

The Structure of the Isocitrate Lyase Gene
of Cucumber and its Transcriptional
Regulation in Higher Plants

Susan Jennifer Reynolds

PhD

University of Edinburgh

1994



Abstract

The glyoxylate cycle operates in oilseeds during postgerminative development when it is involved in the conversion of fatty acids released from lipid reserves, to sucrose, to support growth until the seedling becomes photosynthetically competent. Isocitrate lyase (ICL) and malate synthase (MS) are key enzymes of this cycle, serving to by-pass the carbon dioxide liberating steps of the Krebs' cycle. In cucumber, synthesis of these two enzymes is co-ordinately regulated during seedling development. Using the glyoxylate cycle as a model system, it is anticipated that an insight can be gained into the factors responsible for metabolic and developmental regulation of plant gene expression.

A partial cDNA clone encoding ICL was sequenced and used as a probe in Southern analysis to reveal that the *icl* gene is single copy in cucumber. Transcript levels of both ICL and MS in seedlings up to eight days post-imbibition were analysed by Northern hybridisation and confirm co-ordinate expression of the two genes. A cucumber genomic library was constructed and a clone containing the *icl* gene was isolated. The nucleotide sequence was determined and the structure of the gene analysed. The start of transcription was mapped by primer extension. The 6.5 kb fragment carrying the *icl* gene was used to generate transgenic *Nicotiana plumbaginifolia*. Analysis of these transformants revealed that the transgene was faithfully transcribed following seed germination.

Sequence comparisons of the promoter regions of the *icl* and *ms* genes revealed areas of homology potentially important in the co-ordinate transcriptional regulation of these genes. Gene fusion studies using the β -glucuronidase (GUS) reporter gene were carried out in both stable transformation and transient expression systems. Promoter fragments of lengths 2.9 kb and 0.5 kb were linked to the GUS gene and transferred into *N. plumbaginifolia*. Fluorometric and histochemical analysis of the transformants indicated that, *in vivo*, both promoter fragments faithfully directed the temporal regulation of *icl* gene expression during postgerminative growth. However, sequences important for metabolic control of expression were not present in the shorter promoter fragment. This was confirmed by the findings of the transient expression system, in which several promoter-GUS fusions were introduced into cucumber protoplasts by electroporation. With *icl* promoter fragments of more than 1 kb, the levels of GUS were markedly reduced in the presence of sucrose; this was not the case for those shorter than 1 kb. These results illustrate that there are both developmental and metabolic controls acting on the *icl* gene of cucumber. Functional analysis of the *icl* promoter suggests that separate elements are responsible for these different levels of control.

Acknowledgements

I would like to thank Dr. Steve Smith for all his help and advice during the course of the work for this thesis. I would also like to thank members of the lab and the department, past and present, for their friendship and encouragement.

I am extremely grateful to Prof. Chris Leaver for letting me visit his lab in Oxford to learn about electroporation, and particularly to Ian Graham for his help.

Finally, to Jon, thank you for everything.

Abbreviations

(v/v)	volume:volume ratio
(w/v)	weight:volume ratio
°C	degrees Celsius
A	absorbance
ADP	adenosine 5'-diphosphate
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
CaMV	Cauliflower mosaic virus
CAT	chloramphenicol acetyl transferase
cDNA	complementary deoxyribonucleic acid
Ci	Curie
CoA	coenzyme A
DNA	deoxyribonucleic acid
DPA	days post anthesis
ds	double stranded
EDTA	ethenediaminetetraacetic acid (disodium salt)
g	gramme
g	relative centrifugal force
GO	glycolate oxidase
GUS	β-glucuronidase
h	hour
HPR	hydroxypyruvate reductase
ICL	isocitrate lyase
IMH	Isocitrate lyase - Malate synthase - Homology
kbp	kilobase pair
kDa	kilodalton
l	litre
lb in ⁻²	pounds per square inch
M	molar
m	metre
min	minute
M _r	relative molecular mass
mRNA	messenger RNA
MS	malate synthase

MU	methyl umbelliferone
MUG	methyl umbelliferyl glucuronide
NAD ⁺	nicotinamide adenine nucleotide (oxidised form)
NADH	nicotinamide adenine nucleotide (reduced form)
NPTII	neomycin phosphotransferase, type II
OR	origin of replication
ORF	open reading frame
polyA	polyadenylation
PTS	peroxisomal targeting signal
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
sdd H ₂ O	sterile double distilled water
SDS	sodium dodecyl sulphate
SGAT	serine:glyoxylate amino transferase
Ti	tumour inducing
Tris	tris-(hydroxymethyl)-methylamine
U	units
UTR	untranslated region
v	volume
w	weight

Table of contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Abbreviations	iv
Table of contents	vi
CHAPTER 1	
1.1 Introduction	2
1.2 Peroxisomes	
1.2.1 Structure of peroxisomes	2
1.2.2 Biogenesis of peroxisomes	3
1.2.3 Transition between classes of peroxisomes	3
1.2.4 Targeting of proteins to peroxisomes	4
1.2.4.1 The carboxy terminal signal	4
1.2.4.2 The amino terminal signal	5
1.2.4.3 Internal targeting signals	6
1.2.5 Biochemistry of peroxisomes	6
1.2.5.1 Leaf-type peroxisomes	6
1.2.5.2 Peroxisomes in root nodules	7
1.2.5.3 Glyoxysomes	9
1.3 Synthesis of ICL and MS during plant development	10
1.3.1 Synthesis of ICL and MS during embryogenesis and seed maturation	12
1.3.2 Synthesis of ICL and MS during postgerminative growth	12
1.3.3 Synthesis of ICL and MS during senescence	13
1.4 Metabolic Regulation of gene expression	
1.4.1 Carbon catabolite repression	14
1.4.2 Induction of plant genes by sucrose	16
1.4.3 Metabolic regulation of the <i>icl</i> and <i>ms</i> genes	16
1.5 Control of gene expression	
1.5.1 Regulation of transcription in prokaryotes	18
1.5.2 Transcriptional regulation in eukaryotes	19
1.5.3 Assembly of the preinitiation complex	20

1.5.4 Activation of the preinitiation complex	21
1.5.4.1 DNA binding motifs	22
1.5.4.2 Activation domains	24
1.5.4.3 Topology of transcriptional activation	25
1.5.5 Repression	26
1.5.6 Silencing and genomic imprinting	28
1.5.7 Post-transcriptional regulation	29
1.5.8 Post-translational regulation	31
1.6 Aims of this thesis	32

CHAPTER 2

2.1 Biological material	
2.1.1 Plant material	34
2.1.2 Bacterial strains and genotypes	34
2.1.3 Bacterial plasmids and bacteriophage	35
2.1.4 Isocitrate lyase cDNA clone (partial)	35
2.1.5 Malate synthase cDNA clone	35
2.1.6 Cucumber cDNA library in λ ZAPII	35
2.2 Miscellaneous	
2.2.1 General comments	36
2.2.2 Bacteriological media	36
2.2.3 Tissue culture media	36
2.2.4 Restriction endonucleases and DNA modification enzymes	37
2.2.5 Antibiotics	37
2.2.6 Radiochemicals	37
2.2.7 Autoradiography	38
2.3 DNA isolation, manipulation and analysis	
2.3.1 Small scale plasmid/phagemid isolation	38
2.3.2 Large scale plasmid/phagemid isolation	38
2.3.3 Genomic DNA isolation	39
2.3.4 Bacteriophage DNA isolation	40
2.3.5 Agrobacterium DNA isolation	40
2.3.6 Estimation of DNA and RNA concentration	41
2.3.7 Restriction endonuclease analysis of DNA samples	41
2.3.8 Horizontal gel electrophoresis of DNA	41
2.3.9 Southern blotting	41

2.3.10 DNA sequencing and computer analysis	
2.3.10.1 Sequencing double stranded plasmid DNA	42
2.3.10.2 Computer analysis	42
2.3.11 Generation of nested deletions	43
2.4 RNA isolation and analysis	
2.4.1 Isolation of total RNA from cucumber cotyledons	44
2.4.2 Isolation of total RNA from small amounts of plant tissue	45
2.4.3 Isolation of total RNA from protoplasts	45
2.4.4 Horizontal gel electrophoresis of RNA and Northern blotting	46
2.4.5 Primer extension	46
2.5 DNA fragment subcloning and bacterial transformation	
2.5.1 DNA fragment isolation and preparation for subcloning	47
2.5.2 Vector preparation for the receipt of DNA inserts	47
2.5.3 Ligation of the prepared insert and vector DNA	47
2.5.4 Transformation of <i>E. coli</i> with plasmid or phagemid DNA	48
2.5.5 Screening of bacterial colonies transformed with recombinant plasmid or phagemid DNA	
2.5.5.1 Screening by α -complementation	48
2.5.5.2 Screening of bacterial transformants by colony hybridisation	49
2.6 Construction and screening of libraries	
2.6.1 Construction of cucumber genomic library in λ NM1149	
2.6.1.1 Insert preparation	49
2.6.1.2 Vector preparation (λ NM1149)	
2.6.1.2a Ligation of cohesive ends	50
2.6.1.2b Digestion of λ vector	50
2.6.1.2c Treatment of digested vector with phosphatase	51
2.6.1.3 Ligation of vector and insert DNA	51
2.6.1.4 Packaging of λ DNA	51
2.6.2 Screening of recombinant bacteriophage clones	
2.6.2.1 Preparation of host cells and titration of λ libraries	52

2.6.2.2 Plating and screening recombinant bacteriophage libraries	52
2.6.3 Preparation of plate lysate stocks	53
2.6.4 <i>In vivo</i> excision of recombinant λ ZAPII DNA	53
2.7 Radiolabelling of DNA probes	
2.7.1 Labelling of ds DNA probes by random priming	54
2.7.2 Radiolabelling of oligonucleotides	55
2.8 Hybridisation of radiolabelled DNA probes to membrane bound DNA or RNA	55
2.9 Production and analysis of transgenic plants	
2.9.1 Construction of vectors for plant transformation	56
2.9.2 Transfer of recombinant plasmids into <i>Agrobacterium tumefaciens</i> by conjugation and screening of transformed bacteria	56
2.9.3 <i>Nicotiana plumbaginifolia</i> transformation and regeneration	
2.9.3.1 General conditions	57
2.9.3.2 Preparation of plant material	57
2.9.3.3 Preparation of transformed <i>A. tumefaciens</i> for inoculation of leaf pieces	57
2.9.3.4 Inoculation of leaf pieces and regeneration of transformed plants	57
2.9.3.5 Production and collection of seed	58
2.9.4 Analysis of transgenic <i>N. plumbaginifolia</i> plants	
2.9.4.1 Germination of transgenic <i>N. plumbaginifolia</i> seeds	58
2.9.4.2 Selection of germinating seed at specific developmental stages	58
2.9.4.3 Fluorometric assay of β -glucuronidase (GUS) in germinating seedlings	59
2.9.4.4 Histochemical assay of β -glucuronidase	60
2.9.4.5 Treatment of transformed <i>N. plumbaginifolia</i> seedlings with sucrose	60
2.9.4.6 Northern analysis of germinating transgenic <i>N. plumbaginifolia</i> seeds	60

2.10 Transient expression in cucumber protoplasts	
2.10.1 Construction of vectors for transient expression	60
2.10.2 Isolation of cucumber protoplasts	61
2.10.3 Electroporation and culturing of cucumber protoplasts	61
2.10.4 Analysis of transfected protoplasts	
2.10.4.1 Fluorometric assay of β -glucuronidase in transfected cucumber protoplasts	62
2.10.4.2 Northern analysis of transfected protoplasts	63
2.11 Protein concentration estimation	63
CHAPTER 3	
3.1 Introduction and aims	65
3.2 Determination of gene copy number	65
3.3 Characterisation of the cDNA clone, pBSICL	
3.3.1 Restriction endonuclease analysis	67
3.3.2 Nucleotide sequence analysis	67
3.4 Isolation of a partial cDNA clone encoding isocitrate lyase	71
3.5 Isolation and sequencing of the isocitrate lyase gene	71
3.6 Features of the isocitrate lyase gene	
3.6.1 Features of the 5' flanking region	76
3.6.2 Features of the coding region	79
3.6.3 Features of the 3' flanking region	79
3.7 Features of the isocitrate lyase polypeptide	82
3.8 Isocitrate lyase and malate synthase are coordinately synthesised during germination	87
3.9 Comparison of the sequences of the isocitrate lyase and malate synthase promoters	89
3.10 Conclusion	91
CHAPTER 4	
4.1 Introduction and Aims	93
4.2 Construction of a plant transformation vector containing the isocitrate lyase gene of cucumber	94
4.3 Transfer of the isocitrate lyase gene into <i>N. plumbaginifolia</i>	94
4.4 Analysis of transcription of the isocitrate lyase transgene in <i>N. plumbaginifolia</i>	97

4.5 Temporal expression of the cucumber isocitrate lyase transgene in <i>N. plumbaginifolia</i>	97
4.6 Expression of an isocitrate lyase-glucuronidase fusion gene in <i>N. plumbaginifolia</i>	
4.6.1 Construction of the fusion gene and transfer to <i>N. plumbaginifolia</i>	100
4.6.2 Histochemical localisation of GUS activity in germinating seedlings	105
4.6.3 Temporal expression of the <i>icl</i> -GUS transgene in <i>N. plumbaginifolia</i>	108
4.7 Effect of removal of a 2.3 kb fragment of the 5' flanking region on the expression of the <i>icl</i> -GUS fusion gene in <i>N. plumbaginifolia</i>	
4.7.1 Construction of a 572 bp <i>icl</i> -GUS fusion and transfer to <i>N. plumbaginifolia</i>	111
4.7.2 Analysis of expression of the 572 bp <i>icl</i> -GUS fusion construct in germinating seedlings	111
4.8 Regulation of expression of the <i>icl</i> gene by metabolic status	
4.8.1 Metabolic control of ICL and MS in pro- and eukaryotes	115
4.8.2 Effect of metabolic status on GUS expression in transformed <i>N. plumbaginifolia</i> seedlings	116
4.9 Conclusion	118
CHAPTER 5	
5.1 Introduction and Aims	121
5.1.1 Transfection of protoplasts by electroporation	121
5.1.2 The cucumber protoplast transient assay system	122
5.2 Isocitrate lyase transcript levels in protoplasts cultured on different carbon sources	122
5.3 Expression from the isocitrate lyase promoter in cucumber mesophyll protoplasts	124
5.4 Deletion of the isocitrate lyase promoter and the effect of these deletions on expression from the promoter	
5.4.1 Construction of promoter-GUS fusions	126
5.4.2 Expression from the deletions of the isocitrate lyase promoter in cucumber mesophyll protoplasts	126
5.5 Conclusion	129

CHAPTER 6

6.1 Isolation and characterisation of the isocitrate lyase gene of cucumber	133
6.2 Developmental and metabolic regulation of isocitrate lyase and malate synthase	133
6.3 Further analysis of factors involved in the regulation of isocitrate lyase and malate synthase	138
REFERENCES	141

CHAPTER 1

INTRODUCTION

1.1 Introduction

Plant gene expression through development and in response to environmental changes is by necessity a highly regulated process. Elucidating the factors which regulate expression of plant genes is therefore fundamental to understanding the control of plant development. To this end, specific genes can be isolated, their regulatory regions identified and those important at specific developmental stages, or in response to environmental changes, defined.

The glyoxylate cycle is an ideal system for such studies on gene regulation. Glyoxylate cycle enzyme activities are localised in a defined cellular compartment, the peroxisome, and the cycle functions both at specific developmental stages and in response to changes in the metabolic environment. In plants, two enzymes are unique to this cycle, malate synthase (MS) and isocitrate lyase (ICL), providing excellent candidates for studying the regulation of plant gene expression.

In this introduction, the biogenesis and function of various classes of the peroxisome within plant cells will be discussed, followed by the regulation of the synthesis of ICL and MS. Finally, the various levels at which gene expression may be controlled will be described.

1.2 Peroxisomes

1.2.1 Structure of peroxisomes

Peroxisomes are present in all eukaryotic cells. In plant cells they range in size from 0.1 μm in diameter in undifferentiated parenchyma cells to 1.8 μm in diameter in green leaves, soybean nodules and *Nitella* filaments (Huang *et al.*, 1983). Peroxisomes are bounded by a single membrane and have simple ultrastructure, lacking internal membranes and an organellar genome. All peroxisomes contain characteristic enzymes such as those involved in fatty acid β -oxidation, hydrogen peroxide producing oxidases and catalase (Huang *et al.*, 1983). Different classes of peroxisomes in plants can be defined dependent on their complement of enzymes and hence biochemical function, for example, the glyoxysomes and gerontosomes, leaf peroxisomes and those involved in ureide metabolism in the root nodules of leguminous plants (see section 1.2.5).

1.2.2 Biogenesis of peroxisomes

Although peroxisomes were once thought to form by vesiculation of the endoplasmic reticulum (ER), it is now generally agreed that peroxisome biogenesis takes place through the fission of existing organelles and growth takes place by post-translational uptake of proteins synthesised on free polyribosomes (Lazarow & Fujiki, 1985; Borst, 1989; Subramani, 1993).

Proliferation of the peroxisome number may precede peroxisome growth as is seen in the yeast *Candida boidinii* when moved to a methanol, glucose-free, medium (Veenhuis & Goodman, 1990). In contrast, when the yeast *Hansenula polymorpha* is shifted from glucose to methanol containing medium, growth of peroxisomes occurs until they are 80 times their original volume before dividing into smaller organelles (Veenhuis *et al.*, 1979). The existence of these two pathways is not surprising as each step may take place independently of the other. The *pas8* mutants of the yeast *Pichia pastoris* are deficient in peroxisomal protein import (see section 1.2.4), but the peroxisomes are able to proliferate and segregate to daughter cells (McCollum *et al.*, 1993).

The origin of the first peroxisome is still a matter of speculation. Peroxisomes do not form *de novo*, *Saccharomyces cerevisiae* cells grown on glucose are found to have at least one small peroxisome which grows and proliferates when the organism is moved to a glucose-free medium (Aitchison *et al.*, 1992). The likelihood is that it was an endosymbiont, similar to mitochondria and chloroplasts. Although peroxisomes only have a single membrane and no DNA, it is not inconceivable that, as with mitochondria that have lost 90 % of their DNA to the nucleus, another endosymbiont could lose 100 %. Some endocytic parasites exist that have only one membrane (de Duve, 1983), suggesting that this too is feasible. Analysis of the β -oxidation system of peroxisomes indicates it is more similar to prokaryotic ones than the mitochondrial system, having multifunctional complexes which are not present in mitochondria (Kunau *et al.*, 1988). Similarly, peroxisomal thiolases from many species appear to have evolved from a common ancestor, so lending weight to the endosymbiont theory (Igual *et al.*, 1992).

1.2.3 Transition between classes of peroxisomes

During the lifetime of some plant cells, for example, those in the cotyledons of cucurbits, functional transitions take place. Post-germinative cucumber cotyledons are involved in lipid mobilisation until greening when they become a photosynthetic organ until they senesce (Becker *et al.*, 1978; Huang *et al.*, 1983). The enzyme complement of the peroxisomes of the cells changes radically through these transitions. The

question of whether the change is accompanied by the destruction of one population and the synthesis of a new one, or is the result of the conversion of existing organelles to a new function has been investigated by many workers (for review see Beevers, 1979). The evidence now points clearly to transition from one function to another. Peroxisomes of postgerminative cotyledon cells at the stage of transition from heterotrophy to autotrophy have been shown to contain enzymes typical of both glyoxysomes and leaf peroxisomes by immunogold double labelling of ultra-thin sections (Nishimura *et al.*, 1986; Sautter, 1986; Titus & Becker, 1985). In addition, malate synthase (MS), a marker enzyme of glyoxysomes, can efficiently be imported into both pumpkin leaf peroxisomes and glyoxysomes, but is specifically degraded upon import into peroxisomes isolated from the transition stage (Mori & Nishimura, 1989). No degradation of MS was seen when imported into leaf peroxisomes from a later stage, showing that the degradation machinery is only present during greening.

1.2.4 Targeting of proteins to peroxisomes

Peroxisomal proteins, similar to nuclear-encoded mitochondrial and chloroplast proteins, are synthesised on free poly-ribosomes and imported post-translationally into the organelle (Lazarow & Fujiki, 1985). However, in contrast to mitochondrial and chloroplast import, in the majority of cases, the import of proteins into peroxisomes does not involve proteolytic processing (Scandalios, 1990), including ICL (Maeshima *et al.*, 1988), implying that the targeting signal resides within the mature polypeptide (Lazarow & Fujiki, 1985).

1.2.4.1 The carboxy terminal signal

Peroxisomal targeting signals do not appear to be as well defined as those for other organelles, with more than one signal functioning to direct proteins to peroxisomes. The first signal identified was that residing at the carboxy terminus of the peroxisomal protein, firefly luciferase (Gould *et al.*, 1987), now called a class I peroxisomal targeting signal (PTS). An analysis of the regions necessary for correct import of four other peroxisomal proteins revealed that they all contained the carboxy terminal tripeptide Ser-Lys/His-Leu (Gould *et al.*, 1988). The terminal tripeptide of firefly luciferase, Ser-Lys-Leu, when fused to the cytosolic protein, chloramphenicol acetyl transferase (CAT), was shown to be sufficient to direct this protein to the peroxisomes (Gould *et al.*, 1989). In addition, only a limited number of conservative changes can be made in this tripeptide without it ceasing to act as a targeting signal (Gould *et al.*, 1989,

Miura *et al.*, 1992). The requirement for a carboxy terminal PTS appears to be an amino acid with a small R group at the first position, a basic one at the second and a large non-polar residue at the third (De Hoop & Ab, 1992). By heterologous expression of firefly luciferase in both yeast and plant cells, and of a yeast peroxisomal protein in mammalian cells, it was demonstrated that this terminal tripeptide was conserved as a peroxisomal targeting motif between yeast, plants, insects and mammals (Gould *et al.*, 1990). Furthermore, the use of an antibody recognising the Ser-Lys-Leu motif demonstrated the presence of proteins containing this motif in the peroxisomes of fungi, plants and animals (Keller *et al.*, 1991). The first demonstration that a plant peroxisomal protein is indeed directed by a carboxy terminal tripeptide was by Volokita (1991). Linking the last six amino acids of glycolate oxidase, a leaf peroxisomal protein, to the β -glucuronidase (GUS) reporter protein, caused GUS activity to be detected in the peroxisomes of transgenic tobacco.

In humans, a lethal condition known as Zellweger syndrome results from the inability of certain peroxisomal proteins to be imported, those containing a class I PTS (Walton *et al.*, 1992). The *pas8* mutant of the yeast, *Pichia pastoris*, exhibits this same import deficiency. PAS8 is a peroxisomal membrane protein which, *in vitro*, is capable of binding the PTS1 targeting signal specifically, raising the possibility that PAS8 is a PTS1 receptor (McCollum *et al.*, 1993).

1.2.4.2 The amino terminal signal

Not all peroxisomal proteins contain a class I PTS. Some are synthesised with an amino terminal presequence which is cleaved upon import. This is also the case for mitochondrial and chloroplast proteins where cleavage of the presequence necessary for the targeting of the proteins to those organelles occurs upon import (for review, see Verner & Schatz, 1988). Rat 3-ketoacyl-CoA thiolase, a peroxisomal protein, has been shown to be synthesised with a presequence which is cleaved on import to generate the mature protein. The cleaved presequence has additionally been shown to direct targeting and import of a non-peroxisomal protein, bacterial CAT, to the peroxisomes (Swinkels *et al.*, 1991). A plant peroxisomal thiolase is also made as a precursor, though the importance of the cleaved amino terminus as a targeting signal has not been demonstrated (Preisig-Muller & Kindl, 1993). Watermelon glyoxysomal malate dehydrogenase (gMDH) is also seen to be synthesised in a precursor form which is cleaved to the mature form upon import (Gietl & Hock, 1984; Gietl, 1990). Like plant thiolase however, to date the importance of this sequence in targeting has not actually been demonstrated, merely the fact that it is cleaved upon import. When expressed in

the yeast *H. polymorpha*, watermelon gMDH is targeted and imported to the peroxisomes. However, the apparent molecular mass is equal to that of the precursor form, indicating that in this yeast, processing of the transit peptide does not take place, though the enzyme still displays activity (van der Klei *et al.*, 1993).

1.2.4.3 Internal targeting signals

Some peroxisomal proteins appear to have different mechanisms of targeting, both between organisms and in some cases, in the same one. Acyl CoA oxidase is such a protein. In rat liver, the PTS resides at the carboxy terminus (Miyazawa *et al.*, 1989) and is of the class I type. In the yeast *Candida tropicalis* the carboxy terminus is not Ser-Lys-Leu and does not appear to be important for targeting. Instead there are two areas able to direct targeting independently. The amino terminal residues 1-118 and an internal region from residues 309-427 both contain information that specifically targeted fragments to peroxisomes. When either of these fragments was individually fused to the cytosolic enzyme dihydrofolate reductase (DHFR), this was then also directed to peroxisomes (Small *et al.*, 1988).

1.2.5 Biochemistry of peroxisomes

Peroxisomes within higher plant cells can be classified according to their physiological role. At present, at least three classes are defined, including the leaf peroxisomes found in photosynthetic tissues, those involved in ureide metabolism in the uninfected root nodule cells of legumes, and glyoxysomes found in the fat storing cells of oilseeds.

1.2.5.1 Leaf-type peroxisomes

In photosynthetic tissues, the enzyme ribulose 1,5-*bis*phosphate carboxylase oxygenase (RuBisCO) catalyses both the carboxylation and oxygenation of ribulose 1,5-*bis*phosphate (RuBP). Carboxylation of one molecule of RuBP leads to the formation of two molecules of glycerate 3-phosphate and the continuation of the Calvin cycle of photosynthesis. However, in high oxygen concentrations, oxygenation of RuBP to form glycerate 3-phosphate and glycolate 2-phosphate occurs. Glycolate 2-phosphate cannot be utilised in the Calvin cycle and therefore, if not further metabolised, would represent a loss of assimilated carbon. This process of oxygen incorporation is known as photorespiration and may reduce the rate of carbon dioxide fixation by 10 to 50 % in some plant species (Huang *et al.*, 1983). However, some of

the carbon may be recycled via the glycolate pathway. The leaf-type peroxisomes house many of the enzymes of this pathway, glycolate oxidase (GO), hydroxypyruvate reductase (HPR), serine:glyoxylate aminotransferase (SGAT) and glutamate:glyoxylate aminotransferase. Glycolate is transported from the chloroplasts to the peroxisomes where it is first oxidised to glyoxylate by GO and subsequently transaminated by SGAT or glutamate:glyoxylate aminotransferase to form glycine. Glycine is transported to the mitochondria where the net conversion of two glycine molecules to one serine molecule and one molecule of carbon dioxide takes place. Serine is returned to the peroxisomes where SGAT catalyses the transamination to hydroxypyruvate which is reduced by HPR to form glycerate. This can then be transported to the chloroplast and used in the generation of sugars. This pathway recovers three of the four carbons from two molecules of glycolate 2-phosphate, the fourth being lost as carbon dioxide (figure 1.1; Huang *et al.*, 1983; Ogren, 1984; Tolbert *et al.*, 1981). The peroxisomal enzymes involved in the photorespiratory pathway increase from undetectable levels in the heterotrophic cotyledons of epigeous species to become the major components of the peroxisomes as the cotyledons green and gain photosynthetic capacity. Accumulation of these enzymes is regulated primarily at the transcriptional level and is light dependent (Hondred *et al.*, 1987; Greenler *et al.*, 1989).

1.2.5.2 Peroxisomes in root nodules

In certain legumes, such as soybean and cowpea, a symbiosis with nitrogen fixing bacteria of the *Rhizobium* species occurs in the roots, leading to the formation of root nodules. Nodules are composed of two cell types, infected and uninfected, in equal proportions. Infected cells contain large numbers of bacteroids which are responsible for fixing nitrogen. Uninfected cells lie adjacent to the infected ones and contain many peroxisomes in which the fixed nitrogen is combined with carbon to form the ureides, allantoin and allantoic acid, the major transported nitrogen containing compounds, having a higher nitrogen to carbon ratio than asparagine or glutamine (Huang *et al.*, 1983). The proliferation of the peroxisomes in the uninfected nodule cells is associated with the synthesis of uricase, a hydrogen peroxide producing enzyme involved in purine catabolism (Nguyen *et al.*, 1985). In addition, two other enzymes of this pathway, xanthine dehydrogenase and allantoinase activities are present in nodule extracts (Rawsthorne *et al.*, 1980), but are cytosolic or microsomal respectively rather than peroxisomal (Hanks *et al.*, 1981).

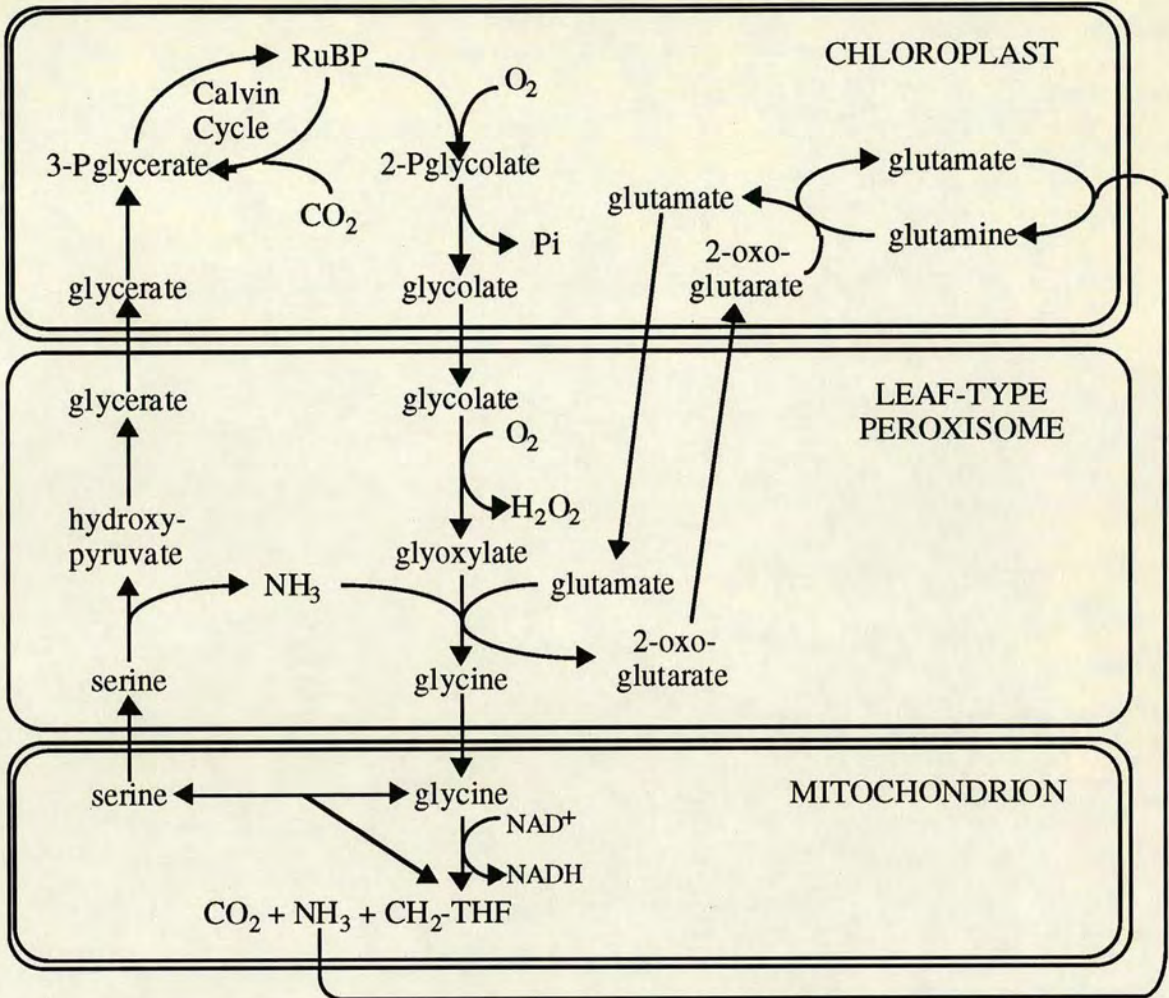


Figure 1.1 The photorespiratory pathway and its subcellular location.

3-Pglycerate = 3-phosphoglycerate; 2-Pglycerate = 2-phosphoglycerate; RuBP = ribulose-1,5-bisphosphate; THF = tetrahydrofolic acid.

1.2.5.3 Glyoxysomes

Glyoxysomes are the class of peroxisomes typically found in the cotyledons and/or endosperm of oilseeds and the scutellum of cereals during postgerminative growth. In addition, they are present in these species during embryogenesis. They are involved in the conversion of fatty acids stored as triacylglycerols to carbohydrate. The triacylglycerols are mobilised from the lipid bodies by lipases and are transported to the glyoxysomes. Within the glyoxysomes, fatty acids are broken down to produce acetyl-Coenzyme A (acetyl-CoA) by the enzymes of β -oxidation. This pathway was first described in glyoxysomes from germinating castor bean endosperm (Cooper & Beevers, 1969a,b). It consists of four steps: (i) activation of the fatty acid to its acyl-CoA derivative by the action of fatty acyl-CoA synthase takes place on the outer face of the glyoxysomes; (ii) following transport across the glyoxysomal membrane, oxidation by the fatty acyl-CoA oxidase generates the enoyl-CoA derivative and hydrogen peroxide, which is then degraded by catalase within the organelle; (iii) conversion of the enoyl-CoA derivative to the 3-keto ester is catalysed by a multifunctional enzyme which shows enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase and 3-hydroxyacyl-CoA epimerase activities (Behrends *et al.*, 1988) and (iv) finally thiolase catalyses the removal of acetyl-CoA from the 3-keto ester by reaction with CoA. The result of these reactions is the formation of a fatty acyl-CoA molecule which is two carbons shorter than the original molecule (Van den Bosch *et al.*, 1992; Kindl, 1993). In animal cells, β -oxidation occurs predominantly in the mitochondria where it contains no multifunctional complex, instead this reaction is catalysed by separate enzymes. Some investigators report β -oxidation activity in the mitochondria of non-fatty plant tissues (Thomas & Wood, 1986; Masterton *et al.*, 1990; Dieuaide *et al.*, 1993), in contrast to previous reports that all activity was associated with the peroxisomes (Gerhardt, 1983).

Acetyl-CoA enters the glyoxylate cycle which serves to condense two molecules of acetyl-CoA to one molecule of succinate, which is then transported to the mitochondrion. The glyoxylate cycle was first described in higher plants in the endosperm of castor bean seedlings (Kornberg & Beevers, 1957). The cycle is catalysed by five enzymes, isocitrate lyase (ICL) and malate synthase (MS), both of which are unique to the cycle, and malate dehydrogenase, citrate synthase and aconitase, which are also part of the TCA cycle. All five enzymes were thought to be present in the glyoxysomes of fatty tissues. However, two recent reports indicate that aconitase is absent from glyoxysomes (Courtois-Verniquet & Douce, 1993; De Bellis *et al.*, 1994).

The mechanism for the reoxidation of the NADH produced by β -oxidation remains controversial. The presence of an electron transport chain within the glyoxysomal membrane which would allow the reoxidation of NADH has been reported (Fang *et al.*, 1987). However, current support lies with the proposal of Mettler & Beevers (1980) that shuttling of reducing equivalents between the glyoxysome and mitochondrion could lead to *in situ* oxidation of NADH. Consequently, the cycle depicted in figure 1.2 is based on this proposal.

1.3 Synthesis of ICL and MS during plant development

The enzymes ICL and MS are unique to the glyoxylate cycle and hence have been studied extensively as marker enzymes for this cycle during plant development. The enzyme activities have been detected in higher plants during embryogenesis and germination of many species (Allen *et al.*, 1988; Comai *et al.*, 1989; Turley & Trelease, 1990; Weir *et al.*, 1980; Zhang *et al.*, 1993), senescence (Graham *et al.*, 1992; Pistelli *et al.*, 1991; De Bellis *et al.*, 1991) and pollen formation (Zhang *et al.*, 1994). In addition, they have also been detected in many microorganisms (Kornberg, 1966; Fernandez *et al.*, 1992; Thurston, 1977) and in some vertebrate tissues (Davis *et al.*, 1990; Davis & Goodman, 1992).

In cucumber, MS is composed of eight identical subunits (Koller & Kindl, 1977) and the single gene has been cloned and sequenced (Graham *et al.*, 1989). ICL is a tetramer, with two subunits of 63 and 61.5 kDa isolated from cucumber (Weir *et al.*, 1980; Reizman *et al.*, 1980). The regulation of both of these enzymes is thought to be primarily at the level of transcription (Comai *et al.*, 1989; Allen *et al.*, 1988) and the levels of enzyme activity reflect changes in levels of mRNA (Smith & Leaver, 1986). Approximately 1 kb of the 5' region of the cucumber malate synthase gene has been shown to direct expression of a reporter gene in transgenic plants in the same temporal and spatial pattern as the gene is regulated in cucumber (Graham *et al.*, 1990). However, regulation may also occur post-transcriptionally as the ratio of MS protein to mRNA is 14-fold higher than that for ICL, a difference which does not seem to be accounted for by differences in the translational efficiencies (Ettinger & Harada, 1990). There have been reports that phosphorylation is important in the regulation of ICL activity in *Escherichia coli* and *Saccharomyces cerevisiae*. In *E. coli*, the enzyme is active in the phosphorylated form, but in *S. cerevisiae*, it is the dephosphorylated form which is active, the enzyme being phosphorylated and therefore inactivated on the addition of glucose to the medium (Hoyt & Reeves, 1988; Robertson & Reeves, 1989; Lopez-Boado *et al.*, 1988). A recent report indicates that ICL of castor bean and

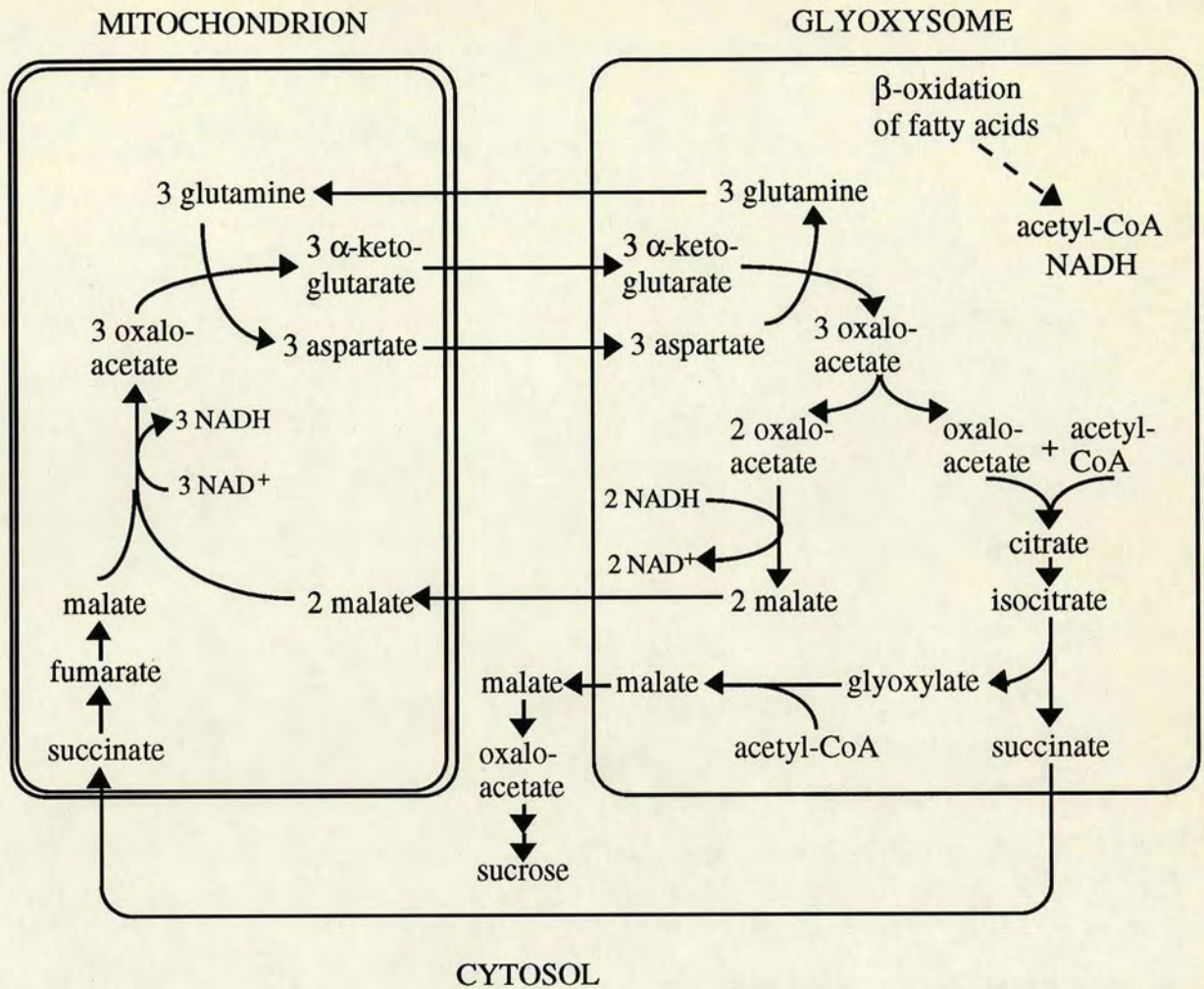


Figure 1.2 The conversion of fatty acids to sucrose, via β -oxidation and the glyoxylate cycle. The malate-aspartate shuttle of reducing equivalents provides for *in situ* oxidation of NADH produced by the oxidation of fatty acids to acetyl-CoA. Based on Mettler & Beevers, 1980.

cucumber is phosphorylated by a glyoxysomal protein kinase. However, the effect of phosphorylation on plant ICL activity has not yet been determined (Finnessy *et al.*, 1994).

1.3.1 Synthesis of ICL and MS during embryogenesis and seed maturation

Glyoxysomal enzyme activities can be detected in developing cotton seeds as early as 17 days post anthesis (DPA; Turley & Trelease, 1990). At this stage, ICL and MS activities occur in a coordinate manner. However, the pattern of expression becomes noncoordinate at later stages, MS exhibiting a significant increase 45 DPA whereas ICL activity does not show a major increase until 12 h post imbibition (Choinski & Trelease, 1978; Miernyk & Trelease, 1981). Other glyoxylate cycle and β -oxidation enzymes all show coordinate development from 22 DPA (Bortman *et al.*, 1981; Choinski & Trelease, 1978; Miernyk & Trelease, 1981). Increases in mRNA levels precede corresponding increases in protein and enzyme activity by approximately 24 h for MS and ICL, indicating that regulation is primarily at the transcriptional level (Turley & Trelease, 1990). ICL and MS mRNAs in *Brassica napus* were shown to be coordinately synthesised in developing seeds from 37 DPA (Comai *et al.* 1989). ICL mRNA was first detected in developing sunflower seeds 19 days after flowering, and levels remained constant through dessication, but western blotting with an antibody to cotton ICL failed to detect protein (Allen *et al.*, 1988). Thus, species specific patterns of expression of MS and ICL appear to exist during embryogenesis.

1.3.2 Synthesis of ICL and MS during postgerminative growth

The existence of the glyoxylate cycle in higher plants was first demonstrated in castor bean endosperm during postgerminative growth (Kornberg & Beevers, 1957). Subsequently, much research to investigate the regulation of the synthesis of ICL and MS during this period has been carried out, both in this species (Martin *et al.*, 1984; Rodriguez *et al.*, 1987) and in the cotyledons of other epigeous oilseeds such as cucumber (Becker *et al.*, 1978; Reizman *et al.*, 1980; Weir *et al.*, 1980), sunflower (Allen *et al.*, 1988), oilseed rape (Comai *et al.*, 1989) and cotton (Turley & Trelease, 1990).

During and post germination in cucumber cotyledons, ICL and MS activities increase from undetectable levels in dry seeds, peaking 3 to 4 days after seed imbibition and thereafter declining (Becker *et al.*, 1978). Monospecific antibodies have been used to

detect these two proteins amongst the products of *in vitro* translation programmed with cotyledonary RNA from cucumber (Lamb *et al.*, 1978; Reizman *et al.*, 1980). Immunoprecipitation of ICL and MS from *in vitro* translation products of cotyledonary RNA isolated during the first seven days of postgerminative growth, indicated that the mRNA levels for each of these enzymes rise and fall in a similar pattern to enzyme activities, preceding them by about one day (Weir *et al.*, 1980). Northern and western blot analysis of MS transcripts and protein, and assays of enzymes activity demonstrated that the regulation of MS synthesis during postgerminative growth is brought about primarily through changes in the levels of transcripts rather than through control of translation (Smith & Leaver, 1986). The rate of decline in MS transcripts, protein and activity was shown to be significantly greater when the seedlings were incubated in the light than in the dark (Smith & Leaver, 1986). Analysis of a transcriptional fusion construct of approximately 1 kb of the *ms* 5' region with the *E. coli* β -glucuronidase (GUS) gene expressed in transgenic *Nicotiana* and *Petunia* seedlings exhibited the same pattern of expression to that of the *ms* gene in cucumber, indicating that transcription is the principal controlling step in the synthesis of MS (Graham *et al.*, 1990).

Similar results have been obtained in other species. In sunflower, the rate of accumulation of ICL mRNA, as well as that of decline, was greater in light grown seedlings than in those incubated in the dark during the first 5 days growth after imbibition. ICL protein was first detectable 24 h after the peak in ICL mRNA levels (Allen *et al.*, 1988), so confirming the observations made in cucumber. Levels of ICL and MS mRNA in *B. napus* (Comai *et al.*, 1989) and mRNA, protein and activity levels in cotton (Turley & Trelease, 1990) also confirm these observations. However, the possibility that post-translational processes may also be important in the regulation of these enzymes should not be forgotten.

1.3.3 Synthesis of ICL and MS during senescence

The presence of the glyoxylate cycle has been suggested in senescent tissue. ICL and MS activities were virtually undetectable in mature barley leaves, but were detected in leaves which had been excised and incubated in the dark (Gut & Matile, 1988). These enzyme activities were coupled with a decline in monogalactosyl diacylglycerol, a thylakoid lipid, suggesting that the glyoxylate cycle may be involved in the degradation of structural lipids. Subsequently, activities of these enzymes have been detected in a range of naturally and induced senescing organs of both mono- and dicotyledonous plants, including petals (De Bellis *et al.*, 1991) and leaves (Pistelli *et al.*, 1991).

Cucumber MS mRNA has been detected in naturally senescing cotyledons, leaves and petals and in detached, dark incubated cotyledons, leaves and roots (Graham *et al.*, 1992; McLaughlin & Smith, 1994). Confirming these observations, ICL and MS activities have also been detected in detached, dark incubated leaves of barley, rice and leaf beet and cotyledons of pumpkin, in which the activities were localised to the peroxisomes (De Bellis *et al.*, 1990; De Bellis & Nishimura, 1991). Thus, a second transition of peroxisomal function takes place in epigeous species, from leaf peroxisome to glyoxysome-like function, which like the first transition, has been demonstrated by double labelling experiments (Nishimura *et al.*, 1993). Transfer of whole plants from light to dark incubation also leads to an induction of MS and ICL in the photosynthetic organs (Birkhan & Kindl, 1990; Graham *et al.*, 1992; Pistelli *et al.*, 1991). A correlation between the physiological changes associated with senescence and an increase in ICL activity, used as a marker of the glyoxylate cycle, has now been demonstrated (Vicentini & Matile, 1993). These results have led to the proposal that the synthesis of glyoxylate cycle enzymes is controlled by the levels of various metabolites, being repressed by high levels of sucrose and induced following the fall of sucrose levels experienced in induced or naturally senescent tissue (Graham *et al.*, 1992). This proposal has now been substantiated by studies with dark incubated cultured cells and protoplasts (Graham *et al.*, 1994; McLaughlin & Smith, 1994).

