Schell J, Van Montagu M, Herrera-Estrella L (1986). Light-inducible and tissue-specific pea lhcp gene expression involves an upstream element combining enhancer- and silencer- like properties. Nature 323 551-554.

Sinden RR, Carlson JO, Pettijohn DE (1980). Torsional tension in the DNA double helix measured with trimethylpsoralen in living E.coli cells: analogous measurements in insect and human cells. Cell 21 773-783.

Spena A, Hain R, Ziervogel R, Saedler H, Schell J (1985). Construction of a heat-inducible gene for plants. Demonstration of heat-inducible activity of the Drosophila hsp70 promoter in plants. EMBO J 2739-2743.

St John TP and Davies RD (1981). The organisation and transcription of the galactose cluster of Saccharomyces. J Mol Biol 152 285-315.

Stark MJR (1987). Multicopy expression vectors carrying the lac repressor gene for the regulated expression of genes in Escherichia coli. Gene 51 253-265.

Stepien PP, Brousseau R, Wu R, Narang S, Thomas DY (1983). Synthesis of a human insulin gene (VI). Expression of the synthetic proinsulin gene in yeast. Gene 24 289-297.

Stinchcomb DT, Struhl K, Davis RW (1979). Isolation and characterisation of a yeast chromosomal replicator. Nature 282 39-43.

Struhl K (1983). The new yeast genetics. Nature 305 391-397.

Struhl K (1985). Naturally occuring poly(dA-dT) sequences are upstream elements for constitutive transcription in yeast. Proc Natl Acad Sci USA 82 8419-8423.

Struhl K (1985a). Negative control at a distance mediates catabolite repression in yeast. Nature 317 822-824.

Struhl K (1987). Promoters, activator proteins and the mechanism of transcription initiation in yeast. Cell 49 295-297.

Struhl K, Stinchcomb DT, Scherer S, Davis RW (1979). High frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc Natl Acad Sci USA 76 1035-1039.

Summers J and Mason WS (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. Cell 29 403-415.

Sutton A and Broach JR (1985). Signals for transcription initiation and termination in the yeast Saccharomyces cerevisiae plasmid 2μm circle. Mol Cell Biol 5 2770-2780.

Sweetser D, Nonet M, Young RA (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc Natl Acad Sci USA 84

1192-1196.

Szostak JW and Blackburn EH (1982). Cloning yeast telomeres on linear plasmid vectors. Cell 29 245-255.

Tague BW and Chrispeels MJ (1987). The plant vacuolar protein, phytohaemagglutinin, is transported to the vacuole of transgenic yeast. J Cell Biol 105 1971-1979.

Tajima M, Nogi Y and Fukasawa T (1986). Duplicate upstream activating sequences in the promoter region of the Saccharomyces cerevisiae GAL7 gene. Mol Cell Biol 6 246-256.

Takatsuji H, Hirochika H, Fukushi T, Ikeda J (1986). Expression of cauliflower mosaic virus reverse transcriptase in yeast. Nature 319 240-243.

Thiele DJ and Hamer DH (1986). Tandemly duplicated upstream control sequences mediate copper-induced transcription of the Saccharomyces cerevisiae copper-metallothionein. Mol Cell Biol 6 1158-1163.

Thiele DJ, Walling MJ, Hamer DH (1986). Mammalian metallothionein is functional in yeast. Science 231 854-856.

Timko PM, Kausch AP, Castresana C, Fassler J, Herrera-Estrella L, Van den Broek G, Van Montagu M, Schell J, Cashmore AR (1985). Light regulation of plant gene expression by an upstream enhancer-like element. Nature 318 579-582.

Toh H, Hayashida H, Miyata T (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature 305 827-829.

Turner DS and Covey SN (1984). A putative primer for the replication of CaMV by reverse transcriptase is virion- associated. FEBS Lett 165 285-289.

Ulrich TH, Chowdhury JB, Widholm JM (1980). Callus and root formation from mesophyll protoplasts of Brassica rapa. Plant Science Letters 19 347-354.

Urdea MS, Merryweather JP, Mullenbach GT, Coit D, Heberlein U, Valenzuela P, Barr PJ (1983). Chemical synthesis of a gene for human epidermal growth factor urogastrone and its expression in yeast. Proc Natl Acad Sci USA 80 7461–7465.

Valenzuela P, Medina A, Rutter WJ, Ammerer G and Hall BD (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. Nature 298 347–350.

Vapnek D and Case M (1982). Gene cloning in Neurospera crassa. Curr Top Microbiol Immunol 96 97–100.

