A Molecular Characterization of Phenylpropanoid Pathway Genes of *Trifolium subterraneum*

A thesis submitted for the degree of
Doctor of Philosophy
of
The Australian National University

by

Paul Arthur Howles

April 1992

Declaration

The research described in this thesis is my own work, except where acknowledgement is made, and has not been submitted for any other degree.

Paul Howles

General acknowledgements

General abbreviations

ATP - adenosine triphosphate
bp - base pair(s)
CHS - chalcone synthase
4CL - 4-coumarate:CoA ligase
CoA - coenzyme A
^{D}C - degrees Celsius
dATP - deoxyadenosine 5' triphosphate
dCTP - deoxycytidine 5' trisphosphate
dGTP - deoxyguanosine 5' triphosphate
dTTP - deoxythymidine 5' triphosphate
DNA - deoxyribonucleic acid
DNase - deoxyribonuclease
EDTA - ethylenediaminetetraacetic acid
g - gram
IPTG - isopropyl-β-D-thio-galactopyranoside
kb - kilobases
kDa - kilodaltons
M - molar
O.D. - optical density
PAL - phenylalanine ammonia-lyase
PCR - polymerase chain reaction
PVP - polyvinylpolypyrrolidine
RNA - ribonucleic acid
mRNA - messenger RNA
RNase - ribonuclease
SDS - sodium dodecyl sulphate (lauryl sulphate)
Tris - tris (hydroxymethyl) aminomethane
UV - ultra-violet (light)
V - volt
Xgal - 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

ABSTRACT

The phenylpropanoid pathway enzyme chalcone synthase (CHS) was found to be encoded by a multigene family in the legume *Trifolium subterraneum* L. cv. Karridale. Restriction mapping and probing of seven clones isolated from a previously constructed genomic gene library of this plant established the existence of eight different genes. Seven of these genes were shown to be present in two clusters, one consisting of four members within 15kb of DNA, and the other of three members within 9kb. It is possible that these two clusters are part of a single one, but this was not verified. Two of the genes present in one of the clusters, CHS2 and CHS3, and one in the second cluster, CHS5, were sequenced in entirety, including approximately 1kb of promoter of each. The promoter and coding sequence of a third gene in the same cluster (CHS4), which was incompletely represented in the characterized genomic clones, was also sequenced. The DNA sequences of these four genes, and those predicted for their corresponding proteins, are highly homologous to those of the CHS genes/proteins of other plants. An intron, situated in a conserved position in all the CHS genes for which the genomic sequences are known, likewise occurs in the same location in the four genes. The promoters of the three genes CHS2, 3 and 4 share five regions in common, while certain elements established (or suspected) as playing a role in the transcriptional regulation of the CHS genes from other species, are present within them, and also in the promoter of CHS5. Primer extension enabled the start point of transcription for CHS3 to be established, but under the conditions used for the harvesting of plant material for the isolation of RNA, the

transcripts of CHS2 and CHS4 could not be detected. Probing and sequencing also found that the genes present in both clusters were orientated in the same direction with respect to the other members, and that one of the CHS genes (CHS7) was truncated (inactive). A brief analysis was also conducted of the molecular organization in clover of the genes coding for the first enzyme of the phenylpropanoid pathway, phenylalanine ammonia-lyase (PAL). This enzyme is also encoded by a multigene family in clover, four genes of which were isolated from the gene library. At least three of the clover PAL genes are clustered. One of these genes, designated PAL1, was sequenced. Transient expression assays of the promoter of CHS3 fused behind the β-glucuronidase reporter gene in *Nicotiana plumbaginifolia* protoplasts failed to demonstrate inducibility of the gene from a basal level of expression; the gene was found to be constitutively active in this system.

Table of Contents

Chapter 1. Introduction

1.1 General introduction

The elucidation of the molecular mechanisms which form the basis of plant-pathogen interactions provides not only an opportunity for understanding the control of gene expression in plants, but is also necessary for any future attempts to construct, via genetic engineering, disease resistant plants. The success or failure of attack by a pathogen is dependent upon a complex series of molecular signals and responses, which is essentially an interplay between host DNA and pathogen DNA (Brettell and Pryor, 1986). In a compatible interaction (host susceptible, pathogen virulent) the pathogen avoids the defence systems of the plant and causes disease, whereas in an incompatible interaction (host resistant, pathogen avirulent) the plant prevents infection (Lamb et al., 1989). Although compatible plant-pathogen interactions are rare in comparison to those which are incompatible, they nevertheless are of immense economic importance (Hahlbrock et al., 1987). The blight infection of potato, which in Europe last century caused death, misery and large-scale emigration, is an example of the devastating effect of such an interaction. The genetically near-homogenous cultivars that characterize many modern crops are obvious targets for new virulent pathogens.

To deal with attack by a pathogen, or exposure to other stress conditions such as irradiation with UV light and wounding, a plant will respond by activating one or more defence mechanisms. The biosynthesis of secondary metabolites is particularly important in this respect (Collinge and Slusarenko, 1987). The characterization of two multigene families of *Trifolium subterraneum,* which are essential to the synthesis of some of these compounds, is the subject of the following study.

1.2 Resistance genes and the hypersensitive response

Analysis of the genetics of plant-pathogen interactions has established that resistance of a plant to disease is frequently inherited as a simple Mendelian trait, with the gene for resistance in the host corresponding to a single gene in the pathogen for avirulence (Ellingboe, 1981). This gene-for-gene hypothesis, first established for flax and the rust *Melamspora lini* (Flor, 1946), holds that the matching between a host resistance gene (R gene) and a pathogen avirulence gene is uniquely specific; frequently both these genes are dominant (Ellingboe, 1981; Brettell and Pryor, 1986; fig. 1.1). As a plant and its pathogen may contain a number of different resistance and avirulence genes respectively, any gene pair interaction that confers incompatibility will be epistatic to others that may specify compatibility. Although a number of bacterial avirulence genes have been isolated and characterized, there is little information on the biochemical mechanisms by which their products act. The cloning of plant resistance genes has proved more difficult and as yet has been unsuccessful. The strategies which are being employed in attempts to clone these genes include the use of restriction fragment length polymorphisms, "shotgun cloning" (selecting for the acquisition of resistance), chromosome walking, and transposon tagging (Gabriel and Rolfe, 1990; Lamb et al., 1989).

A widespread phenomenon that is associated with incompatibility is the hypersensitive response (HR). The descriptive nature of this term refers to the rapid host cell death and cessation in growth of the pathogen that accompanies an incompatible interaction (Collinge and Slusarenko, 1986). Although a compatible interaction will lead to more damage, the time taken to attain this stage is much longer than that taken in the HR.

Viruses, fungi, bacteria and nematodes are all capable of inducing the HR, but some viruses can escape the dead cells and spread to the tissue beyond.

	HOST CULTIVAR	
	A Genotype: R1R1r2r2, or R1r1r2r2	B Genotype: r1r1R2R2, or r1r1R2r2
Race 1 Genotype A1a2	—	**+**
Race 2 Genotype a1A2	**+**	—

PATHOGEN

Fig. 1.1 The gene-for-gene hypothesis holds that the matching between a host resistance gene and a pathogen avirulence gene is specific for that pair of genes. This can be schematically illustrated by the so-called "quadratic check". In the case shown here, for two dominant resistance genes and their complementary dominant avirulence genes, two host cultivars are represented as interacting with two races of a bacterial pathogen. Cultivar A, which has at least one dominant resistance gene R1, produces an incompatible interaction with the pathogen having the corresponding dominant avirulence gene A1, and a compatible interaction with the pathogen that lacks this gene. Conversely, cultivar B, characterized by at least one dominant resistance gene R2, is only able to resist infection to the pathogen carrying the complementary dominant avirulence gene A2. A minus sign indicates resistance by the host, and a plus sign host susceptibility.

(Modified after Collinge and Slusarenko, 1987).

1.3 General plant defence responses

Although R genes may be essential for the resistance of many plants to specific infections, there are other *induced* defence mechanisms which may only be connected indirectly, if at all, to such genes. These include the accumulation of phytoalexins, lignin, hydroxyproline-rich glycoproteins, certain hydrolytic enzymes, and the so-called pathogenesis-related proteins (Collinge and Slusarenko, 1987).

These defence responses of plants can be induced not only by direct pathogen infection, but also by elicitor molecules isolated from fungal and bacterial cell walls and culture fluids. Elicitors include polysaccharides, glycoproteins, peptides and enzymes, and fatty acids (Dixon, 1986). Artificial elicitors, which are structurally unrelated to those isolated from the pathogen, and certain metabolites, can also produce a response in the host plant. Resistance genes may encode receptors (among other molecules involved in the biochemical pathways activated by stress) that interact with pathogen elicitor molecules (Lamb et al., 1989). These receptors are probably not located in the cell walls of plant cells, but in their plasma membranes (Scheel and Parker, 1990).

Phytoalexins are low molecular weight, antimicrobial molecules which are usually chemically and biogenetically related to secondary plant metabolites (Dixon, 1986). They constitute a diverse array of compounds, including isoflavonoids, sesquiterpenes, naphthaldehydes and acetylenes. Any given plant species will produce only a small number of phytoalexins, most of which will be chemically related. However, phytoalexins of different chemical origin have been isolated from some plants. For example, both isoflavonoid and non-isoflavonoid compounds have been obtained from the legumes *Lablab niger*, *Lathyrus odoratus*, and *Vicia faba* (Ingham, 1982). The necrosis of infected cells which

characterizes the HR almost always occurs in conjunction the accumulation of phytoalexins, usually in the adjacent living tissue (Mansfield, 1982).

Lignin is an essential component of vascular plant cell walls, and its biosynthesis in response to infection provides a barrier to further invasion. Cell walls are strengthened by the accumulation of hydroxyproline-rich glycoproteins (HRGPs), which also act as bacterial agglutinins (Dixon et al., 1986). Three major classes of HRGPs - the insoluble extensins, the soluble extensions and the lectin binding proteins - are believed to be specifically involved in plant defence (Graham and Graham, 1991). In the incompatible interaction of cucumber with *Cladosporium cucumerinum*, the accumulation of HRGPs was found to be associated with lignin deposition (Hammerschmidt et al., 1984).

The hydrolytic enzymes chitinase and β-1,3-glucanase are also part of the inducible defence response of plants. The substrates of these two enzymes, chitin and β-1,3-glucan, respectively, are important constituents of the cell walls of many fungi. After treatment of bean leaves with the plant hormone ethylene (whose biosynthesis increases in response to pathogen attack), chitinase and β-1,3-glucanase were found to accumulate almost exclusively in the cell vacuoles (Mauch and Staehelin, 1989). A small amount of β-1,3-glucanase, but no chitinase, was secreted into the cell wall spaces. On the basis of these observations, and taking ethylene treated tissue as being a valid model for stressed plant tissue in general, it has been suggested that chitinase and β-1,3-glucanase act as a last line of plant defence against fungal attack. The lysis of a few host cells would flood the fungus with these potentially lethal enzymes, giving it no time to adapt to their sudden appearance. Oligosaccharide fragments, released from the fungal cell wall by the secreted β-1,3-glucanase, could act as the elicitors that activate phytoalexin accumulation and the other

plant defence responses. However, in *Nicotiana tabacum* (tobacco) it has been found that the vacuolar β-1,3-glucanases are not released upon induction of the HR (van den Bulcke et al., 1989). This suggests that any role in plant defence played by these internally located tobacco enzymes must be restricted to the intracellular environment.

Although the exact cellular functions of the pathogenesis-related (PR) proteins remains largely unclear, it is possible that some of them are responsible for the phenomenon that a plant, following the successful tackling and termination of an infection, will become more resistant to further pathogenic attack (Collinge and Slusarenko, 1987; van Loon, 1985). Many PR proteins are localized extracellularly, but another group lacks signal peptides and hence are probably confined to the cytoplasm (van de Locht et al., 1990).

It appears that these active defence responses can be grouped together into two categories, depending on their relation to the advancing infection front (Graham and Graham, 1991). Some responses, such as the HR, the accumulation of phytoalexins and phenolic polymers, and often the rise in the level of chitinases, are normally confined to a zone adjacent to the area of infection. Other responses, including the accumulation of β-1,3-glucanases and phenylpropanoid gene activation (see below), may occur in cells which are some distance from the infection area, perhaps building up the defence potential of these cells in readiness for the imminent pathogenic assault.

1.4 Phenylpropanoid pathway gene expression and plant defence

The production of the secondary metabolites and proteins connected to plant defence involves a massive change in gene expression. This switch is largely at the level of transcriptional activation of the genes which code

for these molecules directly, or the enzymes that catalyze their formation (Cramer et al., 1985; Lawton and Lamb, 1987). In particular, the genes encoding enzymes of phenylpropanoid biosynthesis are activated.

A variety of plant secondary metabolites are derived from the phenylpropanoid pathway (Dixon, 1986; Collinge and Slusarenko, 1987; fig. 1.2). The first three reactions of the pathway involve the conversion of L-phenylalanine to 4-coumaroyl CoA. This biosynthetic sequence, which is called general phenylpropanoid metabolism, precedes the formation of virtually all phenylpropanoid compounds (Hahlbrock, 1981). The catalysing enzymes, in order of reaction, are phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (CA4H), and 4-coumarate:CoA ligase (4CL).

From 4-coumaroyl CoA branch pathways lead to the production of a range of compounds, including flavonoids and lignins. The oxidation state of the pyran ring can be used to divide flavonoids into twelve classes; the most important of these are the anthocyanins, flavones, flavonols, and isoflavonoids (Harborne, 1980). In the Leguminosae certain members of the isoflavonoid subgroup serve as phytoalexins (Mansfield, 1982). However, flavonoids are involved in many more biochemical processes besides that of defence. They play important roles in higher plants as flower pigments (Hahlbrock, 1981) and protectants against UV damage (Wellmann, 1983), and may possibly influence auxin distribution and concentration (Jacobs and Rubery, 1988). A further role, again specific to the Leguminosae, is their action as signal molecules for the induction of the *Rhizobium* nod genes (Long, 1989). Several naturally occurring isomers of the phenylpropanoid pathway derived dehydrodiconiferyl glucosides, isolated from *Vinca rosea* crown gall tumors, have been shown to stimulate cell division; it is possible that cytokinins may regulate cell division by controlling the availability of these compounds (Binns et

Photosynthesis

Erythrose-4-phosphate + Phosphoenolpyruvate

SHIKIMATE PATHWAY

L-Phenylalanine

↓ PAL

Cinnimate

↓ CA4H

4-Coumarate

↓ 4CL

4-Coumaroyl CoA

GENERAL
PHENYLPROPANOID
METABOLISM

Reduction	+ Malonyl CoA	+ Malonyl CoA
Lignins	Flavonoids	Xanthones and Stilbenes

PHENYLPROPANOID METABOLISM
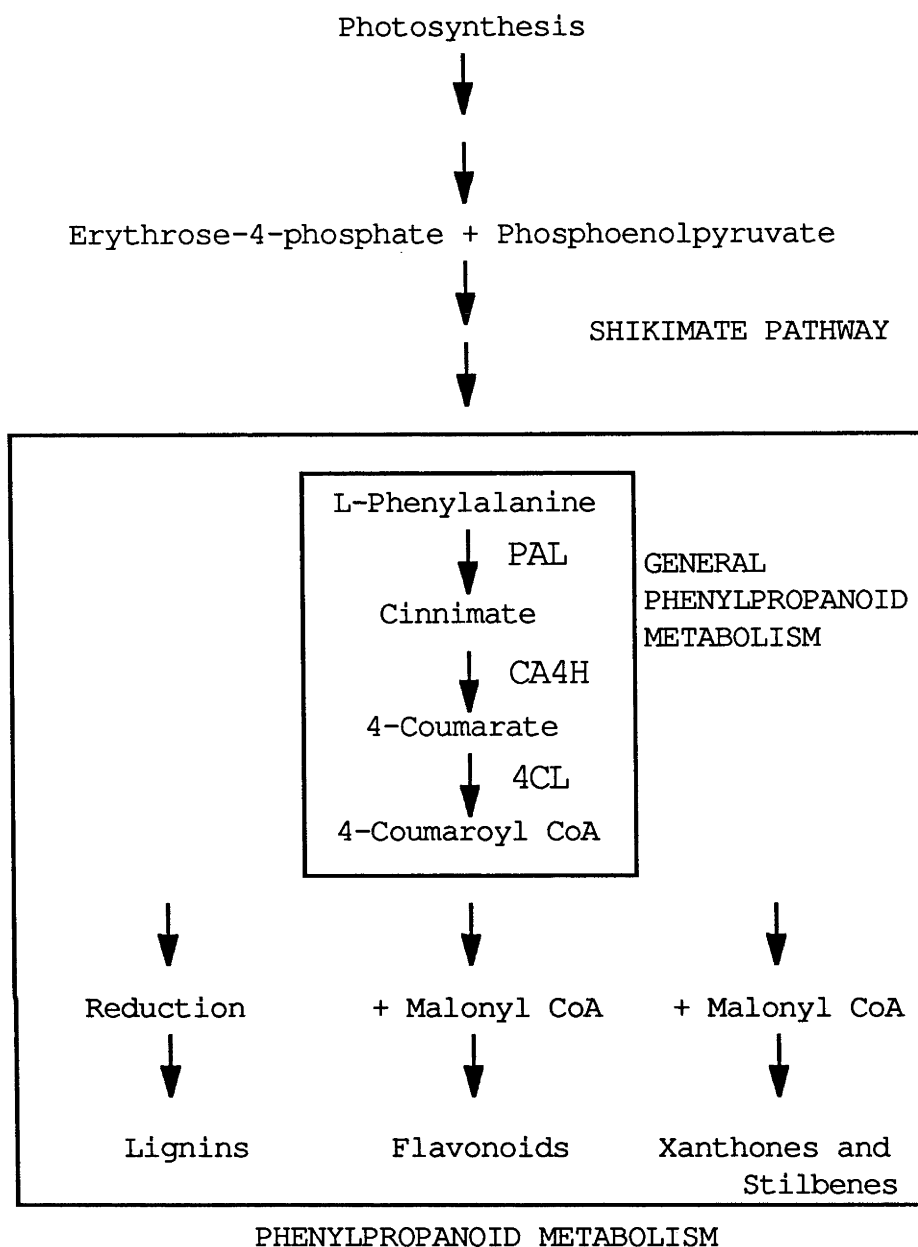
Fig 1.2 Diagram of the phenypropanoid pathway. The first compound in the pathway is L-phenylalanine, which itself is derived from the Calvin cycle via the shikimate (shikimic acid) pathway. The reactions of general phenylpropanoid metabolism (inner box) result in the production of the activated thioester, 4-coumaroyl CoA, from L-phenylalanine. At this stage, branch pathways lead to the synthesis of various aromatic compounds. Only the major groups of these are shown. The aromatic ring of the immediate precursor of 4-coumaroyl CoA, 4-coumarate, can be further modified by the addition of more hydroxyl groups, and the subsequent methylation of these, prior to esterification with CoA (not shown). This is especially common in the synthesis of the lignins (Grisebach, 1981).

al., 1987). The aglycone form of these dehydrodiconiferyl glucosides is able to induce the expression of the virulence genes in *Agrobacterium tumefaciens* (Hess et al., 1991). The induction of phenylpropanoid pathway genes by several of the above factors is discussed below (1.4.1-3).

The first two steps in flavonoid biosynthesis (from 4-coumaroyl CoA) are catalysed by the enzymes chalcone synthase (CHS) and chalcone isomerase (CHI), respectively. *The molecular characterization of the CHS genes present in the legume Trifolium subterraneum forms the basis of this thesis.* The reduction of 4-coumaroyl CoA, and other cinnamoyl CoA esters which are derived from it, is the committed step in lignin biosynthesis (Grisebach, 1981). Two enzymes are required to convert the CoA esters to their corresponding cinnamyl alcohols. The furanocoumarin phytoalexins of parsley (*Petroselinum hortense)* are derived from a branch pathway of phenylpropanoid metabolism which originates at 4-coumaroyl CoA (Hahlbrock et al., 1982).

1.4.1 Induction by elicitors, infection and wounding

An extensive amount of research has demonstrated the induction of the phenylpropanoid pathway genes, as well as other plant defence genes, in response to elicitors, infection by a pathogen, and wounding. Treatment of suspension-cultured bean (*Phaseolus vulgaris* L.) cells with fungal elicitor prepared from *Colletotrichum lindemuthianum* resulted in a marked increase in the transcription of the PAL and CHS genes, and a corresponding increase in the assayable activities of the respective enzymes (Lawton et al., 1983; Cramer et al., 1985; Lawton and Lamb, 1987). The induction of CHS mRNA activity was similarly observed after the treatment of soybean (*Glycine max* L.) cells with a fungal elicitor (Grab et al., 1985). Research using an alfalfa (*Medicago sativa* L.) cell culture system has demonstrated that following treatment with elicitor

the activities of a whole series of isoflavonoid biosynthetic enzymes are increased. Noteworthy is the rapid increase of mRNA species encoding distinct subunit isoforms of PAL and CHS (Dalkin et al., 1990), and the isoflavonoid specific enzyme isoflavone reductase (Paiva et al., 1991). Three isoforms of PAL have been isolated and characterized (Jorrin and Dixon, 1990).

The induction and transcription of PAL and CHS has not only been observed in elicitor treated cell cultures, but also in wounded and infected hypocotyls (i.e. in intact plant tissue). In an incompatible interaction of *C. lindemuthianum* with *P. vulgaris*, the accumulation of mRNA for these genes began approximately 40 hours after the inoculation of the fungus (Bell et al., 1986). It peaked at 69 hours, when the levels (as measured by RNA blot hybridization) were between 80 and 100 fold greater than those found in control uninfected hypocotyls. At the same stage in the compatible interaction there was no significant increase in the mRNA for PAL and CHS. However, later, during attempted limitation of lesion development in this interaction, the levels of mRNA corresponding to the two genes were found to rise markedly. In other plant systems it has been shown that CHS mRNA and protein levels are induced to much higher levels in an incompatible interaction than in a compatible one. For example, cultivars of chickpea (*Cicer arietinum* L.) occur which are resistant and susceptible to infection from the fungus *Ascochyta rabiei*. Treatment of both with an elicitor preparation from this pathogen resulted in a twofold higher induction of CHS mRNA in the resistant cultivar as compared to the susceptible one (Daniel and Barz, 1990).

The kinetics and extent of induction of the bean CHS genes (CHS occurs as a multifamily in this plant; 1.5.2) are dependent on whether they are activated by elicitor, infection, wounding or illumination (Ryder et al., 1987). In elicitor treated bean cells, the increases in the rates of both

PAL and CHS synthesis have been found to be concomitant with the accumulation of the isoflavonoid phytoalexins phaseollin and kievitone (Lawton et al., 1983; Ellis et al., 1989). The synthesis of CHS mRNA transcripts in wounded and infected bean hypocotyls has also been correlated with the production of phytoalexins (Bell et al., 1986). However, whereas with the bean cultivar Canadian Wonder the transcripts of at least four CHS genes were coordinately induced by elicitor from *C. lindemuthianum* (Ryder et al., 1987), this was not observed in the cultivar Imuna (Ellis et al., 1989).

Experiments with a gene fusion of the bean CHS8 promoter to the reporter gene β-glucuronidase (GUS) have provided evidence for the selective transcription of CHS genes in response to external stimuli. In tobacco, under normal (uninduced) conditions, the promoter was most active in the root apical meristem (where flavonoids act as nod gene inducers) and in petals (the site of accumulation of the anthocyanin pigments). However, in the case of mature leaves, upon wounding or treatment with a fungal elicitor, the promoter was observed to become highly active in the region adjacent to the stress site (Schmid et al., 1990).

In addition to stimulating the transcription of the phenylpropanoid pathway genes, elicitors induce those genes coding for HRGPs and chitinase. Bean cells, treated with the elicitor from *C. lindemuthianum*, exhibited a slower, but more prolonged, increase in the transcription of the HRGP genes when compared to the rapid rise of PAL and CHS transcription (Lawton and Lamb, 1987). Transcripts for chitinase, which catalyses the hydrolysis of β-1,4 N-acetylglucosamine linkages in the polysaccharide chitin, were induced 30 fold within twenty minutes of elicitor treatment (Hedrick et al., 1988). The transcriptional activation of the genes coding for HRGPs and chitinase has also been shown to occur when bean hypocotyls are infected with *C. lindemuthianum* (Lawton and

Lamb, 1987; Hedrick et al., 1988). Although the DNA sequences which are involved in the ethylene regulated expression of one bean chitinase gene have been characterized, it is not known whether the induction of this gene by other factors, such as pathogen attack and treatment with elicitors, is due to trans-acting factors regulating these same cis-acting elements (Broglie et al., 1989).

At least 1% of the total mRNA in elicitor induced bean cells can be accounted for by CHS mRNA (Dixon et al., 1986). This represents a major commitment of the protein synthesizing machinery of the cell to the production of one enzyme. The very fast nature in which elicitors induce the transcription of the PAL and CHS genes in bean is comparable to the most rapid gene activation systems (e.g. interferon) found in animal cells (Lawton and Lamb, 1987). This implies that mRNA induction is not a nonspecific secondary effect of elicitor treatment, but a primary response in plant defence (Ryder et al., 1987). The action of an elicitor maybe mediated through several multicomponent transduction pathways which all result in the activation of the same or similar genes (Bolwell et al., 1991). Nevertheless, the number of steps between the initial interaction of an elicitor with a cell and transcription of the appropriate defence genes must be small. It has been suggested that an elicitor-receptor complex could bind to cis-acting regulatory sequences in genes such as PAL and CHS, and thereby facilitate their transcription (Lamb et al., 1989). According to another model, the binding of elicitor to receptor causes the latter to open and form an ion channel; signal transduction is thought to be due, at least in part, to accompanying electrolyte fluxes (Gabriel and Rolfe, 1990). Treatment of parsley cells with the elicitor preparation from *Phytophthora megasperma* f. sp. *glycinea* has been shown to result in the calcium dependent phosphorylation of a number of proteins (Dietrich et al., 1990). Such data indicates that the transduction of elicitor signals is

possibly due to ion fluxes and calcium activated protein kinases (Scheel and Parker, 1990).

1.4.2 Induction by light

A number of studies have established that exposure to light acts to stimulate the transcription of mRNA for the three enzymes of general phenylpropanoid metabolism (PAL, CA4H, and 4CL), and those of the flavonoid branch pathway. Flavonoids absorb UV light in the critical range of 230-280nm, and may therefore provide plants with a UV protective mechanism (Wellmann, 1983). For example, quercetin 3-O-β-D-glucuronide is a UV induced flavonoid of dill (*Anethum graveolens*) which is known to absorb effectively in the UV spectral range (Mohle et al., 1985).

The rates of transcription of the genes coding for PAL and 4CL in parsley have been shown to be induced by both UV irradiation and the elicitor prepared from the cell wall of the fungus *P. megaspora* (Chappell and Hahlbrock, 1984; Kuhn et al., 1984). However, the transcription of CHS is only induced by UV light, and not by treatment with elicitor (Hahlbrock et al., 1982; Kreuzaler et al., 1983; Kuhn et al., 1984). This is presumably due to the absence of the isoflavonoid phytoalexins in parsley. Instead, the activities of enzymes specific to the synthesis of the parsley furanocoumarin phytoalexins, such as dimethylallyl diphosphate: umbelliferone dimethylallyl transferase, are stimulated by elicitors (Tietjen and Matern, 1983). Indirect evidence exists that elicitor induced stimulation of the furanocoumarin specific enzymes is due to increases in the levels of gene transcription (Scheel et al., 1987; Hahlbrock et al., 1987). In the case of carrot cells it has been found that UV light induces the transcription of one particular CHS isoform, as well as increasing the activities of PAL and CHI, these events coinciding with an accumulation

of anthocyanins. However, treatment with elicitor leads to a suppression in the activity of CHS but not PAL (Gleitz et al., 1991). The compound 4-hydroxybenzoic acid, the level of which increases upon elicitation, is presumably not a product of the flavonoid branch pathway.

In *Petunia hybrida* V30, which possesses a CHS multigene family (1.5.2), two CHS genes, A and J, are expressed under normal (uninduced) conditions in the flower tube and corolla (Koes et al., 1986; Koes et al., 1989a). The transcripts of J are 10-fold less than those of A. Treatment of seedlings with UV light was found to induce not only the transcription of these two genes, but also result in the expression of two further members (B and G) of the CHS multifamily (Koes et al., 1989a). Under these conditions, the transcripts of A and J constituted 98% of the total CHS mRNA present.

The response of the phenylpropanoid pathway genes of *P. hybrida* to elicitor induction, as opposed to UV induction, parallels that observed in parsley. UV light induces the expression of CHS mRNA in cell suspension cultures, whereas elicitors do not produce any response (Koes et al., 1989a). Although the phytoalexins of *P. hybrida* are unknown, the fact that elicitors will induce the expression of PAL in this plant suggests that they are derived from the phenylpropanoid pathway, but are not flavonoids.

Light induced CHS mRNA, the enzyme protein itself, and the flavonoid end products have all been localized to the epidermal cells in parsley (Schmelzer et al., 1988). However, in oat (*Avena sativa* L.) CHS, CHI and the terminal enzymes of the flavonoid pathway in that species were all found to be active in the leaf mesophyll (Knogge and Weissenbock, 1986). As between 60 and 70% of the flavonoid products were located in the epidermal layers, they must have been transported there after their biosynthesis.

Light must stimulate the plant defence responses through activation of photoreceptors. In parsley a UV-B light receptor, a blue-light receptor, and phytochrome are all involved in the response of the plant to light (Ohl et al., 1989). However, whereas elicitors induce the phosphorylation of proteins in parsley (1.4.1), UV light does not have this effect (Dietrich et al., 1990). It is hence likely that elicitors and UV light induce the same responses through different mechanisms. In mustard (*Sinapis alba*) CHS mRNA is induced by red-light, a phenomenon that is reversible by subsequent exposure to long-wavelength far-red light; this demonstrates the operation of phytochrome in CHS transcription (Ehmann et al., 1991). Analysis of the light mediated expression of the *Arabidopsis thaliana* CHS gene promoter in transgenic plants has established that blue and UV light cause greater CHS mRNA accumulation than red light; the role of a blue light receptor pathway has been proposed (Feinbaum et al., 1991).

1.4.3 Induction by nodulation and metabolites

The interaction of the Leguminosae with *Rhizobium* involves host-derived flavonoid molecules interacting with the bacterial *nod* D protein, converting it from an inactive form to an active one that can promote transcription of the other *nod* genes (Long, 1989). In the case of the soybean (*Glycine max*) -*Bradyrhizobium japonicum* symbiosis, it has been demonstrated that certain members of both the PAL and CHS multigene families present in the plant are preferentially induced upon infection with a nod⁺fix⁺ bacterial strain. Conversely, when this plant is infected with a nod⁺fix⁻ strain, an event that resembles a pathogenic attack as the symbiont contributes nothing due to its inability to fix nitrogen, different PAL and CHS genes are stimulated (Estabrook and Sengupta-Gopalan, 1991).

The metabolite glutathione (γ-L-glutamyl-L-cysteinyl-glycine - GSH), which probably plays a key role in protecting the cell from oxidative damage, is a further factor known to induce plant defence genes. Upon treatment of bean cell cultures with the reduced form of GSH, the transcription of the genes for PAL, CHS and the HRGPs, increases markedly (Wingate et al., 1988). In fact, the induction of the two phenylpropanoid pathway genes was observed to be much greater than that found with the optimal concentration of *C. lindemuthianum* elicitor. As in the case of elicitor treated cells, however, the induction of PAL and CHS was more rapid but less prolonged when compared to the HRGP genes. Experiments with protoplasts from alfalfa (*Medicago sativa* L.) have also shown that glutathione and the above elicitor have different effects on the expression of CHS; nevertheless, it is believed that both act via the same cis elements and trans-acting factors (Choudhary et al., 1990a/b).

1.5 Chalcone Synthase

As mentioned above (1.4), the essential core of this thesis is a characterization of the *Trifolium subterraneum* CHS genes.

1.5.1 Enzymatic properties of the CHS protein

CHS is the key enzyme of the flavonoid branch pathway of phenylpropanoid metabolism. It catalyses the stepwise addition of three acetate units from malonyl CoA to the phenylpropanoid moiety of 4-coumaroyl CoA, producing the C15 compound 2',4,4',6' tetrahydroxychalcone, also known as 6' hydroxychalcone and naringenin chalcone (Heller and Hahlbrock, 1980; Sutfeld and Wiermann, 1980; fig. 1.3). The sequential addition of the acetate units leads to the formation of a tri-β-ketoacyl (polyketide) thioester open chain compound which cyclizes either spontaneously or by enzyme mediation to form the aromatic A ring of the final product (Hahlbrock, 1981; fig. 1.3). The decarboxylation of malonyl CoA provides the acetyl carbanion that acts as the nucleophile in the condensation reactions.

The A ring hydroxylation pattern of compounds such as the isoflavonoid phytoalexins glyceollin, phaseollin and pisatin , and the chalcone echinatin from *Glycyrrhiza echinata*, implies that they must have the 6' hydroxy group of 2',4,4',6' tetrahydroxychalcone (numbered 5 in the (iso)flavanone structure) removed at some stage during their biosynthesis. Experiments involving the ^{13}C-NMR analysis of the incorporation of [1,2-^{13}C$_2$] acetate into pisatin and phaseollin have shown unambigiously that the three acetate units of the (iso)flavonoid A-ring are orientated in a non-random manner (Stoessl and Stothers, 1979; Dewick et al., 1982). However, in the case of the 5-hydroxy(iso)flavonoids (e.g. kaempferol), the ^{13}C label was found to be incorporated in an entirely

random fashion (Light and Hahlbrock, 1980; Dewick et al., 1982). These observations are consistent with reduction of the hydroxyl group in the former compounds occurring *prior* to the closure of the polyketide chain to form the A ring; hence, only *one* ortho hydroxyl (in the chalcone) remains to become the oxygen present in the flavonoid pyran ring, defining the orientation of the acetates. The compound that results from removal of the hydroxyl is 2',4,4' trihydroxychalcone (6' deoxychalcone). When the hydroxyl group is not removed, rotation is possible about the acyl bond of the A ring, allowing either of the *two* ortho hydroxyls to participate in formation of the pyran ring.

The CHS enzyme which catalyses the formation of 2',4,4',6' tetrahydroxychalcone can thus be more accurately defined as 6' hydroxychalcone synthase. By tracing the fate of the radioactive label in [2-^{14}C] malonyl CoA, the existence of a 6' deoxychalcone synthase activity has been demonstrated in *G. echinata* cells treated with yeast extract (Ayabe et al., 1988a). Incubation of the cells in a medium without NADPH, or with low concentrations (< 0.01mM) of this dinucleotide, produced only naringenin, the flavanone derived from 2',4,4',6' tetrahydroxychalcone. However, in the presence of high concentrations of NADPH (> 0.1mM), not only was naringenin synthesized, but also 2', 4, 4' trihydroxychalcone and its flavanone successor, liquiritigenin. When *G. echinata* protoplasts were incubated with ^{14}C phenylalanine, the label was found in liquiritigenin and the phytoalexin echinatin, but no incorporation could be detected in naringenin (Ayabe et al., 1988b). It was concluded from these experiments that there must exist a 6' deoxychalcone synthase enzyme, comprising 6' hydroxychalcone synthase and "other components" necessary for ketone reduction.

The isolation and characterization of a reductase protein from soybean (*Glycine max*) has provided evidence in support of this theory

(Welle and Grisebach, 1988). The reductase, a 34kDa single polypeptide, was shown to produce 6' deoxychalcone by reducing the appropriate carbonyl group of the polyketide intermediate with the pro-R hydrogen of NADPH. Synthesis was dependent upon the presence of 6' hydroxychalcone synthase, the reductase and NADPH. The two proteins appear to be linked in an ionic manner. It should be kept in mind, then, that the "CHS" protein is (or at least can be) involved in the production of two compounds (fig. 1.4).

The mechanism of action of 6' deoxychalcone synthase is very likely to be similar to that of 6-methylsalicylic acid synthase from *Penicillium patulum* (Kreuzaler and Hahlbrock, 1975; Dimroth et al., 1970). The latter enzyme converts acetyl CoA and malonyl CoA into 6-methylsalicylic acid in the presence of NADPH. It is proposed that the mechanism of this reaction involves reduction of a carbonyl group first to an alcohol, then to a *cis* double bond, and finally cyclization of this intermediate to form the aromatic ring. The reduction step that leads to the synthesis of 2', 4, 4' trihydroxychalcone probably occurs after the addition of the second or third acetate unit (Schroder and Schroder, 1990).

The similarity between CHS and β-ketoacyl-acyl carrier protein (ACP) synthase, the condensing enzyme of fatty acid synthesis in many lower organisms and in higher plants, has raised the possibility that these proteins have a common evolutionary origin (Schuz et al., 1983). With respect to molecular weight, subunit composition and details of reaction mechanism, the two enzymes resemble one another. The antibiotic cerulenin, which inhibits specifically the β-ketoacyl ACP synthase reaction in fatty acid synthesis, is also a potent inhibitor of CHS (Kreuzaler and Hahlbrock, 1975). Furthermore, as might be expected, 6-methylsalicylic acid synthase shares several common properties with β-ketoacyl ACP synthase (Dimroth et al., 1970). The ability of CHS to

catalyse, *in vitro*, chain elongation reactions with aliphatic derivatives of CoA, suggests that it could have evolved its present specificity for the aromatic compound 4-coumaroyl CoA from an enzyme, such as β-ketoacyl ACP synthase, that acts on aliphatic substrates (Schuz et al., 1983).

Finally, it should be noted that not only is the condensation of malonyl CoA with 4-coumaroyl CoA the key reaction of the flavonoid pathway, but it is also the committed step of both stilbene and xanthone biosynthesis. The enzyme resveratrol synthase (RS) catalyses the condensation of three molecules of malonyl CoA with one molecule of 4-coumaroyl CoA (like CHS), but produces 3',4',5'-trihydroxystilbene. RS must thereby liberate four molecules of CO_2 in the reaction, whereas only three are released by CHS (Schroder et al., 1988). Although the mechanism of action of CHS and RS must be similar, they are separate proteins as CHS activity cannot be detected in purified RS (Schoppner and Kindl, 1984). In the case of xanthone biosynthesis, the condensation reaction involves two molecules of malonyl CoA with one molecule of 4-coumaroyl CoA.
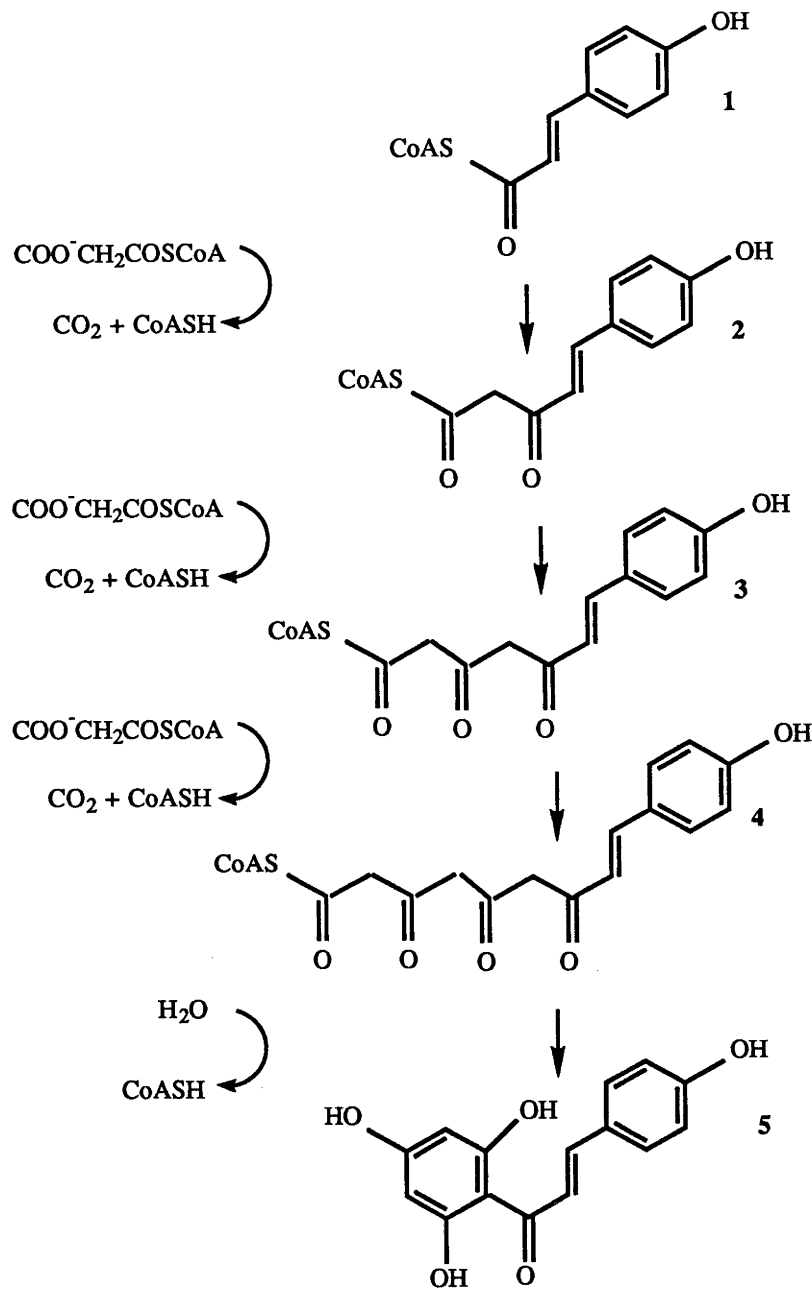
Fig. 1.3 Proposed mechanism for the catalytic action of the phenylpropanoid enzyme chalcone synthase from parsley. This enzyme is responsible for the biosynthesis of 2',4,4',6' tetrahydroxychalcone (5) from 4-coumaroyl CoA (1). The idea of a stepwise addition of the acetate units from three malonyl CoA molecules to the 4-coumaroyl CoA moiety (1), successively resulting in a mono-β-ketoacyl thioester (2), a di-β-ketoacyl thioester (3), and finally a tri-β-ketoacyl thioester (4), is supported by the *in vitro* synthesis of compounds which must be derived from such intermediates (Hrazdina et al., 1976). (figure modified after Hahlbrock, 1981)

PHENYLALANINE

⇩
⇩
⇩

4-COUMAROYL CoA + 3x MALONYL CoA

6'-HYDROXY
CHALCONE
SYNTHASE

6'-DEOXY
CHALCONE
SYNTHASE

1

2

CHALCONE
ISOMERASE

CHALCONE
ISOMERASE

3

4

⇩
⇩
⇩

⇩
⇩
⇩

5-HYDROXY
(ISO)FLAVONOIDS

5-DEOXY
(ISO)FLAVONOIDS
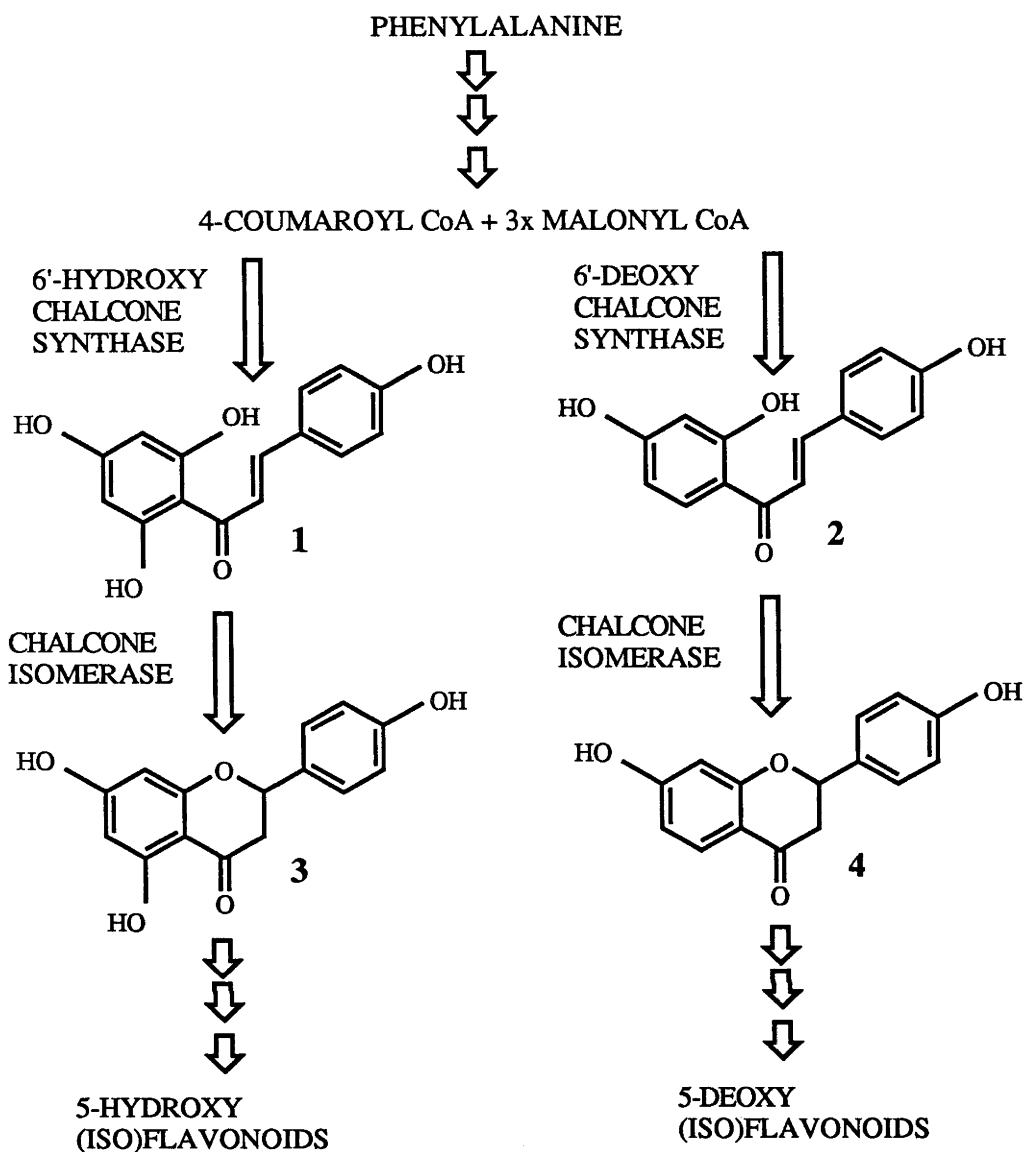
Fig 1.4 Diagram of the two chalcone synthase activities that are found in many plants. The divergence in the general flavonoid pathway at the step following 4-coumaroyl CoA is clearly shown. The enzyme 6' hydroxychalcone synthase is responsible for the synthesis of 2',4,4',6' tetrahydroxychalcone (1), while 6' deoxychalcone synthase produces 2',4,4' trihydroxychalcone (2). The latter compound, as its name implies, does not possess a 6' hydroxy group on the A ring. The two chalcones are converted by chalcone isomerase (CHI), in a stereospecific manner, to the flavonoids naringenin (3) and liquiritigenin (4) respectively. The numbering system for the carbon atoms in these compounds is such that the missing hydroxy group in liquiritigenin is now referred to as being from the 5 position. Further biosynthetic steps convert naringenin to the 5-hydroxy(iso)flavonoids, and liquiritigenin to the 5-deoxy(iso)flavonoids. (modified after Ayabe et al., 1988a)

1.5.2 Structure and molecular organization of CHS genes in plants

The CHS genes (or at least their cDNAs) have been isolated and sequenced from a number of plant species. The nucleotide sequences of these genes (excluding introns) are all approximately 1400bp, coding for a protein of ~ 43,000 daltons. In the case of the single CHS gene of parsley (*Petroselinum crispum*), the technique of *in situ* hybridization has enabled the localization of this gene, in the plant genome, to the long arm of a submetacentric chromosome pair (Huang et al., 1988). Although one 4CL gene in this species has been detected on the short arm of a submetacentric chromosome pair, it is unknown whether this is the same pair as that containing the CHS gene (Huang et al., 1989). If the CHS and 4CL genes do turn out to be linked, then it may have relevance to the fact that both are induced by UV light in parsley (Chappell and Hahlbrock, 1984).

The structure of the CHS gene appears to be conserved in evolution. The presence of an intron at a particular position is a distinctive feature of all CHS genes for which the genomic clones (as opposed to cDNA clones) are known (Schroder and Schroder, 1990). Although it varies in length, the intron is *always* inserted between the first and second nucleotides of a cysteine codon, approximately 180bp (60 amino acids) from the 5' end of the translated region. It is very likely that it reflects the evolutionary origin and domain structure of the protein (Niesbach-Klosgen et al., 1987). The CHS gene of *Antirrhinum majus* (snapdragon) contains a second intron between the codons of amino acids 162 and 163 (Sommer and Saedler, 1986); to date, however, no other CHS gene has been found to possess this second intron. As would be expected, the nucleotide sequence is less conserved between the CHS genes from different species when compared to that for the protein. In one study of the CHS cDNAs from several plant species it was found that they possessed a similarity of 66%

at the nucleotide level and higher than 80% at that of the (derived) amino acid sequence (Niesbach-Klosgen et al., 1987).

In some plants, such as snapdragon, parsley and maize, molecular and genetic analysis indicates the presence of one or two CHS genes per *haploid* genome. In the case of parsley, although there is only one gene per   cell, it occurs in two alleles, the major difference between these being the presence of a transposon in the promoter of one (Herrmann et al., 1988). Both allelic forms are UV light-inducible. However, small CHS multigene families have been discovered in the genomes of many plants, including all legumes examined to date. Those plants found to possess small multigene families (with the number of genes per haploid genome) are the legumes bean (*Phaseolus vulgaris* L., 6-8 genes; Ryder et al., 1987), pea (*Pisum sativum* L., at least 3 genes; Harker et al., 1990), soybean (*Glycine max* L., 6 genes; Wingender et al., 1989), chickpea (*Cicer arietinum* L., 5 genes; Daniel and Barz, 1990), and the non-legumes *Petunia hybrida* L. (at least 8 genes; Koes et al., 1989a/b) and mustard (*Sinapis alba* L., four genes; Batschauer et al., 1991).

The genes of the eight member CHS multigene family present in the strain Violet 30 of *P. hybrida* have been designated A,B,C,D,E,F,G, and J (Koes et al., 1989a/b). Besides these genes, which are each separated into two exons by an intron in the conserved position discussed above, four single exons (H,I,K, and L) have been isolated. It is not clear whether these exons represent isolated gene fragments, being the "by-products" of the evolution of the CHS multigene family, or whether they are parts of complete CHS genes possessing very large (>10kb) introns. One complete gene, D, does not have a translation stop codon in the usual position, but instead has one 30 codons downstream.

Hybridization experiments have revealed the presence of two subfamilies of clustered CHS genes among those constituting the *P.*

hybrida family (Koes et al., 1987). Clusters of CHS genes have also been found in bean (Ryder et al., 1987) and soybean (Wingender et al., 1989), but the spacing between adjacent genes is smaller than that observed in *P. hybrida*. Detailed sequence analysis of the *P. hybrida* CHS genes has shown that the homology between subfamily members is greater than 90%, and extends into both the 5' and 3' untranslated regions, as well as the introns (Koes et al., 1989b). Between those CHS genes which are not linked, homology is approximately 80% and is restricted to the regions coding for the enzyme. Surprisingly, when the predicted amino acid sequences of the UV light inducible genes A, B and G are compared, that of A is found to deviate considerably from the other two. As A is the principle UV inducible gene, it is possible that B and G are being inactivated during the process of evolution.

The introduction of additional CHS genes to *P. hybrida* has been found to result in the coordinate inhibition of both endogenous and introduced CHS gene expression in many transformants. This is possibly due to both the introduced and endogenous CHS copies being inactivated by methylation (Napoli et al., 1990; van der Krol et al., 1990).

Considering the similar reactions catalysed by CHS and RS (1.5.1), it is not surprising that the isolation and characterization of several genomic clones, and a cDNA, of RS from peanut (*Arachis hypogaea*) has revealed that these genes share many features in common with those coding for CHS (Schroder et al., 1988; Lanz et al., 1990). In peanut, RS consists of a multigene family of at least five members. The sequences of these RS genes, and those predicted for their corresponding proteins, possess substantial homology to the known CHS sequences. Moreover, an intron is present in each of the RS genes at exactly the same position as that conserved in the CHS genes. Significantly, however, many of the amino acids that are highly conserved in all CHS proteins are different in the

peanut RS proteins (Schroder et al., 1988). These changes are likely to be responsible for the distinct catalytic properties of the two enzymes.

1.5.3 Promoters of CHS genes

The induction of CHS genes by such diverse stimuli as elicitors, wounding and light, implies they possess important regulatory sequences in their promoters. Expression of an individual gene is very likely the result of a complex integration of the these cis-acting elements, or at least the proteins which bind to them.

The promoter of the snapdragon CHS gene has been subjected to extensive analysis (Sommer and Saedler, 1986; Kaulen et al., 1986; Sommer et al., 1988; Lipphardt et al., 1988; Staiger et al., 1989; Fritze et al., 1991). Fusion of it to a reporter gene, followed by analysis in a callus cell line and protoplasts, has allowed a number of regulatory sequences to be identified. The sequence -39 to -197, which lies immediately 5' to the TATA box, functions in the key role of an orientation independent UV light inducible element (Lipphardt et al., 1988). Deleting this region resulted in a reduction of the expression of the chimeric gene, when induced by white light, to 40% of the wild type level (Kaulen et al., 1986). Within its sequence the hexanucleotide 5'-TACCAT-3' is repeated twice between the TATA and CAAT boxes, and a motif with internal dyad symmetry, 5'-CACGTG-3', has been shown to be recognized by a nuclear protein designated CG-1 (Staiger et al., 1989/1991).

The level of response to UV light is enhanced by the sequence -197 to -357, but this element cannot cause UV induced gene expression on its own (Lipphardt et al., 1988). A third element, lying between -564 and -661, acts to increase transcription under both UV light and control (uninduced) conditions. The major feature of this latter sequence is two near identical (85% homology) 47bp duplications separated by a sequence which is

exactly the same as the simian virus (SV40) core enhancer. The positive regulatory effects of both the -197 to -357 and -564 to -661 sequences have also been demonstrated in transgenic plants (Fritze et al., 1991).

The sequence motifs described above are not unique to the promoter of the CHS gene of snapdragon. Sequences homologous to the SV40 enhancer occur in the promoters of CHS genes from *Arabidopsis thaliana* (Feinbaum and Ausubel, 1988), *P. hybrida* (Koes et al., 1989a), barley (Rohde et al., 1991) and mustard (Batschauer et al., 1991), and also in the promoters of the two 4CL genes of parsley (Douglas et al., 1987). A duplication of the sequence 5'-TACPyAT-3' occurs between the TATA and CAAT boxes of the CHS A gene of *P. hybrida* (Koes et al., 1989a), while a three-fold repetition of the analogous sequence 5'-TA(A)CCAA-3' is located in each of the promoters of the two parsley 4CL genes (Douglas et al., 1987). Two other sequences of the snapdragon promoter, each of approximately 20 nucleotides, have homology with regions of the *P. hybrida* CHS A promoter (van der Meer, et al., 1990).

Footprint analysis and light-induced expression studies of the parsley (*Petroselinum crispum*) CHS promoter have demonstrated the existence of four boxes that are important in UV induced transcriptional activity (Schulze-Lefert et al., 1989a/b; Weisshaar et al., 1991). These boxes, numbered I to IV in order of their proximity to the start of transcription, can be grouped into two light-responsive "units" (boxes I and II, and boxes III and IV). The unit consisting of boxes III and IV is inherently weaker at inducing transcription than that containing boxes I and II, but can partially compensate when the latter is inactivated by mutation. Sequences further upstream from these four boxes are also involved in the response of the promoter to light (Schulze-Lefert et al., 1989b), but boxes I and II are capable on their own of directing light-induced expression (Weisshaar et al., 1991).