1.4 Metabolic Regulation of gene expression

1.4.1 Carbon catabolite repression

It has been known for some years that, in microorganisms, the levels of glucose, or other rapidly mobilisable carbon source, repress the expression of genes encoding enzymes involved in the metabolism of other carbon sources, such as the *SUC*, *MAL* and *GAL* genes, β -oxidation and glyoxylate cycle genes. This phenomenon is known as carbon catabolite repression, and allows the organism to switch rapidly from metabolism of one carbon source to another, as the metabolic environment changes.

Much work has been done to elucidate the mechanism of carbon catabolite repression in yeast. Although many genes which are important have been identified through the use of genetics, the pathway from the initial signal to the final target gene is as yet unclear. The degree of repression varies between genes, at approximately 800-fold for invertase (*SUC2*), 80-fold for MS, 50-fold for ICL to less than 10-fold for aconitase and isocitrate dehydrogenase. The repression of the genes is thought to take place at the

level of transcription or mRNA stability, as the change in enzyme levels is paralleled by a decrease in mRNA (Gancedo, 1992).

Repression is initiated in response to a signal generated when glucose or related sugars are present in the medium. Regulatory proteins which respond to the signal may either act directly on transcription of the regulated gene, or may act on additional regulatory genes or gene products, initiating a regulatory cascade. Repression may be effected either by activation of a repressing protein or by inhibition of an activating protein (see section 1.5.5).

In yeast, catabolite repression may be exerted by glucose, mannose and fructose, and in addition by galactose and maltose. Some genes are repressed by all these sugars, others only by glucose. In order to transmit the 'signal', metabolism through a kinase is necessary, to phosphorylate the sugar (Witt *et al.*, 1966), indicating a possible role for glucose-6-phosphate in the signal transduction pathway.

Many mutants have been isolated which are unable to respond to carbon source changes, for example, in the genes *CAT1,3* and *MSN1*. *CAT1* has protein kinase activity essential for its function and is affected positively by *CAT3*, but in an unknown manner. *MSN1* acts as a transcription activator, but does not itself exhibit DNA binding activity. A second set of mutants unresponsive to repression have been isolated, including those in the genes *HXK2* and *MIG1*. *HXK2* is thought to encode the hexokinase isoenzyme PII (Ma *et al.*, 1989a,b), again implicating hexose-6-phosphates in the control of repression. *MIG1* encodes a DNA binding protein with two Cys₂His₂ zinc finger motifs, and has been shown to bind to the promoters of several genes repressed by glucose. Two sequences in the *SUC2* promoter have been identified which are able to bind *MIG1*, which also binds in the *GALI/GAL10* intergenic region (Nehlin *et al.*, 1991). In addition, similar recognition sequences have been discovered in the *MAL6* promoter. Although some elements have been elucidated, the function of many of the other factors involved is not yet known. The mechanism of regulation varies for different genes, so adding further complexity to the system.

Many genes in other organisms have also been shown to be controlled by catabolite repression. In *Drosophila* larvae, the α -amylase gene is regulated by the levels of glucose that are fed to the larva. Levels of mRNA detected show an approximately 100-fold reduction in larvae fed on 10 % glucose compared to those fed on a diet lacking glucose (Benkel & Hickey, 1987). Chimeric gene studies, using a promoter fragment of the α -amylase gene fused to the alcohol dehydrogenase transcribed region, showed that regulation by glucose is at the level of transcription (Grunder *et al.*, 1993).

In plants, several genes have been reported to be repressed by the presence of glucose or sucrose. In maize, seven genes encoding enzymes involved in photosynthesis are

repressed (Sheen 1990), potentially by metabolic factors related to high carbohydrate content (Krapp *et al.*, 1993). No consensus sequence has been demonstrated to be responsible for the control of transcription. Instead, distinct upstream activating elements mediate repression, rather than basal promoters or negative elements (Sheen, 1990).

1.4.2 Induction of plant genes by sucrose

In addition to those plant genes repressed by sucrose, many are induced by elevated levels of this metabolite. Transcription from the patatin promoter linked to GUS in transgenic potato plants was induced in leaves and stems when cultured on medium containing 0.3 to 0.4 M sucrose. Leaves and stems do not normally express patatin, which is specific to tubers and stolons attached to developing tubers (Wenzler *et al.*, 1989). Sporamin is a tuberous root storage protein of the sweet potato, usually undetectable in other organs. When the stem is grown on sucrose containing medium, it is found to accumulate sporamin (Hattori *et al.*, 1990), and a region of the promoter responding to sucrose induction has been identified. This sequence has also been found in the promoter of the chalcone synthase gene from petunia which is also induced in the presence of 0.3 M sucrose (Tsukaya *et al.*, 1991). The level of potato sucrose synthase mRNA increases in the leaves and petioles when the sucrose concentration is increased (Salanoubat & Belliard, 1989). The two sucrose synthase genes of maize, *Sh1* and *Sus1*, are differentially regulated by sucrose, *Sus1* induced and *Sh1* repressed when excised root tips were incubated in the presence of sucrose (Koch *et al.*, 1992).

1.4.3 Metabolic regulation of the *icl* and *ms* genes

In various organisms, both MS and ICL are regulated by the levels of metabolites. In *E. coli*, the genes encoding these enzymes are located on the *ace* operon, and are required for growth of the organism on acetate or fatty acids as the sole carbon source (Maloy & Nunn, 1982). Expression of this operon is controlled at the level of transcription by the products of the *iclR* and *fadR* genes, the latter also being involved in the regulation of the fatty acid degradation regulon. The repressor, IclR, recognises a 35 bp palindromic sequence which overlaps the -35 recognition site of RNA polymerase. This interaction is impaired by phosphoenolpyruvate, but is insensitive to acetate, acetyl CoA and pyruvate, suggesting that acetate does not act directly (Cortay *et al.*, 1991). Both of these proteins are necessary for full repression of the *ace* operon, independently only exerting a slight degree of repression (Maloy & Nunn, 1982). It is

not known whether the *fadR* gene product exerts its repressive effect by DNA binding to *cis*-acting sequences, or by interacting with IclR (Nunn, 1987).

In *Aspergillus nidulans*, mutants exist, mapped to the *icl* and *ms* genes, which are unable to utilise acetate as the sole carbon source (Armitt *et al.*, 1976). ICL expression is induced by acetate (Armitt *et al.*, 1976) and is also subject to carbon catabolite repression (Kelly & Hynes, 1977). The *facB* gene, the product of which is necessary for acetate dependent induction of ICL and MS, is a *trans*-acting regulatory protein (Armitt *et al.*, 1976). A second gene product, CREA, is also a DNA binding protein, mutations in which lead to derepression of ICL and other enzymes regulated by carbon catabolite repression (Bowyer *et al.*, 1994). A 208 bp fragment of the *icl* promoter has been identified which carries sites recognised by both CREA and the *facB* gene product (De Lucas *et al.*, 1994). As discussed above, the *icl* and *ms* genes of *S. cerevisiae* have been shown to be subject to carbon catabolite repression, with regulation thought to be primarily at the level of transcription (Schöler & Schüller, 1993). The glyoxylate cycle enzymes are essential for growth on C₂ or C₃ substances, the genes being controlled by positively acting derepression genes such as *CAT1* and 3. A 10 bp carbon source-responsive element (CSRE) has been identified in the yeast *icl* promoter which is both necessary and sufficient for carbon source-dependent control of the gene (Schöler & Schüller, 1994). Sequence motifs similar to this CSRE were also found in the upstream regions of other genes involved in gluconeogenesis, including that of the *ms* gene, but this has not yet been shown to have biological significance. The position of the CSRE within the upstream region was shown to be important. When moved 140 bp upstream of its natural position, less than 2 % of the derepressed wild type activity was obtained (Schöler & Schüller, 1994).

In *Euglena*, both ethanol and acetate induce the transcription of the *ms* gene. This induction is greater when the cells are incubated in darkness than in the light and MS mRNA was absent from phototrophic cultures (Woodcock & Merret, 1980; Harrum & Schwartzbach, 1981). Similarly in *Chlorella fusca*, ICL synthesis is subject to catabolite repression, synthesis being repressed if the cells were provided with glucose or allowed to photosynthesize, even in the presence of acetate, which stimulates synthesis of ICL when the cells are incubated in darkness (McCullough & John, 1972). Regulation of the synthesis appears to be primarily at the transcriptional level in both these organisms.

The regulation of *icl* and *ms* expression by metabolites has been demonstrated in suspension cultures of anise. When the cells are grown in medium lacking any source of carbon, ICL and MS activities are detectable. The derepression of these enzymes is

increased when acetate is added to the medium, but when sucrose is added, the activities are no longer detectable (Kudielka & Theimer, 1983a,b).

More recently, the effect of metabolites on synthesis of ICL and MS has been studied in cucumber protoplasts and suspension cultures. ICL and MS proteins are detectable in protoplasts after incubation for 48 hours in the dark in the absence of sucrose. When incubated in the presence of 25 mM sucrose, proteins are not detectable (McLaughlin & Smith, 1994). Starvation of cucumber cell cultures led to induction of transcription of *icl* and *ms*, correlated with a drop in intracellular sucrose, glucose and fructose. Addition to the medium of glucose, fructose and raffinose, one of the major translocated sugars in cucurbits, resulted in repression of *icl* and *ms* transcription. Glucose analogues able to be phosphorylated all resulted in specific repression, but 3-methyl glucose, although taken up by the cells, was unable to effect repression. 3-methyl glucose is not phosphorylatable, leading to the proposal that the signal important for regulation of the *icl* and *ms* genes originates from the flux of hexose sugars into glycolysis (Graham *et al.*, 1994).

1.5 Control of gene expression

1.5.1 Regulation of transcription in prokaryotes

The understanding of regulation of transcription in prokaryotic systems has been achieved by extensive analysis of systems such as bacteriophage λ (Ptashne, 1992). Prokaryotic genes are transcribed by RNA polymerase (pol), which can be regulated by various control proteins. Transcription initiation involves binding of RNA pol to the promoter, followed by isomerisation from the closed to the open form of the complex, resulting in the unwinding of the DNA helix near the transcription start site. Formation of the first few phosphodiester bonds of the RNA transcript leads to release of the polymerase from the promoter (promoter escape). The levels of transcription from any promoter are controlled by DNA binding proteins, serving to enhance or repress the basal levels of transcription attained in the absence of any additional factors. Activation of transcription is usually achieved by a *trans*-activating factor assisting binding of the polymerase to the promoter, as seen with λ repressor (Ptashne, 1992). Transcriptional repressors have been demonstrated to act at each stage of transcriptional initiation. In addition to assisting the polymerase in the transcription of its own gene, the λ repressor also inhibits transcription of *cro* by blocking binding of the polymerase to the promoter (Hawley *et al.*, 1985). The *arc* repressor of bacteriophage P22 blocks the transition from the closed to the open form of the polymerase (Vershon *et al.*, 1987). In *E. coli*,

the formation of the first phosphodiester bonds is blocked by the Gal repressor (Choy & Adhya, 1992) and promoter escape by the Lac repressor at the *lacUV5* promoter (Lee & Goldfarb, 1991).

1.5.2 Transcriptional regulation in eukaryotes

Eukaryotic genes generally consist of the coding sequence flanked by a 5' and 3' non-coding region. The coding sequence, unlike those of prokaryotes, is usually interrupted by introns, though some genes, for example, the maize zein genes, do not contain intervening sequences (Heidecker & Messing, 1986). In the majority of cases, it is the 5' region which contains the sequences necessary for the control of transcription. This region is composed of a number of elements. The promoter carries the TATA box approximately 30 bp upstream of the start of transcription (Joshi, 1987a) and additional orientation specific sequences capable of binding protein factors. Enhancers and silencers (upstream regulatory elements, UREs) occur upstream of the promoter and may exist in either orientation. The UREs serve to regulate the level of transcription from a certain promoter. The 5' region may contain several UREs, which may act independently or in a synergistic manner, differentially regulating expression in different tissues and at different developmental stages. The 3' region is transcribed but not translated and contains sequences which direct processing and polyadenylation of the 3' end of the messenger RNA. These sequences are not highly conserved between plants and animals. In vertebrates, the sequence AAUAAA is essential for correct processing, but often is not recognised in plant systems (Joshi, 1987b; Hunt *et al.*, 1987).

DNA exists within the nucleus as a highly organised structure, packaged first around a histone octamer and then wound into a condensed 30 nm fibre. The degree of packaging affects the availability of a gene for transcription; if highly packaged, the necessary transcription factors will be unable to gain access to the sequences to which they must bind in order to effect activation of transcription. Specific chromosomal regions are normally assembled into specific structural domains. Therefore genes can exist in at least two transcriptional states. Those present in active, or 'euchromatic' regions are accessible to both activational and repressional regulation afforded by the particular set of DNA sequence elements and proteins that comprise that individual gene's regulatory machinery. However, those genes in inactive or 'heterochromatic' domains are not accessible to the regulation normally mediated by their 5' regions (Rivier & Pillus, 1994).

1.5.3 Assembly of the preinitiation complex

Initiation of transcription in eukaryotes is more complex than in prokaryotes, requiring a large number of general transcription factors in addition to RNA polymerase. In eukaryotes, there are three RNA polymerases, RNA polymerase I, II and III. Here only transcription by RNA polymerase II (pol II) will be discussed, which is responsible for transcription of the protein encoding genes. The preinitiation complex is composed of pol II together with several additional polypeptide chains which comprise the general transcription factors (TFs) IIA, B, D, E, F, H and J (for review, Drapkin *et al.*, 1993). Formation of the preinitiation complex starts with the binding of TFIID to the TATA box of the promoter. TFIID is the only factor of the preinitiation complex which has been shown to have DNA binding activity. It is itself composed of several polypeptide chains, the TATA binding protein (TBP) which contacts DNA in the minor groove (see section 1.5.4.1) and TBP associated factors (TAFs). The carboxy terminal DNA binding domain of TBP is highly conserved throughout all species, but the remainder of the protein shows extensive sequence divergence. Some plants contain two genes for TBP implying either a specificity of function at certain pol II promoters or the possibility of one TBP serving pol I and/or pol III transcription exclusively (Gasch *et al.*, 1990; Maass & Feix, 1992). TAFs are thought to contribute to the specificity of activation and may serve as the actual target for transcriptional activators. TFIIA binds to the TBP so stabilising the protein/DNA complex, followed by TFIIB to form the DAB complex. Pol IIa is assisted to the DAB complex by TFIIF, after which binding of TFIIIE, H and J occurs, so forming the assembled preinitiation complex. Pol II exists in two forms, unphosphorylated (IIa) and phosphorylated (IIo), only binding to the DAB complex in the unphosphorylated form. TFIIH shows both DNA helicase (Schaeffer *et al.*, 1993) and protein kinase (Lu *et al.*, 1992) activities. Negative supercoiling of the promoter region has been shown to stimulate transcription (for review, Stanway, 1993) and phosphorylation of pol IIa at the seven residue repeats present at the carboxy terminus is necessary for promoter escape; the phosphorylated form, pol IIo does not associate with TBP. The kinase activity of TFIIH is stimulated by TFIIIE and by DNA containing a TATA box and transcription start site. TFIIF and J are required for elongation in addition to initiation. This preinitiation complex can give rise to a basal level of transcription even in the absence of additional activating factors. Not all promoters require the full set of factors for transcription initiation. When negatively supercoiled, the immunoglobulin heavy chain (IgH) gene promoter can be transcribed in the presence of only TBP, TFIIB and pol II (Parvin & Sharp, 1993). If the DNA was relaxed or linearised, this same promoter requires that all the GTFs are

present. However, the exact requirement for GTFs is specific to the promoter: the adenovirus major late promoter required all the factors, irrespective of the state of the DNA. It should be noted that for all these experiments purified TBP was used rather than TFIID (TBP plus its associated TAFs). *In vivo*, it is possible that the presence of TAFs could expand the requirement for the other GTFs.

1.5.4 Activation of the preinitiation complex

The mechanism of activation of transcription is not understood in detail. Sequence-specific transcriptional activators bind to DNA at the promoter or at enhancers, DNA sequences often present distant to the promoter on which they act, either upstream, downstream or even within the coding sequence (Dietrich *et al.*, 1992; Douglas *et al.*, 1991). As described in section 1.5.1, activation of transcription in prokaryotes does not require the action of proteins bound distant to the promoter, those bound at the operator sites are sufficient to regulate transcription. Although the mechanism is unclear, *trans*-activators presumably act either by stimulating the rate of assembly of the preinitiation complex or by increasing its stability once formed.

Activators often bind to enhancers as multimeric complexes, providing a greater flexibility of regulation than single factors alone could do. Synergism between factors is often found and has been demonstrated for the cauliflower mosaic virus 35S promoter (Benfey & Chua, 1990). They are usually modular proteins, consisting of a DNA binding domain, an activating domain, and those involved in other protein-protein contacts, for example, oligomerisation. The *trans*-activating factor of the *GAL* genes in yeast, GAL4, gives an excellent example of such a modular protein. GAL4 binds as a dimer to the promoters of the genes involved in galactose utilisation and is necessary for transcription of these genes which are only expressed in the presence of galactose and absence of a preferentially utilised carbon source such as glucose. The domains of GAL4 have been mapped to show that two regions at the amino-terminus are needed for DNA binding, one involved in DNA recognition and the other necessary for dimerisation. There are two transcriptional activation domains, either of which can activate transcription when joined with a DNA binding domain. Activation of transcription by GAL4 is inhibited in the absence of galactose by the binding of a second protein, GAL80; residues residing within one of the activation domains are necessary for this inhibition to take place. Activation may require an 'adaptor' protein in addition to the protein containing a DNA binding domain, the 'adaptor' interacting with the factors in the promoter complex, rather than the DNA bound protein making direct contact itself (Martin, 1991).

1.5.4.1 DNA binding motifs

Trans-activating factors recognise DNA through small discrete domains, which usually fit into one of several classes. The first structure of a DNA binding domain to be elucidated was the helix-turn-helix (HTH) motif, found in the bacteriophage λ repressor and subsequently in many eukaryotic transcription factors which play important roles in development, such as the homeobox genes of *Drosophila* encoding Engrailed and Antennapedia. The homeodomain of these proteins consists of a highly conserved sixty amino acid domain folded into three α -helices, the second and third of which form the HTH motif. Helix 3 is responsible for recognition and sits in the major groove where it makes extensive base specific connections. In addition, an amino-terminal arm lies in the minor groove making further contacts (Wright, 1994).

A variant of the homeodomain is the POU (Pit-1, Oct-1 and -2, Unc-86) domain, a bipartite DNA binding domain consisting of the POU-specific (POU_S) domain joined by a linker of variable length to a homeodomain (POU_H). Both domains are required for high affinity and sequence specific binding but are independently stable. POU_H shows high levels of sequence similarity to the classical homeodomain and is assumed therefore to generate a similar three dimensional structure. POU_S consists of four helices which are packed to enclose a highly conserved hydrophobic core. It bears a structural resemblance to the HTH of λ and 434 repressors, helix 3 being involved in recognition, but with the 'turn' extended. In classical homeodomains, the 'turn' consists of four amino acid residues; in POU_S domains, the turn has an additional six amino acids, extending helix 2 by one helical turn (Wright, 1994).

Proteins which exhibit both zinc binding and nucleic acid binding motifs are known collectively as the zinc finger proteins. This class shows considerable diversity, differing in metal coordination strategy, secondary structure and modularity, all of which lead to distinctly different folds (Schmiedeskamp & Klevit, 1994). The first zinc finger protein to be described was *Xenopus* TFIIIA (Miller *et al.*, 1985), a factor required for initiation of transcription of the 5S ribosomal RNA genes by RNA pol III. The zinc ion is contacted by side chains from two cysteine and two histidine residues, this family is described as the Cys₂His₂ fingers. Each finger consists of two irregular β -sheets and an α -helix that pack together around the zinc ion. Three amino acid residues just before and within the α -helix make base-specific contacts to the DNA. However, the specific contacts made by TFIIIA remain obscure. This protein contains nine fingers, not all of which have been shown to be essential for base-specific binding. In addition to the fingers, the linker sequence between them also plays a role

in determining the DNA binding affinity (Berg, 1993). Some retroviral proteins, for example, the HIV-1 nucleocapsid protein, also contain zinc fingers with a Cys₂HisCys motif. Highly conserved hydrophobic and aromatic amino acid residues in the amino terminal of the protein form a cleft which contacts the single stranded nucleic acid (RNA; South & Summers, 1993). The GAL4 protein of yeast is another zinc finger family member. It binds two zinc ions in a Zn₂Cys₆ binuclear cluster. GAL4 binds to its recognition sequence as a dimer, two residues of the amino-terminal α -helix contacting the outermost three base pairs of the 17 base pair consensus half site (Berg, 1993). Other zinc finger proteins include GATA-1, the erythroid transcription factor (ZnCys₄) and the steroid and related hormone receptors, which have two domains, each binding one zinc ion via four cysteine residues (Schmiedeskamp & Klevit, 1994). Many other DNA binding proteins belong to the bZIP (basic region leucine zipper) or bHLH (basic region helix-loop-helix) classes. The DNA binding domain of these classes is a basic region, consisting of two extended α -helices, rich in basic amino acids. The domain is unstructured in solution but forms α -helices when bound to DNA, residues along the 'inner' surface of the helix contacting the edges of the base pairs in the major groove. In addition, neighbouring residues contact the phosphodiester backbone of the binding site. The lack of tertiary structure may give these proteins greater adaptability and flexibility than other DNA recognising helices and may contribute to the diverse specificities of the bZIP and bHLH family members (Ellenberger, 1994). The carboxy terminal domain contains the leucine zipper or HLH sequences which are responsible for dimerisation. Many DNA binding proteins bind as dimers or higher order oligomers, which can contribute to the specificity of control as homodimers or various heterodimers may be formed able to regulate many genes in a tissue-specific or developmental specific manner. In prokaryotes, some proteins interact through β -ribbon recognition elements, for example, the Arc repressor (Harrison, 1991), but none has been identified in eukaryotes to date.

Of the DNA binding proteins for which cocrystal structures have been obtained, most contact the DNA in the major groove. However, TBP, the DNA binding polypeptide of the TFIID complex is seen to contact principally the minor groove of its site. The crystal structure of TBP was first determined from *Arabidopsis thaliana*, and revealed a new highly symmetrical DNA-binding fold resembling a saddle. The DNA binding surface is a curved five-stranded antiparallel β -sheet. When TBP is bound to the DNA via its concave surface, the convex surface of the non-conserved amino-terminal is present for interactions with TAFs and other regulatory proteins (Nikolov *et al.*, 1992). Binding of this protein to the DNA causes bending and leads to partial unwinding of the

helix and widening of the minor groove (Kim J. *et al.*, 1993; Kim Y. *et al.*, 1993; Burley, 1994).

A group of DNA binding proteins, those containing a high mobility group (HMG) domain, also interact with the minor groove. The HMG domain is shared by abundant non-histone components of chromatin and by specific regulators of transcription. This domain is comprised of three α -helices with aromatic residues that are clustered at the junctions thought to mediate the L shaped arrangement of helices II and III. These two helices have clusters of basic amino acids on their outer surfaces which are thought to mediate DNA binding. In addition, the amino-terminal arm also contains basic residues which are implicated in DNA binding. How contacts are made with the minor groove is not known, as in B form DNA this groove is too narrow to accommodate an α -helix (Groschedl *et al.*, 1994).

The majority of plant transcription regulators analysed so far belong to one of the classes described above, mainly the bZIP class. The conservation of so many domains between the plant and animal kingdoms suggests that DNA binding motifs may have originated before the divergence of these two kingdoms. However, some new classes have been discovered which seem predominantly to contain plant transcription factors. The proteins involved in the control of floral organogenesis often contain a motif described as the MADS (MCM1, AG, DEF, SRF) box, a domain consisting of fifty five highly conserved amino acids (Schwarz-Sommer *et al.*, 1990). These proteins are found to bind as heterodimers, such as DEF A and GLO in *Antirrhinum*, where heterodimerisation is essential for binding (Tröbner *et al.*, 1992). The MADS-box is not found exclusively in plants, but is also present in the serum response factor (SRF) in humans, in CArG binding factors and in some yeast factors (e.g. MCM1; for references, see Schwarz-Sommer *et al.*, 1992).

Other motifs discovered in plants are the tri-helix and double tri-helix found in GT-1 and GT-2, two proteins that regulate the expression of light responsive genes in tobacco (Brunelle & Chua, 1993). This domain consists of three putative α -helices, bridged by a short turn, but showing no homology to the HLH or HTH motifs. However, it is possible that structurally they may serve the same function.

1.5.4.2 Activation domains

Once the complexes have assembled at an enhancer, the activation domain of the *trans*-activating factor(s) must interact, directly or indirectly, with components of the preinitiation complex. Less is known of the requirements of an activation domain than DNA binding domains. The same high level of sequence conservation is not shown,

but they can be subdivided into three groups: the proline rich, the glutamine rich and the acidic. Acidic activating regions are universal and have been shown to function in all eukaryotic cells tested (Ptashne, 1988). The structural relationships and mechanisms of specificity are unclear, not all domains of one class contacting the same target (Gill *et al.*, 1994). The predominant amino acids of a domain are not necessarily those important for activation, bulky hydrophobic residues which probably drive protein folding, generating specificity (Gill *et al.*, 1994). Through NMR studies, some activation domains have been shown to lack secondary structure. Therefore, it seems probable that they may undergo a conformational change on contacting a target giving an induced fit. This avoids the problem of unbound activators interacting with the preinitiation complex, and can achieve greater flexibility and specificity (Tjian & Maniatis, 1994).

The strength of activating regions may vary; multiple weak activators may mimic the effect of a single stronger activating region, although individually they are often unable to exert an influence over great distances. Combining many weaker activators allows a greater degree of flexibility in regulation. Many activators show cooperativity, they have a greater activation effect together than the sum of each independently, though often no direct interactions occur.

1.5.4.3 Topology of transcriptional activation

In prokaryotic systems, in order to differentially regulate genes, different promoters are employed, as exemplified by transcription of the gene encoding λ repressor from either P_R or from P_{RM} . However, in eukaryotes, each gene has one promoter region close to the start of transcription to which the preinitiation complex binds. Differential regulation is achieved by the interaction of the preinitiation complex with specific factors binding at sites other than the region neighbouring the transcriptional start; the intervening DNA can form loops, so allowing the two or more protein complexes to interact. In order for activators to contact their target by looping when the two are only separated by a short distance (up to 70 bp), the interaction is assisted if the two sites are separated by an integral number of helical turns. If the two exist on opposite faces, contact could only be made by DNA twisting, an energetically costly process (Ptashne, 1992).

Enhancers contain distinct sets of binding sites, variations in the arrangement providing the potential to create unique protein complexes resulting in specificity. Many *in vitro* experiments led to the belief that the arrangement of elements within an enhancer was unimportant for regulation. However, the activity of many natural enhancers depends on the three dimensional structure produced and requires architectural components in

the formation of stereospecific complexes. Such a complex is found at the mouse T cell receptor α gene enhancer. A T cell specific factor, lymphoid enhancer binding factor 1 (LEF-1) binds to the minor groove and induces a sharp bend, enabling the factors bound to sites on either side to contact each other. LEF-1 alone cannot activate transcription, the additional two factors are essential, as is the arrangement of their binding sites within the enhancer (Tjian & Maniatis, 1994).

One question is how *trans*-factors recognise which promoter complex to target, and not also work on more distal genes. It is possible that eukaryotic chromosomes are arranged so that the lengths of DNA separating the genes and the regulatory sites ensure that regulators work only on 'nearby' genes. Barriers may exist, such as sites of attachment of the chromosome to cellular structures, dividing the chromosome into domains within which, but not between which, loops between activators (or repressors) and targets may form (Ptashne, 1992).

1.5.5 Repression

In addition to DNA binding proteins exerting an activating influence on the initiator complex, there are many examples where the level of transcription is down regulated. In principal, repression of transcription could be exerted at any stage of preinitiation complex assembly and release. There are therefore several mechanisms by which transcriptional inactivation may be achieved, by interfering with *trans*-activators or by interfering with the assembly of general transcription factors.

Interference with the *trans*-activators can take place in many different ways. In order to act as a DNA binding protein, the activator must first be transported to the nucleus. One class of repressors acts by preventing nuclear localisation. An example of this family is $I_{\kappa B}$, which blocks nuclear import of Rel family members such as $NF_{\kappa B}$. Each of these family members has a three hundred amino acid region at the amino-terminus called the Rel homology, which contains the sequences important for nuclear localisation, DNA binding and oligomerisation. $NF_{\kappa B}$ binds as a heterodimer of two Rel proteins, p50 and p65. $I_{\kappa B}$ associates with the region of the Rel domain important for nuclear import, so masking this sequence and retaining the activators in the cytoplasm (Nolan *et al.*, 1991). $I_{\kappa B}$ proteins are regulated by phosphorylation; different phosphorylation cascades presumably therefore control which $I_{\kappa B}$ molecules are active at any particular time and hence which Rel family members are available for *trans*-activation (Link *et al.*, 1992).

Many *trans*-activators bind to DNA as homo- or heterodimers. Inhibition of dimerisation by competing for association with one of the subunits, so preventing

formation of a functional activator, is another level at which repressors may act. In *Drosophila*, bHLH proteins which dimerise before binding, are important in the development of the peripheral nervous system. The cells giving rise to this system are epithelial and also act as the precursors for epidermal cells. The decision to become sensilla cells or epidermal cells is dependent upon the expression of various bHLH proteins. The genes responsible for sensilla organ formation are *daughterless* (*da*) and three in the *achaete-scute* complex (*AS-C*), all of which encode bHLH proteins which heterodimerise to activate transcription. In the epidermal precursors, two genes are expressed, *extramachrochaetae* (*emc*) and *hairy* (*h*), which suppress sensory organ development. *emc* contains an HLH motif for dimerisation, but lacks the basic region which is essential for DNA binding (see section 1.5.4.1), so leading to the proposal that the *emc* gene product can dimerise with those of *da* and *AS-C*, so preventing them from binding and activating transcription (Van Doren *et al.*, 1991).

Repressors may also act by directly competing for a binding site, seen in the formation of the seven stripes expressing the *even-skipped* (*eve*) gene in *Drosophila* embryos. *Eve* encodes a homeodomain protein which is first expressed in *Drosophila* embryos at the nuclear cleavage cycle 12. At this time it is uniformly distributed, but by cycle 14 a gradient has been created and *eve* is no longer detectable at the poles. Within thirty minutes, expression of *eve* is restricted to seven transverse stripes, each five to six nuclei wide. The stripe 2 regulatory element within the *eve* promoter contains sites for both activators, *hunchback* (*hb*) and *bicoid* (*bcd*), and repressors, *Kruppel* (*Kr*) and *giant* (*gt*). *Bcd* and *Kr* are unable to co-occupy their closely linked binding sites, so when the negatively acting proteins are expressed, *eve* expression is repressed (Small *et al.*, 1991; 1992).

Masking of the activation region of the bound *trans*-activating factor can lead to repression. As discussed in section 1.5.4, GAL80 inhibits activation of the *GAL* genes by binding to the activation domain of GAL4 (Ma & Ptashne, 1987). Binding at a neighbouring site may also lead to repression, not by exclusion of the activating factor from its site but by masking of the activating domain. The *c-myc* gene is activated by a widely expressed transcription factor, myc-CF1. A second transcriptional regulator, myc-PRF binds to a site in the *c-myc* promoter, repressing transcription even in the presence of myc-CF1. It has been demonstrated that myc-PRF and myc-CF1 can simultaneously occupy the same promoter, indeed that they interact (Kakkis *et al.*, 1989). Therefore it seems plausible that myc-PRF represses *c-myc* by binding next to, and so masking the activating influence of myc-CF1.

The activating signal is not always transmitted directly by the DNA bound factor, but by an intermediary protein. Overexpression of a repressor that also recognises the

intermediary would lead to down regulation of the gene, a phenomenon described as 'squenching'. This may be the mechanism involved in steroid receptor-mediated repression of induction by other steroid receptors. Overproduction of receptors for progesterone, glucocorticoids or oestrogen interferes with induction of the others, but without heterodimer formation, or any interaction between them, so implying the presence of an intermediary molecule (Meyer *et al.*, 1989).

The methods of repression described above only repress a gene via a single activator. Most eukaryotic genes are controlled by many different factors, therefore if transcription of the gene needs to be shut off completely, it would be more efficient for a repressor to directly affect the general transcription machinery. The SV40 T-antigen binds to three sites in the promoter DNA, so blocking access to it by the general transcription machinery, probably RNA pol II and possibly TFIID (Hanson *et al.*, 1981). Histones can probably prevent binding of TFIID to the TATA box, therefore some repressor proteins may act by directing formation of a positioned nucleosome over the TATA box (Roth *et al.*, 1990). The *Drosophila eve* homeodomain protein represses transcription of genes containing homeodomain binding sites in their promoter by blocking one of the early steps in the formation of the preinitiation complex, TFIID or B being likely targets (Johnson & Krasnow, 1992). No factors have yet been identified which interfere with the late stages of preinitiation complex assembly in eukaryotes.

1.5.6 Silencing and genomic imprinting

Higher order chromatin structures are inaccessible to RNA polymerases and initiation factors. Local alterations of the chromatin structure are known to be involved in the repression of several genes. Silencing is a block to gene expression of regions of a chromosome and shares several characteristics with heterochromatin. It is a mechanism of transcriptional inhibition that is not dependent on the arrangement of positive and negative regulatory elements. The yeast mating type loci provides an example of silencing, the mating type switch genes being silenced at the mating type loci, *HMRa* and *HML α* . The actual mechanism of silencing is not known in detail, but various factors involved have been elucidated. The *HMR* and *HML* loci are flanked by two elements, E and I, which are necessary for silencing of the DNA which lies between them. These elements are termed silencers (Brand *et al.*, 1985), and are in many ways analogous to enhancers, their action being independent of orientation and, within limits, of position. Two transcription factors are known to bind to the *HMR*-E silencer, Rap1p and Abf1p, both of which are necessary for repression. In addition, there is a sequence

homologous to the autonomous replicating sequences (ARS) present in the E silencers , implicating DNA replication in the silencing of these cassettes (Brand *et al.*, 1987). Establishment of the repressed state at *HMR* and *HML* requires passage through the synthesis phase of the cell cycle, suggesting that DNA replication may lead to the formation of the repressive chromatin structure that in turn inactivates transcription (Miller & Nasmyth, 1984; Rivier & Pillus, 1994). In addition, the products of four genes *SIR1* to 4 (silent information regulator) are all required to maintain the cassettes in the repressed state. Sir3p and Sir4p associate with Rap1p, so providing a DNA binding complex competent for repression. Histone H4 is also implicated in silencing, strains containing mutations in histone H4 can result in loss of repression (Herskowitz *et al.*, 1992).

Another example of silencing is the inactivation of one X chromosome in all the cells of female mammals as a dosage compensation mechanism. One X chromosome in each cell is compacted into highly condensed chromatin, making it inaccessible to the preinitiation complex factors (Migeon, 1994). DNA methylation is a global suppressor of gene expression, inhibiting RNA synthesis by preventing the binding of basal transcription factors and by altering the chromatin structure (Eden & Cedar, 1994). It may play a role in both X chromosome inactivation and genomic imprinting. Proteins which bind non sequence specifically to methylated CpGs have been identified in both plants (Zhang *et al.*, 1989) and mammals (Meehan *et al.*, 1989). These proteins may be involved in inhibiting the binding of transcription factors or in altering the chromatin structure, possibly by stabilising the 30 nm solenoid fibre. Imprinting is the process by which a chromosome 'remembers' its parental origin and behaves differently according to whether it is inherited from the mother or the father; it provides both a means for transmitting allele-specific signals from gamete to embryo, and a *cis*-acting mechanism for maintaining the imprinted transcription pattern in each cell (Eden & Cedar, 1994). One example of imprinting is seen in the exclusive inactivation of the paternal X chromosome in marsupials and murine placental tissues (Lyon, 1993). However, imprinting does not necessarily involve the shut down of entire chromosomes; it is thought that each gene is affected independently.

1.5.7 Post-transcriptional regulation

In addition to regulation by the control of transcription of a gene, expression can be controlled at several levels subsequently, post-transcriptionally, translationally or post-translationally.

Post-transcriptional control of gene expression is carried out by mechanisms such as mRNA stability, alternative splicing, 3' end formation and RNA editing. It is implicated in the control of catalase isozyme expression during post-germinative growth in cotton seedlings. Catalase is a tetrameric enzyme, of which five isozymes (A, B, C, D & E) exist, composed of two subunits, SU1 and SU2. The isozymes occur in differing abundancies during the first five days of postgerminative growth. SU1 and SU2 are both expressed, being actively transcribed throughout this period. However, the levels of each isozyme vary significantly, indicating that SU1 and SU2 mRNAs must be post-transcriptionally regulated, either by selective degradation or by modification of the transcripts (Ni & Trelease, 1991). Messengers may be stabilised by protein binding as is seen with the chloroplast *psbD* RNA. Increased stability of this mRNA is gained when the leader sequence is bound by a protein and mutations in the binding site lead to reduced stability (Nickelsen *et al.*, 1994). Messengers may contain sequences in the 3' untranslated region (UTR) which lead to destabilisation of the mRNA. The sequence AUUUA when inserted as tandem repeats into the 3' UTR of reporter gene constructs used to transform tobacco led to an increased rate of degradation of these mRNAs compared to those with no insert or a spacer insert. Such sequences are found in several mammalian protooncogene and lymphocyte mRNAs, suggesting that this degradation pathway is conserved amongst eukaryotes (Sullivan & Green, 1993). Alternative splicing is another important level of regulation and may involve a simple decision whether or not to splice, alternative 5' or 3' splice sites, exon skipping or mutually exclusive exons (McKeown, 1992). RNA editing is a process whereby the nucleotide sequence of an RNA molecule is changed with respect to the nucleic acid molecule that encoded it. It sometimes involves the insertion or deletion of nucleotides as is seen in the mitochondria of kinetoplast protozoa. Alternatively, conversion of one nucleotide to another may take place, or replacement of one residue by another. In plant organelles, the result is usually a change from cytidine to uridine. The result of RNA editing is usually a transcript which would not otherwise have been functional, however mRNAs for mammalian apolipoprotein B are functional both before and after editing (Covello & Gray, 1993).

Translation of eukaryotic mRNAs is initiated by ribosomes entering the mRNA at the capped 5'-end and advancing to the AUG codon by linear scanning (Kozak, 1989). ATP and many protein factors are necessary for the ribosomes to engage the mRNA and the small ribosomal subunit only binds stably to the mRNA after the initiator Met-tRNA has bound. Various mechanistic examples of translational regulation in eukaryotic systems have been elucidated. Repression of translation may be caused by the binding of a protein to the mRNA, exemplified by the inhibition of ferritin

translation by an iron-responsive mRNA binding protein which binds to the 5' region, so blocking ribosome access (Brown *et al.*, 1989). Translation can be modulated via phosphorylation of the eukaryotic initiation factor-2 (eIF-2), which catalyses the binding of Met-tRNA to the 40S ribosomal subunits. Phosphorylation of eIF-2 usually leads to a reduced level of translation, presumably by inhibiting this process. However, translation of the yeast transcription factor GCN4 is increased by phosphorylation of eIF-2. GCN4 regulates transcription of many genes involved in amino acid biosynthesis. The mRNA encoding GCN4 is present constitutively, but translation only occurs when the cells are starved of amino acids, and is dependent on reinitiation of translation. The GCN4 mRNA has four short open reading frames (ORFs) upstream of the initiator AUG. Usually, translation will start at the first AUG and will only initiate at subsequent ones given sufficient time for the 40S subunit to be reprimed with a Met-tRNA. Under normal conditions, the first and fourth ORFs of the GCN4 messenger will be translated, but not the second or third, or that encoding GCN4, as repriming will not have taken place by the time the scanning ribosome reaches the initiator AUG codon. However, under starvation conditions, eIF-2 is phosphorylated, leading to an increase in the time taken to reprime the small subunit. Consequently, ORF4 is not translated, but translation of GCN4 can occur (Kozak, 1992).

1.5.8 Post-translational regulation

In addition to the levels of control discussed above, many enzymes are regulated co- or post-translationally, adding a further point at which regulation of activity may occur. The reactions which regulate at this level take many forms and may be reversible or irreversible. Modification can occur by the addition of new functional groups onto the protein, such as phosphates, sulphates, carbohydrates and lipids. Phosphorylation-dephosphorylation is an important reversible modification in the regulation of many enzymes. The glyoxylate bypass of *E. coli* is thought to be principally regulated by the phosphorylation state of isocitrate dehydrogenase (IDH). During growth on glucose, IDH exists mainly in the active dephosphorylated form. However, during growth on acetate, IDH is partially inactivated by phosphorylation catalysed by isocitrate dehydrogenase kinase/phosphatase, so directing the carbon flux through ICL and the glyoxylate bypass (Borthwick *et al.*, 1984). Phosphorylation of ICL is now thought to also play a role in partitioning between the glyoxylate bypass and the Krebs' cycle (Hoyt & Reeves, 1988; Robertson & Reeves, 1989). Other chemical modifications are irreversible, such as the covalent linkage of a carbohydrate group to egg albumin (Graves *et al.*, 1994).

Some proteins are synthesised as inactive precursors which require cleavage to be activated. Activation of trypsinogen to trypsin requires cleavage of a lysine-isoleucine bond to release a hexapeptide and the active trypsin, and many of the proteins involved in blood clotting are activated on cleavage of a precursor (Graves *et al.*, 1994). Some proteins may only be active when present in an oligomeric form, both ICl and MS are active as oligomers, ICL as a tetramer and MS as an octamer.

Protein degradation may be a very important regulatory step, allowing the re-utilisation of amino acids and the ability to alter the protein content in response to a changing environment, or developmental stage. Degradation must necessarily be specific, not all proteins within a cell are degraded at the same time. One important specific degradation pathway which occurs in the nucleus and the cytoplasm is that involving covalent linkage of ubiquitin to the protein to be degraded. Ubiquitin is probably the most highly conserved of all proteins, exhibiting only three amino acid changes between higher plant and mammalian forms (Hershko & Ciechanover, 1992). Once ubiquitinated, a protein is likely to be degraded, though some may have the ubiquitin polypeptide cleaved and others can exist stably in the cell as conjugates (Vierstra, 1993).

1.6 Aims of this thesis

The analysis of genes expressed at defined stages of development or in response to specific metabolites serves to further the understanding of the mechanisms controlling plant gene expression during development and in response to metabolic status. The *ms* and *icl* genes are expressed during embryogenesis, germination and senescence, and in response to starvation conditions. The *ms* gene of cucumber has been isolated and the regions necessary for its expression during and post germination have been analysed. Recent studies have shown that only 248 bp of the 5' region upstream of the start of transcription are necessary for expression during postgerminative growth (Sarah & Smith, pers. comm.) and for metabolic regulation in protoplasts (Graham *et al.*, 1994a). The *ms* and *icl* genes are coordinately expressed in various species, leading to the suggestion that the mechanisms regulating the genes encoding these enzymes may be shared. The aim of this project was to investigate this intriguing possibility, by isolating and characterising a genomic clone encoding cucumber ICL. This would enable subsequent comparison of the regulatory region of the *icl* gene with that of the *ms* gene, leading to the experimental analysis of the elements responsible for the control of gene expression of this gene during plant development and in response to environmental changes.

CHAPTER 2

MATERIALS AND METHODS

2.1 Biological material

2.1.1 Plant material

Cucumber seeds (*Cucumis sativus* L. cv 'Masterpiece') were obtained from W. K. McNair, Portobello, Edinburgh. Seeds were imbibed at 4 °C overnight in water and sown either in vermiculite, or on damp filter paper, and incubated in a regime of a 16 h day at 25 °C and an 8 h night at 22 °C and an irradiance of 120 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Seeds sown on filter paper were transferred to Levington Universal potting compost 3 to 4 days post imbibition. *Nicotiana plumbaginifolia* seeds were obtained from J. R. Ellis, Department of Botany, University of Durham. Seeds were imbibed in 1 mM gibberellic acid overnight at 4 °C, sown on the surface of Levington Universal potting compost and incubated in a regime of a 16 h day at 25 °C and an 8 h night at 22 °C and an irradiance of 120 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

2.1.2 Bacterial strains and genotypes

<i>Escherichia coli</i> :	JM101	<i>supE thi</i> $\Delta(\text{lac-proAB})$ F'[<i>traD36 proAB+lacI</i> $\phi\text{lacZ}\Delta\text{M15}$] Used as a host for recombinant manipulation.
	DH5 α	<i>supE44</i> $\Delta\text{lacU169}$ ($\phi\text{80lacZ}\Delta\text{M15}$) <i>hsdR17 recA1</i> <i>endA1 gyrA96thi-1 relA1</i> Used as a host for recombinant manipulation.
	HB101	<i>supE44 hsdS20</i> ($r_B^- m_B^-$) <i>recA13 ara-14 proA2</i> <i>IacY 1 gal K 2 rpsL20 xyl- 5 mtl- 1</i> Used in triparental mating with <i>Agrobacterium tumefaciens</i> - carries the helper plasmid pRK2013.
	XL1-Blue	<i>supE 44 hsdR 17 recA 1 endA 1 gyrA 46 thi relA1 lac-</i> F' [<i>proAB+ lacI</i> ϕ <i>lacZ</i> ΔM15 Tn10 (<i>tet^r</i>)] Used to plate and propagate λZAPII bacteriophage.
	NM514	<i>hsdR lyc 7</i> (<i>lyc</i> is an <i>hfl</i> allele) Used to plate and propagate λZAPII and λNM1149 (allows growth of recombinant phage only - selection of $\lambda\text{imm}^{434}\text{CI}$).

Agrobacterium tumefaciens:
LBA4404

Genotype not available.

Carries a cryptic and a 'disarmed' Ti plasmid, the latter lacking the entire T-DNA, but with an intact *vir* region. The bacterial chromosome carries streptomycin resistance and the disarmed Ti plasmid carries rifampicin resistance (Hoekema *et al.*, 1983)

2.1.3 Bacterial plasmids and bacteriophage

Vector	Source	Use
pBS+	Stratagene	General subcloning; sequencing
pBluescript SK +/-	Stratagene	Subcloning; <i>in vivo</i> excision
λ ZAPII	Stratagene	cDNA library vector
R408	Stratagene	<i>in vivo</i> excision
λ NM1149	Murray (1983)	Genomic library vector
pBin19	Bevan (1984)	Plant transformation
pRAJ275	Jefferson <i>et al.</i> , (1986)	Promoter-GUS fusion generation
pRK2013	Ditta <i>et al.</i> , (1980)	Triparental mating

2.1.4 Isocitrate lyase cDNA clone (partial)

The pBS+ plasmid carrying a 1.4 kb cDNA fragment encoding ICL, designated pBSICL. cDNA was synthesised on mRNA template extracted from cucumber cotyledons germinated for 3 days. Gift from W. Becker, University of Wisconsin.