Varmus HE (1982). Form and function of retroviral proviruses. Science 216 812–820.

Volkert FC and Broach JR (1986). Site specific recombination promotes plasmid amplification in yeast. Cell 46 541–550.

Volovitch M, Modjtahedi N, Yot P and Brun G (1984). RNA-dependent DNA polymerase activity in CaMV-infected plant leaves. EMBO J 3 309–314.

Walden RM and Howell SH (1982). Intragenomic recombination events among pairs of defective cauliflower mosaic virus genomes in plants. J Mol Appl Genet 1 447–456.

Wasylyk B, Kédinger C, Corden J, Brison O, Chambon P (1980). Specific in vitro initiation of transcription on conalbumin and ovalbumin genes and comparison with adenovirus 2 early and late genes. Nature 285 367–373.

Wasylyk C and Wasylyk B (1986). The immunoglobulin heavy-chain B lymphocyte enhancer efficiently stimulates transcription in non-lymphoid cells. EMBO J 5 553–560.

Watts F, Castle C and Beggs J (1983). Aberrant splicing of Drosophila alcohol dehydrogenase transcripts in Saccharomyces cerevisiae. EMBO J 2 2085–2091.

Wei R, Wilkinson H, Pfeifer K, Schneider C, Young R, Guarente L (1986).

Two or more copies of Drosophila heat shock consensus sequence serve to activate transcription in yeast. Nucl Acids Res 14 8183–8188.

Weil PA, Luse DS, Segall J, Roeder RG (1979). Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell 18 469–484.

Weintraub H (1984). Histone H1-dependent chromatin superstructures and the suppression of gene activity. Cell 38 17–27.

Weisbrod S (1982). Active chromatin. Nature 297 289–295.

Werr W and Lörz H (1986). Transient gene expression in a Graminaceae cell line. (A rapid procedure for studying plant promoters). Mol Gen Genet 202 471–475.

West RW, Yocum RR and Ptashne M (1984). Saccharomyces cerevisiae GAL1–GAL10 divergent promoter region: location and function of the upstream activating sequence UAS_G. Mol Cell Biol 4 2467–2478.

Wiederrecht G, Shuey DJ, Kibbe WA, Parker CS (1987). The Saccharomyces and Drosophila heat shock transcription factors are identical in size and DNA binding properties. Cell 48 507–515.

Williams JG and Mason PL (1985). Hybridisation in the analysis of RNA. Nucleic Acid Hybridisation: a Practical Approach, 139–160. Ed Hames BD and Higgins SJ. IRL Press.

Willmitzer L (1988). The use of transgenic plants to study plant gene expression. Trends Genet 4 13–18.

Wood CR, Boss MA, Kenten JH, Calvert JE, Roberts NA, Emtage JS (1985). The synthesis and in vivo assembly of functional antibodies in yeast. Nature 314 446–449.

Woolston CJ, Covey SN, Penswick JR and Davies JW (1983). Aphid transmission and a polypeptide specified by a defined region of the cauliflower mosaic virus genome. Gene 23 15–23.

Workman JL and Roeder RG (1987). Binding of transcription factor TFIID to the major late promoter during *in vitro* nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell <u>51</u> 613-622.

Wouldt LP, Smit AB, Mager WM, Planta RJ (1986). Conserved sequence elements upstream of the gene encoding yeast ribosomal protein L25 are involved in transcription activation. EMBO J <u>5</u> 1037-1040.

Wright CF and Zitomer RS (1985). Point mutations implicate repeated sequences as essential elements of the CYC7 negative upstream site in <u>Saccharomyces cerevisiae</u>. Mol Cell Biol <u>5</u> 2951-2958.

Xiong C, Lebeurier G and Hirth L (1984). Detection *in vivo* of a new gene product (gene III) of cauliflower mosaic virus. Proc Natl Acad Sci USA <u>81</u> 6608-6612.

Xiong C, Muller S, Lebeurier G and Hirth L (1982). Identification by immunoprecipitation of cauliflower mosaic virus *in vitro* major translation product with a specific serum against viroplasm protein. EMBO J <u>1</u> 971-976.

Yamashita I, Itoh T, Fukui S (1985). Cloning and expression of the <u>Saccharomycopsis fibuligera</u> α-amylase gene in <u>Saccharomyces cerevisiae</u>. Agric Biol Chem <u>49</u> 3089-3091.

Yamashita I, Itoh T, Fukui S (1985). Cloning and expression of the <u>Saccharomycopsis fibuligera</u> glucoamylase gene in <u>Saccharomyces cerevisiae</u>. Appl Microbiol Biotechnol <u>23</u> 130-133.