The second box is characterized by an octamer with perfect dyad symmetry (5'-CCACGTGG-3'), and the third by a degenerate version of this octamer. However, the respective "partners" of these boxes, I and IV, share no homology to each other. Single base substitutions in box II have shown that the essential nucleotides include only six of the octamer, and a seventh adjacent to it (Block et al., 1990). Moreover, the resulting mutant phenotypes of these substitutions (as assayed by GUS activity) are asymmetric with respect to the centre of the palindromic octamer (i.e. in certain cases a base pair change will produce abolition of light-induced expression, while a change in the symmetrically positioned base pair leads to no change in expression). This implies that the palindrome is not important in the binding of the trans-acting factor, or that the factor may be dimeric with different recognition requirements for the two palindrome halves.

The symmetrical octamer found in box II of the parsley CHS gene promoter, or at least its hexanucleotide core 5'-CACGTG-3', occurs, sometimes with slight variations, in the promoters of CHS genes isolated from snapdragon (mentioned above; Sommer and Saedler, 1986), *Zea Mays* (Niesbach-Klosgen et al., 1987), soybean (Wingender et al., 1989), *Petunia hybrida* (van der Meer et al., 1990) and *Arabidopsis thaliana* (Feinbaum et al., 1991). This sequence is also conserved in the promoters of the small subunit of ribulose 1,5-bisphosphate (SSU) genes from numerous species, where it is termed the "G-box", implying that there may be similar control systems in light regulated genes (Schulze-Lefert et al., 1989a; Staiger et al., 1989). Its occurrence in the promoters of genes known to be under anaerobic, developmental and pathogen control, however, suggests that it is not exclusively responsive to light induced factors. For example, the promoter of the alcohol dehydrogenase gene of *Arabidopsis thaliana*, which is induced by anaerobiosis, and the abscisic

acid regulated wheat *Em* gene, both contain "G-box" sequences (McKendree et al., 1990; Guiltinan et al., 1990). Furthermore, this sequence occurs in the major late promoter of adenovirus and in a number of mammalian promoters (see references within Guiltinan et al., 1990).

Sequences bearing striking homology to the parsley CHS box I occur in many phenylpropanoid gene promoters, including those of the CHS genes of bean, snapdragon, maize, parsley and *Arabidopsis thaliana* (Lois et al., 1989). These promoter sequences are frequently accompanied by a second element, which has the consensus CCA(A/C)C(A/T)AAC(C/T)CC. The common occurrence of the two elements in these plant genes suggests they are important in transcriptional regulation (see chapter 4).

For one member of the bean CHS multifamily, CHS15, analysis of the expression of constructs containing deletions of the promoter has revealed the presence of two controlling regions (Dron et al., 1988). An elicitor regulated activator is located between the TATA box and position -173, and a "silencer sequence" between -173 and -326. Removal of the latter sequence results in enhanced gene expression. A protein factor, designated SBF-1 (see below), is known to bind to three "boxes" which are situated within the silencer; these boxes share the consensus sequence GGTTAA(A/T)(A/T)(A/T), which resembles that bound by the GT-1 protein in the pea SSU-3A gene promoter (Lawton et al., 1991). Deletion analysis of the soybean CHS1 gene promoter has defined the sequences necessary for elicitor inducibility to between -134 and -175 (Wingender et al., 1990). These sequences were also found to be inducible in parsley protoplasts, even though the parsley CHS gene is not activated by elicitor treatment (1.4.3). It is possible that the signalling pathway that responds to elicitors is present in parsley, but does not recognize the "native" CHS gene (cf. also below in relation to the bean CHS8 gene).

The regulatory sequences of the bean CHS genes been further studied by digestion with the enzyme DNase 1. Hypersensitivity to DNase 1 implies local opening of the chromatin structure, which is usually taken as evidence that the region is a cis-acting regulatory sequence to which trans-acting factors bind. In the case of the bean genes, three constitutive DNase 1 hypersensitive sites have been located in the 5' upstream DNA, and two that are induced by GSH in the region that spans the TATA box and the transcription initiation site (Lawton et al., 1990). These are probably binding sites for trans-acting factors involved in the activation of the genes.

The promoter of the bean CHS8 gene has been found to be induced by both wounding and elicitor in tobacco, a plant whose endogenous CHS genes do not respond to such stimuli due to the absence of an isoflavonoid phytoalexin biosynthetic pathway (Schmid et al., 1990). Hence, the bean CHS8 promoter must respond to a stress signal transduction system which is present in solanaceous plants, but is not recognized by their CHS genes. During evolution certain cis-acting elements of genes (such as PAL) which are stress induced in *all* higher plants, may have been incorporated into the promoters of legume CHS genes.

Chimeric genes created by fusing of the *Petunia hybrida* CHS A, J, B and G gene promoters to the GUS gene, and of the A gene promoter to the chloramphenicol acetyl transferase gene, have been observed to maintain the patterns of expression found in the respective wild-type CHS genes of this species (1.4.3; Koes et al., 1990; van der Meer et al., 1990). This suggests that the large differences in expression of the authentic genes is due, at least in part, to their promoters. In the case of the A gene, the two 5'-TACPyAT-3' elements (mentioned above) may be responsible for its flower-specific expression in mature plants (van der Meer et al., 1990).

1.5.4 Trans-acting factors and CHS expression

The protein SBF-1, which binds to sequences within the bean CHS15 promoter (1.5.3), has been purified from (bean) suspension-cell nuclei (Harrison et al., 1991). It has a subunit molecular weight of 95kDa, but may possibly bind to the regulatory sequences as a dimer. Treatment of SBF-1 with alkaline phosphatase eliminated its DNA binding ability; the activity of the protein is hence dependent upon phosphorylation. Interestingly, the AT-1 DNA binding protein, which recognizes certain AT-rich elements in the promoters of SSU and CAB genes, is only active in its unphosphorylated form (Datta and Cashmore, 1989). As phosphorylation appears to be a common feature in the control of animal transcription factors, it is very likely that more and more evidence will be uncovered relating to a similar role in plant gene expression. The bean CHS15 sequences recognized by SBF-1 are very similar to a group found in a number of other plant gene promoters which are bound by a protein factor GT-1 (Harrison et al., 1991).

Three cDNAs coding for proteins which bind to the parsley box II have been isolated from expression libraries of that plant (Weisshaar et al., 1991), while the nuclear factor(s) CG-1 which recognizes the snapdragon CACGTG sequence has been purified from tobacco (Staiger et al., 1991). The derived parsley proteins, called CPRFs (Common Plant Regulatory Factors), all possess the leucine zipper DNA-binding motif. As the mRNA for CPRF-1 accumulates after exposure to light, it is possible that this particular protein is involved in the light induced expression of the parsley CHS gene. A DNA-binding *activity*, designated GBF (G-box binding factor), which recognizes the G-box sequences of the SSU genes from a number of species, has been identified in tomato (*Lycopersicon esculentum*) and *Arabidopsis thaliana* (Giuliano et al., 1988) nuclear extracts. The G-box sequences which occur in the upstream regions of the

Nicotiana plumbaginifolia chlorophyll a/b binding protein (CAB) gene (Schindler and Cashmore, 1990) and the *Arabidopsis thaliana* alcohol dehydrogenase gene (McKendree et al., 1990), are also recognized by GBFs. The protein-DNA interactions of this *Arabidopsis* box, as deduced by footprinting, are identical to those seen for the parsley CHS box II (McKendree et al., 1990; DeLisle and Ferl, 1990). The relationship between the CPRFs, GBFs and CG-1 is not known.

The isolation and sequencing of the gene at the *C* locus of maize has shown that *it* may indeed code for a protein involved in the regulation of CHS transcription (Paz-Ares et al., 1987). The gene for CHS in this plant, which has been given the genetic notation *C2*, is known to be regulated by the loci *C*, *R*, *Vp* and *CIf*. Mutations in any of these loci not only completely eliminate, or greatly diminish, the activity of CHS in the aleurone layer of maize, but they have almost identical effects upon the flavonoid enzyme UDP glucose-3-O-glucosyl transferase (Donner, 1983). The putative protein of gene *C* has a basic amino-terminal region which shows 40% homology to the products of animal *myb* proto-oncogenes (Paz-Ares et al., 1987). Although the functions of these animal proteins are unknown, their basic domains possess a DNA binding activity. If the gene *C* protein is a DNA binding protein, then its acidic carboxy-terminus may function as a transcriptional activator. The activation of gene transcription in yeast is dependent upon acidic domains in its DNA binding proteins.

The transcription of three CHS genes of pea (called CHS1, 2 and 3) is controlled, at least in part, by two regulatory loci, *a* and *a2* (Harker et al., 1990). Two of the genes, CHS1 and CHS3, are expressed in petals, but their activity in this tissue cannot be detected if the plants carry a mutation in the *a2* locus. CHS3, but not CHS1, is expressed in petals in the presence of a mutation in the *a* locus. Induction of the three genes in roots by

treatment with CuCl$_2$ is independent of the genotype of the two loci. It has been suggested that *a* and *a2* may encode transcriptional activators.

1.6 Scope of this thesis

The work presented below is part of a programme aimed at understanding the molecular biology of the defence systems of *Trifolium subterraneum* (subterranean clover), and the way this plant interacts with *Rhizobium*. Subterranean clover (often just called "clover" in this thesis) is the most important legume present in the pastures of south-eastern and western Australia, with much of the cereal and grazing industries in these areas being dependent upon the nitrogen it fixes. The plant is not, however, a native to Australia. Most of the strains now present in Australia probably originated on or near the Atlantic seaboard of western Europe, arriving in Australia last century as a contaminant or admixture in imported agricultural seeds (Gladstones and Collins, 1983). The importance of clover to Australian pastoral industries means that it is not an insignificant subject for investigation. A thorough knowledge of the manner in which it deals with environmental stress may, in the long run, be of benefit to agriculture.

This thesis describes the isolation and characterization of eight members of a CHS multigene family from a previously constructed gene library of *Trifolium subterraneum* L. cv. Karridale. Detailed restriction mapping and probing of a number of clones isolated from this library established that seven of the CHS genes occur in two clusters. Sequencing allowed comparisons of the genes to be made with each other and with the CHS genes isolated from other species. A brief analysis was also conducted on the molecular organization of the PAL gene in clover. The PAL enzyme is encoded by a small multigene family in clover. Transient

expression assays of the promoter from the CHS3 gene in protoplasts showed that the promoter was constitutively active in the system. Primer extension studies demonstrated that at least one gene is expressed in young seedlings in response to light.

Chapter 2 Methods and Materials

2.1 Bacterial strains and Plasmids:

a) The bacterial (*Escherichia coli*) strains used in this study are listed below:

Strain	Characteristics	Reference
NM522	recA+, (supE, thi, Δ(lac-proAB), hsd5, (F',proAB, lacI^q, lacZdM15)	Gough and Murray, 1983
P2392	P2 lysogen of LE392. hsdR514, hsdM, supE44, supF58, lacY1, or Δ(lacIZY)6, galK2, galT22, metB1, trpR55	

b) The plasmids used in this study are listed below:

Plasmid	Characteristics	Reference
Bluescript SK (+)	A 2,958bp plasmid derived from pUC19, SK polylinker, Amp^r	Stratagene Cloning Systems, La Jolla, California, USA.
Lambda Dash 11	43kb genomic cloning vector derived from phage λ	Stratagene Cloning Systems La Jolla, California, USA.

2.2 Chemicals and Reagents:

Agar - Difco laboratories

Agarose - FMC BioProducts

Alpha ^{32}P dATP - Amersham

Ampicillin - Sigma

ATP - Sigma

β—mercaptoethanol - Sigma

5 (and-6) carboxyfluorescein diacetate - Molecular Probes Inc

Deoxyribonucleoside 5' triphosphates - Boehringer Mannheim

Developer and fixer - Agfa-Gevaert

Ethidium bromide - Sigma

Ficoll - Sigma

Glutathione - Sigma

Hybond-N - Amersham

IPTG - Sigma

4-methylumbelliferyl β-D-glucuronide - Sigma

4-methylumbelliferone - Sigma

Mineral salts - Gibco Ltd

PEG - Sigma

Phenol - Wako Pure Chemical Industries Ltd

Polaroid film (type 667) - Polaroid

Polyvinylpolypyrrolidone - Sigma

SDS (lauryl sulphate) - Sigma

Sephadex G50 - Pharmacia

Tris - Boehringer Mannheim

Tryptone - Difco laboratories

Xgal - Sigma

X-ray film - Kodak, Amersham

Yeast extract - Difco laboratories

2.3 Enzymes:

Calf Alkaline Phosphatase - Boehringer Mannheim

Cellulysin/cellulase - Calbiochem

Deoxyribonuclease 1 (from Bovine Pancreas) - Boehringer Mannheim

Klenow fragment of *E.coli* DNA polymerase 1 - Pharmacia

Macerase/pectinase - Calbiochem

Proteinase K - Boehringer Mannheim

Restriction enzymes - Boehringer Mannheim, Pharmacia

Reverse transcriptase - Superscript from BRL

Ribonuclease A (from Bovine Pancreas) - Boehringer Mannheim

RNasin - Promega

S1 nuclease - Boehringer Mannheim

T4 DNA ligase - Boehringer Mannheim, New England Biolabs

Taq DNA polmerase - Pharmacia (for PCR work), Promega (for sequencing)

2.4 Media:

All media were made up in distilled water and autoclaved. Both liquid media and solid media plates (i.e. media which had been autoclaved with agar and set on petri dishes) were usually stored at 4 ^{D}C. For solid media, agar was added at 15g per litre of solution prior to autoclaving. Media were never autoclaved more than once.

(1) Luria Broth (LB) (Miller, 1972).

Composition per 1 litre of water: tryptone (10.0g), yeast extract (5.0g), NaCl (5.0g).

(2) LM

Composition per 1 litre of water: tryptone (10g), yeast extract (5,0g), NaCl (5.0g), maltose(2g), 10ml of 1M $MgCl_2$, 2ml of 1M NaOH.

For 0.7% LM agar, 7g of agar was added to the solution.

(3) Minimal media

E.coli strain NM522 was kept on minimal media agar plates, as were all strains carrying plasmids for which single stranded DNA was required for sequencing.

Composition per 1 litre of water: D-glucose (2.5g), Gibco minimal salts (10g), $MgSO_47H_2O$ (2ml of a 1M solution), thiamine (1ml of a 1mg/ml solution).

(4) Phage Storage Buffer

200mM NaCl, 20mM tris pH 7.4, 0.1/% gelatin, 20mM $MgCl_2$.

(5) Media used to maintain the *Nicotiana plumbaginifolia* cell suspension culture is presented in 2.18.3.

Antibiotics and media additives:

Media were supplemented with antibiotics in order to facilitate the selection of bacteria carrying plasmids with corresponding resistance genes. They were always added to the media after autoclaving.

Antibiotic	*Concentration*
Ampicillin	100mg/l

Ampicillin was made up in water and stored in stocks of 100mg/ml at -20^DC.

Xgal and IPTG were added to provide blue-white colour selection of bacteria that had been transformed with the plasmid bluescript. The β-

galactosidase (*lacZ*) enzyme releases a blue compound after cleaving the colourless Xgal. In an *E.coli* strain which is mutant for the *lacZ* gene, the plasmid bluescript will produce white colonies if its own *lacZ* gene is inactivated by a cloned DNA fragment. IPTG is an artificial inducer of the *lac* operon.

Concentrations: Xgal   0.05mg/ml

                        IPTG   0.12mg/ml

2.5 General buffer solutions:

The following solutions were prepared as 10x stocks (except SSC) and diluted down, when needed, to the desired concentration. They were all prepared in distilled water and were never autoclaved.

(1) 10x *Tris* *E*DTA (TE) buffer:

Composition per 1 litre of water: tris (12.1g), EDTA (4.0g). pH 8.

(2) 10x *Tris* *E*DTA *S*alt (TES) buffer:

Composition per 1 litre of water: tris (12.1g), EDTA (4.0g), NaCl (5.9g). pH 8.

(3) 10x *Tris* *B*oric Acid *E*DTA (TBE) buffer:

Composition per 1 litre of water: tris (108g), EDTA (7.4g), H_3BO_3 (55g). pH 8.

(4) 20x *S*alt *S*odium *C*itrate (SSC):

Composition per 1 litre of water: NaCl (175.3g), sodium citrate (88.2g).

2.6 Other solutions:

(1) Preparation of phenol (Sambrook et al., 1989).

Solid phenol was melted and 8-hydroxyquinoline added to a concentration of 1%. This solution was extracted twice with equal volumes of 1.0M tris pH 8.0, and then twice with equal volumes of 0.1M Tris pH 8.0. The phenol phase was stored under 1xTE.

(2) Chloroform.

Chloroform was always made up with isoamyl alcohol in a ratio of 24:1. The isoamyl alcohol was added to aid separation of the organic and aqueous phases in the extraction of DNA (2.7.3).

(3) Potassium acetate

Potassium acetate, for use in "mini-prep" plasmid isolations, was made up as follows. To 60ml of 5M potassium acetate was added 11.5ml of glacial acetic acid and 28.5ml of water. The potassium was hence 3M and the acetate 5M.

2.7 Standard nucleic acid techniques

2.7.1 Restriction endonuclease digestion of DNA

Restriction enzymes were diluted down to "working" concentrations (usually about 2/3 units/μl) from the highly concentrated forms in which they were purchased. DNA samples were digested with the appropriate enzyme(s) for about 3 hours (longer if previous experience had shown that a particular DNA was difficult to cleave with an enzyme) in a 1x solution of the commercially supplied buffer. If necessary, such as in the

preparation of DNA for ligations, the reactions were terminated by heating at 85^{D}C for 9 minutes.

2.7.2 Ligation of DNA

DNA molecules were ligated overnight, at 4^{D}C, by 1 unit of T$_4$ DNA ligase in a 1x solution of *Boehringer Mannheim* buffer B (100mM tris pH 7.5/50mM MgCl$_2$/1M NaCl/10mM β mercaptoethanol) and 1mM ATP.

2.7.3 Extraction and precipitation of DNA

Phenol and chloroform were used to remove proteins from DNA. The solution of DNA was mixed with an equal volume of phenol and centrifuged for 3 minutes. The aqueous phase was re-extracted with an equal volume of chloroform and the nucleic acid within it precipitated by the addition of 0.1 volume of 3M sodium acetate pH 6.0 (or 0.1 volume of 5M NaCl) and 2.5 volumes of 95% ethanol (or 1 volume of isopropanol). The precipitation was left at least 2 hours, and often overnight, at -20^{D}C and then spun for 20 minutes. The supernatant was removed and the pellet washed in 70% ethanol, dried in a desiccator for 15 minutes, and finally resuspended in an appropriate volume of 1xTE.

2.7.4 Cleaning of DNA preparations

Any DNA sample which responded very poorly to restriction enzyme digestion was cleansed of contaminating substances by using a *Qiagen* (Diagen, West Germany) anion exchange resin. This was particularly the case with λ DNA isolated from the genomic library of *T. subterraneum*. To the preparation of DNA, made up to a volume of 250μl with water, was added 55μl of 5M NaCl and 20μl of 1M MOPS pH 7.0. A Qiagen tip (Tip 20), attached to a 1000 Gilson pipetman, was equilibrated in a solution of buffer C by pipetting 300μl in and out three times. The

DNA was adsorbed to the anion resin by pipetting it gently in and out of the column four times. Washing was accomplished by similarly pipetting a 1.5ml solution of buffer C, and the DNA then eluted by pipetting 200μl of buffer F three times, and repeating this step twice with fresh 200μl aliquots of buffer. The eluates were combined and precipated with isopropanol (2.7.3).

Buffer C: 1M NaCl/50mM MOPS/15% ethanol/pH 7.0

Buffer F: 1.5M NaCl/50mM MOPS/15% ethanol/pH 7.5

2.7.5 "Mini prep" isolation of plasmid DNA from *E.coli*. (Procedure based on Birnboim and Doly, 1979).

The *E.coli* colony containing the required plasmid, or putative recombinant from ligation, was inoculated into 5ml of LB and incubated, under appropriate antibiotic selection pressure (2.4), for at least 3 hours. An Eppendorf tube was filled with 1.5ml of the bacterial solution, spun for 3 minutes, and the pellet resuspended in 100μl of "alkaline lysis buffer" (50mM glucose/10mM EDTA/25mM tris pH 7.5). After standing at room temperature for 5 minutes, 200μl of 0.2M NaOH/1% SDS was added and the solution placed on ice for 5 minutes, followed by the addition of 150μl of potassium acetate (3M/5M; 2.6) and incubation on ice for a similar length of time. Cellular debris was removed by centrifugation for 4 minutes and the supernatant extracted with 300μl of 1:1 phenol/chloroform. The DNA was precipitated on ice for 2 minutes with 0.8ml volume of ethanol and pelleted by spinning for 4 minutes. It was dried in a vacuum desiccator for 20 minutes and resuspended in 1xTE.

2.7.6 Preparation of plasmid DNA from *E.coli* by density gradient (CsCl) centrifugation. (Procedure based on Clewell and Helinski, 1969).

A single colony from the *E.coli* strain containing the desired plasmid was inoculated into 10ml of LB and grown up, under antibiotic selection pressure (2.4), for 4-6 hours. The solution of dense bacteria was poured into a flask containing 500mls of LB (and antibiotic), and incubated overnight. The cells were harvested by centrifugation for 10 minutes (7,000rpm/4^{D}C) in a GS-3 rotor and the pellet resuspended in 9.5ml of 25% sucrose/1xTE. A freshly prepared 1ml solution of 20mg of lysozyme in water, and 5ml of 0.25M EDTA, were added together and incubation conducted on ice for 10 minutes. After the addition of 15ml of 1%Brij 58/0.4% sodium deoxycholate, the solution was left on ice for a further 45 minutes. The cellular debri and the bulk of the chromosomal DNA were removed by centrifugation for 60 minutes (18,000rpm/4^{D}C) in a SS34 rotor, and the plasmid DNA in the supernatant precipitated, for at least 2 hours at 4^{D}C, with 8mls of 50% PEG and 8ml of 5M NaCl. The PEG-DNA precipitate was recovered by centrifugation for 12 minutes (7,000rpm/4^{D}C) in a SS34 rotor, taken up in 6.5ml of 1xTES, and transferred to a corex tube containing 7.9g of CsCl. 0.6ml of a 10mg/ml solution of ethidium bromide was added and the CsCl dissolved by inverting the tube, covered with nescofilm, and then leaving it to stand for 45 minutes on ice. The corex tube was then centrifugated for 30 minutes (10,000rpm/4^{D}C) in a SS34 rotor, the ethidium/DNA/CsCl solution carefully removed from the PEG and placed in a polyallomer Quick-seal centrifuge tube (Beckman). The tube was heat sealed after using paraffin to fill the remaining space within it. Centrifugation was conducted in a 50Ti rotor for 40-48 hours (40,000rpm/20^{D}C), resulting in the plasmid DNA separating into a distinct band *below* that of the chromosomal DNA. The plasmid DNA was removed from the tube with

the aid of a syringe, extracted four times with a solution of n-butanol saturated with 1xTE, and, finally, dialysed against three changes of 1xTE at 4^{D}C.

2.7.7 Phosphatasing of DNA

Calf alkaline phosphatase (CAP) was used to remove the 5' phosphate groups from plasmids to prevent their re-ligation during cloning. After restriction enzyme digestion, the plasmid was placed on ice and one-eighth the digestion volume of 500mM tris pH 9.3/1mM MgCl$_2$ added, together with 1μl of CAP (1 unit/μl). The volume of the solution was adjusted with water to a state where the above buffer was one-tenth its original concentration. The dephosphorylation step was performed for 30 minutes at 37^{D}C and the enzyme then removed by extraction with phenol and chloroform. The DNA was recovered by ethanol precipitation.

2.7.8 Blunting of restriction ends

The Klenow fragment was used to blunt restriction ends where necessary. This proved useful in the construction of a number of sequencing clones: fragments were cleaved out of clones, the ends of the parent molecules blunted, and then "self-ligations" performed. After digestion with the appropriate restriction enzyme(s), the reaction was terminated by heating at 85^{D}C for 9 minutes. To this solution were added Boehringer Mannheim buffer B (1μl), 2mM dNTPs (1μl), Klenow fragment (1 unit/1μl), and water (7μl). Reaction was conducted at 37^{D}C for 10 minutes and then terminated by heating at 65^{D}C for 5 minutes. The molecules were then standardly ligated.

2.7.9 Gel electrophoresis of nucleic acids

Agarose gel electrophoresis of DNA was performed horizontally in perspex chambers. Gels were of two sizes: (1) large (190x140x5mm); (2) small (52x90x5mm). Ethidium bromide was added to the liquid agarose before it set (10μl of a 10mg/ml solution per 150ml of agarose). The large gels, which were run overnight at 16mA/50V, were used to obtain clear separation of DNA molecules and to accurately ascertain their length. When the task at hand did not require such accuracy, it was standard practice to run the DNA for 1-2 hours at 40mA/100V; this was frequently done to gauge the concentration of a DNA sample. In all cases the electrophoresis buffer was 1xTBE and the concentration of agarose 0.8-1% (w/v).

A "loading dye" (10% glycerol/100mM EDTA/0.03% bromophenol blue/pH 8.0) was added to all DNA samples prior to their electrophoresis. Following electrophoresis, the DNA bands in the gel were visualized on a UV transilluminator and photographed.

2.7.10 Isolation of DNA from an agarose gel

DNA was isolated from gels by the use of DEAE cellulose membranes.

(a) Preparation of membranes. The membranes (Schleicher and Schuell, NA45) were first soaked in 10mM EDTA pH 7.6 for 10 minutes and then in 0.5M NaOH for 10 minutes. They were stored in water.

(b) DNA isolation. A groove was cut in the agarose infront of the DNA band of interest and into this was carefully inserted a membrane; the gel was then electrophoresed so that the DNA ran onto the membrane. The membrane was removed from the groove, rinsed in ~250μl of *low salt* solution (20mM tris/0.1mM EDTA/0.15M NaCl/pH 8.0) for 10 minutes at room temperature, and then placed in ~250μl of *high salt* solution (20mM tris/0.1mM EDTA/1.0M NaCl/pH 8.0) for 1 hour at 60^{D}C. Two volumes of

ethanol were added and the DNA precipitated for at least 1 hour at -20^{D}C; it was recovered by spinning and vacuum drying as per usual.

2.8 Sucrose gradients

DNA fragments were sometimes prepared for cloning by purification on a concentration gradient of sucrose. This was set up by carefully layering, in succession, 2.5mls of 40%, 30%, 20%, and 10% sucrose solutions (each in 20mM tris pH 8.0/10mM EDTA/100mM NaCl) into a 14x89mm pollallomer centrifuge tube (Beckman). Centrifugation was performed for ~16 hours at 34,000rpm in a SW41Ti rotor, after which a hole was punched in the base of the tube and 200µl fractions collected in Eppendorf tubes. The position of the DNA within the fractions was assayed by electrophoresing 10µl of each in an agarose gel. The DNA was recovered by diluting the sucrose in the appropriate fractions to less than 10% with 1xTE and then isopropanol precipitation.

2.9 Electroporation of DNA into E.coli:

2.9.1 Preparation of E.coli cells for electroporation

E.coli cells to be used for electroporation were made up in large quantities by the following procedure. The cells were grown up in 1 litre of LB at 37 ^{D}C until the O.D. (600nm) was between 0.5 and 1.0, centrifuged at 4^{D}C (4,000rpm/15 minutes) and resuspended in 1 litre of ice cold millipore water. Three identical centrifugation steps were performed, resuspending the cells successively in 0.5 litres of millipore water, 20ml of 10% glycerol (the glycerol being made up in millipore water), and finally 3ml of 10% glycerol. Aliquots of the cells, each of 100µl, were stored at -70^{D}C until used.

2.9.2 Electroporation

Electroporation was performed using the Gene Pulser instrument of Bio-Rad. The DNA to be electroporated was dissolved in 10ml of millipore water (after being subjected to phenol/chloroform extraction and ethanol precipitation) and mixed thoroughly with half an aliquot of thawed cells (i.e. 50μl). These were inserted into a 2mm gap electroporation cuvette and pulsed once; the instrument was set to 25μF capacitance/2,500V and delivered a pulse of about 12.5kV/cm. A sterile pasteur pipette was used to remove the cells from the cuvette and place them in 5ml of LB. After gentle incubation for 1 hour at 37^{D}C, aliquots were spread on the appropriate selective media, the volume used being determined by the expected efficiency of electroporation.

2.10 Isolation of λ clones from the *T. subterraneum* gene library:

2.10.1 Plaque purification (Slightom and Drong, 1988; Sambrook et al., 1989).

Large LB plates (22x22 cm) containing a confluent lysis of *E.coli* P2392 by the phage were hybridized with the particular probe for 48 hours. By using wide mouthed pipetman tips (created by cleaving off the bottoms of normal tips), each region of lysis shown by autoradiography to contain a hybridizing signal was sucked up and placed into 1ml of phage storage buffer (PSB; 2.4). After vortexing, the phage solutions were placed at 4^{D}C for at least 4 hours. A volume of 1μl was taken from a selected number of these original stocks and diluted 10^{-4} and 10^{-5}. From *these* dilutions, 1μl samples were taken and added to Falcon tubes containing 1ml of LM and 300μl of fresh P2392[*] (O.D.(650nm) ~ 0.5). The solutions were incubated for 25 minutes at 37^{D}C, 3ml of 0.7% LM agar added to each, and plated

on LB. After incubation overnight at 37^{D}C, the plates were blotted to Hybond-N and probed.

Two further rounds of this procedure, in each case increasing the amount of phage from the picked off hybridizing regions added to the P2392, allowed pure positively hybridizing plaques to be obtained. Individual plaques were placed into 1ml aliquots of PSB and stored at 4^{D}C for 24 hours before being used in the following procedure. Falcon tubes containing 3ml of LM, 300µl of P2392*, and 500µl of the stocked phage (a volume large enough to achieve confluent lysis on the host strain) were incubated for 25 minutes at room temperature (23^{D}C), combined with ~8mls of 0.7% LM agar and poured onto large (diameter 15cm) LB plates. The plates were left to set for 20 minutes, then inverted and incubated overnight. The following day each plate was flooded with 15ml of PSB and stored at 4^{D}C for 4 hours with occasional shaking. By tilting the plates at an angle (~20 degrees), the top agar could be scraped to the lower corners, and, together with the PSB, drawn off and placed in SS34 tubes. A further 1ml of PSB was added to each plate, swirled over the surface and combined with the respective previous harvests. The tubes were centrifuged in a SS34 rotor (10,000 rpm/10 mins) and the supernatants removed and placed in 50ml plastic tubes (Sarstedt, Germany). Chloroform was added to 0.3% and the *phage storage stocks* kept at 4^{D}C.

*P2392 was grown up in LM.

2.10.2 Titration of phage storage stocks (Miller, 1987).

In order to calculate the concentration of phage particles in a phage storage stock, the number of plaque-forming units (pfu) per ml of stock was determined. The number of pfu of bacteriophage λ is about half of the number of phage particles, the latter figure usually being in the

range of 10^7-10^{11} particles/ml. A simple dilution series was first performed to obtain a 1ml solution of the phage at a 10^{-6} concentration of that of the original phage storage stock. From this diluted lysate, 1μl (10^{-9}), 10μl (10^{-8}) and 100μl (10^{-7}) aliquots were removed and added separately to Falcon tubes containing 1ml of LM and 100μl of *E.coli* P2392. The tubes were incubated in a 37^DC oven for 20 minutes and then, after the addition of 3ml of 0.7% LM agar, plated on LB petri dishes. A control tube, containing only LM and P2392, was subjected to this procedure to be certain that the host strain had not gained a λ infection from elsewhere. The plates were incubated overnight at 37^DC and the number of plaques counted on each. The number of pfu for the phage storage stock was deduced by taking into consideration the dilution factors.

2.10.3 Isolation of phage DNA (Slightom and Drong, 1988).

A single colony of *E.coli* P2392 was inoculated into 5ml of LM and incubated overnight. From this culture, 500μl was added to 50ml of LM and grown up until the the cells had an O.D. (650nm) of ~0.4 (about 2 hours). To three 2ml aliquots of these cells were added differing amounts of phage (from the phage storage stocks) to include *and* bracket the figure of 4×10^6 pfu. The respective quantities of phage to use, which were determined from the results of the titre calculations (2.10.2), were usually in the range of 10-200μl of the phage storage stocks. After standing in a 37^DC oven for 20 minutes, the three adsorptions were added to separate 50ml aliquots of LM and incubated at 37^DC until lysis of the host strain (~6 hours). The three solutions were combined (i.e. total volume 150ml), spun in a GS-3 rotor (8,000rpm/12 mins), and the supernatant decanted into a fresh tube. DNAse 1 and RNAse A were added to give a final concentration for each enzyme of 1μg/ml; digestion

was allowed to proceed for 30 minutes at 37^DC. The DNA was precipitated overnight with 40ml of 5M NaCl and 40ml of 50% PEG. The following morning the precipitation was centrifuged in a GS-3 rotor (8,000rpm/12 mins), the supernatant discarded and the pellet resuspended in 3ml of 1xTE/10mM NaCl. The solution was made 0.5% (w/v) with respect to SDS (by adding 75μl of 20 % SDS) and digestion of protein conducted for 30 minutes at 65^DC with 300μg of proteinase K. Remaining protein was removed by phenol/chloroform extraction and the DNA precipitated for 2 hours at -20^DC in a corex tube with 2.5 volumes of 96% ethanol and 0.1 volume of 3M NaAcO pH 6.0. The tube was centrifuged in a SS34 rotor (8,000rpm/12 mins) and the pellet washed in 10ml of 70% ethanol by placing nescofilm over the mouth and shaking. After an identical centrifugation, the new DNA pellet was vacuum dried in a desiccator for 20 minutes and resuspended in 150-200μl of 1xTE.

2.11 Membrane transfer (Southern, 1975; Reed and Mann, 1985).

2.11.1 Transfer of DNA to Hybond-N (nylon) membranes

(a) From bacterial colonies/plaques

A Hybond-N membrane (Amersham), marked with orientation dots and cut to the correct size, was placed onto the plate and the position of the dots duplicated onto the plastic surface. After 2 minutes the membrane was peeled off and placed on 3MM paper prewetted with 0.5M NaOH/0.5M NaCl for 10 minutes. It was then transferred to 3MM paper prewetted with 0.5M tris pH 7.2/2.0M NaCl for 10 minutes, washed briefly in 2xSSC, air dried, and finally the DNA crosslinked to the surface with the aid of a UV Stratalinker 1800 (Stratagene) set at 2500

joules x100. In the case of plaques, the plates were placed at 4^DC for at least 30 minutes prior to lifting.

(b) From an agarose gel

Plasmid DNA

The agarose gel containing the DNA to be transferred was soaked for 10 minutes in 0.25M HCl, rinsed briefly in distilled water, and soaked for 45 minutes in 0.5M NaOH/1.5M NaCl. It was then placed face down upon a perspex plate and a sheet of Hybond-N, prewetted in 0.5M NaOH/1.5M NaOH, positioned on top in one move. The membrane was followed by two sheets of 3MM blotting paper, similarly prewetted in the alkaline solution, a pile of paper towels (about 4cm thick), and finally a wooden weight. After blotting overnight, the membrane was washed twice in 2xSSC (10 minutes each time), and the DNA permanently crosslinked with UV light as above. The membrane was wrapped in plastic and stored at 4^DC until used.

Total genomic DNA

The procedure for total genomic DNA from subterranean clover was essentially the same as that for plasmid DNA. The gel was soaked in acid and base as above, but here it was exposed to a final 30 minute neutralization step in 0.5M tris pH 7.2/2.0M NaCl. Blotting was performed with the aid of a wick and two reservoirs of 20xSSC. The membrane was cross-linked as above.

2.11.2 Transfer of DNA to nitrocellulose membranes

Nitrocellulose (Schleicher and Schuell, West Germany) was sometimes used in preference to nylon for the transfer of bacterial colonies. The procedure was identical to that used for Hybond-N (2.11.1), except that after being air dried after the transfer, the filter was baked *in vacuo* for 2 hours at 80^DC to irreversibly bind the DNA to the surface.

2.12 Preparation of radioactive probes

2.12.1 Amplification of DNA

The polymerase chain reaction (PCR) method was used to amplify DNA for subsequent radioactive labelling as a probe (although sometimes DNA isolated from an agarose gel (2.7.10) was directly labelled), and as outlined in chapter 5 for the isolation of a particular clover genomic fragment. To approximately 20ng of DNA starting material was added: buffer - 500mM KCl/100mM tris pH 8.3/15mM $MgCl_2$/0.1% gelatin - (10µl), primers (5µl of each), dATP, dTTP, dGTP, and dCTP (1µl each of 20mM solutions), Taq 1 DNA polymerase (1µl/5.0 units), and water to make the final volume 100µl. In those cases where the sequence to be amplified was cloned into the polylinker of Bluescript, the two primers used were T7 and T3. Paraffin (100µl) was carefully layered on top and the PCR reaction conducted by using a Perkin Elmer Cetus (USA) DNA Thermal Cycler 480. The particular programme used was dependent upon the degree of homology of the primers to the target DNA. Upon the completion of amplification, the paraffin was removed and a small aliquot of DNA (~ 5µl) electrophoresed on an agarose mini-gel to check that the reaction had worked. DNA was recovered by phenol/chloroform extraction and ethanol precipitation. If need be, the flanking promoters were removed by cutting with the appropriate restriction enzyme(s).

2.12.2 Radioactive labelling

DNA was radioactively labelled by random primed synthesis from oligodeoxynucleotides. The Amersham *Multiprime* kit, and protocol provided with it, were standardly employed. The DNA to be labelled, usually in a volume of 10 - 20µl, was denatured by boiling for 2 minutes.

After being placed on ice, the following solutions were added: buffer (5μl), primers (5μl), unlabelled dTTP, dGTP, and dCTP (4μl each of 20mM solutions), α ^{32}P dATP (5μl), 2 units of Klenow fragment of *E.coli* DNA polymerase 1 (2μl), and the appropriate volume of water to make the volume up to 50μl. This reaction mixture was incubated at 37^{D}C for 45 minutes and terminated by the addition of 5μl of loading dye (20% sucrose/5mM EDTA/1% SDS/0.2% bromophenol blue). It was then passed down a Sephadex G-50 column (1x6cm; buffer 1xTES) to separate the unincorporated nucleotides from the actual probe. Unless used immediately, the probe was stored at -20^{D}C.

2.13 Hybridization and autoradiography:

Prehybridization and hybridization were performed in plastic bags, firmly lidded plastic boxes, or in Hybaid mini hybridization ovens (Hybaid Limited, Middlesex, United Kingdom). The prehybridization and hybridization buffer, *per 10ml*, consisted of 20xSSC (3ml)/ millipore water (5ml)/ 50% dextran sulphate (1ml)/ 100x Denhardt's solution (0.5ml)/ 10% SDS (0.5ml). 100x Denhardt's solution was 2% BSA/ 2% Ficoll (400,000)/ 2% polyvinylpyrrolidone (360,000). Membranes were prehybridized for at least 2 hours. After the elapse of this time, the probe, boiled for 2 minutes to separate the DNA strands and placed on ice for 3 minutes to lock them in a disordered state, was added and hybridization carried out for 16-20 hours. In the case of the bean cDNA probe, the temperature for both prehybridization and hybridization was 55^{D}C; homologous hybridizations were carried out at 65^{D}C. After hybridization the probe was sometimes bottled and kept at -20^{D}C for future use. The membranes were washed twice in 2xSSC/0.1% SDS at the hybridization temperature (30 minutes each time) and air dried. Autoradiography was

conducted in X-ray cassettes for 1-7 days at -70^{D}C; the films were developed and fixed in a dark room. Alternatively, the hybridizing bands were visualized by the use of the Molecular Dynamics (California, U.S.A) PhosphorImager.

2.14 Deprobing of filters

Hybond-N filters were deprobed by first incubating in 0.4M NaOH for 30 minutes at 45^{D}C, and then incubating in a solution of 0.1xSSC/0.1% (w/v) SDS/0.2M tris pH 7.5 for a further 30 minutes at 45^{D}C.

2.15 DNA sequencing

2.15.1 Exonuclease III deletions of DNA (Henikoff, 1984; Heinrich et al., 1990) (fig. 2.1)

Deletion of DNA with Exonuclease III was used to obtain overlapping fragments for sequencing. The DNA, cloned into Bluescript SK(+) and purified over a CsCl gradient (2.7.6), was cleaved with two restriction enzymes to give a 5' overhanging end adjacent to the target DNA (the deletion site) and a four base 3' overhanging end protecting the primer site (the block site). After ethanol precipitation the DNA was taken up to 0.1μg/μl in 1xExonuclease III buffer. Approximately 3 to 5μg of DNA was used in an experiment. Exonuclease III was added at the concentration of 150units/pmol of susceptible 3' ends; the conversion factor employed was 1μg of 1kb DNA = 3 pmol 3' ends, an equation that takes into account the length of the DNA to be deleted. Aliquots of 1.5μl were removed at 30 second intervals, each *successive pair* being combined together in a single Eppendorf tube, containing 3μl of water, on dry ice. Upon completion of the time course, the enzyme was inactivated by heating at

70^DC for 12 minutes. To each tube was added 20μl of a solution made from 1300μl of S1 nuclease buffer and 200units S1 nuclease (0.5μl); reaction was conducted for 30 minutes at room temperature. The S1 nuclease activity was terminated by the addition of 5μl of S1 nuclease stop buffer to each tube and heating at 70^DC for 10 minutes. Aliquots of 8μl were electrophoresed on an agarose mini-gel to assess the extent of Exonuclease III deletion. If deletion was successful, the DNA ends of each fraction (or those in the desired size range) were repaired by incubation for 10 minutes (37^DC) with 1unit(1μl) of Klenow fragment and 1μl of dNTP mix. The molecules were then ligated overnight (4^DC) by the addition 14μl of water, 20μl of ligation buffer, 3μl of 10mM ATP and 1 unit of T4 DNA ligase. The fractions were electroporated into *E.coli* NM522 and the mini-prep plasmid isolation procedure used to characterize the clones.

Solutions:

1xExonuclease III buffer. 15mM tris HCl pH 8.0/0.66mM MgCl$_2$.

S1 nuclease buffer. 16mM NaAcO pH 4.6/400mM NaCl/1.6mM ZnSO4/ 8% glycerol.

S1 nuclease stop buffer. 0.8M tris pH8.0/20mM EDTA/80mM MgCl$_2$

dNTP mix. dATP/dTTP/dCTP/dGTP at 0.25mM each.

Ligation buffer. 80mM tris HCl pH 7.5/30mM DTT/20mM MgCl$_2$.

2.15.2 Isolation of single-stranded DNA (ssDNA) (fig. 2.1)

Single stranded DNA was obtained from plasmids which were contained in *E.coli* NM522 strains streaked on minimal media. A single colony of the *E.coli* strain was inoculated into 2ml of LB (containing 100μg of ampicillin), followed by 1μl of VCSM13 helper phage (1×10^{13} p.f.u./ml; Stratagene Cloning Systems, San Diego). After incubation at 37^DC for 1.5 hours, 7μl of a 20mg/ml solution of kanamycin was added

and the culture grown up (at 37^{D}C) for a further 16 hours. The culture was transferred to an Eppendorf tube, spun at 4^{D}C for 5 minutes, the supernatant drawn off and subjected to an identical spin in a fresh tube. It was then placed in a tube previously filled with 300µl of 25% PEG/2.5M NaCl, the solutions mixed by inversion and allowed to stand on ice for 45 minutes. The phage were pelleted by spinning for 5 minutes at 4^{D}C, the supernatant poured off and the residial PEG removed by pipetting. The pellet was resuspended in 100µl of 1xTE and 50µl of TE saturated phenol added. The tube was vortexed for 20 seconds, left at room temperature for 5 minutes, vortexed for 20 seconds again, and spun for 10 minutes at room temperature. 80µl of the supernatant was removed and 4µl of 3M sodium acetate (pH 6) and 180µl of ethanol added. After being left on ice for 20 minutes, the solution was spun, the resulting pellet dried in a vacuum desiccator and resuspended in 30µl of 1xTE.

2.15.3 Sequencing

The single-stranded DNA clones were sequenced by the dideoxy chain termination method (Sanger et al., 1977). A combination of both radioactive and fluorescent dye sequencing was used to process the large number of clones. DNA sequences were compiled and analysed by *SEQ*, a group of computer programmes available at the Research School of Biological Sciences, The Australian National University. Other sequencing programmes used were those of Staden (Staden, 1982/1987) and the University of Wisconsin GCG software package (Devereux et al., 1984).

1. Radioactive

Radioactive sequencing (^{32}P) was conducted using the *Sequenase* version 2 solutions, (United States Biochemical, Cleveland, Ohio 44122,

U.S.A.), and the T7 primer (except in the case of the primer extension experiments where the sequence ladders were created by using the oligonucleotides as the primers). Of particular importance in these reactions was the use of a manganese buffer to facilitate the clear reading of sequences close to the primer (Tabor and Richardson, 1989).

An annealing mixture was first made up: ssDNA (4μl - from the 30μl obtained in the procedure outlined in 2.15.2), sequencing buffer* (2μl), primer (1μl), and distilled water (3μl). This was heated for 2 minutes at $65^D C$, and then allowed to cool slowly for 45 minutes until it was approximately $20^D C$. To the annealed mixture was added: 0.1M DTT (1μl), labelling mixture (2μl of a one in five dilution of the labeling mix*), ^{32}P dATP (1μl of a 370MBq/ml 10MCi/ml solution (Amersham)), and T7 DNA polymerase (1μl of a 1 unit/μl solution). The labelling reaction was conducted for 4 minutes at room temperature, after which four 3.5μl aliquots were removed and placed in separate tubes containing 2.5μl of termination mixture* (G, A, T and C). These tubes were left at $37^D C$ for 3-5 minutes, the reactions being finished by the addition of a stop solution* (4μl to each).

* - These solutions were provided in the Sequenase kit.

2. Fluorescent dye (automatic)

Fluorescent dye sequencing was conducted on Applied Biosystems (Foster City, California 94404, U.S.A.) 370A and 373A machines at the John Curtin School of Medical Research and the Research School of Biological Sciences, respectively, The Australian National University. The *universal primer* (chemically modified with the appropriate fluorescent dyes) was used in every reaction.

Stock solutions: (All volumes in µl)

(1) *Annealing stocks:*

Reagent	A	C	G	T
H2O	2.2	2.2	4.4	4.4
5x Taq buffer	1.8	1.8	3.6	3.6
Dye primer (0.4pmol/ml)	1.0 JOE	1.0 FAM	2.0 TAMRA	2.0 ROX
Total	5.0	5.0	10.0	10.0

These volumes were multiplied by the number of clones which were to be sequenced. JOE, FAM, TAMRA, and ROX were names given to the four fluorescent dyes attached to the primers.

5x Taq buffer: 50mM tris pH8.5/ 50mM MgCl2/ 250mM NaCl.

(2) *Extension stocks:*

Reagent	A	C	G	T
d/ddNTP mix	1.1	1.1	2.2	2.2
Taq dilution	2.2	2.2	4.4	4.4
Total	3.3	3.3	6.6	6.6

These volumes were multiplied by the number of clones which were to be sequenced.

Taq dilution: 10 units of Taq DNA polymerase, 2µl of 10x Taq Dilution Buffer, and water to make the final volume 20µl.

10x Taq Dilution Buffer: 500mM KCl/ 100mM tris pH 8.3/ 15mM MgCl2/ 0.1% (w/v) gelatin.

Reaction steps: (All volumes in µl)

(1) *Annealing reaction:*

Reagent	A	C	G	T
Annealing stock	5.0	5.0	10.0	10.0
Template DNA	1.0	1.0	2.0	2.0

In each case, the annealing stock and template DNA were combined together, heated at 65°C for 5 minutes, and allowed to cool slowly for 45 minutes until the temperature was 20°C. The volumes of ssDNA were those taken from the 30μl obtained in the procedure outlined in 2.15.2.

(2) *Extension reaction 1:*

Reagent	A	C	G	T
Extension stock	3.0	3.0	6.0	6.0
Annealed mixture	6.0	6.0	12.0	12.0

In each case, the annealed mixture from the previous step was combined with the appropriate extension stock, and reaction conducted at 70°C for 5 minutes.

(3) *Extension reaction 2:*

A second extension was subsequently conducted, using the Klenow fragment of DNA polymerase 1, for 5 minutes at 37°C. To the A and C reactions from above, 1μl of a 1unit/μl solution of the Klenow enzyme was added, while 2μl of this solution was added to the G and T reactions.

(4) *Ethanol precipitation:*

Following the second extension, the contents of the A,C,G and T reactions were each added directly to a single 140μl solution of 93% ethanol/0.1M sodium acetate pH4.6. After at least 2 hours at -20°C, the solution was spun down and the pellet dried in a vacuum desiccator for 5 minutes. The pellet was stored at -20°C until just prior to loading on the gel, when it was resuspended in a solution of 3μl of 50mM EDTA pH8.0/deionized formamide (1:5 v/v).

5% acrylamide gels were used for both automatic and manual sequencing; they were electrophoresed in 1xTBE buffer. A 5% acrylamide stock was made from the following: urea (378g), 10xTBE (90mls), 40% acrylamide (112.5mls), water (to give a final volume of *900mls*). The 40% acrylamide consisted of acrylamide (38g) and N,N'-methylenebisacrylamide (2g), made up to 100mls with water.

2.16 Primer extension

2.16.1 Germination and growth of plants

Trifolium subterraneum L. cv. Karridale seeds were placed in a small container, immersed in ethanol for 4 minutes, washed with sterile distilled water three times, immersed in 40% hypochlorite for 15 minutes, and finally again washed three times with water. Seeds which were damaged (e.g. had lost their coats) were removed and discarded. Germination and growth of the seedlings was performed in vermiculite filled magenta jars. The magenta jar method involved fitting the bases of magenta jars (GA7 type from the Magenta Corporation, Chicago) containing vermiculate into the necks of other jars filled with sterilized (liquid) F media. In each case a rope wick was used to connect the two jars, thus allowing passage of the nutrient media from the lower reservoir to the upper container. These "piggy-back" jar pairs were sterilized by autoclaving and cooled to room temperature. Approximately 25 seeds were then planted in the vermiculate. The vermiculite, and the tweezers used to handle the seeds, were baked for at least 2 hours prior to use.

Magenta jars carrying seeds were refrigerated (4^DC) for 16-48 hours and then placed in a incubation cabinet - *without light* - for 7 days

(25^DC/80% humidity). Each jar was covered with a loose fitting plastic top to help prevent evaporation. Plants grown for genomic DNA (2.17) were placed in a normal greenhouse.

F media: Composition per litre of water -

10% $CaCl_22H_2O$ (1ml), 12% $MgSO_47H_2O$ (1ml), 10% KH_2PO_4 (1ml), 15% $Na_2HPO_42H_2O$ (1ml), 0.5% Iron citrate (1ml), Gibsons trace elements (1ml). This was adjusted to pH 6.5-6.8.

Gibson's trace elements. Composition per litre of water - H_3BO_3 (2.86g), $MnSO_44H_2O$ (2.03g), $ZnSO_47H_2O$ (0.22g), $CuSO_45H_2O$ (0.08g), H_2MoO4H_2O (0.09g).

2.16.2 Induction

Light induction

The emerged seedlings (7 days old) were exposed to continuous irradiation (Philips TL 40 lamps; 350-400µE/m^2/sec) with white and UV light. Plants were harvested into liquid nitrogen at 0, 3, 6, 12, 15 and 24 hour time points and immediately frozen away at -80^DC.

2.16.3 RNA isolation (de Vries et al., 1988)[*]

A known quantity of plant material was ground up in liquid nitrogen in a precooled mortar and added to a Sarstedt (Germany) 50ml tube, kept at 90^DC, containing four times (w:v) phenol and four times (w:v) extraction buffer. This solution was left, with occasional swirling, for 5 minutes at 90^DC and then mixed by rotation for 5 minutes. Four times (w:v) of chloroform/isoamyl alcohol was added, the resulting solution mixed for a further 10 minutes by rotation, and centrifuged for 1 hour (5,000rpm/room temperature). The aqueous phase was removed, added to a fresh Sarstedt tube together with two times (w:v) chloroform/isoamyl

alcohol, mixed by rotation for 10 minutes, and centrifuged for 15 minutes (5,000rpm/room temperature). The RNA was precipitated by pipetting 1ml aliquots of the aqueous phase into Eppendorf tubes, adding 1/3 volume 8M LiCl to each, and following gentle inversion leaving them overnight at 4^DC. Recovery of the precipitates was achieved by spinning in a microcentrifuge for 1 hour at 4^DC. The pellets were washed with 70% ethanol and resuspended in sterile millipore water. RNA was stored at -80^DC.

Extraction buffer: 100mM tris (pH 9.0)/100mM LiCl)/10mM EDTA/1% SDS.

* All solutions used in the above procedure were made up in diethyl pyrocarbonate (DEPC) treated water, while all equipment and utensils (which could be) were baked overnight at 180^DC.

2.16.4 Extension reactions

For primer extension the primers used were:

CHS2   5'-TGAACTCAATAGTGG-3'

CHS3   5'-TGAACACAGTTGTT-3'

CHS4   5'-CTCGAACTCAAATAAT-3'

(1) Labelling of the oligonucleotide primers

The primers were labelled at their 5' ends with ^{32}P by polynucleotide kinase. Approximately 100ng of oligonucleotide was mixed with 3μl of 10x buffer (600mM tris pH 7.5/80mM $MgCl_2$), 3μl of $\gamma^{32}P$ ATP (Amersham, England) and the necessary amount of water needed to make the total volume 30μl. To this solution was added, at the 0 and 15 minute time points of a 45 minute reaction conducted at 37^DC, 1unit (1μl) aliquots of polynucleotide kinase. The reaction was terminated by

heating at 70^DC for 10 minutes and the solution loaded onto a column equilibrated with 50mM NaCl/1xTE. Unincorporated γ ^{32}P ATP was washed off with 200mM NaCl/1xTE and the oligonucleotide subsequently eluted by the addition of 1M NaCl/1xTE. The glass column, and the glass wool used to plug it, were siliconized (Sambrook et al., 1989). If not used immmediately, the oligonucleotide was frozen away at -20^DC.

(2) Extension (Koes et al., 1989b; van Tunen et al., 1989).

Approximately 50μg of total RNA was combined with the oligonucleotide (~100,000cpm) and concentrated by ethanol precipitation. After being centrifuged at 4^DC, washed with 70% ethanol, and dried by inversion (not in a desiccator) for 45 minutes, the precipitate was resuspended in 20μl of 100mM tris pH8.3/20mM $MgCl_2$/100mM KCl.

Hybridization was carried out from 3 to 16 hours, and was always initiated by heating at 70^DC for 3 minutes. Temperatures for the annealing step were calculated by evaluating the melting temperatures $(Tm(^DC) = 4(G+C) +2(A+T))$ and taking away five degrees from these numbers (Suggs et al., 1981; Mason and Williams, 1985). For the CHS2, CHS3 and CHS4 oligonucleotides, the annealing temperatures were 37^DC, 33^DC and 37^DC, respectively. The solution of annealed DNA-RNA was placed on ice and a 20μl solution of dATP/dTTP/dGTP/dCTP (each 2mM)/ DTT (2mM)/ (RNasin (3 units)/ reverse transcriptase (50 units) added. Oligonucleotides were extended for 30 minutes at 37^DC, and 30 minutes at 42^DC. Products from the extension were recovered by phenol/chloroform (1:1) extraction and ethanol precipitation. The pellet was washed with 70% ethanol, allowed to dry for 30 minutes (again no vacuum), and resuspended in 3μl of 0.1M NaOH/1mM EDTA. After incubation for 10 minutes at room temperature, 7μl of *Sequenase* loading dye was added. The primer extension sample and size markers, which

were provided by extending the unlabelled oligonucleotide on an appropriate single stranded DNA template using standard *Sequenase* reactions (2.15.3), were heated at 95^{D}C for 2 minutes and loaded onto a 5% polyacrylamide gel at the same time.