2.1.5 Malate synthase cDNA clone

The pBS+ plasmid carrying a 1.9 kb cDNA fragment encoding MS, designated pBSMS1.9. cDNA was synthesised on mRNA template extracted from cucumber cotyledons germinated for 3 days. Gift from J. Hunter and B Schwartz, University of Wisconsin.

2.1.6 Cucumber cDNA library in λ ZAPII

Constructed in this laboratory by Dae-Jae Kim (Kim & Smith, 1994). cDNA was synthesised on mRNA template extracted from senescing cucumber cotyledons.

2.2 Miscellaneous

2.2.1 General comments

Chemicals and reagents were purchased from BDH Chemicals Ltd., or from Sigma Chemical Co. Ltd., unless otherwise stated. All manipulations were carried out at room temperature, unless otherwise stated. Sterilisation of media, chemicals and equipment was performed by autoclaving (120 °C, 20 min, 15 lb in⁻²), by baking (180 °C for at least 6 h) or using disposable filters with a pore size of 0.2 µm (Acrodisc PF or Acrocap, Gelman Sciences, Michigan).

2.2.2 Bacteriological media

Luria-Bertani medium (litre ⁻¹)	10 g bacto-tryptone (Difco laboratories), 5 g bacto-yeast extract (Difco), 10 g NaCl, pH 7.0.
LB agar (litre ⁻¹)	As for LB medium with the addition of 15 g bacto-agar (Difco).
M9 minimal agar (litre ⁻¹)	200 ml 5 X M9 salts (64 g Na ₂ HPO ₄ ·7H ₂ O, 15 g KH ₂ PO ₄ , 2.5 g NaCl, 5.0 g NH ₄ Cl), 20 % (w/v) glucose, 15 g bacto-agar.
2 X YT medium (litre ⁻¹)	16 g bacto-tryptone, 10 g bacto-yeast extract, 5 g NaCl, pH 7.0.
BBL Trypticase agar (litre ⁻¹)	10 g trypticase (Baltimore Biological laboratories), 5 g NaCl, pH 7.2, 15 g bacto-agar.
BBL top (litre ⁻¹)	As for BBL agar, but with 6.5 g bacto-agar.

2.2.3 Tissue culture media

All media were based on the basic medium of Murashige and Skoog (1962; M & S medium, Flow Laboratories, Irvine).

Shooting medium (litre ⁻¹)	4.7 g M & S medium, 10 g sucrose, 0.2 mg N ⁶ Fur-furylaminopurine (kinetin), 2 mg naphthaleneacetic acid (NAA), 8 g bacto-agar, pH 5.6.
Expansion medium (litre ⁻¹)	2.35 g M & S medium, 5 g sucrose, 8 g bacto-agar.
Rooting medium (litre ⁻¹)	2.35 g M & S medium, 8 g bacto-agar.

Protoplast culture medium (litre ⁻¹)	
- sucrose	2.35 g M & S medium, 31.885 g D-mannitol, 1 mg NAA, 0.2 mg benzylaminopurine (BAP), 0.1 mg dichlorophenoxyacetic acid (2-4 D), pH 5.6 with 0.1 M KOH.
+ sucrose	As above except: 30.063 g D-mannitol, 3.423 g sucrose.

2.2.4 Restriction endonucleases and DNA modification enzymes

All restriction endonucleases and DNA modification enzymes were purchased from Northumbria Biologicals Ltd. (NBL), Boehringer Mannheim Biochemicals (BCL), Pharmacia LKB, Cambridge Biolabs (CAMBIO) and Gibco-BRL.

2.2.5 Antibiotics

Antibiotic	Stock conc.	Working conc.
Ampicillin	100 mg ml ⁻¹ in distilled water (d H ₂ O)	100 µg ml ⁻¹
Rifampicin	100 mg ml ⁻¹ in DMSO	100 µg ml ⁻¹
Streptomycin	300 mg ml ⁻¹ in d H ₂ O	300 µg ml ⁻¹
Kanamycin	100 mg ml ⁻¹ in d H ₂ O	100 µg ml ⁻¹ (bacteria) 200 µg ml ⁻¹ (shooting) 200 µg ml ⁻¹ (expansion) 100 µg ml ⁻¹ (rooting)
Carbenicillin	250 mg ml ⁻¹ in d H ₂ O	500 µg ml ⁻¹ (shooting) 250 µg ml ⁻¹ (expansion) 100 µg ml ⁻¹ (rooting)
Tetracycline	5 mg ml ⁻¹ in ethanol	10 µg ml ⁻¹

2.2.6 Radiochemicals

[α-³²P]dCTP (3000 Ci mmol⁻¹), [γ-³²P]dATP (>5000 Ci mmol⁻¹) and [α-³⁵S]dATP (>1000 Ci mmol⁻¹) were purchased from Amersham International plc.

2.2.7 Autoradiography

Cronex 4 X-ray film was used and exposed at -70 °C using intensifying screens (Lightning Plus, Cronex) for detection of [³²P]-labelled nucleotides, or at room temperature for [³⁵S]-labelled nucleotides. Films were developed using an X-Omat developer.

2.3 DNA isolation, manipulation and analysis

2.3.1 Small scale plasmid/phagemid isolation (up to 3 ml culture)

DNA was isolated using a modification of the method of Birnboim and Doly (1979). Bacterial cells were harvested by centrifugation at 12,000 g for 2 min. The pellet was resuspended in 100 µl solution 1 (50 mM glucose, 25 mM Tris-HCl, 10 mM ethylenediaminetetraacetic acid [EDTA], pH 8.0). The cells were lysed by addition of 200 µl solution 2 (0.2 M NaOH, 1 % [w/v] sodium dodecyl sulphate [SDS]) and incubated on ice for 5 min. 150 µl ice-cold solution 3 (3M sodium acetate pH 5.0) were added, the sample mixed by inversion and incubated on ice for a further 5 min, after which time the bacterial debris and genomic DNA was pelleted by centrifugation at 12,000 g for 5 min. Protein was removed from the supernatant containing the plasmid DNA by addition of an equal volume of phenol (buffered to pH 7.4 with 1 M Tris HCl). The sample was mixed, the two phases separated by centrifugation at 12,000 g and the upper, aqueous phase removed to a fresh tube. This process was repeated with an equal volume of chloroform:isoamyl alcohol (24:1 by volume) to remove any traces of phenol, and the DNA precipitated from the solution by addition of 0.1 volumes 3 M sodium acetate, pH 5.0, 2 volumes ethanol and centrifugation at 12,000 g for 10 min at 4 °C. The DNA pellet was washed in 70 % (v/v) ethanol, dried under vacuum and resuspended in 50 µl sterile double distilled water (sdd H₂O).

2.3.2 Large scale plasmid/phagemid isolation (for more than 100 ml culture)

100 to 200 ml of bacterial cells were harvested and DNA isolated as in section 2.3.1 with scaled up volumes: resuspended in 10 ml solution 1, 20 ml solution 2 added, followed by 15 ml solution 3, incubated on ice for at least 30 min and centrifuged at 10,000 g for 15 min. The DNA was precipitated from the supernatant by addition of 0.6 volumes isopropanol for 30 min at -20 °C. The DNA was pelleted by centrifugation at 10,500 g for 30 min and resuspended in sdd H₂O. The volume of the DNA sample

was measured, 0.95 g ml⁻¹ CsCl added, and the solution loaded into 11.5 ml crimp-seal ultracentrifuge tubes (Sorvall) containing 0.45 ml of 10 mg ml⁻¹ ethidium bromide. The tubes were sealed and ultracentrifugation carried out at 46,000 rpm in a 70Ti rotor for 48 h at 20 °C.

To remove the plasmid band from the tube, the top of the tube was pierced and an 18 gauge needle inserted into the tube immediately below the band. The band was withdrawn from the tube into a 2 ml syringe and transferred to a sterile Corex tube. To remove the ethidium bromide from the sample, an equal volume of butan-1-ol, saturated with NaCl, was added, the 2 phases mixed by shaking and allowed to separate. The upper (organic) phase was removed and the process repeated 3 to 4 times until the aqueous phase was free of visible ethidium bromide.

The aqueous phase containing the plasmid DNA was diluted by addition of 2 volumes sdd H₂O and precipitated by addition of 6 volumes ethanol. The sample was placed at -20 °C overnight and the DNA collected by centrifugation at 9,000 g for 30 min at 4 °C. The DNA pellet was resuspended in sdd H₂O, reprecipitated with ethanol and repelleted. The pellet was washed in 70 % (v/v) ethanol, dried and resuspended in sdd H₂O. The concentration and purity of the DNA was estimated as described in section 2.3.6 and diluted to the required concentration.

2.3.3 Genomic DNA isolation

Isolation of genomic DNA was based on the method described by Dellaporta *et al.* (1983). Plant tissue was frozen in liquid nitrogen, ground to a fine powder using a pestle and mortar and transferred to a centrifuge tube. 4.5 ml extraction buffer (0.1 M Tris-HCl pH 8.0, 50 mM NaCl, 10 mM β-mercaptoethanol) per g of tissue were added and the sample incubated at 65 °C until thawed. 0.6 volumes per g of tissue of 10 % (w/v) SDS were added, the sample mixed and incubated at 65 °C for 20 min. Debris was pelleted by centrifugation at 2,500 g for 15 min and the supernatant removed to a fresh tube. Protein was precipitated by the addition of 2.5 volumes 5 M potassium acetate pH 5.5 and the sample incubated on ice for 10 min. Protein was pelleted by centrifugation at 2,500 g for 15 min and the supernatant transferred to a fresh tube. DNA was precipitated by addition of 0.6 volumes isopropanol and incubation at -70 °C for at least 30 min. It was pelleted by centrifugation at 2,500 g for 10 min and resuspended in 4.2 ml sdd H₂O. Reprecipitation was achieved by addition of 0.1 volumes 3M sodium acetate and 0.6 volumes isopropanol and incubation at room temperature for 10 min. DNA was recovered by centrifugation at 2,500 g for 10 min, washed in 70 % (v/v) ethanol and air dried. It was resuspended in 300 µl sdd

H₂O and incubated at 65 °C to aid resuspension. Any insoluble material was discarded. All samples were treated with 10 µg ml⁻¹ RNase prior to gel electrophoresis.

2.3.4 Bacteriophage DNA isolation

Qiagen DNA affinity columns (Qiagen inc) were used in the isolation of phage DNA from plate lysates (see section 2.6.3). 30 µl buffer L1 (20 mg ml⁻¹ RNase A, 6 mg ml⁻¹ DNase1, 0.2 mg ml⁻¹ bovine serum albumin [BSA], 10 mM EDTA, 100 mM Tris-HCl, 300 mM NaCl, pH 7.5) were added to 10 ml of plate lysate and the sample incubated at 37 °C for 30 min. 2 ml ice-cold buffer L2 (30 % [w/v] polyethylene glycol [PEG 6000], 3 M NaCl) were added, the sample mixed gently and incubated on ice for 60 min. The bacteriophage were pelleted by centrifugation at 10,500 g for 10 min and the pellet resuspended in 1 ml buffer L3 (100 mM Tris-HCl, 100 mM NaCl, 25 mM EDTA, pH 7.5). 1 ml buffer L4 (4 % [w/v] SDS) was added, the sample heated to 70 °C for 10 min and cooled on ice. 1 ml buffer L5 (2.55 M potassium acetate pH 4.8) was added, mixed gently and the debris pelleted by centrifugation at 10,500 g for 30 min at 4 °C.

A Qiagen-tip20 was equilibrated with 1 ml buffer QBT (750 mM NaCl, 50 mM 3-[*N*-morpholino]propanesulfonic acid [MOPS], 15 % [v/v] ethanol, pH 7.0, 0.15 % [v/v] Triton X-100). The supernatant from the final centrifugation step was applied and the column washed through two times with 1 ml buffer QC (1.0 M NaCl, 50 mM MOPS, 15 % [v/v] ethanol, pH 7.0). The DNA was eluted from the column with 1.5 ml buffer QF (1.25 M NaCl, 50 mM MOPS, 15 % [v/v] ethanol, pH 8.2) using gravity flow, and the remaining solution forced out using a 'Pipetteman' automatic pipette.

The DNA was precipitated by the addition of 0.6 volumes isopropanol, and collected by centrifugation at 12,000 g for 5 min. It was washed with 70 % (v/v) ethanol, air dried and resuspended in 30 µl TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The concentration was determined as described in section 2.3.6.

2.3.5 *Agrobacterium* DNA isolation

Agrobacterium DNA isolation was performed using the method described by Draper *et al.*, (1988). Bacterial cells were pelleted from 1.5 ml of culture grown for 48 h, and resuspended in 300 µl TE pH 8.0. 100 µl 5 % (w/v) sodium lauryl sarcosinate (sarkosyl) were added and the sample mixed. 25 µg proteinase K in 150 µl sdd H₂O was added, the sample mixed and incubated at 37 °C for 1 h. 500 µl of phenol were

added to the tube, and the viscous mass sheared by passage through a 1 ml Gilson tip five times. The two phases were separated by centrifugation at 12,000 g for 5 min. The aqueous phase was reextracted 3 times with 500 µl phenol:chloroform:isoamyl alcohol (25:24:1 by volume). DNA was precipitated by addition of NaCl to a final concentration of 0.25 M and 3 volumes cold ethanol and storage at -20 °C for 2 h. The DNA was pelleted by centrifugation at 12,000 g for 10 min at 4 °C, washed in 70 % (v/v) ethanol, partially dried under vacuum and resuspended in 50 µl sdd H₂O.

2.3.6 Estimation of DNA and RNA concentration

Spectrophotometric readings of a known dilution of a DNA or RNA sample were taken at 260 nm and 280 nm. An A₂₆₀ reading of 1 corresponds approximately to 50 µg ml⁻¹ for double stranded (ds) DNA and 40 µg ml⁻¹ for single stranded (ss) DNA and RNA. An estimate of sample purity is provided by the ratio of the absorbance value at 260 nm to that at 280 nm, pure DNA and RNA having values of 1.8 and 2.0 respectively.

2.3.7 Restriction endonuclease analysis of DNA samples

Digestions were carried out according to the manufacturers recommended conditions, using buffers supplied by BCL.

2.3.8 Horizontal gel electrophoresis of DNA

DNA samples were analysed using gels prepared from agarose (ultra-pure) at variable concentrations between 0.7 and 1.0 % (w/v) in TAE buffer (0.04 M Tris-Acetate, 1 mM EDTA, pH 8.0). Samples were loaded with 0.1 volumes gel loading dye (0.25 % [w/v] bromophenol blue, 0.25 % [w/v] xylene cyanol FF, 15 % [w/v] Ficoll in water). Ethidium bromide was either included in the gel mix to a final concentration of 0.5 µg ml⁻¹, or the gel was stained using a solution of 0.5 µg ml⁻¹ ethidium bromide for 30 min, and DNA visualised using a short wavelength trans-illuminator (Hybaid). DNA size markers were purchased from BRL.

2.3.9 Southern blotting

DNA samples were fractionated using horizontal agarose gels, visualised and photographed. For gels containing fragments greater than 10 kb, the gel was depurinated by soaking in 0.25 M HCl for 15 min. The gel was transferred to

denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 20 min, followed by 2 periods of 15 min in neutralisation buffer (1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 1 mM EDTA). DNA fragments were transferred to Hybond N filters (Amersham) in 20 X SSC (3 M NaCl, 0.3 M sodium citrate) by capillary action according to the method of Southern (1975). The transferred DNA fragments were crosslinked to the nylon membrane by ultra-violet irradiation of 0.4 J cm⁻² using a Hybaid crosslinker.

2.3.10 DNA sequencing and computer analysis

2.3.10.1 Sequencing double stranded plasmid DNA

DNA was isolated by alkaline lysis as described in section 2.3.1, from 3 ml culture grown overnight. RNA was removed by treating the sample with 10 µg µl⁻¹ RNase at 37 °C for 1 h. The DNA was further purified by use of GeneClean glass bead solution ('glassmilk') (Biolabs 101 Inc.) according to the manufacturers protocol, and eluted from the glassmilk in a total volume of 18 µl sdd H₂O.

The DNA was denatured by addition of 2 µl 2 M NaOH and incubation at 37 °C for 30 min. Annealing of the primer to the template was achieved by addition of 1 µl 10 mM primer, followed by precipitation by addition of 6 µl 3 M potassium acetate pH 5.0 and 150 µl ethanol. The annealed template/primer mix was recovered by centrifugation at 12,000 g for 30 min at 4 °C. After washing in 70 % (v/v) ethanol and drying under vacuum, the annealed template/primer DNA was resuspended in 12 µl sdd H₂O and 2 µl annealing mix (a buffered solution containing MgCl₂ and dithiothreitol [DTT]). The DNA template was then sequenced (labelling and termination reactions) according to the T7Sequencing™ kit protocol (Pharmacia). Primers used were M13 -40 and reverse primers, or ones specific to the ICL genomic clone or to the gene encoding β-glucuronidase (GUS).

The samples were heated at 80 °C for 2 min and resolved on a 6 % (w/v) acrylamide, 7 M urea, 1 X TBE (0.09 M Tris-borate, 2 mM EDTA, pH 8.0) gel. Following electrophoresis, the gel was fixed in 10 % (v/v) methanol, 10 % (v/v) acetic acid and dried under vacuum at 80 °C prior to autoradiography.

2.3.10.2 Computer analysis

Analysis of nucleotide sequence data and derived amino acid sequence was carried out using programmes of the University of Wisconsin Genetics Computer Group

(UWGCG, Devereux *et al.*, 1984) through the VAX 8000 system at Edinburgh University.

2.3.11 Generation of nested deletions

Nested deletions of DNA fragments were required both for DNA sequencing (section 2.3.10) and for use in promoter analysis (section 2.10). Deletions of a fragment were generated using the exonucleases *Exo* III and S1. *Exo* III removes 5' mononucleotides from recessed or blunt 3' termini, protruding 3' termini being completely resistant to the action of the enzyme. To create various nested deletions, the plasmid is digested between the target DNA and the plasmid sequences with two restriction endonucleases. One cleaves nearer the plasmid sequences, generating a protruding 3' terminus, so protecting the plasmid from deletion in this direction, and the other cleaves next to the target DNA generating either a blunt or a recessed 3' terminus. Digestion will therefore occur unidirectionally into the target DNA sequence. The protruding single strands are then removed by digestion with S1 nuclease, and the plasmid recircularised. Digestion with *Exo* III nuclease occurs at a uniform rate so allowing the production of nested deletions of the required size.

10 µg plasmid DNA were digested with two restriction endonucleases for 16 h at 37 °C. The restriction endonucleases were extracted from the sample by addition of an equal volume of phenol, the sample mixed and the two phases separated by centrifugation at 12,000 g for 5 min. The aqueous phase was transferred to a fresh tube and the extraction step repeated by addition of an equal volume of chloroform:isoamyl alcohol (24:1 by volume) to remove any traces of phenol. The DNA was precipitated from the solution by addition of 0.1 volumes 3 M sodium acetate, pH 5.0, 2 volumes ethanol and centrifugation at 12,000 g for 10 min at 4 °C. The pellet was washed in 70 % [v/v] ethanol and dried. It was resuspended in 60 µl 1 X exonuclease III buffer (66 mM Tris HCl pH 8.0, 6.6 mM MgCl₂) and preincubated at 37 °C for 5 min. 2.5 µl were removed and added to 7.5 µl S1 reaction mixture (40 mM potassium acetate, pH 4.5, 0.3 M NaCl, 0.25 mM ZnSO₄, 6 % [v/v] glycerol, 0.3 U µl⁻¹ S1 nuclease) and stored on ice. 150 units (U) per pmol of recessed 3' termini *Exo* III were added to the remaining DNA solution, the sample mixed and incubated at 37 °C. Under these conditions, *Exo* III nuclease removes approximately 200 nucleotides per minute from the recessed 3' end. 2.5 µl samples were removed at minute intervals and added to 7.5 µl S1 reaction mix and stored on ice. After all the samples had been taken, the tubes were incubated at 30 °C for 30 min, after which time the reaction was stopped by the addition of 40 µl S1 stop mixture (60 mM Tris HCl, 10 mM EDTA, pH 8.0). 10 µl

aliquots were analysed by electrophoresis through 1 % [w/v] agarose to check the rate of deletion. The samples of the desired size were extracted by addition of an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 by volume), the two phases separated by centrifugation at 12,000 g for 5 min at 4 °C and the aqueous phase removed to a fresh tube. DNA was precipitated by addition of 0.1 volumes 3 M sodium acetate, pH 5.0, 2 volumes ethanol and centrifugation at 12,000 g for 10 min at 4 °C. The pellet was washed in 70 % [v/v] ethanol, dried and resuspended in 15 µl sdd H₂O. To recircularize the linear DNA, 2 µl 10 X ligase buffer (500 mM Tris HCl pH 7.6, 100 mM MgCl₂, 100 mM DTT, 500 µg ml⁻¹ BSA), 1 µl 0.5 mM dNTPs, 1 U Klenow fragment and 4 Weiss U T4 DNA ligase were added and the samples incubated at room temperature for 16 h. 10 µl were used to transform competent *E. coli* (section 2.5.4) and those colonies containing plasmid of the desired size detected by restriction endonuclease digestion and horizontal gel electrophoresis (sections 2.3.7 & 2.3.8). Deleted fragments were sequenced to map the extent of the deletion cloned into pRAJ275 for promoter analysis (section 2.10 and chapter 5)

2.4 RNA isolation and analysis

2.4.1 Isolation of total RNA from cucumber cotyledons

Isolation of total RNA was performed using a modification of the method described by Parish & Kirby (1966). 1 to 2 g plant tissue was frozen in liquid nitrogen, ground to a fine powder using a baked pestle and mortar and immediately transferred to a cooled baked corex tube, to which 4 volumes extraction buffer (100 mM Tris-HCl pH 8.5, 1 % [w/v] tri-isopropyl naphthalene [TNS], 6 % [w/v] 4-aminosalicylic acid [PAS]) were added. Protein was extracted from the sample by addition of an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 by volume), the sample mixed and the two phases separated by centrifugation at 2,500 g for 10 min at 4 °C. The aqueous phase was transferred to a fresh tube and the extraction step repeated. Nucleic acids were precipitated by the addition of 0.1 volumes 3 M sodium acetate pH 5.0, 2 volumes ethanol and incubation at -20 °C for 3 h. The nucleic acids were pelleted by centrifugation at 10,500 g for 30 min at 4 °C and resuspended in 2 to 3 ml sdd H₂O. Total RNA was selectively precipitated by the addition of 5 M LiCl to give a final concentration of 2 M and incubation at 4 °C for 12 to 16 h. Total RNA was pelleted by centrifugation at 10,500 g for 30 min at 4 °C, washed twice in 70 % (v/v) ethanol, resuspended in sdd H₂O and precipitated with ethanol as above. Centrifugation was repeated, the RNA pellet washed twice in 70 % (v/v) ethanol to remove excess salt,

dried and resuspended in sdd H₂O. Purity and yield were determined as described in section 2.3.6.

2.4.2 Isolation of total RNA from small amounts of plant tissue

To maximise the yield of nucleic acids from small amounts of plant material, the following protocol was used (adapted from Castresan *et al.*, 1988). Plant material was frozen in liquid nitrogen, ground to a fine powder using a baked pestle and mortar, and immediately transferred to a cold 15 ml Corex tube. A minimum of 4 ml guanidium extraction buffer (5 M guanidium thiocyanate, 25 mM sodium citrate, 0.5 % (w/v) sarcosyl, 2 mM EDTA, 50 mM Tris-HCl pH 7.6, 100 mM β -mercaptoethanol) were added and the sample vortexed for 10 s. For amounts of plant material greater than 1 g, 4 ml buffer per g fresh weight were added. After centrifugation at 5,000 g for 10 min at 0 °C, the clarified supernatant was removed to a fresh tube and protein extracted by addition of an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 by volume). The two phases were separated by centrifugation at 10,000 g for 30 min at 0 °C, the aqueous phase removed to a fresh tube and the extraction step repeated. The volume of the aqueous phase was made up to 7.5 ml with sdd H₂O and the nucleic acids precipitated by addition of 2.5 volumes ethanol and incubation at -20 °C overnight. Nucleic acids were pelleted by centrifugation at 15,000 g for 10 min at 0 °C, washed in 70 % (v/v) ethanol and resuspended in 2.5 ml sdd H₂O. Total RNA was precipitated by addition of 5 M LiCl to a final concentration of 2 M and incubation at 4 °C for 12 to 16 h. Total RNA was collected by centrifugation as above and precipitation by the addition of ethanol repeated. Pelleted RNA was washed twice in 70 % (v/v) ethanol, dried and resuspended in sdd H₂O. Purity and yield were determined as described in section 2.3.6.

2.4.3 Isolation of total RNA from protoplasts

Protoplasts were harvested from culture medium by centrifugation at 1000 g and the supernatant removed. 3 ml guanidium extraction buffer (see section 2.4.2) were added per 5×10^6 protoplasts and the sample vortexed for 10 s. 3 ml phenol:chloroform:isoamyl alcohol (25:24:1 by volume) were added, the tube vortexed and the phases separated by centrifugation at 10,000 g for 10 min at 0 °C. The supernatant was removed to a fresh tube and the extraction step repeated. Total nucleic acid was precipitated by addition of 2.5 volumes ethanol, incubation at 20 °C for 3 h and centrifugation at 10,000 g for 30 min at 0 °C. The nucleic acid pellet was washed twice

in 70 % (v/v) ethanol, dried and resuspended in sdd H₂O. This preparation was used for Northern blotting analysis (see section 2.4.4) without removing the DNA.

2.4.4 Horizontal gel electrophoresis of RNA and Northern blotting

RNA was size fractionated on horizontal denaturing agarose gels: 1.3 % (w/v) agarose was melted in boiling water and allowed to cool to 60 °C before addition of 10 X gel running buffer (GRB: 0.2 M MOPS, 50 mM sodium acetate, 10 mM EDTA, pH 7.0), 12.3 M formaldehyde and 10 mg ml⁻¹ ethidium bromide to give final concentrations of 1 X, 2.2 M and 0.05 µg ml⁻¹ respectively.

RNA samples were prepared for loading by addition of an equal volume of formamide sample buffer (10 X GRB:formamide:40 % [w/v] formaldehyde 1:2:1.2 by volume), incubation at 60 °C for 5 min, followed by cooling on ice. 0.25 volumes loading buffer (0.1 M EDTA, 30 % (w/v) Ficoll type 400, 0.25 % (w/v) bromophenol blue) were added and the samples loaded onto the gel. Electrophoresis was carried out in 1 X GRB. After photographing the gel, the RNA was transferred to Hybond-N filters (Amersham) in 20 X SSC by capillary action. The filters were subsequently treated as described for Southern blotting (section 2.3.9).

2.4.5 Primer extension

10 µg total RNA (isolated from 3 day post imbibition cucumber cotyledons using the method described in section 2.4.2) were mixed with 1 or 2 µl radiolabelled oligonucleotide (section 2.7.2) in 500 mM KCl, 50 mM Tris-HCl pH 8.5, 0.5 mM EDTA, 8 mM MgCl₂ in a final volume of 10 µl, denatured by heating at 80 °C for 10 min and annealed by incubation at room temperature for 30 min. Extension of DNA molecules from the primer was achieved by addition of 10 µl 0.5 U µl⁻¹ RNAsin (Promega), 5 mM DTT, 0.5 mM dNTPs, 200 U reverse transcriptase ('Superscript', Gibco) and incubation at 40 °C for 90 min. The nucleic acids were precipitated by addition of 2 µl 3 M sodium acetate, 50 µl ethanol, incubation at -20 °C overnight and centrifugation at 12,000 g for 30 min at 4 °C. The pellet was washed in 70 % (v/v) ethanol, dried and resuspended in 4.5 µl sdd H₂O and 3.5 µl sequencing loading buffer (Pharmacia). The reaction products were heated at 80 °C for 2 min and resolved on a 6 % (w/v) acrylamide, 7 M urea, 1 X TBE (90 mM Tris-borate, 2 mM EDTA, pH 8.0) gel. Sequencing reactions using the radiolabelled oligonucleotide as primer were run alongside the primer extension products, in order that the exact length of the primer

extension products could be determined. Following electrophoresis, the gel was treated as described in section 2.3.10.1

2.5 DNA fragment subcloning and bacterial transformation

2.5.1 DNA fragment isolation and preparation for subcloning

The DNA fragment ('insert') to be subcloned was excised from the plasmid, phagemid or bacteriophage by digestion with the appropriate restriction endonuclease(s). It was separated from other DNA molecules by electrophoresis through a 0.8 % (w/v) agarose gel (section 2.3.8). The portion of the gel containing the fragment was excised from the gel using a scalpel. The insert DNA was isolated from the agarose using GeneClean glass bead solution. The yield of DNA recovery from the gel slice was estimated by electrophoresis of an aliquot through 0.8 % (w/v) agarose.

2.5.2 Vector preparation for the receipt of DNA inserts

The vector (plasmid, phagemid or bacteriophage), into which the DNA insert was to be subcloned, was digested with the appropriate restriction endonuclease(s). To remove the enzyme after digestion, the sample was extracted by addition of, and mixing with, an equal volume of phenol, the phenol was then removed by extraction with chloroform:isoamyl alcohol (24:1 by volume), the DNA precipitated by addition of 0.1 volumes sodium acetate, 2 volumes ethanol and centrifugation at 12,000 g for 15 min at 4 °C. It was resuspended in *sdd* H₂O and the concentration estimated by electrophoresis of an aliquot through 0.8 % (w/v) agarose.

2.5.3 Ligation of the prepared insert and vector DNA

For the ligation of DNA fragments and vectors with cohesive terminals, 100 to 200 ng vector DNA and a 3 fold molar excess of insert were mixed and *sdd* H₂O added to a final volume of 17 µl. To this were added 2 µl 10 X ligation buffer (200 mM Tris-HCl pH 7.6, 50 mM MgCl₂, 50 mM DTT, 5 mM ATP) and 4 Weiss U T4 DNA ligase (1 µl). The reaction mix was incubated at 16 °C for 16 h.

2.5.4 Transformation of *E. coli* with plasmid or phagemid DNA

0.5 ml of a culture of *E. coli* grown overnight was used to inoculate 50 ml LB-medium and was incubated at 37 °C with continuous shaking until the cells reached an A_{600} of 0.6. Cells were pelleted by centrifugation at 5,000 g for 10 min at 4 °C, resuspended in 25 ml 100 mM CaCl_2 and incubated on ice for 30 min. The cells were pelleted again as before and resuspended in 5 ml 100 mM CaCl_2 . Cells were stored at 4 °C and used within 2 days of preparation.

DNA (0.01 μg to 0.10 μg) in 100 μl 100 mM CaCl_2 was added to 200 μl competent cells and left on ice for 30 min. The cells were heat shocked at 42 °C for 90 s and added to 1 ml LB-medium. Incubation at 37 °C for 1 h allowed expression of resistance proteins encoded by the recombinant plasmid, and then the cells were pelleted by centrifugation at 12,000 g for 1 min. The cells were resuspended in 100 μl LB-medium and spread onto LB-agar plates containing the appropriate antibiotic and detection chemicals, to allow the isolation of colonies containing recombinant plasmid, and incubated overnight at 37 °C.

2.5.5 Screening of bacterial colonies transformed with recombinant plasmid or phagemid DNA

2.5.5.1 Screening by α -complementation

The pBS+ and pBluescript phagemids used in this study carry a short segment of DNA that contains the regulatory sequence and the first 146 amino acids of the *E. coli* β -galactosidase (*lacZ*) gene. On transformation with the parent plasmid of certain host strains which produce the C-terminal of the β -galactosidase enzyme encoded on the F' episome, the 2 fragments can interact and produce an active protein (Ullman *et al.*, 1967); this process is termed α -complementation. Bacteria transformed with non-recombinant vector form blue colonies in the presence of 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) on derepression of the *lacZ* gene fragments by isopropylthio- β -D-galactoside (IPTG). Insertion of the recombinant fragment into the cloning site of the vector however disrupts the *lacZ* N-terminal fragment and colonies containing recombinant plasmids are white. The structure of these plasmids is then verified by restriction analysis of minipreparations of plasmid DNA (sections 2.3.1, 2.3.7 & 2.3.8).

Bacteria transformed with vectors allowing α -complementation to be performed were plated onto LB-agar containing 100 $\mu\text{g ml}^{-1}$ X-gal in dimethyl formamide (DMF) and 50 $\mu\text{g ml}^{-1}$ IPTG in H_2O .

2.5.5.2 Screening of bacterial transformants by colony hybridisation

If α -complementation was not possible due to the nature of the construct, a modification of the method of Grunstein & Hogness (1975) was used to screen for bacteria carrying recombinant plasmids. Transformed bacterial colonies were picked using sterile toothpicks and were streaked in duplicate onto Hybond-N gridded membranes (Amersham) placed on LB-agar plates containing appropriate antibiotic. They were also streaked directly onto an LB-agar plate containing the same antibiotic (master plate). The plates were incubated at 37 °C overnight. The master plate was sealed with parafilm and stored at 4 °C. The filters were removed from the plates and placed colony side up on Whatman 3MM paper soaked in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 7 min, transferred to 3MM paper soaked in neutralisation buffer (1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 1 mM EDTA) for 3 min and to fresh neutralisation buffer for a further 3 min. The filters were rinsed in 2 X SSC, dried in an oven at 80 °C for 10 min and the DNA crosslinked to the nylon membrane by ultra-violet irradiation of 0.4 J/cm^2 using a Hybaid crosslinker. Filters were then hybridised to a radioactive probe as described in section 2.8. Following autoradiography, colonies apparently containing recombinant plasmid were picked and grown from the master plate and the structure of the plasmid confirmed by restriction analysis of minipreparations of plasmid DNA (sections 2.3.1, 2.3.7 & 2.3.8).

2.6 Construction and screening of libraries

2.6.1 Construction of cucumber genomic library in λNM1149

2.6.1.1 Insert preparation

Cucumber genomic DNA was isolated as described in section 2.3.3, up to and including precipitation with isopropanol. It was then purified on a CsCl gradient as described in section 2.3.2. The concentration of the purified DNA was estimated by electrophoresis of an aliquot through 0.8 % (w/v) agarose.

55 μg DNA was subjected to digestion with *Hind* III at 37 °C for 16 h. In order to confirm the size of the *Hind* III fragment carrying the cucumber *icl* gene, 5 μg digested

DNA was subjected to horizontal gel electrophoresis, Southern blotting (sections 2.3.8 & 2.3.9) and hybridisation with radiolabelled insert from the ICL cDNA clone (sections 2.7 & 2.8). Once the necessary size had been established, the remaining 50 µg digested DNA was subjected to horizontal gel electrophoresis under exactly replicated conditions. The relevant area of the gel was excised using a scalpel, the DNA isolated from the agarose using GeneClean glass bead solution, eluting the DNA in a total volume of 80 µl sdd H₂O. The concentration was estimated as described in section 2.3.6.

2.6.1.2 Vector preparation (λNM1149)

2.6.1.2a Ligation of cohesive ends

80 µg λNM1149 DNA were incubated for 1 h at 42 °C in 170 µl 0.1 M Tris-HCl pH 7.6, 10 mM MgCl₂, after which time 20 µl 10 X ligase buffer and 10 µl ligase (0.5 Weiss U T4 DNA ligase per µg DNA) were added and the sample incubated for 2 h at 16 °C. The sample was extracted by addition of an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 by volume), and shaken gently for 1 h. The phases were separated by centrifugation at 2,500 g and the upper phase removed to a fresh tube. The extraction step was repeated with an equal volume chloroform:isoamyl alcohol (24:1, by volume) and the DNA precipitated by addition of 0.1 volumes 3 M sodium acetate pH 7.0 and 2 volumes ethanol and incubation at room temperature for 30 min. The precipitated DNA was spooled from the solution using a glass rod, washed in 70 % (v/v) ethanol and allowed to air dry. It was then resuspended in 170 µl TE pH 8.0 at 4 °C for 16 h.

2.6.1.2b Digestion of λ vector

20 µl 10 X buffer B (BCL) were added to the DNA sample prepared as in section 2.6.1.2a and a 1 µl aliquot removed for a test packaging reaction. A 3-fold excess of *Hind* III (150 U) was added, the sample mixed and incubated at 37 °C for 2 h. The reaction was cooled by placing the tube on ice and 1.1 µl removed for the test packaging reaction to check digestion.

The two aliquots removed were packaged, serial dilutions of the packaged bacteriophage made and plated on NM621 cells (methods described in sections 2.6.1.4, 2.6.2.1 & 2.6.2.2). When the digestion has been successful, a drop of 3 orders of magnitude was seen in the packaging efficiency between the two samples.

EDTA was added to a final concentration of 5 mM to the successfully digested DNA and the sample extracted and precipitated as before. It was resuspended to a concentration of 100 $\mu\text{g ml}^{-1}$ in 10 mM Tris-HCl pH 8.3.

2.6.1.2c Treatment of digested vector with phosphatase

0.1 volumes 10 X calf intestinal phosphatase (CIP) dephosphorylation buffer (10 mM ZnCl_2 , 10 mM MgCl_2 , 100 mM Tris HCl pH 8.3) and 0.01 U CIP for every 10 μg λ DNA were added to the digested DNA and incubated at 37 $^\circ\text{C}$ for 1 h. In order to remove all CIP, 10 % (w/v) SDS and 0.5 M EDTA were added to give final concentrations of 0.5 % (w/v) and 5 mM respectively. The sample was mixed well, proteinase K added to a final concentration of 100 $\mu\text{g ml}^{-1}$ and the tube incubated at 56 $^\circ\text{C}$ for 30 min, after which time it was cooled to room temperature, extracted as before and precipitated by addition of 0.1 volumes sodium acetate pH 7.0, 2 volumes ethanol and incubation at 0 $^\circ\text{C}$ for 15 min. DNA was collected by centrifugation at 12,000 g for 10 min at 4 $^\circ\text{C}$, the pellet washed in 70 % (v/v) ethanol, dried and resuspended in TE pH 7.6 at 500 $\mu\text{g ml}^{-1}$.

2.6.1.3 Ligation of vector and insert DNA

Optimum packaging efficiencies are obtained with λ DNAs that are concatemeric. Ligations should therefore be carried out under conditions to favour concatemer formation rather than circular molecule formation containing just one *cos* site. Trial ligations were set up with 0.5 μg vector with 0, 6, 50 and 200 ng insert DNA, the last three giving molar ratios of 8:1, 1:1 and 0.25:1 respectively, 0.5 μl 10 X ligase buffer, 2 U T4 DNA ligase and *sdd* H_2O to 5 μl . All reactions were incubated at 16 $^\circ\text{C}$ for 6 h. Packaging reactions, serial dilutions and plating were carried out as described in sections 2.6.1.4, 2.6.2.1 & 2.6.2.2, and the ligation ratio giving the greatest number of plaque forming units per ml (p.f.u. ml^{-1}) determined.

Ligations were set up at the optimum molar ratio using 1 μg vector for each reaction, and all samples incubated at 16 $^\circ\text{C}$ for 5 h.

2.6.1.4 Packaging of λ DNA

Gigapack II Gold packaging extract (Stratagene) was used to package λ DNA, following the manufacturer's protocol. After the packaging reaction was completed, 500 μl phage dilution (SM) buffer (0.1 M NaCl, 10 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 50 mM Tris



HCl pH 7.5, 0.01 % [w/v] gelatin solution) and 20 μ l chloroform was added, the sample mixed gently and spun briefly to sediment the debris. The supernatant contains packaged λ particles and was stored at 4 °C.

2.6.2 Screening of recombinant bacteriophage clones

2.6.2.1 Preparation of host cells and titration of λ libraries

The *E. coli* cells relevant to the recombinant bacteriophage vector employed (λ ZAPII: XL1-Blue, λ NM1149: NM514) were grown overnight in LB-medium supplemented with 0.2 % (w/v) maltose and 10 mM MgSO₄. The cells were collected by centrifugation at 5,000 g for 10 min and resuspended in 10 mM MgSO₄ to give a final A₆₀₀ of 1. Addition of maltose to the medium induces the maltose operon which includes the *lamB* gene encoding the λ receptor on the surface of the bacterium. Plating cells were stored at 4 °C and used within 2 days of preparation.

10 fold serial dilutions of the recombinant bacteriophage λ stock to be titrated were made in SM. 100 μ l of bacteriophage were mixed with 100 μ l plating cells and incubated at 37 °C for 10 min, to allow the recombinant bacteriophage to adsorb to the *E. coli*. 3 ml molten BBL top agar (48 °C) were added, the sample mixed and poured onto 9 cm diameter BBL-agar plates. The top agar was allowed to set before the plates were inverted and incubated at 37 °C overnight. The p.f.u. ml⁻¹ concentrations of the recombinant bacteriophage were determined from the number of plaques present on the *E. coli* lawn.

2.6.2.2 Plating and screening recombinant bacteriophage libraries

Using the p.f.u. ml⁻¹ value determined by titration of the stock, known dilutions of bacteriophage were mixed with host *E. coli* cells to generate a near confluent plaque density on a bacterial lawn and screened in a similar manner to that described by Benton and Davis (1977). For the cucumber genomic library, 400 cm² plates and for cucumber λ ZAPII cDNA library (constructed by D-J Kim), 100 cm² plates were prepared respectively as detailed above (section 2.6.2.1) with proportionately larger volumes of components.

After incubation at 37 °C overnight and formation of plaques, the plates were placed at 4 °C to harden the top agar. Hybond-N filters, with notches cut to allow future correct orientation, were lowered onto the plaque containing bacterial lawn and the position of the notches marked on the plates. After 30 s the filter was removed and the process

repeated with a second filter which also carried the asymmetric pattern of notches; this filter was removed from the surface of the plate after 1 min. The filters were placed, plaque side up on 3MM filter paper soaked in denaturing solution and treated as for colony hybridisation as described in section 2.5.5.2.

The filters were hybridised with a radioactively labelled DNA fragment (section 2.7 & 2.8). The area of the bacterial lawn seen to contain a plaque hybridising to the radiolabelled probe on both filters was excised from the plate using a glass pipette and the agar plug placed in a microcentrifuge tube containing 1 ml SM buffer and 1 drop of chloroform. The tube was then shaken at 4 °C for 1 h and the resulting phage stock titrated; the whole procedure was repeated until single clonal bacteriophage, which hybridised with the radiolabelled probe, had been isolated.

2.6.3 Preparation of plate lysate stocks

For each 9 cm diameter plate, 10^5 p.f.u. were mixed with 100 µl plating cells (for λ NM1149: NM514), allowed to adsorb for 20 min at 37 °C and plated with 3 ml top agarose (as Qiagen columns were to be used for the subsequent isolation of λ DNA [section 2.3.4], top agarose was used rather than agar, as preparations from agar plates contain polyanionic contaminants that inhibit restriction endonucleases). The plates were incubated at 37 °C until the plaques were touching; 5 ml SM buffer were added to each plate and the plates gently shaken at 4 °C overnight. The SM buffer was collected from the plates and 0.1 ml chloroform was added to the pooled bacteriophage stock. The stock was centrifuged at 5,000 g for 10 min to pellet the bacterial debris and the supernatant was transferred to a fresh tube. A further drop of chloroform was added and the bacteriophage stock stored at 4 °C.

2.6.4 *In vivo* excision of recombinant λ ZAPII DNA

In λ ZAPII, the recombinant cDNA fragment is flanked by the dissected filamentous bacteriophage (f1) origin of replication (OR) present within the pBluescript SK II (-) phagemid with the initiation and termination elements of the f1 OR on either side of the cloned DNA.

Host XL1-Blue cells were grown overnight in LB-medium, harvested by centrifugation at 5,000 g for 10 min and resuspended in 10 mM MgSO_4 to an A_{600} of 1. 200 µl cells were mixed with 200 µl recombinant λ ZAPII (1×10^5 p.f.u. ml^{-1}) and 1 µl 1×10^3 p.f.u. ml^{-1} R408 filamentous f1 "helper 'phage" (Stratagene) and incubated at 37 °C for 15 min. 5 ml 2 X YT medium was added and the sample incubated at 37 °C for 3 h

with shaking. During this period the host cells become infected by both the λ ZAPII bacteriophage and the f1 R408 filamentous bacteriophage. Proteins produced by the f1 R408 helper bacteriophage recognise the f1 OR initiation site in the λ ZAPII genome and nick it at these sites leading to the production of a single-stranded (ss) copy of the DNA downstream of these sites, which includes the cDNA insert and the pBS SK II (-) sequence, until the termination signal in the f1 OR is reached. The ss DNA molecules are circularised by other f1 encoded proteins leading to the re-formation of functional ORs and hence subsequent replication of the recombinant phagemid; the DNA is packaged in coat proteins and secreted from the cell. The cultures were heated at 70 °C for 20 min to kill the host bacteria which were pelleted by centrifugation at 5,000 g for 5 min. The supernatant contains pBluescript SK (-) phagemid packaged as a filamentous bacteriophage particle. 20 μ l phagemid was used to re-infect 200 μ l XL1-Blue plating cells, incubated at 37 °C for 15 min and plated onto LB-agar plates containing ampicillin; on reinfection a ds DNA phagemid molecule is generated in the ampicillin resistant bacterial colonies formed. Single colonies were picked and streaked onto LB-agar plates containing ampicillin at 100 μ g ml⁻¹ from which single colonies were picked and used to inoculate LB-medium from which, after growth at 37 °C overnight, ds phagemid DNA could be isolated (section 2.3.1)

2.7 Radiolabelling of DNA probes

2.7.1 Labelling of ds DNA probes by random priming

Double-stranded DNA was labelled by the random priming method of Feinberg & Vogelstein (1983). 10 to 50 ng (in 34 μ l) DNA insert was heat denatured at 100 °C for 3 min and quenched on ice. To the denatured DNA was added:

5 X oligonucleotide labelling buffer (OLB)	10 μ l
BSA (10 mg ml ⁻¹)	2 μ l
[α - ³² P]dCTP	3 μ l (= 30 μ Ci)
DNA polymerase I - Klenow fragment	1 μ l (= 1 U)

and the labelling reaction allowed to proceed at room temperature overnight.

5 X OLB: 250 mM Tris-HCl pH 8.0, 25 mM MgCl₂, 5 mM β -mercaptoethanol, 2 mM each dATP, dGTP, dTTP, 1 M HEPES (adjusted to pH 6.6 with 4 N NaOH), 1 mg ml⁻¹ random hexanucleotides.

The reaction was stopped by addition of 200 μ l buffer A (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM EDTA pH 8.0, 0.5 % [w/v] SDS) and unincorporated nucleotide

removed by passage through a G-50 Sephadex column. Prior to use, radiolabelled DNA was denatured by heating at 100 °C for 3 min and cooled on ice.