Yamashita I and Fukui S (1983). Molecular cloning of a glucoamylase-producing gene in the yeast <u>Saccharomyces</u>. Agric Biol Chem <u>47</u> 2689-2692.

Yarger JG, Armilei G and Gorman MC (1986). Transcription terminator-like element within a <u>Saccharomyces cerevisiae</u> promoter region. Mol Cell Biol <u>6</u> 1095-1101.

Yocum RR, Hanley S, West R and Ptashne M (1984). Use of lacZ fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10

promoter in <u>Saccharomyces</u> <u>cerevisiae</u>. Mol Cell Biol <u>4</u> 1985-1998.

Zakian VA, Brewer BJ and Fangman WL (1979). Replication of each copy of the yeast 2 micron DNA plasmid occurs during the S phase. Cell <u>17</u> 923-934.

Zaret KS and Sherman F (1982). DNA sequence required for efficient transcription termination in yeast. Cell <u>28</u> 563-573.

Zaret KS and Sherman F (1984). Mutationally altered 3' ends of CYC1 mRNA affect transcript stability and translational efficiency. J Mol Biol <u>177</u> 107-135.

Zhu J, Contreras R and Fiers W (1986). Construction of stable laboratory and industrial yeast strains expressing a foreign gene by integrative transformation using a dominant selection system. Gene <u>50</u> 225-237.

Zinn K and Maniatis T (1986). Detection of factors that interact with the human β-interferon regulatory region *in vivo* by DNAaseI footprinting. Cell <u>45</u> 611-618.

Zitomer RS, Walthall DA, Rymond BC, Hollenberg CP (1984). <u>Saccharomyces</u> <u>cerevisiae</u> ribosomes recognise non-AUG initiator codons. Mol Cell Biol <u>4</u> 1191-1197.

JENNIFER H RICHARDSON   PhD THESIS ABSTRACT, 1988

LEICESTER BIOCENTRE AND DEPARTMENT OF GENETICS, UNIVERSITY OF LEICESTER

CAMV GENE EXPRESSION:
THE ANALYSIS OF TWO CAMV PROMOTERS IN YEAST AND HIGHER PLANTS

The aim of this study was to assess the feasibility of using the budding yeast Saccharomyces cerevisiae as a system in which to analyse plant promoters. The promoters chosen for study were the 19S and 35S promoters of cauliflower mosaic virus (CaMV) which, like cellular plant promoters, are transcribed in the plant nucleus by host cell RNA polymerase II. A complete CaMV genome was introduced into yeast on a 2 micron plasmid-based vector and using Northern blot analysis, several CaMV-hybridising transcripts were detected. More precise information on the activity of the promoters was obtained by constructing gene fusions in which the 19S and 35S promoters were linked to the bacterial lacZ gene. Biochemical assays for β-galactosidase showed that cells harbouring the 19S-lacZ gene expressed β-galactosidase but those harbouring the 35S-lacZ gene did not. The insertion of a yeast transcription termination signal upstream of the 19S promoter did not abolish or diminish expression of the 19S-lacZ gene. β-galactosidase was present at low levels in cells expressing 19S-lacZ, constituting less than 0.01% of total cell protein. The 5'ends of 19S-lacZ transcripts present in yeast were mapped by primer extension. The major RNA species initiated approximately 250bp upstream of the 19S-lacZ coding region, indicating the existence of a fortuitous promoter in this region of the CaMV DNA. Two less abundant RNA species initiated within the 19S-lacZ open reading frame at positions +9 and +25bp and may be produced from the genuine 19S promoter. There is evidence to suggest that one or both of these shorter transcripts is the functional mRNA for β-galactosidase. All three classes of RNA were polyadenylated. Coupling of the 19S-lacZ gene to a yeast enhancer (the GAL UAS) produced a 5-fold increase in β-galactosidase activity. At the transcriptional level, activation of the enhancer resulted in a massive increase in the level of the RNA initiating at -250bp but had a minor influence of the levels of the two RNA species initiating at +9 and +25. A series of deletion mutations within the 19S promoter was constructed using Bal31 nuclease. Analysis of these mutations in yeast revealed that sequences from -500 to -193bp and from -137 to -62bp were not required for 19S promoter function, but a deletion from -62 to -21bp (which removes the putative TATA box) severely reduced 19S-lacZ gene expression. Transgenic tobacco plants containing the 19S promoter deletions fused to a CAT gene were produced by A.tumefaciens-mediated gene transfer but the analysis of these plants was not completed.