2.17 Isolation of total genomic DNA (Murray and Thompson, 1980; Taylor and Powell, 1982)

Total DNA was isolated for Southern blot analysis or PCR. The plants were grown for three to six weeks in a greenhouse, harvested into liquid nitrogen, weighed and frozen away at -80^{D}C until use. Plant tissue from which DNA was to be isolated was removed from the freezer and ground until it was a fine powder with a pestle in a mortar filled with liquid nitrogen. The ground tissue was added to an equal volume (w:v) of extraction buffer (2% CTAB/20mM EDTA/1.4M NaCl/100mM tris pH 8.0), the latter contained in a Sarstedt (Germany) 50ml tube at a temperature of 65^{D}C. The solution was incubated for 45 minutes at 65^{D}C, *placed on ice*, and combined with an equal volume of chloroform/isoamyl alcohol (24:1 v/v). It was mixed by rotation for 10 minutes, and then centrifuged for 10minutes (5,000rpm/room temperature). The top (aqueous) layer was drawn off, re-extracted with an equal volume of chloroform/isoamyl alcohol, and centrifuged in an identical manner to above. After the addition of an equal volume of a precipitation buffer (1% CTAB/10mM EDTA/50mM tris pH 8.0) the aqueous solution was allowed to stand at room temperature for at least 30 minutes, and often overnight. The nucleic acid precipitate was collected by centrifugation for 20 minutes (5,000rpm/room temperature), dissolved in 2.5ml of 1M CsCl[*] and carefully layered above a 2ml cushion of 5.7M CsCl[*]; the interface between the two CsCl layers was marked with a black line.

Centrifugation was performed overnight (30,000rpm/18^{D}C), the proteins partitioning into the CsCl above the interface mark, the DNA below the mark, and RNA pelleting at the bottom. These layers were separated by pipetting them off one at a time. The DNA was recovered by adding three volumes of 70% ethanol, leaving the solution at -20^{D}C for 2 hours, and then briefly spinning (6,000rpm/4^{D}C). Pipetting of DNA in the above procedure was always done with a wide mouthed tip.

* The CsCl solutions were made up in 50mM tris pH 8.0/5mM EDTA/50mM NaCl.

2.18 Transient Expression Assays

2.18.1 Construction of CHS-GUS fusions

CHS-GUS fusions were constructed by cloning the respective CHS promoters 5' to the β-glucuronidase gene. The GUS plasmid from which all transient assay constructs used in this study were ultimately derived, called pQ20, had been created by Di Quiggan of the CSIRO Division of Plant Industry. The Nco1-EcoR1 GUS cassette of the Clonetech vector pRAJ275 was ligated into a vector (CW71; a derivative of pUC118; Vieira and Messing, 1987) such that the GUS gene was preceded by approximately 300 nucleotides of a pea vicilin gene promoter (joined to the GUS gene by the Nco1 site), and flanked at its 3' end by the vicilin termination sequence (the pea promoter was necessary for the work of Di Quiggan). All subterranean clover parental GUS constructs were created by cleaving out the vicilin promoter region of pQ20 and replacing it with the CHS promoter sequence; the termination sequence in all cases was that provided by the vicilin gene.

2.18.2 Oligonucleotide-directed mutagenesis of CHS3

The *Oligonucleotide-directed in vitro mutagenesis system* of Amersham (RPN.2322) was used to construct an *Nco*1 site around the translation start codon of CHS3. This procedure is based upon the observation that certain restriction enzymes cannot cleave phosphorothioate DNA. The mutant strand is "labelled" by the incorporation of a thionucleotide and single strand nicks generated in the non-mutant strand by the action of one of the above enzymes (e.g. *Pvu*1). Exonuclease III is used to digest away part of the nicked strand; the mutant strand can then be used as a template to create a double stranded mutant molecule. The protocol was that supplied by Amersham.

The oligonucleotide used was:

<u>CHS3</u> AGATACACTCACC**ATGG**CTTAATATATCCTTTA

(Nucleotides differing from the wild type CHS sequence (and hence necessary to create the *Nco*1 site) are shown in bold type.

2.18.3 *Nicotiana plumbaginifolia* cell suspension culture (Lorz et al., 1983)

The media, called CSV, which the *Nicotiana plumbaginifolia* cell suspension culture was maintained in, consisted of the following:

CSV:
```
1.  CSV Macro          50ml
2.  CSV Micro           1ml
3.  CSV Iron            1ml
4.  CSV Organic         1ml
5.  MgSO4 7H2O         50ml
6.  CaCl2 2H2O         50ml
7.  Sucrose            30g
8.  Inositol            1g
9.  2,4-D (5mg/μl)    400μl
10. Kinetin (1μg/ml)   50μl
```
Made up to 1 litre with water   pH CSV – 5.8.

Composition of stock solutions:

CSV Macro (per litre): NH_4NO_3 (24.8g), KNO_3 (50.1g), $(NH_4)H_2PO_4$ (9.2g).

CSV Micro (per litre): $MnSO_44H_2O$ (13.0g), H_3BO_3 (5.0g), $ZnSO_47H_2O$ (1.0g), KI (1.0g), $CuSO_45H_2O$ (0.2g), $Na_2MoO_42H_2O$ (0.1g), $CoCl_26H_2O$ (0.1g).

CSV Iron (per litre): Na_2EDTA (20g), $FeSO_47H_2O$ (15.0g).

CSV Organic (per litre): Thiamine HCl (5g), Nicotinic acid (5g), Pyridoxine (0.5g).

The cell suspension culture was grown in continuous light at 25^DC with slow shaking. Cells were usually subcultured twice a week on Mondays and Thursdays by placing 5mls of the old culture into 30-35ml of new medium. Four separate flasks were maintained to prevent the loss of the entire culture if infection occurred.

2.18.4 Isolation and electroporation of *Nicotiana plumbaginifolia* protoplasts (Last et al., 1991).

Approximately 20ml of cell culture was taken from a flask (or flasks) on the day of a subculture and spun down (1,000rpm/10 minutes) in a Sarstedt (Germany) 50ml tube. The pellet was resuspended in 40mls of 1:1 ASW:0.6M mannitol, the cells spun down as before, and taken up in a 25ml solution of 1:1 ASW:0.6M mannitol containing 1.4% cellulysin and 0.5% macerase. Digestion of the cell walls was conducted in the dark for approximately 5-6 hours in a petri dish at 28^DC.

After digestion, the protoplasts were spun down (1,000rpm/10 minutes), washed with 40ml of 1:1 ASW:0.6M mannitol, spun down again and resuspended in 30ml of 1:1 ASW:0.6M mannitol. Incubation was conducted in this medium for 1-2 hours during which the viability of the protoplasts was ascertained by removing 1ml and adding the dye 5 (and-6) carboxyfluorescein diacetate. A haemocytometer and microscope were used to count the viable cells.

For electroporation the protoplasts were spun down (1,000rpm/10 minutes), washed with 30-40ml of 1:1 ASW:0.6M mannitol, spun down again and resuspended in TBS such that the concentration was approximately 1 million/500µl of solution. Protoplasts, usually 400µl, were placed in 4mm gap electroporation (Biorad) cuvettes together with 30µg of CsCl purified (2.7.6) DNA and kept on ice for 2 minutes prior to electroporation. After electroporation, conducted with the Bio-Rad Gene-Pulser, the protoplasts were placed in 2ml of CSV-M and incubated in the dark until for the required time period; their viability was always determined prior to assay.

ASW: (artificial sea water) Composition per litre: NaCl (18.2g), KCl (0.52g), MgSO4 7H2O (4.64g), MgCl2 6H2O (3.39g), CaCl2 2H2O (1.0g), NaHCO3 (0.147g), MES (2.13g). ASW was usually made up 1:1 with 0.6M mannitol (i.e. mannitol 0.3M in final solution). pH 5.8

TBS: 30mM tris, 150mM NaCl, 6mM CaCl2, 253mM mannitol, pH 9.0.

CSV-M was the same as CSV except it contained: (1) 0.4M mannitol and 1% sucrose; (2) 1ml/litre of K Organic Acids; (3) 1ml/litre of K Sugars and Sugar Alcohols.

K Organic Acids (per litre) - sodium pyruvate (0.5g), citric acid (1g), malic acid (1g), fumaric acid (1g).
K Sugars and Sugar Alcohols (per litre) - sucrose (12.5g), fructose (12.5g), xylose (12.5g), rhamnose (12.5g), sorbitol (12.5), mannitol (12.5g).

2.18.5 Fluorometric GUS assay (Jefferson et al., 1987; Jefferson, 1987).

Protoplasts were pipetted into a Falcon tube and spun down (1,000rpm/15 minutes). The supernatant was removed, the protoplasts resuspended in 250µl of lysis buffer, sonicated for 5 seconds and then centrifuged for 10 minutes (12,000rpm). Protein concentration was

determined by removing 10μl of the supernatant and adding to it 790μl of sterile distilled water and 200μl of the Bio-Rad Protein Assay dye reagent concentrate, the latter itself diluted 1:4 with water; optical density was measured at 595nm and a standard curve used to estimate concentration. Usually 100μg of protein was used in each assay. A number of Eppendorf tubes were each filled with the required amount of supernatant and made up to 100μl with water. To each was added 100μl of assay buffer. A zero time point and usually two other time points were taken; the final time point was rarely taken after 1.5 hours. In each case the reaction was stopped by the addition of 1ml of 0.2M Na_2CO_3. A Hitachi F-3000 Fluorescence Spectrophotometer was used to measure fluorescence (excitation 365nm; emission 455nm; excitation bandpass 5nm; emission bandpass 10nm). A standard graph of fluorescence against concentration of 4-methyl umbelliferone (MU) was used to the calculate nM concentration of product in each of the reactions.

That this assay was working was evaluated by making crude protein extracts from a transgenic tobacco plant containing the cauliflower mosaic virus promoter fused to the GUS gene and analysing these for the presence of the β-glucuronidase enzyme. On all occasions when this control was performed, the GUS assay was observed to be working. The tobacco plant, called 1103 "smiley", was kindly provided by Dr. Phil Larkin, C.S.I.R.O. Division of Plant Industry, Canberra.

lysis buffer: 50mM NaPO4, 10mM EDTA, 0.1% triton X100, 0.1% sarkosyl, 10mM β-mercaptoethanol, pH 7.0.

assay buffer: To 20ml of the above lysis buffer was added 7mg (1mM) of 4-methylumbelliferyl β-D-glucuronide (MUG).

The β-mercaptoethanol was only added to these buffers just before their use - 7μl of 14.4M stock per 10ml of lysis/assay buffer.
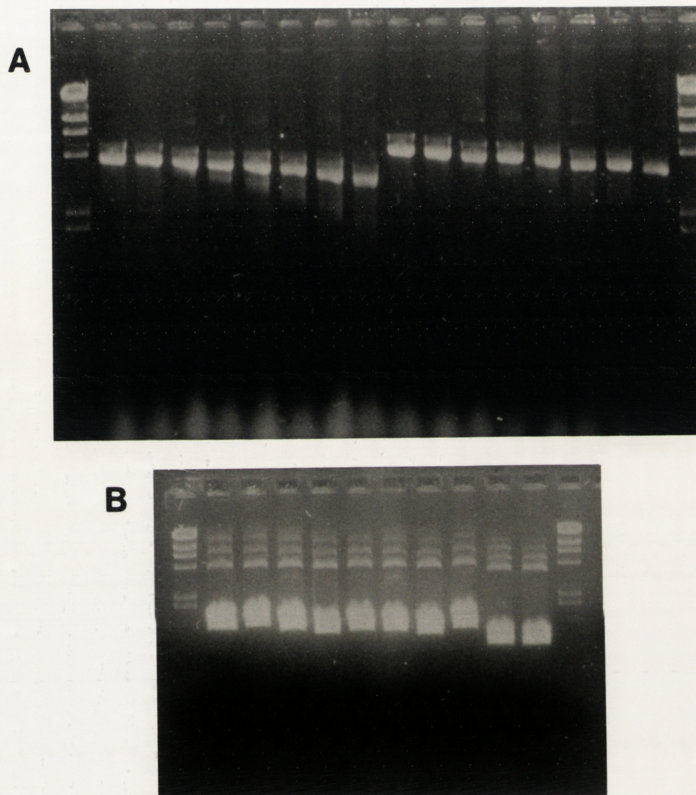
Fig. 2.1

A. Examples of two Exonuclease 111 deletion series into a linearized DNA molecule. Fractions were taken at 30 second intervals and approximately one third of each electrophoresed on a 0.8% agarose gel (shown). The outer lanes contain λ/*Hin*d111 standard markers.

B. Examples of ssDNA produced from ten sequencing clones. In each case one fifteenth of the DNA was electrophoresed on a 0.8% agarose gel (shown). The outer lanes contain λ/*Hin*d111 standard markers.

Chapter 3 Genomic organization of the clover CHS genes

3.1 Introduction

The cloning of a *Petroselinum hortense* CHS cDNA (Kreuzaler et al., 1983; Reimold et al., 1983) has led, both directly and indirectly, to an explosion in our knowledge of the molecular structure of the CHS genes in plants. The techniques of hybrid-arrested and hybrid-selected translation were used to verify that this UV irradiation-induced cDNA was specifically derived from CHS mRNA. It has since been used as a hybridization probe to identify the (genomic) CHS genes in a number of species, including *Antirrhinum majus* (Wienand et al., 1982), *Petunia hybrida* (Reif et al., 1985), bean (Ryder et al., 1984), and soybean (Wingender et al., 1989). In turn, the CHS genes of pea have been isolated by probing with the heterologous bean CHS cDNA (Harker et al., 1990). The CHS gene of *Zea mays* (maize), however, has been cloned by making use of a transposable element as a "gene tag" (Wienand et al., 1986).

This chapter describes the isolation and characterization of the CHS genes of *Trifolium subterraneum* L. (subterranean clover) cv. Karridale. The near full length cDNA of the bean CHS1 gene (Ryder et al., 1984; Ryder et al., 1987) was employed as a hybridization probe to detect and isolate the genes from a genomic gene library of the plant. The high degree of homology between the CHS genes of different plant species (Niesbach-Klosgen et al., 1987), and the success of the previous studies using CHS sequences from other species as probes, meant that it was not expected that there would be any problems in using this cDNA as a probe in conditions of low stringency.

The gene library of clover had been constructed in the vector Lambda Dash 11 by T. Arioli. Genomic DNA, partially digested with *Sau*3A to

give fragments in the size range of 15-20kb, was ligated into the *Bam*H1
site of the vector (fig. 3.1). Such large fragments of genomic DNA would
not only contain the entire sequence of any gene, but also more than one
member of any clustered CHS multigene family. After cloning, the
inserts could be cleaved out of the phage vector, in entirety, with the
enzymes *Eco*R1, *Not1, Sal*1 and *Xba*1, the restriction sites for which
remained in the adjacent flanking vector DNA at both ends. Restriction
mapping was performed with the aid of a computer programme that
allowed the sizes of unknown fragments to be determined with the use of
the known lengths of λ size standards.
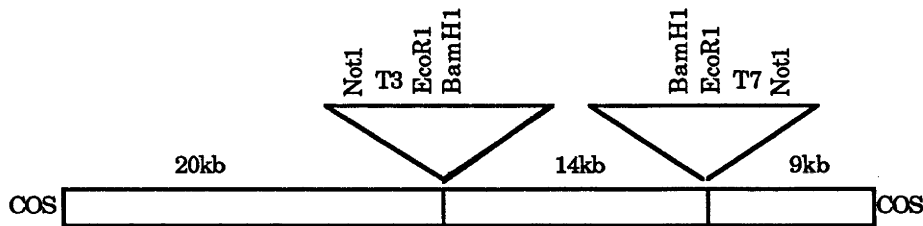
Fig. 3.1 **Map of the vector Lambda Dash 11.** The gene library of *Trifolium
subterraneum* was constructed in this vector by cloning partial *Sau*3A DNA fragments
into the *Bam*H1 site; the internal 14kb was lost in this procedure. T7 and T3 refer to the
two phage promoters, and COS to the *cos* sequence of λ.
(modified from the Stratagene Cloning Systems 1990 Product Catalog)

3.2 Results

3.2.1 Restriction mapping of the λTsCHS clones

The *initial* probing of the bean CHS1 cDNA onto part of the amplified
clover gene library, spread out on 30 large plates (2.10.1), revealed
approximately 200 hybridization signals. Of these, 150 were stocked in
phage storage buffer as a "library" of (inpure) clover clones containing
potential CHS genes. Plaque purification was performed for 12 members
of this "library", and phage storage stocks made of these λTsCHS clones.

DNA was isolated from the 12 clones λTsCHS1.1, λTsCHS1.2, λTsCHS1.3, λTsCHS1.4, λTsCHS1.5, λTsCHS1.6, λTsCHS4.1, λTsCHS4.3, λTsCHS8.3, λTsCHS8.4, λTsCHS26.2, and λTsCHS26.4. The plaques of ten of these showed intense hybridization to the bean CHS1 cDNA probe, while those of the remaining two (λTsCHS1.6 and λTsCHS26.4) hybridized only weakly. It was reasoned that the clones which hybridized strongly were likely to contain an entire CHS gene, or perhaps a cluster of such genes, while the poorly hybridizing clones would be characterized by incomplete genes or only single isolated genes. In each case, the phage DNA was restricted with the enzymes *Eco*R1, *Hin*d111 and *Bam*H1 to gain a general picture of its profile. Great difficulty was encountered in trying to digest these clones. Although the use of Qiagen columns to clean the DNA improved enzyme digestion markedly, it was never entirely satisfactory. This was possibly due to proteins remaining tightly bound to the λ DNA. The restriction digests were blotted and probed with the bean CHS1 cDNA to confirm the existence of sequences homologous to this clone.

The inserts in the 12 λTsCHS clones were all cloned into the vector Bluescript in order to enable a more thorough restriction analysis to be performed, free of the problems of digesting the phage DNA. This was achieved by making use of the *Not*1 sites flanking the inserts (fig. 3.1). The recognition sequence of *Not*1 is 8bp long, and assuming a random distribution of nucleotides within DNA, should occur approximately every 66kb. Each of the 12 λ clones was determined to lack a *Not*1 sequence; they were therefore cleaved with *Not*1 and ligated into the *Not*1 site of Bluescript (hence called "pTsCHS1.2" etc). Plasmid clones for the 12 genomic fragments were isolated and subjected to single and double digestions with *Eco*R1, *Hin*d111, and *Bam*H1. The gels of these digests

were, as above, blotted and probed with the bean CHS1 cDNA probe (fig. 3.2).

The digestion and probings of the phage and plasmid clones allowed an analysis to be made of whether any two or more possessed similar restriction bands, suggesting either the presence of overlapping fragments, or even identical fragments, from the same region of the genomic DNA. It was found that a large number of the clones were in fact identical, with only seven distinct genomic fragments being present in the 12 clones. The seven fragment groups were as follows - in each case the clone which will from now on designate the fragment is listed in italics first: (1) *λTsCHS1.2*/λTsCHS8.4; (2) *λTsCHS4.3*/λTsCHS1.4; (3) *λTsCHS4.1*/λTsCHS1.1/λTsCHS1.3/λTsCHS8.3; (4) *λTsCHS26.4*; (5) *λTsCHS1.5*; (6) *λTsCHS26.2*; (7) *λTsCHS1.6*. The large number of identical clones must result from the amplification of the gene library.

Restriction maps for the enzymes *Eco*R1, *Hind*111, and *Bam*H1 were established in six of the pTsCHS clones (fig. 3.3). The regions of hybridization within these clones, and hence the positions of the (putative) CHS gene(s), were established from the bean CHS1 cDNA probing. The seventh clone, pTsCHS1.6, was discovered to have such a complicated digestion pattern with these three enzymes, both alone and in double digests, that no attempt was made to map all their restriction sites within the fragment. However, it was established that the region of hybridization to the bean probe was confined to a 0.8kb *Hind*111-*Bam*H1/*Eco*R1 fragment at one end of the clone (fig 3.3). As the sequences of the CHS genes from other species (excluding introns) are all approximately 1400bp, it was concluded that a clover CHS gene was incompletely represented on this clone. The DNA fragment of this clone, and that of pTsCHS26.4, contained the smallest regions of hybridization
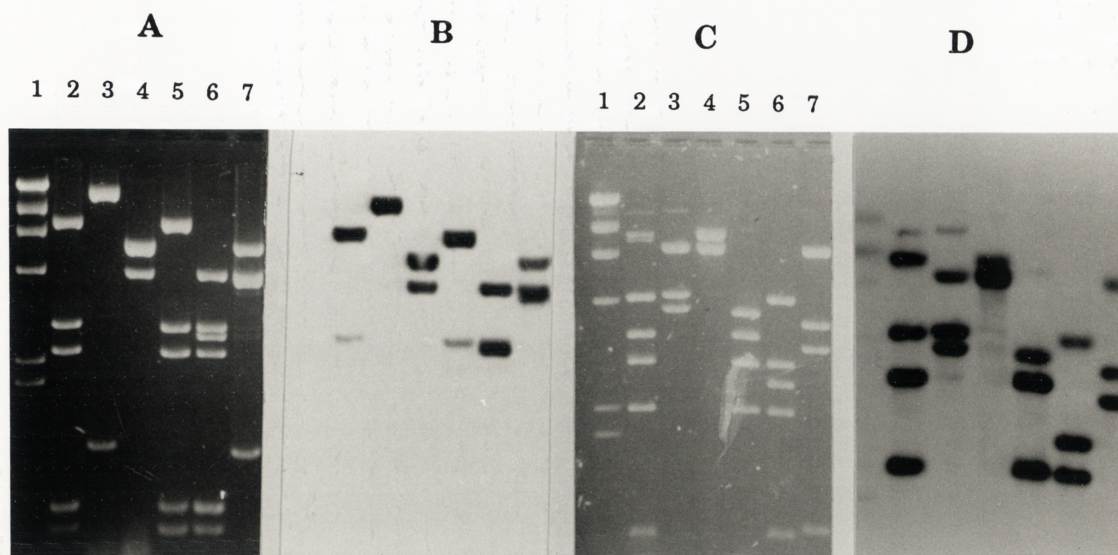
Fig. 3.2

Examples of the *initial* restriction mapping and probing of the pTsCHS clones (λ Dash 11 fragments in Bluescript SK(+)). Three restriction enzymes were used : *Eco*R1, *Bam*H1 and *Hin*d111.

A

<u>*pTsCHS4.3*</u>

Lane 1: λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Lanes 2,3,4,5,6, and 7: Restriction of pTsCHS4.3 with *Eco*R1, *Hin*d111, *Bam*H1, *Eco*R1/*Hin*d111, *Eco*R1/*Bam*H1, and *Bam*H1/*Hin*d111 respectively.

C

<u>*pTsCHS26.2*</u>

Lane 1: λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0)

Lanes 2,3,4,5,6, and 7: Restriction of pTsCHS26.2 with *Eco*R1, *Hin*d111, *Bam*H1, *Eco*R1/*Hin*d111, *Eco*R1/*Bam*H1 and *Bam*H1/*Hin*d111 respectively.

B and **D**

Hybridization of the bean CHS1 cDNA probe to the Southern transfers of the gels in A and C respectively.

λTsCHS 1.2/8.4

λTsCHS 4.3/1.4

λTsCHS 4.1/1.1/1.3/8.3

λTsCHS 26.4

λTsCHS 1.5

λTsCHS 26.2
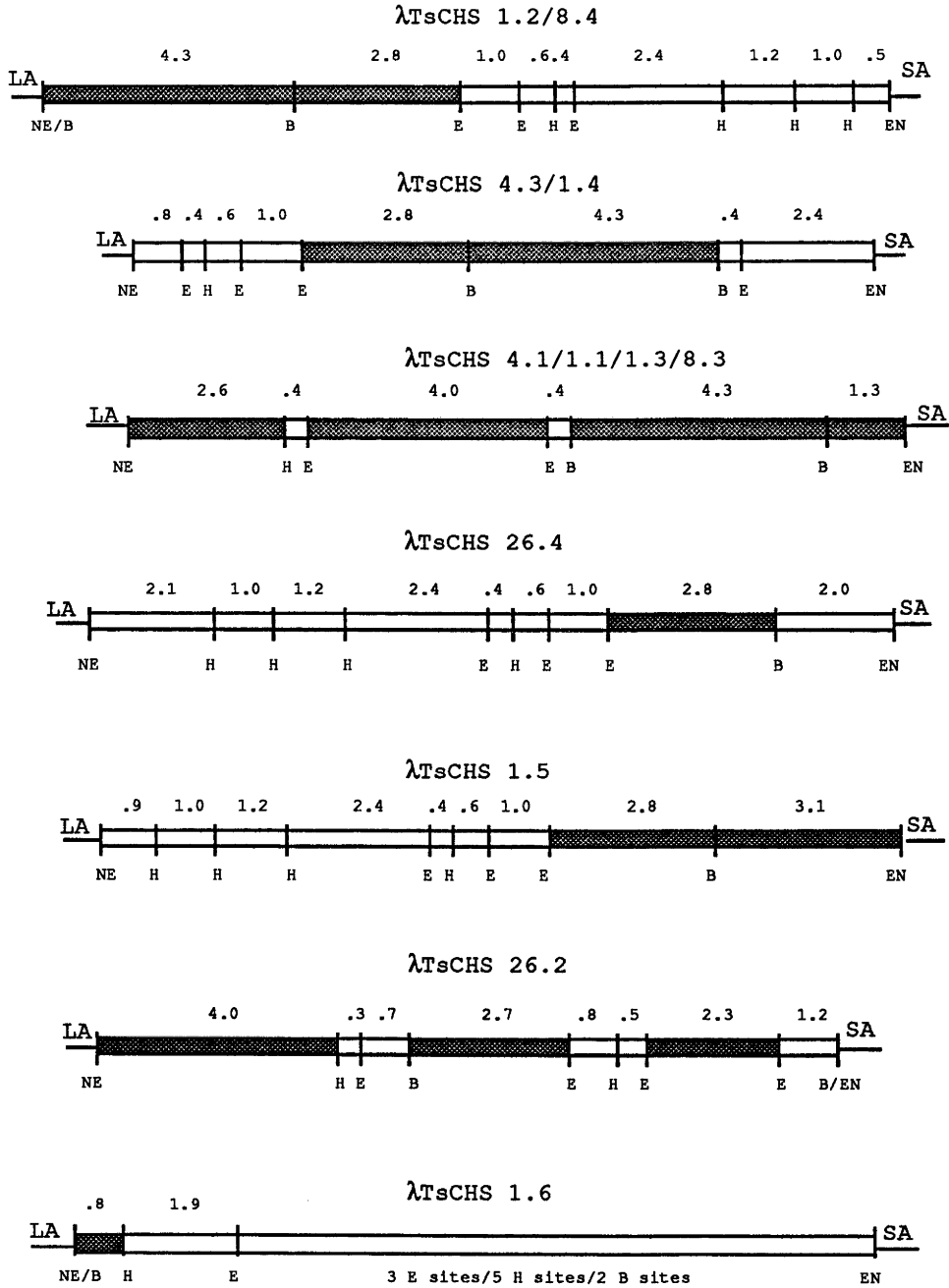
λTsCHS 1.6

Fig. 3.3

Simple restriction maps for the seven subterranean clover DNA fragments λTsCHS1.2 (8.4), λTsCHS4.3 (1.4), λTsCHS4.1 (1.1, 1.3, 8.3), λTsCHS26.4, λTsCHS1.5, λTsCHS26.2, and λTsCHS1.6. The fragments are represented by oblong boxes. Their orientations within the λ Dash 11 vector clones, from which the Bluescript clones were derived, are indicated by LA and SA, referring to the long arm and short arms of the phage, respectively. Regions of hybridization to the bean CHS1 cDNA probe are indicated by shading. Restriction enzymes: E - EcoR1, H - Hind111, and B - BamH1. In the case of λTsCHS1.6, the order and positions of many of the restriction sites for these enzymes were not established.

to the bean probe of the seven clones, providing the expected explanation for their faint hybridization during plaque purification (see above).

The initial restriction maps (fig. 3.3) indicated that the five fragments λTsCHS1.2, λTsCHS4.3, λTsCHS4.1, λTsCHS26.4, and λTsCHS1.5 probably represented different *Sau*3A digests of the same region of DNA. A 4.3kb *Bam*H1 fragment was found to be a common feature of λTsCHS1.2, λTsCHS4.3, and λTsCHS4.1, and in each case it hybridized to the bean CHS1 cDNA probe. Furthermore, a 2.8kb *Eco*R1-*Bam*H1 fragment, showing hybridization to the probe, was discovered to lie adjacent to the 4.3kb *Bam*H1 fragments of λTsCHS1.2 and λTsCHS4.3. Seemingly "isolated" 2.8kb *Eco*R1-*Bam*H1 fragments were also found to be present in λTsCHS26.4 and λTsCHS1.5, two clones which were obviously only slight variations on each other (fig. 3.3). Comparison of the restriction profiles of the regions adjacent to λTsCHS26.4 and λTsCHS1.5 on the one hand, and λTsCHS1.2 on the other, suggested that the 4.3kb *Bam*H1 fragment had been cleaved in two by *Sau*3A, at different positions, to create one end of each of the former two clones.

In order to confirm that these five clones were indeed overlapping, two *non*-hybridizing fragments, the 2.4kb *Eco*R1-*Hin*d111 of λTsCHS1.2 and the 2.4kb *Eco*R1 of λTsCHS4.3, were cloned into Bluescript and their amplified PCR products hybridized onto blots of appropriate restriction digests: the *Eco*R1-*Hin*d111 fragment was hybridized to a blot of the *Eco*R1/*Hin*d111 digests of each of pTsCHS1.2, pTsCHS4.3, and pTsCHS26.4 (acting also for the similar clone pTsCHS1.5), while the *Eco*R1 fragment was used to probe the *Eco*R1-*Bam*H1 digests of pTsCHS4.1 and pTsCHS4.3. As expected from the restriction maps, the *Eco*R1-*Hin*d111 probe hybridized to the 2.4kb *Eco*R1-*Hin*d111 fragments of pTsCHS1.2 and pTsCHS26.4, and to the 0.8kb *Eco*R1 fragment of pTsCHS4.3 (fig. 3.4). However, although the *Eco*R1 probe hybridized to

the 2.4kb and 4.0kb *Eco*R1 fragments of pTsCHS4.3 and pTsCHS4.1 respectively, it was also observed to hybridize faintly to the 4.3kb *Bam*H1, 2.8kb *Eco*R1-*Bam*H1, and 3.0kb *Eco*R1 fragments (fig. 3.5). This anomaly, implying homologous sequences between the five fragments, of which only the latter four hybridized to the bean probe, was ultimately resolved by sequencing (chapter 4), but a determination of the number of CHS genes present in this region of DNA allowed a guess at its solution to be made (see below). Hence, together the five fragments, λTsCHS1.2, λTsCHS4.3, λTsCHS4.1, λTsCHS26.4, and λTsCHS1.5 were established to represent 23.2kb of contiguous clover DNA. The two remaining clones, λTsCHS26.2 and λTsCHS1.6, could neither be matched to this region of clover DNA, nor to each other. The region represented in the five clones is referred to as "I", while that contained in λTsCHS26.2 (which possessed extensive hybridization to the probe) as "II".

The overlapping nature of the five region I fragments, and their respective hybridization patterns to the bean CHS1 cDNA probe, made it possible to narrow down the regions of hybridization within the 4.3kb *Bam*H1 and the 4.0kb *Eco*R1 fragments. As mentioned above, the right hand ends of λTsCHS26.4 and λTsCHS1.5 (as drawn in fig.3.3) must have been produced by *Sau*3A cleaving at two recognition sites within the 4.3kb *Bam*H1 fragment. As no evidence was obtained that the 2.0kb of the 4.3kb *Bam*H1 fragment surviving in λTsCHS26.4 hybridized to the bean CHS1 cDNA probe (fig.3.2), it could be concluded that the region of this fragment which was homologous to the probe was located in the 2.3kb of DNA furthest from the adjoining 2.8kb *Eco*R1-*Bam*H1 fragment. This was supported by the hybridization of the 3.1kb *Bam*H1-*Eco*R1 fragment of λTsCHS1.5; it was reasoned that this fragment must contain at least part of the hybridizing sequence of the 4.3kb *Bam*H1 fragment. In an analogous fashion, comparison of λTsCHS4.3 and λTsCHS4.1

A   **B**

1 2 3 4
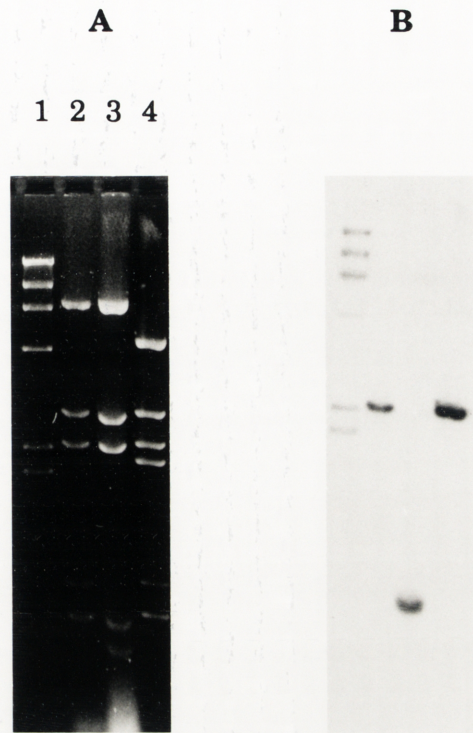
Fig. 3.4

Establishing the overlapping nature of the five genomic clones λTsCHS1.2, λTsCHS4.1, λTsCHS4.3, λTsCHS26.4 and λTsCHS1.5 (1). The 2.4kb *Eco*R1-*Hin*d111 fragment of λTsCHS1.2 was cloned, labelled as a probe, and hybridized onto a Southern transfer of the DNA from pTsCHS1.2, pTsCHS4.3, and pTsCHS26.4 (acting also for TsCHS1.5), each cleaved with *Eco*R1/*Hin*d111. Hybridization of the probe to a (predicted) band in each digestion proved the clones were from the same region of subterranean clover DNA.

A

Lane 1: λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Lanes 2, 3, and 4: Restriction of pTsCHS1.2, pTsCHS4.3, and pTsCHS26.4, respectively, with *Eco*R1/*Hin*d111.

B

Hybridization of the 2.4kb *Eco*R1-*Hin*d111 fragment to the Southern transfer of the gel in A.

The hybridizing fragments are: Lane 1 - 2.4kb. Lane 2 - 0.8kb. Lane 3 - 2.4kb.

A   **B**

1 2 3
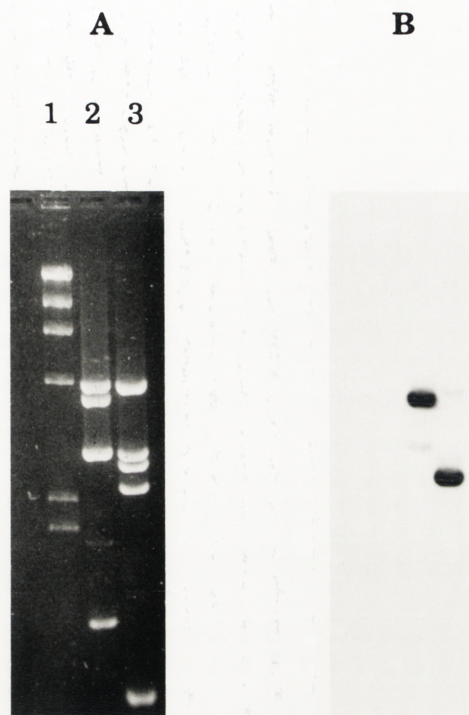
Fig. 3.5

Establishing the overlapping nature of the five genomic clones λTsCHS1.2, λTsCHS4.1, λTsCHS4.3, λTsCHS26.4 and λTsCHS1.5 **(2)**. The 2.4kb *Eco*R1 fragment of λTsCHS4.3 was cloned, labelled as a probe, and hybridized onto a Southern transfer of the DNA from pTsCHS4.1 and pTsCHS4.3, each cleaved with *Eco*R1/*Bam*H1. Hybridization of the probe to the 2.4kb *Eco*R1 and 4.0kb *Eco*R1 fragments of pTsCHS4.3 and pTsCHS4.1 was as expected. However, the fainter hybridization to the 4.3kb *Bam*H1, 2.8kb *Eco*R1-*Bam*H1, and 3.0kb *Eco*R1 fragments was not.

A

Lane 1: λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Lanes 2 and 3: Restriction of pTsCHS4.1 and pTsCHS4.3, respectively, with *Eco*R1/*Bam*H1.

B

Hybridization of the 2.4kb *Eco*R1 fragment to the Southern transfer of the gel in A.

demonstrated that the bean probe did not recognize the 2.4kb of the 4.0kb *Eco*R1 fragment of λTsCHS4.1 situated closest to the 4.3kb *Bam*H1 fragment. The hybridizing region of the 4.0kb *Eco*R1 fragment was thus established to lie in the 1.6kb not present in λTsCHS4.3.

The extent of hybridization to the bean CHS1 cDNA as revealed by the restriction mapping and probing made it highly likely that there were a number of CHS genes present in the large region of DNA covered by the five overlapping clones. For example, intense hybridization was present thoroughout virtually the entire 13kb of λTsCHS4.1 (figs. 3.3). On top of this, most of the separate fragment λTsCHS26.2 hybridized to the probe. To gain a general idea of the number of CHS genes present, use was made of the fact that all CHS genes, for which the genomic sequences are known, possess a conserved intron approximately 180bp from the 5' end of the *translated* region (Schroder and Schroder, 1990). By using a published restriction map of the bean CHS1 cDNA (Ryder et al., 1984), it was possible to cleave this DNA to generate a probe specific for the second exon (i.e. that to the 3' of the intron). Restriction of an available Bluescript SK(+) clone of the bean CHS1 cDNA with *Pst*1 generated two fragments, one of which, 0.7kb in length, would be expected to be specific for any genomic sequences which lie in the second exon of a CHS gene (assuming only one intron in the genomic gene). The restriction map of the bean cDNA did not, however, suggest any DNA sequence situated entirely to the 5' of the intron that could be easily isolated.

The clones pTsCHS1.2, pTsCHS4.3, pTsCHS4.1, pTsCHS26.4, and pTsCHS1.5 were cleaved with *Eco*R1/*Bam*H1, pTsCHS26.2 with *Eco*R1/*Hin*d111/*Bam*H1, and pTsCHS1.6 with *Eco*R1/*Hin*d111, electrophoresed, blotted and probed with the 0.7kb *Pst*1 fragment. Restriction of the Bluescript clones with these respective enzyme combinations liberated those fragments that showed hybridization to the

bean CHS1 cDNA probe, but were *separated* from each other by regions of non-hybridization (fig. 3.3; taking into account the information presented above that the 2.8kb *Eco*R1-*Bam*H1 fragment was separated by a region of non-hybridization from the hybridizing region within the 4.3kb *Bam*H1 fragment). If the regions of non-hybridization reflected introns, then it would be expected that some of the fragments generated by these digestions showing hybridization to the entire bean cDNA would *not* hybridize to the 0.7kb 3' probe.

The hybridization patterns were found to be identical to those of the clones when probed with the entire bean CHS1 cDNA. This data strongly *implied* that each of the 2.8kb *Eco*R1-*Bam*H1, 4.3kb *Bam*H1, 4.0kb *Eco*R1, and 3.0kb *Eco*R1 fragments of region I, and the 4.0kb *Eco*R1-*Hind*111, 2.7kb *Eco*R1-*Bam*H1, and 2.3kb *Eco*R1 fragments of region II, as well as the sequence in pTsCHS1.6, contained the 3' sequence of a CHS gene (i.e. a sequence to the 3' of the expected intron), and as a consequence there were eight CHS genes in the characterized genomic clones. It also offered a solution to the unexpected phenomenon of the 2.4kb *Eco*R1 fragment of λTsCHS4.3, which did not hybridize to the bean probe, itself hybridizing to all these fragments. As the bean CHS1 probe was a cDNA, it was devoid of all promoter sequences and most of that 3' to the gene. Thus, conceivably the 2.4kb *Eco*R1 fragment contained such sequence (from the gene present in the 4.0kb *Eco*R1 fragment) which was able to hybridize to the analogous regions of the other genes. Nucleic acid sequencing (chapter 4) would establish this to be the case.

3.2.2 Restriction mapping of the CHS genes

In order to obtain more advanced restriction maps of the various *Eco*R1/*Hind*111/*Bam*H1 fragments showing hybridization to the bean CHS1 cDNA probe (3.2.1), they were separately cloned into Bluescript.

Digestion of the clones with a range of 10 enzymes (*Eco*R1, *Hind*111, *Bam*H1, *Pst*1, *Sac*1, *Bgl*11, *Acc*1, *Xba*1, *Kpn*1, and *Cla*1), and *then* double and triple digestions with those enzymes established to have restriction sites (examples in figs. 3.6 and 3.7), provided the data to construct comprehensive restriction maps of these seven fragments. The parent genomic clones were also digested with these enzymes to confirm the order and number of sites. Restriction maps for the two regions of clover DNA are presented in fig. 3.8.

If each isolated hybridizing area contained a separate CHS gene, it would be expected that this would be apparent in their respective restriction maps. In bean, for example, a *Kpn*1 site has been found in all the CHS genes (Lawton et al., 1990). Upon comparing the restriction maps of the seven hybridizing regions (fig. 3.8), the only potentially common site appeared to be that of *Pst*1. All of the separate areas of hybridization to the bean probe were found to possess a *Pst*1 site, and in most cases this was flanked by a *Sac*1 site on one side (approximately 300bp distant), and by a *Xba*1 site on the other side (approximately 600bp distant). The stretches of DNA flanking the *Xba*1-*Sac*1 fragments, although having no restriction enzyme profile similarities, nevertheless always faintly hybridized to the probe. In the case of the 3.0kb *Eco*R1 fragment of region I, because the area of intense hybridization was contained at one end of the clover DNA covered by the five genomic clones constituting this region, it was possible that the absence of an *Xba*1 site in this fragment was due to the sequence of the potential CHS gene not being fully present. Other restriction enzyme similarities between the seven hybridizing areas, such as the frequent occurrence of *Acc*1 sites adjacent to the *Pst*1 sites, were also evident.

Hence, prior to nucleic acid sequencing, an overwhelming amount of evidence suggested that four CHS genes were located in region I, and

a further three more in region II. *Sequencing confirmed this to be the case* (chapter 4). The seven genes were named *CHS1, 2, 3, 4, 5, 6,* and *7* respectively. The two "regions" of clover DNA are from henceforth called cluster I and cluster II. The incompletely represented gene found in λTsCHS1.6, which did not show any obvious relationship to these other CHS genes, was named *CHS8*.

One interesting feature to arise at this point concerned the most left hand (as drawn in fig. 3.8) gene present on λTsCHS26.2, CHS7. As the three hybridizing regions of λTsCHS26.2 were all mapped to be entirely *within EcoR1* fragments (fig. 3.8; if the flanking phage *EcoR1* sites, lying adjacent to the *Not1* sites, are taken into account), it would have been expected that digestion with *EcoR1* would produce three bands with equivalent intensities of hybridization to the bean CHS1 cDNA probe. However, the initial *EcoR1/Hind111/BamH1* mapping gel of λTsCHS26.2 (fig. 3.2) did not suggest that the three *EcoR1* fragments contained identical hybridizing regions. One *EcoR1* fragment (that on the left in fig. 3.8) was observed to produce fainter hybridization than the other two (fig. 3.2: C and D, lane 2; note that the largest band is undigested DNA). The possible incomplete nature of CHS7 was confirmed by sequencing (4.2.6).
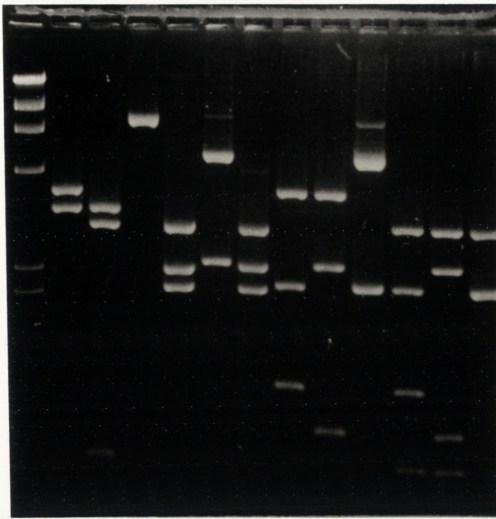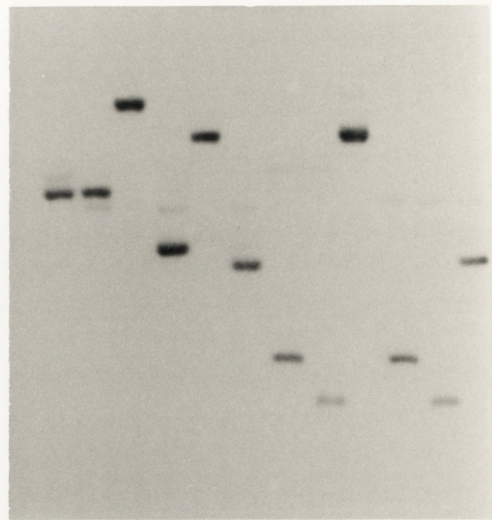
Fig. 3.6

Example of the detailed restriction mapping and probing of the (sub) fragments from within the subterranean clover genomic clones.

The 4.3kb *Bam*H1 fragment. Digestion and probing of clone pTs11Z (the 4.3kb fragment in Bluescript).

A

Lane 1: λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Lanes 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13: Restriction of pTs11Z with the following enzymes - *Xba*1, *Xba*1/*Bam*H1, *Bgl*11, *Bgl*11/*Bam*H1, *Sac*1, *Sac*1/*Bam*H1, *Xba*1/*Bgl*11, *Xba*1/*Sac*1, *Bgl*11/*Sac*1, *Bam*H1/*Xba*1/*Bgl*11, *Bam*H1/*Xba*1/*Sac*1, and *Bam*H1/*Bgl*11/*Sac*1 - respectively.

B

Hybridization of the bean CHS1 cDNA probe to the Southern transfer of the gel in A.

In order to verify the results of A and B, a Bluescript clone carrying the 4.3kb fragment in the reverse orientation to pTs11Z was also digested and probed in an identical manner (not shown). The faint hybridization of the probe to the 0.3kb *Xba*1 fragment within the 4.3kb *Bam*H1 fragment was established by a longer exposure of the autoradiograph shown in B.
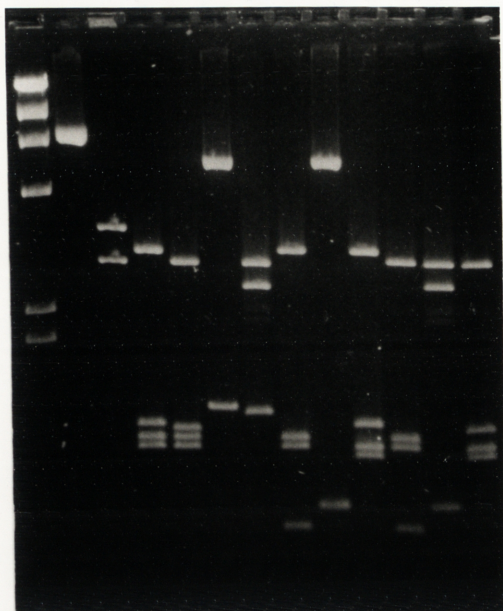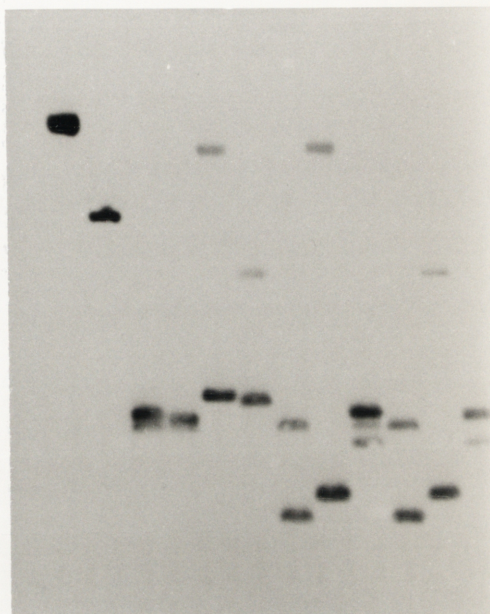
Fig. 3.7

Example of the detailed restriction mapping and probing of the (sub) fragments from within the subterranean clover genomic clones.

The 4.0kb *Eco*R1 fragment. Digestion and probing of clone pTs5F (the 4.0kb fragment in Bluescript).

A

Lane 1: λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Lanes 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13: Restriction of pTs5F with the following enzymes - *Xba*1, *Xba*1/*Eco*R1, *Acc*1, *Acc*1/*Eco*R1, *Kpn*1, *Kpn*1/*Eco*R1, *Xba*1/*Acc*1, *Xba*1/*Kpn*1, *Acc*1/*Kpn*1, *Eco*R1/*Xba*1/*Acc*1, *Eco*R1/*Xba*1/*Kpn*1, and *Eco*R1/*Acc*1/*Kpn*1 - respectively.

B

Hybridization of the bean CHS1 cDNA probe to the Southern transfer of the gel in A.

In order to verify the results of A and B, a Bluescript clone carrying the 4.0kb fragment in the reverse orientation to pTs5F was also digested and probed in an identical manner (not shown).
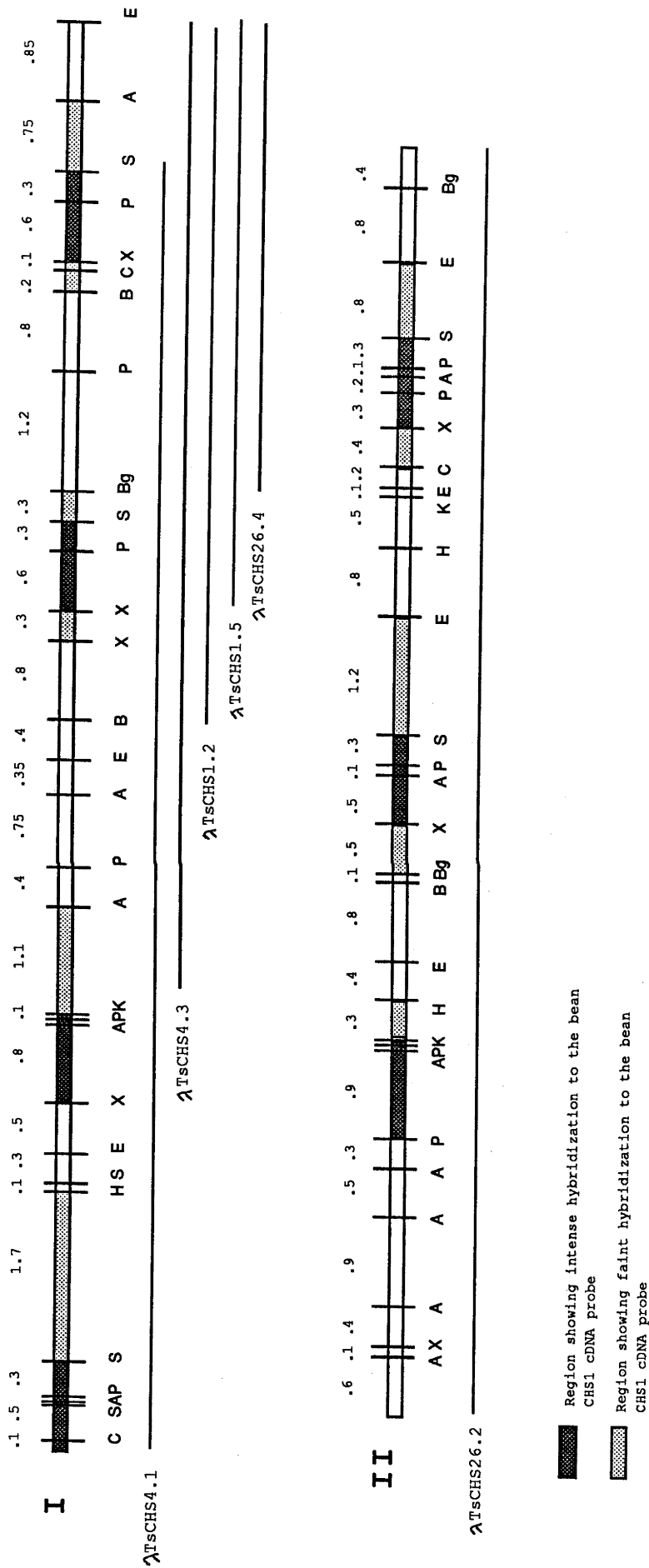
Fig. 3.8

Detailed restriction map for the two regions of subterranean clover DNA, numbered I and II, containing seven separate areas of hybridization to the bean CHS1 cDNA probe. These areas are indicated by shading. The positions of the five genomic clones which covered region I are shown beneath the restriction map for that region. Note that λTsCHS1.2, λTsCHS1.5, λTsCHS4.3, and λTsCHS26.4 extended beyond the EcoR1 site at the right hand end. Only one genomic clone covered region II. A - Acc1, B - BamH1, Bg - Bgl11, Cla - Cla1, E - EcoR1, H - Hind111, K - Kpn1, P - Pst1, S - Sac1, X - Xba1.

3.2.3 Clover CHS copy number

By observing the mapped restriction sites in relation to the CHS genes, it was reasoned that the only useful enzyme with which to cleave clover total genomic DNA for copy number estimation by Southern analysis was *Eco*R1. This enzyme did not possess internal restriction sites in any of the characterized CHS genes, but did possess sites which nicely separated all of the members of the clusters (except CHS1 and CHS2). DNA was isolated from subterranean clover cv. Karridale, digested with *Eco*R1, electrophoresed and Southern transferred to Hybond-N.

The filter was hybridized with the *Pst*1-*Xba*1 fragment of CHS2, which probing with the bean CHS1 cDNA, and sequence analysis (chapter 4), showed to contain entirely chalcone synthase sequence (specifically that from the second exon). It was to be expected, then, that only the coding sequences of CHS genes would hybridize. The probing (fig. 3.9) revealed all of the expected bands, and several others, among which were presumably CHS4, 7, and 8 (all of which only possessed one flanking *Eco*R1 site each on the characterized clones; fig. 3.8).

The identification of the *Eco*R1 bands corresponding to CHS4, 7, and 8 was ascertained by hybridizing probes to *Eco*R1 cleaved genomic DNA which were expected to be specific for the gene (CHS4 and CHS7), or preferentially pick it out in relation to the other CHS genes (CHS8). CHS4, 7, and 8 were found to be situated on ~6.0kb, ~10.0kb, and ~4.0kb *Eco*R1 fragments, respectively (fig. 3.10); CHS8 shares the same *Eco*R1 fragment size as CHS3. The fragments containing CHS4 and CHS7 were observed to always faintly hybridize to the exon2 probe (fig. 3.9). The presence of a deletion in the second exon of CHS7 (chapter 4) may explain the poorer hybridization of this gene.
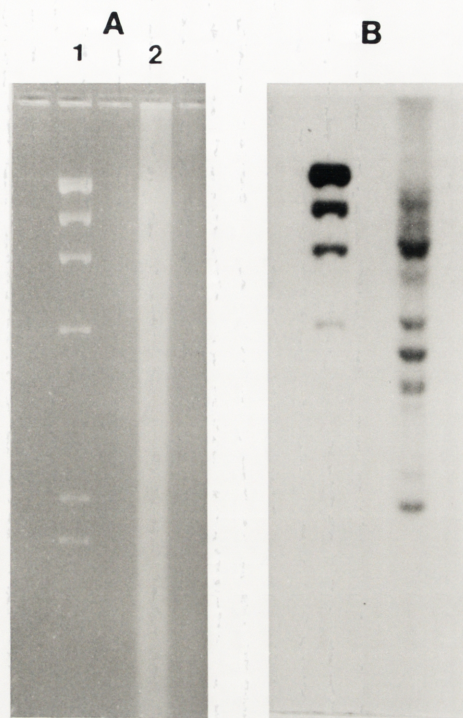
Fig. 3.9 Determination of the copy number of the gene chalcone synthase (CHS) in *Trifolium subterraneum* L. cv. Karridale.

A

Lane 1 - λ/*Hind*111 standard markers (23.1 9.4, 6.6, 4.3, 2.3, 2.0, 0.5kb).
Lane 2 - *Eco*R1 digested *Trifolium subterraneum* L. cv. Karridale genomic DNA.