2.7.2 Radiolabelling of oligonucleotides

Synthetic oligonucleotides are synthesised without a phosphate group at the 5' terminal and therefore can be labelled by the transfer of [γ -³²P]dATP to the 5' terminal using the bacterial T4 polynucleotide kinase (PNK) in the following reaction mixture:

Oligonucleotide	1 μ l (= 20 ng)
10 X PNK buffer	1 μ l
[γ - ³² P]dATP	5 μ l (= 50 μ Ci)
PNK	1 μ l
H ₂ O	2 μ l

10 X PNK buffer: 0.5 M Tris-HCl pH 7.6, 0.1 M MgCl₂, 50 mM DTT, 1 mM spermidine, 1 mM EDTA.

The reaction mix was incubated at 37 °C for 1 h and unincorporated radionucleotide removed by passage through a Biogel P-2 (Biorad) column. The labelled oligonucleotide was collected from the column in a volume of 100 μ l.

2.8 Hybridisation of radiolabelled DNA probes to membrane bound DNA or RNA

After UV crosslinking, membrane-bound nucleic acids were prehybridised in 5 X SSPE (20 X SSPE: 3.6 M NaCl, 0.2 M sodium phosphate, 0.02 M EDTA pH 7.7), 2 X Denhardtts (100 X Denhardtts solution: 2 % [w/v] BSA, 2 % [w/v] Ficoll, 2 % [w/v] polyvinyl-pyrrolidone [PVP]), 0.1 % (w/v) SDS for at least 3 h at 65 °C in glass tubes in a rotisserie oven (Hybaid).

After prehybridisation, the solution was replaced with fresh solution, the denatured radiolabelled probe (section 2.7.1) added and the filters hybridised overnight at 65 °C. The filters were washed at 65 °C for 20 min each wash in progressively more stringent, *i.e.* lower salt, conditions. After the 2 X SSPE wash, the level of background radioactivity attached to the filter was assessed using a handheld monitor, and the necessity and lengths of subsequent washes varied accordingly.

Washes: 5 X SSPE, 0.1 % (w/v) SDS
2 X SSPE, 0.1 % (w/v) SDS
1 X SSPE, 0.1 % (w/v) SDS
0.1 X SSPE, 0.1 % (w/v) SDS

After washing, the filters were partially air dried, wrapped in Saran wrap and exposed to film in autoradiography cassettes as described in section 2.2.7.

2.9 Production and analysis of transgenic plants

2.9.1 Construction of vectors for plant transformation

Insert fragments of the DNA sequence of interest were cloned into the plant transformation vector pBin19 and used to transform *E. coli* as described in sections 2.5.1 to 4. Recombinant plasmids carrying the desired fragments were identified by restriction analysis of minipreparations of plasmid DNA (sections 2.3.1, 2.3.7 & 2.3.8).

2.9.2 Transfer of recombinant plasmids into *Agrobacterium tumefaciens* by conjugation and screening of transformed bacteria

The conjugations that are required to transfer the plasmid of interest from *E. coli* to *Agrobacterium tumefaciens* are carried out by a triparental mating. This involves the donor *E. coli* carrying the recombinant pBin19 plasmid, the recipient *A. tumefaciens* and *E. coli* HB101 carrying a 'helper' plasmid, pRK2013. The helper plasmid provides mobilisation and transfer functions in *trans*, which act on a specific origin of transfer and an activation site in the cloning vector. The method used is described in Draper *et al.* (1988).

Single colonies of donor and helper strains of *E. coli* and *A. tumefaciens* LBA4404 were picked and used to inoculate 5 ml LB-medium containing 100 µg ml⁻¹ kanamycin for the two *E. coli* strains and 100 µg ml⁻¹ rifampicin and 300 µg ml⁻¹ streptomycin for *A. tumefaciens*. The *E. coli* strains were incubated at 37 °C and the *A. tumefaciens* strain at 28 °C overnight with shaking. 100 µl of each of the three cultures was pipetted onto an LB-agar plate, mixed by spreading with a glass rod and incubated at 28 °C overnight. Duplicate plates were made for each pBin19 construct. Two streaks of cells were removed from each plate, resuspended in 500 µl LB-medium and plated onto LB-agar plates containing 100 µg ml⁻¹ kanamycin, 100 µg ml⁻¹ rifampicin and 300 µg ml⁻¹

streptomycin. The plates were inverted and incubated at 28 °C for 4 days, during which time single colonies of *A. tumefaciens* grew. These were used to inoculate 5 ml LB-medium containing antibiotics as detailed above and DNA isolated as described in section 2.3.5. Colonies were screened by restriction digestion and Southern hybridisation (sections 2.3.7, 2.3.8, 2.3.9 & 2.8).

2.9.3 *Nicotiana plumbaginifolia* transformation and regeneration

2.9.3.1 General conditions

Cultivation of all genetically manipulated plants was carried out according to the code of practice of the local Genetic Manipulation Safety Committee. All tissue culture operations were carried out in a laminar flow cabinet using standard aseptic technique. Plant material in culture was kept in growth rooms with 24 h illumination from white fluorescent tubes (Thorn 3500) providing an irradiance of 30 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at a constant temperature of 25 °C.

2.9.3.2 Preparation of plant material

N. plumbaginifolia seeds were germinated and grown as described in section 2.1.1. Fully expanded leaves were excised, washed briefly in 70 % (v/v) ethanol and sterilised in a 4 % (v/v) solution of sodium hypochlorite for 15 min. The leaves were rinsed six times in sterile distilled water. Leaf pieces were cut, avoiding the mid rib and major veins, and placed in 100 ml shooting medium (section 2.2.3) without agar.

2.9.3.3 Preparation of transformed *A. tumefaciens* for inoculation of leaf pieces

A single colony of *A. tumefaciens* LBA4404 carrying the recombinant pBin19 plasmid of interest was used to inoculate 5 ml LB-medium containing 25 $\mu\text{g ml}^{-1}$ kanamycin; it was grown with shaking at 300 rpm for 24 to 48 h at 28 °C.

2.9.3.4 Inoculation of leaf pieces and regeneration of transformed plants

2 ml of *A. tumefaciens* culture were added to the leaf pieces in shooting medium and left at room temperature for 15 min with intermittent mixing. After this time, leaf pieces

were transferred onto shooting medium plates (0.8 % [w/v] agar) and cultured at 25 °C as detailed in section 2.9.3.1 for 2 days. The pieces were transferred onto shooting medium plates containing 500 µg ml⁻¹ carbenicillin, 200 µg ml⁻¹ kanamycin and culturing continued under the same conditions. The pieces were transferred onto fresh shooting medium plates every 3 weeks until shootlets began to develop, approximately 5 to 6 weeks after inoculation. Once large enough to be excised, developed shootlets were transferred onto plates of expansion medium (section 2.2.3) containing 250 µg ml⁻¹ carbenicillin and 200 µg ml⁻¹ kanamycin. After 2 weeks, shootlets that continued to expand were transferred into pots containing rooting medium with 100 µg ml⁻¹ carbenicillin and 100 µg ml⁻¹ kanamycin. Shootlets which developed roots in the selection medium were transferred to Levington Universal potting compost and incubated at 25 °C in a 16 h photoperiod at an irradiance of 100 µmol m⁻² s⁻¹. High humidity was initially maintained by covering the pots with plastic bags, which were gradually pierced to allow accustomisation to the growth room environment and eventually removed after approximately 2 weeks. Growth was continued until the plants produced seed.

2.9.3.5 Production and collection of seed

N. plumbaginifolia plants were allowed to self-fertilise and seed was collected from each plant and stored at room temperature.

2.9.4 Analysis of transgenic *N. plumbaginifolia* plants

2.9.4.1 Germination of transgenic *N. plumbaginifolia* seeds

Seeds were surface sterilised by treatment with 10 % (v/v) sodium hypochlorite for 10 min, after which time the solution was removed and the seeds rinsed 6 times with sdd H₂O. Seeds were imbibed in 1 mM gibberellic acid overnight at 4 °C, rinsed the following day in sdd H₂O and transferred to rooting medium plates without antibiotic. The plates were incubated in a 16 h photoperiod at 25 °C.

2.9.4.2 Selection of germinating seed at specific developmental stages

Germination of *N. plumbaginifolia* occurs in a non-synchronous manner; therefore it was necessary to select seedlings at specific stages for analysis. Stages were selected as

depicted in figure 4.5 using a binocular microscope and stored at -70 °C, so that complete sets of assays with all stages represented could be performed.

2.9.4.3 Fluorometric assay of β -glucuronidase (GUS) in germinating seedlings

The assay was performed as described in Jefferson (1987). GUS is able to cleave a number of glucuronides and this assay takes advantage of its ability to cleave 4-methylumbelliferyl β -D-glucuronide (MUG), a non-fluorescent substrate, to yield glucuronic acid and 7-hydroxy-4-methylcoumarin (4-methylumbelliferone, MU), a fluorescent product that is maximally fluorescent in the presence of a basic buffer.

40 seeds from each stage post imbibition were harvested into liquid nitrogen. Once all stages had been collected, the seeds were homogenised in 100 μ l GUS extraction buffer (50 mM sodium phosphate pH 7.0, 10 mM β -mercaptoethanol, 0.1 % [w/v] Triton X-100, 1 mM EDTA). Cell debris was pelleted by centrifugation at 12,000 g for 2 min, and 50 μ l clarified extract was used per GUS enzyme assay. The remaining extract was frozen in liquid nitrogen and later used for protein concentration estimation (section 2.11).

50 μ l of extract was diluted with a further 750 μ l GUS extraction buffer and prewarmed at 37 °C. To start the reaction, 200 μ l 5 mM MUG in GUS extraction buffer at 37 °C was added to the extract, giving a final MUG concentration of 1 mM. After mixing, 200 μ l was immediately removed to a tube containing 800 μ l 'stop buffer' (0.2 M Na_2CO_3) and the remaining reaction mix incubated at 37 °C. Further time points were taken at 5, 10, 15 and 30 min for seedling sets showing high levels of GUS activity or at 30, 60 and 120 min for those showing lower levels.

The concentration of MU, the fluorescent product of the reaction, was determined fluorometrically using a Perkin-Elmer LS series spectrofluorimeter with settings of excitation at 365 nm and emission at 455 nm. Readings taken with a spectrofluorimeter are relative fluorescence, so for each use, the spectrofluorimeter was calibrated with standards of MU such that 500 units relative fluorescence was equivalent to 500 nM MU. Thus relative fluorescence of the samples could be read directly as nM MU. The rate of the reaction was calculated from the change in MU concentration with time and the results presented as fKat MU seedling⁻¹.

2.9.4.4 Histochemical assay of β -glucuronidase

Histochemical staining was carried out using a modified version of the method described by Jefferson (1987). Whole seedlings which had been grown for 3 days (stages 3 to 4) post imbibition were incubated in 1 mg ml⁻¹ 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc: prepared by dissolving the powder in DMF at a concentration of 100 mg ml⁻¹ with subsequent dilution to 1 mg ml⁻¹ in 50 mM sodium phosphate) at 37 °C for between 30 min and 16 h, depending on the degree of GUS activity. Prior to incubation, the testa and endosperm were removed from the seedling using fine forceps. After staining, the seedlings were transferred to 50 mM sodium phosphate and stored at 4 °C.

2.9.4.5 Treatment of transformed *N. plumbaginifolia* seedlings with sucrose

Seeds were germinated as described in section 2.9.4.1 and grown for 18 days. They were transferred to petri dishes containing filter paper soaked in distilled H₂O or 25 mM sucrose and incubated either in the light (conditions as in section 2.9.4.1), or in the same growth room but wrapped in aluminium foil and black polythene bags. After 5 days, 40 seedlings from each of the four treatments were harvested into liquid nitrogen and assayed for GUS activity as described in section 2.9.4.3.

2.9.4.6 Northern analysis of germinating transgenic *N. plumbaginifolia* seeds

Seeds were germinated and grown for varying periods (figure 4.3) up to 7 days post imbibition as described in section 2.9.4.1. They were harvested into liquid nitrogen, RNA isolated and analysed by Northern blotting (sections 2.4.2 & 2.4.4)

2.10 Transient expression in cucumber protoplasts

2.10.1 Construction of vectors for transient expression

Insert fragments of the DNA sequence of interest were cloned into the vector pRAJ275 into which a *Nos* terminator had also been inserted. The recombinant plasmids were used to transform *E. coli* (sections 2.5.1 to 4). Recombinant plasmids carrying the desired fragments were identified by restriction analysis of small scale preparations of

plasmid DNA (sections 2.3.1, 2.3.7 & 2.3.8). Large scale preparation of recombinant plasmid DNA were carried out as described in section 2.3.2, concentration of DNA determined and each plasmid diluted to a concentration of 2.5 mg ml⁻¹.

2.10.2 Isolation of cucumber protoplasts

Cucumber plants were grown as described in section 2.1.1 in Levingtons Universal potting compost until the second true leaf was 10 x 10 cm². Second true leaves were excised, washed in 70 % (v/v) ethanol for 30 s and transferred to 2 % (v/v) sodium hypochlorite containing 0.05 % (v/v) Tween 80, for 15 min. The leaves were washed six times in sdd H₂O. Leaves were sliced into approximately 5 x 5 mm² and placed immediately in petri dishes containing 0.35 M D-mannitol pH 5.6, 0.6 % (w/v) cellulase onozuka R10 (R. W. Unwin & Co. Ltd., Welwyn, Herts), 0.35 % (w/v) macerozyme R10 (R. W. Unwin & Co. Ltd.) and 10 µg ml⁻¹ tetracycline. The petri dishes were sealed, wrapped in aluminium foil and incubated for 12 to 14 h at 25 °C. The solution was removed from the petri dish using a wide bore pipette, avoiding the leaf tissue, and passed through 2 layers of muslin into a 50 ml sterile tube. Approximately 1 ml 0.35 M D-mannitol was added to the petri dish, the protoplasts released from the leaf pieces by gentle mashing and the solution transferred as above to a sterile tube. A further 10 to 20 ml 0.35 M D-mannitol was added to the petri dish to collect further protoplasts and pooled with the previous solution. All tubes were centrifuged at 100 g for 5 min. The supernatant was removed, the protoplasts gently resuspended in a small volume of 0.35 M D-mannitol and the volume made up to 40 to 50 ml with 0.35 M D-mannitol. The centrifugation step was repeated, and the whole washing procedure repeated 4 to 5 times, until the supernatant was virtually colourless. After the final wash, the protoplasts were resuspended in a small volume of 0.35 M D-mannitol, counted using a haemocytometer slide and diluted to give a concentration of 1 x 10⁶ protoplasts per 0.6 ml.

2.10.3 Electroporation and culturing of cucumber protoplasts

Cell membranes, when exposed to high-intensity electric field pulses, can be temporarily destabilised in specific regions of the cells. During the destabilisation period, transient pores in the cell membrane allow entry of exogenous macromolecules such as DNA present in the surrounding medium (Wong & Neumann, 1982; Fromm *et al.*, 1985).

Electroporation of cucumber protoplasts was carried out using a PG201 ProGenetor II (Hoefer Scientific Instruments) combined with a PG250 safety chamber and PG220P-5 ring electrode, designed for use with 16 mm wells of 4 well culture trays.

0.6 ml (1×10^6) protoplasts per well were transferred to 4 well plates on ice. 100 μ g sterile herring sperm DNA and 25 μ g recombinant plasmid DNA were added, the sample gently mixed and electroporated immediately. Electroporation conditions were optimised using a construct containing the cauliflower mosaic virus 35S promoter fused to the gene encoding β -glucuronidase. The conditions selected were: 1 pulse of 340 V, giving a field strength of 567 V cm⁻¹, 760 μ F capacitance and 10 msec. pulse time. These conditions, using the timed discharge function, simulate a square wave pulse which yields better transfection efficiencies with plant protoplasts than does exponential decay.

After electroporation, the protoplasts were left on ice for 10 min without movement, to allow the nucleic acids time to move across the plasmalemma and also to maintain the viability of the protoplasts. 0.5 ml protoplast culture medium (with or without sucrose, section 2.2.3) were slowly added, the sample mixed gently and transferred to a sterile 15 ml screw top tube containing 5 ml culture medium. The plate well was washed with 0.5 ml culture medium, which were pooled with the rest of the sample. All tubes were firmly closed and cultured horizontally for 48 h at 25 °C in the dark.

2.10.4 Analysis of transfected protoplasts

2.10.4.1 Fluorometric assay of β -glucuronidase in transfected cucumber protoplasts

The transfected protoplasts were pelleted at 1,000 g for 10 min and all but 50 μ l of the culture medium removed. The protoplasts were resuspended in GUS extraction buffer (section 2.9.4.3) and transferred to a microcentrifuge tube containing approximately 100 μ l glass beads (Sigma: 150 - 212 microns) and vortexed for 1 min. The debris and glass beads were pelleted by centrifugation at 12,000 g for 2 min and 100 μ l of the clarified extract transferred to a fresh Eppendorf containing a further 700 μ l GUS extraction buffer. The rest of the assay was carried out exactly as for transformed *N. plumbaginifolia* seedlings (section 2.9.4.3) with time points taken at 0, 30, 60 and 120 min after addition of the substrate.

2.10.4.2 Northern analysis of transfected protoplasts

Protoplasts were harvested, RNA isolated as described in section 2.4.3 and analysed by Northern blotting (section 2.4.4).

2.11 Protein concentration estimation

The protein concentration of plant extracts was determined by the method of Bradford (1976) using the Bio-Rad protein kit and a range of known BSA concentrations as standards (0 to 20 $\mu\text{g ml}^{-1}$). Extracts were diluted with distilled H_2O to a final volume of 800 μl . 200 μl concentrated protein binding solution was added, the sample mixed well and incubated at room temperature for at least 5 min. The absorbance at 595 nm was determined using a Beckman DU-62 spectrophotometer. All protein concentration values were determined in duplicate.

CHAPTER 3

THE ISOCITRATE LYASE GENE OF CUCUMBER

3.1 Introduction and aims

Previously a cDNA clone from *C. sativus* encoding part of the ICL polypeptide had been isolated in W. Becker's laboratory (University of Wisconsin), but it had not been characterised. Both cDNA and genomic clones encoding malate synthase (MS) have previously been isolated and fully characterised and 1089 bp of the upstream region shown to be sufficient to direct faithful transcription in transgenic *Nicotiana plumbaginifolia*. (Graham *et al.*, 1989, 1990). The aim of this project was to isolate both full length cDNA and genomic clones encoding ICL, to fully characterise them by restriction endonuclease digestion and DNA sequencing, in order to enable a study of the structure of the gene and the deduced amino acid sequence to be undertaken. Comparison with the genes and polypeptides of other organisms could then be carried out. In addition, a comparison of the regulatory sequences of the *ms* gene with those of the *icl* gene would be undertaken, to provide the foundation for further studies to analyse the elements responsible for the control of gene expression of this gene during plant development.

3.2 Determination of gene copy number

In order to determine the copy number of the gene encoding isocitrate lyase (ICL) in cucumber, Southern blot analysis was carried out. Genomic DNA was isolated from young leaves as described in section 2.3.3 and 5 µg aliquots were digested with the restriction endonucleases *Eco* RI, *Hind* III and *Xba* I (section 2.3.7). The digested DNA was subjected to horizontal agarose gel electrophoresis and transferred to Hybond N as described in sections 2.3.8 and 2.3.9. The membrane was hybridised firstly with the 1.4 kb insert of the cDNA clone, pBSICL, radiolabelled with [³²P]dCTP (figure 3.1a), and subsequently, following removal of the first probe, with the 800 bp *Eco*-*Xba* fragment from the 5' end, similarly radiolabelled (figure 3.1b). The simple banding pattern generated indicated that only a single gene showed extensive homology to the cDNA probe used. The occurrence of two bands in the *Xba* I digested lane indicates the presence of a *Xba* I site within the gene. This was confirmed by the second hybridisation, in which only a single band was observed in this lane. The malate synthase gene of cucumber is also present in a single copy (Graham *et al.* 1989). This is not the case in other species, for example, *B. napus*, where there are at least four classes of *ms* genes (Comai *et al.*, 1992) and approximately six *icl* genes falling into two classes (Zhang *et al.*, 1993), and *Glycine*

(a)

(b)

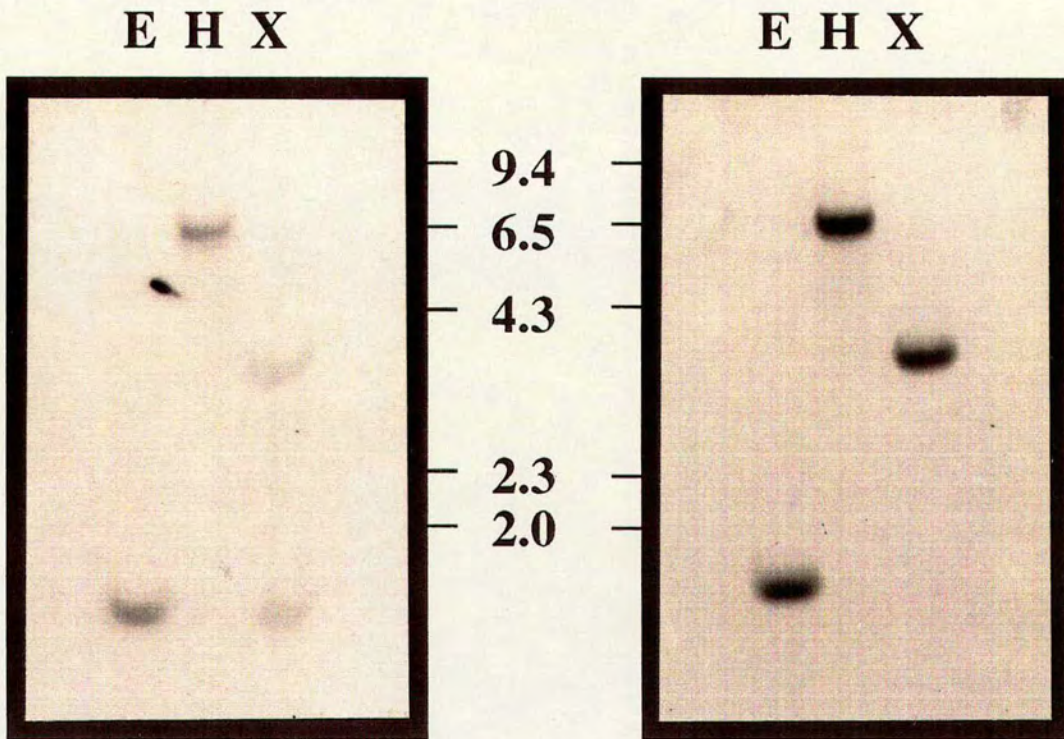


Figure 3.1 Southern blot analysis of cucumber genomic DNA digested with *Eco* RI (E), *Hind* III (H) or *Xba* I (X), fractionated through 0.8 % (w/v) agarose and probed with (a) the 1.4 kb insert of pBSICL and (b) the 800 bp *Eco* - *Xba* fragment from the 5' end. Size markers (kbp) are λ DNA digested with *Hind* III.

max in which there are at least two genes encoding each of these enzymes (F. Widmer, pers. comm.). Study of these genes and the regulation of their expression in cucumber therefore avoids the complication of studying the regulation of a multigene family as is the case in other species.

In previous investigations, two subunits for ICL have been detected on protein gels of immunoprecipitated products of an *in vitro* translation (Reizman *et al.*, 1980; Weir *et al.*, 1980). Production of two subunits from a single gene could be the result of post-transcriptional processing. There have been reports of ICL being phosphorylated in *E. coli* (Hoyt & Reeves, 1988; Robertson & Reeves, 1989), *S. cerevisiae* (Lopez-Boado *et al.*, 1988) and recently in cucumber and castor bean (Finnessy *et al.*, 1994). Phosphorylation of a polypeptide would be sufficient to cause this difference in subunit size. The importance of phosphorylation in the regulation of ICL activity in higher plants is not yet clear.

3.3 Characterisation of the cDNA clone, pBSICL

3.3.1 Restriction endonuclease analysis

Digestion of the cDNA clone with various restriction endonucleases revealed that it was bounded by *Eco* RI sites and contained *Sac* I and *Xba* I sites approximately 200 and 800 bp from the 5' end respectively, agreeing with the data from Southern blotting (section 3.2) that the gene contained an *Xba* I site. The restriction map is displayed in figure 3.2.

3.3.2 Nucleotide sequence analysis

The cDNA clone was fully sequenced using the Sanger dideoxy chain termination procedure (Sanger *et al.*, 1977) employing the universal -40 and reverse primers. Subclones were created using the internal *Xba* I and *Sac* I sites, generating four subclones carrying inserts of 800 and 600 bp (*Xba* I clones) and 200 and 1200 bp (*Sac* I clones). The strategy for sequencing is shown in figure 3.2. The cloned insert was shown to be 1421 nucleotides in length, encoding 473 amino acids of the ICL polypeptide. Comparison of the nucleotide sequence with that from the previously published cotton ICL cDNA sequence (Turley *et al.*, 1990) revealed that the clone lacked 86 and 221 nucleotides at the 5' and 3' ends of the protein coding sequence respectively, corresponding to 29 and 74 amino acids (figure 3.3d).

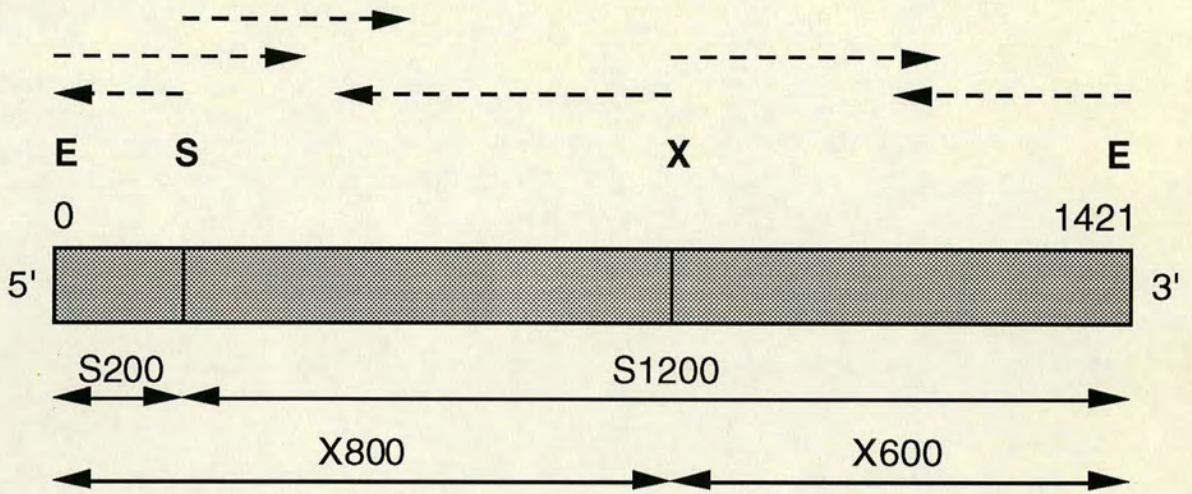


Figure 3.2 Restriction endonuclease map and strategy for sequencing the cucumber cDNA clone, pBSICL. Restriction endonuclease map of the 1.4 kbp insert of the cDNA clone, pBSICL. Restriction enzymes used were: E = *Eco*RI, X = *Xba*I, S = *Sac*I. Four subclones created using these sites are depicted by double headed arrows. The extent of the sequence information obtained from each clone is depicted by dotted arrows.

Figure 3.3

(a) Restriction endonuclease map of a 0.5 kbp genomic fragment in λ NM1149 containing the *icl* gene. Restriction enzymes used were: H = *Hind* III, X = *Xba* I, S = *Sal* I, N = *Nco* I, E = *Eco* RI, C = *Sac* I.

Broad line represents the extent of the nucleotide sequence determined.

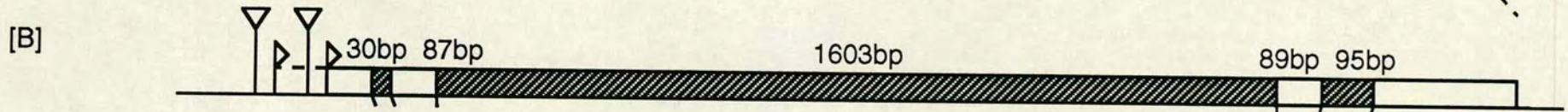
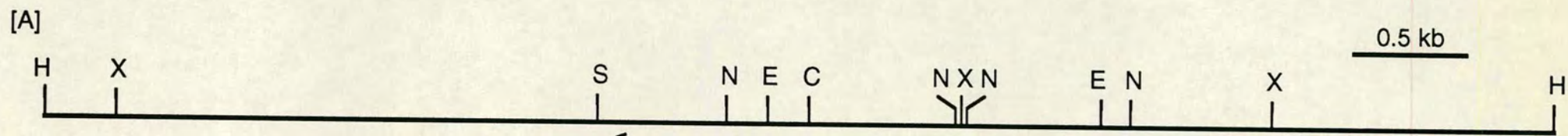
(b) Diagrammatic representation of the *icl* gene structure. Hatched boxes represent the protein coding region, open boxes represent 5' and 3' untranslated regions and introns. The vertical lines topped by a triangle represent putative TATA boxes (see figure 3.5) and the predicted transcription initiation sites are represented by flags.

(c) Mature ICL mRNA. The AUG start and UAA termination codons delimit a 1728 bp ORF (hatched box). The known polyadenylation site, as determined by sequence analysis of a cDNA clone, is indicated by 'polyA'.

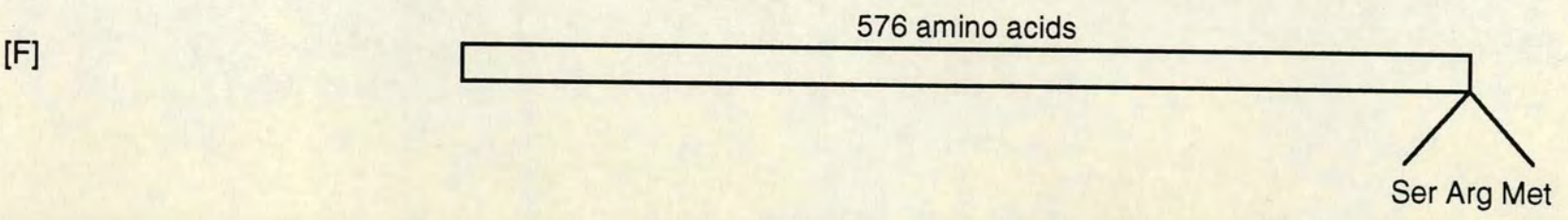
(d) 1421 bp cDNA, pBSICL, lacking 86 and 221 nucleotides of the coding region at the 5' and 3' ends respectively.

(e) 1768 bp cDNA, pBSICL1.7, lacking 235 nucleotides of the coding region at the 5' end. This cDNA has a 3' untranslated region of 275 bp and a poly(A) tail of 25 nucleotides.

(f) ICL protein with predicted size as indicated. The putative type 1 peroxisomal targeting signal Ser-Arg-Met is indicated at the carboxy terminal end.



70



3.4 Isolation of a partial cDNA clone encoding isocitrate lyase

In order to gain information about the 3' end of the gene (both coding and untranslated regions), further cDNA clones encoding ICL were isolated from a λ ZAPII library constructed using mRNA template extracted from senescing cucumber cotyledons by Dae-Jae Kim (University of Edinburgh). Eight clones were isolated, and following *in vivo* excision and subsequent restriction endonuclease digestion, were all shown to carry inserts of approximately 1.7 kb.

Three clones were picked and the 5' and 3' ends sequenced using the universal -40 and reverse primers. All clones started at the same position, 235 bp downstream of the start of the coding region and terminated 275 bp downstream of the translation termination codon, carrying a 25 nucleotide poly A tail (figure 3.3e). They are therefore presumably sibling clones and will subsequently be referred to as pBSICL1.7.

3.5 Isolation and sequencing of the isocitrate lyase gene

A cucumber genomic library was constructed from *Hind* III digested leaf DNA as described in section 2.6.1. It was plated and screened for clones carrying the *icl* gene using the radiolabelled 1.4 kb insert from pBSICL as a probe. Ten phage were purified, DNA isolated from three of them and the inserts subcloned into pBS and analysed by restriction endonuclease digestion. All three clones carried a *Hind* III insert of approximately 6.5 kb and generated the same restriction map, though one contained the insert in the opposite orientation relative to the other two. The restriction map is shown in figure 3.3a.

In order to sequence the genomic clone, subclones were generated as depicted in figure 3.4. Deletion series of the longer subclones were constructed using the *Exo* III/S1 nuclease system and the nucleotide sequence of all clones determined using the Sanger dideoxy chain termination procedure. The sequence is displayed in figure 3.5. Comparison with the sequence determined from pBSICL shows that the two differ in two places, leading to one amino acid substitution in each case (see figure 3.9a). The library from which the cDNA clone was isolated was constructed from mRNA isolated from a different cultivar to that used throughout the rest of this thesis. Consequently the differences in sequence seen could merely be due to the use of different cultivars. In addition to the protein coding sequence, the nucleotide sequence of 2.7 kbp of the 5' and 400 bp of the 3' flanking regions was determined.

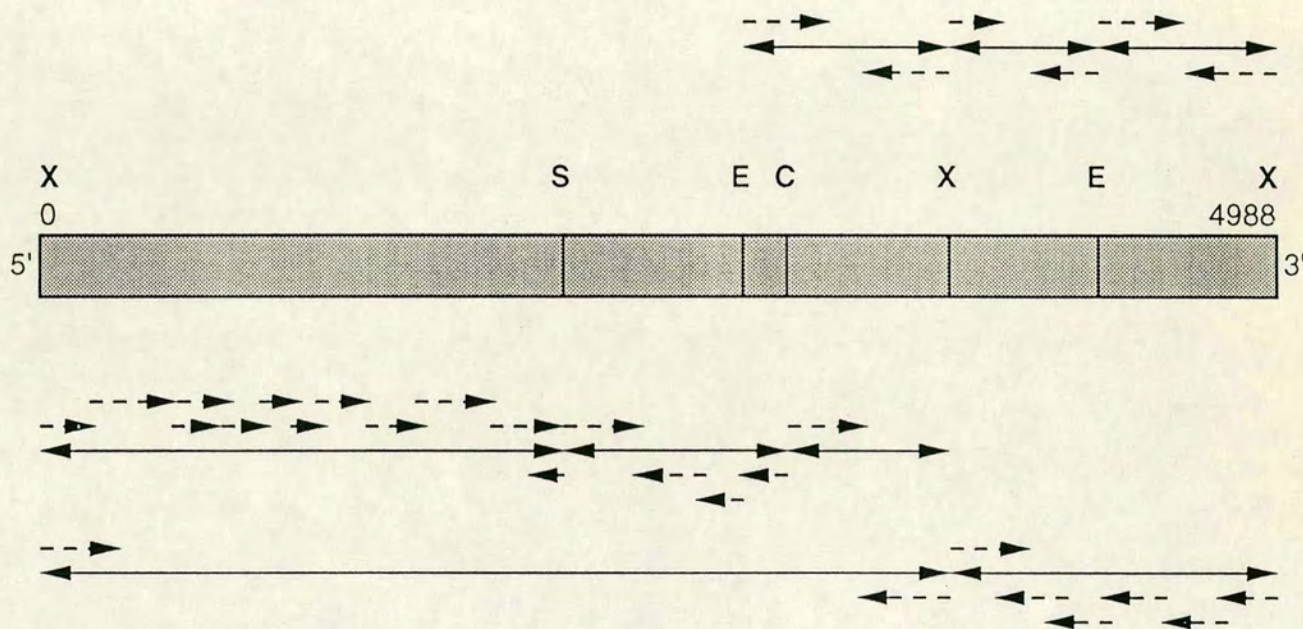


Figure 3.4 Strategy for sequencing the *icl* genomic clone. Restriction endonuclease map of the 4988 bp of the genomic clone for which the nucleotide sequence was determined. Sites shown are X = *Xba* I, S = *Sal* I, E = *Eco* RI, C = *Sac* I. Subclones utilising these sites were created as indicated by solid double headed arrows. Extent of the sequence information obtained is indicated by dotted arrows.

-2659

tctagatttacatgctatatagaaagatggttgaaaaacaacacttgaatatggttaggaagggaagtagtaaaacaaa
aaggatatacaacattaaaaatgatgtggatgaggtctaggaacaaatgaagtaaatgaagaagagaatgtag
aaaggttttgagatactcctcgaagtgtcaaaaaaagggttaaccaatctaaggggatggacaaggctgcaattcaat
ggcactcttggccccatttccaccgcccttgaagataaccaattttattctcacacttcccgaggactaactaca
tcaaaccgatttttagtccatccatttccatttttaggcattcatttttttaaaaaaaactttcatcattttactta
aataaatagataacaacaaaagtacgttacacctgtaacaaaagctatatgacttcaccaaatataacaacacttttc
aataagaaccttaactaaaggatgctcaactatagttggcttaacaaactatcaactttattccataagttgaaaag
gaccatttatcaattttattagtaaaacaatgaacaagattcatgcaaatataaaactacaattatttactgtgtgt
agaagatccacggctttgaaagtaaaattgatacgacctcaatatttaaacattttgataaaaaatggtcatatctaa
attgaatataaaaaacatttaaaaaatttaactaattgagtagttgttctacatattgtagatgaatcatattatgt
aggaagcagttcctatatatttctgttcaaccattgtaaggttctcaaaaagagccttaaatgaaaaaatatttattgtt
ttattcctgatataagaataaatttgacaagataaatacaataataataaaaaaagaattaggaatgcaatgaaaa
gttattttgattcataatagaagtagctatggttggaagcgggtgggaaggtgtgtgtgatgtaaatcaagttgtt
atcgggtaagttcatagggatattataaaggagtagtcatggttgacgcagcctctcttcttcgatccctccac
cccctacggatcatttaatttcaatttcaatttccatactctatccaaatttgaaatactttgagggcttttttctaat
acatacgccaatttttcaattcaagatagattgaatttttgaacacacaccccaaccttatccatttccatcctctttt
aattaaccatcaactcctccaaaccccaactatgtattttatataattttctcccctcttagttatgcccgaatata
aaagagtaccctgattttcaattcacttaaatatctctatcattacgtcctttatgacgtcattggatgatttaag
gataaaatagagaactaaatatctttttgttgggtaaaataagagatgatgagtaattactgttatttttatatgcatc
actaattgaatttttatgggtaattatataatgatgtttcaaatattcattgaacttgaatatcgatatcaaaaccttat
aatgatttattttgtcaattattgactataatcaatataatctttgtttgtttcccctcaaagttatagtataa
attttaattttgataatgatcctagatttttttaaaaaaatatttctcttaaaaaatagataaatgaacttgaaga
attagaagaatattggaatctacaatatttaacttaattgattctactatacgtaaaaaatgttaactgaaatgat
aaacaagatttgaaggatagaagagagagaataaaaaaatcataggttaggggaacgtttaagttataagaaaaaata
atgcaaatatggtatctgataatcgcgaaacttgaatggttacctttatgatttcaattacgacacatcacaca
tcacacatccatcttaaaactgccttcaactgtgtcgcacaaataattacgacacgtcggctccttacatttgaaggagg
agattccgatgaaatcttctgcatttctcatcagactgcctattttgaccacagaaaaactttaccgtatgttatac
acataaattcttagcttatacagaataagaatcatatttaacaaaaaaaccgtaaaaactattaccgaagattatg
gttattgagaagatgagagaagacaagatggcagcagagagagaggatgggttcccttacgtgcccgagttgc
acctcaaacagacaagttccatcaactctgaagataacaatgaacaaaattcaaatgtattttctttgtaacttg
ttttcttggttctgtagctctcttttcttatcaaatccataattaacactcttttacctatctcaagggttctga
tcattctctccccttataaatacaaccccatcaacaaactcttctgcattgagttgattggcttctttctgtattgtcga

-52
ggaatatagaactaaagaacctaaagaagcagatataaactcagctctaccATG GCT GCC TCC TTC TCT GTT
M A A S F S V

22
CCT TCG ATG gtgaatacaaatccatttgaattttgctgtaatgcaggtactaataataactaataactaattttt
P S M

98
tagatttcttgttggatagATA ATG GAA GAA GAA GGA AGA TTT GAA GCG GAA GTT GCA GAG GTT
I M E E E G R F E A E V A E V

163
CAG GCA TGG TGG AAT TCA GAG AGG TTC AAG CTA ACA CGA CGG CCG TAC ACG GCA AAA GAC
Q A W W N S E R F K L T R R P Y T A K D

223
GTG GTG TCA CTA CGT GGG AGC CTG AGA CAA AGC TAC GCT TCA AAC GAT TTA GCT AAG AAG
V V S L R G S L R Q S Y A S N D L A K K

283
CTG TGG CGA ACC CTC AAA ACC CAC CAA GCT AAC GGC ACA GCC TCG AGG ACA TTC GGA GCT
L W R T L K T H Q A N G T A S R T F G A

343
CTT GAC CCT GTT CAA GTG ACG ATG ATG GCT AAG CAT TTG GAC ACC ATT TAT GTC TCC GGT
L D P V Q V T M M A K H L D T I Y V S G

403
TGG CAA TGT TCT TCA ACT CAC ACC TCC ACT AAC GAA CCT GGC CCG GAC CTC GCC GAT TAC
W Q C S S T H T S T N E P G P D L A D Y

463
CCA TAC GAC ACT GTT CCG AAC AAG GTC GAG CAT CTT TTC TTC GCT CAA CAG TAC CAT GAC
P Y D T V P N K V E H L F F A Q Q Y H D

523
AGG AAA CAG AGG GAA GCC AGG ATG AGT ATG AGT CGA GAG GAA CGA GCC AAA ACT CCT TAC
R K Q R E A R M S M S R E E R A K T P Y
583
ATC GAC TAT CTC AAA CCG ATT ATC GCC GAC GGT GAC ACC GGC TTC GGC GGC ACC ACC GCT
I D Y L K P I I A D G D T G F G G T T A
643
ACA GTG AAG CTC TGT AAG CTA TTT GTG GAA CGC GGT GCC GCC GGT GTT CAT ATT GAG GAT
T V K L C K L F V E R G A A G V H I E D
703
CAA TCC TCT GTT ACG AAG AAA TGC GGT CAT ATG GCG GGA AAA GTG TTG GTA GCG GTT AGC
Q S S V T K K C G H M A G K V L V A V S
763
GAA CAC ATC AAT CGG CTC GTA GCG GCT CGA TTG CAG TTC GAT GTG ATG GGG GTT GAG ACG
E H I N R L V A A R L Q F D V M G V E T
823
ATT CTC GTG GCT CGA ACG GAT GCG GTT GCA GCT ACT TTG ATT CAA ACG AAT GTT GAT AAA
I L V A R T D A V A A T L I Q T N V D K
883
CGA GAC CAT CAA TTT ATT TTG GGG GCG ACG AAC CCA AAT TTG AGA GGG AAG AGT TTA GCT
R D H Q F I L G A T N P N L R G K S L A
943
GGG GCT TTA GCA GAA GCC ATG GCG GCT GGA AAA ACA GGA GCG GAG CTT CAA GCT CTA GAA
G A L A E A M A A G K T G A E L Q A L E
1003
GAT CAA TGG ATT TCC ATG GCC CAA TTG AAA ACA TTT TCG GAA TGT GTA ACG GAC GCA ATC
D Q W I S M A Q L K T F S E C V T D A I
1063
ATG AAT ACA AAC GCA ACA GAG AAT GAA AAG AGG AGG AAA TTA GAC GAA TGG ATG AAC CAT
M N T N A T E N E K R R K L D E W M N H
1123
TCC AGT TAC GAG AAA TGT ATA TCG AAC GAA CAA GGA CGT GAA ATC GCA GAG AAA TTA GGA
S S Y E K C I S N E Q G R E I A E K L G
1183
CTG AAG AAT CTC TTC TGG GAT TGG GAT TTA CCA AGA ACT AGA GAA GGG TTT TAC AGA TTC
L K N L F W D W D L P R T R E G F Y R F
1243
AAA GGT TCG GTA ATG GCG GCG ATT GTT CGT GGT TGG GCT TTC GCA CCA CAT GCA GAC TTA
K G S V M A A I V R G W A F A P H A D L
1303
ATC TGG ATG GAA ACT TCA AGC CCA GAT TTG GTT GAA TGC ACA ACA TTC GCA AAA GGG ATG
I W M E T S S P D L V E C T T F A K G M
1363
AAA TCC ATA CAT CCA GAA ACA ATG TTA GCT TAT AAT CTA TCC CCA TCT TTC AAC TGG GAT
K S I H P E T M L A Y N L S P S F N W D
1423
GCA TCA GGA ATG AGC GAT AAG CAA ATG GAA GAG TTC ATC CCG AGG ATC GCG AGG CTG GGT
A S G M S D K Q M E E F I P R I A R L G
1483
TTC TGC TGG CAG TTC ATA ACA CTG GCG GGA TTC CAC GCG GAT GCG TTG GTG GTG GAC ACA
F C W Q F I T L A G F H A D A L V V D T
1543
TTT GCA AGA GAT TAC GCG AGA AGG GGA ATG TTG GCA TAC GTG GAG AGA ATT CAA AGA GAA
F A R D Y A R R G M L A Y V E R I Q R E
1603
GAA AGA AAC AAT GGA GTG GAT ACA TTG GCA CAT CAG AAA TGG TCT GGT GCT AAC TAC TAC
E R N N G V D T L A H Q K W S G A N Y Y
1663
GAT CGG TAT CTG AAG ACG GTT CAG GGG GGG ATT TCG TCC ACT GCG GCC ATG GGA AAA Ggc
D R Y L K T V Q G G I S S T A A M G K
1723
aagcaaatgtttatcattagatgtttgatattaatgttaaaaaactgttgaagtgtgtttatcaagtgtttgaattttg

1802

aaatgcag GG GTA ACG GAA GAG CAG TTC AAG GAG TCA TGG ACG AGG GAA GGA GCA GTG AAC
G V T E E Q F K E S W T R E G A V N

1863

TTG GGA GAA GAA GGG AAC GTG GTG GTA GCC AAG TCC AGA ATGtaatgaagagcaaaagcaaagcaa
L G E E G N V V V A K S R M

1929

gttctaagaataataatagtagtggcggttgagatttcaatatttttattggttgttgggacaatctgttttatttat
tactttgggttttcttttgttttttcaaatttgaagtaccagcattatgaaatatcattttgttgcttgattgtat
tgtgtgtcccaactcctccactttttttttttcttttttccgctgtgttattgtcatgtatgaaatgaatgaattgga
agctcttcttctgtttttactttccgccatttacctctcgctactgttttgtttactgcaattgaaccatcttggtt
tctcttcaaaatttcaaaagccttctgattttctatttgtagatcattgcttccaacaccaaaaatgttcatttttt
tctaga 2329

Figure 3.5 Primary structure of the *icl* gene. The nucleotide sequence presented is that of the two *Xba* I fragments shown in figure 3.3a. Protein coding sequence only is in upper case letters. The derived amino acid sequence is shown in single letter code. Nucleotides are numbered in the 5' to 3' direction, beginning with the first residue of the initiator codon ATG. The putative TATA boxes and putative polyadenylation signals are underlined, the predicted nucleotides representing the start of transcription are in bold and underlined, as is the adenine nucleotide which acts as a site of polyadenylation for the cDNA clone, pBSICL1.7.

3.6 Features of the isocitrate lyase gene

3.6.1 Features of the 5' flanking region

Primer extension analysis was carried out using a synthetic oligonucleotide (20 mer) complementary to positions +11 to +30 of the genomic sequence. Several products were generated between 79 and 220 nucleotides in length, the strongest signal being given by the shortest product (figure 3.6). There are several possible reasons why multiple products were generated:

(i) the oligonucleotide used as a primer could possibly be mispriming at other sites: this could be determined by annealing at a higher temperature, so reducing the chances of mispriming.