B

Autoradiograph of the Southern transfer of the gel shown in A after probing with the 0.6kb *Pst*1-*Xba*1 fragment of *Trifolium subterraneum* CHS2. This fragment consists entirely of clover CHS exon2 sequence.

Bands corresponding to CHS genes for which the *Eco*R1 fragment size was determined by restriction mapping of the isolated genomic clones:
7.5kb (CHS1, CHS2), 4.0kb (CHS3), 3.5kb (CHS6), 2.3kb (CHS5).
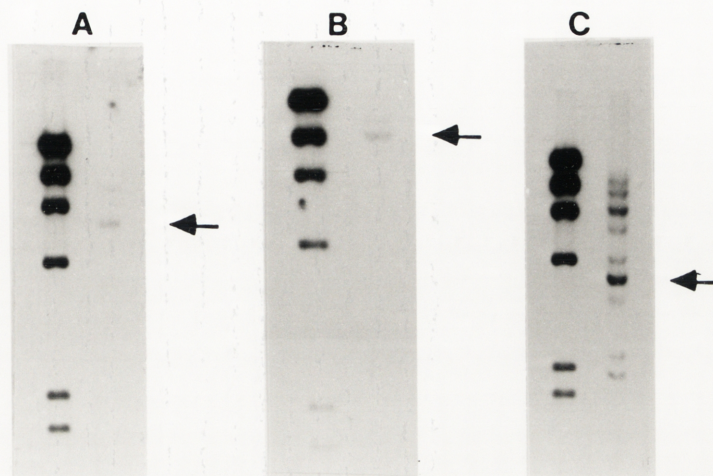
Fig. 3.10 Location of CHS4, CHS7 and CHS8 on the *Eco*R1 digest of genomic DNA from *Trifolium subterraneum* L. cv. Karridale.

Three autoradiographs obtained after separately hybridizing to Southern transfers of *Eco*R1 digested *Trifolium subterraneum* L. cv. Karridale genomic DNA probes selected to identify the fragments carrying CHS4, CHS7 and CHS8. For CHS4 the probe was the 0.4kb *Hin*d111-*Eco*R1 fragment of λTsCHS4.1, for CHS7 the 0.4kb *Hin*d111-*Eco*R1 fragment lying adjacent to the hybridizing region of this gene, and for CHS8 the 0.8kb *Hin*d111-*Eco*R1 fragment of λTsCHS1.6.

A. Position of CHS4 in the *Eco*R1 digest of *T. subterraneum* L. cv. Karridale genomic DNA - ~5.5kb.

B. Position of CHS7 in the *Eco*R1 digest of *T. subterraneum* L. cv. Karridale genomic DNA - ~10kb.

C. Position of CHS8 in the *Eco*R1 digest of *T. subterraneum* L. cv. Karridale genomic DNA - ~4.0kb.

Note that the probe used for the fragment containing CHS8 consisted of CHS coding sequence. This can easily be seen by its hybridization to all the CHS genes; nevertheless, the preferential hybridization of the CHS8 gene is obvious. The standard markers are λ/*Hin*d111 (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Several other bands hybridizing to the exon2 probe could not be accounted for by the the isolated and characterized CHS genes. In particular, a band of ~4.3kb was found to hybridize at the same intensity as those for other single CHS genes (except CHS4 and 7 as mentioned just previously). Two other bands, at ~2.6kb and ~3.2kb, hybridized less intensely to the probe, suggesting incomplete CHS sequences, partial fragments of CHS genes containing internal *Eco*R1 sites, or sequences which fortuitously bear resemblance to at least part of the probe.

The overall result from this experiment was that the genome of subterranean clover contains at least nine CHS genes. The possibility of some of these being alleles in this diploid plant cannot be excluded, but as the copy number of CHS in other characterized legumes is known to be approximately six per haploid genome, a number of about nine for clover is not to be unexpected (Ryder et al., 1987; Wingender et al., 1989).

3.3 Discussion

By using the bean CHS1 cDNA clone as a hybridization probe, it was possible to identify, isolate, and characterize a CHS multigene family in subterranean clover. Seven individual clones containing CHS genes were isolated from a gene library of clover by their hybridization to a bean CHS cDNA probe. Five of these were established to be overlapping fragments from the same region of genomic DNA. The result obtained from using the "3'" probe suggested that four separate genes were present in the genomic DNA present in these clones, and three more in a separate region covered by a single clone (λTsCHS26.2). The hybridization pattern of the seventh genomic clone suggested an "isolated" gene was present on one end of it. Hybridization with the 3' probe was based on the assumption that only a single intron, at a

conserved position, would be present in each of the CHS genes. As in only one instance, that of *Antirrhinum majus* (snapdragon), has a CHS gene been found to have more than one intron (Sommer and Saedler, 1986), the data obtained from this probing experiment was worthy of respect.

Restriction mapping, and associated probing with the (entire) bean CHS1 cDNA, provided further evidence for the existence of seven genes within the stretches of clover DNA covered by λTsCHS1.2, λTsCHS1.5, λTsCHS4.1, λTsCHS4.3, and λTsCHS26.4, and λTsCHS26.2. The seven areas of hybridization were found to be confined to *specific* sites within the various *Eco*R1/*Bam*H1/*Hin*d111 fragments. Moreover, each was composed of a part which hybridized intensely to the probe, and at least one other region which did only faintly (fig. 3.8). As stated above (3.2.2), only sequencing (next chapter) finally confirmed each area of hybridization to represent a CHS gene (fig. 3.11).

CHS1 is flanked (to its right as drawn in fig. 3.11) by approximately 9kb of DNA, stretching to one limit of the characterized clones, which is devoid of any CHS gene; this suggests that it constitutes one end of the cluster. However, as CHS4 occurs at precisely the other limit of cluster I, it is possible that other members of this gene cluster exist (i.e. to the "left" of this gene). Clusters I and II could conceivably be separate "profiles" of the same gene cluster, with the end of CHS4 being very close to one of the termini of cluster II. A single gene cluster of such a large number of CHS genes has not, as of yet, been reported for any plant. Sequencing provided evidence both for and against clusters I and II being part of a single locus (next chapter).

Although the direction of transcription of the seven genes was only established by sequencing, and is as indicated in fig. 3.11, the restriction mapping provided a tentative answer. The repeated occurrence of the

*Sac*1-*Pst*1 and *Pst*1-*Xba*1 restriction motifs in members of the two clusters, and the identical orientation of each motif with respect to the others, suggested that the four genes in cluster I, and the three in cluster II, were transcribed in the same direction (the orientation of the two clusters may not necessarily be identical if they occur on the same chromosome).

Families of CHS genes have been found in the genomes of all legumes that have so far been examined, and also in the non-legumes *Petunia hybrida* and mustard (1.5.2). Hence, the discovery of a CHS gene family in subterranean clover is not surprising, and it would appear that such families are a universal phenomenon in the Leguminosae. Multigene families are found for many plant genes besides CHS (Sutliff et al., 1991, and references there-in), but in higher plants it appears that most genes occur as a single copy (Tanksley and Pichersky, 1988). As far as gene family sizes go, the CHS gene families are very small in comparison to eukaryotic transfer RNA, ribosomal RNA, and histone gene families (Ohta, 1983). These gene families frequently contain between 100 and several thousand members.

The four CHS genes of cluster I, and the three of cluster II, are tightly linked. In cluster I, CHS1 and CHS2 are separated by ~2.0kb, CHS2 and CHS3 by ~3.4kb, and CHS3 and CHS4 by ~2.5kb; similar intergenic distances are observed in cluster II. The close proximity of the clover CHS genes is similar to that found with three of the CHS genes of soybean (Wingender et al., 1989). These soybean genes, two of which are divergently transcribed, are separated by regions of only 2.3 and 3.5kb. Some of the bean CHS genes are clustered, and in at least one case two such genes are orientated in a 5' to 5' fashion (Ryder et al., 1987). However, although some of the CHS genes of *Petunia hybrida* are

clustered, they are separated by distances in excess of 10kb (Koes et al., 1987; 1989b).

Multiple copies of genes which code for abundant proteins are frequently found in clusters (Pichersky, 1990). For example, clusters of chlorophyll a/b binding protein (CAB) genes have been discovered in tomato, *Arabidopsis thaliana*, and *Petunia* (Mitchell) (Pichersky et al., 1985; Leutwiler et al., 1986; Dunsmuir et al., 1983). Clusters of SSU genes have been identified in a number of species, including tomato and *Petunia* (Sugita et al., 1987; Dean et al., 1985). In *Petunia* (Mitchell), four genes for SSU are located within a 20kb stretch of DNA (Dean et al., 1985), while in the same plant two subfamilies of CAB genes each contain two members separated by only 2-3kb (Dunsmuir et al., 1983). It has been suggested that the high CHS copy number in the Leguminosae may be the result of the need for extra genes to "cope" with synthesis of the isoflavonoid phytoalexins which are characteristic to this plant family (Ryder et al., 1987).

It is believed that tandemly linked genes are produced by homologous but unequal crossing-over (Smith, 1974; Jeffreys and Harris, 1982). The initial duplication of a single gene may be determined by the fortuitous location of repetitive sequences, which can undergo such recombination, in the flanking DNA at both ends (fig. 3.12). In the case of a rice α-amylase multigene cluster, probings indicate the presence of repetitive DNA surrounding the genes; these sequences are suggested to be responsible for the duplications (Sutliff et al., 1991). It is also possible that such events as nonhomologous chromosomal breakage and reunion are important in the evolution of multigene families (Maeda and Smithies, 1986).

Once a tandem repeat is created, the greater sequence similarities between the two homologous chromosomes that each bear the gene

repetition may lead to further duplications of the initial gene pair (Pichersky, 1990). The two clusters, or possibly one, of CHS genes present in clover could well be the result of such a process. If these two clusters of the clover CHS gene family (as well as CHS8) are not linked, then the origin of the family cannot be explained entirely by unequal crossing-over. Although there are no well documented mechanisms available for the origin of unlinked duplicate genes, it is possible they are created by the breakage and insertion elsewhere in the genome of linear DNA fragments (Pichersky, 1990). A further point to consider is that the two gene clusters are allelic. However, as subterranean clover is a diploid plant which is highly inbred, such a scenario is unlikely. Furthermore, the restriction enzyme sites flanking the genes in the two clusters are not the same.

As CHS is represented by multigene families in all members of the Leguminosae examined to date, selection must obviously act to control the copy number. Either duplication of these genes is by some virtue a genetic event that occurs readily, and selection only acts when the copy number exceeds a certain "limit", or there is selection pressure to *maintain* a multigene family (Tanksley and Pichersky, 1988). In the last case, of course, selection may act to prevent the copy number from becoming too high.

In the Solanaceae, it is possible that the members of the SSU gene families in each species, except those from the genus Nicotianeae, map to three independent loci (Tanksley and Pichersky, 1988; Jamet et al., 1991). The copy number of this gene has not, however, been conserved between species. It is attractive to speculate that the many genes within each species of the Leguminosae are also confined to a defined number of loci. If this were indeed to be the case, then these loci would need to have been present in the last common ancestor of all the Leguminosae. A

great amount of research, focusing on the genetic structure of the CHS genes in the Leguminosae, would be needed to confirm this hypothesis. If clusters I and II are part of the same locus, then conceivably CHS8 is located close by. Such a situation would parallel that of the SSU multigene family in pea, the five members of which are clustered at a single locus (Polans et al., 1985).

The probing of the subterranean clover gene library for CHS, and the characterization of the CHS genomic clones, was conducted in conjunction with T. Arioli.
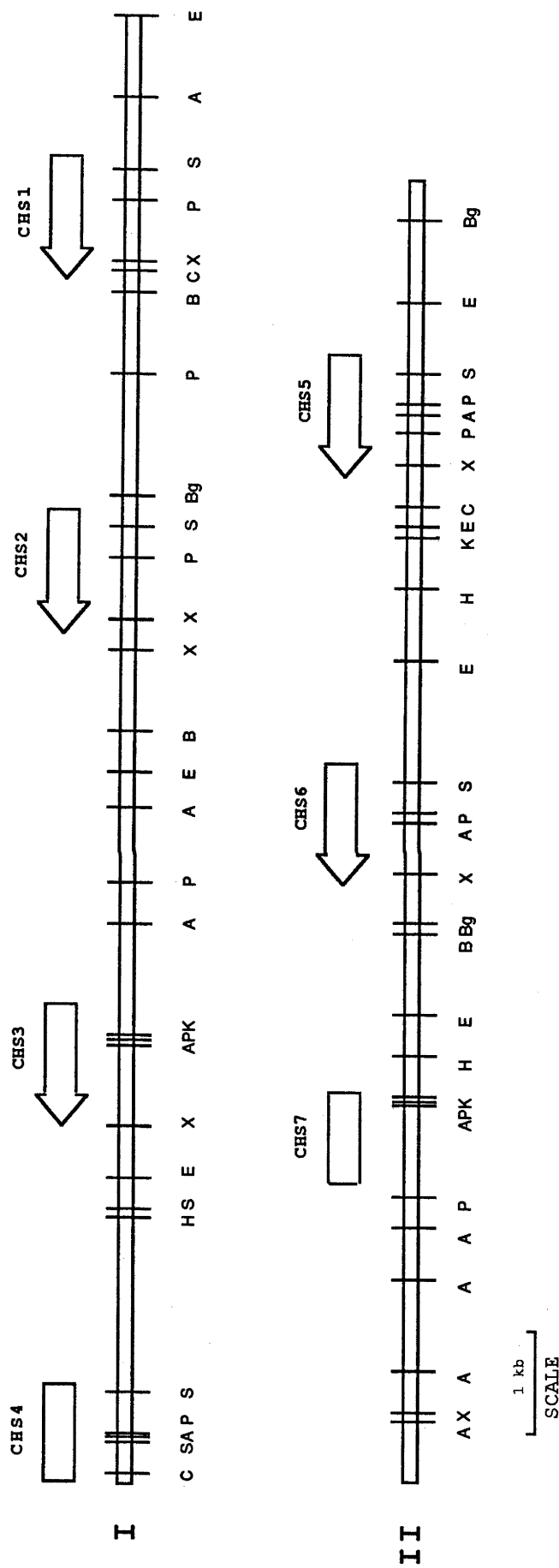
Fig. 3.11

Position of seven CHS genes, grouped into two clusters (I and II), in the stretches of subterranean clover DNA covered by six of the seven characterized genomic (λ) clones. The genes, and their directions of transcription (which were established by sequencing) are indicated by arrows. As CHS4 is incompletely represented in cluster I, it is shown as a box; it is, however, orientated in the same direction as the other three CHS genes of this cluster. CHS7 is also shown as a box as sequencing found it to be truncated (inactive).

A - *Acc*1, B - *Bam*H1, Bg - *Bgl*11, Cla - *Cla*1, E - *Eco*R1, H - *Hind*111, K - *Kpn*1, P - *Pst*1, S - *Sac*1, X - *Xba*1.
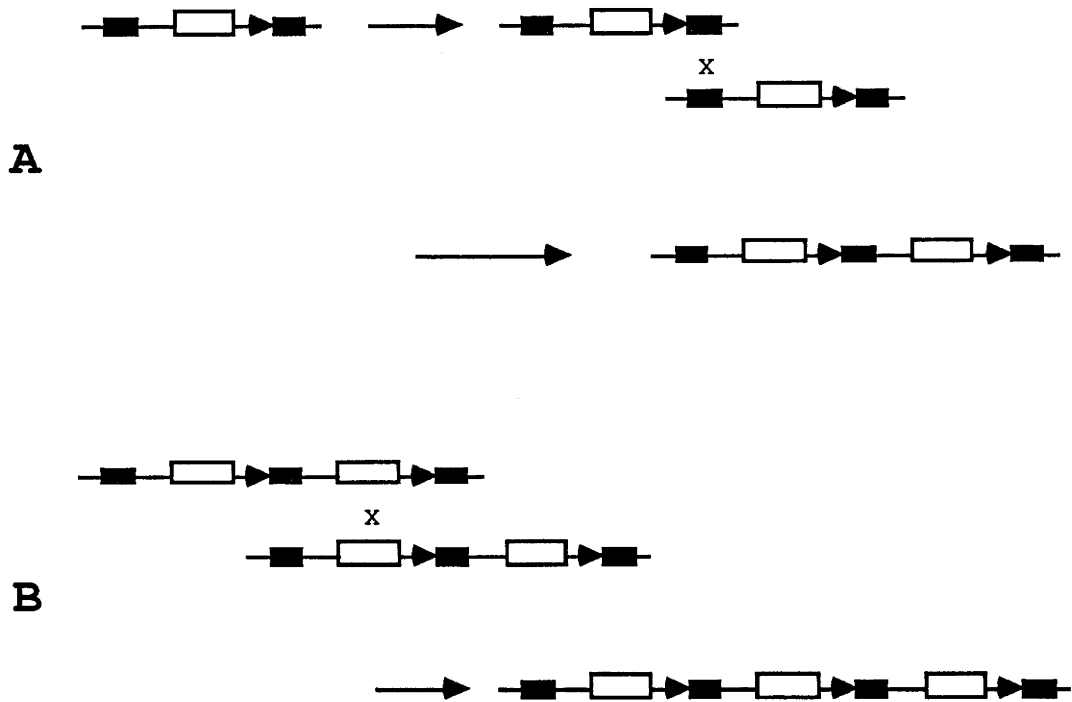
Fig. 3.12 Diagram showing one possible mechanism by which the subterranean clover CHS gene cluster(s) could have been generated. Short repeated sequences (black boxes) at either end of an initial single gene (unshaded box) may have allowed unequal crossing-over between homologous chromosomes to occur at meiosis (A). Later, the new duplicated gene could have undergone recombination in an similar manner, leading to an even larger cluster (B). Modified after Jeffreys and Harris, 1982.

Chapter 4. Sequence analysis of the clover CHS gene family

4.1 Introduction

In order to study the CHS gene family of subterranean clover in more depth, the DNA sequences of at least some of the genes were necessary. Sequence data would establish both the sequence similarity between the genes and enable the direction of transcription of each member of the characterized cluster (in relation to the others), to be determined. Furthermore, the sequences of these genes could be compared with those previously published for the CHS genes isolated from several plants (chapter 1). Of special interest would be the degree of homology between the clover promoters, and the possibility that certain elements within them could be present in the promoters of other CHS genes, and indeed other plant genes. These elements, if not already identified as such, would be prime candidates as the sequences essential for transcriptional regulation. This chapter presents the entire sequences of CHS2, CHS3, and CHS5, and that available for CHS4, including approximately 1kb of each promoter, and an analysis of the structure of each. Sequencing of a small amount of CHS7 verified it was incomplete (truncated). A single clone of the gene CHS8 was sequenced to orientate the coding sequence on the genomic fragment λTsCHS1.6.

The sequences of CHS2, 3, 4, and 5 hybridizing to the bean CHS1 cDNA probe were contained, in entirety, in the 4.3kb *Bam*H1 fragment, the 4.0kb *Eco*R1 fragment, the 3.0kb *Eco*R1 fragment, and the 2.3kb *Eco*R1 fragments respectively (figs. 3.3/8). These four fragments had been cloned, each in both orientations, into Bluescript SK(+) during restriction mapping (chapter 3). For each gene, a series of sequencing clones was created for both orientations (directions) by deleting into their

two respective Bluescript clones with Exonuclease III. The M13/T7 primer region was in every case that protected from Exonuclease digestion. The restriction enzyme "block" sites (i.e. those containing a 3' overhanging end) were either *Kpn*1 or *Apa*1, depending upon the fragment. Other sequencing clones were formed by restriction digests and the subcloning of fragments.

The sequencing strategy for CHS2 is shown in fig. 4.1; the sequences of the other three genes were established by a similar array of overlapping clones.
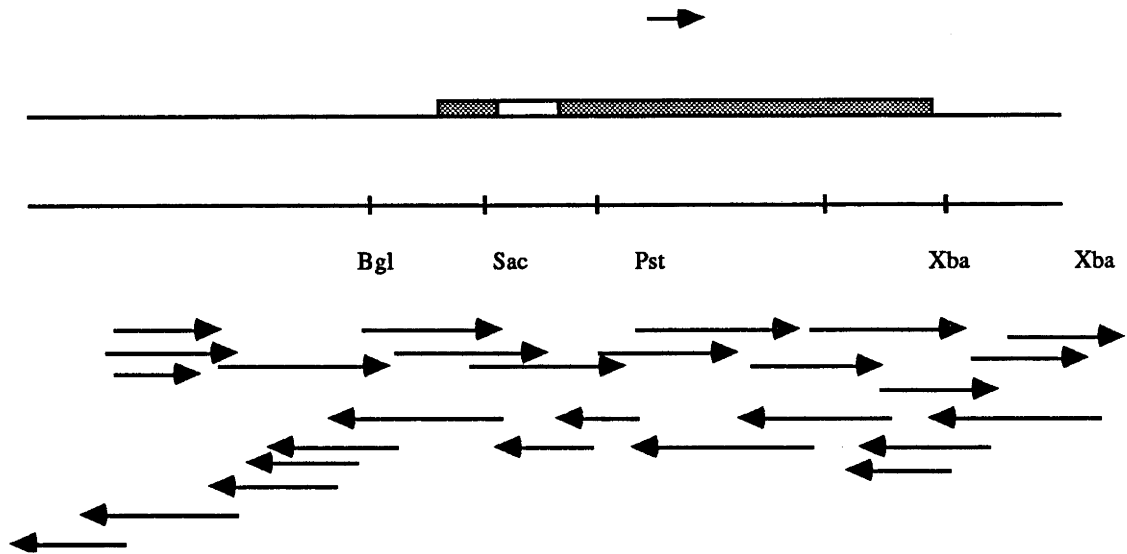
Fig. 4.1 Sequencing strategy for subterranean clover CHS2. The structural arrangement of the gene (as deduced by the sequencing) is represented at the top: the two exons (shaded boxes) are separated by a single intron (unshaded). Major restriction sites (cf. fig. 3.8) are indicated beneath (Bgl - *Bgl*11; Pst - *Pst*1; Sac - *Sac*1; Xba - *Xba*1). The arrows at the bottom show the length and direction of sequencing clones obtained by both Exonuclease III deletions and cloning of restriction fragments, while the single arrow at the top indicates the direction of transcription.

4.2 Results

4.2.1 General examination of the sequences of CHS genes 2, 3, 4 and 5

The nucleotide sequences and predicted amino acid sequences of subterranean clover CHS2, CHS3, CHS4 and CHS5 are shown in figs. 4.2 and 4.3. Analysis of the sequences by Staden's (1984) "Positional Base Preference" programme and Fickett's (1982) "Testcode" programme (fig. 4.4), together with comparison to published sequences of CHS genes (see below), established that the four genes are each composed of two exons separated by a single intron. The coding sequences of CHS2, 3, and 5 are 1170bp (including the stop codons), corresponding to proteins of 389 amino acids (~ 43kDa). CHS4 is missing its final 236 nucleotides in the isolated genomic clone λTsCHS4.1. Translation of the mRNAs derived from the genes almost certainly begins at the ATG codons indicated in fig. 4.2. This can be concluded not only by homology with CHS protein sequences from other plant species (especially those of the Leguminosae; see below), but also from the absence of any in-phase methionine codons in the adjacent 5' flanking regions of DNA. As is the case with all CHS genes for which the genomic structure is known, the introns of the four clover genes occur between the first and second nucleotides of a cysteine codon (amino acid 60; Schroder and Schroder, 1990; 1.5.2).

All three CHS genes (and CHS1; T. Arioli personal communication) in cluster I are orientated in the same direction with respect to transcription. The position and orientation of CHS3 within the 4.0kb EcoR1 fragment is such that part of its promoter must be present on the 2.4kb EcoR1 fragment of λTsCHS4.3 (fig. 3.3), hence providing a solution to the problem of the latter fragment hybridizing to the 2.8kb EcoR1-BamH1, 4.3kb BamH1 and 3.0kb EcoR1 fragments (3.2.1). Homology between the promoters of the four CHS genes must be high enough for

that of *one* to readily hybridize to the others; the sequence data presented below underlines this.

A prediction of the secondary structure of the CHS2 protein using the algorithms of Chou and Fasman (1978) is shown in fig. 4.5. However, with all such exercises, the calculated structure may only weakly resemble that of the actual protein. Data obtained from X-ray crystallography would be necessary to firmly establish the structure of the CHS protein.

Fig. 4.2 (overleaf, next three pages) Nucleotide sequences of *Trifolium subterraneum* L. cv. Karridale CHS2 (2), CHS3 (3), CHS4 (4), and CHS5 (5). For each gene, except CHS5, the sequences of approximately 1100bp of 5' non-coding DNA and 300bp of 3' non-coding DNA are also shown; these sequences are slightly shorter for CHS5. In the case of the coding sequence (exons), only the sequence of CHS2 is indicated; the nucleotide sequences of the other three genes are indicated only when they differ from that of CHS2. The full intron sequences of the four genes are shown. The 3' end of the coding sequence of CHS4 was not present in the characterized clones; the "last" nucleotide of this gene is indicated by a #. Putative TATA boxes are underlined. Numbers in the 5' non-coding sequence refer to positions with respect to the start points of *translation*. The transcription initiation point for CHS3, as mapped by primer extension (chapter 6), is shown in bold face type.

5' NON-CODING SEQUENCES

```
-1100
2 GATTTTGTTT TTAACATTTG ACTATGACAG CTTTGGTGAC TAAATTAATG TCCTTGTCCA TGTTCATCCC TTTTAATCTT
3 TACCAGTAGC AGGAAGGATT CGCCCAACTA GGAAGAGTTA CATGGTCCTC AACTGATTAT GAAACTAGGA AAAAGTTATA
4 CAGGAAGAAT TACATGGCCC TCAATTGATT ATGAAACTAG GAATAAGTTA TATGGTCTTT GATTTGAGTA TAAAAGTGTA
-1020
2 TACTAAATTG CTTGAAACTA ACTGAATGGA AACCAGCATT TGTTTGCTTA TGGAGGCTGA ATTTTACTAC TAATTTATTT
3 TGGTCTTCAA TTTGAGTATG ATACAAGGCA CGTAGTTTTT GAGTTGATTC TTCTAGTTCT AAGTAGTTGG CATACAAAAT
4 AAAGTGTATT TTTAACAAGG CACACATAGT TTTTGAGTTG ATTCTTCTAG TTCTAAGTAG TTGGCCTACA AAATGTTTTG
-940
2 ATTTTTGGTA CAAAGGGTGG TTGTTAGTAA GCTAACATGC CAAAATTATT GAAATAAGTT GAAAATAATT TATAAACTGT
3 GTTTTGTAGT ATGTTGCTTT TAACAATTTG ACTATGATAG ATTTAGTGAC TAAATTAATG TCCTTGTCCA TGTTGATCCC
4 TAGGATGGTT TTTTTAGCAA TTTGAATATG ACAGCTTTAG TGACTAAATT AATGTCCTTG TCCATGTTCA TCCCTTTTAA
-860
2 TATTATAAGT TTTCAAAAAC AGTGTCACAA GTATTTATTT CAGTAGATAA ACTCAAATAA GTTATTTCAA ACGACCAAAA
3 TTTTAATCTT TTCACTTTGT TGTTTATGAA CATGTTATAT AGGGTGTTTG GTATTCGAC ACGTAAAGCA AAATTGTGGC
4 TCTTTTCACT TTGTTTGTTA AATTGCTTGA AATTAATTGA ATGGAAATCA ATATTTGTTT GCTTTATGGA GATTCTACTA
-780
2 ACAATGATAT TTGAACACTC TTATTTTTCC GACACTCATT CGCTCTATCT TTTCCTTCCT ATCACATAAC CAACATACTA
3 AAATAATTGT GTAGACGTTG AAGCTGTAAT ATATAGCTTC TCTATTTTCA CGTCAATATT ACCTTGGGAC ACTTTCATGT
4 TTAATGGTAG TTTATGGAAA TAAGTTCAAC AGGTTATATA GGATGTTTGG TACTGGGACA AGTTAAGCAA AATCGCGGCA
-700
2 CAACATTCTC CCTATTTATT TCTCTCTCAA AGTATCAATT GATGTTTAGG AGTTCGTTTG ATCTACCAAA TGTAAGTACT
3 TAATATTACC CCGGGACACG AAAATGTAGC TCAAATTGCA TATCCACACA TACCAATAAT CCGATATTAT ATGCTCTACA
4 AAGAATCACT TAGATGTGGA AGCTGCAATG TATAGCTTCT TTGTTTTCAC GTCAATATTA CCTCGTGACA CGAAAATGTA
-620
2 AGACAGAACA GTACATGAGA AGATAAAGTT TAAGGTATTA AACAAATTTT GTCTTGTACA TGTTTGGTGG ACAAAATGTT
3 AAATAGTGTC AGAGTGTTTA TGTTAGTAAA AAAATTCAAA TAAGTATATT CAAACAGACG CATTACTCAA GAATGATATT
4 GCTCAAACCG CATATCCAAA CATACCAATA ATTCAATATT ATATGTTCTA CAAAATAGTG TCAGAGCGTT TACGTCAATA
5 TGATTAATTA AATTTAGTTA AAGACTAATT GTTCATGAAA AGTTTAAATC TAACTGTAAC AGATTGATCA AATTTTATTT
540
2 ATTTTGATTT TAGACAACTT GTGCAGAAGA CAAAAAATTG TTTAGTCCCT TGATTTCCGG TTTTTCTAGT CACAATTTTA
3 TAGACACTCT TAATTTTCAC TACTCATTCT ATATATTTCT TTTCTTTCTT ATCACATCAC CAATATACCA CAACATTCTA
4 GAAAAATTCA AATAAATAAA TTCAAACAGA GTATTAATCA ACAATGATAT TTAGACACTC TTAATTTTCG CCACTCATTC
5 GTGAACTCTA AATTATCACA TTTTAATTAT TTTCTCATAG AACAAGAGCG TTACAATCAA TTTTACAACT CTAAAAATGA
-460
2 TTTTTAAATC AATCATAGTA CAACTAAAAT ATTCCTGTCC AATCCCTTGT ATTTTAACAA ATCAAACACA TCACTTAAAC
3 CCTATTCTTT TCACTCTCAC AGTAATCATT AAACATTTTC CTTAATTAAT TGCAATTATT TGCAGCCCGG TTGTGTTGCA
4 TCTCATTCTT TTTTTTAATA TTTTATATTA ACATACCAAT AACAGATTTT CTAAAAATGT CATATGCAAT TATGCAGCCG
5 TTTTGTCTTA TTCAAAAAAT GGAATCTATT ATACATTAAC GTAACATATT CTTATACAAC GGTATTAGTA TGATTCATGT
-380
2 ATTTTCCTTA ATTAATTGTA ATTATTTGCA GCCGGTTTGT GATGTATGTA ACTTGAGATT TAACAAAAAT AAGTGCACAA
3 CGTGACTTAA GATTCACAA AAATAAGTGG ATAAATAAAA CAAAATACTA GTCAAATTGT TTACAAAATA ATAATTAAAA
4 GTTGTGTTGC ATGTGACTTA AGATTTAACA AATAAGTGCA AAAATAAAAC AAAAGACTAG TCAAATTGTT TACAAAATAA
5 ACCTTAAGAT CCCCTAAAAT AATTAAGTAA TAAAATATAA CCAATAATTC CTTAATAAAA TAAACAGAAG AAAAAAAACT
-300
2 ATAAAACAAA ATGCTGGTCA AATTGTTTAC AAAATAATAA TTAAAATAAA AGAGGGTCTT CAAAATCATG GTGGTAGTTG
3 TAAAAGTGGT TGGACTAGAG AGGGTCTTCC AAGTCATGCT GCTATTTTCA ACCAAACTCC TACATGTCAC AAACACAACC
4 TAATTAAAAT AAAAGGGGTT GGACTAAAGA GGGTCTTTGA AGGCATGCGG GTATTTTCCA CCAAACTCAC TACATATCAG
5 ACAAAATTAG GTAAAGACTA AAAGAAATTG TTAAGTTAAA AAAAGAGACT AAATCATGCA AGTATTTGCC ACCAAACTCC
-220
2 CCACCAGACT CCTACATGTC ACAAAATCAA CCAAGCAACA AGTTGTTGTG TTGAAGCACA TGATGTTCAT TCAGCTACCT
3 AACTGACCAA GCAGCCATAG TATTATTATT AGCAGCCAAG TTGTTTTGCT GAAGCAGCCA ATATTCACTA TCACCTCTTC
4 AAGCAGCCAT AGTTGAAAGC ACGTGACTGA TTTTATATAT TCAGCTACCC TCCATTTCAA TACTCATTAT CAACTTTCCA
5 TAGGTACATC TACATGGCAC AAACACAACC AGCCAACCAT ATTATTATTA AAAATCAACT TGTGTTGAAG CACGTGAAGT
-140
2 TCCAATATTC ACTATATATA ACCATCTCAT CCCACAGATC TTTTCACCAC AACACACTTC AGCCTTTGCT TTTCTACAAC
3 TTCCAATCTC CCCTTCATTC ATATATATAT AACCATCTCA TCCCACACAT TTCCACCACA ACACTCTTCA TCTCATTCCT
4 CTTTTCCCTT CTTATAAAAC CATCTCACCC CACACATCTT TCCACCACAA CACTCTTCAA CCTTTGCTTT TCTACAAACA
5 TCAGCTACCC TCCAATATTC ACTACCTCTA CCCTTCATAT ATATATAACC ATCTCACCCT TAACATATTT CATAACAATA
-60
2 AACTTCTGTA ATAAACCACT ATTGAGTTCA ATTTACATAG AAACTACTAA AGATATTAAC
3 TTTCTATAAA AACAACTGTG TTCAAATTAC ACATTGAAAC TACTAAAGAT ATATTAAGAT
4 ACTTCTGTAA TAAACCATTA TTTGAGTTCG AGTTACATAG AAACTACTAA AGATATTAAC
5 AGTTCCTGTA ATAAACCAAT TCAAATTACA TACATAGCAG GAACTACTAA TTCTATCAAT
```

FIRST EXONS

```
2 ATGGTTAGTG TGTCTGAAAT TCGTAAGGCT CAAAGGGCAG AAGGCCCTGC AACCATTTTA GCCATTGGTA CTGCTAATCC
3        G          A                     C
4                              C
5        C          AG                    C                                T

2 AGCAAACCGT GTTGAGCAGG CCACATATCC TGATTTTTAC TTCAAAATTA CTAACAGTGA GCACAAGGTT GAGCTCAAAG
3     T     AAG                    C                    C          A        T          A
4     T     AAG                    C                    C                   T
5     C     AAG                    C                    C                   T                 G

2 AGAAATTCCA GCGCATGT
3
4
5         T
```

INTRONS

```
2 GTAAGTATTA AGCAACTTCT GATATTGTGT GTAAAAAAAA AGCAACTTCT GATATTAAGT TTTCCTACCA AATTTATGCA
3 GTAAGTATTA TTATTTGTAG AAGCATTAAA AGACATATTA TCATTATTAT GTGATTGACT TTGGTATTGC CCAAAAATTG
4 GTAAGTATTA AGCAACTTCT GATATTTAAT TTTCCTACCA AATTTATACA AATTATTATA TATGTGATTG TGATTGTCTT
5 GTAAGTATTT GCCTACCAAA TTGTTACAAA TTATTTGCAT GAAAAGATAT ATTATACATG GCACCTAAAT ATATTGTAGT

2 AATTATTTGC ATCAAACAAC ATATTATATA TGTGATTGGC TTTGGTATTA CCATTATCCA AAAATTGTTA TACTAATGTC
3 TTGTACTGAT GTCATTTTTT GTTATAATTA ATTCAG
4 TGGTACTACT ATTATCCAAA AATTGTTGTA CTAATGTCAT TTTTCTTATA ACTAATTCAG
5 ACTAATATGA TTTTTTTTAA CTAATTCAG

2 ATTTTTCTTA TAACTAATTC AG
3
4
5
```

SECOND EXONS

```
2 GTGATAAATC TATGATCAAG AGTAGATACA TGTATCTAAC AGAAGAGATT TTGAAAGAAA ATCCAAGTGT TTGTGAGTAC
3                       C                    T                    G        T  C        A
4            C                                                    G        T
5                                            G

2 ATGGCACCTT CATTGGATGC AAGACAAGAC ATGGTGGTGG TAGAAGTACC TAGACTAGGA AAGGAGGCTG CAGTGAAGGC
3                      T                     G                     G  A
4                      T  G                                        G
5                      T  G                                        G  A

2 TATTAAAGAA TGGGGACAAC CAAAGTCAAA GATTACTCAC TTAATCTTTT GTACTACAAG TGGTGTAGAT ATGCCTGGAG
3                    T                        A  T  G            C  C                  C         T
4 C  C                C          A            T  G                                     C
5 C  C                C          A  T         T  G            C                        C

2 CTGATTACCA ACTTACAAAA CTCTTAGGTC TTCGCCCGTA CGTAAAAAGG TATATGATGT ACCAACAAGG TTGTTTTGCA
3           G  C                            A  T  G  G          C                  A  C
4              C                         T  A  T     G          C                  A  C
5              C                            T  G  G                                G     C  T

2 GGAGGCACGG TGCTTCGTTT GGCTAAAGAT TTGGCTGAGA ACAACAAAGG TGCTCGTGTG CTTGTTGTTT GTTCTGAAGT
3 T  G       T                    A  C                  G                T  G
4 T  G       T          C  G                          G  A              T  G
5 T  G                  C  G       C                                    T  G

2 TACCGCAGTC ACATTCCGCG GTCCTAGTGA CACTCATTTA GACAGTCTTG TTGGACAAGC ACTATTCGGA GATGGAGCTG
3 A                   T  C                  C  G                              T
4 A                T  T  C                  C  G                    T     T              C
5 C  T                T  C                  C  G                    T

2 CTGCTCTTAT TGTTGGTTCT GATCCAGTGC CAGAAATTGA AAAACCAATA TTTGAGATGG TTTGGACTGC ACAAACAATT
3           C                                A            G         T          A
4    A  C                           G        G                      A
5           C                                G

2 GCTCCAGATA GTGAAGGAGC CATTGATGGT CATCTTCGTG AAGCTGGACT AACATTTCAT CTTCTTAAAG ATGTTCCTGG
3           C          T
4           C          T           C
5           C          T                                                                    C
```

```
2 GATTGTCTCA AAGAACATTG ATAAAGCATT GGTTGAGGCT TTTCAACCAT TAAACATCTC TGATTACAAT TCAATCTTTT
3          A                              C          C          GGGA  T           C
4                   C                                C          GGGA  T           C
5                            G          A          A  C

2 GGATTGCTCA CCCAGGTGGT CCTGCAATTC TAGACCAAGT AGAACAAAAG CTAGCCTTGA AGCCTGAAAA GATGAAGGCC
3          A                              T          G                                      A T
4          A    T                    CT     T#
5                       A                              G          T G       A       A

2 ACTAGGGAAG TTCTAAGTGA ATATGGAAAC ATGTCAAGTG CATGTGTATT GTTCATCTTA GATGAGATGA GAAAGAAATC
3        A      C  C  C          T  T                    T          T            A
5        T    A              T

2 AGCTCAAAAT GGACTTAAGA CAACAGGAGA AGGACTTGAA TGGGGTGTAT TGTTTGGCTT TGGACCAGGA CTTACTATTG
3 A  A  G          G          T          T          G    C                    C
5                A    C          T          G                    T          C

2 AGACTGTTGT TTTGCACAGT GTGGCCATTT GA
3                    C  T  C
5 A        G    C A T    T  T  A
```

3' NON-CODING SEQUENCES

```
2 GAGGCTTACT TATTTTATTT TCATGTATCA TTTTTAAATT TGCTTGATTT TTATGTAACC GTGAAAAACT CATCTAGAGT
3 GATGCGCTTC ATTATTTTAT TTTCATGTAT TATTTTTAAA TTTGCTTGAT TTTTATGTAA GCATGGAAAA CTCATCTAGA
5 GATTGAGAGA CTTTTGATTG TTTTTTAATT GTATTGTATT ACTTTACTTT TAATTAATCT TGCTTGAACT TCCATTTAAG

2 TCAACATGGA CAATCATATT AAAATAATAT TTGATTAACT TTTGTATTCT AAGATAGTTA CTTACGTGAT AAAATATAAT
3 GTTCAACATG GACAATCATA TTAAAATAAT ATTCGATTAA CTTTGTATTC TAAGATAGTT ATGTGATAAA ATATAATTTA
5 AATAAATATG GAGTTCAATC TGTACCATCA ATGTTAAAAT AATATATCGT TAATAGCTTT TATTTATAGT GTCTGCTTTT

2 TTATGATTAA TAAGAAAATC GTTTTAATAT ATTAAAGATG ATGAAGATGA TTGTAACCAG TGAATTGTTA GAAATTCGGT
3 CAATTAATGA GAAAATCATG TTGACATATT AAAGACAATG AAGATGATTG TAACTGGTGA ATTACTATAT TTATAATTAA
5 TCTTCTTTCT AAACTATTTT ATTTTAGTAT TTGTTATTTA TTGATTTAAA ATAAACACTG TCCTCTTAAC TGATAAATAA

2 ATGATAATCC TGGATATCTT CGAAGAATCG TGTTCGTTCA CTTTCCGAAC AAAGTATTTC GACCCTATTC TTGTCTGGGT
3 TGAGAAAATC ATGCTCTTAT AACTTTTTTT TTCCAAGCAT CCACGAAAAG TGGTCTGTAT GATGTGACAA TGAGGTTGAG
```

```
          .              .              .              .              .
2  MVSVSEIRKAQRAEGPATILAIGTANPANRVEQATYPDFYFKITNSEHKVELKEKFQ
3              Q                         K
4                                        K
5        A    Q                          K

       .              .              .              .              .              .
2  RMC*DKSMIKSRYMYLTEEILKENPSVCEYMAPSLDARQDMVVVEVPRLGKEAAVKA
3                                    L
4
5

          .              .              .              .              .              .
2  IKEWGQPKSKITHLIFCTTSGVDMPGADYQLTKLLGLRPYVKRYMMYQQGCFAGGTV
3
4
5

       .              .              .              .              .              .
2  LRLAKDLAENNKGARVLVVCSEVTAVTFRGPSDTHLDSLVGQALFGDGAAALIVGSD
3
4
5

       .              .              .              .              .              .
2  PVPEIEKPIFEMVWTAQTIAPDSEGAIDGHLREAGLTFHLLKDVPGIVSKNIDKALV
3                V
4
5

       .              .              .              .              .              .
2  EAFQPLNISDYNSIFWIAHPGGPAILDQVEQKLALKPEKMKATREVLSEYGNMSSAC
3        G
4        G                             #
5                                                          D

          .              .              .              .
2  VLFILDEMRKKSAQNGLKTTGEGLEWGVLFGFGPGLTIETVVLHSVAI
3                TKD              D
5                                 D
```

Fig. 4.3 Predicted amino acid sequences for *Trifolium subterraneum* CHS2 (2), CHS3 (3), CHS4 (4) and CHS5 (5). The entire sequence of CHS2 is shown; in the case of CHS3, CHS4 and CHS5, only amino acid differences with respect to CHS2 are indicated. The last amino acid that could be predicted from the incomplete genomic sequence of CHS4 is indicated by a #. An asterisk shows the cysteine whose corresponding DNA codon is split between the first and second nucleotides by an intron in all four genes.
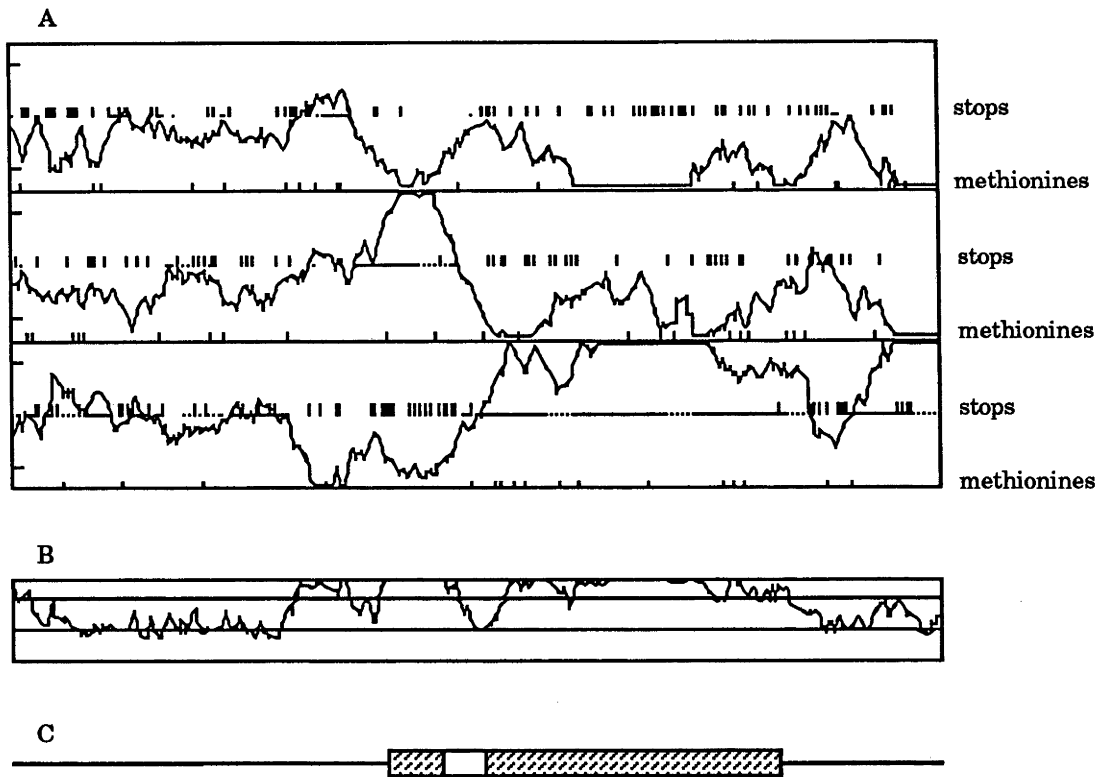
Fig. 4.4 Analysis of the nucleotide sequence of CHS2 shown in fig. 4.3 to establish the correct reading frames. (A) Probabilities of the three reading frames containing protein coding sequence as predicted by Staden's "Positional Base Preference" programme (Staden, 1984). In each box, the short vertical lines across the centre, and those along the bottom, represent stop codons and methionine residues respectively. The first exon occurs in the middle box, and the second exon in the lowest box; they are recognized by their high peaks. No coding sequence occurs in the top reading frame. (B) Calculation of the non-randomness of the sequence performed by Fickett's "Testcode" programme (Fickett, 1982). The presence of a gene is again seen by the high (off scale) peaks; these correspond to the two exons of CHS2. (C) Schematic representation of CHS2. The two exons are shown as shaded boxes, and the single intron as an unshaded box.

2 MVSVSEIRKAQRAEGPATILAIGTANPANRVEQATYPDFYFKITNSEHKVELKEKFQ
  HHHHHHHHHHHHHH   BBBBBB   TTTTTHHHHHH TTBBBBBBBHHHHHHHHHHHBB

2 RMC*DKSMIKSRYMYLTEEILKENPSVCEYMAPSLDARQDMVVVEVPRLGKEAAVKA
  BBB TTT   BBBBBHHHHHHH TTBBBBB TTHHHHHHBBBBBB   HHHHHHH

2 IKEWGQPKSKITHLIFCTTSGVDMPGADYQLTKLLGLRPYVKRYMMYQQGCFAGGTV
  HHHH   TTT BBBBBBBBTTT   TTTTBBBBBBB   BBBBBBBBBBBBBBTTBB

2 LRLAKDLAENNKGARVLVVCSEVTAVTFRGPSDTHLDSLVGQALFGDGAAALIVGSD
  BBBHHHHHHHTTTTBBBBBBB BBBBBBTTTTT   BBBBBBBTTTHHHHHH TT

2 PVPEIEKPIFEMVWTAQTIAPDSEGAIDGHLREAGLTFHLLKDVPGIVSKNIDKALV
  TT   HHHHHHHHHHHHHHHHTTTT   TTHHHHHHHHHHHHHHH   TTHHHHHH

2 EAFQPLNISDYNSIFWIAHPGGPAILDQVEQKLALKPEKMKATREVLSEYGNMSSAC
  HHHH   TTTT BBBBBBTTTHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH TT BB

2 VLFILDEMRKKSAQNGLKTTGEGLEWGVLFGFGPGLTIETVVLHSVAI
  BBBBBHHHHHHHHHTTTHHHHHHHHHH BBBBB TTBBBBBBBBBBBBB

Fig. 4.5. Secondary structure prediction for subterranean clover CHS2 using the method of Chou and Fasman (1978). H - a helix, B - b pleated sheet, T - turn.

4.2.2 Coding sequences and introns of clover CHS genes 2, 3, 4 and 5

The nucleotide sequences of the three "complete" genes, CHS2, CHS3 and CHS5, are approximately 92% homologous, while their derived protein sequences are over 97.5% homologous (fig. 4.6). The high homology between the predicted proteins of these three genes, and also that of the CHS4 gene, suggests that they are all functionally active. If one or more of these genes was not active it would be expected that the sequences would show greater divergence.

The great majority of the nucleotide substitutions between the four CHS genes are in the third positions of codons (as would be predicted), resulting in the amino acid differences between the genes being small in number (fig. 4.6/7). Interestingly, CHS2 bears greater homology to CHS5 than it does to CHS3, its close neighbour; this is discussed below (4.3). The GC contents of the coding sequences of CHS2, 3 and 5 are 41.5%, 42.8% and 43.4%, respectively. These values are slightly below those found for the CHS genes from several other dicotyledons (45.5-53.9%), but are obviously representative of these type of plants when compared to the figures for monocotyledons (65.7-69.3%; Niesbach-Klosgen et al., 1987).

Of the nine amino acid differences that do exist between CHS2 and CHS3, five are conserved (with respect to both charge and size) according to a group classification published by Miyata et al., 1979 (fig. 4.7). The partially available protein sequence of CHS4 (first 311 amino acids) shows two differences to that of CHS2, and three to that of CHS3; these differences involve only those positions *and* amino acids which are variant between the sequences of CHS2 and CHS3. Although the C-terminal portion of the CHS4 protein is not known, the high homology of the known amino acid sequence with those of CHS2 and CHS3 strongly implies that the corresponding CHS4 gene is not a pseudogene. The

derived protein sequence of the CHS5 gene from cluster II does not vary in any significant way from the sequences of the cluster I proteins; most of the differences it does have with respect to the latter sequences involve conservative substitutions (fig. 4.7).

	CHS2	CHS3
CHS3	91.9%/ 97.7%	
CHS5	93.1%/ 98.7%	92.1%/97.9%

	CHS2	CHS3	CHS4
CHS3	9/3/83/**95**		
CHS4	5/2/61/**68**	5/1/38/**44**	
CHS5	7/1/73/**81**	10/2/80/**92**	4/1/49/**54**

Fig. 4.6 Pairwise comparison of divergence among subterranean clover CHS2, CHS3, CHS4, and CHS5. The Upper table shows absolute percentage homologies between the three "complete" genes; the left figures are for the nucleotide (coding) sequences and the right hand ones for those of the derived proteins. The Lower table provides the data for nucleotide differences between all four genes. In each case the first three figures refer to nucleotide differences in the 1st, 2nd and 3rd positions of codons, and the last figure is the total differences between the two sequences.

Position in the sequence	Amino acid in CHS2	Amino acid in CHS3	Amino acid in CHS4	Amino acid in CHS5
* 5	S	S	S	A
9	K	Q	K	Q
* 30	R	K	K	K
* 83	V	L	V	V
238	E	V	E	E
291	N	G	G	N
* 329	E	E	–	D
* 354	A	T	–	A
355	Q	K	–	Q
* 356	N	D	–	N
* 366	E	D	–	D

Fig. 4.7 Amino acid differences between clover CHS2, CHS3, CHS4 and CHS5. The asterisks indicate those positions where the differences between the four sequences can be classified as conserved (with respect to charge and size) on the basis of the results arising from the study by Miyata et al., 1979. The dashes show the positions in the sequence of CHS4 for which the amino acid is not known.

The methionine start codons of CHS2, CHS3 and CHS4 are preceded in the -3 position by a purine (A G, and A, respectively), a feature which is characteristic of eukaryotic translation initiation sites in both plants and animals (Kozak, 1984; Joshi, 1987a). However, the sequences which surround the start codons of the three genes have only poor homology with a consensus sequence derived for the translation start regions of plant genes (fig. 4.8; Joshi, 1987a).

Consensus sequence for the'

translation start site   T A A A C A **A T G** G C T

CHS2   A T T A A C **A T G** G T T

CHS3   T A A G A T **A T G** G T G

CHS4   A T T A A C **A T G** G T T

Fig. 4.8 Comparison of the nucleotides surrounding the translation start codons of clover CHS2, CHS3 and CHS4 with a consensus sequence derived for such regions in plant genes (Joshi, 1987a).

The clover CHS introns are all greater than ~70bp and rich in AT nucleotides (~76%); both these features are characteristic of the introns present in the genes of dicot plants, and appear necessary for efficient intron excision from the pre-mRNA (Goodall and Filipowicz, 1989/1990/1991). The monocot maize, however, can splice out introns which are GC-rich (Goodall and Filipowicz, 1991). The actual sizes of the CHS2, CHS3, CHS4 and CHS5 introns - 182bp, 116bp, 140bp, and 109bp - respectively, are similar to those for the introns present in other legume CHS genes. For example, the intron of the CHS1 gene from soybean is 122bp (Wingender et al., 1989), while that from the CHS8 gene of bean is 103bp (Schmid et al., 1990). The large size of some of the introns in the CHS genes of *Petunia hybrida* (greater than 2000bp; Koes et al., 1989b) does not seem to be a characteristic found among the Leguminosae. A

second intron found in the coding sequence of the CHS gene from snapdragon is as yet an isolated phenomenon (Sommer and Saedler, 1986).

The nucleotides of the 5' and 3' splice junctions of the four introns conform closely to those of consensus sequences published for plant genes (fig. 4.9). A consensus sequence has also been derived for a putative plant intron "branchpoint", the region surrounding the adenosine residue which forms an ester bond with the phosphate group of the 5' terminus of the intron during the splicing process (Brown, 1986). However, as no plant branchpoint has (1988) been mapped in a homologous *in vitro* or *in vivo* system, the published consensus can only be accepted tentatively (Krainer and Maniatis, 1988). Nevertheless, sequences in the introns of the four clover CHS genes 2, 3, 4, and 5 match the given consensus (fig. 4.9).

Besides the splice junction regions, the sequences of the four introns share substantial homology, especially towards their 3' ends (fig. 4.10). As the intron of CHS2 is the longest of the six "complete' clover CHS genes (i.e. excluding CHS7 and 8, and including sequencing results for CHS1 and CHS6 by T. Arioli) it must be presumed that it has gained some DNA, or the other introns have lost some, since the duplication event that resulted in their origin as separate entities in the genome of subterranean clover. It is especially interesting that the intron of this gene has a direct repeat of 17bp (AAGCAACTTCTGATATT), the two sequences being separated by 13bp. A single form of this sequence occurs in the intron of CHS4. As transposition in plants is frequently characterized by the excision of the transposon from one location and its insertion in a new sequence (i.e. not replicative transposition), but the retainment of at least part of the duplicated repeat in the original position, the 17bp repeat present in CHS2 may indicate such an event

(Saedler and Nevers, 1985). This type of scenario has been proposed to explain the existence of repeats in the introns of Nicotianeae SSU genes (Jamet et al., 1991).

The lack of homology between the introns of clover CHS2, 3, 4, and 5 on the one hand, and those belonging to such plants as soybean (a fellow legume; Akada et al., 1990a/b;1991) and *Petunia hybrida* (a non-legume: Koes et al., 1989b) on the other, implies there is virtually no selection pressure to maintain any intron sequence (let alone length) across species.

Consensus sequence for 5'
exon-intron splice junctions: C/A A G : G T A A G T

CHS2   T   G T : G T A A G T

CHS3   T   G T : G T A A G T

CHS4   T   G T : G T A A G T

CHS5   T   G T : G T A A G T

Consensus sequence for 3'

intron-exon splice junctions:	T T T T/Pu	T T T/Pu	T/Pu	T/Pu	T/Pu	T G C A G : G
CHS2	T T A T	A A C	T	A	A	T T C A G : G
CHS3	T T A T	A A T	T	A	A	T T C A G : G
CHS4	T T A T	A A C	T	A	A	T T C A G : G
CHS5	T T T T	A A C	T	A	A	T T C A G : G

Consensus sequence for branchpoint   T/Pu T/Pu C T Pu A Py

CHS2	151 T	A C T A A T	157
CHS3	84 T	A C T G A T	90
CHS4	109 T	A C T A A T	115
CHS5	80 T	A C T A A T	86

Fig. 4.9 Comparison of the 5' and 3' splice junctions and putative branchpoints of CHS2, CHS3, CHS4 and CHS5 to consensus sequences published for plant exon/intron boundaries and intron branchpoints by Brown, 1986. Numbers for the branchpoint sequences refer to distances from the most 5' intron nucleotide.

```
CHS2          GCATCAAACAACATATTATA.....TATGTGATTGGCTTTGGT......ATTACC.
CHS3          GCATTAAAAGACATATTATCATTATTATGTGATTGACTTTGGT......ATTGCCC
CHS4                                    TGTGATTGTCTTTGGTACTACTATTATCC


CHS2          ATTATCCAAAAATTGTTATACTAATGTCATTTTTC.TTATAACTAATTCAG
CHS3          .......AAAAATTGTTGTACTGATGTCATTTTTTGTTATAATTAATTCAG
CHS4          .......AAAAATTGTTGTACTAATGTCATTTTTC TTATAACTAATTCAG
CHS5          AATATATTGTAGTACTAATATGATTTTTTTT...AACTAATTCAG
```

Fig. 4.10 Alignment of the last 94 nucleotides of the introns of CHS2 and CHS3, and the last 72 and 42 nucleotides, respectively, of the introns of CHS4 andCHS5. The dots represent spaces placed in to improve the homology.

A codon frequency table compiled from the clover genes CHS2, 3, 4 and 5 is presented in fig. 4.11. As has been observed for all organisms, in the case of those amino acids which have more than one triplet coding for them, there is a pronounced preference for one or more of these to be used over the others (Grantham et al., 1986). However, it would be unwise to conclude that the codon usage table in fig. 4.11 is indicative of clover as a whole as it is based on such a small number of sequences, all of which come from a single gene family. Nevertheless, some interesting features can be observed in the codon table. For most of those amino acids which have four triplets coding for them which differ in the last base (alanine, glycine, leucine, proline, threonine, valine), there is an overwhelming preference for the final position to be occupied by an adenine (A) or thymine (T). The GCG and CTG codons for alanine and leucine, respectively, are especially avoided. This trend does not hold for arginine, where among the CGN codons, those ending in thymine and cytosine (C) are favoured.

GGG	(G)	11	GAG	(E)	49	GCG	(A)	0	GTG	(V)	31
GGA	(G)	45	GAA	(E)	55	GCA	(A)	47	GTA	(V)	29
GGC	(G)	13	GAC	(D)	20	GCC	(A)	26	GTC	(V)	12
GGT	(G)	45	GAT	(D)	55	GCT	(A)	62	GTT	(V)	49
AGG	(R)	12	AAG	(K)	54	ACG	(T)	4	ATG	(M)	49
AGA	(R)	14	AAA	(K)	54	ACA	(T)	34	ATA	(I)	5
AGC	(S)	2	AAC	(N)	26	ACC	(T)	11	ATC	(I)	25
AGT	(S)	39	AAT	(N)	13	ACT	(T)	31	ATT	(I)	59
CGG	(R)	0	CAG	(Q)	12	CCG	(P)	3	CTG	(L)	1
CGA	(R)	0	CAA	(Q)	42	CCA	(P)	36	CTA	(L)	22
CGC	(R)	9	CAC	(H)	12	CCC	(P)	1	CTC	(L)	15
CGT	(R)	20	CAT	(H)	15	CCT	(P)	37	CTT	(L)	36
TGG	(W)	15	TAG	(*)	0	TCG	(S)	0	TTG	(L)	40
TGA	(*)	3	TAA	(*)	0	TCA	(S)	21	TTA	(L)	22
TGC	(C)	4	TAC	(Y)	27	TCC	(S)	1	TTC	(F)	22
TGT	(C)	19	TAT	(Y)	16	TCT	(S)	19	TTT	(F)	31

Fig. 4.11 Codon frequency table compiled from subterranean clover CHS genes 2, 3, 4, and 5. The one letter abbrieviations for the amino acids are given after each codon.

4.2.3 The 5' and 3' flanking sequences of clover CHS genes 2, 3, 4 and 5

The promoters (5' flanking sequences) of CHS2, CHS3 and CHS4 share substantial homology in five regions of the approximately 1000bp of each that was sequenced (fig. 4.12). As would be expected, one of these regions lies in between the presumptive TATA boxes and the translation start sites. The other four regions occur further upstream, in the same order, but not in the same absolute positions with respect to the TATA boxes. The region numbered "1" begins at the most 5' nucleotide of CHS2 which was sequenced, implying that the homology between the three promoters extends further than a kilobase upstream. The CHS genes 1 and 2 of soybean also share large regions of homology between their promoters (Wingender et al., 1989). Frequently the regions of homology between two of the genes, but not the third, extend well beyond the "ends"

```
1
Ts2               -1100 GATTTTGTTTTTAACA.TTTGACTATGACAGCTTTGGTGACTAAATTAATGTCCTTG
Ts3   -941 TGTTTTGTAGTATGTTGCTTTTAACAATTTGACTATGATAGATTTAGTGACTAAATTAATGTCCTTG
Ts4   -947 TGTTTTGTAGGATGGTTTTTTTAGCAATTTGAATATGACAGCTTTAGTGACTAAATTAATGTCCTTG
con        TGTTtTGTaGgATgtTgTTTTTAaCAATTTGAcTATGAcAGcTTTaGTGACTAAATTAATGTCCTTG

Ts2 TCCATGTTCATCCCTTTTAATCTTT   -1020
Ts3 TCCATGTTGATCCCTTTTAATCTTT   -850
Ts4 TCCATGTTCATCCCTTTTAATCTTT   -856
con TCCATGTTcATCCCTTTTAATCTTT

2
Ts2 -869 ATAAACTGTTATTATAAGTTTTCAAAAACAGTGTCACAAGTATTTATTTCAGTAGATAAACTCAAATA
Ts3 -645 ATAATCCGATATTATATGCTCTACAAAATAGTGTCAG.AGTGTTTATGTTAGTAAAAAAATTCAAATA
Ts4 -593 ATAATTCAATATTATATGTTCTACAAAATAGTGTCAG.AGCGTTTACGTCAATAGAAAAATTCAAATA
con      ATAAtccgaTATTATAtGtTcTacAAAAtAGTGTCAg.AGtgTTTAtgTcAgTAgAaAAAtTCAAATA

Ts2 AGTTATTTCAAA -790
Ts3 AGTATATTCAAA -567
Ts4 AATAAATTCAAA -515
con AgTaaaTTCAAA

3
Ts2 -778 AATGATATTTGAACACTCTTATTTTTCCGACACTCATTCGCTCTAT..CTTTT -728
Ts3 -549 AATGATATTTAGACACTCTTAATTTTCACT.ACTCATTCTATATATTTCTTTT -498
Ts4 -498 AATGATATTTAGACACTCTTAATTTTCGCC.ACTCATTCTCTCATTCTTTTTT -447
con      AATGATATTTagACACTCTTAaTTTTC.c..ACTCATTCtcT.taT.TcTTTT

4
Ts2 -345 TTTGTGATGTATGTAACTTGAGATTTAACAAAAATAAGTGCACAAATAAAACAAAATGCTGGTCAAATT
Ts3 -391 GTTGTGTTGCACGTGACTTAAGATTTCACAAAAATAAGTGGATAAATAAAACAAAATACTAGTCAAATT
Ts4 -380 GTTGTGTTGCATGTGACTTAAGATTTAACAAA..TAAGTGCAAAAATAAAACAAAAGACTAGTCAAATT
con      gTTGTGtTGcAtGTgACTTaAGATTTaACAAAAATAAGTGcA.AAATAAAACAAAAtaCTAGTCAAATT

Ts2 GTTTACAAAATAATAATTAAAATAAAAG.............AGGGTCTTCAAAATCATGGTGGTAGTTGCCAC
Ts3 GTTTACAAAATAATAATTAAAATAAAAGTGGTTGGACTAGAGAGGGTCTTCCAAGTCATGCTGCTATTTTCAAC
Ts4 GTTTACAAAATAATAATTAAAATAAAAGGGGGTTGGACTAAAGAGGGTCTTTGAAGGCATGCGGGTATTTTCCAC
con GTTTACAAAATAATAATTAAAATAAAAG.GGTTGGACTA.AGAGGGTCTTc.AAgtCATGctGgTAtTTtCcAC

Ts2 CAGACTC.CTACATGTCAC -199
Ts3 CAAACTC.CTACATGTCAC -231
Ts4 CAAACTCACTACATATCAG -221
con CAaACTC.CTACATgTCAc

5
Ts2 -128 TATATATAACCATCTCATCCCACAGATCTTTTCACCACAACACACTTCAGCCTTTG.CTTTTCTACAA.
Ts3 -117 TATATATAACCATCTCATCCCACACA..TTTCCACCACAACACTCTTCATCTCATTCCTTTTCTATAA.
Ts4 -128  TATAAAACCATCTCACCCCACACATCTTTCCACCACAACACTCTTCAACCTTTG.CTTTTCTACAAA
con      TATATAtAACCATCTCAtCCCACACATCTTTcCACCACAACACtCTTCA.CcttTg.CTTTTCTAcAA.

Ts2 CAACTTCTGTAATAAACCACTATT.GAGTTCAATTTA..CATAGAAACTACTAAAGATATTAAC -1
Ts3 AAACAACTGT................GTTCAAATTACACATTGAAACTACTAAAGATATATTAAGAT -1
Ts4 CAACTTCTGTAATAAACCATTATTTGAGTTCGAGTTA..CATAGAAACTACTAAAGATATTAAC -1
con cAACttCTGTAATAAACCA.TATT.GAGTTcaA.TTA..CATaGAAACTACTAAAGATATtaac...
```

Fig. 4.12 Alignment of those promoter sequences (five in number) of CHS2 (Ts2), CHS3 (Ts3) and CHS4 (Ts4) which share the greatest homology. Numbers refer to the positions of the adjacent bases with respect to the start points of translation. The dots are placed in the sequences to represent the best alignments. A consensus sequence (con) is provided beneath. It should be noted that often the homology between *two* promoters alone extends outside the regions shown.

shown in fig. 4.12.

The approximately 600bp of 5' flanking sequence of CHS5 which was established does not possess the same five large areas of homology. There are, however, some stretches of sequence, each about 20bp, which are obviously homologous between the promoters of CHS5 and CHS2; these are all located very close to the respective TATA boxes. Nevertheless, the striking finding from this analysis is undoubtedly the high homology of the promoters of the three cluster I genes, especially when compared against the cluster II CHS5.

At the other ends of the coding sequences, CHS 2 and 3 share a high degree of homology for over 200bp following their respective stop codons (fig. 4.13). This phenomenon is also observed with the two above mentioned soybean genes (Wingender et al., 1989). Yet again, the analogous region of CHS5 bears little homology to the cluster I genes.

```
Ts2  GAG..GCTTACTTATTTTTATTTTCATGTATCATTTTTAAATTTGCTTGATTTTTATGTAACCGTGAAAAACTCA
Ts3  GATGCGCTTCATTATTTTTATTTTCATGTATTATTTTTTAAATTTGCTTGATTTTTATGTAAGCATGGAAAACTCA

Ts2  TCTAGAGTTCAACATGGACAATCATATTAAAATAATATTTGATTAACTTTTGTATTCTAAGATAGTTACTTACG
Ts3  TCTAGAGTTCAACATGGACAATCATATTAAAATAATATTCGATTAACTTT.GTATTCTAAGATAGTTAT....G
```

Fig. 4.13 Comparison of the first 146 and 143 nucleotides downstream form the stop codons of CHS2 (Ts2) and CHS3 (Ts3), respectively. Gaps are introduced to maximalize alignment. The homology between the two sequences continues for about a further 80 nucleotides beyond that shown here.

The sequence AATAAA is required for the efficient polyadenylation of mammalian mRNA molecules (Proudfoot, 1984). In plants, an unaltered AATAAA sequence is rare, the last A often being replaced by a different nucleotide (Joshi, 1987b). Furthermore, in some plant genes this sequence, or a close variation of it, is absent altogether (Mogen et al., 1990). CHS2 and CHS3 each have a putative AATAAA-like signal, AATAAT, beginning 103 and 105 nucleotides downstream from their

respective stop codons. Both the sequences AATAAA and AATAAT occur in the 3' flanking sequence of CHS5, situated 81 and 118 nucleotides, respectively, downstream from the stop codon. A sequence CGTGTTCG, found 269 nucleotides away from the stop codon of CHS2, bears a striking resemblance to YGTGTTYY, a (consensus) sequence which is believed to play a role in 3' end formation in mammalian mRNAs (McLauchlan et al., 1985; Joshi, 1987b). Besides the necessity of the AATAAA sequence for the termination of transcription in at least some plant genes, there appears to be no other *highly* conserved sequences that play such a role in plants. Nevertheless, elements that are essential for the formation of correct 3' termini in individual plant genes have been identified (Mogen et al., 1990).

4.2.4 Comparison of the clover CHS genes (and derived protein sequences) with those from other species:
(a) coding sequences

Much of the discussion in this section uses CHS2 as a "model" subterranean clover CHS gene. A comparison of the nucleotide and amino acid sequences of CHS2 with CHS genes from bean, soybean, snapdragon, parsley and *Petunia hybrida* is presented in figures 4.14 and 4.15. In the case of bean, soybean and *P. hybrida*, which all possess large CHS multigene families, only one member of each family was selected for this purpose. The plants from which the CHS sequences are derived represent a broad cross-section, from both within and outside the Leguminosae. As would be expected, the nucleotide and predicted protein sequences of the CHS genes isolated from the legumes bean and soybean share the greatest homology to those of the CHS2 gene of clover (fig. 4.16; see below).

```
Ph                                                ATGGCAAATCATCAT

         *         * ** *       * *     **      **    ** ** **    *    *  *
Ts2   ATGGTTAGTG TGTCTGAAAT TCGTAAGGCT CAAAGGGCAG AAGGCCCTGC AACCATTTTA
Pv17       G        A     G  C AC                    T  A           CC T
Gm         G        TGAA  G             G        C T          C   TG CA G
Am         G C      TGAG  GG        G         G      TG  G A G C   AG    G
Ph    AT CAGAAA     TGAG     CA G  CAGG       G  A  CC      T      C AT  AC
PhyA       G CA     CGAG  GTA           A       C T  T      T  A   C   TG CA G

      * ** ** * ** ** **  * *        ** **    ** ** **
Ts2   GCCATTGGTA CTGCTAATCC AGCAAACCGT GTTGAGCAGG CCACATATCC TGATTTTTAC
Pv17         A       A CC   T T    T        T  A G          C    AC
Gm      T    C    C  C C    TC C    T C   G  T   A GT  C        C A     T
Am      G    A    A    C    T  C  TT C       T  AA G   T            A   T
Ph      T    C       C C    TT C    T     CT C     TGAT           AC
PhyA         A    A  C CA   TA     T         T  AA G   T  C        A    T

      **    ** *  * **     **   ***     *  ** ** **    ** ** **  * * *****
Ts2   TTCAAAATTA CTAACAGTGA GCACAAGGTT GAGCTCAAAG AGAAATTCCA GCGCATGTGT
Pv17       G  C  A            T ACC   C              G  T
Gm         CGC C    C    C    T ACC              A      A A
Am         CG  C       T    A  T ACG          G  A      A            C
Ph         TCGT        C   TCC    A  T ACC    T      C TC      A      A
PhyA       TCGT C                   AC   T  T  G         TA

      ** ** **      **     *    * ** ***  ** * ** * * **      *     ** **
Ts2   GATAAATCTA TGATCAAGAG TAGATACATG TATCTAACAG AAGAGATTTT GAAAGAAAAT
Pv17      C  G  G       A   A G     T   C  T G AC  G     AC    G G
Gm          G  G       T   A GC             CT  AC          CC   G G
Am             A  AT        AC T        C  CT   C     A   CC      G  C
Ph       A     A           GA A GC T    C  A         G ATA C        C
PhyA     A     A       T    A A  G      C  CT         G A  C      G

      **    * * * *  **     *** ** **  ** ** **    * ** **   *  ** ** **
Ts2   CCAAGTGTTT GTGAGTACAT GGCACCTTCA TTGGATGCAA GACAAGACAT GGTGGTGGTA
Pv17    T ACA G      CT              T          G  G           A
Gm      C           CA  T           G          G              TA G
Am      TGCCA G    C              C    T C    C G             T    C G
Ph      C A    A   C CA    GA         C    C    C TC  C      C    T    G
PhyA    T   A G      A              T C T      T    G         A        T

      ** ** **      ** ** **  *  ** **      ** ** ** ** ** **  *** ** **
Ts2   GAAGTACCTA GACTAGGAAA GGAGGCTGCA GTGAAGGCTA TTAAAGAATG GGGACAACCA
Pv17    G      A AGT    G   A              C    A G  G       G
Gm      G      A AGT G      A          ACT    A  C G         T       C
Am      G T  C  T C  C  C   A A     C  CAA A CA        G     C  G   C
Ph      G C  A  G       C   A A        TCC  A  C  C      G   C       T
PhyA    G C     AA  T  C    A     A  C CAA        C  G       C  G   C

      **      *  * ** **  *   * ** *** ** **    *  * ** ***** *** **
Ts2   AAGTCAAAGA TTACTCACTT AATCTTTTGT ACTACAAGTG GTGTAGATAT GCCTGGAGCT
Pv17             A          G A    C   C  C           C C           T
Gm        C          C  TC C          C    T         C C           T
Am      A C  A  C C         GG       C C  C C  C      C C       C G A
Ph      A C     C        C  C T C  C  C  TTC          G C          C
PhyA      C  A     C  T     GT       C  A  T          G C          GTG

      ** ** ** *  * ** **    ** ** **  ** *  **      ** *   ********
Ts2   GATTACCAAC TTACAAAACT CTTAGGTCTT CGCCCGTACG TAAAAAGGTA TATGATGTAC
Pv17         G   C  C          G  A    G  C  T  G  G          C
Gm        T  G   C  T        A     C   T  C C     C  GC T     C
Am        T  G   C  C          C  C  A       T C  C    C T T C
Ph      C    G   C  C   G     C T  C  C  T  C CT  C  GC C T C
PhyA    C  T     C  T  G    AC T  G       T  A CG  C  G    CT C
```

Fig. 4.14 (This page and over) Alignment of the nucleotide coding sequences of the CHS genes from *Trifolium subterraneum* (Ts; CHS2), bean (Pv17; Ryder et al., 1987), soybean (Gm; Akada et al., 1990), snapdragon (Am; Sommer and Saedler, 1986), parsley (Ph; Reimold et al., 1983), and *Petunia hybrida* (PhyA; Koes et al., 1989b). Note that the parsley sequence is longer than the others; in order to show that it is homologous to the other CHS sequences, it is represented as beginning earlier and ending later. Only nucleotide replacements with respect to the *Trifolium* sequence are given. The dashes represent the deletion of three nucleotides in the soybean sequence. The stop codons are underlined. The asterisks indicate those nucleotides which were found to be conserved in a comparative study of eight CHS nucleotide sequences, including those of Am and Ph (Niesbach-Klosgen et al., 1987); this "consensus" is hence *always* represented by the Am/Ph nucleotide.

```
        ** ** ** *  **** ** **  ** ** **   ** **  * * * ***** *   ** ** ***
Ts2     CAACAAGGTT GTTTTGCAGG AGGCACGGTG CTTCGTTTGG CTAAAGATTT GGCTGAGAAC
Pv17             A  C                  T      A        C G
Gm                C  C         C    T                   C        CC  C      A
Am             G  C         T    T T         G    A     G  G  C      G  A
Ph             G  C         T    C    T C    C    C     G  CC  C  C
PhyA              C         T    T         T       G  A  C  G  C      G  A


        **    ** * * ** **    *  ** ** **  ** ** ** * * ** ** **  ***** **
Ts2     AACAAAGGTG CTCGTGTGCT TGTTGTTTGT TCTGAAGTTA CCGCAGTCAC ATTCCGCGGT
Pv17             C            G G         GA C    T     G  T      T G
Gm           G         C         C            GA C                        C
Am         TGCC              T  G  C          C   A C   T  C T  T     G  C
Ph         TGCC      A  A  C  C       C  C    C    GA C  T  C     T     T  C
PhyA        G  C      A  C                    A   A C   G        C     T  G


        **       *    ** * **        * *** ** ** **  * ** ** **  ** ** *
Ts2     CCTAGTGACA CTCATTTAGA CAGTCTTGTT GGACAAGCAC TATTCGGAGA TGGAGCTGCT
Pv17    A          C  CC            G        G  T  G T            A
Gm      A C        C   C T  T  C       G        CT  G  T         T A  C
Am      GGCG          C  G  T  T G  G     T       TT  G      T   T A  A
Ph       TC    T      CC T  TTCC  A     T  G      T  T  T       T A
PhyA    A A    T        G     T   T A    C        C  T  T  T   G  A GC


        **    *     * **  * ** **    ** **  * ** ** **  * *  *
Ts2     GCTCTTATTG TTGGTTCTGA TCCAGTGCCA GAAATTGAAA AACCAATATT TGAGATGGTT
Pv17    AG G             C   A T   C G       G   G  TT G       T
Gm       G C         A  A    C  CT A   ---G          G  TT G  C  C T  C
Am      GG    A       A  C   C  G  TGTC  GGG G  GC GC GC T   CC     C
Ph      AG C    C  C  G  A      G AT T  TCCG G  GC GT   C T   C  C CA A
PhyA    GA C    A  A           A T       G G C  G  GG  TT G  C    C C


          *  *     *  * *  * *       ** **   * ** ** * * ** ** **  *  * **
Ts2     TGGACTGCAC AAACAATTGC TCCAGATAGT GAAGGAGCCA TTGATGGTCA TCTTCGTGAA
Pv17             G  C              C        T  T  T            A  C      C
Gm               C  G    CCT       C          G  T       A    C     C
Am      ACCG G  G       GC  CT G  C  C    C T  T  A     C  G    T GA G
Ph      CCG G  C              TT A  C  CTC    C  G  G     C  G       C
PhyA    CAG A  C       TC  CT C         C  CT  T  T       C         C


        ** ** ** * * ** **   *   * ** **   ** *  **  * ** ** **  ** **
Ts2     GCTGGACTAA CATTTCATCT TCTTAAAGAT GTTCCTGGGA TTGTCTCAAA GAACATTGAT
Pv17    T       C  G    C   C    C  G                                   GA
Gm      T  T  C  T  C       C  C  G         AC  CA    C      T        G
Am      TG  G  G  T       T  GT G  G      C    C  GA  G               G
Ph      TG  C  T  C  C               A  G  CT  AA    G  A  TA A     A
PhyA    T  G  T     C  CT  A  C       C  C T  T  T     C  GA      A  T   G


              *      ** ** **     ** *  ** **    * *****    * ** ***
Ts2     AAAGCATTGG TTGAGGCTTT TCAACCATTA AACATCTCTG ATTACAATTC AATCTTTTGG
Pv17     G  C TT        C    CA C    G         A        C    C  C      C
Gm       G  C        A  C    C     C  G GGA    C         T    CA
Am      GAGT    A AG         G  C  G  G GGT  T  G     GG     GG G
Ph      GT G    A AG  A      GGG  TA    GG    A  A  C GG  C   TT A
PhyA    GAGCC T  AG  A  A    A    T G GG    T         GG  C   TC A  C


        * ** ** ** *  * ** **  ** ** **   * **     ** * *       * *
Ts2     ATTGCTCACC CAGGTGGTCC TGCAATTCTA GACCAAGTAG AACAAAAGCT AGCCTTGAAG
Pv17        A       T     A            G         T  G      T  G GTC     A
Gm         A       T     A  C     TG        G    GGCT T  G   G GG      G
Am            T    G     C  G T    C      G  T  GG G   T   G GG       G
Ph      A  C  T   G     C  A T   TG   T G          CTG    T G GCT  A
PhyA          T         G         T G        T     AT     T G GCA


          *        *     * **   *  * **    *    * * * ** ** ** *** ** **
Ts2     CCTGAAAAGA TGAAGGCCAC TAGGGAAGTT CTAAGTGAAT ATGGAAACAT GTCAAGTGCA
Pv17                     A  G A T        A T  G    G       T     G       G
Gm           A       G A  T    C T  G    C  C G      T
Am        C  G  C    G T G  CC AC G  A    G      G     G       G     T
Ph      GAA    A      CG    CC AC G  G T G    T
PhyA      C  G  AC  T     T   A   AT G T      C        T            T


        ** **  * * * **   * **  ** *** *  **      **    * ** **      **
Ts2     TGTGTATTGT TCATCTTAGA TGAGATGAGA AAGAAATCAG CTCAAAATGG ACTTAAGACA
Pv17       GC A         G         G  G              G      G        A
Gm         GC A         G     C A  C G       A TAG                 GGC
Am         GC      T  TC C           G    TCT  G  GA GG A   GA G GC  T
Ph         G       T  T          A        G  TA T G  G A     GAAAGCA
PhyA        C      T  T  G        A        GCC        CA  G A   TT AGGA  T


        ** ** ** * *  * ** **  *** ** *  ** ** ** *  * ** **  *  ** ** **
Ts2     ACAGGAGAAG GACTTGAATG GGGTGTATTG TTTGGCTTTG GACCAGGACT TACTATTGAG
Pv17          T                         T          T          T      C C
Gm      C  C        C     C            GC A    T  C T  T  T   C  G
Am      T          GT A  T    T       C  G  C  G  G  T       C  GG  C
Ph      T          TT G  T    TC T   C  G  C  G  G  T        GG  A
PhyA    T  T       G     G    TC T        A      G  T  G   A  AG


        ** ** **  * ** **  *   *
Ts2     ACTGTTGTTT TGCACAGTGT GGCCATTTGA
Pv17      C  C    C  C         C  A ATAA
Gm               AC  C G      CA TG CTAA
Am      G  G  GC  G  T        C GC AATTAA
Ph      C       G       T  C  AC  GCTACTTTTACTCACTGA
PhyA          CC  C           T  T C TAA
```

```
            *   * * ***     ******   * *  ***** * ** * *    ** ******
Ts2    MVSVSEIRKA QRAEGPATIL AIGTANPANR VEQATYPDFY FKITNSEHKV ELKEKFQRMC
Pv17            Q                        T S C  D S   Y   R      MT D
Gm        E                  VM          T P C  D S   Y   R      MT      K
Am      T E V R              V           T   C  D S   Y   R      MT      K
Ph   MANHHNAEIE  NR   Q   N              T S C  Y AD  Y   R      MT D  L K
PhyA    T E Y                VM          T T C  D S   Y   R        T D   K


            ** *  *** * **    ** *  * * *** ****** *** *** *** **  ******* *
Ts2    DKSMIKSRYM YLTEEILKEN PSVCEYMAPS LDARQDMVVV EVPRLGKEAA VKAIKEWGQP
Pv17          K    H N        NM A          I     K
Gm           K     N          A             M   K       T
Am      N R   H            Y  AM            I           Q
Ph    E  RK   HI   Y          N A E         L           S
PhyA  E  K    H               M             I     K     Q


          * **** ** ********   ****** *** ***  * *** ********** ** *******
Ts2    KSKITHLIFC TTSGVDMPGA DYQLTKLLGL RPYVKRYMMY QQGCFAGGTV LRLAKDIAEN
Pv17
Gm                                     S
Am               V                     S  F                  M
Ph                                     S  F
PhyA            F           C          S  L


          * ******** **********   *   ****** *********  *   * **     *****
Ts2    NKGARVLVVC SEVTAVIFRG PSDTHLDSLV GQALFGDGAA ALIVGSDPVP EIEKPIFEMV
Pv17             I                               V    I Q     L L
Gm              I          T                     V    L -V   L QL
Am      A       I          A                     V      V GV R L QI
Ph      A       I            S                   V L   DL SV R L QLI
PhyA            I          N                G  I I    I GV R L L


           *   ***  ********* ********** * *******   * ** *  **  **  **
Ts2    WTAQTIAPDS EGAIDGHLRE AGLTFHLLKD VPGIVSKNID KALVEAFQPL NISDYNSIFW
Pv17            D          V                G  F N
Gm         L              V          LI    E          G         R
Am      TA LL   H          V          LI    E S K D   G    W V
Ph    SA L   D            V           LI    E S K G I G  W L
PhyA  SA LL   H            V           LI    E S E  K   G    W L


          *** ***** * ** * *     ** * ** ******* ***** ****  * *   *  *
Ts2    IAHPGGPAIL DQVEQKLALK PEKMKATREV LSEYGNMSSA CVLFIIDEMR KKSAQNGLKT
Pv17           G          D    D   G          R EK
Gm        A G          E   H                Q    IE   G
Am        E G          LRS Q                S  KE MS
Ph        L G   E    R   Q  D                    IEE KA
PhyA      I G          L   N   D              A  KE G


          ** * **** ********** ******
Ts2    TGEGLEWGVL FGFGPGLTIE TVVLHSVAI
Pv17
Gm        D          V    R TV
Am        D          V    PLN
Ph        D          V    PATFTH
PhyA                 V      T
```

Fig. 4.15 Alignment of the predicted amino acid sequences for CHS proteins from *Trifolium subterraneum* (Ts2; CHS2), bean (Pv17; Ryder et al., 1987)), soybean (Gm; Akada et al., 1990), snapdragon (Am; Sommer and Saedler, 1986), parsley (Ph; Reimold et al., 1983) and *Petunia hybrida* (PhyA; Koes et al., 1989b). Only amino acid replacements with respect to the *Trifolium* sequence are shown. An amino acid deletion in the soybean protein is represented by a dash. The asterisks indicate those amino acids which were found to be conserved in a comparative study of the deduced amino acid sequences of eight CHS nucleotide sequences, including those of Am and Ph (Niesbach-Klosgen et al., 1987); this "consensus" is hence *always* represented by the Am/Ph amino acid.

	% Homology at the nucleotide level	% Homology at the amino acid level
Pv17	82.5	90.4
Gm	77.7	87.7
Am	71.0	82.5
Ph	68.7	81.2
PhyA	74.5	83.8

Fig. 4.16 The degree of homology, at both the nucleotide and amino acid levels, between subterranean clover CHS2 *and* the CHS genes of bean (Pv17; Ryder et al., 1987)), soybean (Gm; Akada et al., 1990a), snapdragon (Am; Sommer and Saedler, 1986), parsley (Ph; Reimold et al., 1983) and *Petunia hybrida* (PhyA; Koes et al., 1989b). The genes selected from the multigene families of bean and *Petunia* are as indicated. The number of nucleotides/amino acids identical to the respective reference sequences of CHS2 are expressed as a percentage of the total number of nucleotides/ amino acids compared. Only the coding region of CHS2 (excluding the stop codon; 1167 nucleotides/389 amino acids) was used in the comparison.

The comparison of the amino acid sequences presented in fig. 4.15 shows that the highest regions of conservation occur in positions ~90-~220, ~250-~280, and at the C-terminal end of the protein. Presumably these stretches of sequence are those which are essential to the catalytic activity of the CHS protein.

In a previously published study, the nucleotide and predicted amino acid sequences for eight CHS cDNAs (none of which were derived from legumes) were examined and "consensus sequences", constituting the conserved bases/residues, were deduced (Niesbach-Klosgen et al., 1987). The nucleotide and amino acid sequences of clover CHS2, bean CHS17 (Ryder et al., 1987), and a soybean CHS (Akada et al., 1990a) have few differences to these sequences (figs. 4.14, 4.15 and 4.17). At both the nucleotide and amino acid levels, the respective sequences of clover are

| | | Nucleotide/amino acid sequence of: | | |
		Ts2	Pv17	Gm
Total differences to the Consensus:		24/14	18/11	12/8
Differences to the Consensus that agree with those of:	Ts2	–	–	–
	Pv17	13/8	–	–
	Gm	7/3	7/3	–
Differences to the Consensus that all 3 agree on:		5/2		

Fig. 4.17 Table comparing the nucleotide and predicted amino acid sequences of clover CHS2 (Ts2), bean CHS17 (Pv17; Ryder et al., 1987), and a soybean CHS (Gm; Akada et al., 1990a) with CHS consensus sequences (for both the DNA and protein levels) derived from examining eight CHS cDNAs from seven plant species (Niesbach-Klosgen et al., 1987). In each case, the number on the left refers to the nucleotide differences, and that on the right to the amino acid differences. The nucleotide consensus sequence consists of 600 bases, while that for the protein is 270 amino acids. These consensus sequences are indicated in figures 4.14 and 4.15.

Position in the sequence of CHS2	Amino acid in Consensus	Amino acid in CHS2
26	T	N
39	Y	F^*
57	K	Q
71	H	Y
153	S	Y
193	I	V^*
234	R	K^*
236	L	I^*
261	V	A
274	L	I^*
275	I	V^*
291	G	N
295	W	Y^*
379	V	I^*

Fig. 4.18 Amino acid differences of clover CHS2 from a published consensus sequence for the predicted CHS proteins from seven plants (Niesbach-Klosgen et al., 1987). The asterisks indicate those amino acids which are conserved (with respect to charge/size) according to a study by Miyata et al., 1979.

approximately 95% homologous to these (no-longer) "consensus" sequences. According to the group classification of amino acids mentioned above (4.2.2; Miyata et al., 1979), over half of the differences which the predicted protein sequence of CHS2 has to the consensus of Niesbach-Klosgen et al., 1987 are conserved (fig. 4.18).

Arabidopsis thaliana, mustard (*Sinapis alba*), and *Matthiola incana*, share five amino acid differences to the above protein consensus sequence (Epping et al., 1990; Ehmann and Schafer, 1988). On the basis of these observations, it has been proposed that there are some "Brassicaceae-specific (CHS) features" (Ehmann and Schafer, 1988). The two amino acid differences to the consensus which are shared by CHS2, the bean CHS17, and the soybean CHS protein (fig. 4.17; at positions 234 and 295 in the clover sequence), are also found in CHS3, 4 and 5, and in the bean CHS1, 4, 5, and 14 proteins (Ryder et al., 1987). Although this number of shared differences between the three legumes is only two, they may nevertheless possibly constitute "Leguminosae-specific features".

Of special interest is that a comparison of the number of differences between the nucleotide and derived amino acid sequences of the three Leguminosae CHS genes reveals that clover and bean have the greatest homology, and of these two plant species the former diverges more from soybean (fig. 4.19). The soybean CHS proteins also contain a deletion of one amino acid at position 231 (Akada et al., 1990a/b; 1991), a feature not observed in the bean (Ryder et al., 1987) and clover CHS proteins (or alternatively the latter proteins contain a one amino acid insertion). This relationship of the three Leguminosae CHS genes is reflected in their respective differences to the consensus sequences of Niesbach-Klosgen (fig. 4.17). The nucleotide and amino acid sequences of soybean are closest to the consensus sequences, whereas those of clover contain the

greatest number of differences; bean lies in-between the two. It is not clear, however, why the sequences of soybean (instead of those of clover) should bear the greatest resemblances to the two consensus sequences. Such an analysis would suggest that bean and clover possess a more closer evolutionary relationship to each other than either to soybean. This does not, however, correspond to the findings of classical taxomony; both bean and soybean are grouped in the tribe Phaseoleae, whereas subterranean clover occurs in the tribe Trifolieae (fig. 4.20). Any understanding of the evolutionary origins of the various members of the Leguminosae from the point of view of DNA sequence would, of course, have to be based upon much more data than that analysed here.

```
                                      Nucleotide/amino acid sequence of:


                                        Ts2        Pv17        Gm


Differences to the
nucleotide/amino acid
sequence of:            Ts2            –           –           –
                        Pv17           204/37      –           –
                        Gm             260/48      229/42      –
```

Fig. 4.19 Table comparing the number of differences between the nucleotide and predicted amino acid sequences of clover CHS2 (Ts2), bean CHS17 (Pv17; Ryder et al., 1987), and a soybean CHS (Gm; Akada et al., 1990a). In each case, the number on the left refers to the nucleotide differences, and that on the right to the amino acid differences.

FAMILY   Leguminosae
                                               |
SUBFAMILY   Papilionoideae
                          ┌────────────────────┴────────────────────────┐
TRIBE   Trifolieae   Phaseoleae
                          | ┌──────────┴──────────┐
SUBTRIBE | Glycininae   Phaseolinae
                          | | |
GENUS   Trifolium   Glycine   Phaseolus
SPECIES   subterraneum L.   max L.   vulgaris L.
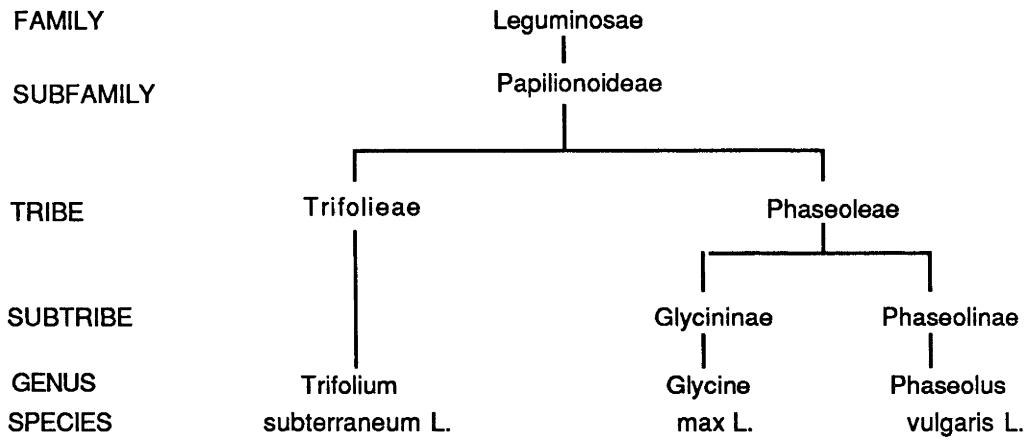
Fig. 4.20 Diagram of the taxomonic relationships of *Trifolium subterraneum* L.(subterranean clover), *Glycine max* L. (soybean) and *Phaseolus vulgaris* L. (bean) within the Leguminosae. The information used to compile the diagram was taken from Allen and Allen, 1981.

4.2.5 Comparison of the clover CHS genes (and derived protein sequences) with those from other species:

(b) promoter sequences

Several sequences which occur in the promoters of CHS2, CHS3, CHS4 and CHS5 have homologies with motifs found in the CHS and PAL genes from other plant species. Some of these may be spurious, and only the most promising are presented in point form below:

1. A large number of genes involved in phenylpropanoid metabolism from different plant species have one, or both, of two motifs conserved within their promoters (Lois et al., 1989; Ohl et al., 1990; Becker-Andre et al., 1991; Hauffe et al., 1991; fig. 4.21). In the case of the parsley (*Petroselinum crispum*) PAL1 gene, these sequences both display *in vivo* footprints after exposure to UV light and treatment with an elicitor (Lois et al., 1989), while that one of the sequences which is present in the parsley 4CL-1 promoter also has been shown to bear footprints (Hauffe et al., 1991). Part of box I within the parsley CHS promoter (1.5.3) corresponds to one of these motifs. The promoters of the clover CHS2, 3, 4 and 5 genes each contain sequences which are highly homologous to that motif having the consensus CCA(A/C)C(A/T)AAC(C/T)CC; other sequences in these clover promoters bear some resemblance to the second motif (fig. 4.21).

2. The promoters of the bean CHS15 and PAL2 genes share two conserved sequences situated to the 5' of their respective TATA boxes (Dron et al., 1988; Cramer et al., 1989). That sequence which in both cases is closest to the TATA box (-74 to -52 in the CHS gene, and element "D" in the PAL gene) contains one of the two motifs mentioned just previously, and *hence* is homologous to a sequence within the clover CHS promoters (fig. 4.21). The other sequence, which is AT-rich and part of a series of such motifs found upstream from the bean CHS and PAL

genes, does not have high homology with any parts of the CHS2, CHS3 and CHS4 promoters. However, a number of AT-rich motifs are situated in the clover CHS promoters, their overall positions being similar to those in the bean promoters (see discussion below). These motifs are also characteristic of analogous regions of the soybean CHS promoters (Wingender et al., 1989).

3. AC-rich elements occur between the TATA boxes and translation start sites in the promoters of the bean CHS15 gene (Dron et al., 1988) and the *Arabidopsis thaliana* PAL1 gene (Ohl et al., 1990), and the TATA boxes of of the bean PAL 2 and PAL3 genes are each preceded by one such sequence (element A; Cramer et al., 1989). After treatment of bean cells with glutathione, the bean CHS sequence becomes hypersensitive to digestion with DNase 1, implying it plays a role in transcriptional regulation (Lawton et al., 1990). Each of the of the three clover CHS genes 2-4 contains two AC-rich sequences located downstream from their TATA boxes, and another such sequence can be found in the soybean CHS1 promoter (fig. 4.22). In the case of the two conserved clover sequences, that which in each of the genes is closest to the translation start point bears the greatest resemblance to the bean CHS and *A. thaliana* PAL sequences. The CHS5 gene does not possess any AC-rich stretches between the TATA box and start codon.

4. Four regions of high homology, which may hence be important in the regulation of transcription, are found *between* the promoters of several of the soybean CHS genes (Wingender et al., 1989). Three of these (located between the site of translation initiation and the TATA box) are characterized by a repetition, usually non-contiguous, of the motif CTT(T)(T). This element occurs a number of times in the leader sequences of the CHS2, CHS3 and CHS4 promoters (fig. 4.22). It is also found, again in a repetitive fashion in the leader sequence of the

```
Consensus       T/A  C  T  C/A  A  C  C  T  A  C/A  C  C/A
Ts CHS2 -148     A   G  C   T   A  C  C  T  T   C   C   A  -137
Ts CHS3 -152     T   A  T   C   A  C  C  T  C   T   T   C  -141
Ts CHS4 -178     A   G  C   T   A  C  C  C  T   C   C   A  -167
Ts CHS5 -138     A   G  C   T   A  C  C  C  T   C   C   A  -127


Consensus       C  C  A  A/C  C  A/T  A  A  C  C/T  C  C
Ts CHS2 -220    C  C  A   C   C   A   G  A  C   T   C  C  -209
Ts CHS3 -252    C  A  A   C   C   A   A  A  C   T   C  C  -241
Ts CHS4 -243    C  C  A   C   C   A   A  A  C   T   C  A  -232
Ts CHS5 -232    C  C  A   C   C   A   A  A  C   T   C  C  -221
```

Fig. 4.21 Comparison of elements from the promoters of clover (Ts) CHS2, CHS3, CHS4 and CHS5 with two consensus sequences derived for conserved motifs found in the promoters of a number of phenylpropanoid genes (Lois et al., 1989). CHS, PAL, and 4-CL genes from a several different plant species were used to establish the consensus sequences. Numbers refer to the relationships of the clover sequences to the translation start sites.

```
Pv   CHS15   GCTAACTCCTCCAACATCTTCAAACACACAACTTCAGCAACCAAACCAACAACCCTCTCT

Gm   CHS1    TCCACCTTTCCAAATCACTTTCCAACATCCACCCCCATCATCATATCATACCCTTTCTAT

At   PAL1    CCACCCTTCATTCATCTAATTTTCCTCAGGAACAAATACAATTCCTTAACCAACAATATT

Ts   CHS2    ACCATCTCATCCCACAGATCTTTTCACCACAACACACTTCAGCCTTTGCTTTTCTACAAC

Ts   CHS3    ACCATCTCATCCCACACATTTCCACCACAACACTCTTCATCTCATTCCTTTTCTATAAAA

Ts   CHS4    ACCATCTCACCCCACACATCTTTCCACCACAACACTCTTCAACCTTTGCTTTTCTACAAA
```

Fig. 4.22 Comparison of the 60 nucleotides downstream from the presumptive TATA boxes of the CHS15 gene of bean (Pv; Dron et al., 1988), the CHS1 gene of soybean (Gm; Wingender et al., 1989), the PAL1 gene of *Arabidopsis thaliana* (At; Ohl et al., 1990) and those of clover (Ts) CHS2, CHS3 and CHS4. The AC-rich elements are overlined; the recurring CTT(T)(T) motifs are underlined; a further seven of these elements occur closer to the translation start codon in the leader sequence of the *Arabidopsis thaliana* gene (not shown).

Arabidopsis thaliana PAL1 gene (Ohl et al., 1990).

5. The sequence CACGTG, which is commonly found in the regulatory regions of light inducible genes (the core sequence of the G-box; 1.5.3), occurs once in each of the CHS3, CHS4 and CHS5 promoters (at positions -382 to -377, -201 to -196, and -151 to -145 upstream from the respective translation start sites), but not in the CHS2 promoter. Although these three clover elements are not similarily located in their promoters, a sequence CATGTG does occur at position -371 to -366 in the promoter of CHS4. This 5' upstream region is highly conserved between the CHS3 and CHS4 genes (part of that numbered *4* in fig. 4.12), and as such it appears that the CHS4 CATGTG element is a one base variant of the CHS3 CACGTG sequence. The promoter of CHS3, however, does not contain any sequence analogous to that of CACGTG at the approximate position of the CHS4 -201 to -196 sequence. Although the CHS2 promoter does not contain the sequence CACGTG, it does possess an element CACATG at -164 to -159. Both this one base variant of CACGTG, and the one just mentioned above, involve purine-purine or pyrimidine-pyrimidine substitutions.

6. A sequence, with the consensus GGTTAA(A/T)(A/T)(A/T), occcurs in three boxes of the bean CHS15 gene which are bound by a presumptive transcription factor (1.5.3; Lawton et al., 1991). Two elements which have a high degree of homology to this consensus are located in the promoter of CHS4 (fig. 4.23).

```
Bean consensus sequence        G  G  T  T  A  A  A/T A/T A/T
CHS4                     -845  T  G  T  T  A  A   A   T   T   -837
CHS4                     -749  G  G  T  T  A  T   A   T   A   -741
```

Fig. 4.23 Comparison of two sequences, both from the CHS4 promoter, with the consensus sequence for a site which occus four times in three bean CHS15 promoter boxes bound by the transcription factor SBF-1 (Lawton et al., 1991). The numbers refer to the positions of the two sequences with respect to the translation start site.

7. Elements fitting the consensus sequence for animal virus (such as SV40) core enhancers (Weiher et al., 1983) have been found in the promoters of the phenylpropanoid genes from a number of species (1.5.3), and also in the promoters of the SSU genes from several higher plants (Timko et al., 1985; Sugita et al., 1987). A sequence reminiscent of, and one conforming to the core consensus, is located in the promoters of clover CHS2 and CHS3, respectively (fig. 4.24). These two sequences are not situated at identical positions with respect to the start points of translation. It has not been established for any plant species (as of yet) if these "enhancer" sequences play an important role in the regulation of CHS transcription.

```
Animal virus enhancer
core consensus sequence        (G)  T  G  G  A/T  A/T  A/T  (G)
CHS2                   -554    G    T  G  G  A    C    A    A  -547
CHS3                   -354    G    T  G  G  A    T    A    A  -347
```

Fig. 4.24 Comparison of two sequences, one from each of the promoters of CHS2 and CHS3, with the animal virus core enhancer consensus sequence (Weiher et al., 1983). The numbers refer to the positions of the two sequences with respect to the translation start sites.

4.2.6 CHS7 and CHS8

The truncated state of CHS7 was confirmed by sequencing. This "gene" does not possess the first CHS exon and contains at least one deletion in the second exon (this exon was not sequenced in entirety; figs. 4.25/4.26); it obviously must be a "byproduct" of the genetic events that have led to the present clover gene family. The orientation of this "gene" is the same as that of CHS5 and CHS6 (fig. 3.11). A similar incomplete gene is found in a cluster of tomato CAB genes; as with the clover CHS7 gene, this tomato sequence contains an internal deletion (Pichersky et al., 1985).

```
AAGCTTAATATTTTAAAATATAAAATAAAAATCAATATATTATTTATCATTAAATTAAGACCAGACAAT
TATCAATAATCTAATTAAATGCGTAAAAACCAGGAAAAAATCAATATATTACTATTATATGATTGACTT
TATTAATTTATTATAGTACTAATATGATTTTTTTTTATAACTAATTCACGTGATAAATCTATGATCAAA
AGTAGATACATGTATCTAACAGAAGAGATTTTGAAAGAAAATCCAAGTGTTTGTGAATACATGGCACCT
TCATTGGATGCTAGGCAAGACATGGTGGTGGTAGAGGTACCTAGACTAGGGAAAGAGGCTGCAGTGAAG
GCAATAAAAGAATGGGGTCAACCAAAATCAAAGATCACTCACTTAATCTTTTGCACCACAAGTGGTGTA
GACATGCCTGGTGCTGATTACCAGCTCACAAAACTCTTAGGTCACATTTGGCGGCCCTAGTGATACTCA
CTTGG
```

Fig. 4.25 Part of the sequence of the truncated gene CHS7. The beginning of the coding sequence of the second exon is indicated by a bold face G. No evidence was obtained of the presence of a first exon in the preceding DNA.

Sequencing of one clone from the CHS8 sequence established that the gene is orientated such that transcription would proceed from left to right as λTsCHS1.6 is drawn in fig. 3.3. As the hybridizing region present on λTsCHS1.6 is so small, the clone must be devoid of part of the 5' sequence and promoter of CHS8. Although the predicted amino acid sequences of CHS5 (this study) and CHS6 (T. Arioli, personal communication) did not reveal any *susbstantial* differences to those of the cluster I proteins, the derived amino acid sequence for the single clone used to orientate CHS8 within λTsCHS1.6 was found to possess four unique differences to the CHS2-5 proteins within 20 residues (fig. 4.27). The nucleotide sequence of this part of the CHS8 gene contains between 2 and 3 more differences to the corresponding regions of the CHS2-5 genes than these latter sequences do to each other at this point. However, all the amino acid changes are conservative (according to the group classification of Miyata et al., 1979), suggesting that the CHS8 gene is under selection pressure and may well be actively transcribed.
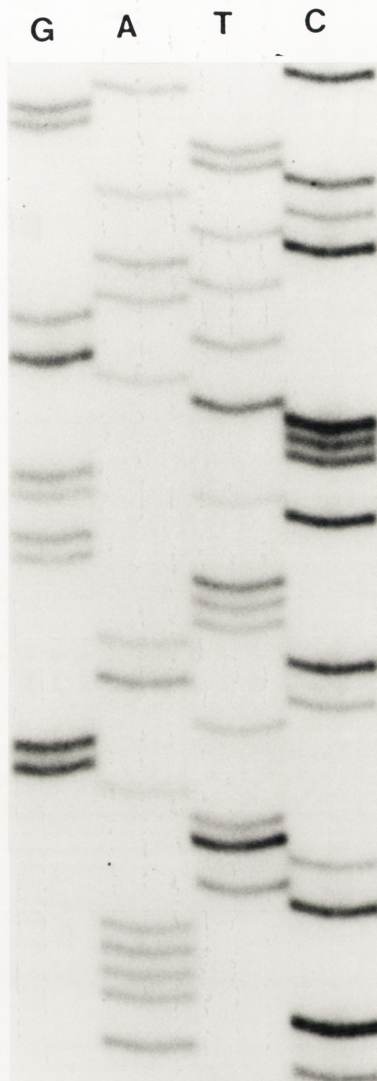
Fig. 4.26 Part of a sequencing gel showing the region of CHS7 which contains a 140bp deletion. A comparison of this sequence against that of CHS2 is given below. Numbers flanking the CHS2 sequence refer to the relationship of the end nucleotides to the start point of translation. The deletion in CHS7 is indicated by asterisks.

```
CHS2  614  TACAAAACTCTTAGGTC  140bp  ACATTCCGCGGTCCTAGTGACACTCATTTAGAC  803
CHS7       CACAAAACTCTTAGGTC  *****  ACATTTGGCGGCCCTAGTGATACTCACTTGGAC
```

Fig. 4.27 Part of the sequence of the gene CHS8. The *complement* of the nucleotide sequence is shown below, together with the derived amino acid sequence. The corresponding region of the CHS2 protein (amino acids 272-299) is shown for comparison; differences in the CHS8 sequence are italicised.

5'TTCCTGGACTCATCTCAAAGAACATTGAAAAAGCTCTTGTTGAGGCCTTTCAACCTTTGGGTGTATC
TGATTACAATTCCATAT 3'

CHS8 P G *L I* S K N I *E* K A L V E A F Q P L G *V* S D Y N S I
F
CHS2 P G I V S K N I D K A L V E A F Q P L N I S D Y N S I
F

4.3 Discussion

The detailed analysis of clover CHS2, CHS3, CHS4 and CHS5 presented here shows that not only are they extremely similar, but they are also highly homologous to the CHS genes (or their cDNAs) which have been isolated from other plants. The sequences of these genes contain no obvious features, such as internal stop codons or frameshift mutations, that are characteristic features of pseudogenes. Furthermore, their amino acid sequences and hence molecular weights are all very similar. However, at least in the case of carrot, it appears that isoforms of CHS can differ by at least 2 kDa and still be expressed (Gleitz and Seitz, 1989).

Comparison of the deduced amino acid sequences of the four genes with those predicted for the proteins of several resveratrol synthase (RS) genes from *Arachis hypogaea* (Schroder et al., 1988; Lanz et al., 1990), suggests that the subterranean clover genomic clones are indeed those coding for CHS proteins and *not* RS proteins. As mentioned previously (1.5), the amino acid sequences of these RS proteins contain a number of residues which are different from those highly conserved among CHS proteins. Although the clover proteins contain some differences to the consensus of Niesbach-Klosgen et al., 1987, none of these correspond to those that set the RS protein distinctly apart from the CHS one. Attempts to isolate resveratrol from fungus inoculated leaflets of two subspecies of *T. subterraneum* have proved unsuccessful (Ingham, 1990). It appears that within the genus *Trifolium*, the compound resveratrol is confined to the section Chronosemium, of which *T. subterraneum* is not a member (Ingham, 1978/1990). Hence, the RS gene may not even be present in clover.

The high ratio of "silent" to "non-silent" nucleotide substitutions between the coding sequences of the clover CHS2, 3, 4, and 5 genes (fig. 4.6/7) suggests that, at least until recently, the genes have been subject to direct selection pressure. However, no two clover CHS genes (including those for which the sequences are not presented in this thesis) are entirely identical, as is the case with two of the tomato SSU genes, whose exon as well as intron sequences are invariant (Sugita et al., 1987).

Of great interest is the finding that the nucleotide and derived amino acid sequences of CHS2 are closer to CHS5 than to CHS3 (fig. 4.7). If clusters I and II represent two different loci, it would normally be expected that the tightly linked CHS2, 3, and 4 genes would possess higher homology to each other than to CHS5. In the case of the *Petunia hybrida* CHS genes, homology between those genes which are tightly linked is much greater than that between non-linked genes, prompting the classification of the linked genes into subfamilies (Koes et al., 1987/1989b). The same phenomenon has been observed among the two clusters of tomato CAB genes (Pichersky et al., 1985), while three tandemly arranged members of the tomato SSU gene family share much greater homology to each other than to the two remaining SSU genes of that plant, which are located separately at individual loci (Sugita et al., 1987). It is hence possible that clusters I and II are part of a single locus, with CHS4 and CHS5 lying adjacent to each other as suggested in the previous chapter (3.3). However, the large stretches of homology between the 5' and 3' flanking sequences of CHS2, 3, and 4, and their absence from CHS5, would suggest that the two clusters, I and II, are separate loci.