(ii) there may in fact be numerous starts of transcription for the *icl* gene used at different developmental stages. However, this is unlikely to be the explanation in this experiment as the mRNA used as template was isolated from a uniform population of cotyledons harvested three days post imbibition.

(iii) the formation of secondary structures within the message may have caused the reverse transcriptase to pause and 'fall off' before the end of the message was reached. The extension reaction was performed at a relatively low temperature, so potentially allowing the formation of secondary structure within the RNA template. In this case the longest product would be the genuine one.

In addition to the longest (220 nucleotides) and the shortest (79 nucleotides), two other products of 113 and 116 nucleotides are seen in P1. Both the doublet at 113/116 nucleotides (-83/-86) and the longest of 220 nucleotides (-190) have a putative TATA box approximately 30 nucleotides upstream of the predicted transcript initiation site (**TATAAATA** and **TATAACTT** respectively), both fitting closely to the consensus for enzymes compiled by Joshi (1987a) of **TATA(A/T)ATA** (figure 3.7). The context for initiation of the two products does not exactly fit the consensus in either case, **TGCATTGAG** for the doublet and **GCTCTC** for the longer putative leader sequence (figure 3.7). The major product of 79 nucleotides lacks sequences resembling a TATA box in the expected position and the consensus around the initiation site. Most transcripts are found to initiate at an adenine residue, however this is not always the case and should not be used to discount the second sequence from being a transcription initiation site. In common with those from many other plant genes analysed, the leader sequences from both these putative starts of transcription are A+T rich, having a composition of 64 % A+T. The scanning mechanism proposed by Kozak (1986) is facilitated by a high A+T composition, as a more labile secondary structure is formed

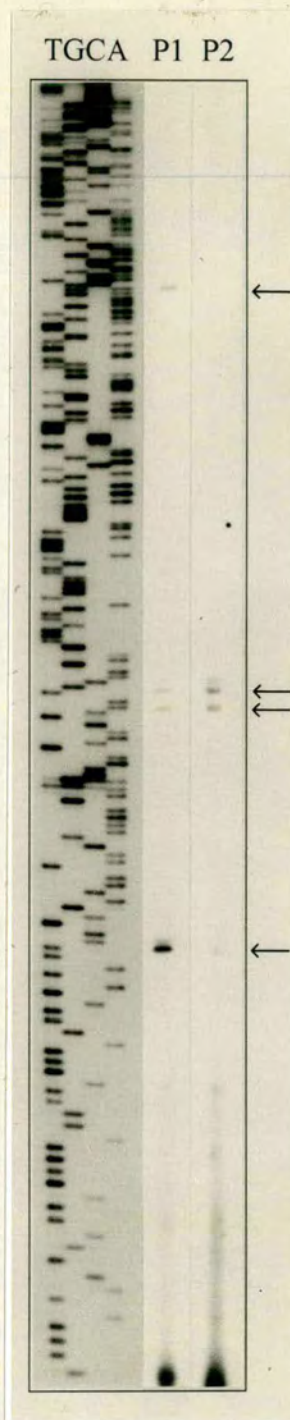


Figure 3.6 Primer extension analysis to identify the start of transcription. A 20mer oligonucleotide, complementary to position +11 to +30 of the genomic sequence, end-labelled using $[\gamma\text{-}^{32}\text{P}]\text{dATP}$ was used as a primer in both sequence analysis and primer extension studies. The sequence of the strand complementary to that presented in figure 3.5 is shown next to the products of the primer extension reaction. The products of lengths 79, 113, 116 and 220 nucleotides are indicated by arrows. (A = adenine, C = cytosine, G = guanine, T = thymine, P1, P2 = primer extension products).

(a)

<i>icl 1</i>	-122	C	C	C	T	T	A	T	A	A	A	T	A	C	-110
<i>icl 2</i>	-223	T	C	T	T	T	A	T	A	A	C	T	T	G	-211
Enzyme consensus		C ₆₀	C ₄₀	A ₄₀	C ₇₀	T ₈₀	A ₈₀	T ₇₀	A ₇₀	A ₅₀	A ₅₀	T ₆₀	A ₇₀	G ₄₀	
											T ₅₀				
Total consensus		T ₃₇	C ₄₂	A ₃₈	C ₄₇	T ₉₆	A ₉₇	T ₉₀	A ₉₄	T ₅₃	A ₉₅	T ₆₃	A ₇₁	G ₄₄	

(b)

<i>icl 1</i>	-89	T	G	C	A	T	T	G	A	G	-81
<i>icl 2</i>	-193	G	C	T	C	T	C	T			-187
Enzyme consensus		C ₇₀	C ₄₀	C ₆₀	A ₆₀	N ₃₀	C ₄₀	C ₄₀			
			T ₄₀				T ₄₀				
Total consensus		C ₂₉	T ₃₄	C ₅₉	A ₆₅	T ₅₄	C ₄₃	A ₄₂			
		T ₂₉									

Figure 3.7 Contexts for (a) TATA box and (b) transcription start of the 116 nucleotide (*icl1*) and 220 nucleotide (*icl2*) products generated by primer extension. Enzyme consensus derived from sequences of 10 plant genes encoding enzymes and total consensus derived from 79 plant genes (from Joshi, 1987a). The base with maximum frequency of occurrence at each position is shown in the consensus. An alternative base is shown when two bases have identical maximum frequency. N denotes the situation where three bases have identical maximum frequency of occurrence. The subscript figure following a base is the maximum frequency of occurrence. The TATA box and the nucleotide at which transcription is initiated are shown in bold type. The numbers at the ends of lines indicate the position of the sequence relative to the start of translation.

than in G+C rich sequences, therefore being more easily melted by ribosomal subunits. For subsequent discussion, the longest product of 220 nucleotides will be taken as the start of transcription, giving a 5' untranslated region of 190 nucleotides. This fits well with the estimate from Northern analysis of the size of ICL mRNA at 2.1 kb. The sequence of the 5' flanking region was scanned for the presence of known binding sequences for transcription factors. Two homologies, one to a cAMP responsive element (CRE; -988 to -983), and one to the CARG box to which the MADS box family of regulators bind (-2238 to - 2229) were found, which had previously been noticed in the *ms* promoter. However, when these sequences in the *ms* promoter were mutated, there was no effect on the level or pattern of expression directed by this promoter (Graham *et al.*, 1994a), therefore it seems unlikely that the factors which are known to bind to such sequences play a role in the regulation of *ms* and *icl* gene expression.

3.6.2 Features of the coding region

By comparison with known coding sequences from other plant ICL cDNAs, the *icl* gene of cucumber was seen to consist of three exons of 30, 1603 and 95 bp and two introns of 87 and 89 bp (figure 3.3b). The gene isolated from oilseed rape has four introns, three of which are much longer than those of cucumber. The position of the two cucumber introns is identical to that of the first and fourth oilseed rape introns. A direct comparison between the genomic sequences of oilseed rape and cucumber is displayed as a dotplot in figure 3.8, readily showing the positions of the introns and the high degree of homology between the two coding sequences at the nucleotide level (76 %). The coding regions of three of the five known fungal gene sequences (*S. cerevisiae*, *C. tropicalis* and *Yarrowia lipolytica*) are uninterrupted, whereas those of *A. nidulans* and *Neurospora crassa* are both interrupted by two introns. These are in corresponding positions to each other, but do not line up with the positions of any of the plant introns.

3.6.3 Features of the 3' flanking region

The 3' end of pBSICL1.7 isolated from the senescing cotyledon library terminates with an adenine residue at +2179 to give 275 nucleotides of 3' untranslated region. The vertebrate polyadenylation signal 5' AAUAAA 3' is absent from the region immediately upstream of the cleavage site. However, this exact sequence was only found in 40 % of plant nuclear genes (Joshi, 1987b), many others containing similar

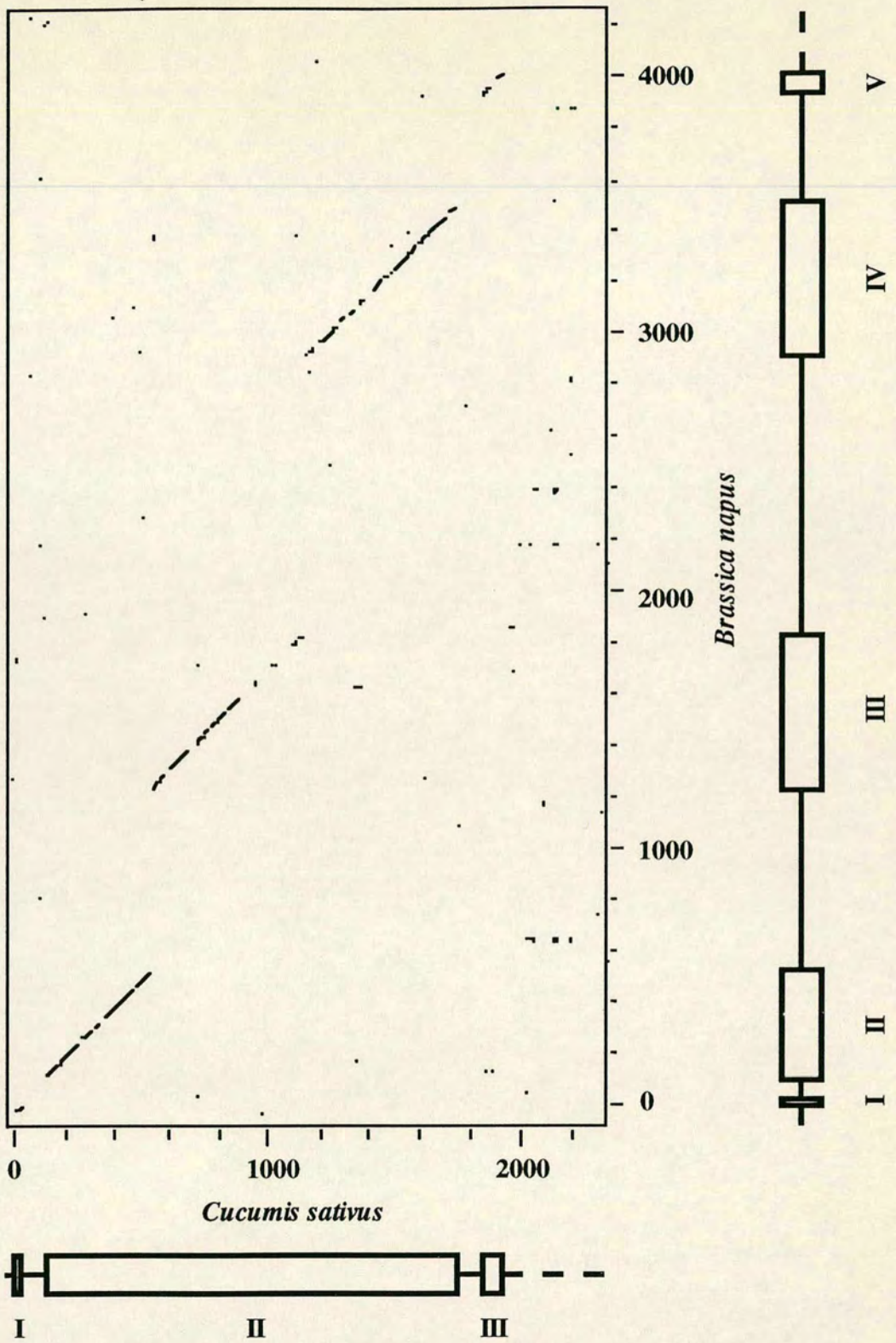


Figure 3.8 A comparison of the nucleotide genomic sequences of *C. sativus* and *B. napus*. The two sequences were aligned using the Compare and Dotplot programmes of the UWGCG package, using a window of 18 and stringency of 14. The gene structure of each is shown along the axis to which it relates. Boxes represent exons, solid lines represent introns and dotted lines represent 5' and 3' non coding sequences.

sequences with one or two base differences. Such a sequence, AAUGAA is present in two overlapping copies 28 and 24 nucleotides upstream of the cleavage site found in the clone isolated from senescing cotyledons. In vertebrate genes, in addition to this polyadenylation signal, other sequences are involved: AAUAAA occurs within the coding region and introns without causing processing around these sites (for reviews see Birnstiel *et al.*, 1985; Wahle & Keller, 1992). A G+U rich sequence (consensus YGUGUUY) lying approximately 30 nucleotides downstream of AAUAAA is also necessary for efficient processing (Wahle & Keller, 1992). In addition, the sequence CAYUG is implicated in processing, often being found close to the site of polyA addition. This last sequence is thought to base pair with a small nuclear RNA, U4, implicating snRNPs in the cleavage and polyA addition processes (Berget, 1984). In plant genes, the process of cleavage and polyadenylation is still poorly defined compared to animal and viral genes. The signals important in vertebrate systems do not appear to be conserved to any degree in plant systems (Joshi, 1987b) and indeed are not properly recognised in plants (Hunt *et al.*, 1987). The vertebrate signal AAUAAA is not necessarily active as part of the polyadenylation signal even when it is present (Sanfaçon, 1994), though it has been shown by deletion to be an important element in the cauliflower mosaic virus promoter (Sanfaçon & Hohn, 1990; Sanfaçon *et al.*, 1991), but tolerating a degree of mutation. In addition, the downstream UG-rich region present in vertebrate systems is not usually present. Multiple elements appear to be involved in plant systems, including some a considerable distance upstream of the cleavage site (Mogen *et al.*, 1990, 1992; Sanfaçon *et al.*, 1991; Sanfaçon, 1994), but in the majority of cases these have not yet been clearly defined. The repeated element UUUGUA has been identified to be important in processing of the cauliflower mosaic virus polyadenylation signal when present a defined distance upstream of the AAUAAA motif, though other upstream sequences are also necessary for efficient processing (Rothnie *et al.*, 1994). Recent studies suggest that both elements close to the processing site and others further upstream are necessary for efficient processing of plant mRNA 3'-end formation (Mogen *et al.*, 1992; Sanfaçon, 1994). Many plant genes have been reported to have multiple polyadenylation sites (Dean *et al.*, 1986; Graham *et al.*, 1989; Schaller *et al.*, 1991); only one was identified for the *icl* gene, though there may be others also utilised as yet unidentified. In general, one site tends to be used at a higher frequency than the others, but not to the same extent as in animal systems where a single site is predominantly used (Birnstiel *et al.*, 1985).

3.7 Features of the isocitrate lyase polypeptide

The coding region of the *icl* gene in cucumber is 1728 nucleotides long, translation yielding a polypeptide of 576 amino acids with a predicted relative molecular mass of 64,618. Comparison of the predicted amino acid sequence with those of cotton (*Gossypium hirsutum*, Turley *et al.*, 1990), oilseed rape (*B. napus*, Zhang *et al.*, 1993), castor bean (*Ricinus communis*, Beeching & Northcote, 1987) and *E. coli* (Matsuoka & McFadden, 1988) reveals similarities of 94.8 %, 90.8 %, 92.7 % and 57.3 % respectively, showing this protein is conserved to a very high degree between higher plant species. Amino acid comparisons between the predicted amino acid sequences of ICL from the four higher plant species, the five fungal species and *C. sativus*, *S. cerevisiae* and *E. coli* are displayed in figures 3.9a, b and c respectively. Several features of interest are apparent from these comparisons.

A hexapeptide, K-K-C-G-H-M, at positions 211 to 216 of the predicted cucumber amino acid sequence (boxed in figure 3.9c) is conserved throughout all species for which the sequence of ICL is known. This area has been identified to be important for the catalytic activity of the enzyme, the cysteine being the putative active site residue in *E. coli*, where it has been assigned to the glyoxylate subsite (Nimmo *et al.*, 1989; Ko & McFadden, 1990). This is the only cysteine conserved between the sequences of the plant, fungal and bacterial enzymes. A second sequence, Y-N-L-S-P-S-F-N-W, conserved throughout eukaryotes and with a single substitution in *E. coli*, has been implicated as being present at the active site (boxed in figure 3.9c): in *E. coli*, the two serine residues have been shown to be important for activity of the enzyme (Ko *et al.*, 1992).

The mechanism of targeting proteins to the peroxisomes appears to differ from protein to protein and even within the same protein between different species. Experiments investigating targeting mechanisms first demonstrated that the terminal tripeptide acts as a peroxisomal targeting signal. It was shown that the conserved tripeptide S-K-L at the carboxy terminus of firefly luciferase was both necessary and sufficient to direct a non-peroxisomal protein to the peroxisome (Gould *et al.* 1987, 1989; Miura *et al.*, 1992). None of the ICL polypeptides for which the sequence is known have this terminal tripeptide. However, all the higher plant enzymes, including that from cucumber, have either S-R-M or A-R-M at the carboxy terminus (boxed in figure 3.9a), both of which are conserved variants of S-K-L and have been shown to be functional as targeting signals in *Trypanosoma brucei* (Sommer *et al.*, 1992). Using deletion constructs expressed in transgenic *Arabidopsis*, Olsen *et al.* (1993) have demonstrated that this tripeptide is indeed necessary to localise oilseed rape ICL to the glyoxysomes and in

Figure 3.9 Amino acid alignments of predicted polypeptide sequences of ICL from (a) higher plants, (b) fungi and (c) cucumber, *S. cerevisiae* and *E. coli*. The top line shows the complete sequence for ICL from that organism, with differences in other species marked underneath. Numbers refer to the top sequence only. Dots indicate gaps introduced to generate an improved alignment. In (a), the conserved carboxyl tripeptide is boxed, and the positions of the two amino acid differences between the cucumber genomic and cDNA sequences marked with an asterisk. At these positions the cDNA residue is a S; in (c), the two conserved regions identified as being at the active centre of *E. coli* are boxed, and asterisks indicate the residues completely conserved between the higher plant sequences known. CS = *Cucumis sativus* (genomic sequence), GH = *Gossypium hirsutum* (Turley *et al.*, 1990), RC = *Ricinus communis* (Beeching & Northcote, 1987), BN = *Brassica napus* (Zhang *et al.*, 1993), NC = *Neurospora crassa* (Gainey *et al.*, 1992), AN = *Aspergillus nidulans* (Gainey *et al.*, 1992), SC = *Saccharomyces cerevisiae* (Schöler & Schüller, 1993), CT = *Candida tropicalis* (Atomi *et al.*, 1990), YL = *Yarrowia lipolytica* (Barth & Scheuber, 1993) and EC = *Escherichia coli* (Matsuoka & McFadden, 1988).

(a)

1 60
 CS MAASFSVPSM IMEEGRFEA EVAEVQAWWN SERFKLTRRP YTAKDVVSLR GSLRQSYASN
 GH T S R A K G
 RC G R A N K
 BN T S R A H K G

61 120
 CS DLAKKLWRTL KTHQANGTAS RTFGALDPVQ VTMMAKHLDT IYVSGWQCSS THTSTNEPGP
 GH EM T S T
 RC E S T
 BN EM S

121 180
 CS DLADYPYDTV PNKVEHLFFA QQYHDRKQRE ARMSMSREER AKTPYIDYLK PIIADGDTGF
 GH R V
 RC R V
 BN FV

181 240
 CS GGTATVKLC KLFVERGAAG VHIEDQSSVT KKCGHMAGKV LVAVSEHINR LVAARLQFDV
 GH
 RC I
 BN

241 * 300
 CS MGVETILVAR TDAVAATLIQ TNVDKRDHQF ILGATNPPLR GKSLAGALAE AMAAGKTGAE
 GH V T NM G N PQ
 RC L E N S T V TL T G N
 BN T V S I S V S SSL G V NN PA

301 * 360
 CS LQALDQWIS MAQLKTFSEC VTDAIMNTNA TENEKRRKLD EWMNHSSYEK CISNEQGREGI
 GH I N LA I M KSM I D RMN D L A
 RC T N LA P M K M G D RMN T D L Y N
 BN I L S R M DA VE LKRM L S S RVN L AR N L L

361 420
 CS AEKLGKLNLF WDWDLPRTRE GFYRFKGSVM AAIVRGWAF PHADLIWMET SSPDLVECTT
 GH R Q R M R
 RC DRM V R I AK FA A
 BN A VTD Q T V QI A N Q

421 480
 CS FAKGMKSIHP ETMLAYNLSP SFNWDASGMS DKQMEEFIPR IARLGFQWQF ITLAGFHADA
 GH E V M I T EH RD K
 RC E V M I T E RD G
 BN E V KT V T Q M Y

481 540
 CS LVVDTFARDY ARRGLMAYVE RIQREERNNG VDTLAHQKWS GANYDYDRYLK TVQGGISSTA
 GH T F K F
 RC I K K
 BN K S

541 576
 CS AMGKGVTEEQ FKESWTREGA VNLGEEGNVV VAKSRM
 GH T P G I S L A
 RC T P MEM SA SE A
 BN T P AGM GTSL

(b)

1 58
 NC .MAANNMVPN AVDPLEDEL FAKEVEEVKK WWSDSRWRQT KRPFTAQIV SKRGNL.KIE
 AN MSYI E QR YWD A. N K Y A
 SC ..MPIPVG T KN FAALQAK LDADAA IE SK NYS RD A VR TFPP
 CT AYT KI INQ EAD Q A I EP K IYS ED A K L A
 YL MQVSHQTFC VLTQCPNRQR NN I SP KH VYSP D A R TI VP

59 118
 NC YASNAQAKKL WKILEDRFAK RDASYTYGCL EPTMVTQMAK YLDTVYVSGW QSSSTASSSD
 AN P V G RN NK .E F D T
 SC P SVM R F V KHHNE GTV K F A D VQIS I I C T N
 CT P SQ SD F L KHD KSV F F A D IH A SI C T N
 YL Q SQ D F L QEHEKN HT F AS D VQ SI C T N

119 176
 NC EPGPDLADYP YTTCPNKVGH LFMAQLFHDR KQRQERLSVP KDQREKL..A NIDYLRPIVA
 AN S MN V N W E MTT H V T V I
 SC MD V E K LEA SKAK SQEELDEMG P T
 CT S MD V E WF E NMT EE ANTP . Y F I
 YL SS GG MD V E WF E NE L ESEPIQAPR RV I

177 236
 NC DADTGHGGLT AVMKLTKLFI EKGAAGIHIE DQAPGTTKCG HMAGKVLVPI QEHINRLVAI
 AN V R S
 SC A F M R M TSTN RCVI V V T
 CT I II R V
 YL E V M R I

237 296
 NC RAQADIMGSD LLCIARTDAE AATLITTTID PRDHAFILGC TNPDLLEPLAH LMMKAEAEKG
 AN T A S S H P I S IQ ND VM QA
 SC MC H IVV S SS T Y V A NI F E VLND IMS A
 CT S F N AV S S H Y I A ESGD A AE K I
 YL S FALN A S SS Y Y A A K AGH VD V VA R Q

297 355
 NC TGAQLQAIED DWLAKADLKR FDEAVLDVIA KGKFSNAKDL AAKYQAAVKG .KQISNREAR
 AN N E E G L ND V A N NSPLP K AA IE LTQS . . L
 SC S QE AD Q K CRD G L H I E E RSAL KQE IK FTSK GP LTET H K
 CT Y DE AR T E TK G L H I E K A NY KEA IK FTDK NP LSHT HK K
 YL A P V E NR GV L H FA EVN D SY KAE I EFNKK TP LSNTPDI D

356 415
 NC AIARQLLQOE IFFDWESPRT REGYYRLKGG CDCSINRAIS YAPYCDAIWM ESKLPDYAQA
 AN KEIA TD Y A YQ TQ A VA FA L K
 SC KL KEI H L V L YR TQ M RA F A LV NY FQ
 CT KL KE T KD Y N DVA A YQ TQ AVM GRA F A L A N
 YL YL AR KD Y N AA V YQ TQ AV GY A L

416 474
 NC EEFKAGPR.V WPEQKLAYNL SPSFNWKTAM GRDDQETYIR RLAKLGYCWQ FITLAGLHTT
 AN K D VHA K P E K GA A
 SC K E VKEK F D W PK SV E H F Q GD I N
 CT K D VKAA V D W NK PA E K GQ V
 YL K E VKNA V H W T SPE S V N

475 534
 NC ALISDQFAKA YSKIGMRAYG ELVQEPEIDN GVDVVKHQKW SGATYVDELQ KMTGGVSST
 AN T A Q MA T N NML I
 SC AVHN SRD FA D K A QN QR M D L E I G L LAQ A
 CT AV D NQ Q QT Q EK E N I G L R S T
 YL KSC ER K GEI Q Q CE E I GIL R IT

535 547
 NC AAMGKGVTED QFH.....
 AN KS.....
 SC T KENGVKK
 CT A KETKAKV
 YL A KSKL...

addition is sufficient to localise non-glyoxysomal proteins to that organelle. Of the five known fungal sequences, only that from *Y. lipolytica* contains this conserved tripeptide, in this case S-K-L. Behari and Baker (1993) have studied the import of castor bean ICL in an *in vitro* system using glyoxysomes isolated from the cotyledons of sunflower. They have demonstrated that deletions from the carboxy terminus do not abolish import of this protein. Indeed, a construct encoding the 168 N-terminal amino acids generates a polypeptide that is still import competent in this system.

Two sequences have been shown to be involved in targeting of ICL to the glyoxysomes of higher plants in different systems, the amino terminal region and the conserved carboxy terminal tripeptide. A third region potentially important for targeting is the central domain found in the protein sequences of the higher plant and the five fungal *icl* genes already sequenced, but not present in the *E. coli* protein, where no sequestration takes place. ICL from *Chlorella fusca* is smaller than that from higher plants at 47 kDa (Nicholl & John, 1986), approximately the same size as the *E. coli* protein, but the sequence is unknown. The algal protein is cytosolic (Pacy & Thurston, 1987); it is therefore interesting to speculate that it lacks the same region as the *E. coli* protein, and that this region is important for targeting to the peroxisomes. This central domain is the least conserved area of the enzyme; this however does not necessarily exclude it from a role in targeting as such sequences are often not highly conserved.

3.8 Isocitrate lyase and malate synthase are coordinately synthesised during germination

To investigate the regulation of *icl* and *ms* transcription during germination, RNA was isolated from cucumber cotyledons germinated from 0 to 8 days post imbibition under both light and dark conditions (section 2.4.1). Northern blot analysis was carried out, using 5 µg aliquots of each sample (section 2.4.4). The 1.4 kb insert from pBSICL and the 1.9 kb insert from pBSMS1.9 were radiolabelled and used as probes. The filter was first hybridised with the 1.4 kb insert of the cDNA clone, pBSICL, radiolabelled with [³²P]dCTP and subsequently, following removal of the first probe, rehybridised with the 1.9 kb insert of the cDNA clone, pBSMS1.9, similarly radiolabelled with [³²P]dCTP. The results are displayed in figure 3.10. The ICL transcript is approximately 2.1 kb long, which agrees with the predicted size from the data gained from analysis of the genomic clone. The MS transcript is about 200 nucleotides shorter at approximately 1.9 kb.

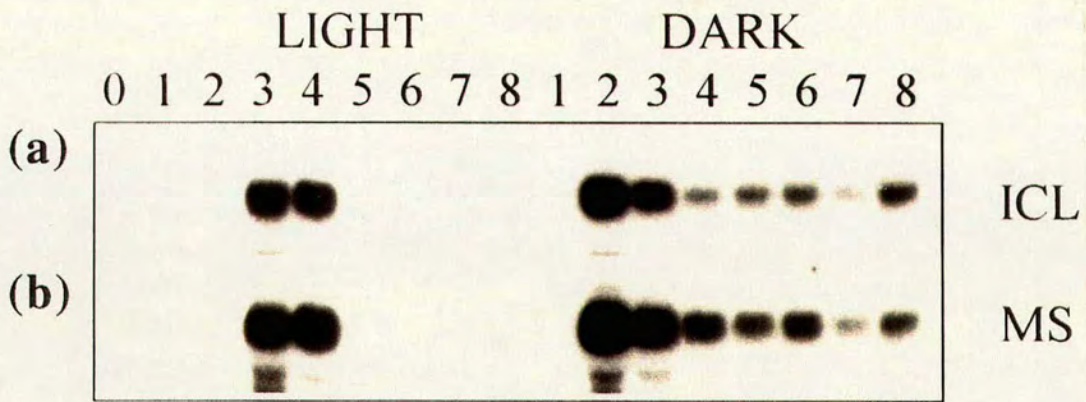


Figure 3.10 Northern blot analysis of cucumber total RNA, isolated from cotyledons of seedlings grown under light or dark conditions, from 0 to 8 days post imbibition, fractionated through a 1.2 % (w/v) denaturing agarose gel and hybridised with (a) the ICL cDNA clone, pBSICL, showing the 2.1 kb ICL transcript and (b) the MS cDNA clone, pBSMS1.9, showing the 1.9 kb MS transcript.

The two genes are transcribed in a coordinate manner in cucumber, confirming previous observations that the levels of translatable mRNAs and the enzyme activities of the two are coordinately regulated (Becker *et al.*, 1978; Weir *et al.*, 1980). Transcripts are easily detectable in light grown seedlings after 3 days, persisting through day 4 but thereafter the levels decline rapidly until day 5 when they are no longer detectable. In dark grown seedlings, transcript levels are detectable after 2 days, peaking around this time and from then on declining at a much slower rate than in light-grown seedlings, still being detectable after 8 days. Analysis of the levels of transcripts encoding ICL and MS throughout the life of a cotyledon reveals that, although undetectable in green cotyledons, the transcripts encoding both genes are once again abundant in the cotyledons during senescence in a temporally coordinated manner (Kim & Smith, submitted). They are also coordinately induced in dark-incubated detached cotyledons (McLaughlin & Smith, 1994).

3.9 Comparison of the sequences of the isocitrate lyase and malate synthase promoters

In view of the fact that ICL and MS synthesis is coordinately regulated at specific developmental stages, it was considered possible that the two genes may share common *cis*-acting elements in their 5' flanking regions or possibly elsewhere within the genes. To investigate this, the sequences of the 5' flanking regions were compared using the Bestfit programme from the UWGCG package. A 248 bp fragment of the *ms* promoter has been shown to direct faithful transcription of the *ms* gene during germination in transgenic *N. plumbaginifolia* (Sarah & Smith, pers. comm.), therefore this region was used for comparison with the *icl* flanking region. Various elements were identified that showed a degree of homology between the two. One element showing a high degree of homology lies at -163 to -176, in the reverse orientation in the *icl* promoter and at -215 to -202 in the *ms* promoter relative to the start of transcription. This sequence was designated RT. A second copy of RT is present in the *icl* promoter, at positions -325 to -312. These are depicted in figure 3.11a. It has also been detected in the promoter sequences of *icl* and *ms* genes from *A. nidulans*, *N. crassa* and *S. cerevisiae* (Graham *et al.*, 1994a). A 191 bp *Alu* I fragment of the *ms* promoter, which contains the RT sequence, was used in gel retardation assays and shown to bind a factor, which was specifically competed by addition of non-radioactive *Alu* I fragment. Gel retardation was also abolished by addition of a synthetic double stranded oligonucleotide, specific to the RT sequence, suggesting that

(a) RT (IMH1):

<i>icl1</i>	-163	TCATCTTCTcaATa	-176
<i>icl2</i>	-325	aaATCTTCTgcATT	-312
<i>ms</i>	-215	TCATCTTCTtgATT	-202

(b) IMH2:

<i>icl</i>	-1158	AAcCCCAaCCT	-1148
<i>ms</i>	-227	AAaCCCAcCCT	-217

(c) IMH3:

<i>icl</i>	-1087	TTTATAtATTTTCTcCCcTCTT	-1066
<i>ms</i>	-297	TTTATAaATTTTCTtCctTCTT	-276

(d) IMH4:

<i>icl</i>	-216	AAcAAAAAAAAACCGTAAAAaCT	-195
<i>ms</i>	-368	AAgAAAAAAAAACctTAAAAcCT	-347

Figure 3.11 Conserved sequences in the 5' flanking regions of the cucumber *icl* and *ms* genes. Conserved bases are shown in upper case and those differing between the genes in lower case letters. Numbers relate to the start of transcription. In (a), *icl1* is in the reverse orientation relative to the other sequences.

this portion of the promoter may bind a factor which is important in the regulation of the transcription of these genes (Graham *et al.*, 1994a). Further comparisons of the *icl* and *ms* flanking regions revealed additional conserved sequences (figure 3.11b-d), designated IMHs (ICL-MS Homology). However, the significance of any sequence conservation needs to be tested experimentally, before a role in regulation of transcription can be assigned.

3.10 Conclusion

The *icl* gene of cucumber has been shown to be present in a single copy in contrast to *B. napus* which has approximately six (Zhang *et al.*, 1993). The structure and sequence of the gene have been determined and the predicted amino acid sequence shown to be highly conserved between cucumber and those from other higher plants. The genes encoding ICL and MS are coordinately regulated at the transcriptional level in cucumber, so raising the possibility that they may share common *cis*-acting elements in their 5' regions. Comparison of these regions yields several highly conserved elements, which may be involved in the transcriptional regulation of these genes. To test the significance of these sequences, the activity from transcriptional fusions of the *icl* 5' region with a reporter gene in a stable transformation and a transient assay system were analysed as described in the following chapters.

CHAPTER 4

ANALYSIS OF *CIS*-ACTING REGULATORY ELEMENTS OF THE *ICL* GENE

4.1 Introduction and Aims

In the preceding chapter, the isolation and characterisation of the *icl* gene of cucumber was described. The gene was shown to be present in a single copy in cucumber, so like the *ms* gene, all the DNA elements for the control of gene expression must be found within this one gene. Previous studies in *B. napus* have indicated that expression of both the *icl* and *ms* genes is regulated primarily at the transcriptional level (Comai *et al.*, 1989) and it has been demonstrated that 3.5 kb of the 5' flanking region of the *icl* gene contains sufficient regulatory elements for the correct spatial and temporal expression of the gene in embryos and seedlings of *B. napus* (Zhang *et al.*, 1993). Changes in the steady state levels of the mRNAs during germination in cucumber indicates that it is likely that regulation also takes place at this level in this species (Smith & Leaver, 1986). The *ms* gene has indeed been shown by analysis of a reporter gene under the control of the *ms* promoter in transgenic plants to be regulated at the transcriptional level, by sequences present in the 5' flanking region of the gene, during and post germination (Graham *et al.*, 1990). By computer comparison with the *ms* gene 5' flanking region, elements in the *icl* gene promoter were identified that were common to the two, and therefore perhaps important for the coordinate regulation of the two genes throughout the development of the plant. To investigate this experimentally, the aim was to construct promoter-reporter gene fusions and an analysis of reporter gene expression carried out in transgenic plants. The reporter gene used in these studies encodes β -glucuronidase (GUS), an enzyme which has little endogenous activity in higher plant tissues and which can easily be assayed by both fluorometric and histochemical means (Jefferson, 1987).

Such analyses would ideally be carried out in the homologous host plant. However, in the case of cucumber, transformation using *A. tumefaciens* is not yet routine. Cucumber transformation is possible using the hairy root system of *Agrobacterium rhizogenes* (Trulson *et al.*, 1986), but for this project it was decided to use the easily transformed *Nicotiana plumbaginifolia* as the host for the transgene. Use of a heterologous host avoids the potential problem of *trans*-inactivation of homologous sequences which may be experienced when transgenes are introduced into the homologous host plant. Such *trans*-inactivation is thought to be due to genomic imprinting of the homologous sequence, possibly mediated by methylation of the regulatory regions (Matzke & Matzke, 1993). *N. plumbaginifolia* seeds contain storage lipid, present in the cotyledons and the endosperm, so the analysis of a gene normally expressed during and post germination of oilseeds should be possible.

N. plumbaginifolia has already been used successfully for such experiments using a *ms*-GUS transgene (Graham *et al.*, 1990).

The expression of the *icl* gene following germination has been described in the preceding chapter (section 3.8). In addition to an analysis of factors important for controlling regulation at this stage of development, an analysis of the metabolic control of *icl* gene expression was also investigated (section 4.8).

4.2 Construction of a plant transformation vector containing the isocitrate lyase gene of cucumber

In order to determine if the 6.5 kb *Hind* III fragment carrying the *icl* gene contains all the information necessary to direct faithful transcription in *N. plumbaginifolia*, it was subcloned into the binary vector pBIN19 for plant transformation (Bevan, 1984), using the *Hind* III site present in the polylinker of the plasmid (figure 4.1). Only one orientation of the fragment relative to the neomycin phosphotransferase gene (NPTII) was selected for subsequent transformation, as orientation of an entire gene fragment had previously been shown not to affect the level of expression when using this vector system (Graham *et al.*, 1990). The recombinant plasmids, identified by restriction endonuclease analysis, were transferred from *E. coli* to *A. tumefaciens* by a triparental mating (section 2.9.2). DNA was isolated from transformed bacterial colonies and used in Southern hybridisation to ensure that no rearrangement had occurred during the conjugation event (figure 4.2).

4.3 Transfer of the isocitrate lyase gene into *N. plumbaginifolia*

The *icl* gene was transferred into *N. plumbaginifolia* using the leaf disc transformation system as described in section 2.9.3. Leaf pieces were cut and inoculated with *A. tumefaciens* carrying the recombinant plasmid containing the *icl* gene. Shootlets were formed and regenerated in the presence of kanamycin, eventually producing entire plants which were grown to maturity, allowed to self-fertilise and generate seed.

Four independent transformants were grown to maturity and seed was collected from each one. In order to show that the plants were genuinely transformed, seeds were germinated on rooting medium with kanamycin at 400 µg ml⁻¹. Untransformed seeds lack the NPTII gene which is carried on the pBIN19 plasmid, and consequently they do not grow normally beyond the first five days. Non-resistant plants turn white in the presence of kanamycin, whereas those carrying the NPTII gene can continue to grow normally and generate green cotyledons and leaves. Seeds from all four independent

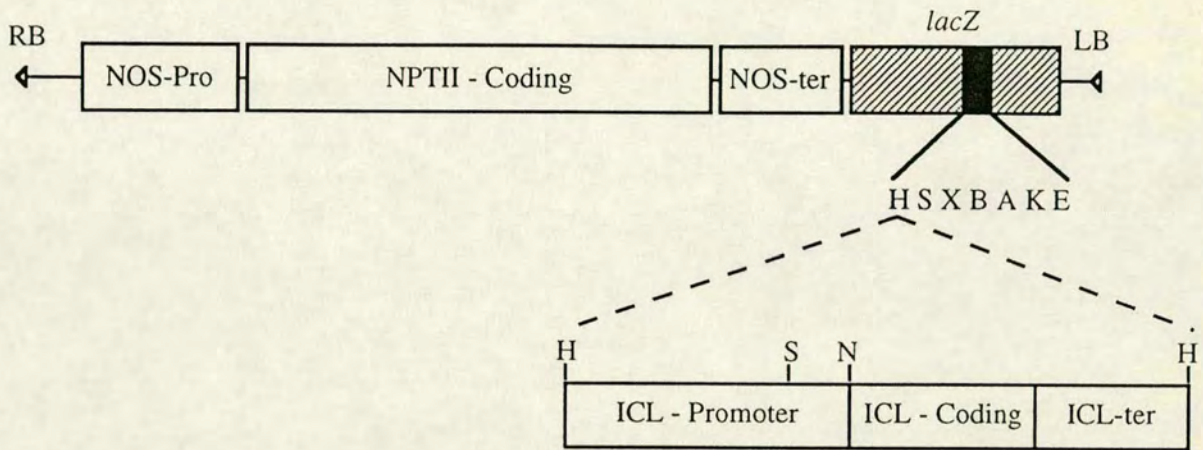


Figure 4.1 Constructs used for transformation. Organisation of the transferred DNA in pBIN19 (Bevan, 1984) and the 6.5 kb *Hind* III fragment carrying the *icl* gene of cucumber. The T-DNA of pBIN19 is bounded by imperfect repeats, the right and left borders (RB and LB respectively). The shaded region (*lacZ*) represents the α -complementation region of the *lac* operon with the solid black box within it representing the polylinker. Unique sites only are shown. The *icl* gene fragment inserted into pBIN19 is shown divided into the promoter, coding and terminator regions. H = *Hind* III, S = *Sal* I, X = *Xba* I, B = *Bam* HI, A = *Ava* I, K = *Kpn* I, E = *Eco* RI, N = *Nco* I; NOS-Pro and NOS-Ter = nopaline synthase gene promoter and terminator sequences respectively; NPTII-Coding = neomycin phosphotransferase-II coding sequence from Tn5; ICL-Promoter, ICL-Coding and ICL-Ter = the isocitrate lyase promoter, coding and terminator regions respectively.



Figure 4.2 Southern hybridisation to confirm plasmid arrangement of pBIICL after transfer from *E. coli* to *A. tumefaciens* by triparental mating. *Agrobacterium* DNA was isolated, digested with *Hind* III, and size fractionated through 0.8 % (w/v) agarose. The filter was probed with the 1.4 kb insert of pBSICL. Lanes 1 to 4 = DNA isolated from individual clones (30 minute exposure). P = *Hind* III digested pBIICL DNA (5 minute exposure).

transformants showed a ratio of resistance of three to one, showing that kanamycin resistance is transmitted through meiosis in a Mendelian fashion.

4.4 Analysis of transcription of the isocitrate lyase transgene in *N. plumbaginifolia*

Seeds from two of the transformants were germinated as described in section 2.9.4.1 and grown for three days under light conditions. They were harvested and RNA isolated from entire seedlings, as the cotyledons are too small to dissect at this stage. Northern analysis was carried out, using the radiolabelled 1.4 kb insert of pBSICL as a probe. RNA isolated from three day old cucumber cotyledons was used as a positive control, and that isolated from 3 day old untransformed *N. plumbaginifolia* seedlings as a negative. Both transformed plants contained transcripts of the correct size, though the degree of hybridisation was variable, transformant 2 giving a higher level of hybridisation, though still considerably lower than that given by RNA from three day old cucumber cotyledons (results not shown). No transcripts were detected in the untransformed *N. plumbaginifolia* seedlings, indicating that the cucumber probe does not detect the endogenous *N. plumbaginifolia* ICL transcripts.

4.5 Temporal expression of the cucumber isocitrate lyase transgene in *N. plumbaginifolia*

To analyse changes in the levels of ICL transcripts, seeds of transformant 2 were germinated and grown for 0, 1, 3, 5 and 7 days (see figure 4.3), RNA was isolated from each, and in addition from the mature leaf tissue of this plant, and used in Northern hybridisation, using the radiolabelled 1.4 kb insert of pBSICL as a probe. ICL mRNA increased in amount to day three, and then rapidly declined (figure 4.4) in a manner similar to that seen in cucumber and to that seen for MS both in cucumber and for the transgene in *N. plumbaginifolia* and *Petunia* (Graham *et al.*, 1990). Transcripts of ICL were undetectable in the sample isolated from mature leaf tissue and from untransformed seedlings. RNA samples from 14 day old cucumber cotyledons, which are by then fully green, also show no detectable ICL transcripts when used in Northern hybridisation (McLaughlin & Smith, 1994). The enzyme activity of ICL in untransformed *N. plumbaginifolia* seedlings was analysed to determine if it followed the same temporal pattern of accumulation and decline as in cucumber (S. M. Smith, University of Edinburgh, pers. comm.). *N. plumbaginifolia* seeds frequently germinate in a non-synchronous manner, therefore several stages of development through the first



Figure 4.3 *N. plumbaginifolia* seedlings at different developmental stages following the onset of germination. Seeds were imbibed overnight at 4 °C in 1 mM gibberellic acid, and germinated on rooting medium under a 16 h photoperiod for the number of days indicated over each developmental stage. Magnification is x 10

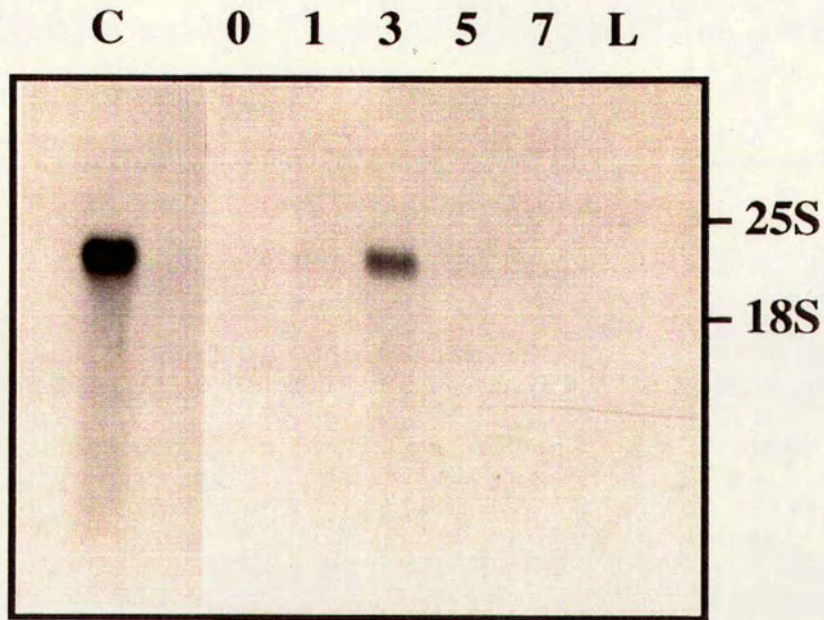


Figure 4.4 Northern blot analysis to investigate the temporal expression of the cucumber *icl* transgene during germination of *N. plumbaginifolia* transformant ICL2 seedlings in the light and in mature leaf tissue. Total RNA was isolated from seed batches germinated for 0 day post imbibition to 7 days post imbibition (0, 1, 3, 5, 7) and from leaves (L). 10 μ g total RNA was loaded per lane. C = 10 μ g total RNA isolated from cucumber cotyledons, 3 days post imbibition. C was exposed for a shorter time than the other lanes.

five days of growth were identified and seedlings selected according to these stages rather than selecting at days post imbibition. Seeds were germinated and grown for 0 to 5 days, stages harvested as depicted in figure 4.5 and the extracts assayed for ICL activity. The results depicted in figure 4.6 show that indeed the temporal pattern of enzyme activity is very similar in these two plant species. Stage 4 is reached after three days of growth post imbibition; this peak is therefore consistent with the results of the northern analysis displayed in figure 4.4. It has previously been demonstrated that activity of ICL in cucumber shows this pattern of expression, peaking at 3 days post imbibition and subsequently declining (Weir *et al.*, 1980).