The location of introns in the four clover CHS genes at the same position as those found in the CHS genes of other species, reinforces the well observed phenomenon of the structural conservation of these genes

between species. Because intron loss and insertion are relatively common events in evolution, the retention of this intron may reflect its necessity for the biogenesis of stable CHS mRNA and/or the transport of the message across the nuclear membrane to the cytoplasm (Krainer and Maniatis, 1988). In some instances, intron containing genes are not expressed when their introns are removed by genetic engineering, implying an essential role for this internal non-coding DNA; in other systems, however, removal of introns does not appear to influence gene expression (Kuhlemeier et al., 1988). It is possible that the universal prevalence of the CHS intron may be due to an inefficient mechanism for its removal and/or the absence of any selective advantage resulting from its loss (Krainer and Maniatis, 1988).

The promoters of CHS2, CHS3 and CHS4 share five regions of extensive homology between themselves. The conservation of these sequences may reflect their importance in transcriptional regulation, and/or be the result of the duplication events which led to the formation of cluster I. However, of greater importance is the occurrence of shorter sequences within the clover promoters which strongly resemble those found in other plant gene promoters. The sequence CACGTG, the core element of the G-box (Giuliano et al., 1988) and the parsley CHS box II (Schulze-Lefert et al., 1989a/b), is widely recognized to be involved in the transcriptional regulation of many genes. The clover promoter sequences share not only the core CACGTG motif (except for the single nucleotide substitutions in one of the CHS4 sequences and that present in the CHS2 promoter), but also the flanking nucleotides, giving rise to a consensus *GCACGTGA*.

If the two CHS4 "G-boxes" sequences (4.2.5) indeed play important roles in the regulation of the expression of this gene, perhaps in relation to induction by light, then this could imply that the promoter has a

structure similar to that of the parsley CHS gene. Four regions are important in the light regulated expression of the single CHS gene present in that plant. The sequence CACGTG, and a one base variant of it, are the central features of the box II and box III regions respectively (1.5.3; Schulze-Lefert et al., 1989a/b). Moreover, it has been found that the activity of box II requires the presence of box I at a defined distance (Block et al., 1990). As mentioned above (4.2.5; fig. 4.21), the latter box contains a motif which is homologous to sequences from other phenylpropanoid genes, including clover CHS4 (-178 to -167). Interestingly, the distance between *this* clover CHS4 element and the neighbouring (upstream) CACGTG sequence is approximately the same as that separating boxes I and II in the parsley promoter (20bp). Two sequences in the promoter of the mustard CHS1 gene, which resemble parsley boxes I and II, are likewise separated by a similar distance (Batschauer et al., 1991). In the soybean CHS1 promoter a sequence bearing a variant of CACGTG, and one analogous to the parsley box I, both reside in regions known to be essential for light responsiveness, but these are separated by approximately 100bp (Wingender et al., 1990). Other research work has provided support for the role of a CACGTG-like motif acting in conjunction with a cis-element *lacking* such a sequence to promote transcription. Sequences resembling that of box II have been found in the promoters of non phenylpropanoid pathway plant genes, and many of these occur in a similar juxtaposition to other cis elements as that seen in parsley; these other elements do not bear sequence homologies to box I (see references within Block et al., 1990). It is speculated that diverse gene families may have selected the box II/G-box element to act as a transcriptional enhancer *with* various cis-acting partners (Schulze-Lefert et al., 1989a; Block et al., 1990). A summary of

some of the presumptive cis-acting elements present in the subterranean clover CHS4 promoter is shown in fig. 4.28.

The two presumptive CHS4 G-boxes are separated by 170bp. This figure is very close to 177bp, the distance separating two possible G-boxes located in the *Arabidopsis thaliana* CHS promoter (Feinbaum et al., 1991); the parsley boxes II and III, however, are separated by a much shorter distance (Schulze-Lefert et al., 1989a/b). More interestingly, the number 170 represents an exact integer of helical repeats (16), assuming a repeat of 10.6bp (Peck and Wang, 1981; Rhodes and Klug, 1981). Although this may very well be coincidental, it could reflect the binding of two or more regulatory proteins alongside each other. It is believed that a group of transcription factors interact with each other on the same side of the helix in the SV40 early promoter (Takahashi et al., 1986). Alteration of the distance between the SV40 enhancer and the "21bp repeat region", and between the latter and the TATA box, by the insertion of DNA sequences that were a multiple of a full helix turn resulted in only a minor reduction in the efficiency of transcription when compared to insertions that changed the orientation of the SV40 elements to each other by half a turn. If this does imply the binding of the SV40 transcription machinery on the same face of the helix, then it is possible that such a situation may occur with the protein factors that regulate the expression of CHS4. The three CPRFs which binds to the parsley box II (G-box) are of the leucine zipper class (Weisshaar et al., 1991), but a DNA-binding protein possessing a helix-loop-helix domain, as well as a potential leucine zipper, recognizes the promoter sequence CACGTG in a mammalian system (Carr and Sharp, 1990).

On a more general level, a characteristic feature of the clover CHS promoters, and the promoters of phenylpropanoid genes from many species (4.2.5), is the occurrence of AT-rich elements. In other systems,

certain sequences of such nucleotide constitution are known to have unusual stereochemical conformations which may be important in protein recognition and binding (Nussinov, 1990 and references therein). Poly(dA)-poly(dT) elements are structurally unique in the respect that they possess a helix repeat of 10bp instead of the normal 10.6bp (Peck and Wang, 1981; Rhodes and Klug, 1981). Physical-chemical studies of these sequences have shown that they are capable of causing DNA bending (Nussinov, 1990). An AT-rich sequence of 12bp in the promoter of the high mobility group protein of trout (HMG-T) has been shown to be both important in the control of transcription and capable of adopting a cruciform structure when torsionally stressed by negative supercoiling; it is likely the two are linked (Wright and Dixon, 1986). In the minor groove of β DNA, AT and GC base pairs differ only by the presence of the hydrogen atom in adenine instead of the amino group of guanine. It is possible that at least some proteins which recognize AT-rich sequences do so by making extensive contacts in the minor grooves of such regions, being unable to bind to GC-rich sequences due to the "interfering" amino groups (Solomon et al., 1986).

In a number of cases it has been demonstrated that specific proteins, which function in the control of transcription, recognize AT-rich elements in the promoters of plant genes. Two distinct regions to the 5' of the soybean leghaemoglobin lbc3 gene, which have been shown to bind the same trans-acting factor, both consist almost entirely of A and T base pairs (Jensen et al., 1988). This factor, together with two others, also recognizes AT-rich regions in the promoter of the soybean nodulin gene N23 (Jacobsen et al., 1990). The promoter of the lectin gene from this plant has two AT-rich elements which are the targets for an embryo DNA binding protein (Jofuku et al., 1987). A similar situation occurs with the gene encoding the γ subunit of glutamine synthetase in bean;

again two AT-rich elements have been identified that are bound by protein factors (Forde et al., 1990). Interestingly, these factors also have a high affinity for at least one (that tested) of the soybean lbc3 AT regions. Analysis of the pea SSU promoters has shown that at least two different AT-rich regions are bound by proteins (Datta and Cashmore, 1989; Lam et al., 1990). One of these, called the AT-1 box, occurs in the promoters of the SSU and CAB genes from other species.

The unusual structural properties associated with AT-rich sequences, together with evidence that they are important sites for regulatory control in other genes (in both plants and animals), *suggests* that their occurrence in the clover CHS promoters is connected directly to a role in the regulation of transcription. A number of studies have produced evidence which would support the notion that it is the structure of these elements, rather than strict DNA base sequence, which is of greater importance to the binding of transcription factors (Pedersen et al., 1991, and references there-in). It has been suggested that there may be an evolutionary conserved group of DNA binding proteins which are specific for AT-rich sequences (Forde et al., 1990).

A new amino acid consensus sequence, superseding that deduced by Niesbach-Klosgen et al., 1987, can be derived for the CHS protein by taking into consideration the CHS sequences which have been published in the mean-time (fig. 4.29). It differs little from that of Niesbach-Klosgen et al., indicating that the sequences used in this original study were sufficient to enable the major conserved parts of the protein (and hence those regions essential for catalytic activity) to be discovered. Thirty-eight positions for which a single amino acid was recorded in the original study were found to be able to accept another residue. Of these changes, twenty-four (63%) are conserved according to the amino acid classification of Miyata et al., 1979. This suggests that in the future,

changes to this "new" consensus will probably involve conservative substitutions. It should be remembered, of course, that the *pattern* of hydrophobic and hydrophilic residues may be the essential structural feature of a protein sequence, rather than the actual residues themselves (Bowie et al., 1990).

It will be recalled that the catalytic mechanism of the CHS protein is very likely similar to that of β-ketoacyl-ACP-synthase (1.5.1). The active site for the β-ketoacyl-ACP-synthase protein has been experimentally determined for a number of species (see references within Schroder and Schroder, 1990). Comparisons of CHS amino acid sequences with those from β-ketoacyl-ACP-synthase proteins have suggested two alternative sequence motifs which may contain the CHS active site, and specifically the important cysteine residue (Kauppinen et al., 1988). One of these sequence motifs, that closest to the carboxyl end of the protein, is favoured as the active site. Both of the motifs remain conserved in the new amino acid consensus sequence presented in fig. 4.29. It has been suggested that the CHSB protein of *Petunia hybrida*, whose gene is expressed in certain circumstances at a low level (Koes et al., 1989a), is functionally inactive as a (polar) cysteine residue occurs adjacent to the "active" cysteine in the second sequence motif (Kauppinen et al., 1988).

However, recent site-directed mutagenesis studies of the mustard (Sinapis alba) protein point to a different cysteine residue (cys-169 in the mustard sequence) as being the active site (Schroder et al., 1991). Separate conversion of all the known conserved cysteines (from a comparison of derived CHS protein sequences) into alanines showed that only cys-169 was absolutely essential for catalytic activity. As the clover proteins CHS2, CHS3 CHS4 and CHS5 all possess this latter candidate for the "active" cysteine (and also the initial two alternatives suggested by Kauppinen et al., 1988), then a further sign that their respective genes

could be pseudogenes, or in the "process" of being inactivated, is eliminated.
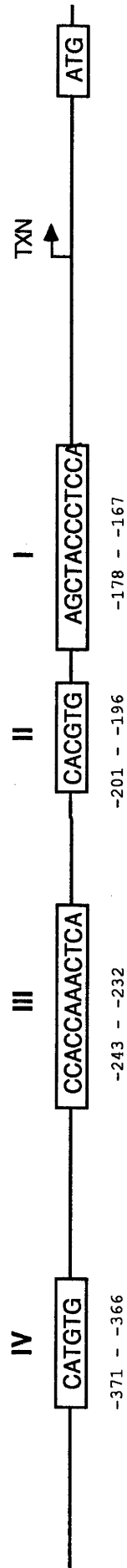
Fig. 4.28 Diagram showing four "boxes" (I - IV) which *may* be important in the transcriptional regulation of the subterranean clover CHS4 gene. Boxes II and IV bear sequences analogous to boxes II and III of the parsley CHS promoter (Schulze-Lefert et al., 1989a/b), and to regulatory sequences present in many gene promoters (Guiltinan et al., 1990 and references there-in). Boxes I and III contain sequences resembling those found in many phenylpropanoid pathway gene promoters (Lois et al., 1989). The (presumptive) start point of transcription is marked by TXN. Numbers refer to distances from the methionine start codon.

```
---------** --R--Q-A-G PA-***IGTA TP*--V-Q*- YPDYYF*IT-
---------*- --R--Q-A-G PA-***IGTA -P---V-Q*- YPD*YF*IT-

S*H-**LK-K FKRMC*KS-I **RYMH*TE* -*-ENP-*C- Y-APSLDARQ
S*H--*LK-K F-R*C*KS-I *-R-M-*-E* -*-*NP-*C- Y-APSLD*RQ

D*VVVEVP*L GK-AA-KAIK EWG-P*S*IT HL*FCTTSGV DMPG-DYQLT
D*VV-EVP*L GK-AA--AIK EWG-P*S*IT H*-FCTTSGV DMPG-DYQLT

K-LGLR*S-- R-MMYQQGCF AGGTVLR*AK DLAENN-GAR VLVVCSEITA
K-LGLR*--- R-MMYQQGCF AGGTVLR*AK DLAENN-GAR V*VVCSE*TA

VTFRGP-**H LDSLVGQALF GDGA*A***G *DP----ERP LF*****Q*
VTFRGP-**H *DSLVGQALF *DGA*A---G *D*----E*P **-**-**Q*

*-PDS-GAID GHLREVGLTF HLLKDV-GLI SKNI-**L-* AF-P*GI--W
*-PDS-GAID GHL-E-GL*F HLLKDV-G** SKNI-**L-- AF-P*-I--*

N-*FW*AHP* GPAILD-VE- K*-L---*** *TR-VLS*YG NMSSACVLFI
N--F-*AH-* GPAILD-VE- K*-L---**- *TR-VLS*YG NMS*ACVL-*

*DEMRK-S-- *G--TTG*G* *WGVLFGFGP GLTVETVVLH S**--------
*D*MR*-*-- -G--TTG*G* *WGVLFGFGP GLT*ETVVL* S*---------
```

Fig 4.29 Comparison of the amino acid consensus sequence for the CHS protein calculated by Niesbach-Klosgen et al., 1987 (top lines), with that deduced by taking into account sequences published since then (bottom lines). All the amino acid sequences used as the basis to construct both consensus sequences were derived from nucleotide sequences. Those used in the study of Niesbach-Klosgen came from *Antirrhinum majus, Hordeum vulgare, Magnolia liliiflora, Petroselinum hortense, Petunia hybrida, Ranunculus acer*, and *Zea mays*. The *further* sequences used to produce the new consensus were those from *Phaseolus vulgaris* (CHS17, 1, 5, 4, and 14; Ryder et al., 1987), *Sinapis alba* (Ehmann and Schafer, 1988), *Arabidopsis thaliana* (Feinbaum and Ausubel, 1988), *Petunia hybrida* (CHSA and J; Koes et al., 1989b), *Matthiola incana* (Epping et al., 1990), *Glycine max* cv. Merr. (Akada et al., 1990a/b;1991), *Glycine max* cv. Prize (Estabrook and Sengupta-Gopalan, 1991), *Zea mays* (Franken et al., 1991), *Hordeum vulgare* (Rohde et al., 1991) and *Trifolium subterraneum* (CHS2, CHS3, CHS4 and CHS5; this study). In the case of *P.hybrida* only the most actively transcribed genes were used in the analysis. Asterisks indicate those positions where more than one amino acid is found, but the residues present are conserved with respect to the classification of Miyata et al., 1979; dashes represent those positions where no conservation is found. The two sequence motifs suggested by Kauppinen et al., 1988 to contain the active site are underlined. The more likely candidate for the active site cysteine, as suggested by the work of Schroder et al., 1991, is double underlined.

Chapter 5. Brief examination of the clover PAL multigene family

5.1 Introduction

Phenylalanine ammonia-lyase (PAL) catalyses the first reaction of phenylpropanoid metabolism, converting phenylalanine to trans cinnamic acid. The structure and molecular organization of the subterranean clover genes for this enzyme was subjected to a much more briefer analysis than that conducted for CHS.

Probing of the clover gene library with the bean PAL5 cDNA (Edwards et al., 1985) at the non-stringent temperature of 55^DC failed to produce any evidence of hybridization. Presuming that the level of homology between the bean and clover PAL sequences was not high enough for one to be successfully used as a probe to locate the other, it was decided to use the technique of the polymerase chain reaction (PCR) to obtain a clover PAL probe, and inturn a clover genomic PAL clone (Mullis and Faloona, 1987). An examination of all the derived PAL protein sequences known at the time - sweet potato (Tanaka et al., 1989), rice (Minami et al., 1989), parsley (Lois et al., 1989; Schulz et al., 1989), bean (Cramer et al., 1989) and soybean (Estabrook and Sengupta-Gopalan, 1991) - revealed that there were a number of stretches of residues that were conserved in each species. The strategy was to design primers (oligonucleotides) which would hybridize to the corresponding clover nucleotide sequences and be extended by the polymerase enzyme so as to amplify the region between. By such a procedure an homologous DNA probe for the clover PAL genes could be obtained.

5.2 Results

5.2.1 Design of PCR primers

In selecting the conserved sequences to form the basis of the primers, three important factors were taken into consideration. Firstly, it was advantageous to have as many amino acids as possible which had only one codon (e.g. methionine), and conversely to avoid those amino acids which had several alternate codons (Compton et al., 1990). In the latter case, the primer would have to be made degenerate at the positions where variation was conceivable, meaning that its specificity for the PAL sequence would be reduced. Sequence including the residues leucine, serine and arginine was especially undesirable, as these amino acids are all specified by six codons. The limitations of the triplet genetic code mean, of course, that the codons of such amino acids vary not only in the third position, but in other positions. For instance, leucine is designated by CTN (N being any nucleotide) *and* TTA/G. With this amino acid it is quite possible that a *single* mutation could not only occur in the third codon position, and still not change the residue, but also in the first position (C to T, or vice versa). Any primer complementary to a DNA sequence containing such an amino acid must tolerate all these alternatives. Secondly, to enhance the binding of the primers, it was also best that their sequences consisted of a random base distribution and were devoid of any polynucleotide tracts (Saiki et al., 1990). The third consideration was to check the primers against each other to be sure that they were not complementary.

As would be expected, it proved impossible to entirely satisfy the above three conditions when examining the amino acid sequences and the possible nucleotide sequence combinations. Nevertheless, two stretches of residues, approximately 500bp apart, were selected as those which would

form the basis for primers that would enable part of the PAL gene from not only clover, but from any species having these conserved residues, to be isolated (fig. 5.1). These two sequences are present in the second exon of the PAL gene (this gene has a structure of two exons and one intron in all cases for which a genomic clone has been characterized). The primers were also designed with (different) restriction enzyme sites at their ends (*Eco*R1 and *Bam*H1) so that any PCR products could be cleaved to produce restriction ends and easily cloned into Bluescript (Higuchi, 1989). A PAL cDNA from alfalfa, published after the synthesis and use of the primers in this study, proved to have the same derived amino acid sequence in the two priming regions as those found in the other plants (Gowri et al., 1991). The conservation and sensibility of these two regions as PCR priming sites was hence reinforced.

5.2.2 Cloning and molecular analysis of the clover PAL genes

Total genomic DNA from clover (approximately 0.5µg) was combined with the primers (50ng of each) and the necessary nucleotides and buffer (2.12.1), and subjected to thirty-five cycles of denaturation (94^{D}C/2 mins), annealing (55^{D}C/2 mins) and extension (72^{D}C/2 mins). The first time that this was attempted a major product of approximately the correct size was obtained, together with a number of higher molecular weight species (fig. 5.2). This mixture was phenol/chloroformed, cleaved with the two restriction enzymes which had cleavage sites at the 5' ends of the primers (*Bam*H1 and *Eco*R1) and ligated with Bluescript SK(+) cleaved with the same enzymes. Plasmids were isolated from four white colonies (the colour indicating the presence of an insert; 2.4) by the mini prep procedure (2.7.5) and restricted with *Bam*H1 and *Eco*R1. All four were found to possess inserts.

The inserts of the four PCR clones were subjected to sequence analysis.

```
 _____
|          |░░░░░░░░░░░░░░░░░░░░░░░░░░░░|                  |
|_____|░░░░░░░░░░░░░░░░░░░░░░░░░░░░|_____|
 1                        976          1496             2175
```

```
Alfalfa      AAAIMEHILDGSSYVKAAKK...137 amino acids...NPVTTHVQSAEQHNQDV

Bean1        ...............M.D...                     ....S............
Bean2        ...............I.....                     ....S............
Bean3        ................N...                      ....S....T.......
Parsley1     ............A.....Q.                      ....N............
Parsley4     ................Q.                        ....N............
Rice         .........A...FMSH...                      ..I.N............
Soybean      .........E..........                      ....S............
S. Potato    ................Q.                        ....N............

                               A         G   C
PAL-1 5'-AA GGATCC GCI GCI GCI ATT ATG GA  CA  AT-3'
                               C         A   T


              G   T   A   A   C   C
PAL-2 5'-AA GAATTC AC  TC  TG  TT  TG  TG  TC-3'
              A   C   G   G   T   T
```
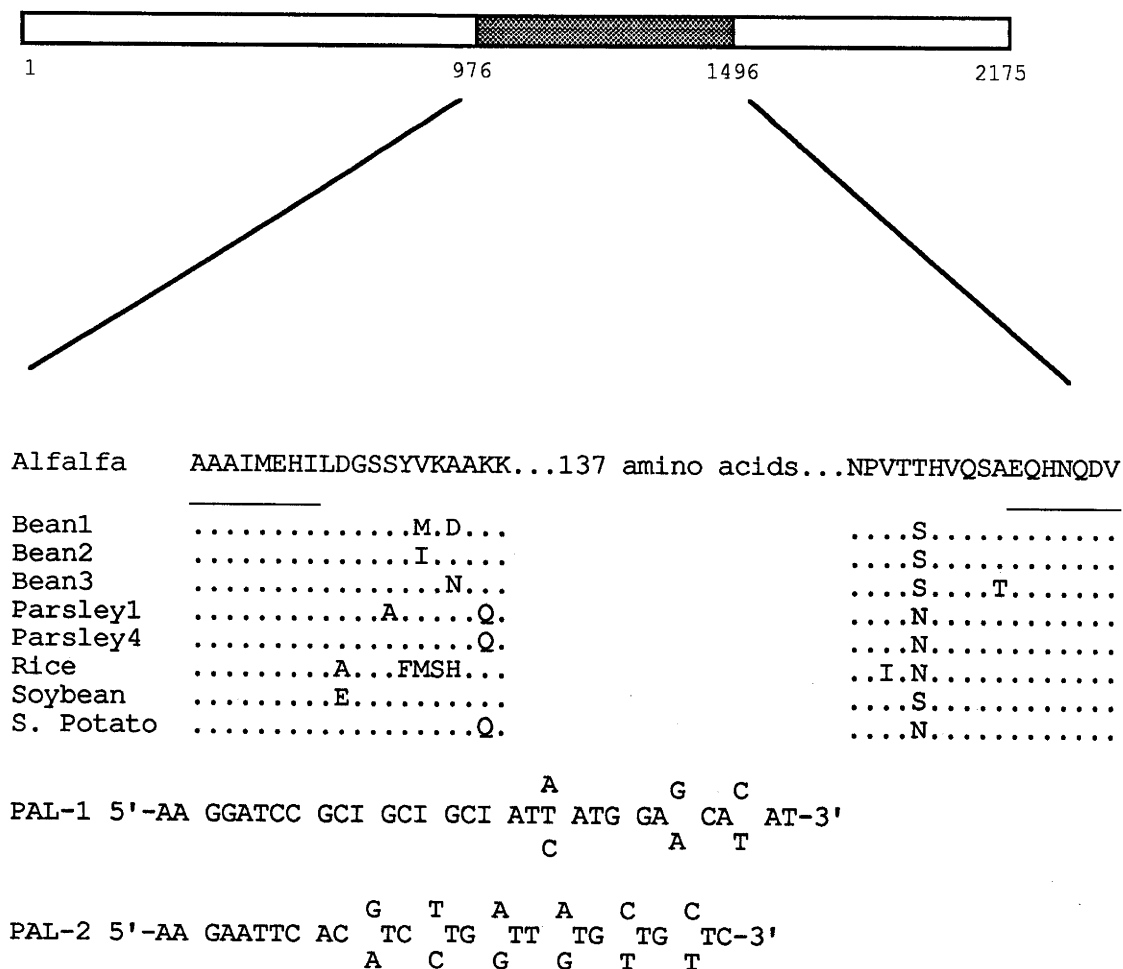
Fig 5.1 Diagram showing the amino acid sequences used to design the PCR primers, and the sequences of these oligonucleotides. At the top is depicted a generalized structure of a PAL gene (minus the intron which seems to be a universal feature of these genes). The numbers refer to the positions in the alfalfa pAPAL1 cDNA nucleotide sequence (Gowri et al., 1991). The derived amino acid sequence of the alfalfa gene in the two regions from which the primers were designed, together with some of the (internally) flanking residues, is shown beneath. The differences of bean gPAL1 (1), gPAL2 (2) and gPAL3 (3) (Cramer et al., 1989), parsley PAL1 (Lois et al., 1989), parsley PAL4 (Schulz et al., 1989), sweet potato (Tanaka et al., 1989), rice (Minami et al., 1989) and soybean (Estabrook and Sengupta-Gopalan, 1991) to the alfalfa sequence are indicated. The amino acids included fully, or partially, in the correponding sequences of the primers, are underlined. The nucleotide sequences of the two primers are shown beneath. Each was flanked at its 5' end by a restriction sequence (*Bam*H1 in the case of PAL-1 and *Eco*R1 for PAL-2) for easy cloning of amplified products into a plasmid vector. Nucleotides separated by strokes were alternatives used in the primers. I is the base inosine, which can pair with any of the four regular bases (Sambrook et al., 1989).
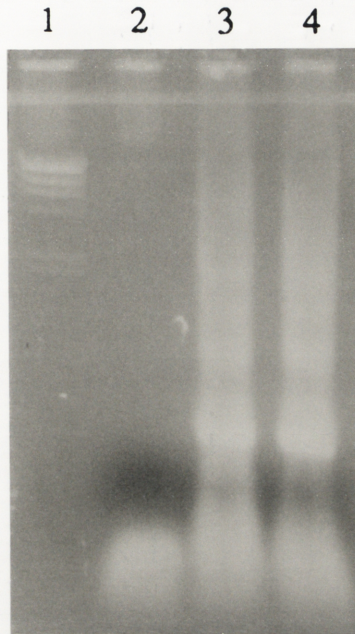
Fig. 5.2 Result of a PCR of *Trifolium subterraneum* L. cv. Karridale genomic DNA using primers designed to amplify a segment of any PAL gene.

Lane 1 - λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Lane 2 - Control PCR without any genomic DNA.

Lane 3 - PCR of genomic DNA.

Lane 4 - PCR of genomic DNA.

The PCR conditions in each case were denaturation - 94^DC, annealing - 55^DC, extension - 72^DC, for 35 cycles. By comparing the PAL nucleotide sequences from other plants, the amplified clover fragment should be 537bp (521bp of actual clover sequence plus 16bp of flanking artificial restriction sites for cloning); the major band in both lanes 3 and 4 is approximately this size.

Of the four clones, only one was found to contain a PAL sequence (fig. 5.3). Fortuitously, an internal *Bam*H1 site within the PAL sequence was not cleaved, and the fragment was cloned into the vector by using the *Bam*H1 site at the primer end. Comparison of the 521bp clover PAL sequence against that of the bean PAL5 cDNA (gPAL1) revealed a 86% homology at the nucleotide level, while that of 96.6% at that of the protein. Hence, most of the nucleotide changes were conservative, suggesting that the sequence which had been cloned was that of a clover gene, and furthermore, that the *Taq* polymerase had made few errors, if any, in the extension reactions. The sequences of the other three products arising out of the PCR bore no resemblance to that of a PAL gene. Presumably they resulted from the primers binding to analogous sequences elsewhere in the genome of the plant; they were not analysed further.

```
GCGGCGGCGATAATGGAGCATATTTTGGATGGGAGTGCTTATGTCAAAGATGCTAAGAAGTTG
 A   A   A   I   M   E   H   I   L   D   G   S   A   Y   V   K   D   A   K   K   L
CATGAGATGGACCCTTTGCAGAAGCCAAAGCAAGATAGATATGCACTTAGAACTTCGCCGCAA
 H   E   M   D   P   L   Q   K   P   K   Q   D   R   Y   A   L   R   T   S   P   Q
TGGCTTGGTCCTTTGATTGAAGTGATTAGATTCTCTACAAAGTCAATTGAGAGAGGGATCAAC
 W   L   G   P   L   I   E   V   I   R   F   S   T   K   S   I   E   R   G   I   N
TCTGTCAATGACAACCCTTTGATTGATGTTTCAAGAAACAAGGCTTTGCACGGTGGAAATTTT
 S   V   N   D   N   P   L   I   D   V   S   R   N   K   A   L   H   G   G   N   F
CAAGGAACACCTATCGGAGTATCCATGGATAATACACGTTTGGCTCTTGCATCAATTGGTAAA
 Q   G   T   P   I   G   V   S   M   D   N   T   R   L   A   L   A   S   I   G   K
CTTATGTTTGCTCAATTCTCCGAGCTTGTCAATGATTTTTACAACAATGGGTTGCCTTCAAAT
 L   M   F   A   Q   F   S   E   L   V   N   D   F   Y   N   N   G   L   P   S   N
CTCTCTGCTAGTAGAAATCCTAGCTTGGATTATGGGTTCAAGGGATCCGAAATTGCCATGGCT
 L   S   A   S   R   N   P   S   L   D   Y   G   F   K   G   S   E   I   A   M   A
TCCTATTGTTCTGAGTTACAATATCTTGCAAATCCAGTTACAACCCATGTCCAAAGTGCCGAA
 S   Y   C   S   E   L   Q   Y   L   A   N   P   V   T   T   H   V   Q   S   A   E
CAACACAACCAAGATGT
 Q   H   N   Q   D   V
```

Fig. 5.3 Nucleotide sequence and derived amino acid sequence of a 521bp product obtained after a PCR of total genomic DNA from *Trifolium subterraneum* L. cv. Karridale with the two PAL primers PAL-1 and PAL-2 (fig. 5.1). The sequences show that the amplified product is indeed part of a clover PAL gene.

The (potential) clover PAL fragment, designated pcrPAL, was amplified up from the Bluescript plasmid by the use of the PAL primers and used as a hybridization probe to screen the subterranean clover genomic library. In the first round of hybridization seven signals were observed. Subsequent plaque purification resulted in two of these being identified as definitely positive. Phage DNA for each of these was isolated from stocks of singly hybridizing plaques, cleaved with *Not*1, and cloned immediately into Bluescript SK(+); the two genomic fragments were designated λTsPAL311 and λTsPAL511. Detailed restriction analysis, combined with probing using pcrPAL and the sequencing of single clones from within each of the hybridizing regions, enabled the position and orientation of the PAL genes within the two fragments to be established. λTsPAL511 was found to possess three PAL genes (PAL1, 2, and 3) clustered together in tandem, while a single gene (PAL4) was located on λTsPAL311 (fig. 5.4). The PAL genes within λTsPAL511 are all orientated in the same direction.

The gene PAL1 was sequenced in entirety (fig. 5.5). Comparison of the nucleotide sequence with those from other plants showed that it indeed belongs to a PAL gene, which consists of two exons separated by a 617bp intron. Excluding the intron and final stop codon, the coding sequence is 2175bp, corresponding to a predicted protein of 725 amino acids (~ 80 kDa). The presence and position of the intron, splitting an arginine codon between the second and third bases, is identical to that found for all PAL genes which have had their sequence established from genomic clones. Use of Staden's (1984) "Positional Base Preference" and Fickett's (1982) "Testcode" programmes confirmed the existence of the two exon/one intron structure (fig. 5.6). The intron is AT-rich (79%), which as noted in chapter 4 is a characteristic of these sequences in plants (4.2.2); the nucleotides of the 5' exon-intron and 3' intron-exon splice junctions

conform closely to consensus sequences published for these regions (Brown, 1986). A potential polyadenylation signal lies 52 to 57 nucleotides downstream from the stop codon.

The homology between the PAL nucleotide and predicted amino acid sequences presented here, and those of other plants, is much greater in the second exon than in the first, a fact noted in previous comparisons of PAL sequences (Cramer et al., 1989; Gowri et al., 1991). For instance, the clover PAL1 gene and alfalfa APAL1 genes (Gowri et al., 1991) bear 85% and 91% homology at the nucleotide and protein levels in the first exon, respectively; in the second exon the two genes are 90% and 95% homologous at the nucleotide and protein levels, respectively. As would be expected, the nucleotide and derived amino acid sequences of PAL1 are much less homologous to non-legumes. The lack of conservation of the first exon between PAL genes is even more evident in this case, with the sequences of rice (Minami et al., 1989) and sweet potato (Tanaka et al., 1989) being approximately 60bp shorter. The second exons of clover PAL1 and this rice PAL gene are about 60% and 70% homologous at the nucleotide and protein levels respectively.

In the ~500bp of 5' non-coding sequence of the clover PAL1 gene which was sequenced, there are two motifs, TCAACTAACCCA (-346 to -335, with respect to the start point of translation) and CCCTACCTACCA (-296 to -285), which have substantial homology to sequences in other phenylpropanoid gene promoters (Lois et al., 1989; see 4.2.5; fig. 4.21). The occurrence of these motifs in both the clover CHS and PAL genes (cf. 4.2.5) suggests that they are important in the regulation of transcription.
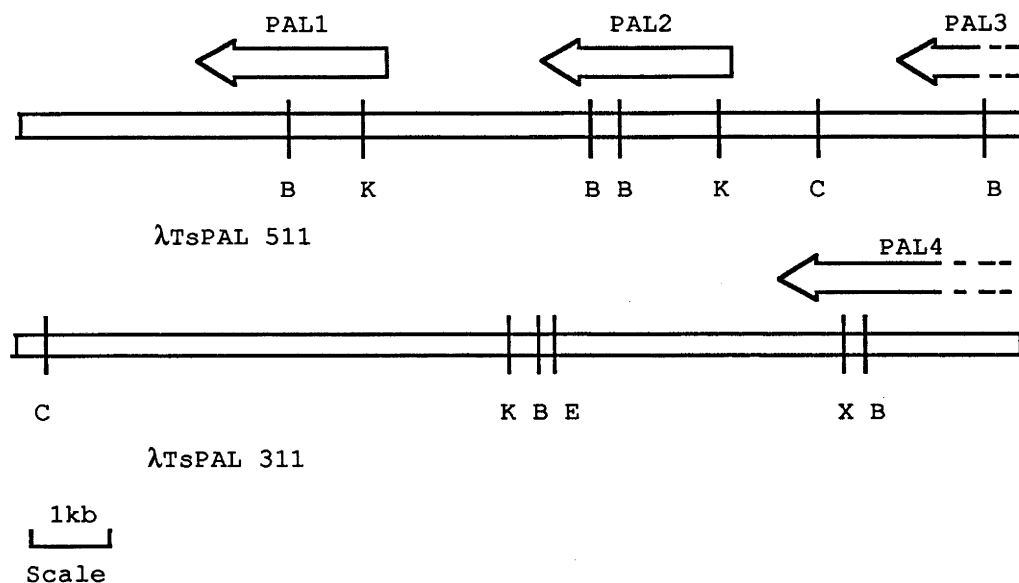
Fig. 5.4 Restriction maps for two subterranean clover genomic clones containing, in total, four different PAL genes. The positions of genes within the clones, and their directions of transcription, are indicated by the arrows. PAL3 is not completely represented upon clone λTsPAL511 (part of the second exon and the first exon are absent); the presence or absence of the first exon of PAL4 on λTsPAL311 was not established. B - BamH1, C - Cla1, E - EcoR1, K - Kpn1, X - Xba1. Other BamH1 and Xba1 sites lie towards the left of the Kpn1 site of clone λTsPAL311; these were not mapped accurately, and as such are not shown.

Fig. 5.5 (overleaf) Nucleotide and derived amino acid sequences of subterranean clover PAL1. The gene consists of two exons separated by an intron; the latter is shown in italics. Approximately 350bp of 5' non-coding sequence and 150bp of 3' non-coding sequence are also shown. The potential TATA box and polyadenylation sites are underlined, as are the two upstream promoter elements conforming to the consensus sequences presented in Lois et al., 1989. Numbers refer to the positions of the adjacent nucleotides in relation to the start point of translation.

```
CTCATAATTAATTTTTCAACTAACCCACTTATTTTCTCTACGTACTGCTTGTGCAACTTGTCTCTCCCTACCTACCAAACCCACACATGCA -271
TAATAATAAGAGAGAGTTAATAATAATTACAATAATGCATATTAATGTAGCCTCCAAAATATACTTTATATTTTATTTTATTTTGATGCCA -181
AACACACCTCTATATAAAGCTCAACAACTCTTCCTTCACATCTCAAGAAATTCACAACTCTTTCATTCTTTCATAATCATTTGTTTTAAA  -91
TTAATTTCCATTCTCTCTCAAAATATTCTATAGCTACCAAAACATAACAAAGTAATTAACACCTATTACTATTATTAGTATCAATTTAAG   -1
ATGGAAGTAGTAGCAGCAGCCATACTAAAAAACAACATTAATGATTATGATTCATTTTGCTTAACTCATGCTAATGCTAATAACATGAAA   90
 M  E  V  V  A  A  A  A  I  L  K  N  N  I  N  D  Y  D  S  F  C  L  T  H  A  N  A  N  N  M  K

GTGAATGCTGCTGATCCTTTGAATTGGGGTGGTGGCGGCTGAGGCAATGAAGGGGAGTCACTTGGATGAGGTGAAGCGTATGGTGGAGGAG  180
 V  N  A  A  D  P  L  N  W  G  V  A  A  E  A  M  K  G  S  H  L  D  E  V  K  R  M  V  E  E

TACCGGAAACCAGTTGTCCGTCTTGGTGGTGAGACACTGACGATTTCTCAGGTGGCTGCCATTGCTGCACACGATGGTGCCACGGTGGAG  270
 Y  R  K  P  V  V  R  L  G  G  E  T  L  T  I  S  Q  V  A  A  I  A  A  H  D  G  A  T  V  E

CTGTCCGAATCTGCTAGAGCTGGCGTTAAGGCGAGCAGTGATTGGGTTATGGAGAGTATGAACAAAGGTACCGACAGTTATGGTGTCACC  360
 L  S  E  S  A  R  A  G  V  K  A  S  S  D  W  V  M  E  S  M  N  K  G  T  D  S  Y  G  V  T

ACAGGGTTCGGCGCTACCTCACACCGCCGAACAAAACAAGGTGGTGCTTTGCAGAAAGAACTCATAAGGTAATATACAATGTTATATATG  450
 T  G  F  G  A  T  S  H  R  R  T  K  Q  G  G  A  L  Q  K  E  L  I  R

AAGTTTTATAGAATAATTTTTTTTTTCTTAGCCCCCTAATTAATATTGGAGTATAGAGAACTATAGGGATGCTAGAGAGATGATGAAAGGA  540
AAATAAGTACAAAAAAAATGATAAGTGATGTGATACGGAAAAAAGCGAATACAAATTAAAGGAATGTTAAAGATTAATGAATAACATTTT  630
CCTGAAAAAATATGAAATTAATGTATATTAATTAGTAAGTGGTTTAAAAAAAAGTATATTAGTAATTATAACTTTTTTAGTTTTTTTTTT  720
TTTTGGAAAGGCCAAGAAAAAATATGAAAGTATATTAATAAAAAAAATTTTGAACTCAAAGTATATTAATAATTATAATTATAACTTTTT  810
TATTTGCCTTTTTTATTTTTTATGTAAGTGACTTCAGGTTTAACTTAGGGTAATTATATTTCTTTTAAATTATTAAATCTCTAATTGTA  900
CATCTAATAATTCATTTATTCATTTTTTGAATTATAATTACATCAATCATCAATCGTGAATCTCTAAGATGGATTTATTTTATGCTTAAA  990
AGAATATATTAAACAAAAGGTCTTAATATTAATTCTTCTCCCTTCATTTTTGCAGGTTTTTGAATGCTGGAATATTTGGAAACGGAACCG 1080
                                                              F  L  N  A  G  I  F  G  N  G  T

AGTCAAACCACACCCTACCACACACAGCAACACGAGCAGCCATGTTAGTAAGAATCAACACACTTCTCCAAGGTTATTCAGGAATTAGAT 1170
 E  S  N  H  T  L  P  H  T  A  T  R  A  A  M  L  V  R  I  N  T  L  L  Q  G  Y  S  G  I  R

TTGAAATATTAGAAGCTATAACCAAGCTCCTTAACAACAACATTACTCCATGTTTACCACTTCGCGGTACAATCACAGCTTCAGGAGATT 1260
 F  E  I  L  E  A  I  T  K  L  L  N  N  N  I  T  P  C  L  P  L  R  G  T  I  T  A  S  G  D

TAGTCCCACTTTCTTACATTGCTGGTTTACTAACTGGTAGATCAAATTCAAAAGCTCATGGACCCTCTGGAGAAATGCTTAATGCAAAGG 1350
 L  V  P  L  S  Y  I  A  G  L  L  T  G  R  S  N  S  K  A  H  G  P  S  G  E  M  L  N  A  K

AAGCTTTTCAATTAGCAGGAATCAATGCCGAGTTCTTTGAATTACAACCGAAAGAAGGTCTTGCACTTGTTAATGGAACTGCGGTTGGTT 1440
 E  A  F  Q  L  A  G  I  N  A  E  F  F  E  L  Q  P  K  E  G  L  A  L  V  N  G  T  A  V  G

CTGGTTTAGCTTCTATTGTTCTGTTTGAGGCTAACATATTGGCAGTGTTGTCTGAAGTTTTATCTGCAATTTTCGCTGAAGTTATGCAAG 1530
 S  G  L  A  S  I  V  L  F  E  A  N  I  L  A  V  L  S  E  V  L  S  A  I  F  A  E  V  M  Q

GGAAACCCGAATTTACTGATCATTTGACACATAAGTTGAAGCACCACCCTGGTCAAATTGAGGCCGCTGCTATTATGGAACACATTTTGC 1620
 G  K  P  E  F  T  D  H  L  T  H  K  L  K  H  H  P  G  Q  I  E  A  A  A  I  M  E  H  I  L

ATGGGAGTGCTTATGTCAAAGATGCAAAGAAGTTGCATGAAATGGATCCTTTGCAGAAACCAAAACAAGATAGATATGCACTACGAACTT 1710
 H  G  S  A  Y  V  K  D  A  K  K  L  H  E  M  D  P  L  Q  K  P  K  Q  D  R  Y  A  L  R  T

CGCCACAATGGCTTGGTCCTTTGATTGAAGTCATTAGATTCTCTACCAAGTCAATTGAGAGGGAGATTAACTCAGTCAATGATAACCCTT 1800
 S  P  Q  W  L  G  P  L  I  E  V  I  R  F  S  T  K  S  I  E  R  E  I  N  S  V  N  D  N  P

TGATTGATGTTTCAAGAAACAAAGCTTTGCACGGTGGAAATTTTCAAGGAACACCTATCGGAGTATCCATGGATAATACACGTTTGGCTC 1890
 L  I  D  V  S  R  N  K  A  L  H  G  G  N  F  Q  G  T  P  I  G  V  S  M  D  N  T  R  L  A

TTGCATCAATTGGAAAACTTTTTGTTTGCTCAATTCTCTGAGCTTGTCAATGATTTTTACAACAATGGGTTGCCTTCAAATCTCTCTGCTA 1980
 L  A  S  I  G  K  L  L  F  A  Q  F  S  E  L  V  N  D  F  Y  N  N  G  L  P  S  N  L  S  A

GTAGAAATCCGAGCTTGGATTATGGGTTCAAGGGATCTGAAATTGCCATGGCTTCCTATTGTTCCGAGTTACAATATCTTGCAAATCCAG 2070
 S  R  N  P  S  L  D  Y  G  F  K  G  S  E  I  A  M  A  S  Y  C  S  E  L  Q  Y  L  A  N  P

TTACAACCCATGTCCAAAGTGCCGAGCAACACAACCAAGATGTGAATTCTTTGGGTTTAATTTCTTCTCGAAAAACAAAAGAAGCCATTG 2160
 V  T  T  H  V  Q  S  A  E  Q  H  N  Q  D  V  N  S  L  G  L  I  S  S  R  K  T  K  E  A  I

AGATTCTTCAACTCATGTCTTCCACATTCTTGATTGCACTTTGCCAAGCAATTGATTTAAGACATTTGGAGGAGAACTTGAAAAACTCAG 2250
 E  I  L  Q  L  M  S  S  T  F  L  I  A  L  C  Q  A  I  D  L  R  H  L  E  E  N  L  K  N  S

TCAAAAATACCGTAAGCCAAGTGGCCAAAAAGACCTTAACTATAGGTGTCAGTGGAGAACTTCATCCTTCAAGATTTTGCGAAAAAGACT 2340
 V  K  N  T  V  S  Q  V  A  K  K  T  L  T  I  G  V  S  G  E  L  H  P  S  R  F  C  E  K  D

TGTTGAAAGTGGTTGATAGGGAGCATGTGTTTTCCTACATTGACGATCCTTGTAGTGCTACATACCCATTGGCACAAAAACTCAGGCAAG 2430
 L  L  K  V  V  D  R  E  H  V  F  S  Y  I  D  D  P  C  S  A  T  Y  P  L  A  Q  K  L  R  Q

TGCTAGTAGATCATGCTTTAGTTAATGGAGAAAGTGAGAAAAAATTCAAACACATCAATCTTCCAAAAGATTGCAACTTTTGAGGAAGAGT 2520
 V  L  V  D  H  A  L  V  N  G  E  S  E  K  N  S  N  T  S  I  F  Q  K  I  A  T  F  E  E  E

TGAAGACCCTCTTGCCAAAAGAGGTTGAAAGTGCAAGGACTGCATATGAGAATGGAAATTCAACAATTGCAAACAAGATTAATGGATGTA 2610
 L  K  T  L  L  P  K  E  V  E  S  A  R  T  A  Y  E  N  G  N  S  T  I  A  N  K  I  N  G  C

GATCTTATCCGCTTTATAAGTTTGTGAGAGAGGAGTTGGGAACAAGTTTGCTAACCGGAGAACGTGTCATTTCACCAGGTGAAGAGTGTG 2700
 R  S  Y  P  L  Y  K  F  V  R  E  E  L  G  T  S  L  L  T  G  E  R  V  I  S  P  G  E  E  C

ACAAACTATTCACAGCTATGTGTCAAGGGAAAATCATTGATCCTCTTCTTAAATGTTTAGGAGAGTGGAATGGTGCTCCTCTTCCAATTT 2790
 D  K  L  F  T  A  M  C  Q  G  K  I  I  D  P  L  L  K  C  L  G  E  W  N  G  A  P  L  P  I

GTTAACTTTTTTAGTTAAGTTTAAAAAATGTTTTATTTGTATCTATGCAAGTGTTCCAATAATCATTTGATTTTTCATGCCTTTTTTTTTT 2880
 C

ATAAATGATTTTTCATGCCTTTGACAAAATTAATACATTGAAAGTAAGCTTCATTTCAAT
```

pcrPAL (fig. 5.7). The most useful hybridization, as far as copy number was concerned, was that to the digest of the genomic DNA with *Bam*H1 (lane 3). The two *Bam*H1 subfragments of λTsPAL511 (4.0kb and 4.8kb) were observed to strongly hybridize (as expected); a band of similar intensity at 3.5kb *may* indicate the remainder of λTsPAL511 to the left of the 4.0kb BamH1 fragment, as a terminal *Bam*H1 site exists at the left end of λTsPAL511 [i.e. a *Bam*H1 site at the end of a genomic fragment may be produced by a genuine *Bam*H1 site in the genomic DNA being cleaved by *Sau*3A and reformed upon cloning into the λ Dash 11 *Bam*H1 site, or by a sequence containing only "part" of a *Bam*H1 site (e.g. TGATCC) being cloned into the phage *Bam*H1 site]. Hybridization and brief sequence analysis had shown that the sequence of pcrPAL straddled a *Bam*H1 site within PAL4 (that close to the *Xba*1 site in fig. 5.4). The majority (~90%) of the PAL4 sequence corresponding to pcrPAL was on the 2.1kb fragment, at the right hand end of λTsPAL311; this may be represented by the 2.8kb hybridizing fragment. The remainder of the pcrPAL sequence of PAL4 is located on the 4.3kb *Bam*H1 fragment of λTsPAL311, which in the genomic digest and blot lies faintly between the λTsPAL511 4.0kb and 4.8kb *Bam*H1 fragments.

Other than those just discussed, no other significantly hybridizing *Bam*H1 fragments could be observed. Characterization of the two genomic fragments indicates that the clover PAL multigene family must consist of at least four members, while the Southern blot data suggests that the family may contain no further members. PAL has also been found to be a multigene family in rice (Minami et al., 1989), bean (Cramer et al., 1989), parsley (Lois et al., 1989), *Arabidopsis thaliana* (Ohl et al., 1990), soybean (Estabrook and Sengupta-Gopalan, 1991) and alfalfa (Gowri et al., 1991).
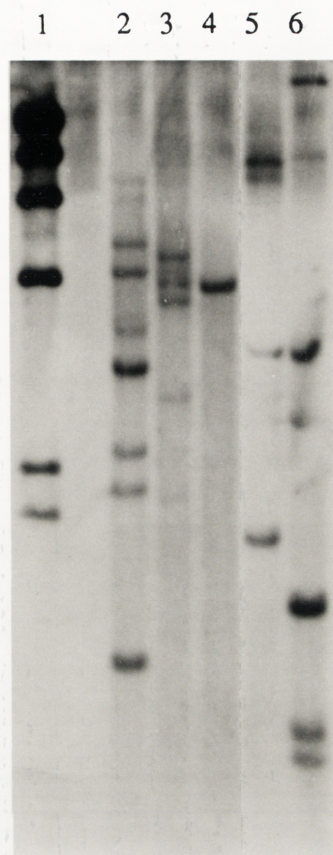
Fig. 5.7 Hybridization of the 537bp *Trifolium subterraneum* L. cv. Karridale PAL (PCR) fragment onto a Southern transfer of genomic DNA isolated from the same plant and cleaved with a range of restriction enzymes.

Lane 1: λ/*Hin*d111 standard markers (23.1, 9.4, 6.6, 4.3, 2.3, 2.0).

Lane 2 - *Bgl*11; lane 3 - *Bam*H1; lane 4 - *Kpn*1; lane 5 - *Eco*R1; lane 6 - *Hin*d111.

5.3 Discussion

Initial probing of the clover gene library with the bean PAL5 cDNA did not result in any positive hybridization. This was not expected, as the same probe had been successfully used to isolate a PAL gene from a gene library of the more distantly related *Arabidopsis thaliana* (Ohl et al., 1990). It is possible that the temperature used for the hybridization (55^DC) was still too stringent for hybridization of this bean sequence to the analogous clover sequences. However, the nucleotide homology of the second exons of the bean and clover PAL genes is similar to that of the the bean CHS sequences and their clover counterparts (the bean PAL5 cDNA covering most of the second exon of a PAL gene). Although unlikely, it should also be considered that no clover PAL gene was present in the plaques which were probed with the bean sequence (a different plating of the library was probed with pcrPAL). Most likely, plain inexperience explains the failure of the bean probe to be used successfully to isolate a clover PAL sequence.

As has been found with many other plants, PAL is present in a multigene family in clover. However, this is the first time that PAL genes have been reported to exist in a cluster. In the case of bean, although at least some of the CHS genes are clustered like those of clover, no evidence of PAL gene clustering could be obtained (Ryder et al., 1987; Cramer et al., 1989). It was hence suggested that the difference in clustering between the bean CHS and PAL genes may be due to the *latter* evolving into a gene family at a earlier date, hence allowing more time for the locational divergence of the constituent genes (Cramer et al., 1989). If this is indeed correct, and presuming independent origins of the PAL multigene families of clover and bean (i.e. they are not inherited from a common

ancestral legume), the PAL genes of clover may either have been generated at approximately the same time as the CHS genes, or at an earlier stage but have not all separated. It would be interesting to obtain the sequences of all the clustered clover PAL genes and ascertain whether they are more similiar in sequence to each other than the sequences of the non-clustered bean PAL genes are to each other.

Protein purification studies have demonstrated that the"native" PAL enzyme in plants exists as a tetramer (Jorrin and Dixon, 1990; Gowri et al 1991 and references there-in). In some plants with multigene PAL families, different charged forms of the tetramer have been found; these could result from the subunits coded for by any individual gene only associating with each other, and/or combinations of subunits coded for by different genes. It is likely that a number of PAL protein isoforms will be discovered in clover. At least in the case of parsley, the PAL subunits from one gene can form an active enzyme (tetramer), excluding the possibility that the different genes in the PAL family of this plant are all needed to produce a functional protein (Schulz et al., 1989).

Finally, comparison of the sequence of the 521 bp pcrPAL with that of the analogous region of PAL1 revealed that there were 30 differences (i.e. 94% homology). This would suggest that the PCR product did not arise from PAL1, but from another member of the clover PAL gene family.

Chapter 6. Clover CHS gene expression

6.1 Primer extension

As noted in chapter 1 (1.4.1-3), CHS gene expression is frequently induced by environmental stress factors such as exposure to excessive light and attack by a pathogen. The former of these was selected as a regime under which to grow subterranean clover plants (2.16), the RNA from which was isolated and examined for CHS expression by primer extension. Primers complementary to the leader sequences of CHS2, CHS3 and CHS4 were designed, care being taken that each did not hybridize to any of the other CHS sequences (2.16.4).

The only gene and time point for which a product could be observed using these conditions was CHS3 at 15 hours. The site of transcription initiation for this gene is 99+/-1 nucleotides upstream from the methionine start codon (fig. 6.1/2). This number is that for the furthest extension product, but as fig. 6.2 shows, other bands of smaller molecular weight were also detected. There are at least four possible reasons for the appearance of shorter transcripts (Williams and Mason, 1985): (1) the reverse transcriptase may be terminating at the methylated nucleotide adjacent to the cap site; (2) multiple sites for transcription initiation; (3) cross hybridization to another member of the CHS multigene family (however, as noted above, the primer was checked against the sequences of the other genes and was found not to be complementary); (4) premature termination of the reverse transcriptase due to secondary structure.

This result, of course, demonstrates that the CHS3 gene is light inducible. However, under these conditions used for the isolation of RNA, no transcripts could be detected for CHS2 and CHS4. The inability to pick

up transcripts for these genes does not necessarily imply that they are pseudogenes, especially considering that their nucleotide and derived amino acid sequences are highly homologous to those of CHS3 (4.2/3). In an analogous primer extension study of three genes of the six member soybean CHS multigene family, a transcript for only one of the genes was detected (Wingender et al., 1989). However, six CHS proteins are found in cell cultures (Welle and Grisebach, 1987). Likewise, the transcript for one member of the peanut RS multigene family was never detected despite a number of inducing conditions being tried (Lanz et al., 1990). It is likely that CHS2 and CHS4 are expressed in response to other stress factors, or at specific developmental stages not examined in this study.

It should also be noted that no positive control was available for the primer extension (i.e. the transcription start point for no other clover gene has been mapped in this laboratory). Hence, the inability to pick up transcripts for CHS3 at other time points, and for the other genes, may be due to the extension experiments not working consistently. The quality of RNA in some isolations was possibly poor (i.e. at least partially degraded), and at the time of writing clover CHS expression has not been demonstrated by northerns. Obviously, further work is needed to establish reliable northern/primer extension methods for examining gene expression in clover.

5'**TATATATATAA**CCATCTCATCCCACACATTTCCACCACAACACTCTTCATCTCATTCCTTTTCTAT

AAAAACAACTGTGTTCAAATTACACATTGAAACTACTAAAGATATATTAAGAT**ATG** 3'
 3'TTGTTGACACAAGT 5'

Fig. 6.1 Diagram showing the leader sequence of CHS3 and the position where the primer extension oligonucleotide (underlined) was designed to hybridize. Extension of the oligonucleotide with reverse transcriptase and deoxynucleoside 5' triphosphates would occur towards the TATA box. Both the TATA box and methionine translation start codon of CHS3 are shown in bold type.
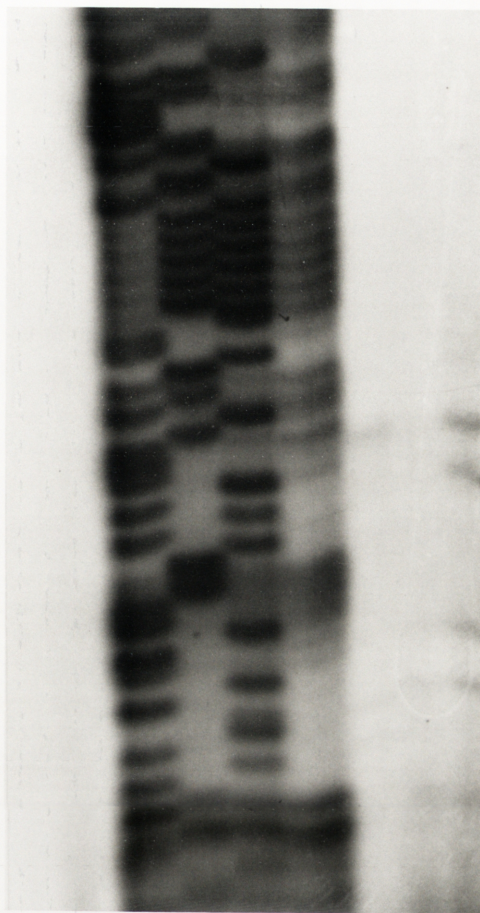
Fig. 6.2 Primer extension of subterranean clover CHS3. RNA was isolated from plants grown without light for seven days (time point zero), and after 15 hours of exposure to light. The extension products were electrophoresed against standard sequencing markers obtained from using the primer to sequence DNA from the appropriate orientation of CHS3.