4.6 Expression of an isocitrate lyase-glucuronidase fusion gene in *N. plumbaginifolia*

4.6.1 Construction of the fusion gene and transfer to *N. plumbaginifolia*

In order to determine if the elements important for controlling expression of the *icl* gene are contained within the 5' flanking region, fusions of the *icl* promoter with the β -glucuronidase (GUS) reporter gene in plasmid pRAJ275 (Jefferson *et al.*, 1986), were employed. Only fusions with the promoter in the correct orientation were constructed as it had previously been shown that the *ms* promoter, when fused in the reverse orientation was unable to direct transcription of the GUS gene (Graham *et al.*, 1990). pRAJ275 contains the GUS gene from *E. coli* with the context around translation start altered to give a 'consensus' translational initiator and so generating a unique *Nco* I site (CCATGG) at the initiator ATG codon (figure 4.7). The *icl* gene of cucumber has an *Nco* I site at the start of translation, therefore fusion constructs made with pRAJ275 contain promoter and 5' untranslated sequences derived solely from the cucumber *icl* gene. This is therefore an ideal system to use for this study. pRAJ275 lacks a transcription terminator, therefore the nopaline synthase (Nos) terminator was inserted into the plasmid by substitution of the *Sna* BI-*Eco* RI fragment from pRAJ275 with that from pBI101 (Jefferson *et al.*, 1987) which carries the Nos terminator at the 3' end of the GUS gene. The 2900 bp *Hind* III-*Nco* I fragment of the *icl* gene, containing approximately 2700 bp upstream of the transcriptional start, was fused to the 5' end of the GUS gene, to create pRAJ2900. The construction of this fusion is shown in figure 4.7. Recombinant plasmids were identified by restriction endonuclease digestion and the fusion junctions were confirmed by DNA sequencing, using oligonucleotide primers specific to the GUS gene sequence. The promoter-GUS-terminator cassette was then transferred from the pUC based vector to the binary vector pBIN19, using the

0 1 2 3 4 5 6 7

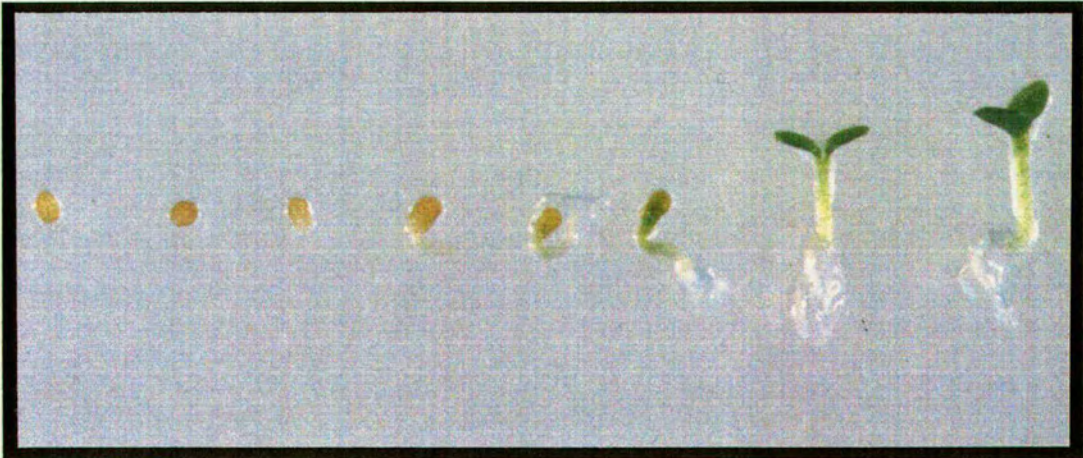


Figure 4.5 *N. plumbaginifolia* seedlings at different developmental stages (0 to 7) following the onset of germination in the light, from 0 to 5 days post imbibition. Magnification is x 8

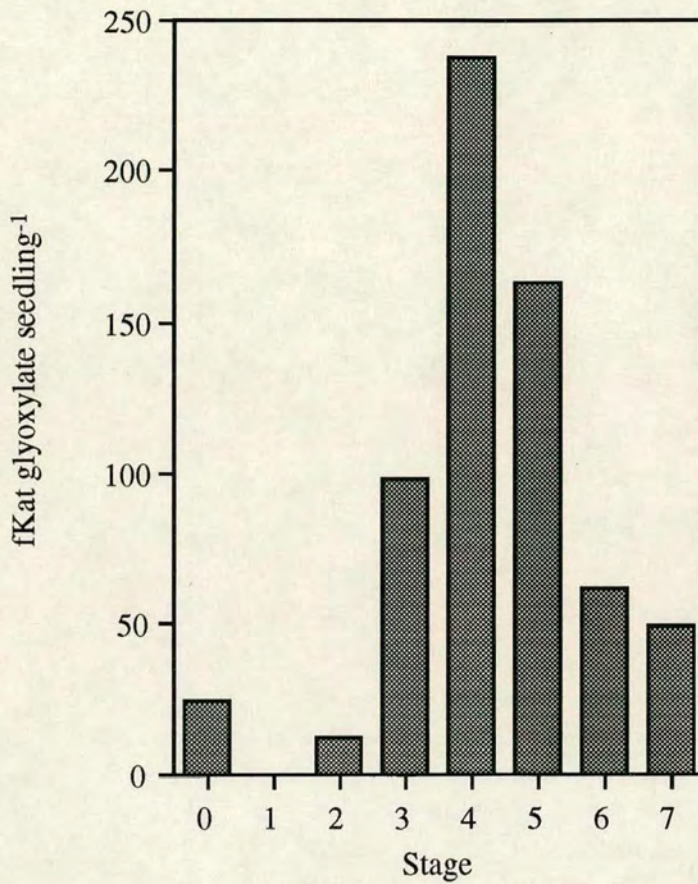


Figure 4.6 ICL activity during germination and post germinative growth in *N. plumbaginifolia*. Seeds from untransformed *N. plumbaginifolia* were germinated and grown in the light for up to 5 days post imbibition. Stages were selected and isocitrate lyase activity determined (S. M. Smith, pers. comm.) in cell-free extracts from 40 seedlings at each stage.

Figure 4.7 Organisation of the promoter-GUS fusion genes.

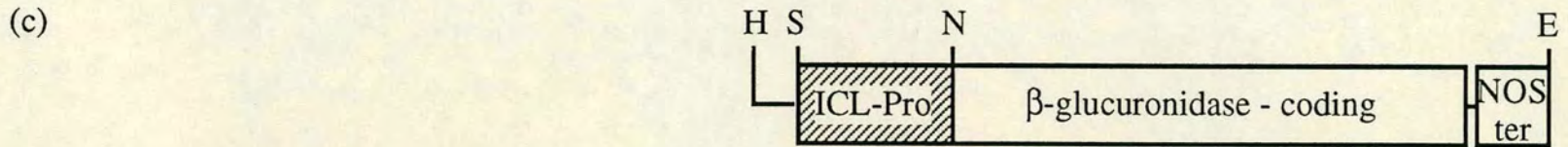
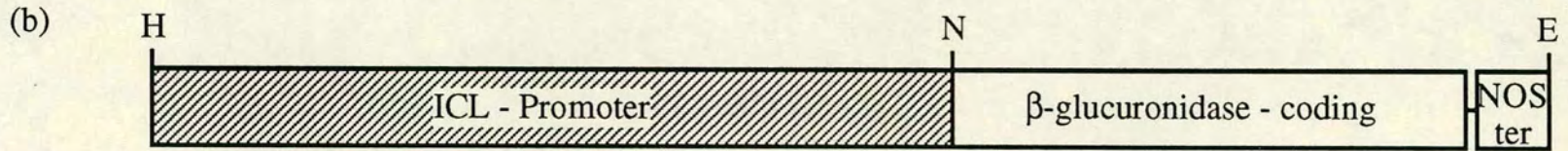
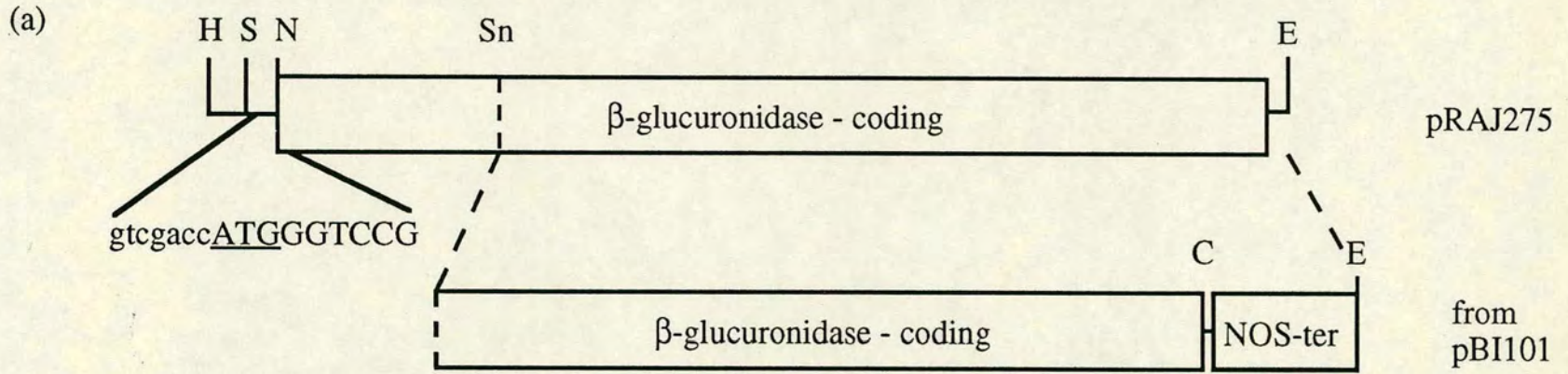
(a) The pUC based plasmid pRAJ275 (Jefferson *et al.*, 1986), showing the consensus translational initiation sequence and *Nco* I site at the start of the coding region. pRAJ275 lacks a terminator, so the *Sna* BI-*Eco* RI fragment was substituted by that fragment from pBI101 (Jefferson *et al.*, 1987), thus inserting the terminator from the nopaline synthase gene.

(b) The plasmid pRAJ2900: A 2900 bp *Hind* III-*Nco* I fragment of the 5' flanking region fused to the GUS gene in the altered pRAJ275 plasmid.

(c) The plasmid pRAJ572: A 572 bp *Sal* I-*Nco* I fragment of the 5' flanking region fused to the GUS gene in the altered pRAJ275 plasmid.

H = *Hind* III, S = *Sal* I, N = *Nco* I, Sn = *Sna* BI, E = *Eco* RI, C = *Sac* I; β -glucuronidase - coding = the coding region of the β -glucuronidase from *E. coli*; NOS-ter = nopaline synthase gene terminator region; ICL - Promoter (or Pro) = the promoter from the isocitrate lyase gene of cucumber.

(a) is drawn at twice the scale of (b) and (c).



Hind III and *Eco* RI sites which flank the cassette, to generate pBI2900 (figure 4.7), and recombinant plasmids were identified by restriction endonuclease digestion. The recombinant plasmid was transferred from *E. coli* into *A. tumefaciens* by a triparental mating (section 2.9.2), DNA was isolated from transformed colonies and used in Southern hybridisation to ensure that no rearrangement had occurred during the conjugation event (figure 4.8).

To transfer the chimeric gene into *N. plumbaginifolia*, leaf disc transformation was carried out as described in section 2.9.3. Plants were regenerated, grown to maturity, allowed to self fertilise and produce seed. Seed was collected from eleven independent transformants. Seeds from each transformant were germinated and grown on rooting medium containing 400 $\mu\text{g ml}^{-1}$ kanamycin, in order to confirm that they were genuinely transformed. Approximately 75 % of the seedlings from all eleven independent transformants were resistant to kanamycin, indicating that they were all transformed and that kanamycin resistance had been inherited in a Mendelian fashion.

4.6.2 Histochemical localisation of GUS activity in germinating seedlings

The histochemical assay for GUS activity can be used to identify which tissues and even which individual cells are expressing the fusion construct. In cucumber, ICL is expressed principally in the cotyledons which are the main storage reserve of the seed. In *N. plumbaginifolia*, the cotyledons are also one of the major storage organs, the other being the endosperm. The histochemical assay was employed to determine if GUS under the control of the *icl* promoter was being expressed in *N. plumbaginifolia*. Transformed *N. plumbaginifolia* seeds from the eleven independent transformants were germinated for 3 days and seedlings representative of that age selected. The testa and endosperm were carefully removed and the seedlings incubated in 1 mg ml^{-1} X-gluc at 37 °C (section 2.9.4.4), until the development of blue staining was observed. Staining of the cotyledons was detected in the seedlings from eight of the eleven independent transformants. No other parts of the seedling gave the same intensity of staining, suggesting that GUS expression is restricted to the cotyledons (figure 4.9). Some staining of the root hairs was detected in the seedling sets where an intense degree of staining was observed in the cotyledons. This could have been due to the uptake of the blue colour from the incubation medium by the root hairs. In the three transformants which had no potential to stain, it is likely that integration of the transgene occurred in an area of the genome which is not accessible to the transcription factors at this stage of development, due to the state of the chromatin or the methylation

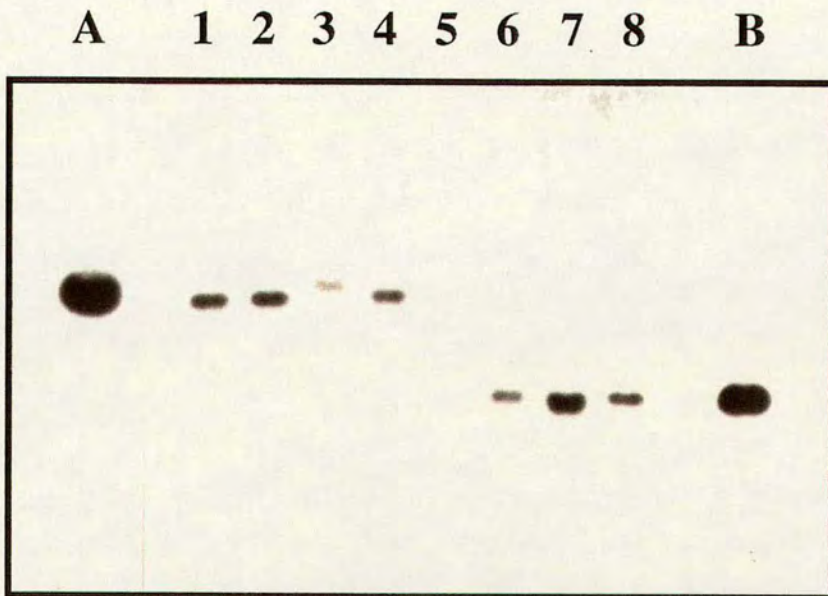


Figure 4.8 Southern hybridisation to confirm plasmid arrangement of pBI2900 and pBI572 after transfer from *E. coli* to *A. tumefaciens* by triparental mating. *Agrobacterium* DNA was isolated, digested with *Hind* III/*Eco* RI, and size fractionated through 0.8 % (w/v) agarose. The filter was probed with the 572 bp *Sal* I-*Nco* I promoter fragment (figure 3.3). Lanes 1 to 4 = DNA isolated from individual clones containing pBI2900, lanes 5 to 8 = DNA isolated from individual clones containing pBI572 (30 minute exposure). A = *Hind* III/*Eco* RI digested pBI2900 DNA (5 minute exposure); B = *Hind* III/*Eco* RI digested pBI572 DNA (30 minute exposure).

2900



572

ICL

Figure 4.9 Histochemical staining of transgenic *Nicotiana* seedlings. Seeds were germinated and grown until they reached stage 4. The testa and endosperm were removed and the seedlings stained (section 2.9.4.4). 2900 seedlings were stained for 60 minutes, 572 and ICL for 16 hours.

state, or that the inserted sequence had been rearranged. Staining of the endosperm was not investigated in these experiments. In previous work with a *ms*-GUS transgene, Graham *et al.*, (1990) showed that selected cells of the endosperm isolated 1 to 2 days post imbibition were stained (those closest to the radicle pole), and occasionally staining throughout the endosperm was observed, but the potential to stain was rapidly lost after 2 to 3 days of germination. As with the *icl*-GUS transgene, the cotyledons of plants transformed with the *ms*-GUS transgene stained intensely at days 2 to 3 of germination.

4.6.3 Temporal expression of the *icl*-GUS transgene in *N. plumbaginifolia*

The variation in levels of expression of the transgene during and immediately post germination were analysed using a fluorometric assay. Seeds from the eight independent plants transformed with pBI2900 which had been shown to have GUS activity by histochemical analysis were grown for 0 to 5 days post imbibition and forty representative seedlings selected for every stage depicted in figure 4.5. Seedling extracts were obtained (section 2.9.4.3) and assayed for GUS activity. The activity detected in the seedlings rose to a peak around stages 5 to 6 and thereafter declined sharply to stage 7 (figure 4.10), thus reflecting the pattern of endogenous ICL activity detected in untransformed seedlings (figure 4.6). However, the decline in GUS activity after stages 5 to 6 is much less than for ICL, which also peaks slightly earlier. This can be explained by the stability of GUS in plant cells (Jefferson *et al.*, 1987), giving rise to a slower turnover than that of ICL. The pattern of expression reported here will subsequently be referred to as the 'germination response'. Seedlings from untransformed plants were also tested for GUS activity, but no significant levels were detected throughout the germination series (figure 4.11).

The level of activity varied between independent transformants, the peak ranging from 1041 to 6783 fKat seedling⁻¹ at stages 5 to 6. The copy number for the transgene in each of these transformants was not determined; however, previous work on the malate synthase gene and other systems has indicated that the level of expression in such a system is not directly proportional to the number of transgenes carried (Graham *et al.*, 1990; Willmitzer, 1988). Plant transgenes are thought to insert randomly into the genome, though a preference towards actively expressed regions has been proposed, due to the greater accessibility of such regions. Therefore, it is thought that positional effects rather than number of copies of a transgene are more important in determining the level of expression within a particular independent transformant. The qualitative

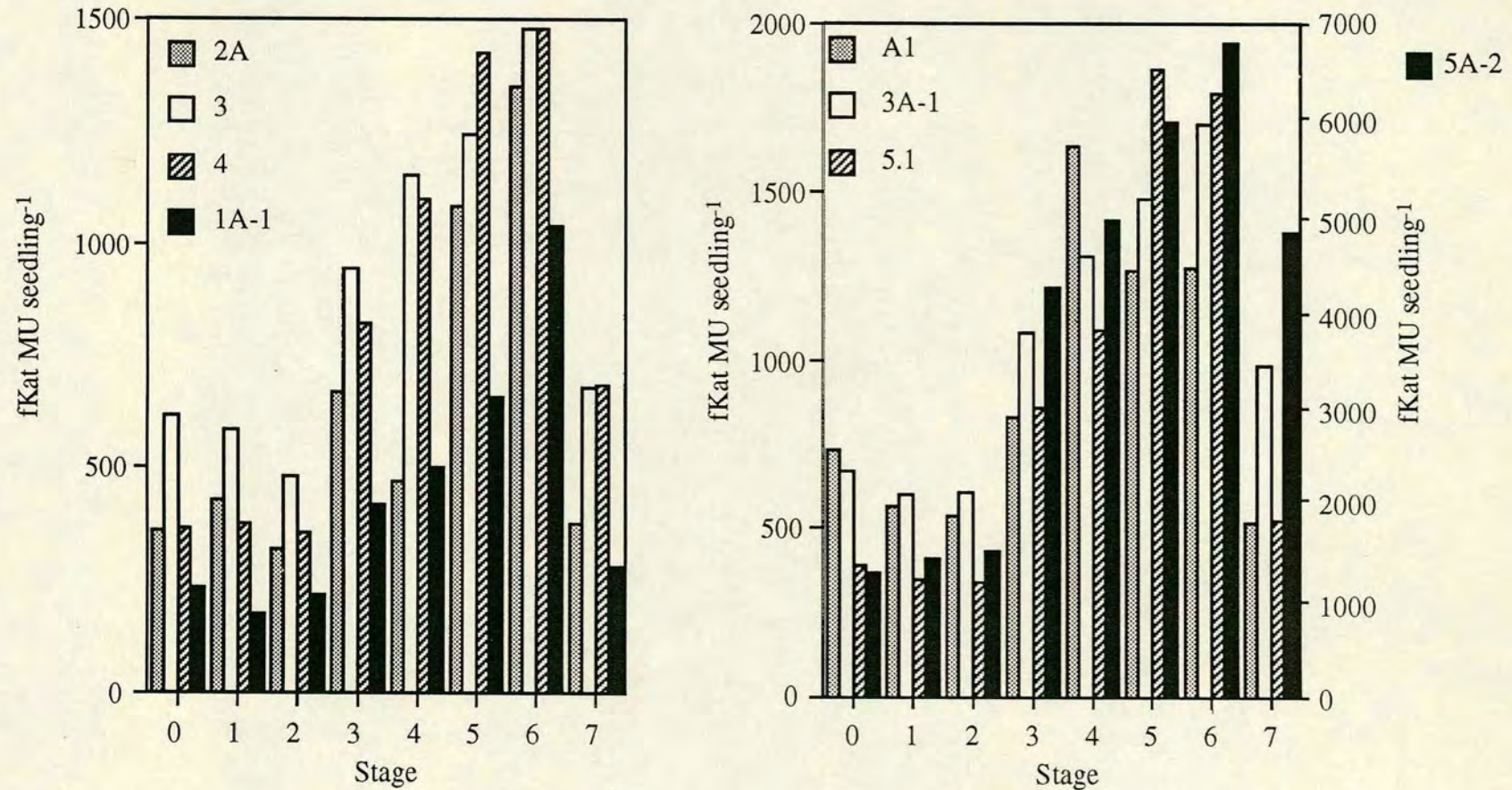


Figure 4.10 Glucuronidase activity during and post germination of *N. plumbaginifolia* transformed with pBI2900. Forty seeds/seedlings per stage from eight independent transformants (see legends beside axes) were analysed for levels of GUS activity. The stages selected are depicted in figure 4.5.

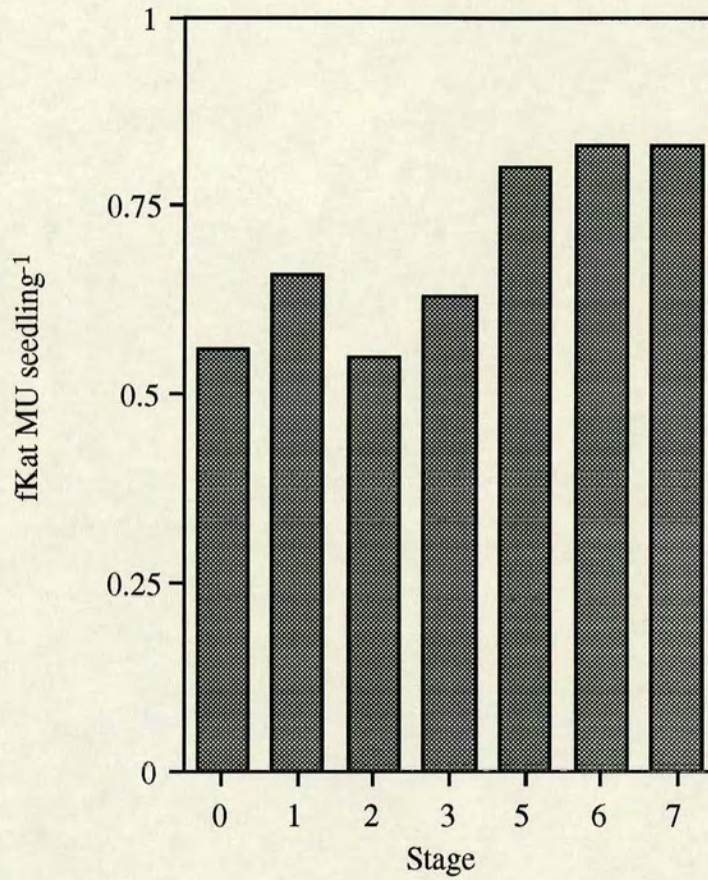


Figure 4.11 Glucuronidase activity during germination of untransformed *N. plumbaginifolia*. 40 seeds per stage were analysed for levels of GUS activity. The stages selected are depicted in figure 4.5

pattern of GUS activity observed here is very similar to that found using the *ms*-GUS transgene (Graham *et al.*, 1990), as would be anticipated in view of the coordinate expression of the two genes during this period of development in cucumber.

4.7 Effect of removal of a 2.3 kb fragment of the 5' flanking region on the expression of the *icl*-GUS fusion gene in *N. plumbaginifolia*

4.7.1 Construction of a 572 bp *icl*-GUS fusion and transfer to *N. plumbaginifolia*

The fusion gene was constructed in the same manner as previously: a 572 bp *Sal* I - *Nco* I fragment was fused to the GUS gene plus *Nos* terminator in pRAJ275 to generate pRAJ572 (figure 4.7), which contains the 382 bp immediately upstream of the start of transcription. The fusion junctions were confirmed by DNA sequencing, the *Hind* III-*Eco* RI cassette transferred to pBIN19 to generate pBI572 and the recombinant plasmid identified by restriction endonuclease mapping. It was then transferred to *A. tumefaciens* by a triparental mating and plasmid arrangement confirmed by Southern hybridisation as for pBI2900 (figure 4.8). Leaf disc transformation was carried out as described in section 2.9.3, plants were regenerated and seed collected from 10 independent transformants which were shown to be genuinely transformed by their resistance to kanamycin as described previously (section 4.6.1).

4.7.2 Analysis of expression of the 572 bp *icl*-GUS fusion construct in germinating seedlings

Transformed seeds were germinated and localisation and qualitative analysis of expression levels carried out by histochemical means as described previously (section 4.6.2). As for seedlings transformed with pBI2900, staining of the cotyledons was observed in seedlings from 6 of the 10 independent transformants (figure 4.9). No staining was observed in the radicle of these seedlings. However, seedlings transformed with pBI572 showed very weak staining compared to those transformed with pBI2900. Incubation for 16 h was necessary before staining was detectable to the naked eye with seedlings containing the 572 bp *icl*-GUS fusion whereas a 30 min incubation was sufficient with those containing the 2900 bp *icl*-GUS fusion, and the colour generated was much more intense in the latter. Although low, the level of staining observed in the 572 bp *icl*-GUS fusion containing seedlings was above that

found in untransformed 3 day old seedlings or in seedlings transformed with pBIICL stained for 16 h, in which no colouration was detectable (figure 4.9).

GUS levels were analysed in more detail using the fluorometric assay. As for the pBI2900 transformants, seeds of the six pBI572 independent transformants were germinated for 0 to 5 days, the stages selected as depicted in figure 4.5 and seedling extracts assayed (section 2.9.4.3). The results of these assays are presented in figure 4.12.

As predicted from the histochemical analysis, the levels of expression are very much lower than those found in 2900 bp *icl*-GUS fusion containing seedlings, at only 20 to 30 fKat seedling⁻¹ at stage 4 to 5. However, this is significantly above the level found in untransformed seedlings, which had its greatest value at stage 7 of 0.8 fKat seedling⁻¹ (figure 4.11). The qualitative pattern of expression is also somewhat different from that seen with seedlings containing the 2900 bp *icl*-GUS fusion, GUS activity at day 0 being the highest value for 5 of the 6 transformants analysed. However, 5 of the 6 (all but transformant 2.1) do show a peak around stages 4 to 5, as does endogenous ICL activity, which presumably corresponds to the 'germination response' (figure 4.13), although it occurs slightly earlier than the peak in seedlings transformed with pBI2900. The significance, if any, of this slight difference in timing is unclear. Any sign of a peak around this time point is completely absent in the untransformed seedlings (figure 4.11). This indicates that the proximal 382 bp of the flanking region contains information necessary and sufficient to direct faithful transcription of the GUS gene during germination albeit at very low levels. The additional 2300 bp in pBI2900 must contain an element(s) important for the enhancement of the germination response to the higher levels normally detected. The regulation of transcription during germination may demand a number of elements which together serve to produce faithful expression of the gene.

One possible explanation for the unexpectedly high levels of expression in stage 0 seedlings is that ICL (and MS) is expressed during embryogenesis (Comai *et al.*, 1992; Turley & Trelease, 1990) and the GUS activity seen at day 0 remains from the expression of the fusion construct during embryogenesis. GUS is a very stable enzyme, and so it is not unlikely that some activity would remain from a previously high level at an earlier stage of development. It is of note that endogenous ICL activity in *N. plumbaginifolia* determined here is also at a higher level in stage 0 seedlings than in those from stages 1 and 2. The *Nicotiana* seeds used in all these experiments were relatively fresh, that is they had not been allowed to lie dormant for more than three months, hence the likelihood of finding activity remaining from embryogenesis may be higher in these seeds than in those which had had greater periods of dormancy. The

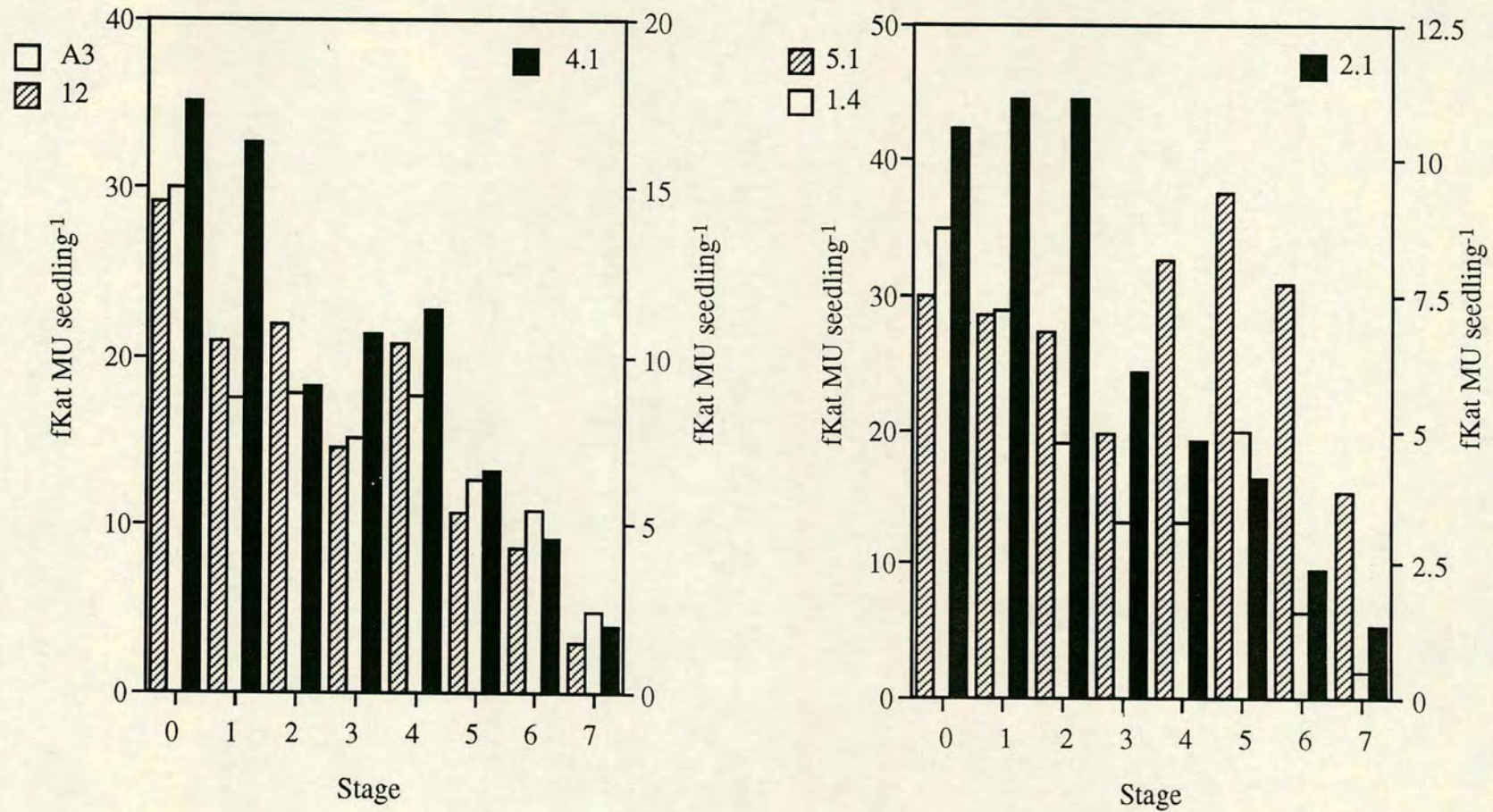


Figure 4.12 Glucuronidase activity during and post germination of *N. plumbaginifolia* transformed with pBI572. Forty seeds/seedlings per stage from six independent transformants (see legends beside axes) were analysed for levels of GUS activity. The stages selected are depicted in figure 4.5.

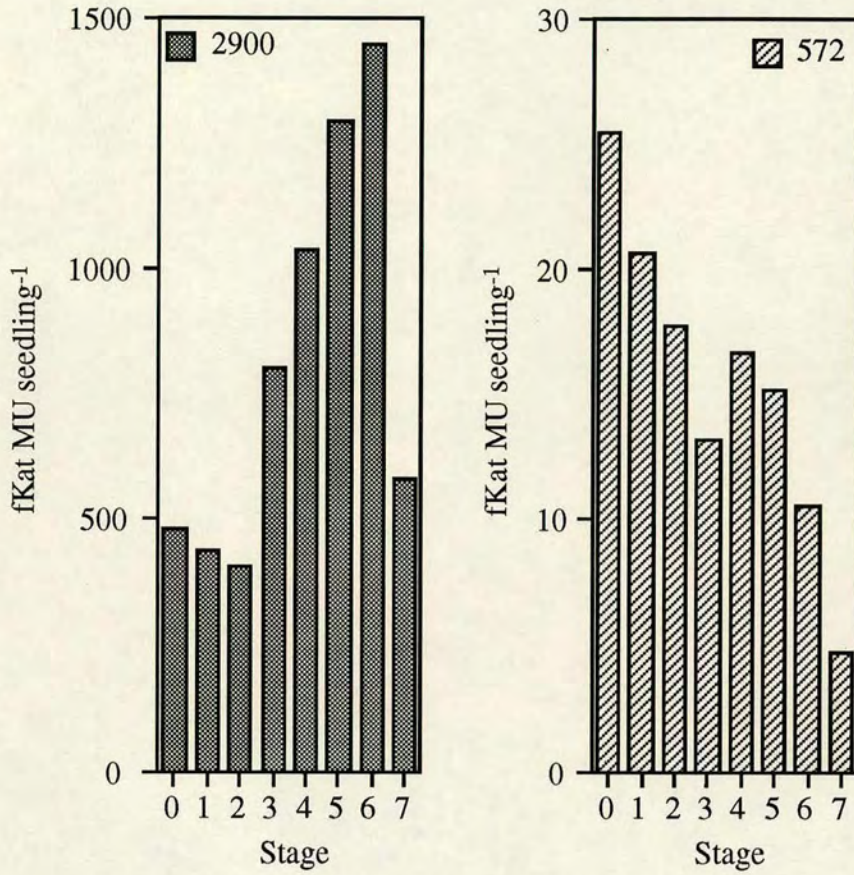


Figure 4.13 Average glucuronidase activity during germination of *N. plumbaginifolia* transformed with either pBI2900 or pBI572. The GUS activity of eight (pBI2900) or six (pBI572) independent transformants was averaged for each stage.

immediate decline of activity post imbibition could be due to the renewed activity of proteases in the imbibed seed.

4.8 Regulation of expression of the *icl* gene by metabolic status

4.8.1 Metabolic control of ICL and MS in pro- and eukaryotes

Numerous studies of the effect of various metabolites on the expression of the enzymes involved in the glyoxylate cycle have been made. The glyoxylate bypass genes (*icl* and *ms*) of microorganisms allow growth on media containing acetate and other two-carbon compounds, when depleted of glucose or other sugars. Expression of these genes is repressed by the presence of glucose or related metabolites, but induced when the organism is grown on medium containing two-carbon compounds as the sole carbon source (Kornberg, 1966; Sjogren & Romano, 1967; McCullough & John, 1972). Acetate non-utilising mutants of ascomycete fungi have been shown to lack the genes encoding either MS or ICL (Armitt *et al.*, 1976; Ballance and Turner, 1986), showing that these genes are necessary for growth on this carbon source.

Metabolic control of the *icl* and *ms* genes has also been demonstrated using suspension cultures of anise. Removal of sucrose from the growth medium led to the appearance of ICL and MS activities. This effect was enhanced by the inclusion of acetate in the medium, but only in the absence of sucrose (Kudielka & Theimer, 1983a). However, if sucrose was reintroduced to the cells, the activities of the two enzymes was lost, 75 % of the ICL activity disappearing within 8 hours after addition of sucrose (Kudielka & Theimer, 1983b). It has been proposed that the metabolic status is an important factor in the regulation of both *ms* and *icl* gene expression during plant development (Graham *et al.*, 1992). Using a cucumber cell culture, it was demonstrated that starvation of the cells results in an induction of *icl* and *ms* gene expression, which is closely correlated with a drop in the levels of intracellular sucrose, glucose and fructose below threshold levels. Addition of glucose, fructose and raffinose results in a repression of expression as do mannose and 2-deoxyglucose, hexoses which can be phosphorylated, but not further metabolised. 3-methyl glucose, which can not be phosphorylated, does not lead to repression (Graham *et al.*, 1994b). It has therefore been proposed that the intracellular concentration of hexose sugars, or their flux into glycolysis, is the important factor regulating expression of these genes (Graham *et al.*, 1994b). The pathway by which changes in metabolite levels result in the induction of expression of the *icl* and *ms* genes remains to be elucidated.

In addition to controlling the expression of the *icl* and *ms* genes by addition to or removal from the surrounding medium of various metabolites, these genes can also be induced by starvation, by placing whole plants, detached photosynthetic organs or protoplasts in the dark (Gut & Matile, 1988; Birkhan & Kindl, 1990; Graham *et al.*, 1992; McLaughlin & Smith, 1994). Dark incubation has the effect of removing the plants own source of sugar, so necessitating the induction of another metabolic pathway to supply the energy requirements. Expression of the glyoxylate cycle genes during natural senescence of plant organs has also been widely demonstrated in a variety of plant species (Pistelli *et al.*, 1991; De Bellis *et al.*, 1991; Graham *et al.*, 1992) where it is thought to be involved in the utilisation of structural lipids. Senescence is a period of the plants life cycle when the levels of photosynthate will fall below those normally found in mature plant tissue. Hence, it would appear that levels of sugars do indeed affect expression of the *icl* gene. To assess whether this regulation also occurs at the level of transcription, the effect of metabolic status on expression of the *icl*-GUS fusion constructs in transformed *N. plumbaginifolia* seedlings was investigated.

4.8.2 Effect of metabolic status on GUS expression in transformed *N. plumbaginifolia* seedlings

Seeds of *N. plumbaginifolia* transformants pBI2900 5.1 or pBI572 5.1 were germinated and grown for 18 days in a 16 h photoperiod at 25 °C. After this time, seedlings were transferred to filter papers soaked either in sdd H₂O or 25 mM sucrose and incubated either in the light (16 hour photoperiod) or in the dark at 25 °C (section 2.9.4.5). Forty seedlings from each treatment were harvested after 5 days and assayed for GUS activity as described previously. The results are displayed in figure 4.14. For seedlings transformed with pBI2900, those in which starvation had been induced, by incubation in the absence of light and a carbon source, show levels of GUS activity considerably higher (approximately 4 fold) than the levels seen in those maintained in the light but without a source of carbon provided. The light incubated sample shows reduced activity compared to the day 18 sample, whereas the dark incubated sample shows an induction of activity. A comparison of those incubated in dark or light in the presence of sucrose again reveals a 4 fold higher level in the dark incubated sample. Comparing samples incubated in the dark, with or without sucrose, reveals a higher level of expression in those incubated in the absence of sucrose, as does the similar comparison of the two light incubated samples (both 1.5 fold higher in the absence of sucrose). These results indicate that removal of the photosynthetic capability of a plant

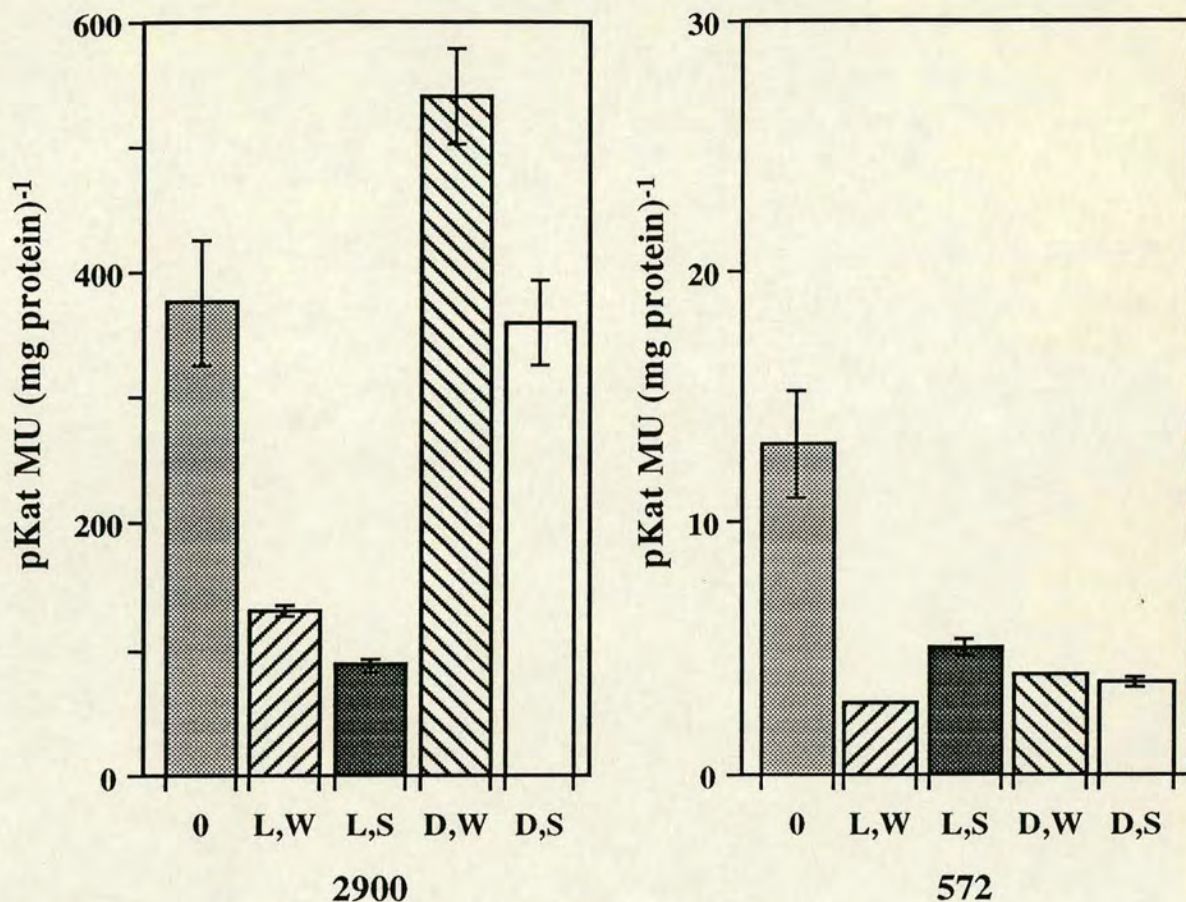


Figure 4.14 Glucuronidase activity of *N. plumbaginifolia* transformed with either pBI2900 or pBI572, under differing metabolic conditions. Seeds from transformant pBI2900 5.1 and pBI572 5.1 were germinated, grown for 18 days, transferred to filters and incubated under different conditions for 5 days. Forty seedlings for each treatment were analysed for levels of GUS activity, and the results from five experiments averaged. D,W = dark incubated on water; D, S = dark incubated on 25 mM sucrose; L,W = light incubated on water; L, S = light incubated on 25 mM sucrose.

(by dark incubation) and hence its source of sugar, leads to an induction of transcription from the *icl* gene promoter. This induction is partially inhibited by the external addition of sucrose and will be referred to as the 'starvation response'. In contrast, the results from pBI572 transformant 5.1 show no great difference in the levels of GUS activity between any of the treatments, implying that metabolic status has no effect on the expression when directed solely by the proximal 382 bp of the flanking region. This fragment presumably does not contain the necessary information to direct the starvation response. This implies that at least two elements must exist which control expression of the *icl* gene under different circumstances: an element responsible for the germination response which is present within the 382 bp closest to the start of transcription, and the element responsible for directing the starvation response, which is found within the 2700 bp of the promoter analysed here.

4.9 Conclusion

The results presented in this chapter show that the 6.5 kb fragment of the cucumber genome, isolated from a λ library, contains a fully functional *icl* gene. In cucumber, the *icl* gene is single copy, therefore all the elements necessary for the regulation of its transcription throughout development and in response to any metabolic or environmental signals must be contained within this one gene.

The experiments using *icl* promoter-GUS fusions have shown that 2700 bp of 5' flanking region contains all the elements necessary to direct qualitative transcription of the *icl* gene both during and post germination and also when the plant is starved, it contains both germination and starvation response elements. However, the level of expression is considerably lower in *N. plumbaginifolia* than in cucumber, so it is possible that some quantitative elements may be lacking from this region. This need not be the case. The quantitative expression of transgenes in foreign hosts is often seen to be below the levels observed in the natural genomic environment (Willmitzer, 1988).

The shorter promoter-GUS fusion with just 382 bp of the 5' flanking region also showed a qualitative germination response. However, the levels of activity from this construct were severely reduced compared to the longer one indicating that further elements for enhancing the levels of expression are contained within the additional 2300 bp present in pBI2900. No starvation response was seen in plants containing this short promoter-GUS fusion. Therefore, at least two distinct elements exist which are necessary for regulation of the gene during either the germination response or the starvation response, present within different regions of the promoter.

Other stages of plant development during which the *icl* and *ms* genes are known to be expressed are embryogenesis, senescence and pollen formation (Graham *et al.*, 1992; Zhang *et al.*, 1993; 1994). The expression from the fusion constructs has not been analysed during these developmental stages. From the high levels of GUS activity present within seeds immediately after imbibition, it would appear likely that the 2700 bp fragment upstream of the transcriptional start in pBI2900 contains the information necessary to direct a response during embryogenesis. Relatively high levels of GUS activity at day 0 were also detected in the seedlings transformed with pBI572. Therefore this element also may be predicted to carry an embryogenesis response element, but again probably with reduced transcriptional strength than would be produced from the longer fragment. This remains to be investigated.

The induction of the glyoxylate cycle genes during senescence is likely to be in response to a fall in the levels of photosynthate within the cells. It is possible that the response seen in senescent tissue could be controlled by the same *trans*-acting factors recognising the same *cis*-acting sequences as in the starvation response seen here. However, two other enzymes involved in lipid metabolism, phosphoenolpyruvate carboxykinase and glyoxysomal malate dehydrogenase are both induced in senescence, but are not induced by starvation (Kim & Smith, 1994). Therefore, at least in some genes, the elements controlling these two responses are unlikely to be the same.

So far, two regions of the cucumber *icl* gene have been broadly defined as containing important *cis*-acting elements, the 2300 bp fragment containing element(s) for the germination response and element(s) essential for metabolic repression, and the proximal 382 bp containing an element necessary for the qualitative aspects of the germination response. In the next chapter, experiments to further define the elements necessary for metabolic control of gene expression will be described.

CHAPTER 5

METABOLIC REGULATION OF THE *ICL* GENE

5.1 Introduction and Aims

It has been proposed previously that expression of the *ms* and *icl* genes is regulated indirectly by the availability of metabolites such as sucrose, so enabling their activities to increase in response to starvation conditions (Graham *et al.*, 1992). This has been demonstrated in cucumber cell cultures (Graham *et al.*, 1994b) and for MS in a cucumber protoplast transient assay system (Graham *et al.*, 1994a). In the preceding chapter, experiments using a stable transformation system were described which defined two broad areas of the *icl* promoter carrying elements responsible for the germination and the starvation responses (sections 4.6, 4.7 & 4.8). Using a transient expression system in cucumber protoplasts, the aim was to analyse further the element(s) necessary for control of *icl* gene expression in response to changes in metabolic status. The transient expression system has several advantages over the stable transformation system. Firstly it is an extremely rapid method: GUS expression from the introduced DNA can be measured within a few days as compared to the months required for stable transformation studies. Secondly, it is possible to use cucumber protoplasts, the analysis therefore taking place in the homologous host rather than in an unrelated species. The aim was to create deletions of the *icl* promoter, and to analyse the levels of GUS expression directed by each of them in protoplasts cultured either in the presence or in the absence of sucrose.

5.1.1 Transfection of protoplasts by electroporation

Electroporation is a method for introducing DNA into large numbers of protoplasts by means of electrically induced pores in the membrane. The general method was developed by Fromm *et al.* (1985), based on previously reported methods for electrical transfection of mouse cells (Wong & Neumann, 1982). Membrane permeabilisation is achieved when short DC pulses are applied, creating electropores. Macromolecules in the suspension medium are then able to enter or leave the cell, during the period in which the pores remain open. The extent of poration is dependent on the electrical parameters used and the constitution of the electroporation buffer, and the closing of the pores may be slowed by decreasing the temperature (Potter *et al.*, 1984). The uptake of DNA by the cells is an active process, faster than can be accounted for by diffusion alone (Dimitrov & Sowers, 1990), and the efficiency of uptake is dependent on the physical form of the DNA, single stranded giving rise to higher efficiencies than double stranded. For electroporation of plant protoplasts, the shape of the pulse delivered can affect the efficiency of transfection. Square wave pulse generators have a broader range

of efficient working conditions than do those generating an exponential decay pulse. However, transformation is only about 60 % efficient under optimal conditions, therefore transformants generated by a square wave pulse will be contaminated with significant numbers of viable parental protoplasts (Saunders *et al.*, 1989a). Minimising manipulation of the protoplasts directly after electroporation allows the nucleic acids time to move across the plasmalemma and also helps to maintain the viability of the protoplasts (Saunders *et al.*, 1989b).

5.1.2 The cucumber protoplast transient assay system

Transient expression systems can be useful for rapid analysis of DNA sequences important for transcriptional regulation of gene expression. However, the method involves the introduction of large numbers of plasmid molecules per cell, which are not integrated into the genome and so may influence the results by not accurately representing the true situation within a cell. Therefore, the accuracy of the system must be confirmed by analysis of the endogenous gene, to ensure that transcription directed by the introduced sequences reflects the pattern of transcription of the endogenous gene.

5.2 Isocitrate lyase transcript levels in protoplasts cultured on different carbon sources

Before using the transient expression system to analyse the effect of metabolic status on expression from the *icl* promoter, it was important to demonstrate that regulation of the introduced promoter is the same as that of the endogenous gene. Therefore, cucumber protoplasts were isolated as described in section 2.10.2 and cultured for 48 h in the dark either in the presence of 0.35 M mannitol or with 0.33 M mannitol plus 20 mM sucrose. It has been demonstrated that in a cucumber protoplast transient expression system, the same degree of repression from the *ms* promoter is obtained with 5 mM as with 60 mM sucrose (Graham *et al.*, 1994a), therefore 20 mM sucrose is a suitable level to use for the analysis of the *icl* promoter. The results are shown in figure 5.1. In the presence of sucrose, the level of expression of the endogenous cucumber gene falls to approximately 50 % of that seen when the protoplasts are cultured in mannitol alone. Similar results were seen when the levels of MS transcripts were analysed both in mesophyll protoplasts (Graham *et al.*, 1994a) and in cucumber cell cultures (Graham *et al.*, 1994b). Transcripts of both ICL and MS are detectable at low levels in mesophyll protoplasts incubated in the presence of sucrose, whereas the transcripts are not

detected in cucumber cell cultures incubated in the same concentration of sucrose. It is thought possible that the levels of transcripts seen in the protoplast system may be due to other factors such as stress affecting expression (Graham *et al.*, 1994a).

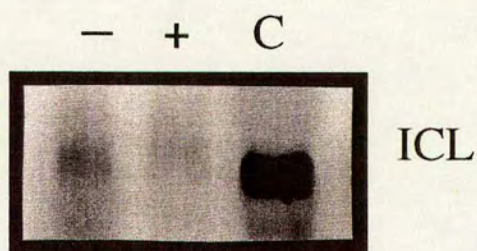


Figure 5.1 Northern blot analysis showing the effect of sucrose on ICL transcript levels in cucumber protoplasts. Protoplasts were isolated overnight in 0.35 M mannitol, then half of the preparation was transferred to 0.33 M mannitol and 20 mM sucrose (+), the other half remaining in 0.35 M mannitol (-), and incubated at 25 °C for 48 h. Total RNA was isolated and 10 µg loaded per lane. It was hybridised with the radiolabelled insert form pBSICL1.4. C = 10 µg total RNA from cucumber cotyledons, 3 days post imbibition.

5.3 Expression from the isocitrate lyase promoter in cucumber mesophyll protoplasts

The 2700 bp of the *icl* flanking region present in pRAJ2900 has been shown to be sufficient to direct faithful transcription in transgenic *N. plumbaginifolia* seedlings (chapter 4), and is therefore suitable for use in a transient expression system. In order to assess whether the *icl* promoter is affected by metabolic status in the cucumber protoplast system in the same manner as it was in *Nicotiana* seedlings treated with sucrose, pRAJ2900 was introduced into protoplasts by electroporation as described in section 2.10. To enable a comparison to be made, the cauliflower mosaic virus (CaMV) 35S promoter, which is not specifically regulated by the carbon source, linked to GUS was also introduced into protoplasts. One million protoplasts were used per sample which was electroporated using the conditions described in section 2.10.3. After electroporation, the protoplasts were diluted with culture medium containing either 0.35 M mannitol or 0.33 M mannitol plus 20 mM sucrose and incubated at 25 °C in the dark for 48 h. The protoplasts were pelleted, lysed and the extract assayed for GUS activity as described in section 2.10.4.1. The results from two experiments are presented in figure 5.2.

In the experiment depicted in figure 5.2a, GUS expression from the CaMV 35S promoter is found to be approximately 1.5 times higher when the protoplasts were incubated in the presence of 20 mM sucrose, than when incubated in mannitol alone; in that depicted in figure 5.2b, expression is 4 times higher in the presence of sucrose. This trend was seen in all experiments carried out and could be due to a general decrease in transcription which is likely to occur when the carbohydrate supply becomes limiting. However, when the levels of GUS activity directed from the *icl* promoter are observed, it can be seen that the trend is very different. GUS expression is approximately 6 times higher in the absence of sucrose than when the protoplasts were incubated in 20 mM sucrose. This trend was observed in numerous experiments. These results clearly confirm those findings with *N. plumbaginifolia* seedlings transformed with pBI2900 when incubated in the presence or absence of sucrose (section 4.8.2); in both cases, the level of GUS expression under the control of the 2.7 kb *icl* promoter is reduced in the presence of 20 mM sucrose. These changes in the levels of GUS expression reflect the changes seen in the levels of ICL transcripts in cucumber protoplasts cultured either in the presence or absence of sucrose. This transient expression system is suitable therefore for further analysis of the sequences necessary for the transcriptional regulation of the *icl* gene.

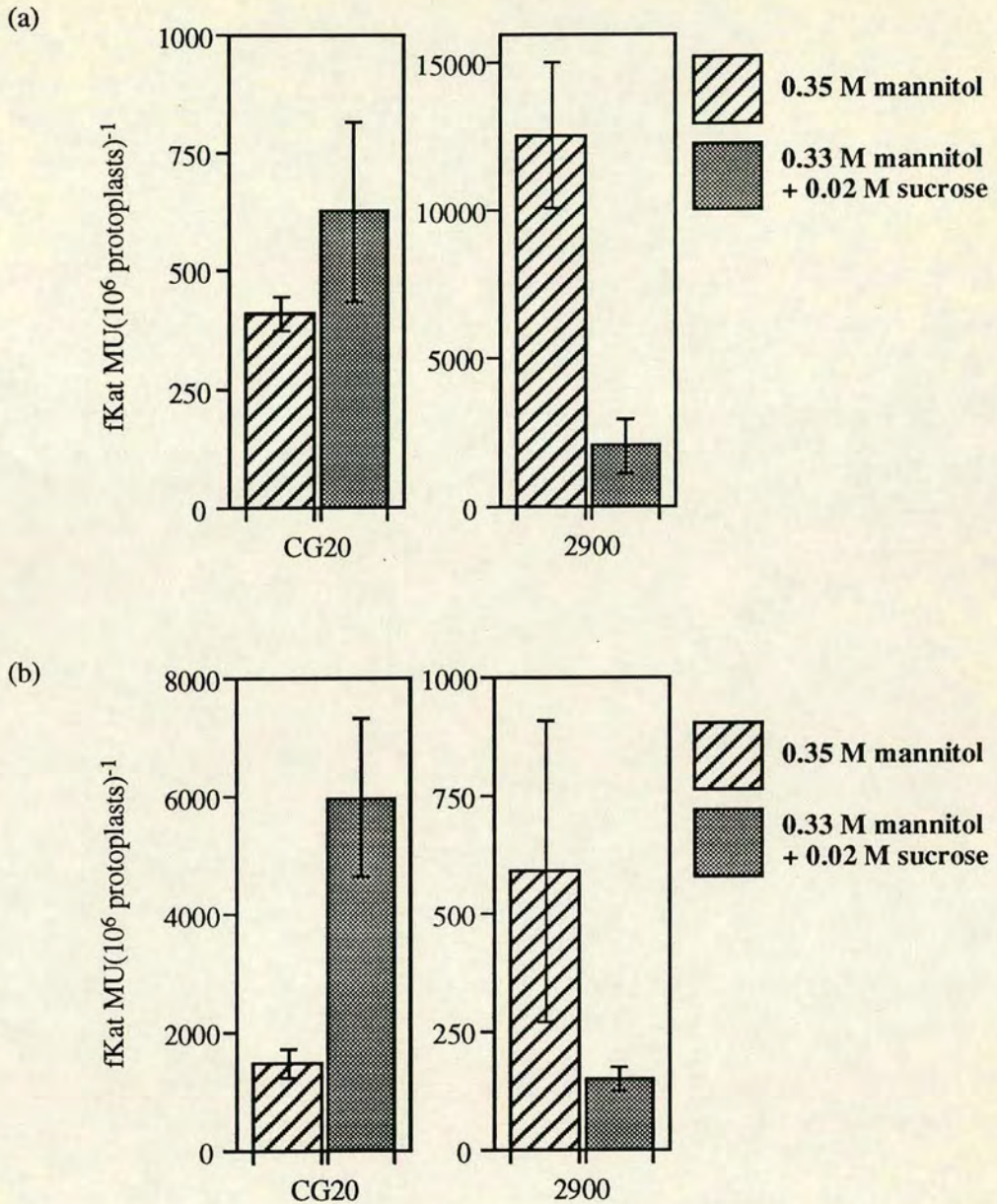


Figure 5.2 Analysis of GUS activity showing the effect of sucrose on transcription from the CaMV 35S (CG20) and *icl* (2900) promoters in cucumber protoplasts cultured in 0.35 M mannitol (striped bars) or 0.33 M mannitol plus 20 mM sucrose (stippled bars). The results from two experiments are presented (a and b), with each treatment having been repeated four times.

5.4 Deletion of the isocitrate lyase promoter and the effect of these deletions on expression from the promoter

5.4.1 Construction of promoter-GUS fusions

In order to further define the region(s) of the *icl* promoter important for controlling expression of the *icl* gene under differing metabolic conditions, a number of deletions of the 2.9 kb fragment were generated and fused to GUS allowing analysis in the transient assay system using cucumber protoplasts. Deletions of the promoter were created using the *Exo* III/S1 nuclease system as described in section 2.3.11 in the plasmid pBS containing the 2.1 kb *Xba* I-*Sal* I fragment (see figure 3.3). It was digested with *Sac* I to generate a protected 3' protruding terminus and with *Bam* HI to generate a 3' recessed terminus susceptible to *Exo* III digestion. Deletions were created and those of the desired size sequenced to map the exact extent of the deletion. Fragments of lengths 1570, 1091 and 570 nucleotides were selected for analysis in the transient assay system. The fragments were excised from pBS using *Hind* III and *Sal* I and inserted into pRAJ572, constructed as described in section 4.7.1, to generate pRAJ2142, pRAJ1663 and pRAJ1142, which contain the promoter fragments extending 1952, 1473 and 952 nucleotides upstream of the start of transcription. These constructs are depicted in figure 5.3.

5.4.2 Expression from the deletions of the isocitrate lyase promoter in cucumber mesophyll protoplasts

The effect of the deletions on the regulation of expression from the *icl* promoter was analysed in the cucumber protoplast transient assay system as described above. The results from two experiments are depicted in figure 5.4. Absolute levels of GUS activity were found to be variable between experiments, both between the same constructs and in the ratios of expression levels of the constructs. This is probably due to the nature of the transient expression system used in which multiple copies of plasmid were introduced into the cell, so making quantitative analysis difficult. However, an inspection of the ratios of GUS expression levels in protoplasts cultured in the absence of sucrose to those of protoplasts cultured in the presence of sucrose reveals that inhibition of transcription from the *icl* promoter occurred in those protoplasts into which constructs containing at least 1473 bp upstream of the start of transcription had been introduced (constructs pRAJ2900, 2142 and 1663), but that the starvation response was lost by deletion to within 952 bp of the transcriptional start

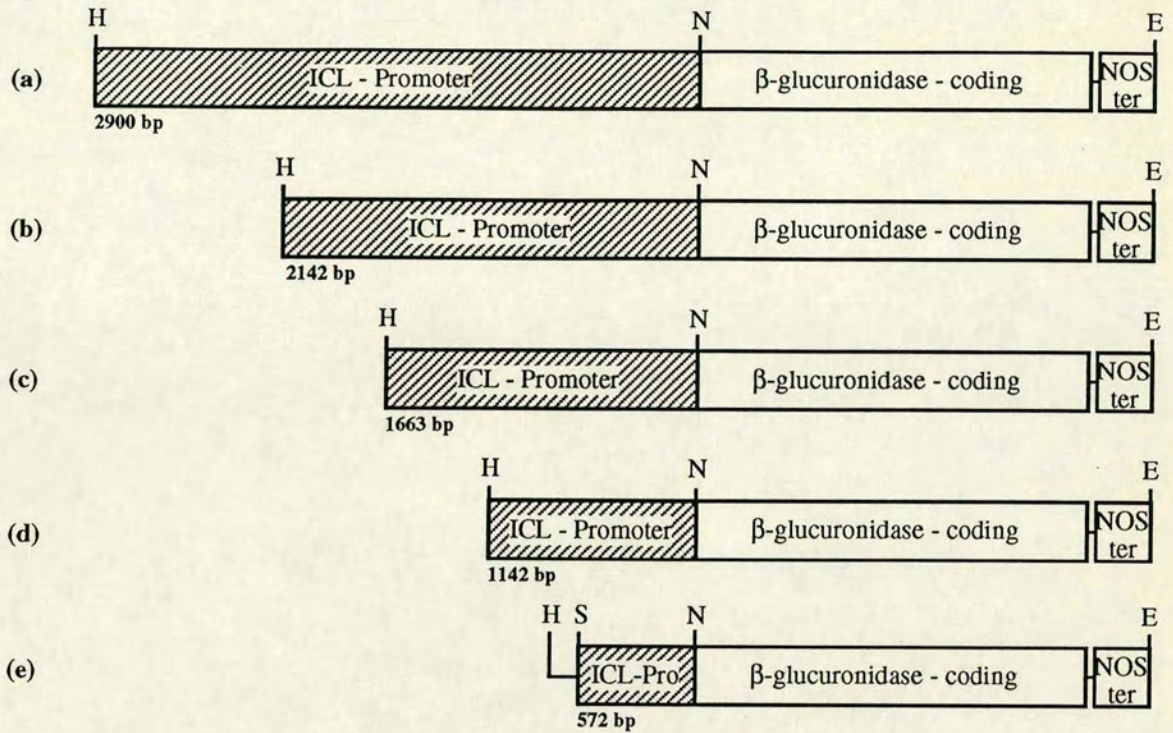


Figure 5.3 Diagram of the *icl* promoter deletion series generated using *Exonuclease III*. Numbers at the left hand end of each construct denote the number of nucleotides from the 5' end of the promoter to the start of translation. H = *Hind III*, N = *Nco I*, E = *Eco RI*. NOS ter = nopaline synthase terminator sequence.

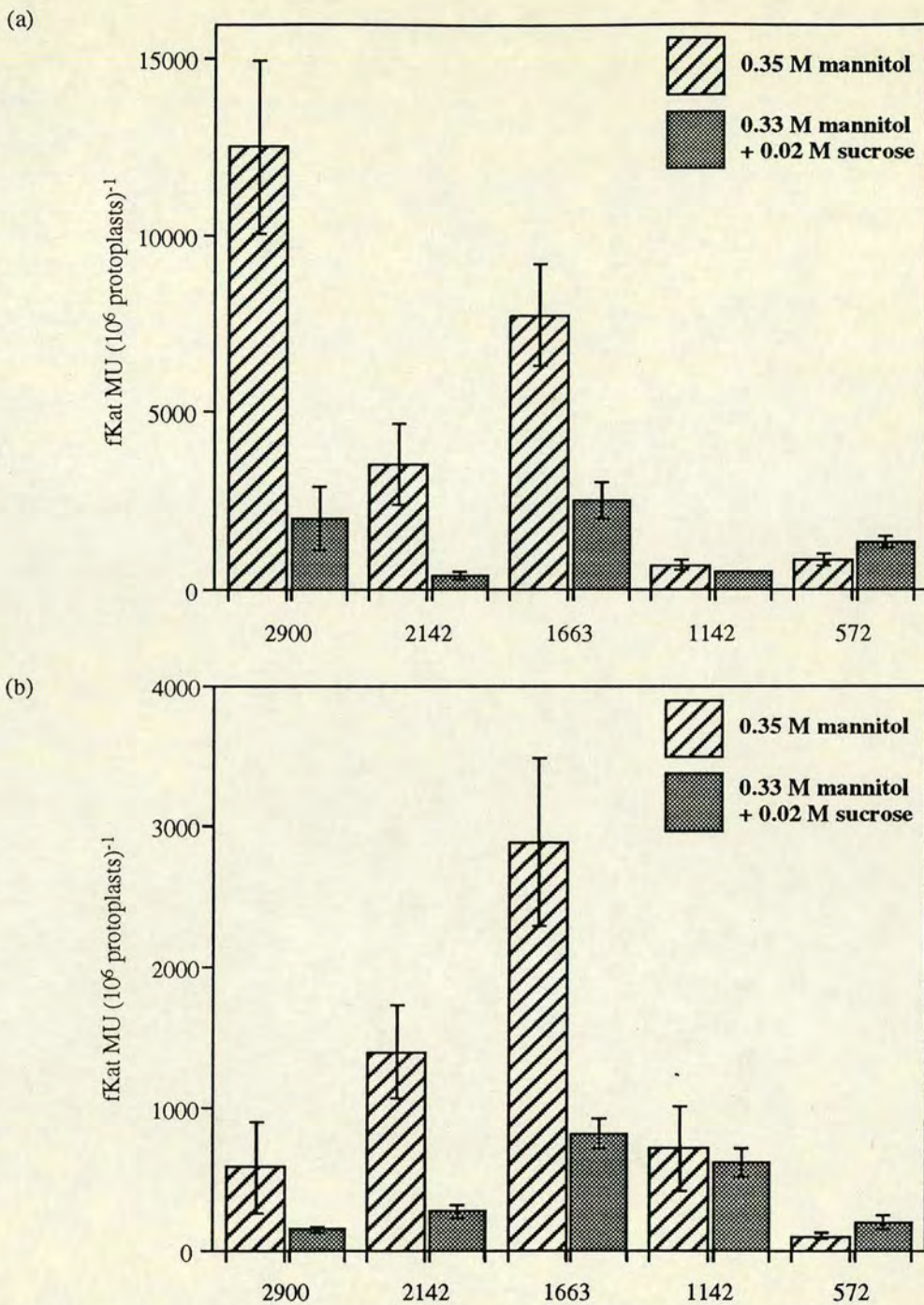


Figure 5.4 Transient expression analysis of deletions of the *icl* promoter showing the GUS activity of the different constructs in protoplasts cultured in 0.35 M mannitol (striped bars) or 0.33 M mannitol plus 20 mM sucrose (stippled bars). Numbers below the bars represent the construct depicted in figure 5.3. The results from two experiments are shown, each treatment was repeated four times.

(table 5.1). The 520 bp removed from pRAJ1663 to generate pRAJ1142 are therefore important for the metabolic regulation of the *icl* promoter in protoplasts. Deletions of the 572 bp fragment showed very low levels of GUS activity with no starvation response (results not shown).

Construct	CG20	pRAJ2900	pRAJ2142	pRAJ1663	pRAJ1142	pRAJ572
expt 1	2:3	6:1	9:1	3:1	1:1	1:2
expt 2	1:4	4:1	5:1	4:1	1:1	1:2

Table 5.1: Ratio of expression levels of GUS in cucumber protoplasts electroporated with different *icl* promoter-GUS fusions cultured in the absence of sucrose to those electroporated with the same construct and incubated in the presence of 20 mM sucrose. CG20 = CaMV 35S promoter linked to GUS, other constructs as in the text.

These results confirm and extend the findings of the stable transformation system reported in chapter 4. Under conditions of starvation, transgenic *N. plumbaginifolia* seedlings expressed the GUS gene when directed by 2700 bp of the *icl* promoter, but no increase in GUS expression in response to starvation was detected when directed by just 382 bp of the *icl* promoter. In this transient expression system, GUS expression directed by 2700 bp of the promoter is repressed when in the presence of sucrose, but when deprived of it, so simulating starvation conditions, expression is induced. This does not occur when expression is directed solely by the proximal 382 bp of the *icl* promoter.

5.5 Conclusion

In this chapter, metabolic regulation of both the endogenous *icl* gene and also of introduced promoter-GUS fusion constructs has been demonstrated in cucumber protoplasts. The patterns of expression from the transcriptional fusion constructs in protoplasts maintained on different carbon sources show that the *icl* gene is subject to metabolic regulation of transcription. Similar results have been obtained in studies using the *ms* promoter (Graham *et al.*, 1994a). Repression of GUS expression directed by 248 bp of the *ms* promoter was seen when the transfected protoplasts were cultured

in the presence of sucrose and also in the presence of phosphorylatable hexose sugars. A more extensive study of the metabolites able to cause repression has been carried out in cucumber cell cultures (Graham *et al.*, 1994b) where it was demonstrated that all phosphorylatable hexose sugars tested gave rise to a repression of transcription of the *ms* and *icl* genes, but that hexose sugars which were not phosphorylatable had no effect on the levels of transcription. The hypothesis proposed by Graham *et al.*, (1994b) that hexose sugars or the flux of hexose sugars into glycolysis via hexokinase are important for signalling nutritional status and hence induction of genes which can support the organism when deprived of these sources, is confirmed by the results presented here.

There is evidence for such a metabolic control of the expression of the *icl* and *ms* genes in other organisms. In yeast, both these genes, and in addition those encoding alcohol dehydrogenase and malate dehydrogenase, are regulated by carbon catabolite repression. Expression of the genes is repressed by the presence of glucose in the medium, and also by both 2-deoxyglucose and glucosamine: both of these compounds are phosphorylated but are not further catabolised in yeast (Witt *et al.*, 1966).

Genes encoding photosynthetic enzymes have also been shown to be repressed by sucrose and glucose, though their repression only occurs at much higher sugar concentrations than is needed for the repression of the *icl* and *ms* genes (Sheen, 1990; Krapp *et al.*, 1993). The metabolic repression of these photosynthetic genes overrides other regulatory signals, such as developmental stage specific signals. Which form of regulation for the *icl* gene is dominant, the germination response or the metabolic response, has so far not been determined.

The area of greatest homology to the RT sequence discussed in chapter 3 lies within the proximal 382 bp of the *icl* promoter. This region appears not to contain information sufficient to direct transcription in response to starvation, though this region does direct the germination response albeit at extremely low levels. However, there are several additional regions showing considerable homology to the RT sequence of the *ms* promoter, three of which lie in the 520 bp which have been shown to be necessary for sucrose repression, two lying on the sense strand, the third in the reverse orientation. Experiments using the RT sequence from the *ms* promoter have shown it to bind specifically a protein factor which has been partially purified (Graham *et al.*, 1994a). Whether this same factor would also bind the *icl* promoter remains to be elucidated. The role, if any, of this factor in the regulation of transcription of either promoter is yet to be determined, and its importance for developmental or metabolic regulation of transcription would need to be demonstrated.

Of the other three IMH sequences shown in figure 3.11, both IMH2 and IMH3 are in the region shown to be necessary for metabolic regulation, but are not essential for the

germination response. IHM4 is present in all the constructs analysed, therefore no conclusions as to its function can yet be drawn without further experimentation.

An upstream activating sequence has been defined in the 5' region of the *icl* gene of *S. cerevisiae*, that is necessary for the derepression of this gene when the organism is transferred onto medium lacking glucose or related sugars (Schöler & Schüller, 1994). This sequence has also been detected upstream of several other genes involved in the gluconeogenic pathway, such as the *ms* and fructose-1,6-bisphosphatase genes. However, the consensus sequence which was derived from these comparisons, 5' CATYCRTCCG 3', is not present in the 5' region of the cucumber *icl* gene.

The role of metabolites in the regulation of genes involved in the switch from carbohydrate to fatty acid metabolism is conserved between a variety of species and the molecular factors important for this regulation have begun to be elucidated, though much remains to be determined.

CHAPTER 6

SUMMARY AND FUTURE WORK

6.1 Isolation and characterisation of the isocitrate lyase gene of cucumber

The objective of the work presented in chapter 3 was to isolate and subsequently characterise the *icl* gene of cucumber. Southern analysis showed there to be a single gene encoding ICL in cucumber. Nucleotide sequence analysis of the genomic clone and comparison of the derived amino acid sequence with those from other higher plants and fungi revealed that ICL is a highly conserved protein. Analysis of the predicted amino acid sequence revealed the presence of a putative PTSI signal at the carboxy terminus, indicating that cucumber ICL may be targeted to the glyoxysomes by this mechanism rather than another involving sequences elsewhere within the protein. However, this sequence has been demonstrated to be unnecessary for the correct localisation of castor bean ICL to the glyoxysomes in an *in vitro* import system (Behari & Baker, 1993). Therefore further experimentation is needed before the mechanism of cucumber ICL import to glyoxysomes is established. By analysis of transcript levels of ICL and MS, the two genes were shown to be coordinately expressed during postgerminative growth. This raises the possibility that the regulatory regions of the two genes contain similar *cis*-acting sequences responsible for the control of gene expression at this stage of development. Since there is a single *icl* gene in cucumber, all the *cis*-acting sequences necessary for the regulation of its expression must be present within the one unit. Comparison of the 5' flanking region with that necessary for the regulation of *ms* gene expression revealed the presence of several highly conserved sequences, designated RT (IMH1) and IMH2 to 4. The significance of these sequences in the regulation of *icl* gene expression was therefore tested by use of promoter-reporter gene fusions.

6.2 Developmental and metabolic regulation of isocitrate lyase and malate synthase

The aim of the work presented in chapters 4 and 5 was to extend the knowledge of the factors involved in the regulation of the *icl* gene which was known to be expressed at several stages of development and in a number of tissues. In addition to its developmental regulation, the *icl* gene has been shown to be expressed when the cell is subjected to starvation conditions. The results presented here have shown that regulation of the *icl* gene is complex, with more than one control element being needed to regulate correctly *icl* gene expression.

Regulation of any gene showing a complex pattern of expression would be expected to involve multiple elements, as demonstrated by two recent reports. Regulation of the gene encoding the tuber protein, patatin, has been shown to need different *cis*-acting sequences and *trans*-acting factors for sucrose-responsiveness to those needed for tissue-specific expression (Grierson *et al.*, 1994). Similarly, several promoter elements regulate the maize alcohol dehydrogenase gene, different combinations being required for tissue- and developmental-specific expression (Kyojuka *et al.*, 1994).

Using a stable transformation system in combination with a transient assay system, it has been shown that more than one element is responsible for controlling *icl* gene expression during postgerminative growth and under sugar starvation conditions. The elements necessary for controlling expression under these different situations are physically separable. Two regions of the *icl* promoter have been identified that are important either for metabolic or for developmental regulation of the *icl* gene. An element essential for the starvation response has been localised to between -1473 and -952 bp from the start of transcription. Both constructs analysed in transgenic *Nicotiana*, pBI2900 and pBI572 (approximately 2700 and 382 bp from the transcriptional start respectively), showed the germination response. Therefore, an element(s) necessary for the qualitative, though not the quantitative, germination response is present to within -382 bp of the transcriptional start.

Similar analyses using *ms* promoter-GUS fusions have been carried out. Using the same transient assay system, a 123 bp fragment of the *ms* promoter, from -248 bp to -125 bp relative to the start of transcription, has been shown to be responsible for transcriptional control of the *ms* gene. This element is sufficient to lead to the induction of GUS gene expression when protoplasts electroporated with this promoter fragment fused to the GUS gene were cultured in the absence of metabolisable sugars (Graham *et al.*, 1994a). A series of deletions of this 123 bp fragment have been constructed, linked to the GUS gene, and analysed in transgenic *Nicotiana*, both during postgerminative growth and under conditions of starvation. Deletions up to and including -199 bp show the germination response, but the starvation response is very weak after deletion to -233 bp and completely absent from -216 bp and subsequent deletions (Sarah & Smith, pers. comm.).

Gel retardation studies with a fragment containing one of the conserved sequences, RT, has shown it to bind a protein; the protein interaction with the RT element is sequence specific as it is competed when an unlabelled RT oligonucleotide is included in the reaction. The RT element lies at -215 bp to -202 bp relative to the start of transcription in the *ms* promoter, and therefore has been shown not to be essential for the germination response of the *ms* gene, but is perhaps important for metabolic regulation.

However, this element is present in all the *icl*-GUS constructs analysed here, so although it may be necessary for metabolic regulation, it is not sufficient, as deletion to -952 bp from the transcriptional start abolishes the starvation response. IMH2 lies in the regions of both promoters shown to be necessary for metabolic regulation, and may therefore be considered a candidate for the metabolic response element. Two of the other conserved sequences, IMH3 and IMH4, both lie in the region shown to be essential for the metabolic response of the *icl* gene. However, both lie outside the region in the *ms* promoter necessary for a metabolic response, both in the transient assay system (Graham *et al.*, 1994a) and in transgenic *Nicotiana* (Sarah & Smith, pers. comm.). This does not necessarily exclude these sequences from a role in metabolic regulation. Activation of transcription frequently requires the binding of more than one factor to the regulatory regions, the factors often working synergistically to produce a greater response in combination than the sum of the individual responses. Although the *ms* promoter deletions analysed show a metabolic response, it is not so great as that seen when a 1033 bp fragment linked to the GUS gene is similarly analysed (Sarah & Smith, pers. comm.). Therefore, it is possible that elements present within the 233 bp fragment are sufficient alone to produce a weak metabolic response, but others present further upstream are also necessary for the level of response normally found within the plant. Likewise, when deletion of a segment of DNA abolishes a particular response, it does not necessarily demonstrate that all the sequences required for this response are present in that region, merely that an element(s), without which the response does not occur, resides in that region. Therefore, care must be taken when interpreting such results and in designing future experiments. The arrangement of regulatory elements and conserved sequences in the *icl* and *ms* promoters is summarised in figure 6.1.

In order to define more precisely the regions of the *icl* promoter needed for the different levels of regulation, further deletions need to be constructed and analysed. A metabolic response element lies between -1473 and -952 bp from the transcriptional start. Deletions of this region can be linked to GUS and analysed both in the transient assay system and in a stable transformation system, either using *Nicotiana* as in this thesis, or using *Agrobacterium rhizogenes* to generate transformed hairy roots of cucumber to which metabolites may be fed. The proximal 382 bp of the promoter have been shown to contain a germination response element. Further definition of this sequence will be less easy as the levels of GUS activity directed by the 382 bp directly upstream of the transcriptional start are already extremely low. The level of GUS activity directed by this promoter fragment may be enhanced at the level of translation through the inclusion of translational enhancers such as that present within the alfalfa mosaic virus 4 RNA (Anderson *et al.*, 1989).

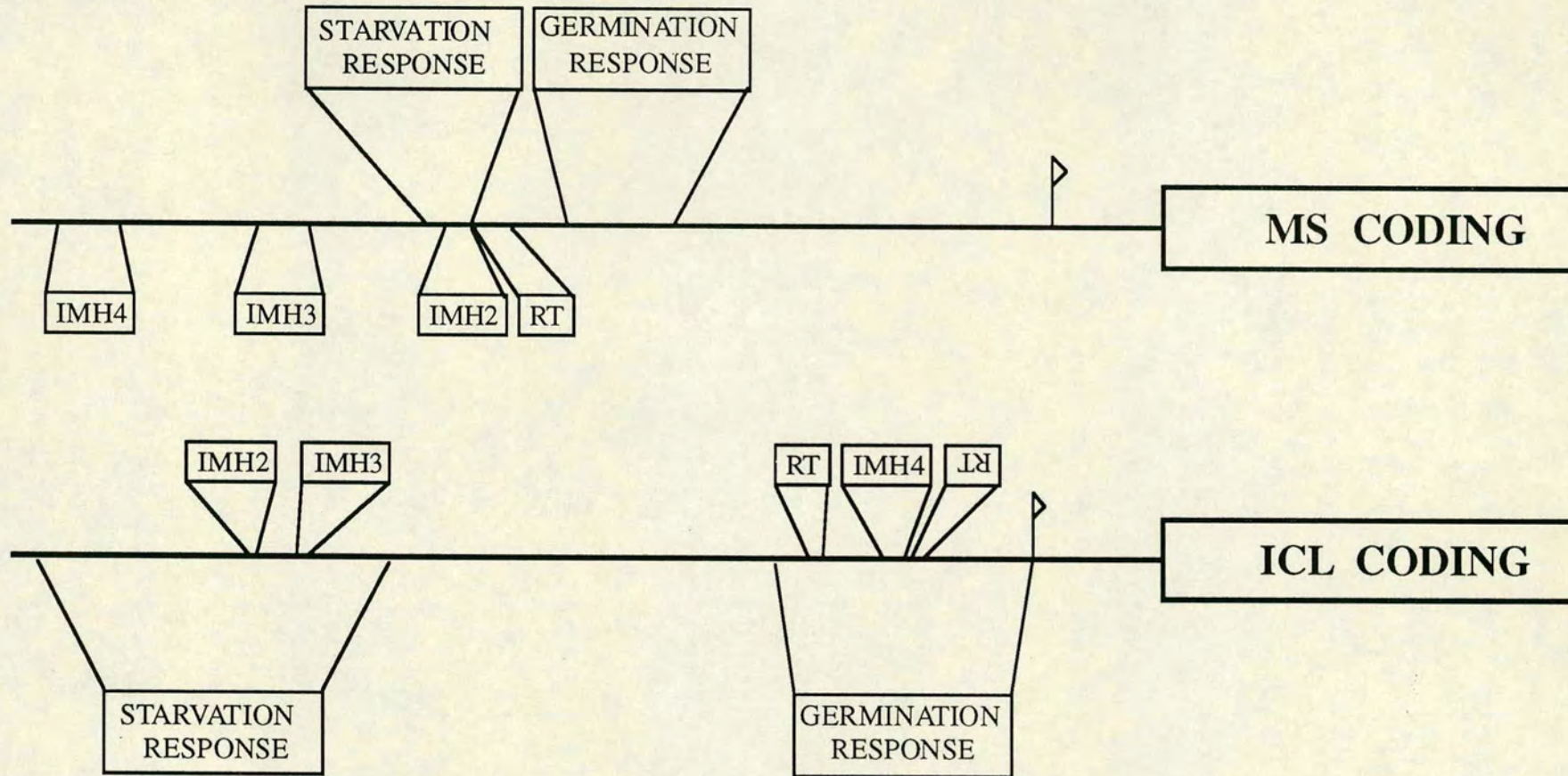


Figure 6.1 Summary diagram of the arrangement of germination and starvation response elements and conserved sequences in the *icl* and *ms* promoters. The start of transcription is indicated by a flag. The scale of the *ms* promoter is 4x that of the *icl* promoter.

Investigations into the effect of glucose on the activity of ICL and MS in germinating seedlings has indicated the possibility that a hierarchy of control occurs in the regulation of the *icl* and *ms* genes. In peanut cotyledons imbibed for seven days in 0.11 M glucose, the activity of ICL was inhibited by 65 % and that of MS by 45 % compared to controls incubated in water (Longo & Longo, 1970). In excised cotyledons of germinating castor bean seedlings, ICL activity is inhibited by 75 % when incubated in the presence of 0.1 M glucose for 8 hours compared to water controls (Lado *et al.*, 1968). A similar study using excised cotyledons of germinating squash revealed 48 % inhibition of ICL activity after 24 hours and 28 % inhibition after 48 hours (Lado *et al.*, 1968).

These results in combination with the findings presented in this thesis suggest that metabolic regulation is independent of, and dominant to, developmental regulation. In a similar manner, metabolic regulation has been shown to override developmental regulation of some photosynthetic genes (Sheen, 1990). The levels of ICL transcripts in postgerminative cucumber cotyledons rise to a peak around day 3 and thereafter decline rapidly when incubated in the light. If the cotyledons are incubated in darkness, this decline is not seen; instead, transcripts persist until at least day 8 (see section 3.8). The developmental response in this situation is for transcript levels, and hence enzyme activity, to rise and subsequently decline. The persistence of transcripts beyond day 4 in dark-incubated, therefore 'starved' seedlings could be an indication of metabolic regulation overriding developmental regulation at this time. The germination response in dark incubated transgenic *Nicotiana* seedlings was not tested in this thesis. If the metabolic response does override the developmental one, it would be anticipated that seedlings containing the 2900 bp *icl*-GUS fusion would show GUS activity persisting, but that those containing the 572 bp *icl*-GUS fusion would exhibit the same pattern of GUS activity in the light and dark. Such analyses could also be carried out on plants into which the constructs containing the 1473 bp and 952 bp upstream of transcriptional start had been introduced. These constructs gave higher activity than the proximal 382 bp fragment in the transient assay system and consequently would be easier to analyse.

Additionally, the effect of germinating seedlings in the presence of a metabolisable sugar could be investigated with *Nicotiana* plants transformed with the various promoter-GUS fusions. The two studies detailed above (Lado *et al.*, 1968; Longo & Longo, 1970) used around 100 mM glucose to produce the inhibitory effects. This is very high in comparison to the concentration of sucrose used in the work presented in this thesis, particularly since 5 mM sucrose was shown to effect maximal inhibition of *ms* gene expression in cucumber protoplasts (Graham *et al.*, 1994a). Such work would

therefore be better carried out with, for example, 20 mM glucose or sucrose as in this thesis. Transgenic *Nicotiana* containing constructs already shown to contain a starvation response element would be expected to show inhibition, whereas those containing shorter promoter fragments such as the 1142 or 572 bp fusions would be expected to exhibit the same pattern of GUS activity in the presence or absence of metabolisable sugar.

Expression patterns of *icl* and *ms* genes appear to be conserved when the genes are expressed in heterologous hosts. The genes encoding MS from *A. nidulans* and *N. crassa* are correctly expressed when transferred from one species to another. In addition, they still show the same patterns of inducibility (Sandeman *et al.*, 1991). Likewise, *icl* and *ms* gene expression patterns are maintained when transferred from cucumber (Graham *et al.*, 1990) and *Brassica* (Comai *et al.*, 1992; Zhang *et al.*, 1994) into *Nicotiana*. This suggests that the *cis*-acting sequences and *trans*-acting factors responsible for the regulation of these genes are conserved at least within fungi and within higher plants. Therefore, comparison of the promoter sequences of *icl* and *ms* with those from other higher plants could yield interesting information. However, to date no other higher plant *icl* or *ms* promoter sequences are available.

6.3 Further analysis of factors involved in the regulation of isocitrate lyase and malate synthase

In addition to postgerminative growth and in response to starvation, ICL and MS are also synthesised during pollen formation (Zhang *et al.*, 1994), late embryogenesis (Comai *et al.*, 1989; Turley & Trelease, 1990) and senescence (De Bellis *et al.*, 1991; Graham *et al.*, 1992; Pistelli *et al.*, 1991). Analysis of regulation directed by the cucumber *icl* promoter in transgenic *Nicotiana* during these stages of plant development has not yet been undertaken. Both plants containing the 2900 bp *icl*-GUS fusion and those containing the 572 bp *icl*-GUS fusion almost certainly show an embryogenesis response as GUS activity was detectable in seeds tested at day 0 which declined before the major peak seen at stage 4 of postgerminative growth. This would imply that the proximal 382 bp of the *icl* promoter contains elements to direct an embryogenesis response albeit at very reduced levels. Further delineating the sequences necessary would pose the same problem as defining those necessary for the germination response as GUS activity directed by this fragment is so low. Again, the construction of a vector with a sequence to enhance the level of the response would assist such analysis (see section 6.2).

ICL has been detected in all pigmented tissues during senescence. Expression of GUS directed by the various length promoter fragments could also be analysed in senescent tissues of *Nicotiana*. It seems unlikely that the senescence response is a simple metabolic response, as the genes encoding phosphoenolpyruvate carboxykinase (Kim & Smith, 1994) and glyoxysomal malate dehydrogenase (Kim & Smith, pers. comm.) are coordinately expressed with those encoding *icl* and *ms* during senescence, but are not induced by starvation of tissues. Therefore, further elements for directing a senescence specific response would be anticipated, in addition to those controlling the metabolic response.

Once sequences have been sufficiently defined, site directed mutagenesis may help to determine which residues are important for making specific contacts with DNA binding proteins. Although SDM analysis allows firm statements to be made about the relevance of a control element without necessitating global disruption of the promoter region, limitations to this approach exist. This is well exemplified in analysis of the contacts made by the yeast GAL4 transcriptional activator, where in a 17 bp half site, specific contacts are only made with the outer 3 residues (Ptashnè, 1992).

Future efforts should concentrate on characterising *trans*-acting factors which associate with the defined *cis*-acting sequences and the specific contacts made in the DNA-protein complexes. One factor shown to bind to RT has already been identified and partially purified, but which level of regulation this factor may be responsible for has not yet been determined. Nuclear extracts from the different developmental stages and from tissues which have been starved could each be tested using oligonucleotides specifying the defined *cis*-acting sequence. Electrophoretic mobility shift assays ('band shifts'), performed both with and without unlabelled oligonucleotide as specific competitor, would identify sequence element interactions with regulatory proteins (Garner & Revzin, 1981). Sequence specific DNA binding proteins may be identified from crude extracts by ultraviolet (uv) crosslinking. After formation of a DNA-protein complex, irradiation with uv light produces purine and pyrimidine free radicals. If a protein molecule is in close proximity to the free radical, a covalent bond is formed, so crosslinking the protein to the DNA. The molecular weight of the bound protein may be accurately determined by SDS-PAGE of the DNA-protein complex, after digestion of free DNA. The specificity of the reaction may be determined by the inclusion of competitor DNA to compete for binding sites (Chodosh *et al.*, 1986).

DNase I footprint analysis ('footprinting') can be used to locate the specific binding sites of proteins on the DNA elements. The method involves the protection by bound protein of the phosphodiester backbone of an end labelled DNA fragment (probe) from DNase I catalysed hydrolysis. After hydrolysis of bound and free probe, binding sites

are determined by autoradiography of the electrophoresis products (Galas & Schmitz, 1978). One limitation of footprinting is that if binding is incomplete, or if protein exchange takes place during digestion, cutting of the protected site may occur. Methylation interference may be used to obtain more information about the specific residues of a sequence which are in close proximity to the protein. Dimethyl sulphate is used to methylate guanine and adenine residues, at an average of one site per molecule, in an end-labelled oligonucleotide. Methylation of the residues inhibits close binding of the protein to the DNA. Bound probe is then separated from free probe by bandshifts, the DNA cleaved with piperidine, separated by gel electrophoresis and subjected to autoradiography. Guanine and adenine residues that interfere with binding when methylated are identified by their absence in the retarded complex when compared to piperidine-cleaved free probe (Hendrickson & Schlieff). A similar method has been developed using depurinated or depyrimidated DNA as a probe, thus also yielding information on contacts made to pyrimidines within the binding site (Brunelle & Schlieff, 1987). The advantage of these two methods is that they detect specific residues rather than just a general area. However, they only demonstrate that residues are in close proximity to a bound protein, not that they are responsible for making direct contacts.

Subsequent purification and sequence analysis will identify the binding proteins and allow comparison with sequences of known transcription factors. Transcription factors responsible for regulation can be purified biochemically on the basis of their ability to bind the *cis*-acting elements. A combination of such techniques may be employed in conjunction with the approaches described in this thesis to further elucidate the many factors involved in regulation of the expression of the *icl* and *ms* genes of cucumber.