 GATC - standard sequencing markers.

 1,2 and **3** - Three results for extending 15 hour light RNA with the CHS3 primer.

 4 - Result for extending zero time point RNA with the CHS3 primer.

CHS3 primer - 5' TGAACACAGTTGTT 3'.

The sequence of the standard markers, reading from the three As situated one third of the way up from the bottom, is

5' AAATGTGTGGGATGAGATGGTTATATATATATG 3' etc.

6.2 Analysis of promoter activity of CHS3 in transient expression assays

6.2.1 Introduction

To characterize the role of a plant promoter in controlling transcription it is now commonplace to fuse the promoter to a reporter gene and assay for the protein product of this gene. *Transient gene expression* involves introducing the recombinant constructs into protoplasts by electroporation and/or polyethylene glycol mediated transformation. Alternatively, binary vectors can be used to introduce the promoter-reporter gene constructs stably into the genome, thus creating a transgenic plant. The most obvious advantage of transient gene expression is that the the activity of the reporter gene, and hence the ability of the fused promoter to activate transcription, can be measured in days, whereas months are usually needed before the activity of a promoter in a transgenic plant can be evaluated. Furthermore, the expression of a gene in protoplasts is not influenced by genome position effects, a drawback that can lead to transgenic plants only expressing poorly the introduced gene. However, transient gene expression suffers from the major problem that it can only give limited, if any, information on the tissue and developmental expression of a gene; such data can be easily forthcoming from analysis of transgenic plants (Bustos et al., 1991). At present the chloramphenicol acetyltransferase (CAT), neomycin phosphotransferase (NPT), luciferase and β-glucuronidase (GUS) genes are the most important reporter genes available for transient expression assays in plant protoplasts (Bates et al., 1988).

To obtain an insight into the expression and inducibility of the clover CHS3 gene, transient expression studies involving the GUS reporter gene were conducted in a protoplast system. Originally, the idea was to isolate

protoplasts from clover, or Lucerne (a fellow legume) cell suspension cultures, but the extrememly slow and poor growth of such cultures meant that an alternative system had to be found. A cell suspension culture of the plant *Nicotiana plumbaginifolia* was obtained from the C.S.I.R.O. Division of Plant Industry and proved not only fast growing but also very amenable to the isolation of protoplasts (fig. 6.3). The CHS3 promoter was cloned behind the GUS gene of pQ20 and a series of deletions into this parental construct created ("CHS-GUS-VIC" plasmids). These were analysed in the protoplast system, as was the effect of the addition of the compound glutathione, which previous studies have found to be capable of inducing the expression of CHS promoters in protoplasts (Dron et al., 1988; Choudhary et al., 1990a/b). Glutathione may act as an antioxidant and scavenger of electrophiles that are released during the elicitation process (Edwards et al., 1991).

Phenylpropanoid pathway genes are known to be induced by such factors as microbial attack, wounding, and exposure to UV light (chapter 1). It would be expected that in an ideal protoplast system that the endogenous phenylpropanoid pathway genes, and any introduced reporter gene under the control of a such a gene promoter, would possess low if any activity until induced by environmental factors. This is indeed the case with parsley protoplasts, whose defence responses are not activated by the protoplast preparation process, but *subsequently* can be by exposure to elicitor or UV light (Dangl et al., 1987). However, the preparation and culture of protoplasts from soybean has been found to result in the stress responses that are normally observed after microbial attack or treatment with an elicitor (Mieth et al., 1986). The problem of CHS induction by endogenous elicitors, or factors released during the protoplast preparation procedure, is a central feature of the results presented in this section.
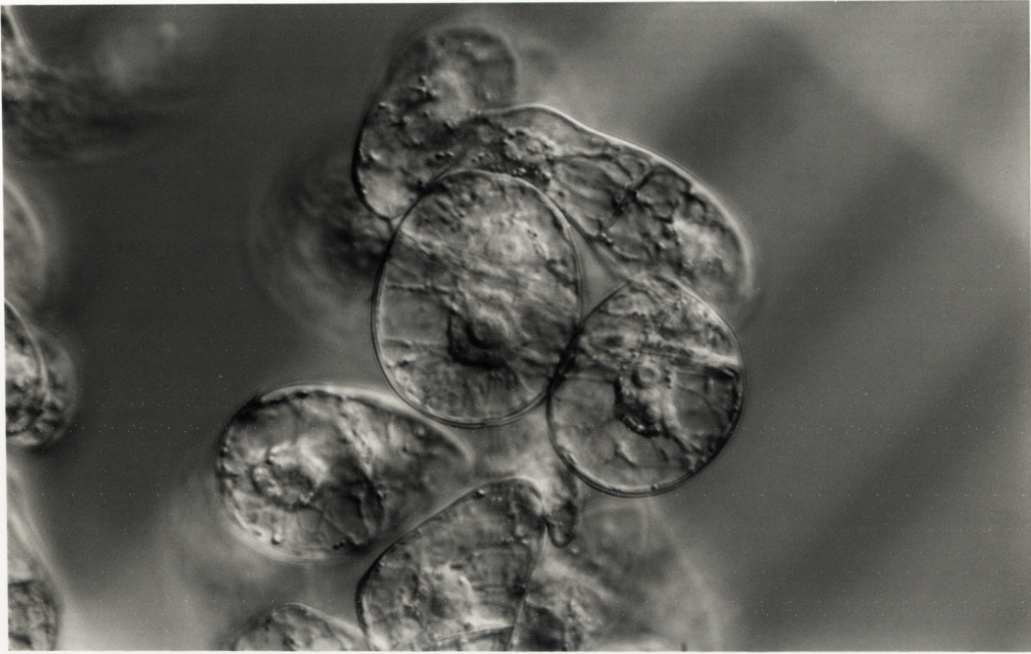
Fig. 6.3

Top - *Nicotiana plumbaginifolia* suspension culture cells

Bottom - Protoplasts isolated from the above cell suspension culture

Both photographs kindly taken by Dr. Robert Ridge.

6.2.2 Methods

6.2.2.1 Construction of parental GUS plasmid

A Bluescript clone of the subterranean clover CHS3 gene containing approximately 2,200bp of promoter was used as the source of the sequence which was cloned in front of the GUS gene. Oligonucleotide directed mutagenesis was used to create a *Nco*1 site at the translation start codon of CHS3. A *Hind*111-*Nco*1 fragment containing the entire 2,200bp of CHS3 5' flanking sequence was cloned in front of the GUS gene of plasmid pQ20 (2.17.1), producing a plasmid called pCHS300. By using this method the entire sequence of the CHS3 promoter between the methionine start codon and the TATA box could be preserved intact. In many fusion experiments, restriction sites which are fortuitously situated in this region are used as the cloning site which joins the promoter to the reporter gene, and frequently they are not very close to the start codons. To give just one example from regulatory studies of phenylpropanoid pathway genes, the *Arabidopsis thaliana* PAL1 promoter was cloned in front of the GUS gene by making use of a *Bgl*11 site ~50 bp upstream form the start codon; in this situation it always remains possible that sequences in the omitted leader sequence could be important in the regulation of expression. The mutagenesis method employed here overcomes this potential problem. The *Nco*1 fusion between the CHS3 promoter and the GUS gene was confirmed by sequencing (fig. 6.4).

6.2.2 2 Construction of nested deletions.

A series of promoter deletions of pCHS300 were created by making use of conveniently situated restriction sites in the sequence (fig. 6.5). The fusion between the promoter sequence and GUS gene in each of these constructs did not have to be checked by sequencing as they were all

```
T A A A  G A T A T A T  T A A  G C C A T G G  T C C  G T C C  T G  T A
T A A A  G A T A T A T  T A A  G C C A T G G  T C C  G T C C  T G  T A
```

Fig. 6.4 Confirming, by (fluorescent dye) DNA sequencing, the construction of a *Nco*1 site at the translation start point of CHS3 and its correct fusion to the GUS gene. The relevant sequence is :

5' TAAAGATATATTAAG**CCATGG**TCCGTCCTGTA 3'

The *Nco*1 site is printed in bold. That sequence to its 5' (left) is from the leader region of CHS3, while that to its 3' (right) is of the GUS gene (Jefferson et al., 1986; Clontech (Palo Alto, California) catalog (1989/90), pg. 18.6).

derived from the parent, which as stated above was analysed for the correct (re)formation of the *Nco*1 site.
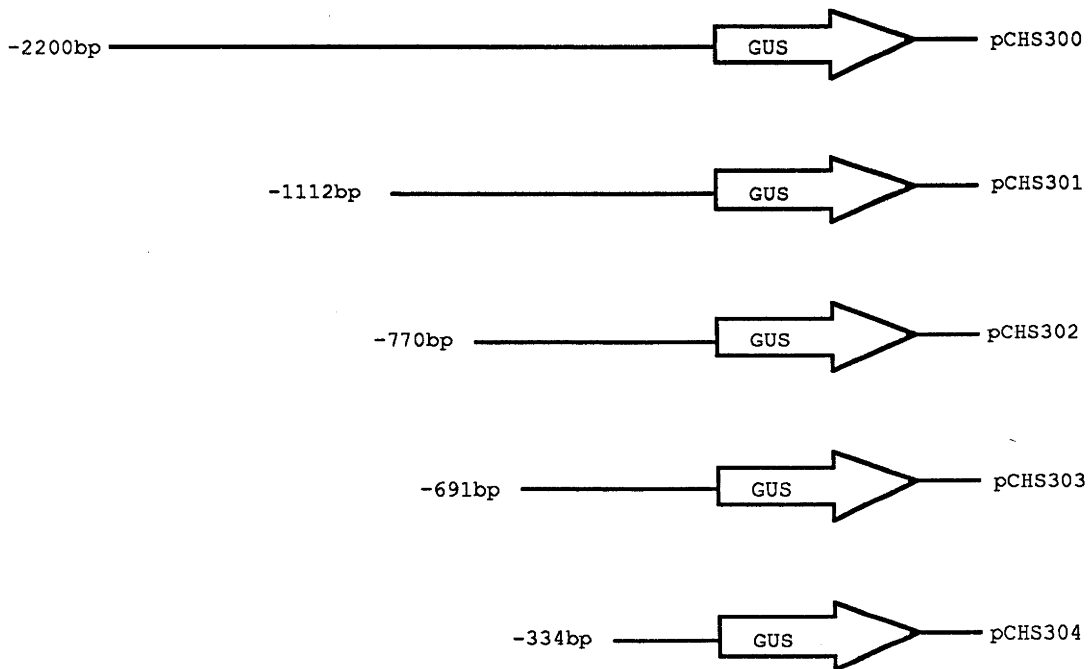
Fig. 6.5 Diagram of the CHS-GUS-VIC plasmids showing the lengths of clover CHS3 promoter which were fused to the GUS gene.

6.2.2.3 Construction of control plasmids

Two control plasmids were constructed. A promoterless plasmid, p113, was made by digesting pQ20 with *Pst*1 and *Nco*1, filling in the ends with the Klenow fragment and ligating them together (2.7.8). The 302bp vicilin promoter segment (2.18.1) was hence removed (fig. 6.6). Sequencing was used to verify the survival of the GUS gene's ATG start codon in the construct. A positive control, p115 was provided by cleaving the plasmid pBI121 (Clonetech Laboratories Inc, California, U.S.A.) with *Hind*111 and *Bcl*1 and cloning the fragment containing the cauliflower mosaic virus (CMV) promoter into pQ20 which had also been cleaved with this pair of enzymes. The reformation of the *Bcl*1 site, within the GUS gene, was   confirmed by sequencing. Hence the CMV promoter,

which is known to be highly expressed / (Jefferson, 1897), was placed

5' *to* the GUS gene bearing the same 3' termination sequence as the

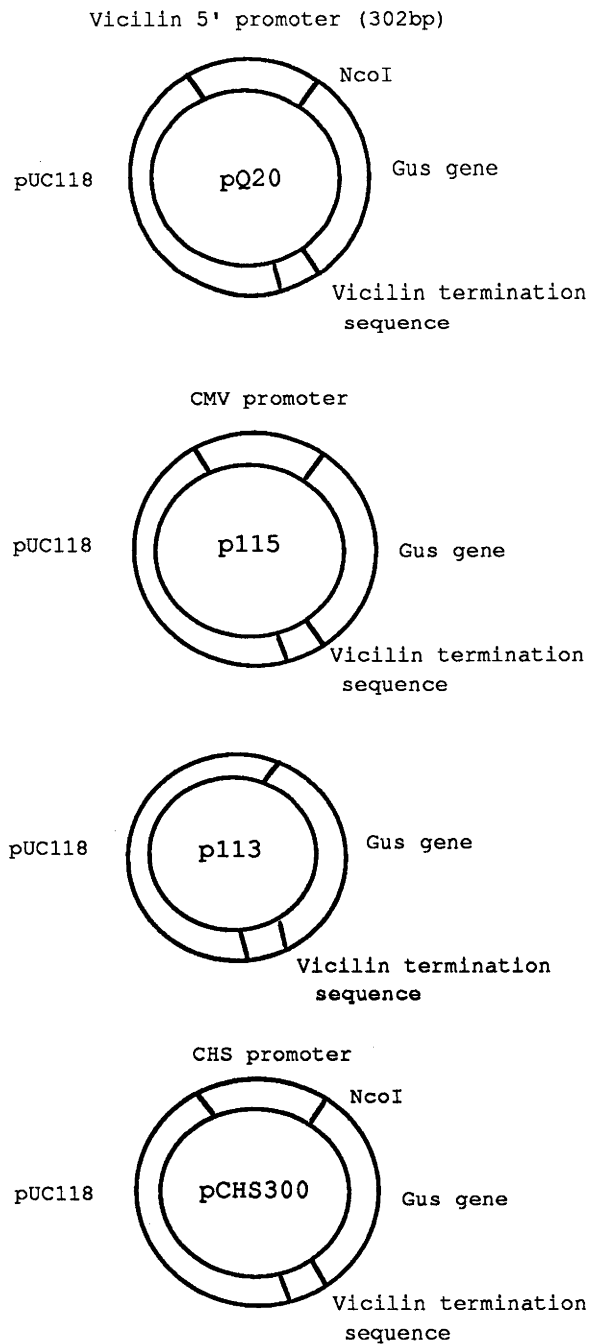CHS promoter constructs (fig. 6.6).

6.2.3 Results

6.2.3.1 Electroporation conditions

The electroporation protocol for *N. plumbaginifolia* protoplasts, which

formed the background to these experiments, had been developed at the

C.S.I.R.O. Division of Plant Industry, Canberra (Taylor and Larkin,

1988). Although the recommended electroporation buffer could easily be

made up, the "optimal" electrical conditions could not be duplicated on

the Biorad Gene Pulser machine. Hence, in the initial series of

experiments, the object was to establish the conditions which produced

the highest activity of the introduced GUS genes as possible.

The C.S.I.R.O. protocol involved using three 300V pulses (delay of

100msec), each with a length of 5msec, through a 24μF capacitance and

having a 0.4cm gap between the electrodes. Just prior to electroporation,

the protoplasts were resuspended in a high salt buffer called TBS (30mM

tris, 150mM NaCl, 6mM $CaCl_22H_2O$, 253mM mannitol) at pH9.0, which

experiments had shown to produce high levels of reporter gene

expression (Taylor and Larkin, 1988). With the Biorad Gene Pulser it is

not possible to obtain delays between pulses of 100msec, and the "time

constant" of this machine does not correlate with the pulse length

mentioned above.

When *N plumbaginifolia* protoplasts were suspended in the above

buffer, pulsed once and three times (delay of ~5 seconds) in 0.4cm Biorad

cuvettes with the Biorad machine using conditions of 300V/25μF, the

levels of p115 GUS expression the next day (~16 hours), following 2 hours

Vicilin 5' promoter (302bp)
NcoI
pUC118   pQ20   Gus gene
Vicilin termination
sequence

CMV promoter
pUC118   p115   Gus gene
Vicilin termination
sequence

pUC118   p113   Gus gene
Vicilin termination
sequence

CHS promoter
NcoI
pUC118   pCHS300   Gus gene
Vicilin termination
sequence

Fig. 6.6 Diagram of the plasmids pQ20, p113, p115, pCHS300 used in the transient expression assays. Only the most important restriction sites are shown. p113 and p115 were negative and positive controls, respectively. pCHS300 was derived from pQ20 by excising the vicilin gene promoter from the latter and cloning in ~2.2kb of subterranean clover CHS3 promoter. The other clover promoter GUS fusion plasmids (pCHS301, pCHS302, pCHS303, and pCHS304) were created by deleting upstream fragments from pCHS300. All plasmids possessed ~200bp of vicilin 3' non-coding sequence at the end of the GUS gene.

of incubation with the MUG substrate, were not measurable above background. The time constants gained for each pulse were approximately 1.5msec.

In order to increase the time constant, and thus hopefully gene expression, the capacitance was increased from 25µF to 125µF. Separate aliquots of protoplasts were electroporated, in the presence of p115 and p113, with either a single pulse of 300V with a capacitance of 125µF across 0.4cm (a field strength of 750V/cm), or three pulses under these conditions. Time constants of ~5-6 msec were obtained. The viability of protoplasts assayed the following day (16 hours) was ~70% for a single pulse, but less than ~30% for those which had been exposed to three pulses. As such, the GUS activity for the latter protoplasts was extremely low; however, higher levels of expression were recorded for those protoplasts which received only one pulse (not shown). Similar conditions of voltage and capacitance (with a single pulse) were employed by one group of researchers working with *N. tabacum* (tobacco) protoplasts (Fujiwara et al., 1991). However, these researchers also favoured a low salt buffer (5mM MES, 70mM KCl, 300mM mannitol, pH5.8). Changing to this buffer with the conditions of 300V/125µF etc, produced a time constant of less than 2msec, and again very low expression. Hence, the TBS buffer proved much more satisfactory with the protoplasts and electroporation machine available, even although the high salt must absorb some of the charge. In one study of the best conditions for the expression of the CAT gene in carrot protoplasts, it was also found that a high salt media with a field strength of 750V/cm was ideal (Bates et al., 1988).

To examine the expression of the GUS plasmids in *N. plumbaginifolia* protoplasts, it was decided to use the TBS pH9.0 buffer and 750V/cm field strength (300V/125µF), and evaluate the activity of the CHS-

A

		1	2	3	4	5
p115	No inducer	19.6	61.1	28.1	69.5	13.6
	Inducer present	30.4	46.7	54.1	98.4	16.2
p113	No inducer	1.7	1.7	2.4	3.4	0.3
	Inducer present	3.0	1.8	2.5	2.3	3.3
p300	No inducer	7.1	*	15.7	13.0	*
	Inducer present	11.8	*	20.3	15.0	*
p301	No inducer	17.5	25.3	52.7	74.7	23.6
	Inducer present	31.0	97.5	78.6	89.2	33.7
p302	No inducer	*	18.9	27.8	62.3	20.1
	Inducer present	*	45.9	29.8	67.6	26.6
p303	No inducer	*	*	27.8	68.4	11.9
	Inducer present	*	*	40.4	79.2	15.2
p304	No inducer	*	*	21.3	53.3	*
	Inducer present	*	*	23.4	61.9	*

B

		6	7	8	9	10	11	12
p115	No inducer	16.0	71.8	35.0	17.8	21.3	8.8	13.1
	Inducer present	13.4	102	35.8	9.3	21.4	12.3	19.7
p113	No inducer	6.1	3.4	2.0	2.7	5.1	1.4	0.3
	Inducer present	6.2	3.1	2.5	3.2	5.2	1.2	1.8
p300	No inducer	*	4.2	*	5.6	*	9.2	8.8
	Inducer present	*	7.2	*	10.4	*	7.7	12.7
p301	No inducer	48.4	25.5	104	28.9	81.6	25.4	9.7
	Inducer present	40.7	131	119	64.5	114	24.0	14.5
p302	No inducer	36.6	10.2	68.8	20.1	94.8	14.7	18.6
	Inducer present	38.9	15.7	99.1	26.6	94.6	25.2	23.3
p303	No inducer	33.0	26.2	*	21.0	*	27.6	21.3
	Inducer present	27.9	39.1	*	18.7	*	42.5	31.0
p304	No inducer	39.4	14.3	*	14.8	87.9	27.0	17.6
	Inducer present	26.1	63.4	*	24.3	90.9	37.5	30.7

Fig. 6.7 Results for expression of the GUS constructs in *N. plumbaginifolia* protoplasts 16 hours (A) and 22 hours (B) after electroporation; the inducer, glutathione, was added 9 and 6 hours prior to harvesting, respectively. Each vertical column within the 16 and 22 hour timepoints corresponds to an independent protoplast batch; there is no relationship between the protoplasts from the two timepoints. Expression data are given as pmoles of 4-methylumbelliferyl-β-D-glucuronide converted to 4-methylumbelliferone/min/mg of protein. Asterisks indicate that the genetic construct (clone) was not tested in that experiment.

three-fold. The presence of glutathione was consistently observed to enhance the expression of the introduced constructs (including the negative control p113). As the relative pattern of expression of the constructs was repeatable between different protoplast batches, an analysis of variance was performed. This statistical analysis (figs. 6.8/9) showed that there was a significant difference in the levels of expression between the positive control (p115) and the CHS-GUS-VIC constructs on the one hand, and the negative control (p113) on the other. Furthermore, the addition of glutathione was shown to cause increased levels of expression for all the constructs; this induction was essentially the same for each construct. As the expression of not only the CHS-GUS-VIC constructs, but also the controls, was observed to be increased to the same level, it would appear that glutathione was influencing the transcription of genes in a general manner (i.e. it was not specifically influencing the expression of phenylpropanoid/defence genes). In such a situation, it is tenuous to state that there was any stress induced expression of the CHS-GUS-VIC constructs; only if the level of induced expression of these constructs was above the "background" induction could it confidently concluded that glutathione was causing stress related phenylpropanoid gene activity.

6.2.4 Discussion

A precondition for the activity of any promoter in a protoplast system is the presence of trans-acting factors which can bind to it and allow transcription of the fused reporter gene to take place. The activity of the subterranean clover CHS promoter GUS fusions in *Nicotiana plumbaginifolia* protoplasts implies that this latter plant contains the trans-acting factors necessary to activate a legume phenylpropanoid

Fig. 6.8

A Results for the expression of the genetic constructs in *N. plumbaginifolia* protoplasts at 16 hours after a statistical analysis of variance was performed on the data presented in fig. 6.7. Data is expressed as the log of the response (log(rep)). It demonstrates a significant clone effect (P<0.001) and a significant inducer (P=0.005) effect. Note that the lack of expression data for pCHS303 and pCHS304 in a number of experiments in fig. 6.7 meant these two clones were excluded from the analysis.

```
clone       p115    p113    p300    p301    p302
log (rep) 3.590   0.659   2.413   3.792   3.446
Standard error of differences:   0.1871
```

Induction caused by the addition of glutathione was uniform across the constructs:

```
Absence of inducer (log (rep)):   2.579
Presence of inducer (log (rep)):  2.981
Standard error of differences:    0.0757
```

B Graph of the log of the response against each clone for the 16 hour timepoint.
C Graph of the log of the response against the absence and presence of glutathione for the 16 hour timepoint.
In each graph LSD (least significant difference) is twice the standard error of differences.
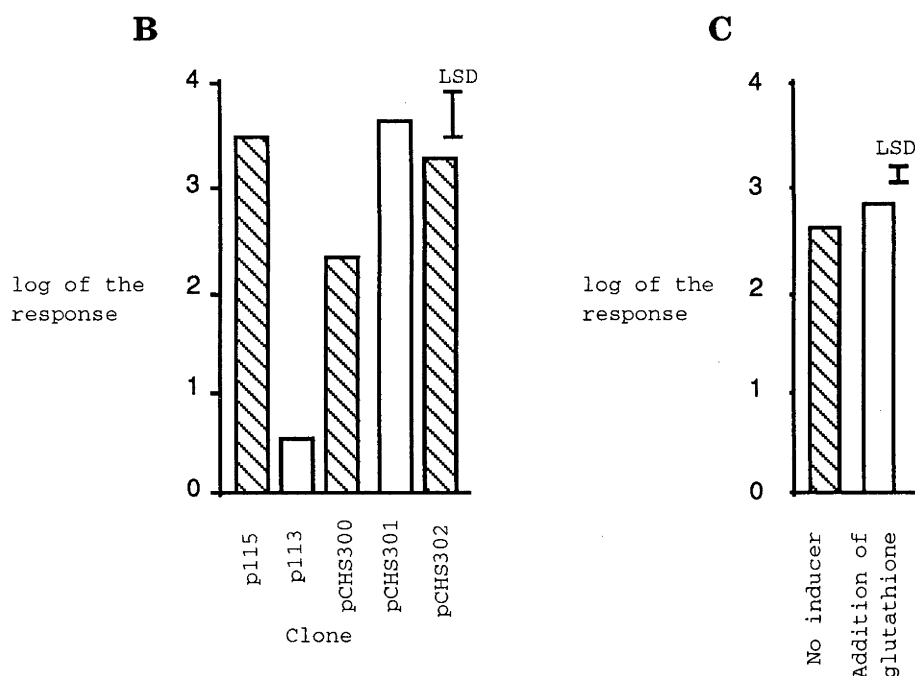
Fig. 6.9

A Results for the expression of the genetic constructs in *N. plumbaginifolia* protoplasts at 22 hours after a statistical analysis of variance was performed on the data presented in fig. 6.7. Data is expressed as the log of the response (log(rep)); it demonstrates a significant clone effect (P<0.001) and a significant inducer effect (P<0.001).

```
clone       p115   p113   p300   p301   p302   p303   p304
log (rep) 3.062  0.924  2.392  3.796  3.433  3.584  3.581
Standard error of differences: 0.2814
```

Induction caused by the addition of glutathione was uniform across the constructs:

```
Absence of inducer (log (rep)):  2.831
Presence of inducer (log (rep)): 3.104
Standard error of differences:   0.0717
```

B Graph of the log of the response against each clone for the 22 hour timepoint.
C Graph of the log of the response against the absence and presence of glutathione for the 22 hour timepoint.
In each graph LSD (least significance difference) is twice the standard error of differences.
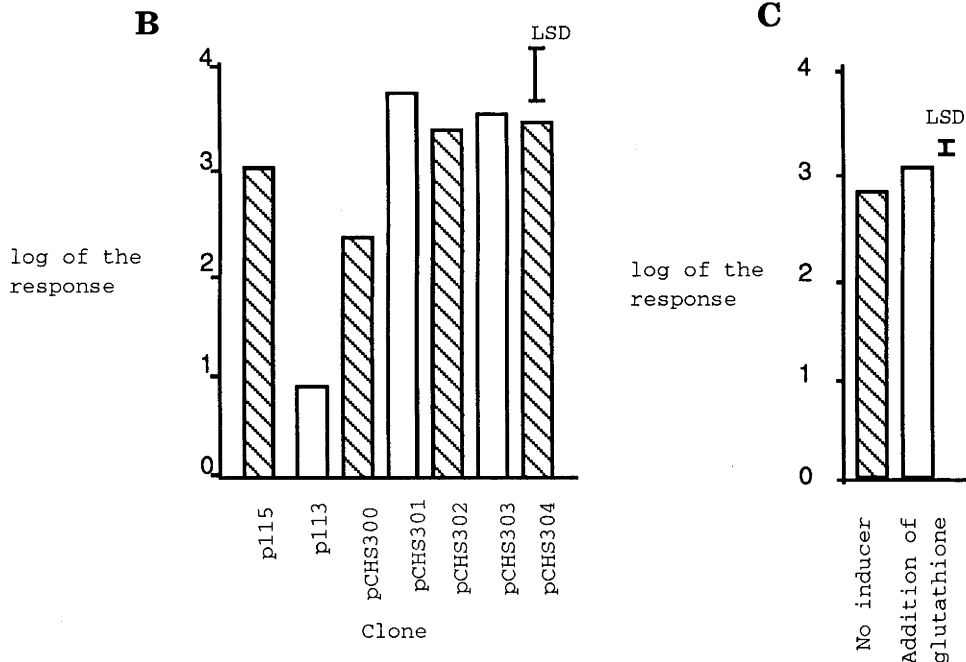
pathway gene. This phenomenon has been observed in tobacco (*Nicotiana tabacum*) protoplasts, where a bean CHS promoter was observed to be expressed (mentioned above; 6.2.3.2), and in the activity of a bean CHS promoter in transgenic plants (Schmid et al., 1990). As the clover and bean CHS promoter gene fusions are presumably activated by stress induced factors in protoplast systems, and the bean promoter can be induced by elicitor or wounding in transgenic tobacco, it is interesting to note that the tobacco phytoalexins (and very likely those of *N. plumbaginifolia*) are not flavonoids. The clover and bean CHS genes must respond to a signal system that is not recognized by the endogenous CHS genes (1.5.3; Schmid et al., 1990).

The levels of GUS expression observed in these *N. plumbaginifolia* protoplasts were always very low when compared to other systems, such as parsley and soybean (Schulze-Lefert et al., 1989a/b; Wingender et al., 1990). However, the expression of p115 was only slightly lower than that recorded for a construct containing the cauliflower mosaic virus promoter fused to the GUS gene with the nopaline synthase gene termination sequence in *N. plumbaginifolia* protoplasts (Last et al., 1991). This implies that the conditions used with the Biorad Gene Pulser were not as good at creating transient expression as those employed in the original C.S.I.R.O. protocol (Taylor and Larkin, 1988), and/or the vicilin termination sequence gives reduced expression. Nevertheless, if the addition of glutathione was going to have a significant influence upon phenylpropanoid pathway gene expression *only*, it should have been observed with the electroporation conditions utilized.

The high levels of constitutive expression of the CHS-GUS-VIC constructs may possibly due to the fact that the cell suspension culture from which the protoplasts were isolated had been maintained for many years at the C.S.I.R.O. Division of Plant Industry before being obtained

for use, and was subsequently subcultured for a number of months during the experiments outlined here. It is known that cell suspension cultures of alfalfa that have been through over 10 serial subcultures since callus have very high levels of constitutive expression of any introduced CHS promoter, with only weak levels of induction with glutathione (Choudhary et al., 1990a); these findings parallel those presented in this study. However, in protoplasts isolated from alfalfa cultures which had only been subcultured 4-8 times, the basal level of CHS expression could be reduced dramatically, whilst a ten-fold induction by glutathione was possible (Choudhary et al., 1990a). Large differences in the levels of expression, both basal and induced, between different protoplast batches, were also observed in this alfalfa system. The physiological state of two protoplast preparations is never the same, and as such it is to be expected that the levels of transient expression in each case will be different. The phase of the cell cycle which the majority of the protoplasts are in, and their metabolic activity, will be very important in defining expression of introduced constructs (Lepetit et al., 1991).

On the whole, it has to be concluded that the *Nicotiana plumbaginifolia* cell suspension/protoplast system which was employed in the experiments detailed here was not an ideal system for analysing the molecular biology of plant defence gene expression. Any system which is to be used successfully in such experiments must be able to obtain low levels of expression of any phenylpropanoid pathway gene promoter before the addition of an inducing compound. Although the addition of glutathione caused an increase in the levels of expression of the CHS-GUS-VIC constructs, as noted above, even this small increase is compromised by an identical increase in the expression of both the positive and negative controls.

One interesting point to come out of these studies was that the construct pCHS300, containing approximately 2.2kb of promoter fused to the GUS gene, produced levels of expression (regardless of the presence or absence of glutathione) much lower than those constructs containing 1.1kb or less of promoter (figs 6.8/6.9, note how the difference between the log of the response for pCHS300 and the other CHS-GUS-VIC constructs is greater than the LSD). This phenomenon has also been found to occur in transient studies involving the PAL-1 gene of *Arabidopsis thaliana*, where it has been observed that deletion of the promoter from -1816 to -832 causes an increase of 30% in GUS activity (Ohl et al., 1990). It is possible that a negative regulatory element (a "silencer" sequence) is situated between 1.1 and 2.2kb upstream of the translation start site of CHS3. In the case of the bean CHS15 gene, a negative regulatory sequence is known to occur much closer to the methionine start codon (Dron et al., 1988).

The plasmid pCHS301, containing just over 1100bp of promoter, produced the highest levels of GUS activity in those protoplasts harvested after both 16 and 22 hours. However, the statistical difference between the expression (log of the response) of this plasmid and those of pCHS302, pCHS303, and pCHS304 was not significant (not greater than the LSD). Hence, it cannot be concluded that important regulatory sequences are removed upon deletion of ~340bp from it to create the next shorter construct (pCHS302).

Chapter 7. Final Discussion

1. Introduction

This thesis has presented the isolation and extensive molecular characterization of a multigene chalcone synthase (CHS) family of *Trifolium subterraneum* L. cv. Karridale, and a less in-depth study of a similar multigene family of phenylalanine ammonia-lyase (PAL). The CHS gene family was found to consist of at least nine members, seven of which occur in two clusters of four and three members; it is possible that these two clusters constitute a single group of genes, but this was not established. Sequencing of some of these genes showed that they possessed highly conserved coding sequences, as well as having homologies between their promoters. As would be expected, the DNA and derived protein sequences bore striking resemblances to the corresponding sequences of CHS genes isolated from other plant species.

Multigene families of CHS and PAL are a characteristic of legumes. This is possibly due to their induction by a wide range of stimuli (e.g. wounding, pathogen attack, light), and the involvement of their respective enzymes in the biosynthesis of a variety of compounds. In contrast, the phenylpropanoid pathway enzyme isoflavone reductase, which is *specific* for isoflavonoid biosynthesis, appears to be encoded by only a single gene in alfalfa (Paiva et al., 1991).

2. Differential expression of the individual members of the CHS gene family

In those plants that possess PAL and CHS multigene families, recent research has established that only certain members are activated in response to different inducing conditions, their activity being also dependent upon the tissue and the developmental stage of the plant

(Ryder et al., 1987; Liang et al., 1989; Wingender et al., 1989; Koes et al., 1989a; Harker et al., 1990; Lois and Hahlbrock, 1992). For example, two members of the *Petunia hybrida* multigene family, those designated A and J, are expressed almost exclusively in the flower organs upon exposure to light (Koes et al., 1989a). Two of the pea CHS genes, CHS1 and CHS3, are expressed in both petal and root tissue, whereas the third, CHS2, is active only in roots (Harker et al., 1990). Besides the phenylpropanoid pathway genes, other plant defence genes that occur in families are apparently subject to differential activation. The enzyme 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR), which is essential to the production of phytoalexins and wound response compounds in potato, is encoded by a multigene family in that plant. Infection with the pathogen *Erwinia carotovora* only activates one isogene of the HMGR family, which is distinct from the isogene(s) responsive to wounding (Yang et al., 1991).

The synthesis of the 5 deoxyisoflavonoid phytoalexins in legumes appears to be linked to the induction of certain members of multigene CHS families. Upon exposure of cell suspension cultures of bean (cultivar Imuna) to elicitor from *Colletotrichum lindemuthianum*, the transcripts of the (bean) CHS1 gene continued to steadily rise during the period when total CHS mRNA remained fairly constant (1.5-22.5 hours), only reaching a maximum level after 24 hours. As the accumulation of the 5-deoxyisoflavonoid phytoalexin phaseollin (i.e. a compound which must be derived from 2', 4', 4-trihydroxychalcone) occurred in the period immediately after that in which the level of CHS 1 transcripts peaked, it is proposed that this gene is specifically involved in its biosynthesis (Ellis et al., 1989). Similarly, treatment of cell suspension cultures of chickpea (*Cicer arietinum* L.) with an elicitor from the fungus *Ascochyta rabiei* resulted in the accumulation of the 5-deoxyisoflavonoid derived

phytoalexins medicarpin and maackiain, but only two of the five CHS isoenzymes were induced (Daniel and Barz, 1990). These genes, which are specific for the biosynthesis of 5 deoxyisoflavonoids, can be thought of as 6'-deoxychalcone synthases (1.5.1; fig. 1.4).

Isolation and examination of the phytoalexins from the genus *Trifolium* has shown that for *T. subterraneum* L. subsp. *brachycalycinum* and *T. subterraneum* L. subsp. *yanninicum*, the two major phytoalexins are medicarpin and its biogenetic successor vestitol (fig. 7.1; Ingham 1978/1990). Both of these compounds are 5 deoxyisoflavonoids. Although the phytoalexins of *T. subterraneum* L. cv. Karridale have not been determined, it is very likely that they consist of at least medicarpin, and probably vestitol. In one study of the genus *Trifolium*, 93% of the examined species were found to have medicarpin, and this frequently occurred in conjunction with vestitol (Ingham, 1990). The synthesis of these 5 deoxyisoflavonoid phytoalexins in clover may well be connected to the activity of only one or two of the CHS genes.

Exposure of cell suspension cultures of soybean to elicitor has been shown to result in the *concomitant* induction of both the CHS *and* reductase (1.5.1) mRNAs and proteins (Welle and Grisebach, 1989). It is possible that in those plants which possess multigene CHS families, such as the legumes, the transcription of only those genes involved in 5 deoxyflavonoid formation are coordinated to that of the reductase, and as such lead to the synthesis of 2', 4', 4-trihydroxychalcone and its successor compounds. Even if this is not the case, it has been suggested that in the chickpea system the two CHS isoenzymes which are induced by elicitor could have a "preferential affinity" for the reductase (Daniel and Barz, 1990). However, the soybean reductase can coact with the CHS proteins from other plant species, such as bean and parsley, to produce 2', 4', 4-trihydroxychalcone (Welle and Grisebach, 1989). The reductase must

hence be able to recognize virtually any CHS protein, especially considering that 5-deoxy-(iso)flavonoids are not produced in parsley, and presumably neither is 2', 4', 4-trihydroxychalcone. As the amino acid sequences of the clover CHS proteins do not suggest any obvious differences, the presumptive clover reductase protein (considering the probable phytoalexins of this legume, it very likely possesses the enzyme) should be able to "interact" with all the isoforms. It should also be noted, incidently, that the synthesis of the reductase could be an essential regulatory mechanism in the formation of the clover phytoalexins and *nod*D inducers.

Other members of the subterranean clover CHS multigene family may be directly linked to the production of the flavonoid inducers of the *Rhizobium nod*D protein. Although the *Rhizobium* stimulatory compounds of *Trifolium subterraneum* have not been determined, they have been for the closely related *Trifolium repens* and also for a number of other legumes (Rolfe, 1988). In the case of *T. repens*, these substances are 7,4'-dihydroxyflavone (DHF), 7,4'-dihydroxy-3'-methoxyflavone (geraldone) and 4'-hydroxy-7-methoxyflavone, all 5 deoxy(iso)flavonoids (fig. 7.1). It is thus likely that some, if not all, of the subterranean clover phenylpropanoid compounds which stimulate the *Rhizobium nod* genes are 5 deoxyflavonoids. As mentioned in chapter 1 (1.4.3), it has been postulated that, at least in the case of the legume soybean, only certain CHS (and PAL) genes are involved in the production of these signal molecules (Estabrook and Sengupta-Gopalan, 1991).

If indeed some of the clover CHS genes are specific to the biosynthesis of the 5 deoxy(iso)flavonoid phytoalexins, while *others* have a role in the production of 5 deoxy(iso)flavonoid *nod*D inducers, then the remaining family members may be involved in the production of 5 hydroxy-(iso)flavonoids. These natural products could serve some or all of the

A

B

C
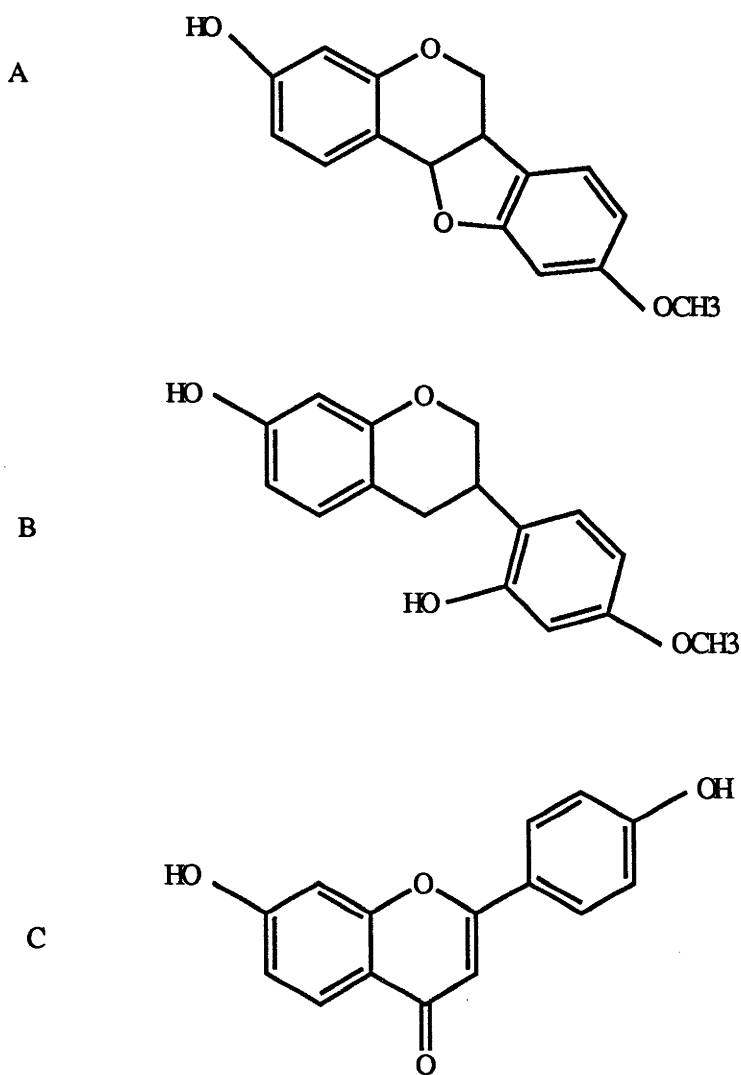
Fig. 7.1 Structures of medicarpin (A; 3-hydroxy-9-methoxypterocarpan), vestitol (B; 7,2'-dihydroxy-4'-methoxyisoflavan), and 7,4'-dihydroxyflavone (C; DHF). The first two compounds are common *Trifolium* phytoalexins. The third is a flavonoid of *Trifolium repens* which induces the expression of the *Rhizobium nod* genes.

remaining flavonoid roles outlined in 1.4, such as acting as flower pigments and being essential to the control of auxin distribution.

A great deal of research would be needed to verify any specific roles for the members of the clover CHS family. The expression of some of the genes will almost certainly be dependent upon the tissue in question and the age of the plant. It is possible that no clear cut role for any gene will be

found, and a picture will emerge of certain genes being responsive to a wide range of stimuli. In the case of the bean genes, for reasons that are not (and may never be) clear, one member, that numbered 8, is very highly expressed, and due to this has been selected for work with transgenic plants (Schmid et al., 1990; Doerner et al., 1990). Similarly, the soybean CHS1 gene has been observed to be strongly expressed under more than one induction regime (Wingender et al., 1989). Nevertheless, for each individual gene, the cis-acting elements must respond to and integrate the developmental and environmental cues (trans-acting proteins) present in the cell (see below; Liang et al., 1989).

In brief, a thorough examination of the conditions needed to induce the various genes of the clover CHS family is a necessity. Techniques such as primer extension, S1 nuclease protection analysis, and the probing of northerns with oligonucleotides designed exclusively for each of the CHS genes, would provide a wealth of useful data in this respect.

Although much research has elucidated the mechanisms which lead to increased phenylpropanoid pathway gene expression, little is known about the cessation of transcription and/or translation. However, the level of *trans*-cinnamic acid, the product of the reaction catalysed by the PAL enzyme, has been linked to a down regulation of at least some of these genes. Addition of this compound to elicited bean cells has been found to rapidly reduce the translatable activites of mRNAs coding for PAL and CHS (Bolwell et al., 1988; Mavandad et al., 1990). In elicitor induced alfalfa protoplasts, low concentrations (5×10^{-5}M) of *trans*-cinnamic acid slightly increased the expression of a bean CHS gene promoter fused to a reporter gene, whereas high concentrations (10^{-4}M) reduced expression (Loake et al., 1991). On the basis of these results, it has been suggested that *trans*-cinnamic acid may act as a modulator of gene expression in the phenylpropanoid pathway.

3. Control of CHS gene expression at the transcriptional level

One of the most exciting findings of this study is that the four clover genes CHS2, 3, 4 and 5 have a number of sequences in their respective promoters that are highly homologous to elements known to be intimately involved in the regulation of phenylpropanoid pathway genes in other plant species. Perhaps the most important of these are the sequences which resemble the two consensus sequences for regulation elements presented by Lois et al., (1989), and those which contain the core sequence of the G-box/parsley box II. The research presented in this thesis should be extended by trying to establish the importance (if any) of these clover sequences to the expression of the CHS genes. The corresponding G-box sequences in the promoter of the single CHS gene of parsley (boxes II and III) were confirmed to be essential for efficient light induction by mutating them, assaying the mutant and wild-type sequences with GUS fusions, and also by footprint analysis (Schulze-Lefert et al., 1989a/b). Similarly, the two elicitor and light responsive elements of the parsley PAL-1 promoter were identified by footprinting (Lois et al., 1989). Techniques such as these could be employed to define the roles of the homologous clover elements, and to locate any further cis-acting elements.

It is now well established that in mammalian systems promoters (and enhancers) are composed of discrete elements, sometimes called *modules*, each about 7-20bp in length, which contain recognition sequences for trans-acting factors (Dynan, 1989). The recent explosion of knowledge about the elements which control the expression of plant genes (including CHS), supports the view that plant promoters have, like their mammalian counterparts, a modular constitution. The extensive studies which have been undertaken with the parsley and bean PAL and

CHS promoters have elucidated the complex manner with which these promoter elements interact with each other (or at least the proteins which recognize these sequences do) to activate mRNA transcription. Such a situation, of course, is to be expected in each of the clover CHS promoters. The "ubiquitous" G-box may play a general role in gene expression (DeLisle and Ferl, 1990). It has been suggested that in the case of the regulation of transcription by light, the GBF may interact with protein factors which bind to other light regulatory sequences, such as that designated GT-1 (Weising and Kahl, 1991). Likewise, AT-rich sequences, which are present in so many plant gene promoters, including those of the clover CHS genes (4.3), may be recognized by a group of DNA binding proteins which combine with other protein factors to promote transcription (Forde et al., 1990). The finding that the parsley CPRFs (1.5.4) are leucine zipper proteins has raised the possibility that the dimerization of different DNA-binding factors could be important in plant gene regulation (Weisshaar et al., 1991). It is to be expected that as with other genes, the regulation of clover CHS expression will be controlled by gene *specific* signals and other signals that are of a more *general* nature (Nussinov, 1990).

Further studies into the sequences governing clover CHS gene expression should not overlook the role that enhancers may play in regulation. An enhancer, by definition, may be situated a great distance from the gene in question but yet be able to influence its transcription (Atchison, 1988). It is possible that proteins which bind to enhancers interact with those which bind promoter elements by the looping out of the intervening DNA sequences. Enhancers are sometimes found to the 3' of genes, but as of yet no sequence lying in such a position has been implicated in CHS expression. However, in the case of the *Petunia* (Mitchell) SSU genes, and the potato proteinase inhibitor II gene, it has

been discovered that sequences in the DNA immediately 3' to the coding regions influence RNA transcription (Dean et al., 1989; An et al., 1989).

A startling discovery has been made in studies involving the parsley 4CL-1 gene. Research involving the expression of the native gene, and chimeric genes consisting of the promoter fused to the GUS reporter gene, have established that *exonic* sequences are necessary for high levels of elicitor and light induced expression (Douglas et al., 1991). This mode of expression, however, is novel for phenylpropanoid pathway genes. All studies of CHS and PAL promoters linked to reporter genes have found that the elements necessary and sufficient for elicitor and light induced expression are available in these 5' flanking sequences (e.g. Ohl et al., 1989; Schmid et al., 1990; van der Meer et al., 1990; Fritze et al., 1991).

4. Control of CHS gene expression at the post-transcriptional level

For a number of mammalian genes it has been demonstrated that altering the stability of the RNA message is an important mechanism in the regulation of expression (see references within Elliott et al., 1989). Although most plant genes are controlled at the transcriptional level (Kuhlemeier et al., 1987), in some systems it is known that the stability of the message, or its ability to be read by the ribosomes, is at least partly responsible for determining gene expression (Wanner and Gruissem, 1991, and references there-in). For instance, the regulatory elements in the 3' flanking sequence of the potato gene mentioned above are thought to increase the stability of the transcript (An et al., 1989). Although the pea ferredoxin 1 (*Fed-1*) gene is "switched on" by light, it has been found that neither the promoter, nor the 3' flanking sequences, are capable on their own of causing light responsiveness of the GUS gene. As a consequence of these observations, it has been suggested that light influences the

expression of the *Fed-1* gene by increasing the stability of the mRNA (Elliott et al., 1989). Light is known to cause an increase in the transcription of the CHS genes from plants such as parsley and *Petunia hybrida* (Kuhn et al., 1984; Koes et al., 1989a). Perhaps in certain stress situations control of CHS expression in these plants is mediated by changing mRNA stability, as in the case of the *Fed-1* gene. It has been postulated that CHS mRNA is particularly unstable, being degraded shortly after its synthesis (Ingelbrecht et al., 1989). Prolonging the life of various CHS mRNA species in response to environmental stimuli would be one method of increasing expression if the message is lacking in stability.

A post-transcriptional form of regulation, very likely at the stage of translation, is believed to operate in the case of the CHS gene encoded by the *Whp* locus of maize (Franken et al., 1991). *Whp* mRNA can be detected when the maize plant contains the dominant or recessive forms of the intensifier allele *In*, whereas the presence of the protein product of this gene is *dependent* upon this allele being present in the homozygous recessive form (*in/in*). It is possible that the *In* allele encodes a protein which prevents the ribosomes from moving along the *Whp* mRNA, or alternatively controls the action of another protein which acts at the level of translation.

A study which may be instructive in considering gene expression in phenylpropanoid pathway gene families has been conducted with the SSU gene family of tomato (Wanner and Gruissem, 1991). Among the members of the tomato gene family, control of expression appears primarily to be at the level of transcription, but superimposed upon this is a level of post-transcriptional control. For instance, in etiolated seedlings the mRNA from one gene was most abundant, but another of the genes was more highly transcribed. As above, the stability of the message is

playing a significant role in gene expression. It is suggested that the 3'
untranslated regions of the various tomato SSU genes determine their
stability in different organs and at different developmental stages of the
plant. Such a situation may well be found among members of the clover
CHS (and PAL) gene families.

5. Concerted evolution of the CHS multigene family

As can be easily deduced from comparing the CHS sequences from
different legumes, the level of intraspecific gene variation is much lower
than that of interspecific variation. The sequence analysis presented in
chapter 4 showed that the subterranean clover CHS genes are
approximately 92% homologous at the nucleotide level (fig. 4.6), but only
about 80% homologous to the bean and soybean sequences (fig. 4.16). One
simple explanation for this phenomenon is that the genome of each
species of the Leguminosae contains a "recently" duplicated family of
CHS genes which have not been in existence long enough for substantial
sequence divergences to occur (Tanksley and Pichersky, 1988).

However, the (seemingly) universal nature of multigene CHS families
in the Leguminosae suggests that the lineages of at least some of these
families, if not all, are very "ancient". It is likely that an initial
duplication event, or events, occurred prior to the major divergences
between genera. In such a situation, it might be expected that the
members of a particular CHS multigene family would show much
greater variation than is observed, and that this intraspecific variation
would be similar to the level of that seen between CHS genes isolated from
different species.

The absence of high levels of sequence variation in the members of
multigene families which are not of recent origin is now recognized to be
a common phenomenon; the genes are seen as not evolving

independently, but interacting with each other and evolving as a single unit (Arnheim, 1983; Li and Graur, 1991). This process, called *concerted evolution* (also known as coincidental or horizontal evolution), can obliterate the differences that gradually build up between genes, resulting in the history of any gene family subjected to it being one of a continual series of homogenization/regeneration events (Maeda and Smithies, 1986). It should also be noted that the CHS gene families in each species of the Leguminosae could all be of ancient but independent origins, and the homogeneity of each maintained by concerted evolution.

Two genetic mechanisms, unequal crossing-over and gene conversion, are at present considered to be mainly responsible for concerted evolution (Ohta, 1983; Li and Graur, 1991). The former of these, which has also been linked to the initial formation of gene families (3.3), occurs when homologous genes (or non-coding DNA sequences) are tandemly arranged on chromosomes. Repeated rounds of unequal crossing-over can expand and contract the number of genes in a cluster and lead to one gene becoming the dominant one; this is called crossover fixation (Lewin, 1987). Gene conversion is a process whereby the formation of a heteroduplex at meiosis or mitosis can result in one gene "converting" an homologous gene, or part of the sequence of such a gene, to itself. The advantages of this latter mechanism is that it probably has a much greater chance of acting not only on tandem gene clusters, but also on dispersed repeats, and it does not lead to a change in gene number (Li and Graur, 1991). When the rate of the occurrence of unequal crossing-over and gene conversion is low, a gene family will acquire variability, whereas conversely when it is high the family will tend to remain more uniform. In time a gene may "escape" a gene family and become functionally distinct (i.e. coding for a protein catalyzing a slightly

different reaction) by accumulating such a number of point mutations that it can no longer participate in these mechanisms (Walsh, 1987).

Concerted evolution could also theoretically be driven almost entirely by selection acting at the protein level (Pichersky et al., 1986; Tanksley and Pichersky, 1988). It is undeniable that any change in the nucleotide sequence which produces a protein which cannot fold properly or perform its required enzymatic activity will be selected against. Moreover, if the protein interacts with other proteins or cellular components, its structure will by necessity be more restrained (Meagher et al., 1989).

Gene conversion has been implicated as the major mechanism acting in the concerted evolution of a number of plant multigene families. The absolutely identical sequences of the tomato 3A and 3C SSU genes, which extends from just prior to the first ATG codons and ends directly following the stop codons, together with the near identical sequence of the 3B gene, is believed to favour gene conversion (Sugita et al., 1987). In a study of the SSU genes from a number of organisms, it was found that the members of a SSU multigene family from the same species were more similar to each other than to those from other species; the movement of homologous sequences between chromosomes was interpreted as being due to gene conversion (Meagher et al., 1989). Both gene conversion and unequal crossing-over have been linked to the high sequence homologies, both coding and non-coding, of the petunia and pea CAB genes (Pichersky et al., 1985). As there is no non-coding sequence homology between the tomato CAB genes, it was suggested that these genes were last homogenized at a more ancient date than those in the above two plants (Pichersky et al., 1985). Subterranean clover CHS2, 3, and 4 could be of recent origin (whether by unequal crossing-over or gene conversion) due to the high 5' and 3' non-coding sequence homologies (4.2.3).

The amount of homogenization that the gene conversion process can produce between two sequences is dependent upon the actual length of sequence "converted", and the repair mechanism, the latter of which usually does not result in two absolutely identical sequences. Because of this, the action of gene conversion in the concerted evolution of any multigene family can be viewed from two extreme perspectives (Meagher et al., 1989). First, if the conversion does produce two identical or near identical sequences sometime in the past, the observed (i.e. as seen today) differences between these sequences might accurately represent the amount of divergence since the event itself, and even its time. Second, the divergence between two gene sequences from a multigene family could be a reflection of the differences remaining from a conversion, but *without* time for any subsequent changes. Presuming gene conversion to be at least partly responsible for the high homologies of the clover CHS genes, but without any knowledge of the extent of homogenization it has produced between any two genes, it is impossible to say which of these two alternatives is a better explanation of the differences between genes.

None of the subterranean clover CHS sequences show such high homology as seen between the three tomato SSU genes mentioned above. Hence, any obvious evidence of gene conversion, or at least as interpreted as such in the above study, is lacking in the clover CHS genes. Nevertheless, it is intriguing that CHS2, 3, and 4 have five regions of substantial homology between their promoters (fig. 4.12), and also high sequence homology in the regions immediately adjacent to their termination codons. If gene conversion is responsible for the origin of these three genes, the formation of heteroduplexes will presumably have begun outside the coding sequences. At the level of the coding sequences alone, it is of great interest that CHS2 and CHS5, situated (on the present data) in different clusters, are more homologous than CHS2 is to the

adjacent CHS3. It could be concluded from this phenomenon, and yet again assuming the action of gene conversion in the evolution of these sequences, that there is no simple relationship specifying that the closer two genes the higher the rate of gene conversion. For a family of kallikrein serine proteases from rat, which is believed to be acted upon by gene conversion, the frequency of gene conversion cannot be related to the linkage of genes (Wines et al., 1991).

On the basis of the high non-coding homology between CHS2, 3 and 4, it is tempting to classify them as a clover CHS "subfamily". Although the promoter of CHS1 does not possess any of the long promoter sequences which are shared between these three genes, its 3' flanking sequence is strikingly homologous to those of CHS2 and CHS3 (data not shown). Hence, whether CHS1 should be included in this subfamily is not clear. In one study of the degree of divergence between SSU genes from three species of Nicotianeae, which were differentiated into two subfamilies on the basis of their active/pseudogene status and sequence homologies, it was found that genes belonging to different species, but from the same subfamily, were closer in sequence than genes from the two subfamilies in any one species (Jamet et al., 1991). It would be extremely interesting to investigate the possibility that any subfamily of CHS genes within *Trifolium subterraneum* may have a counterpart in other species from this genus (such as *Trifolium repens*), and that the levels of interspecific homology between such subfamily members may be higher than that which they possess with CHS genes in different subfamilies of their respective species.

If the CHS genes of some of the genera within the Leguminosae can indeed be grouped into common subfamilies, it would provide evidence that the lineages of these genes are more ancient than the independent genera. Such a situation has been found with the SSU genes of the

Solanaceae, where there are three gene lineages (3.3; Pichersky et al., 1986; Tanksley and Pichersky, 1988).

In summary, the organization of the CHS genes in subterranean clover, as presumably in other legumes, is likely the result of numerous unequal crossovers and gene conversions. As pointed out above, this concerted evolution can thwart any attempts to establish gene lineages, meaning that the origin and history of the clover CHS genes probably will never be known. Nevertheless, this study has provided a thorough molecular characterization of the CHS genes from this plant, and any future work on their expression and evolution must use it as a basis.

REFERENCES

Akada, S., Kung, S.D., and Dube, S.K. (1990a). Nucleotide sequence of one member of the soybean chalcone synthase multi-gene family. *Nucl. Acids Res.* **18**, 3398.

Akada, S., Kung, S.D., and Dube, S.K. (1990b). Nucleotide sequence of gene 3 of the soybean chalcone synthase multigene family. *Nucl. Acids Res.* **18**, 5899.

Akada, S., Kung, S.D., and Dube, S.K. (1991). Nucleotide sequence of gene 1 of the soybean chalcone synthase multigene family. *Plant Mol. Biol.* **16**, 751-752.

Allen, O.N., and Allen, E.K. (1981). The Leguminosae, A Source Book of Characteristics, Uses, and Nodulation. Macmillan Publishers Ltd, London.

An, G., Mitra, A., Choi, H.K., Costra, M.A., An, K., Thornburg, R.W., and Ryan, C.A. (1989). Functional analysis of the 3' control region of the potato wound-inducible proteinase inhibitor II gene. *Plant Cell* **1**, 115-122.

Arnheim, N. (1983). Concerted evolution of multigene families. In: Evolution of Genes and Proteins. M. Nei and R. K. Koehn (eds), Sinauer Associates Inc., Sunderland, Mass., U.S.A., pp38-61.

Atchison, M.L. (1988). Enhancers: mechanisms of action and cell specificity. *Ann. Rev. Cell Biol.* **4**, 127-153.

Ayabe, S., Udagawa, A., and Furuya, T. (1988a). NAD(P)H-dependent 6'-deoxychalcone synthase activity in *Glycyrrhiza echinata* cells induced by yeast extract. *Arch. Biochem. Biophys.* **261**, 458-462.

Ayabe, S., Udagawa, A., and Furuya, T. (1988b). Stimulation of chalcone synthase activity by yeast extract in cultured *Glycyrrhiza echinata* cells and 5-deoxyflavanone formation by isolated protoplasts. *Plant Cell Reports* **7**, 35-38.

Bates, G.W., Piastuch, W., Riggs, C.D., and Rabussay, D. (1988). Electroporation for DNA delivery to plant protoplasts. *Plant Cell, Tissue and Organ Culture* 12, 213-218.

Batschauer, A., Ehmann, B., and Schafer, E. (1991). Cloning and characterization of a chalcone synthase gene from mustard and its light-dependent expression. *Plant Mol. Biol.* 16, 175-185.

Becker-Andre, M., Schulze-Lefert, P., and Hahlbrock, K. (1991). Structural comparison, modes of expression, and putative *cis*-acting elements of the two 4-coumarate:CoA ligase genes in potato. *J. Biol. Chem.* 266, 8551-8559.

Bell, J.N., Ryder, T.B., Wingate, V.P.M., Bailey, J.A., and Lamb, C.J. (1986). Differential accumulation of plant defense gene transcripts in a compatible and an incompatible plant-pathogen interaction. *Mol. Cell Biol.* 6, 1615-1623.

Binns, A.N., Chen, R.H., Wood, H.N., and Lynn, D.G. (1987). Cell division promoting activity of naturally occurring dehydrodiconiferyl glucosides: Do cell wall components control cell division? *Proc. Natl. Acad. Sci. USA* 84, 980-984.

Birnboim, H.C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7, 1513-1523.

Block, A., Dangl, J.L., Hahlbrock, K., and Schulze-Lefert, P. (1990). Functional borders, genetic fine structure, and distance requirements of cis elements mediating light responsiveness of the parsley chalcone synthase promoter. *Proc. Natl. Acad. Sci. USA* 87, 5387-5391.

Bolwell, G.P., Mavandad, M., Millar, D.J., Edwards, K.J., Schuch, W., and Dixon, R. (1988). Inhibition of mRNA levels and activities by *trans*-cinnamic acid in elicitor-induced bean cells. *Phytochemistry* 27, 2109-2117.

Bolwell, G.P., Coulson, V., Rodgers, M.W., Murphy, D.L., and Jones, D. (1991). Modulation of the elicitation reponse in cultured French bean cells and its implication for the mechanism of signal transduction. *Phytochemistry* **30**, 397-405.

Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A., and Sauer, R.T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306-1310.

Brettell, R.I.S., and Pryor, A.J. (1986). Molecular approaches to plant and pathogen genes. In: <u>A Genetic Approach to Plant Biochemistry</u>. A D. Blonstein and P.J. King (eds), Springer-Verlag, Vienna and New York, pp 233-246.

Broglie, K.E., Biddle, P., Cressman, R., and Broglie, R. (1989). Functional analysis of DNA sequences responsible for ethylene regulation of a bean chitinase gene in transgenic tobacco. *Plant Cell* **1**, 599-607.

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

Bustos, M.M., Begum, D., Kalkan, F.A., Battraw, M.J., and Hall, T.C. (1991). Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. *EMBO J.* **10**, 1469-1479.

Carr, C.S., and Sharp, P.A. (1990). A helix-loop-helix protein related to the immunoglobin E box-binding proteins. *Mol. Cell. Biol.* **10**, 4384-4388.

Chappell, J., and Hahlbrock, K. (1984). Transcription of plant defence genes in response to UV light or fungal elicitor. *Nature* **311**, 76-78.

Chou, P.Y., and Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology* **47**, 45-148.

Choudhary, A.D., Kessmann, H., Lamb, C.J., and Dixon, R.A. (1990a). Stress responses in alfalfa (*Medicago sativa* L.) IV Expression of defense gene constructs in electroporated suspension cell protoplasts. *Plant Cell Reports* **9**, 42-46.

Choudhary, A.D., Lamb, C.J., and Dixon, R.A. (1990b). Stress responses in alfalfa (*Medicago sativa* L.) VI Differential responsiveness of chalcone synthase induction to fungal elicitor or glutathione in electroporated protoplasts. *Plant Physiol.* **94**, 1802-1807.

Clewell, D.B., and Helinski, D.R., (1969). Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. *Proc. Natl. Acad. Sci.* USA **62**, 1159-1166.

Collinge, D.B., and Slusarenko, A.J. (1987). Plant gene expression in response to pathogens. *Plant Mol. Biol.* **9**, 389-410.

Compton, T. (1990). Degenerate primers for DNA amplification. In: PCR Protocols. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White (eds), Academic Press, San Diego, pp39-45.

Cramer, C.L., Ryder, T.B., Bell, J.N., and Lamb, C.J. (1985). Rapid switching of plant gene expression induced by fungal elicitor. *Science* **227**, 1240-1242.

Cramer, C.L., Edwards, K., Dron, M., Liang, X., Dildine, S.L., Bolwell, G.P., Dixon, R.A., Lamb, C.J., and Schuch, W. (1989). Phenylalanine ammonia-lyase gene organization and structure. *Plant Mol. Biol.* **12**, 367-383.

Dalkin, K., Edwards, R., Edington, B., and Dixon, R.A. (1990). Stress responses in alfalfa (*Medicago sativa* L.). I Induction of phenylpropanoid biosynthesis and hydrolytic enzymes in elicitor-treated cell suspension cultures. *Plant Physiol.* **92**, 440-446.

Dangl, J.L., Hauffe, K.D, Lipphardt, S., Hahlbrock, K., and Scheel, D. (1987). Parsley protoplasts retain differential responsiveness to u.v. light and fungal elicitor. *EMBO J.* **6**, 2551-2556.

Daniel, S., and Barz, W. (1990). Elicitor-induced metabolic changes in cell cultures of chickpea (*Cicer arietinum* L.) cultivars resistant and susceptible to *Ascochyta rabiei*. *Planta* **182**, 279-286.

Datta, N., and Cashmore, A.R. (1989). Binding of a pea nuclear protein to promoters of certain photoregulated genes is modulated by phosphorylation. *Plant Cell* **1**, 1069-1077.

Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P., and Bedbrook, J. (1985). Linkage and homology analysis divides the eight genes for the small subunit of petunia ribulose 1,5-bisphophate carboxylase into three gene families. *Proc. Natl. Acad. Sci. USA* **82**, 4964-4968.

Dean, C., Favreau, M., Bond-Nutter, D., Bedbrook, J., and Dunsmuir, P. (1989). Sequences downstream of translation start regulate quantitative expression of two Petunia *rbc*S genes. *Plant Cell* **1**, 201-208.

DeLisle, A.J., and Ferl, R.J. (1990). Characterization of the *Arabidopsis Adh* G-box binding factor. *Plant Cell* **2**, 547-557.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

De Vries, S., Hoge, H., and Bisseling, T. (1988). Isolation of total and polysomal RNA from plant tissues. In: <u>Plant Molecular Biology Manual</u>. S. B. Gelvin and R.A. Schilperoort (eds), Kluwer Academic Publishers, Dordrecht, Section B6: 1-13.

Dewick, P.M., Steele, M.J., Dixon, R.A., and Whitehead, I.M. (1982). Biosynthesis of isoflavonoid phytoalexins: incorporation of sodium [1,2-$^{13}C_2$] acetate into phaseollin and kievitone. *Z. Naturforsch.* **37c**, 363-368.

Dietrich, A., Mayer, J.E., and Hahlbrock, K. (1990). Fungal elicitor triggers rapid, transient, and specific protein phosphorylation in parsley cell suspension cultures. *J. Biol. Chem.* **265**, 6360-6368.

Dimroth, P., Walter, H., and Lynen, F. (1970). Biosynthese von 6-methylsalicylsaure. *Eur. J. Biochem.* **13**, 98-110.

Dixon, R.A. (1986). The phytoalexin response: elicitation, signalling and control of host gene expression. *Biol. Rev.* **61**, 239-291.

Dixon, R.A., Bailey, J.A., Bell, J.N., Bolwell, G.P., Cramer, C.L., Edwards, K., Hamdan, M.A.M.S., Lamb, C.J., Robbins, M.P., Ryder, T.B., and Schuch, W. (1986). Rapid changes in gene expression in response to microbial elicitation. *Phil. Trans. R. Soc. Lond. B* **314**, 411-426.

Doerner, P.W., Stermer, B., Schmid, J., Dixon, R.A., and Lamb, C.J. (1990). Plant defense gene promoter-reporter gene fusions in transgenic plants: tools for identification of novel inducers. *Bio/technology* **8**, 845-848.

Dooner, H.K. (1983). Coordinate genetic regulation of flavonoid biosynthetic enzymes in maize. *Mol. Gen. Genet.* **189,** 136-141.

Douglas, C., Hoffmann, H., Schulz, W., and Hahlbrock, K. (1987). Structure and elicitor or u.v.-light-stimulated expression of two 4-coumarate: CoA ligase genes in parsley. *EMBO J.* **6**, 1189-1195.

Douglas, C.J., Hauffe, K.D., Ites-Morales, M.-E., Ellard, M., Paszkowski, U., Hahlbrock, K., and 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.

Dron, M., Clouse, S.D., Dixon, R.A., Lawton, M.A., and Lamb, C.J. (1988). Glutathione and fungal elicitor regulation of a plant defense gene promoter in electroporated protoplasts. *Proc. Natl. Acad. Sci. USA* **85**, 6738-6742.

Dunsmuir, P., Smith, S.M., and Bedbrook, J. (1983). The major chlorophyll a/b binding protein of petunia is composed of several polypeptides encoded by a number of distinct nuclear genes. *J. Mol. Appl. Gen.* **2**, 285-300.

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

Edwards, K., Cramer, C.L., Bolwell, G.P., Dixon, R.A., Schuch, W., and Lamb, C.J. (1985). Rapid transient induction of phenylalanine ammonia-lyase mRNA in elicitor-treated bean cells. *Proc. Natl. Acad. Sci. USA* **82**, 6731-6735.

Edwards, R., Blount, J.W., and Dixon, R.A. (1991). Glutathione and elicitation of the phytoalexin response in legume cell cultures. *Planta* **184**, 403-409.

Ehmann, B., and Schafer, E. (1988). Nucleotide sequences encoding two different chalcone synthases expressed in cotyledons of SAN 9789 treated mustard (*Sinapis alba* L.). *Plant Mol. Biol.* **11**, 869-870.

Ehmann, B., Ocker, B., and Schafer, E. (1991). Development- and light-dependent regulation of the expression of two different chalcone synthase transcripts in mustard cotyledons. *Planta* **183**, 416-422.

Ellingboe, A.H. (1981). Changing concepts in host-pathogen genetics. *Ann. Rev. Phytopathol.* **19**, 125-143.

Elliott, R.C., Dickey, L.F., White, M.J., and Thompson, W.F. (1989). *cis*-Acting elements for light regulation of pea ferredoxin 1 gene expression are located within transcribed sequences. *Plant Cell* **1**, 691-698.

Ellis, J.S., Jennings, A.C., Edwards, L.A., Mavandad, M., Lamb, C.J., and Dixon, R.A. (1989). Defense gene expression in elicitor-treated cell suspension cultures of french bean cv. Imuna. *Plant Cell Reports* **8**, 504-507.

Epping, B., Kittel, M., Ruhnau, B., and Hemleben, V. (1990). Isolation and sequence analysis of a chalcone synthase cDNA of *Matthiola incana* R. Br. (Brassicaceae). *Plant Mol. Biol.* **14**, 1061-1063.

Estabrook, E.M., and Sengupta-Gopalan, C. (1991). Differential expression of phenylalanine ammonia-lyase and chalcone synthase during soybean nodule development. *Plant Cell* **3**, 299-308.

Feinbaum, R.L., and Ausubel, F.M. (1988). Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. *Mol. Cell. Biol.* **8**, 1985-1992.

Feinbaum, R.L., Storz, G., and Ausubel, F.M. (1991). High intensity and blue light regulated expression of chimeric chalcone synthase genes in transgenic *Arabidopsis thaliana* plants. *Mol. Gen. Genet.* **226**, 449-456.

Fickett, J.W. (1982). Recognition of protein coding regions in DNA sequences. *Nucl. Acids Res.* **10**, 5303-5318.

Flor, H.H. (1946). Genetics of pathogenicity in *Melamspora lini*. *J. Agric. Res.* **73**, 335-337.

Forde, B.G., Freeman, J., Oliver, J.E., and Pineda, M. (1990). Nuclear factors interact with conserved A/T-rich elements upstream of a nodule-enhanced glutamine synthetase gene from French Bean. *Plant Cell* **2**, 925-939.

Franken, P., Niesbach-Klosgen, U., Weydemann, U., Marechal-Drouard, L., Saedler, H., and Wienand, U. (1991). The duplicated chalcone synthase genes *C2* and *Whp* (white pollen) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by anthocyanin intensifying gene *in*. *EMBO J.* **10**, 2605-2612.

Fritze, K., Staiger, D., Czaja, I., Walden, R., Schell, R., and Wing, D. (1991). Developmental and UV light regulation of the snapdragon chalcone synthase promoter. *Plant Cell* **3**, 893-905.

Fujiwara, T., Naito, S., Chino, M., and Nagata, T. (1991). Electroporated protoplasts express seed specific gene promoters. *Plant Cell Reports* **9**, 602-606.

Gabriel, D.W., and Rolfe, B.G. (1990). Working models of specific recognition in plant-microbe interactions. *Annu. Rev. Phytopathol.* **28**, 365-391.

Giuliano, G., Pichersky, E., Malik, V.S., Timko, M.P., Scolnik, P.A., and Cashmore, A.R. (1988). An evolutionary conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. Natl. Acad. Sci. USA* **85**, 7089-7093.

Gladstones, J.S., and Collins, W.J. (1983). Subterranean clover as a naturalized plant in Australia. *The Journal of the Australian Institute of Agricultural Science* **49**, 191-202.

Gleitz, J., and Seitz, H.U. (1989). Induction of chalcone synthase in cell suspension cultures of carrot (*Daucus carota* L. ssp. *sativus*) by ultraviolet light: evidence for two different forms of chalcone synthase. *Planta* **179**, 323-330.

Gleitz, J., Schnitzler, J.-P., Steimle, D., and Seitz, H.U. (1991). Metabolic changes in carrot cells in response to simultaneous treatment with ultraviolet light and a fungal elicitor. *Planta* **184**, 362-367.

Goodall, G.J., and Filipowicz, W. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58**, 473-483.

Goodall, G.J., and Filipowicz, W. (1990). The minimum functional length of pre-mRNA introns in monocots and dicots. *Plant Mol. Biol.* **14**, 727-733.

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

Gough, J.A., and Murray, N.E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.* **166**, 1-19.

Gowri, G., Paiva, N.L., and Dixon, R.A. (1991). Stress responses in alfalfa (*Medicargo sativa* L.) 12. Sequence analysis of phenylalanine ammonia-lyase (PAL) cDNA clones and appearance of PAL transcripts in elicitor-treated cell cultures and developing plants. *Plant Mol. Biol.* **17**, 415-429.

Grab, D., Loyal, R., and Ebel, J. (1985). Elicitor-induced phytoalexin synthesis in soybean cells: changes in the activity of chalcone synthase mRNA and the total population of translatable mRNA. *Arch. Biochem. Biophys.* **243**, 523-529.

Graham, T.L., and Graham, M.Y. (1991). Cellular coordination of molecular responses in plant defense. *Molecular Plant-Microbe Interactions* **4**, 415-422.

Grantham, R., Perrin, P., and Mouchiroud, D. (1986). Patterns in codon usage of different kinds of species. *Oxford surveys in evolutionary biology*. **3**, R. Dawkins and M. Ridley (eds), 48-81.

Grisebach, H. (1981). Lignins. In: The Biochemistry of Plants. Vol 7. E.E. Conn (ed), Academic Press, pp 457-478.

Guiltinan, M.J., Marcotte, W.R., and Quatrano, R.S. (1990). A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* **250**, 267-271.

Hahlbrock, K. (1981). Flavonoids. In: The Biochemistry of Plants Vol 7. E.E. Conn (ed), Academic Press, pp 425-456.

Hahlbrock, K., Kreuzaler, F., Ragg, H., Fautz, E., and Kuhn, D.N. (1982). Regulation of flavonoid and phytoalexin accumulation through mRNA and enzyme induction in cultured plant cells. In: Biochemistry of Differentiation and Morphogenesis. L. Jaenicke (ed), Springer-Verlag, Berlin, pp 34-43.

Hahlbrock, K., Cretin, C., Cuypers, B., Fritzemeier, K.-H., Hauffe, K.-D., Jahnen, W., Kombrink, E., Rohwer, F., Scheel, D., Schmelzer, E., Schroder, M., and Taylor, J. (1987). Tissue specificity and dynamics of

disease responses in plants. In: <u>Plant Molecular Biology</u>. D. von Wettstein and N.-H. Chua (eds), Plenum Press, pp 399-406.

Hammerschmidt, R., Lamport, D.T.A., and Muldoon, E.P. (1984). Cell wall hydroxyproline enhancement and lignin deposition as an early event in the resistance of cucumber to *Cladosporium cucumerinum*. *Physiol. Plant Pathol.* **24**, 43-47.

Harborne, J.B. (1980). Plant phenolics. In: <u>Secondary Plant Products</u>. E.A. Bell and B.V. Charlwood (eds), Springer Verlag, Berlin, pp 329-402.

Harker, C.L., Ellis, T.H.N., and Coen, E.S. (1990). Identification and genetic regulation of the chalcone synthase multigene family in pea. *Plant Cell* **2**, 185-194.

Harrison, M.J., Lawton, M.A., Lamb, C.J., and Dixon, R.A. (1991). Characterization of a nuclear protein that binds to three elements within the silencer region of a bean chalcone synthase gene promoter. *Proc. Natl. Acad. Sci. USA* **88**, 2515-2519.

Hauffe, K.D., Paszkowski, U., Schulze-Lefert, P., Hahlbrock, K., Dangl, J.L., and Douglas, C.J. (1991). A parsley 4CL-1 promoter fragment specifies complex expression patterns in transgenic tobacco. *Plant Cell* **3**, 435-443.

Hedrick, S.A., Bell, J.N., Boller, T., and Lamb, C.J. (1988). Chitinase cDNA cloning and mRNA induction by fungal elicitor, wounding, and infection. *Plant Physiol.* **86**, 182-186.

Heinrich, P., Nixon, B.T., and Voytas, D. (1987). Construction of nested deletions for DNA sequencing. In: <u>Current Protocols in Molecular Biology</u>, vol. 1. F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl (eds), John Wiley & Sons, unit 7.3.

Heller, W., and Hahlbrock, K. (1980). Highly purified "flavanone synthase" from parsley catalyzes the formation of naringenin chalcone. *Arch. Biochem. Biophys.* **200**, 617-619.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-359.

Herrmann, A., Schulz, W., and Hahlbrock, K. (1988). Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. *Mol. Gen. Genet.* **212**, 93-98.

Hess, K.M., Dudley, M.W., Lynn, D.G., Joerger, R.D., and Binns, A.N. (1991). Mechanism of phenolic activation of *Agrobacterium* virulence genes: development of a specific inhibitor of bacterial sensor/response systems. *Proc. Natl. Acad. Sci. USA* **88**, 7854-7858.

Higuchi, R. (1989). Using PCR to engineer DNA. In: PCR Technology, H.A. Erlich (ed), Stockton Press, New York, pp 61-70.

Hrazdina, G., Kreuzaler, F., Hahlbrock, K., and Grisebach, H. (1976). Substrate specificity of flavanone synthase from cell suspension cultures of parsley and structure of release products *in vitro*. *Arch. Biochem. Biophys.* **175**, 392-399.

Huang, P.-L., Hahlbrock, K., and Somssich, I.E. (1988). Detection of a single-copy gene on plant chromosomes by in situ hybridization. *Mol. Gen. Genet.* **211**, 143-147.

Huang, P.-L., Hahlbrock, K., and Somssich, I.E. (1989). Chromosomal localization of parsley 4-coumarate: CoA ligase genes by in situ hybridization with a complementary DNA. *Plant Cell Reports* **8**, 59-62.

Ingelbrecht, I.L.W., Herman, L.M.F., Dekeyser, R.A., Van Montagu, M.C., and Depicker, A.G. (1989). Different 3' end regions strongly influence the level of gene expresion in plant cells. *Plant Cell* **1**, 671-680.

Ingham, J.L. (1978). Isoflavonoid and stilbene phytoalexins of the genus *Trifolium*. *Biochem. Syst. Ecol.* **6**, 217-223.

Ingham, J.L. (1982). Phytoalexins from the Leguminosae. In: Phytoalexins. J.A. Bailey and J.W. Mansfield (eds), Blackie, Glasgow, pp 21-80.

Ingham, J.L. (1990). A further investigation of phytoalexin formation in the genus *Trifolium. Z. Naturforsch* **45c**, 829-834.

Jacobs, M., and Rubery, P.H. (1988). Naturally occurring auxin transport regulators. *Science* **241**, 346-349.

Jacobsen, K., Laursen, N.B., Jensen, E.O., Marcker, A., Poulsen, C., and Marcker, K.A. (1990). HMG 1-like proteins from leaf and nodule nuclei interact with different AT motifs in soybean nodulin promoters. *Plant Cell* **2**, 85-94.

Jamet, E., Parmentier, Y., Durr, A., and Fleck, J. (1991). Genes encoding the small subunit of Rubisco belong to two highly conserved subfamilies in Nicotianeae. *J. Mol. Evol.* **33**, 226-236.

Jefferson, R.A., Burgess, S.M., and Hirsh, D. (1986). β-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. USA* **83**, 8447-8451.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: β-Glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901-3907.

Jefferson, R.A. (1987). Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Molecular Biology Reporter* **5**, 387-405.

Jeffreys, A.J., and Harris, S. (1982). Processes of gene duplication. *Nature* **296**, 9-10.

Jensen, E.O., Marcker, K.A., Schell, J., and de Bruijn, F.J. (1988). Interaction of a nodule specific, *trans*-acting factor with distinct DNA elements in the soybean leghaemoglobin lbc3 5' upstream region. *EMBO J.* **7**, 1265-1271.

Jofuku, K.D., Okamuro, J.K., and Goldberg, R.B. (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. *Nature* **328**, 734-737.

Jorrin, J., and Dixon, R.A. (1990). Stress responses in alfalfa (*Medicago sativa* L.). 2 Purification, characterization, and induction of phenylalanine ammonia-lyase isoforms from elicitor-treated cell suspension cultures. *Plant Physiol.* **92**, 447-455.

Joshi, C.P. (1987a). An inspection of the domain between 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**, 9627-9640.

Kaulen, H., Schell, J., and Kreuzaler, F. (1986). Light-induced expression of the chimeric chalcone synthase-NPTII gene in tobacco cells. *EMBO J.* **5**, 1-8.

Kauppinen, S., Siggaard-Andersen, M., and von Wettstein-Knowles, P. (1988). β-ketoacyl-ACP synthase 1 of *Escherichia coli:* nucleotide sequence of the *fabB* gene and identification of the cerulenin binding residue. *Carlsberg Res. Commun.* **53**, 357-370.

Knogge, W., and Weissenbock, G. (1986). Tissue-distribution of secondary phenolic biosynthesis in developing primary leaves of *Avena sativa* L. *Planta* **167**, 196-205.

Koes, R.E., Spelt, C.E., Reif, H.J., van den Elzen, P.J.M., Veltkamp, E., and Mol, J.N.M. (1986). Floral tissue of *Petunia hybrida* (V30) expresses only one member of the chalcone synthase multigene family. *Nucl. Acids Res.* **14**, 5229-5239.

Koes, R.E., Spelt, C.E., Mol, J.N.M., and Gerats, A.G.M. (1987). The chalcone synthase multigene family of *Petunia hybrida* (V30): sequence

homology, chromosomal localization and evolutionary aspects. *Plant Mol. Biol.* **10**, 375-385.

Koes, R.E., Spelt, C.E., and Mol, J.N.M. (1989a). The chalcone synthase multigene family of *Petunia hybrida* (V30): differential, light regulated expression during flower development and UV light induction. *Plant Mol. Biol.* **12**, 213-225.

Koes, R.E., Spelt, C.E., van den Elzen, P.J.M., and Mol, J.N.M. (1989b). Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. *Gene* **81**, 245-257.

Koes, R. E., van Blokland, R., Quattrocchio, F., van Tunen, A.J., and Mol, J.N.M. (1990). Chalcone synthase promoters in *Petunia* are active in pigmented and unpigmented cell types.| *Plant Cell* **2**, 379-392.

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

Krainer, A.R., and Maniatis, T. (1988). RNA splicing. In: Transcription and Splicing. B.D. Hames and D.M. Glover (eds). IRL Press Limited, Oxford and Washington D.C., pp 131-206.

Kreuzaler, F., and Hahlbrock, K. (1975). Enzymic synthesis of an aromatic ring from acetate units. *Eur. J. Biochem.* **56**, 205-213.

Kreuzaler, F., Ragg, H., Fautz, E., Kuhn, D.N., and Hahlbrock, K. (1983). UV-induction of chalcone synthase mRNA in cell suspension cultures of *Petroselinum hortense*. *Proc. Natl. Acad. Sci. USA* **80**, 2591-2593.

Kuhlemeier, C., Green, P.J. and Chua, N.-H. (1987). Regulation of gene expression in higher plants. *Ann. Rev. Plant. Physiol.* **38**, 221-257.

Kuhlemeier, C., Fluhr, R., and Chua, N.-H. (1988). Upstream sequences determine the difference in transcript abundance of pea *rbc*S genes. *Mol. Gen. Genet.* **212**, 405-411.

Kuhn, D.N., Chappell, J., Boudet, A., and Hahlbrock, K. (1984). Induction of phenylalanine ammonia-lyase and 4-coumarate:CoA ligase mRNAs in cultured plant cells by UV light or fungal elicitor. *Proc. Natl. Acad. Sci. USA* **81**, 1102-1106.

Lam, E., Kano-Murakami, Y., Gilmartin, P., Niner, B., and Chua, N-H. (1990). A metal-dependent DNA-binding protein interacts with a constitutive element of a light-responsive promoter.| *Plant Cell* **2**, 857-866.

Lamb, C.J., Lawton, M.A., Dron, M., and Dixon, R.A. (1989). Signals and transduction mechanisms for activation of plant defenses against microbial attack. *Cell* **56,** 215-224.

Lanz, T., Schroder, G., and Schroder, J. (1990). Differential regulation of genes for resveratrol synthase in cell cultures of *Arachis hypogaea* L. *Planta* **181**, 169-175.

Last, D.I., Brettell, R.I.S., Chamberlain, D.A., Chaudhury, A.M., Larkin, P.J., Marsh, E.L., Peacock, W.J., and Dennis, E.S. (1991). pEmu: an improved promoter for gene expression in cereal cells. *Theor. Appl. Genet.* **81**, 581-588.

Lawton, M.A., Dixon, R.A., Hahlbrock, K., and Lamb, C. (1983). Rapid induction of the synthesis of phenylalanine ammonia-lyase and of chalcone synthase in elicitor-treated plant cells. *Eur. J. Biochem.* **129,** 593-601.

Lawton, M.A., and Lamb, C.J. (1987). Transcriptional activation of plant defense genes by fungal elicitor, wounding, and infection. *Mol. Cell. Biol.* **7,** 335-341.

Lawton, M.A., Clouse, S.D., and Lamb, C.J. (1990). Glutathione-elicited changes in chromatin structure within the promoter of the defense gene chalcone synthase. *Plant Cell Reports* **8**, 561-564.

Lawton, M.A., Dean, S.M., Dron, M., Kooter, J.M., Kragh, K.M., Harrison, M.J., Yu, L., Tanguay, L., Dixon, R.A., and Lamb, C.J. (1991). Silencer region of a chalcone synthase promoter contains multiple binding sites for a factor, SBF-1, closely related to GT-1. *Plant Mol. Biol.* **16**, 235-249.

Lepetit, M., Ehling, M., Gigot, C., and Hahne, G. (1991). An internal standard improves the reliability of transient expression studies in plant protoplasts. *Plant Cell Reports* **10**, 401-405.

Leutwiler, L.S., Meyerowitz, E.M., and Tobin, E.M. (1986). Structure and expression of three light-harvesting chlorophyll a/b binding protein genes in *Arabidopsis thaliana*. *Nucl. Acids Res.* **14**, 4051-4064.

Lewin, B. (1987). Genes. 3rd edition. John Wiley and Sons, New York.

Li, W.-H., and Graur, D. (1991). Fundamentals of Molecular Evolution. Sinauer Associates Inc., Sunderland, Mass., U.S.A.

Liang, X., Dron, M., Cramer, C.L., Dixon, R.A., and Lamb, C.J. (1989). Differential regulation of phenylalanine ammonia-lyase genes during plant development and by environmental cues. *J. Biol. Chem.* **264**, 14486-14492.

Light, R.J. and Hahlbrock, K. (1980). Randomization of the flavonoid A ring during biosynthesis of kaempferol from $[1,2-^{13}C_2]$ acetate in cell suspension cultures of parsley. *Z. Naturforsch* **35c**, 717-721.

Lipphardt, S., Brettschneider, R., Kreuzaler, F., Schell, J., and Dangl, J.L. (1988). UV-inducible transient expression in parsley protoplasts identifies regulatory *cis*-elements of a chimeric *Antirrhinum majus* chalcone synthase gene. *EMBO J.* **7**, 4027-4033.

Loake, G.J., Choudhary, A.D., Harrison, M.J., Mavandad, M., Lamb, C.J., and Dixon, R.A. (1991). Phenylpropanoid pathway intermediates regulate transient expression of a chalcone synthase gene promoter. *Plant Cell* **3**, 829-840.

Lois, R., Dietrich, A., Hahlbrock, K., and Schulz, W. (1989). A phenylalanine ammonia-lyase gene from parsley: structure, regulation, and identification of elicitor and light responsive *cis*-acting elements. *EMBO J.* **8**, 1641-1648.

Lois, R., and Hahlbrock, K. (1992). Differential wound activation of members of the phenylalanine ammonia-lyase and 4-coumarate: CoA ligase gene families in various organs of parsley plants. *Z. Naturforsch* **47c**, 90-94.

Long, S.R. (1989). *Rhizobium*-legume nodulation: life together in the underground. *Cell* **56**, 203-214.

Lorz, H., Larkin, P.J., Thomson, J., and Scowcroft, W.R. (1983). Improved protoplast culture and agarose media. *Plant Cell Tissue Organ Culture* **2**, 217-226.

Maeda, N., and Smithies, O. (1986). The evolution of multigene families: human haptoglobin genes. *Ann. Rev. Genet.* **20**, 81-108.

Mansfield, J.W. (1982). The role of phytoalexins in disease resistance. In: Phytoalexins. J.A. Bailey and J.W. Mansfield (eds), Blackie, Glasgow, pp 253-288.

Mason, P.J., and Williams, J.G. (1985). Hybridisation in the analysis of recombinant DNA. In: Nucleic Acid Hybridization - A Practical Approach. B.D. James and S.J. Higgins (eds), IRL Press, pp113-137.

Mauch, F., and Staehelin, L. A. (1989). Functional implications of the subcellular localization of ethylene-induced chitinase and β-1,3-glucanase in bean leaves. *Plant Cell* **1**, 447-457.

Mavandad, M., Edwards, R., Liang, X., Lamb, C.J., and Dixon, R.A. (1990). Effects of *trans*-cinnamic acid on expression of the bean phenylalanine ammonia-lyase gene family. *Plant Physiol.* **94**, 671-680.

McKendree, W.L., Paul, A.-L., DeLisle, A.J., and Ferl, R.J. (1990). In vivo and in vitro characterization of protein interactions with the dyad G-box of the *Arabidopsis Adh* gene.\   *Plant Cell* **2**, 207-214.

McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. (1985). The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucl. Acids Res.* **13**, 1347-1368.

Meagher, R.B., Berry-Lowe, S., Rice, K. (1989). Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. *Genetics* **123**, 845-863.

Mieth, H., Speth, V., and Ebel, J. (1986). Phytoalexin production by isolated soybean protoplasts. *Z. Naturforsch* **41c**, 193-201.

Miller, H. (1987). Practical aspects of preparing phage and plasmid DNA: growth, maintenance and storage of bacteria and bacteriophage. In: <u>Methods in Enzymology</u>, vol. **152**, S.L. Berger and A.R. Kimmel (eds), Academic Press, Orlando (Florida), pp 145-170.

Miller, J.H. (1972). <u>Experiments in Molecular Genetics</u>. Cold Spring Harbor, New York.

Minami, E., Ozeki, Y., Matsuoka, M., Koizuka, N., and Tanaka, Y. (1989). Structure and some characterization of the gene for phenylalanine ammonia-lyase from rice plants. *Eur. J. Biochem.* **185**, 19-25.

Miyata, T., Miyazawa, S., and Yasunaga, T. (1979). Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.* **12**, 219-236.

Mogen, B.D., MacDonald, M.H., Graybosch, R., and Hunt, A.G. (1990). Upstream sequences other than AAUAAA are required for efficient messenger RNA 3'-end formation in plants.| \   *Plant Cell* **2**, 1261-1272.

Mohle, B., Heller, W., and Wellmann, E. (1985). UV-induced biosynthesis of quercetin 3-O-β-D-glucuronide in dill cell cultures. *Phytochemistry* **24**, 465-467.

Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. In: Methods in Enzymology, vol. **155**, R. Wu (ed), Academic Press, San Diego, pp 335-350.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucl. Acids Res.* **8**, 4321-4325.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into *Petunia* results in reversible co-suppression of homologous genes *in trans.* *Plant Cell* **2**, 279-289.

Niesbach-Klosgen, U., Barzen, E., Bernhardt, J., Rohde, W., Schwarz-Sommer, Z., Reif, H.J., Wienand, U., and Saedler, H. (1987). Chalcone synthase genes in plants: a tool to study evolutionary relatioships. *J. Mol. Evol.* **26**, 213-225.

Nussinov, R. (1990). Sequence signals in eukaryotic upstream regions. *Critical Reviews in Biochemistry and Molecular Biology* **25**, 185-224.

Ohl, S., Hahlbrock, K., and Schafer, E. (1989). A stable blue-light-derived signal modulates ultraviolet-light-induced activation of the chalcone synthase gene in cultured parsley cells. *Planta* **177**, 228-236.

Ohl, S., Hedrick, S.A., Chory, J., and Lamb, C.J. (1990). Functional properties of a phenylalanine ammonia-lyase promoter from *Arabidopsis.* *Plant Cell* **2**, 837-848.

Ohta, T. (1983). On the evolution of multigene families. *Theoretical Population Biology* **23**, 216-240.

Paiva, N.L., Edwards, R., Sun, Y., Hrazdina, G., and Dixon, R.A. (1991). Stress responses in alfalfa (*Medicago sativa* L.) 11. Molecular cloning and expression of alfalfa isoflavone reductase, a key enzyme of isoflavonoid phytoalexin biosynthesis. *Plant Mol. Biol.* **17**, 653-667.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A., and Saedler, H. (1987). The regulatory *c1* locus of *Zea Mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* **6**, 3553-3558.

Peck, L.J., and Wang, J.C. (1981). Sequence dependence of the helical repeat of DNA in solution. *Nature* **292**, 375-378.

Pedersen, T.J., Arwood, L.J., Spiker, S., Guiltinan, M.J., and Thompson, W.F. (1991). High mobility group chromosomal proteins bind to AT-rich tracts flanking plant genes. *Plant Mol. Biol.* **16**, 95-104.

Pichersky, E., Bernatzky, R., Tanksley, S.D., Breidenbach, R.W., Kausch, A.P., and Cashmore, A.R. (1985). Molecular characterization and genetic mapping of two clusters of genes encoding chlorophyll a/b binding proteins in *Lycopersicon esculentum* (tomato). *Gene* **40**, 247-258.

Pichersky, E., Bernatzky, R., Tanksley, S.D., and Cashmore, A.R. (1986). Evidence for selection as a mechanism in the concerted evolution of *Lycopersicon esculentum* (tomato) genes encoding the small subunit of ribulose-1,5-bisphospahte carboxylase/oxygenase. *Proc. Natl. Acad. Sci. USA* **83**, 3880-3884.

Pichersky, E. (1990). Nomad DNA - a model for movement and duplication of DNA sequences in plant genomes. *Plant Mol. Biol.* **15**, 437-448.

Polans, N.O., Weeden, N.F., and Thompson, W.F. (1985). Inheritance, organization, and mapping of *rbcS* and *cab* multigene families in pea. *Proc. Natl. Acad. Sci. USA* **82**, 5083-5087.

Proudfoot, N. (1984). The end of the message and beyond. *Nature* **307**, 412-413.

Reed, K.C., and Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucl. Acids Res.* **13**, 7207-7221.

Reif, H.J., Niesbach, U., Deumling, B., and Saedler, H. (1985). Cloning and analysis of two genes for chalcone synthase from *Petunia hybrida*. *Mol. Gen. Genet.* **199**, 208-215.

Reimold, U., Kroger, M., Kreuzaler, F., and Hahlbrock, K. (1983). Coding and 3' non-coding nucleotide sequence of chalcone synthase mRNA and assignment of amino acid sequence of the enzyme. *EMBO J.* **2**, 1801-1805.

Rhodes, D., and Klug, A. (1981). Sequence-dependent helical periodicity of DNA. *Nature* **292**, 378-380.

Rohde,W., Dorr, S., Salamini, F., and Becker, D. (1991). Structure of a chalcone synthase gene from *Hordeum vulgare*. *Plant Mol. Biol.* **16**, 1103-1106.

Rolfe, B.G. (1988). Flavones and isoflavones as inducing substances of legume nodulation. *Biofactors* **1**, 3-10.

Ryder, T.B., Cramer, C.L., Bell, J.N., Robbins, M.P., Dixon, R.A., and Lamb, C.J. (1984). Elicitor rapidly induces chalcone synthase mRNA in *Phaseolus vulgaris* cells at the onset of the phytoalexin defense response. *Proc. Natl. Acad. Sci. USA* **81**, 5724-5728.

Ryder, T.B., Hedrick, S.A., Bell, J.N., Liang, X., Clouse, S.D., and Lamb, C.J. (1987). Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol. Gen. Genet.* **210**, 219-233.

Saedler, H., and Nevers, P. (1985). Transposition in plants: a molecular model. *EMBO J.* **4**, 585-590.

Saiki, R.K. (1990). Amplification of genomic DNA. In: PCR Protocols. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White (eds), Academic Press, San Diego, pp13-20.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning. A Laboratory Manual. 2nd edition. Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.* USA **74**, 5463-5467.

Scheel, D., Dangl, J.L., Douglas, C., Hauffe, K.D., Herrmann, A., Hoffmann, H., Lozoya, E., Schulz, W., and Hahlbrock, K. (1987). Stimulation of phenylpropanoid pathways by environmental factors. In: Plant Molecular Biology. D. von Wettstein and N.-H. Chua (eds), Plenum Press, pp 315-326.

Scheel, D., and Parker, J.E. (1990). Elicitor recognition and signal transduction in plant defense gene activation. *Z. Naturforsch* **45c**, 569-575.

Schindler, U., and Cashmore, A.R. (1990). Photoregulated gene expression may involve ubiquitous DNA binding proteins. *EMBO J.* **9**, 3415-3427.

Schmelzer, E., Jahnen, W., and Hahlbrock, K. (1988). *In situ* localization of light-induced chalcone synthase mRNA, chalcone synthase, and flavonoid end products in epidermal cells of parsley leaves. *Proc. Natl. Acad. Sci. USA* **85**, 2989-2993.

Schmid, J., Doerner, P.W., Clouse, S.D., Dixon, R.A., and Lamb, C.J. (1990). Developmental and environmental regulation of a bean chalcone synthase promoter in transgenic tobacco. *Plant Cell* **2**, 619-631.

Schoppner, A., and Kindl, H. (1984). Purification and properties of a stilbene synthase from induced cell suspension cultures of peanut. *J. Biol. Chem.* **259**, 6806-6811.

Schroder, G., Brown, J.W.S., and Schroder, J. (1988). Molecular analysis of resveratrol synthase: cDNA, genomic clones, and relationship with chalcone synthase. *Eur. J. Biochem.* **172**, 161-169.

Schroder, J., and Schroder, G. (1990). Stilbene and chalcone synthases: related enzymes with key functions in plant-specific pathways. *Z. Naturforsch* **45c**, 1-8.

Schroder, J., Lanz, T., Tropf, S., and Schroder, G. (1991). Key enzymes in phytoalexin biosynthesis: the active site in resveratrol and chalcone synthase. In: Molecular Biology of Plant Growth and Development. Program and Abstracts of the Third International Congress of Plant Molecular Biology, Tucson, Arizona, U.S.A. # Poster abstract 1100.

Schulz, W., Eiben, H.-G., and Hahlbrock, K. (1989). Expression in *Escherichia coli* of catalytically active phenylalanine ammonia-lyase from parsley. *FEBS letters* **258**, 335-338.

Schulze-Lefert, P., Dangl, J.L., Becker-Andre, M., Hahlbrock, K., and Schulz, W. (1989a). Inducible *in vivo* DNA footprints define sequences necessary for UV light activation of the parsley chalcone synthase gene. *EMBO J.* **8**, 651-656.

Schulze-Lefert, P., Becker-Andre, M., Schulz, W., Hahlbrock, K., and Dangl, J.L. (1989b). Functional architecture of the light-responsive chalcone synthase promoter from parsley. *Plant Cell* **1**, 707-714.

Schuz, R., Heller, W., and Hahlbrock, K. (1983). Substrate specificity of chalcone synthase from *Petroselinum hortense. J. Biol. Chem.* **258**, 6730-6734.

Slightom, J.L., and Drong, R.F. (1988). Procedures for constructing genomic clone banks. In: Plant Molecular Biology Manual. S.B. Gelvin and R. A. Schilperoort (eds), Kluwer Academic Publishers, Dordrecht, Section A8: 1-42.

Smith, G.P. (1974). Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symposia On Quantitative Biology* **38**, 507-513.

Solomon, M.J., Strauss, F., and Varshavsky, A. (1986). A mammalian high mobility group protein recognizes any stretch of six AT base pairs in duplex DNA. *Proc. Natl. Acad. Sci. USA* **83**, 1276-1280.

Sommer, H., and Saedler, H. (1986). Structure of the chalcone synthase gene of *Antirrhinum majus*. *Mol. Gen. Genet.* **202**, 429-434.

Sommer, H., Bonas, U., and Saedler, H. (1988). Transposon-induced alterations in the promoter region affect transcription of the chalcone synthase gene of *Antirrhinum majus*. *Mol. Gen. Genet.* **211**, 49-55.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucl. Acids Res.* **10**, 4731-4751.

Staden, R. (1984). Measurements of the effects that coding for a protein has on a DNA sequence and their use for finding genes. *Nucl. Acids Res.* **12**, 551-567.

Staden, R. (1987). Computer handling of DNA sequencing projects. In: Nucleic acid and protein sequence analysis: a practical approach. M.J. Bishop and C.J. Rawlings (eds). IRL Press, Oxford, pp 173-217.

Staiger, D., Kaulen, H., and Schell, J. (1989). A CACGTG motif of the *Antirrhinum majus* chalcone synthase promoter is recognized by an evolutionary conserved nuclear protein. *Proc. Natl. Acad. Sci.* USA **86**, 6930-6934.

Staiger, D., Becker, F., Schell, J., Koncz, C., and Palme, K. (1991). Purification of tobacco nuclear proteins binding to a CACGTG motif of the chalcone synthase promoter by DNA affinity chromatography. *Eur. J. Biochem.* **199**, 519-527.

Stoessl, A., and Stothers, J.B. (1979). The incorporation of [1,2-^{13}C$_2$] acetate into pisatin to establish the biosynthesis of its polyketide moiety. *Z. Naturforsch* **34c**, 87-89.

Suggs, S.V., Hirose, T., Miyake, T., Kawashima, E.H., Johnson, M.J., Itakura, K., and Wallace, R.B. (1981). Use of synthetic oligodeoxyribonucleotides for the isolation of specific cloned DNA sequences. In: Developmental Biology Using Purified Genes. ICN-UCLA Symp. Mol. Cell. Biol. vol 23, D.D. Brown (ed), Academic Press, New York, pp 683-693.

Sugita, M., Manzara, T., Pichersky, E., Cashmore, A., and Gruissem, W. (1987). Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. *Mol. Gen. Genet.* **209**, 247-256.

Sutfeld, R., and Wiermann, R. (1980). Chalcone synthesis with enzyme extracts from tulip anther tapetum using a biphasic enzyme system. *Arch. Biochem. Biophys.* 201, 64-72.

Sutliff, T.D., Huang, N., Litts, J.C., and Rodriguez, R.L. (1991). Characterization of an α-amylase multigene family in rice. *Plant Mol. Biol.* **16**, 579-591.

Tabor, S., and Richardson, C.C. (1989). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase 1. *Proc. Natl. Acad. Sci. USA* **86**, 4076-4080.

Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M., and Chambon, P. (1986). Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter. *Nature* **319**, 121-126.

Tanaka, Y., Matsuoka, M., Yamanoto, N., Ohashi, Y., Kano-Murakami, Y., and Ozeki, Y. (1989). Structure and characterization of a cDNA clone for phenylalanine ammonia-lyase from cut-injured roots of sweet potato. *Plant Physiol.* **90**, 1403-1407.

Tanksley, S.D., and Pichersky, E. (1988). Organization and evolution of sequences in the plant nuclear genome. In: Plant Evolutionary Biology. L.D. Gottlieb and S.K, Jain (eds), Chapman and Hall, London, pp55-83.

Taylor, B. and Powell, A. (1982). Isolation of plant DNA and RNA. *Focus* **4:3**, 4-6.

Taylor, B.H., and Larkin, P.J. (1988). Analysis of electroporation efficiency in plant protoplasts. *Aust. J. Biotech.* **1**, 52-57.

Tietjen, K.G., and Matern, U. (1983). Differential response of cultured parsley cells to elicitors from two non-pathogenic strains of fungi. *Eur. J. Biochem.* **131**, 409-413.

Timko, M.P., Kausch, A.P., Hand, J.M., Cashmore, A.R., Herrera-Estrella, L., Van den Broeck, G., and Van Montagu, M. (1985). Structure and expression of nuclear genes encoding polypeptides of the photosynthetic apparatus. In: <u>Molecular Biology of the Photosynthetic Apparatus</u>. K.E. Steinback, S. Bonitz, C.J. Arntzen, and L. Bogorad (eds), Cold Spring Harbor Laboratory, New York.

van den Bulcke, M., Bauw, G., Castresana, C., van Montagu, M., and Vandekerckhove, J. (1989). Characterization of vacuolar and extracellular β(1,3)-glucanases of tobacco: evidence for a strictly compartmentalized plant defense system. *Proc. Natl. Acad. Sci. USA* **86**, 2673-2677.

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N.M., and Stuitje, A.R. (1990). Flavonoid genes in *Petunia*: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2**, 291-299.

van de Locht, U., Meier, I., Hahlbrock, K., and Somssich, I.E. (1990). A 125bp promoter fragment is sufficient for strong elicitor-mediated gene activation in parsley. *EMBO J.* **9**, 2945-2950.

van Loon, L.C. (1985). Pathogenesis-related proteins. *Plant Mol. Biol.* **4**, 111-116.

van der Meer, I.M., Spelt, C.E., Mol, J.N.M., and Stuitje, A.R. (1990). Promoter analysis of the chalcone synthase (chsA) gene of *Petunia*

hybrida: a 67bp promoter region directs flower-specific expression. *Plant Mol. Biol.* **15**, 95-109.

van Tunen, A.J., Hartman, S.A., Mur, L.A., and Mol, J.N.M. (1989). Regulation of chalcone flavanone isomerase (CHI) gene expression in *Petunia hybrida*: the use of alternative promoters in corolla, anthers and pollen. *Plant Mol. Biol.* **12**, 539-551.

Vieira, J., and Messing, J. (1987). Production of single-stranded plasmid DNA. In: <u>Methods in Enzymology</u>, vol. **153**, R. Wu and L. Grossman (eds), Academic Press, San Diego, pp 3-11.

Walsh, J.B. (1987). Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics* **117**, 543-557.

Wanner, L.A., and Gruissem, W. (1991). Expression dynamics of the tomato *rbc*S gene family during development. *Plant Cell* **3**, 1289-1303.

Weiher, H., Konig, M., and Gruss, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. *Science* **219**, 626-631.

Weising, K., and Kahl, G. (1991). Towards an understanding of plant gene regulation: the action of nuclear factors. *Z. Naturforsch* **46c**, 1-11.

Weisshaar, B., Armstrong, G.A., Block, A., da Costa e Siva, O., and Hahlbrock, K. (1991). Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. *EMBO J.* **10**, 1777-1786.

Welle, R., and Grisebach, H. (1987). Purification and properties of chalcone synthase from cell suspension cultures of soybean. *Z. Naturforsch* **42c**, 1200-1206.

Welle, R., and Grisebach, H. (1988). Isolation of a novel NADPH-dependent reductase which coacts with chalcone synthase in the biosynthesis of 6'-deoxychalcone. *FEBS lett.* **236**, 221-225.

Welle, R., and Grisebach, H. (1989). Phytoalexin synthesis in soybean cells: elicitor induction of reductase involved in biosynthesis of 6'-deoxychalcone. *Arch. Biochem. Biophys* **272**, 97-102.

Wellmann, E. (1983). UV radiation in photomorphogenesis. In: Photomorphogenesis. W. Shropshire Jr., and H. Mohr (eds), Springer-Verlag, Berlin, pp 745-756.

Wienand, U., Sommer, H., Schwarz, Z., Shepherd, N., Saedler, H., Kreuzaler, F., Ragg, H., Fautz, E., Hahlbrock, K., Harrison, B., and Peterson, P.A. (1982). A general method to identify plant structural genes among genomic DNA clones using transposable element induced mutations. *Mol. Gen. Genet.* **187**, 195-201.

Wienand, U., Weydemann, U., Niesbach-Klosgen, U., Peterson, P.A., and Saedler, H. (1986). Molecular cloning of the c2 locus of *Zea mays*, the gene coding for chalcone synthase. *Mol. Gen. Genet.* **203**, 202-207.

Williams, J.G., and Mason, P.J. (1985). Hybridisation in the analysis of RNA. In: Nucleic Acid Hybridization - A Practical Approach. B.D. James and S.J. Higgins (eds), IRL Press, pp139-160.

Wines, D.R., Brady, J.M., Southard, E.M., and MacDonald, R.J. (1991). Evolution of the rat kallikrein gene family: gene conversion leads to functional diversity. *J. Mol. Evol.* **32**, 476-492.

Wingate, V.P.M., Lawton, M.A., and Lamb, C.J. (1988). Glutathione causes a massive and selective induction of plant defense genes. *Plant Physiol.* **87**, 206-210.

Wingender, R., Rohrig, H., Horicke, C., Wing, D., and Schell, J. (1989). Differential regulation of soybean chalcone synthase genes in plant defence, symbiosis and upon environmental stimuli. *Mol. Gen. Genet.* **218**, 315-322.

Wingender, R., Rohrig, H., Horicke, C., and Schell, J. (1990). *cis*-Regulatory elements involved in ultraviolet light regulation and plant defense. *Plant Cell* **2**, 1019-1026.

Wright, J.M., and Dixon, G.H. (1986). Induction by torsional stress of a cruciform conformation 5' upstream of the gene for a high mobility group protein from trout and the specific binding to flanking sequences by the gene product. *J. Cell Biol.* **103**, 43a.

Yang, Z., Park, H., Lacy, G.H., and Cramer, C.L. (1991). Differential activation of potato 3-hydroxy-3-methylglutaryl coenzyme A reductase genes by wounding and pathogen challenge. *Plant Cell* **3**, 397-405.