REFERENCES

- Aitchison J. D., Nuttley W. M., Szilard R. K., Brade A. M., Glover J. R. & Rachubinski R. A. (1992) Peroxisome biogenesis in yeast. *Mol. Microbiol.* 6: 3455 - 3460
- Allen R. D., Trelease R. N. & Thomas T. L. (1988) Regulation of isocitrate lyase gene expression in sunflower. *Plant Physiol.* 86: 527 - 532
- Anderson S., Davis D. N., Dahlbach H., Jornvall H. & Russell D. W. (1989) Cloning, structure and expression of the mitochondrial cytochrome P450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *J. Biol. Chem.* 264: 8222 - 8229
- Armitt S., McCullough W. & Roberts C. F. (1976) Analysis of acetate non-utilising (*acu*) mutants in *Aspergillus nidulans*. *J. Gen. Microbiol.* 92: 263 - 282
- Atomi H., Ueda M., Hikida M., Hishida T., Teranishi Y. & Tanaka A. (1990) Peroxisomal isocitrate lyase of the n-alkane-assimilating yeast *Candida tropicalis*: gene analysis and characterisation. *J. Biochem.* 107: 262 - 266
- Ballance D. J. & Turner G. (1986) Gene cloning in *Aspergillus nidulans*: isolation of the isocitrate lyase gene (*acuD*). *Mol. Gen. Genet.* 202: 271 - 275
- Barth G. & Scheuber T. (1993) Cloning of the isocitrate lyase gene (ICL1) from *Yarrowia lipolytica* and characterisation of the deduced protein. *Mol. Gen. Genet.* 241: 422 - 430
- Becker W. M., Leaver C. J., Weir E. M. & Reizman H. (1978) Regulation of glyoxysomal enzymes during germination of cucumber. 1. Developmental changes in cotyledonary protein, RNA and enzyme activities during germination. *Plant Physiol.* 62: 542 - 549
- Beeching J. R. & Northcote D. H. (1987) Nucleic acid (cDNA) and amino acid sequences of isocitrate lyase from castor bean. *Plant Mol. Biol.* 8: 471 - 475
- Beevers H. (1979) Microbodies in higher plants. *Ann. Rev. Plant Physiol.* 30: 159 - 193
- Behari R. & Baker A. (1993) The carboxy terminus of ICL is not essential for import into glyoxysomes in an *in vitro* system. *J. Biol. Chem.* 268: 7315 - 7322
- Behrends W., Engeland K. & Kindl H. (1988) Characterisation of two forms of the multifunctional protein acting in fatty acid β -oxidation. *Arch. Biochem. Biophys.* 263: 161 - 169
- Benfey P. N. & Chua N-H. (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* 250: 959 - 966
- Benkel B. I. & Hickey D. A. (1987) A *Drosophila* gene is subject to glucose repression. *Proc. Natl. Acad. Sci. USA* 84: 1337 - 1339
- Benton W. D. & Davis R. W. (1977) Screening λ gt recombinant clones by hybridisation to single plaques *in situ*. *Science* 196: 180 - 182

- Berg J. M. (1993) Zinc finger proteins. *Curr. Opin. Struct. Biol.* 3: 11 - 16
- Berget S. M. (1984) Are U4 small nuclear ribonucleoproteins involved in polyadenylation? *Nature* 309: 179 - 182
- Bevan M. W. (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucl. Acids Res.* 12: 8711 - 8721
- Birkhan R. & Kindl H. (1990) Re-activation of the expression of glyoxysomal genes in green plant tissue. *Z. Naturforsch* 45: 107 - 111
- Birnboim H. C. & Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7: 1513 - 1523
- Birnstiel M. L., Busslinger M. & Struhl K. (1985) Transcription termination and 3' processing: the end is in site! *Cell* 41: 349 - 359
- Borst P. (1989) Peroxisome biogenesis revisited. *Biochim. Biophys. Acta* 1008: 1 - 13
- Borthwick A. C., Holms W. H. & Nimmo H. G. (1984) The phosphorylation of *Escherichia coli* isocitrate dehydrogenase in intact cells. *Biochem. J.* 222: 797 - 804
- Bortman S. J., Trelease R. N. & Miernyk J. A. (1981) Enzyme development and glyoxysome characterisation in cotyledons of cotton seeds. *Plant Physiol.* 68: 82 - 87
- Bowyer P., De Lucas J. R. & Turner G. (1994) Regulation of expression of the isocitrate lyase (*acuD*) gene of *Aspergillus nidulans*. *Mol. Gen. Genet.* 242: 484 - 489
- Bradford M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248 - 254
- Brand A. H., Breeden J., Abraham R., Sternglanz R. & Nasmyth K. (1985) Characterisation of a 'silencer' in yeast: A DNA sequence with properties opposite to those of an transcriptional enhancer. *Cell* 41: 41 - 48
- Brand A. H., Micklem G. & Nasmyth K. (1987) A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. *Cell* 51: 709 - 719
- Brown P. H., Daniels-McQueen S., Walden W. E., Patino M. M. & Gaffield L. (1989) Requirements for the translational repression of ferritin transcripts in wheat by a 90 kDa protein from rabbit liver. *J. Biol. Chem.* 264: 13383 - 13386
- Brunelle A. N. & Schlieff R. (1987) Missing contact probing of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* 84: 6673 - 6676
- Brunelle A. N. & Chua N-H. (1993) Transcription regulatory proteins in higher plants. *Curr. Opin. Genet. Dev.* 3: 254 - 258
- Burley S. K. (1994) DNA-binding motifs from eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* 4: 3 - 11
- Castresan C., Garcia-Luque I., Alonso E., Malik U. S. & Cashmore A. R. (1988) Both positive and negative regulatory elements mediate expression of a photoregulated CAB gene from *Nicotiana plumbaginifolia*. *EMBO J.* 7: 1929 - 1936

- Chodosh L. A., Carthew R. W. & Sharp P. A. (1986) A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. *Mol. Cell. Biol.* 6: 4723 - 4733
- Choinski J. S. Jr. & Trelease R. N. (1978) Control of enzyme activities in cotton cotyledons during maturation and germination II. Glyoxysomal enzyme development in embryos. *Plant Physiol* 62: 141 - 145
- Choy H. E. & Adhya S. (1992) Control of *gal* transcription through DNA looping: inhibition of the initial transcribing complex. *Proc. Natl. Acad. Sci. USA* 89: 11264 - 11268
- Comai L., Dietrich R. A., Maslyar D. J., Baden C. S. & Harada J. J. (1989) Coordinate expression of transcriptionally regulated isocitrate lyase and malate synthase genes in *Brassica napus* L. *Plant Cell* 1: 293 - 300
- Comai L., Matsudaira K. L., Heupel R. C., Dietrich R. A. & Harada J. J. (1992) Expression of a *Brassica napus* malate synthase gene in transgenic tomato plants during the transition from late embryogeny to germination. *Plant Physiol.* 98: 53 - 61
- Cooper T. G. & Beevers H. (1969a) Mitochondria and glyoxysomes from castor bean endosperm. *J. Biol. Chem.* 244: 3507 - 3513
- Cooper T. G. & Beevers H. (1969b) β -oxidation in glyoxysomes from castor bean endosperm. *J. Biol. Chem.* 244: 3514 - 3520
- Cortay J-C., Nègre D., Galinier A., Duclos B., Perrière G. & Cozzone A. J. (1991) Regulation of the acetate operon in *Escherichia coli*: purification and functional characterisation of the IclR repressor. *EMBO J.* 10: 675 - 679
- Courtois-Verniquet F. & Douce R. (1993) Lack of aconitase in glyoxysomes and peroxisomes. *Biochem J.* 294: 103 - 107
- Covello P. S. & Gray M. W. (1993) On the evolution of RNA editing. *TIG* 9: 265 - 268
- Davis W. L., Goodman D. B. P., Crawford L. A., Cooper O. J. & Matthews J. L. (1990) Hibernation activates glyoxylate cycle and gluconeogenesis in black bear brown adipose tissue. *Biochim. Biophys. Acta* 1051: 276 - 278
- Davis W. L. & Goodman D. B. P. (1992) Evidence for the glyoxylate cycle in human liver. *Anat. Rec.* 234: 461 - 468
- De Bellis L., Picciarelli P., Pistelli L. & Alpi A. (1990) Localisation of glyoxylate-cycle marker enzymes in peroxisomes of senescent leaves and green cotyledons. *Planta* 180: 435 - 439
- De Bellis L. & Nishimura M. (1991) Development of enzymes of the glyoxylate cycle during senescence of pumpkin cotyledons. *Plant Cell. Physiol.* 32: 555 - 561
- De Bellis L., Tsugeki R. & Nishimura M. (1991) Glyoxylate cycle enzymes in peroxisomes isolated from petals of pumpkin (*Curcubita* sp.) during senescence. *Plant Cell Physiol.* 32: 1227 - 1235

- De Bellis L., Hayashi M., Biagi P. P., Hara-Nishimura I., Alpi A. & Nishimura M. (1994) Immunological analysis of aconitase in pumpkin cotyledons: the absence of aconitase in glyoxysomes. *Physiologia Plantarum* 90:757 - 762
- De Duve C. (1983) Microbodies in the living cell. *Sci. Amer.* 248: 74 - 84
- De Hoop M. J. & Ab G. (1992) Import of proteins into peroxisomes and other microbodies. *Biochem J.* 286: 657 - 669
- De Lucas J. R., Gregory S. & Turner G. (1994) Analysis of the regulation of the *Aspergillus nidulans acuD* gene, encoding isocitrate lyase, by construction of a hybrid promoter. *Mol. Gen. Genet.* 243: 654 - 659
- Dean C., Tamaki S., Dunsmuir P., Favreau M., Katayama C., Dooner H. & Bedbrook J. (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nucl. Acids Res.* 14: 2229 - 2240
- Dellaporta S. L., Wood J. & Hicks J. B. (1983) A plant DNA miniprep: Version II. *Plant Mol. Biol. Rep.* 1: 19 - 21
- Devereux J., Haeberli P. & Smithies O. (1984) A comprehensive set of sequence analysis programmes for the VAX. *Nucl. Acids Res.* 12: 387 - 395
- Dietrich R. A., Radke S. E. & Harada J. J. (1992) Downstream DNA sequences are required to activate a gene expressed in the root cortex of embryos and seedlings. *Plant Cell* 4: 1371 - 1382
- Dieuaide M., Couée I., Pradet A. & Raymond P. (1993) Effects of glucose starvation on the oxidation of fatty acids by maize root tip mitochondria and peroxisomes: evidence for mitochondrial fatty acid β -oxidation and acyl-CoA dehydrogenase activity in a higher plant. *Biochem J.* 296: 199 - 207
- Dimitrov D. S. & Sowers A. E. (1990) Membrane electroporation - fast molecular exchange by electroosmosis. *Biochim. Biophys. Acta* 1022: 381 - 392
- Ditta G., Stanfield S., Corbin D. & Helinski D. R. (1980) Broad host range cloning system for gram negative bacteria - Construction of a gene bank of *Rhizobium melilotis*. *Proc. Natl. Acad. Sci. USA* 77: 7347 - 7351
- Douglas C. J., Hauffe K. D., Ites-Morales M-E., Ellard M., Paszkowski U., Hahlbrock K. & Dangl J. L. (1991) Exonic sequences are required for elicitor and light activation of a plant defense gene, but promoter sequences are sufficient for tissue-specific expression. *EMBO J.* 10: 1767 - 1775
- Draper J., Scott R., Armitage P. & Walden R. (1988) Plant genetic transformation and gene expression: A laboratory manual. Blackwell Scientific Publications.
- Drapkin R., Merino A. & Reinberg D. (1993) Regulation of RNA polymerase II transcription. *Curr. Opin. Cell Biol.* 5: 469 - 476
- Eden S. & Cedar H. (1994) Role of DNA methylation in the regulation of transcription. *Curr. Opin. Gen. Dev.* 4: 255 - 259

- Ellenberger T. (1994) Getting a grip on DNA recognition: structures of the basic region leucine zipper and the basic region helix-loop-helix DNA-binding domains. *Curr. Opin. Struct. Biol.* 4: 12 - 21
- Ettinger . F. & Harada J. J. (1990) Translational or posttranslational processes affect differentially the accumulation of isocitrate lyase and malate synthase proteins and enzyme activities in embryos and seedlings of *Brassica napus*. *Arch. Biochem. Biophys.* 281: 139 - 143
- Fang T. K., Donaldson R. P. & Vigil E. L. (1987) Electron transport in purified membranes from castor bean endosperm. *Planta* 172: 1 - 13
- Feinberg A. P. & Vogelstein B. (1983) A technique for radio-labelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132: 6 - 13
- Fernandez E., Moreno F. & Rodicio R. (1992) The *ICL1* gene from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 204: 983 - 990
- Finnessy J. J., Trelease R. N. & Randall D. D. (1994) Glyoxysomal isocitrate lyase is a phosphoprotein. *Plant Physiol. (Suppl.)* 105: 92
- Fromm M., Taylor L. P. & Walbot V. (1985) Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proc. Natl. Acad. Sci. USA* 82: 5824 - 5828
- Gainey L. D. S., Connerton I. F., Lewis E. H., Turner G. & Ballance D. J. (1992) Characterisation of the glyoxysomal isocitrate lyase genes of *Aspergillus nidulans* (*acuD*) and *Neurospora crassa* (*acu-3*) *Curr. Genet.* 21: 43 - 47
- Galas D. & Schmitz A. (1978) DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucl. Acid Res.* 5: 3157 - 3170
- Gancedo J. M. (1992) Carbon catabolite repression in yeast. *Eur. J. Biochem.* 206: 297 - 313
- Garner M. M. & Revzin A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucl. Acids Res.* 9: 3347 - 3360
- Gasch A., Hoffmann A., Horikoshi M., Roeder R. G. & Chua N-H. (1990) *Arabidopsis thaliana* contains two genes for TFIID. *Nature* 346: 390 - 394
- Gerhardt B. (1983) Localisation of β -oxidation enzymes in peroxisomes isolated from non-fatty plant tissue. *Planta* 159: 238 - 246
- Gietl C. & Hock B. (1984) Import of in-vitro-synthesised glyoxysomal malate dehydrogenase into isolated watermelon glyoxysomes. *Planta* 162: 261 - 267
- Gietl C. (1990) Glyoxysomal malate dehydrogenase from watermelon is synthesised with an amino terminal transit peptide. *Proc. Natl. Acad. Sci. USA* 87: 5773 - 5777
- Gill G., Pascal E., Tseng Z. H. & Tjian R. (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAF_{II}110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. *Proc. Natl. Acad. Sci. USA* 91: 192 - 196

- Gould S. J., Keller G-A. & Subramani S. (1987) Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *J. Cell. Biol.* 105: 2923 - 2931
- Gould S. J., Keller G-A. & Subramani S. (1988) Identification of peroxisomal targeting signals located at the carboxy terminus of four peroxisomal proteins. *J. Cell. Biol.* 107: 897 - 905
- Gould S. J., Keller G-A., Hosken N., Wilkinson J. & Subramani S. (1989) A conserved tripeptide sorts proteins to peroxisomes. *J. Cell. Biol.* 108: 1657 - 1664
- Gould S. J., Keller G-A., Schneider M., Howell S. H., Garrard L. J., Goodman J. M., Distel B., Tabak H. & Subramani S. (1990) Peroxisomal protein import is conserved between yeast, plants, insects and mammals. *EMBO J.* 9: 85 - 90
- Graham I. A., Smith L. M., Brown J. W. S., Leaver C. J. & Smith S. M. (1989) The malate synthase gene of cucumber. *Plant Mol. Biol.* 13: 673 - 684
- Graham I. A., Smith L. M., Leaver C. J. & Smith S. M. (1990) Developmental regulation of expression of the malate synthase gene in transgenic plants. *Plant Mol. Biol.* 15: 539 - 549
- Graham I. A., Leaver C. J. & Smith S. M. (1992) Induction of malate synthase gene expression in senescent and detached organs of cucumber. *Plant Cell* 4: 349 - 357
- Graham I. A., Baker C. & Leaver C. J. (1994a) Analysis of the cucumber malate synthase gene promoter by transient expression and gel retardation assays. *Plant J.* In Press.
- Graham I. A., Denby K. J. & Leaver C. J. (1994b) Carbon catabolite repression regulates glyoxylate cycle gene expression in cucumber. *Plant Cell* 6: 761 - 772
- Graves D. J., Martin B. L. & Wang J. H. (1994) Co- and post-translational modification of proteins. Chemical principals and biological effects. Oxford University Press.
- Greenler J., Sloan J. S., Schwartz B. W. & Becker W. M. (1989) Isolation, characterisation and sequence analysis of a full-length cDNA clone encoding NADH-dependent hydroxypyruvate reductase from cucumber. *Plant Mol. Biol.* 13: 139 - 150
- Grierson C., Du J-S., de Tórees Zabala M., Beggs K., Smith C., Holdsworth M. & Bevan M. (1994) Separate *cis* sequences and *trans* factors direct metabolic and developmental regulation of a potato tuber storage protein gene. *Plant J.* 5: 815 - 826
- Groschedl R., Giese K. & Pagel J. (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *TIG* 10: 94 - 100
- Grunder A. A., Loverre-Chyivlia A. & Hickey D. A. (1993) Expression of an amylase-alcohol dehydrogenase chimeric gene in transgenic strains of *Drosophila melanogaster*. *Genome* 36: 954 - 961
- Grunstein M. & Hogness D. S. (1975) Colony hybridisation: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* 72: 3961 - 3965

- Gut H. & Matile P. (1988) Apparent induction of key enzymes of the glyoxylic acid cycle in senescent barley leaves. *Planta* 176: 548 - 550
- Hanks J. F., Tolbert N. E. & Schubert K. R. (1981) Localisation of enzymes of ureide biosynthesis in peroxisomes and microsomes of nodules. *Plant Physiol.* 68: 65 - 69
- Hanson U., Tenen D. G., Livingston D. M. & Sharp P. A. (1981) T antigen repression of SV40 early transcription from two promoters. *Cell* 27: 603 - 612
- Harrison S. C. (1991) A structural taxonomy of DNA-binding domains. *Nature* 353: 715 - 719
- Harrum M. A. & Schwartzbach S. D. (1981) Nutritional regulation of organelle biogenesis in *Euglena*. *Plant Physiol.* 68: 430 - 434
- Hattori T., Nakagawa S. & Nakamura K. (1990) High-level expression of tuberous root storage protein genes of sweet potato in stems of plantlets grown *in vitro* on sucrose medium. *Plant Mol. Biol.* 14: 595 - 604
- Hawley D. K., Johnson A. D. & McClure W. R. (1985) Functional and physical characterisation of transcription initiation complexes in the bacteriophage λ O_R region. *J. Biol. Chem.* 260: 8618 - 8626
- Heidecker G. & Messing J. (1986) Structural analysis of plant genes. *Ann. Rev. Plant Physiol.* 37: 439 - 466
- Hendrickson W. & Schlieff R. (1985) A dimer of AraC protein contacts three adjacent major groove regions at the Ara I DNA site. *Proc. Natl. Acad. Sci. USA* 82: 3129 - 3133
- Hershko A. & Ciechanover A. (1992) The ubiquitin system for protein degradation. *Ann. Rev. Biochem.* 61: 761 - 807
- Herskowitz I., Rine J. & Strathern J. (1992) in *The molecular and cellular biology of the yeast Saccharomyces: gene expression Vol II.* (J. R. Broach, E. W. Jones, J. R. Pringle, Eds.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Hoekema A., Hirsch P. R., Hooykaas P. J. J. & Schilperoort R. A. (1983) A binary plant vector strategy based on separation of *vir* and T-region of the *A. tumefaciens* Ti-plasmid. *Nature* 303: 179 - 180
- Hondred D., Wadle D. M., Titus D. E. & Becker W. M. (1987) Light-stimulated accumulation of the peroxisomal enzymes hydroxypyruvate reductase and serine:glyoxylate aminotransferase and their translatable mRNAs in cotyledons of cucumber seedlings. *Plant Mol. Biol.* 9: 259 - 275
- Hoyt J. C. & Reeves H. C. (1988) *In vivo* phosphorylation of isocitrate lyase from *Escherichia coli* D₅H₃G₇. *Biochem. Biophys. Res. Comm.* 153: 875 - 880
- Huang A. H. C., Trelease R. N. & Moore T. S. Jr. (1983) *Plant peroxisomes.* Academic Press. New York
- Hunt A. G., Chu N. M., Odell J. F., Nagy F & Chua N-H. (1987) Plant cells do not properly recognise animal gene polyadenylation signals. *Plant Mol. Biol.* 8: 23 - 35

- Igual J. C., Gonzalez-Bosch C., Dopazo J. & Perez-Ortin J. E. (1992) Phylogenetic analysis of the thiolase family. Implications for the evolutionary origin of peroxisomes. *J. Mol. Evol.* 35: 147 - 155
- Jefferson R. A., Burgess S. M. & Hirsh D. (1986) β -glucuronidase from *Escherichia coli* as a gene fusion marker. *Proc. Natl. Acad. Sci. USA* 83: 8447 - 8451
- Jefferson R. A. (1987) Assaying chimeric genes in plants: The GUS gene fusion system. *Plant Mol. Biol. Rep.* 5: 387 - 405
- Jefferson R. A., Kavanagh T. A. & Bevan M. W. (1987) GUS-fusions: β -glucuronidase as a sensitive and versatile gene marker in higher plants. *EMBO J.* 6: 3901 - 3907
- Johnson F. B. & Krasnow M. A. (1992) Differential regulation of transcription preinitiation complex assembly by activator and repressor homeodomain proteins. *Genes Dev.* 6: 2177 - 2189
- Joshi C. P. (1987a) An inspection of the domain between the putative TATA box and translation start site in 79 plant genes. *Nucl. Acids Res.* 15: 6643 - 6653
- Joshi C. P. (1987b) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucl. Acids Res.* 15: 9627 - 9641
- Kakkis E., Riggs K. J., Gillespie W. & Calame K. (1989) A transcriptional repressor of *c-myc*. *Nature* 339: 718 - 721
- Keller G-A., Krisans S., Gould S. J., Sommer J. M., Wang C. C., Schliebs W., Kunau W., Brody S. & Subramani S. (1991) Evolutionary conservation of a microbody targeting signal that targets proteins to peroxisomes, glyoxysomes and glycosomes. *J. Cell Biol.* 114: 893 - 904
- Kelly J. M. & Hynes M. J. (1977) Increased and decreased sensitivity to carbon catabolite repression of enzymes of acetate metabolism in mutants of *Aspergillus nidulans*. *Mol. Gen. Genet.* 156: 87 - 92
- Kim D-J. & Smith S. M. (1994) Molecular cloning of cucumber phosphoenolpyruvate carboxykinase and developmental regulation of gene expression. *Plant Mol. Biol.* In Press
- Kim J. L., Nikolov D. B. & Burley S. K. (1993) Co-crystal structure of TBP recognising the minor groove of a TATA element. *Nature* 365: 520 - 527
- Kim Y., Geiger J. H., Hahn S. & Sigler P. B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365: 512 - 520
- Kindl H. (1993) Fatty acid degradation in plant peroxisomes: Function and biosynthesis of the enzymes involved. *Biochimie* 75: 225 - 230
- Ko Y. H. & McFadden B. A. (1990) Alkylation of isocitrate lyase from *Escherichia coli* by 3-bromopyruvate. *Arch. Biochem. Biophys.* 278: 373 - 380
- Ko Y. H., Cremo C. R. & McFadden B. A. (1992) Vanadate-dependent photomodification of serine 319 and 321 in the active site of isocitrate lyase from *Escherichia coli*. *J. Biol. Chem.* 267: 91 - 95

- Koch K. E., Nolte K. D., Duke E. R., McCarty D. R. & Avigne W. T. (1992) Sugar levels modulate differential expression of maize sucrose synthase genes. *Plant Cell* 4: 59 - 69
- Koller W. & Kindl H. (1977) Glyoxylate cycle enzymes of the glyoxysomal membranes from cucumber cotyledons. *Arch. Biochem. Biophys.* 181: 236 - 248
- Kornberg H. L. & Beevers H. (1957) The glyoxylate cycle as a stage in the conversion of fat to carbohydrate in castor beans. *Biochim. Biophys. Acta* 26: 531 - 537
- Kornberg H. L. (1966) The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem J.* 99: 1 - 11
- Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44: 283 - 292
- Kozak M. (1989) The scanning model for translation: an update. *J. Cell Biol.* 108: 229 - 241
- Kozak M. (1992) Regulation of translation in eukaryotic systems. *Ann. Rev. Cell Biol.* 8: 197 - 225
- Krapp A., Hofmann B., Schäfer C. & Stitt M. (1993) Regulation of the expression of *rbcS* and other photosynthetic genes by carbohydrates: A mechanism for the 'sink regulation' of photosynthesis? *Plant J.* 3: 817 - 828
- Kudielka R. & Theimer R. (1983a) Derepression of glyoxylate cycle enzyme activities in Anise suspension cultures. *Plant Sci. Letts.* 31: 237 - 244
- Kudielka R. & Theimer R. (1983b) Repression of glyoxysomal enzyme activities in anise (*Pimpinella anisum* L.) suspension cultures. *Plant Sci. Letts.* 31: 245 - 252
- Kunau W. H., Buehne S., de la Garza M., Kionka C., Mateblowski M., Schultz-Borchard U. & Thieringer R. (1988) Comparative enzymology of beta oxidation. *Biochem Soc. Trans.* 16: 418 - 420
- Kyozuka J., Olive M., Peacock W. J., Dennis E. S. & Shimamoto K. (1994) Promoter elements required for developmental expression of the maize *Adh1* gene in transgenic rice. *Plant Cell* 6: 799 - 810
- Lado P., Schwendimann M. & Marre E. (1968) Repression of isocitrate lyase synthesis in seeds germinated in the presence of glucose. *Biochim. Biophys. Acta* 157: 140 - 148
- Lamb J. E., Reizman H., Becker W. M. & Leaver C. J. (1978) Regulation of glyoxysomal enzymes during germination of cucumber. 2. Isolation and immunological detection of isocitrate lyase and catalase. *Plant Physiol.* 62: 754 - 760
- Lazarow P. B. & Fujiki Y. (1985) Biogenesis of peroxisomes. *Ann. Rev. Cell Biol.* 1: 489 - 530
- Lee J. & Goldfarb A. (1991) *lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. *Cell* 66: 793 - 798

- Link E., Kerr L. D., Schreck R., Zabel U., Verma I. & Baeuerle P. (1992) Purified I κ B-6 is inactivated upon dephosphorylation. *J. Biol. Chem.* 267: 239 - 246
- Longo C. P. & Longo G. P. (1970) The development of glyoxysomes in peanut cotyledons and maize scutella. *Plant Physiol.* 45: 249 - 254
- Lopez-Boado Y. S., Herrero P., Fernandez T., Fernandez R. & Moreno F. (1988) Glucose-stimulated phosphorylation of yeast isocitrate lyase *in vivo*. *J. Gen. Microbiol.* 134: 2499 - 2505
- Lu H., Zawel L., Fisher L., Egly J. M. & Reinberg D. (1992) Human general transcription factor IIIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358: 641 - 645
- Lyon M. (1993) Epigenetic inheritance in mammals. *TIG* 9: 123 - 128
- Ma J. & Ptashne M. (1987) The carboxy-terminal 30 amino acids of Gal4 are recognized by Gal80. *Cell* 50: 137 - 142
- Ma H., Bloom L. M., Walsh C. T. & Botstein D. (1989b) The residual enzymatic phosphorylation activity of hexokinase II mutants is correlated with glucose repression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9: 5643 - 5649
- Ma H., Bloom L. M., Zhu Z., Walsh C. T. & Botstein D. (1989a) Characterisation of mutations in the *HXK2* gene of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9: 5630 - 5642
- Maass M. M. & Feix G. (1992) Two different cDNAs encoding TFIID proteins of maize. *FEBS letts.* 301: 294 - 298
- Maeshima M., Yokoi H. & Asahi T. (1988) Evidence for no proteolytic processing during transport of isocitrate lyase into glyoxysomes. *Plant Cell Physiol.* 29: 381 - 384
- Maloy & Nunn (1982) Genetic regulation of the glyoxylate shunt in *E. coli* K12. *J. Bacteriol.* 149: 173 - 180
- Martin C., Beeching J. R. & Northcote D. H. (1984) Changes in levels of transcripts in endosperms of castor beans treated with exogenous gibberellic acid. *Planta* 162: 68 - 76
- Martin K. J. (1991) The interactions of transcription factors and their adaptors, coactivators and accessory proteins. *BioEssays* 13: 499 - 503
- Masterton C., Wood C. & Thomas D. R. (1990) β -oxidation enzymes in the mitochondria of *Arum* and oilseed rape. *Planta* 182: 129 - 135
- Matsuoka M. & McFadden B. A. (1988) Isolation, hyperexpression and sequencing of the *aceA* gene encoding isocitrate lyase in *Escherichia coli*. *J. Bacteriol.* 170: 4528 - 4536
- Matzke M. & Matzke A. J. M. (1993) Genomic imprinting in plants: Parental effects and *trans*-inactivation phenomena. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 44: 53 - 76

- McCullum D., Monosov E. & Subramani S. (1993) The *pas8* mutant of *Pichia pastoris* exhibits the peroxisomal protein import deficiencies of Zellweger syndrome cells - the PAS8 protein binds to the COOH-terminal tripeptide peroxisomal targeting signal, and is a member of the TPR protein family. *J. Cell Biol.* 121: 761 - 774
- McCullough W. & John P. C. L. (1972) Control of *de novo* isocitrate lyase synthesis in *Chlorella*. *Nature* 239: 402 - 405
- McKeown M. (1992) Alternative mRNA splicing. *Ann. Rev. Cell Biol.* 8: 133 - 155
- McLaughlin J. C. & Smith S. M. (1994) Metabolic regulation of glyoxylate-cycle enzyme synthesis in detached cucumber cotyledons and protoplasts. *Planta In Press*
- Meehan R. R., Lewis J. D., McKay S., Kleiner E. L. & Bird A. P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58: 499 - 507
- Mettler I. J. & Beevers H. (1980) Oxidation of NADH in glyoxysomes by a malate-aspartate shuttle. *Plant Physiol.* 66: 555 - 560
- Meyer M-E., Grinemeyer H., Turcoite B., Bocquel M-T., Tasset D. & Chambon P. (1989) Steroid hormone receptors compete for factors that mediate their enhancer function. *Cell* 57: 433 - 442
- Miernyk J. A. & Trelease R. N. (1981) Control of enzyme activities in cotton cotyledons during maturation and germination. *Plant Physiol.* 67: 341 - 346
- Migeon B. (1994) X-chromosome inactivation: molecular mechanisms and consequences. *TIG* 10: 230 - 235
- Miller A. M. & Nasmyth K. A. (1984) Role of DNA replication in the repression of the silent mating type loci in yeast. *Nature* 312: 247 - 251
- Miller J., McLachlan A. D. & Klug A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* 4: 1609 - 1614
- Miura S., Kasuya-Arai I., Mori H., Miyazawa S., Osumi T., Hashimoto T. & Fujiki Y. (1992) Carboxyl-terminal consensus Ser-Lys-Leu related tripeptide of peroxisomal proteins functions *in vitro* as a minimal peroxisomal-targeting signal. *J. Biol. Chem.* 267: 14405 - 14411
- Miyazawa S., Osumi T., Hashimoto T., Ohno K., Miura S. & Fujiki Y. (1989) Peroxisome targeting signal of rat liver acyl-CoA oxidase resides at the carboxy terminus. *Mol. Cell. Biol.* 9: 83 - 91
- Mogen B. D., MacDonald M. H., Graybosch R. & Hunt A. G. (1990) Upstream sequences other than AAUAAA are required for efficient mRNA 3'-end formation in plants. *Plant Cell* 2: 1261 - 1272
- Mogen B. D., MacDonald M. H., Leggewie G. & Hunt A. G. (1992) Several distinct types of sequence elements are required for efficient mRNA 3'-end formation in a pea *rbcS* gene. *Mol. Cell. Biol.* 12: 5406 - 5414
- Mori H. & Nishimura M. (1989) Glyoxysomal malate synthetase is specifically degraded in microbodies during greening of pumpkin cotyledons. *FEBS Letts.* 244: 163 - 166

- Murashige T. & Skoog P. (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol. Plant.* 15: 473 - 497
- Murray N. (1983) Lambda II. eds. Hendrix R.W., Roberts J.W., Stahl F.W. & Weisberg R.A. Cold Spring Harbor Laboratory.
- Nehlin J. O., Carlberg M. & Ronne H. (1991) Control of yeast *GAL* genes by *MIG1* repressor: a transcriptional cascade in the glucose response. *EMBO J.* 10: 3373 - 3377
- Nguyen T., Zelechowska M., Foster V., Bergmann H. & Verma D. P. S. (1985) Primary structure of the soybean nodulin-35 gene encoding uricase II localised in the peroxisomes of uninfected cells of nodules. *Proc. Natl. Acad. Sci. USA* 82: 5040 - 5044
- Ni W. & Trelease R. N. (1991) Post-transcriptional regulation of catalase isozyme expression in cotton seeds. *Plant Cell* 3: 737 - 744
- Nicholl D. S. T. & John P. C. L. (1986) Estimation of the Mr of isocitrate lyase messenger RNA from *Chlorella fusca*. *J. Gen. Microbiol.* 132: 1423 - 1427
- Nickelson J., van Dillewijn J., Renhire M. & Rochaix J-D. (1994) Determinants for stability of the chloroplast *psbD* RNA are located within its short leader region in *Chlamydomonas reinhardtii*. *EMBO J.* 13: 3182 - 3191
- Nikolov D., Hu S-H., Lin J., Gasch A., Hoffman A., Horikoshi M., Chua N-H., Roeder R. & Burley S. (1992) Crystal structure of TFIID TATA box binding protein. *Nature* 360: 40 - 46
- Nimmo H. G., Douglas F., Kleanthous C., Campbell D. G. & MacKintosh C. (1989) Identification of a cysteine residue at the active site of *Escherichia coli* isocitrate lyase. *Biochem. J.* 261: 431 - 435
- Nishimura M., Yamaguchi J., Mori H., Akazawa T. & Yokota S. (1986) Immunocytochemical analysis shows that glyoxysomes are directly transformed to leaf peroxisomes during greening of pumpkin cotyledons. *Plant Physiol.* 81: 313 - 316
- Nishimura M., Takeuchi Y., De Bellis L. & Hara-Nishimura I. (1993) Leaf peroxisomes are directly transformed to glyoxysomes during senescence of pumpkin cotyledons. *Protoplasma* 175: 131 - 137
- Nolan G.P., Ghosh S., Liou H-C., Tempst P. & Baltimore D. (1991) DNA binding and I κ B inhibition of the cloned p65 subunit of NF κ B, a Rel-related polypeptide. *Cell* 64: 961 - 969
- Nunn W. D. (1987) *Escherichia coli* and *Salmonella typhiurium*: Cellular and molecular biology. In Ingraham J. L., Low B., Magasanik B., Schaechter M. & Umberger H. E. (eds). American Society of Microbiology. Washington DC. vol I pp 285 - 301
- Ogren W. L. (1984) Photorespiration: pathways, regulation, and modification. *Ann. Rev. Plant Physiol.* 35: 415 - 442
- Olsen L. J., Ettinger W. F., Damsz B., Matsudaira K., Webb M. A. & Harada J. J. (1993) Targeting of glyoxysomal proteins to peroxisomes in leaves and roots of a higher plant. *Plant Cell* 5: 941 - 952

- Pacy J. & Thurston C. F. (1987) Distribution of isocitrate lyase in induced and de-adapting cells of the green alga *Chlorella fusca*. J. Gen. Microbiol. 133: 3341 - 3346
- Parish J. H. & Kirby K. S. (1966) Reagents which reduce interactions between ribosomal RNA and rapidly labelled RNA from rat liver. Biochim. Biophys. Acta 129: 554-562
- Parvin D. P. & Sharp P. A. (1993) DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. Cell 73: 533 - 540
- Pistelli L., De Bellis L. & Alpi A. (1991) Peroxisomal enzyme activities in attached senescing leaves. Planta 184: 151 - 153
- Potter H. Weir L. & Leder P. (1984) Enhancer-dependent expression of human α immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. Proc. Natl. Acad. Sci. USA 81: 7161 - 7165
- Preisig-Muller R. & Kindl H. (1993) Thiolase mRNA translate *in vitro* yields a peptide with a putative N-terminal presequence. Plant Mol. Biol. 22: 59 - 66
- Ptashne M. (1988) How eukaryotic transcriptional activators work. Nature 335: 683 - 689
- Ptashne M. (1992) A genetic switch. Cell press & Blackwell Scientific Publications
- Rawsthorne S., Minchin F. R., Summerfield R. J., Cookson C. & Coombs J. (1980) Carbon and nitrogen metabolism in legume root nodules. Phytochemistry 19: 341 - 355
- Reizman H., Weir E. M., Leaver C. J., Titus D. E. & Becker W. M. (1980) Regulation of glyoxysomal enzymes during germination of cucumber. *In vitro* translation and characterisation of four glyoxysomal enzymes. Plant Physiol. 65: 40 - 46
- Rivier D. H. & Pillus L. (1994) Silencing speaks up. Cell 76: 963 - 966
- Robertson E. F. & Reeves H. C. (1989) Phosphorylation of isocitrate lyase in *Escherichia coli*. Biochimie 71: 1065 - 1070
- Rodriguez D., Dommes J. & Northcote D. H. (1987) Effect of abscisic and gibberellic acids on malate synthase transcripts in germinating castor bean seeds. Plant Mol. Biol. 9: 227 - 235
- Roth S. Y., Dean A. & Simpson R. T. (1990) Yeast $\alpha 2$ repressor positions nucleosomes in TRP1/ARS1 chromatin. Mol. Cell. Biol. 10: 2247 - 2260
- Rothnie H. M., Reid J. & Hohn T. (1994) The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3'-end formation in plants. EMBO J. 13: 2200 - 2210
- Salanoubat M. & Belliard G. (1989) The steady-state level of potato sucrose synthase mRNA is dependent on wounding, anaerobiosis, and sucrose concentration. Gene 84: 181 - 185

- Sandeman R. A., Hynes M. J., Fincham J. R. S. & Connerton I. F. (1991) Molecular organisation of the malate synthase genes of *Aspergillus nidulans* and *Neurospora crassa*. *Mol. Gen. Genet.* 228: 445 - 452
- Sanfaçon H. & Hohn T. (1990) Proximity to the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. *Nature* 346: 81 - 84
- Sanfaçon H., Brodmann P. & Hohn T. (1991) A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev.* 5: 141 - 149
- Sanfaçon H. (1994) Analysis of figwort mosaic virus (plant pararetrovirus) polyadenylation signal. *Virology* 198: 39 - 49
- Sanger F., Nicklen S. & Coulson J. R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. (USA)* 74: 5463 - 5467
- Saunders J. A., Matthew B. F. & Miller P. D. (1989a) Plant gene transfer using electrofusion and electroporation. p343 - 354 *in* *Electroporation and electrofusion in cell biology.* (E. Neumann, A. Sowers & C. Jordan Eds.) Plenum Press, New York.
- Saunders J. A., Smith C. R. & Kaper J. M. (1989b) Effects of electroporation pulse wave on the incorporation of viral RNA into tobacco protoplasts. *BioTechniques* 7: 1124 - 1131
- Sautter C. (1986) Microbody transition in greening watermelon cotyledons: double immunocytochemical labeling of isocitrate lyase and hydroxypyruvate reductase. *Planta* 167: 491 - 503
- Scandalios J.G. (1990) Targeting, import and processing of nuclear gene-encoded proteins into mitochondria and peroxisomes. *Isozymes: structure, function and use in biology and medicine.* p515 - 544 Wiley-Liss, Inc.
- Schaeffer L., Roy R., Humbert S., Moncollin V., Vermeulen W., Hoeijmakers J. H. J., Chambon P. & Egly J. M. (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* 260: 58 - 63
- Schaller A., Schmid J., Leibinger U. & Amrhein N. (1991) Molecular cloning and analysis of a cDNA coding for chorismate synthase from the higher plant *Corydalis sempervirens* Pers. *J. Biol. Chem.* 266: 21434 - 21438
- Schmiedeskamp M. & Klevit R. (1994) Zinc finger diversity. *Curr. Opin. Struct. Biol.* 4: 28 - 35
- Schöler A. & Schüller H. J. (1993) Structure and regulation of the isocitrate lyase gene *ICL1* from the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* 23: 375 - 381
- Schöler A. & Schüller H. J. (1994) A carbon source-responsive element necessary for activation of the isocitrate lyase gene *ICL1* is common to genes of the gluconeogenic pathway in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 14: 3613 - 3622
- Schwarz-Sommer Z., Huijser P., Nacken W., Saedler H. & Sommer H. (1990) Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250: 931 - 936

- Schwarz-Sommer Z., Hue I., Huijser P., Flor P. J., Hansen R., Teten F., Lönnig W-E., Saedler H. & Sommer H. (1992) Characterisation of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* 11:251 - 263
- Sheen J. (1990) Metabolic repression of transcription in higher plants. *Plant Cell* 2: 1027 - 1038
- Sjogren R. E. & Romano A. H. (1967) Evidence for multiple forms of isocitrate lyase in *Neurospora crassa*. *J. Bacteriol.* 93: 1638 - 1643
- Small G. M., Szabo L. J. & Lazarow P. B. (1988) Acyl-CoA oxidase contains two targeting sequences each of which can mediate protein import into peroxisomes. *EMBO J.* 7: 1167 - 1173
- Small S., Kraut R., Hoey T., Warrior R. & Levine M. (1991) Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* 5: 827 - 839
- Small S., Blair A. & Levine M. (1992) Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J.* 11: 4047 - 4057
- Smith S. M. & Leaver C. J. (1986) Glyoxysomal malate synthase of cucumber: molecular cloning of a cDNA and regulation of enzymes synthesis during germination. *Plant Physiol.* 81: 762 - 767
- Sommer J. M., Cheng Q. L., Keller G. A. & Wang C. C. (1992) *In vivo* import of firefly luciferase into the glycosomes of *T. brucei* and mutational analysis of the C-terminal targeting signal. *Molecular Biology of the Cell* 3: 749 - 759
- South T. L. & Summers M. F. (1993) Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the psi-site analog, dACGCC. *Protein Sci.* 2: 3 - 19
- Southern E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503 - 17
- Stanway C. (1993) Simplicity amidst complexity in transcriptional initiation. *BioEssays* 15: 559 - 60
- Subramani S. (1993) Protein import into peroxisomes and biogenesis of the organelle. *Ann. Rev. Cell. Biol.* 9: 445 - 478
- Sullivan M. J. & Green P. J. (1993) Post-transcriptional regulation of nuclear encoded genes in higher plants: the roles of mRNA stability and translation *Plant Mol. Biol.* 23: 1091 - 1104
- Swinkels B. W., Gould S. J., Bodnar A. G., Rachubinski R. A. & Subramani S. (1991) A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-ketoacyl-CoA thiolase. *EMBO J.* 10: 3255 - 3262
- Thomas D. R. & Wood C. (1986) The two β -oxidation sites in pea cotyledons. *Planta* 168: 261 - 266
- Thurston C. F. (1977) Control of isocitrate lyase synthesis in *Chlorella fusca* var. *vacuolata*. *Biochem J.* 164: 147 - 151

- Titus D. E. & Becker W. M. (1985) Investigation of the glyoxysome-peroxisome transition in germinating cucumber cotyledons using double label immunoelectron microscopy. *J. Cell Biol.* 101: 1288 - 1299
- Tjian R. & Maniatis T. (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77: 5 - 8
- Tolbert N. E. (1981) Metabolic pathways in peroxisomes and glyoxysomes. *Ann. Rev. Biochem.* 50: 133 - 157
- Tröbner W., Ramirez L., Motte P., Hue I., Huijser P., Löhnig W-E., Saedler H., Sommer H. & Schwarz-Sommer Z. (1992) *Globusa*: a homeotic gene which interacts with *deficiens* in the control of *Antirrhinum* floral organogenesis. *EMBO J.* 11: 4693 - 4704
- Trulson A. J., Simpson R. B. & Shahin E. A. (1986) Transformation of cucumber (*Cucumis sativus* L.) plants with *Agrobacterium rhizogenes*. *Theor. Appl. Genet.* 73: 11 - 15
- Tsukaya H., Ohshima T., Naito S., Chino M. & Komeda Y. (1991) Sugar-dependent expression of the *CHS-A* gene for chalcon synthase from petunia in transgenic *Arabidopsis*. *Plant Physiol.* 97: 1414 - 1421
- Turley R. B., Choe S. M. & Trelease R. N. (1990) Characterisation of a cDNA clone encoding the complete amino acid sequence of cotton isocitrate lyase. *Biochim. Biophys. Acta* 1049: 223 - 226
- Turley R. B. & Trelease R. N. (1990) Developments and regulation of three glyoxysomal enzymes during cotton seed maturation and growth. *Plant Mol. Biol.* 14: 137 - 146
- Ullman A., Jacob F. & Monod J. (1967) Characterisation by *in vitro* complementation of a peptide corresponding to an operator-proximal segment of the β -galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.* 24: 339 - 343
- Van den Bosch H., Schutgens R. B. H., Wanders R. J. A. & Tager J. M. (1992) Biochemistry of peroxisomes. *Ann. Rev. Biochem.* 61: 157 - 197
- Van der Klei I. J., Faber K. N., Keizer-Gunnick I., Gietl C., Harder W. & Veenhuis M. (1993) Watermelon glyoxysomal malate dehydrogenase is sorted to peroxisomes of the methylotrophic yeast, *Hansenula polymorpha*. *FEBS Letts.* 334: 128 - 132
- Van Doran M., Ellis H. M. & Posakony J. W. (1991) The *Drosophila extramachrochaetae* protein antagonises sequence-specific DNA binding by *daughterless/achaete-scute* protein complexes. *Development* 113: 245 - 255
- Veenhuis M., Keizer I. & Harder W. (1979) Characterisation of peroxisomes in glucose-grown *Hansenula polymorpha* and their levels after transfer of cells to methanol-containing media. *Arch. Microbiol.* 120: 167 - 175
- Veenhuis M. & Goodman J. M. (1990) Peroxisomal assembly: membrane proliferation precedes the induction of abundant matrix proteins in the methylotrophic yeast *Candida boidinii*. *J. Cell Sci.* 96: 583 - 590

- Verner K. & Schatz G. (1988) Protein translocation across membranes. *Science* 241: 1307 - 1313
- Vershon A. K., Liao S-M., McClure W. R. & Sauer R. T. (1987) Interaction of the bacteriophage P22 Arc repressor with operator DNA. *J. Mol. Biol.* 195: 323 - 331
- Vicentini F. & Matile P. (1993) Gerontosomes, a multifunctional type of peroxisome in senescent leaves. *J. Plant Physiol.* 142: 50 - 56
- Vierstra R. D. (1993) Protein degradation in plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 44: 385 - 410
- Volokita M. (1991) The carboxy-terminal end of glycolate oxidase directs a foreign protein into tobacco leaf peroxisomes. *The Plant J.* 1: 361 - 366
- Wahle E. & Keller W. (1992) The biochemistry of 3'-end cleavage and polyadenylation of mRNA precursors. *Ann. Rev. Biochem.* 61: 419 - 440
- Walton P. A., Gould S. J., Feramisco J. R. & Subramani S. (1992) Transport of microinjected proteins into peroxisomes of mammalian cells: inability of Zellweger cell lines to import proteins with the SKL tripeptide peroxisomal targeting signal. *Mol. Cell. Biol.* 12: 531 - 541
- Weir E. M., Reizman H., Grienenberger J-M., Becker W. M. & Leaver C. J. (1980) Regulation of glyoxysomal enzymes during germination of cucumber. Temporal changes in translatable mRNAs for isocitrate lyase and malate synthase. *Eur. J. Biochem.* 112: 469 - 477
- Wenzler H. C., Mignery G. A., Fisher L. M. & Park W. D. (1989) Analysis of a chimeric class I patatin-GUS gene in transgenic potato plants: High level expression in tubers and sucrose-inducible expression in cultured leaf and stem explants. *Plant Mol. Biol.* 12: 41 - 50
- Willmitzer L. (1988) The use of transgenic plants to study plant gene expression. *TIG* 4: 13 - 18
- Witt I., Kronau R. & Holzer H. (1966) Repression von alkoholdehydrogenase, malatdehydrogenase, isocitratlyase und malatsynthase in hefe durch glucose. *Biochim. Biophys. Acta* 118: 522 - 537
- Wong T. K. & Neumann E. (1982) Electric field mediated gene transfer. *Biochem. Biophys. Res. Comm.* 107: 584 - 587
- Woodcock E. & Merret M. J. (1980) Malate synthase messenger RNA in relation to enzyme derepression in *Euglena gracilis*. *Arch. Microbiol.* 124: 33 - 38
- Wright P. E. (1994) POU domains and homeodomains. *Curr. Opin. Struct. Biol.* 4: 22 - 27
- Zhang D., Ehrlich K. C., Supakar P. C. & Ehrlich M. (1989) A plant DNA-binding protein that recognises 5-methyl cytosine residues. *Mol. Cell. Biol.* 1351 - 1356
- Zhang J. Z., Gomez-Pedrozo M., Baden C. S. & Harada J. J. (1993) Two classes of ICL genes are expressed during late embryogeny and postgermination in *Brassica napus* L. *Mol. Gen. Genet.* 238: 177 - 184

Zhang J. Z., Laudencia-Chingcuanco D. L., Comai L., Li M. & Harada J. J. (1994)
Isocitrate lyase and malate synthase genes from *Brassica napus* L. are active in pollen.
Plant Physiol. 104: 857 - 864