STUDIES ON THE GENES FOR THE ENZYMES OF THE SHIKIMATE PATHWAY FROM *Pisum sativum*

Stewart P. Granger

Submitted for the degree of Doctor of Philosophy

in the Faculty of Science, Univerisity of Glasgow

June 1989

Acknowledgements

I would like to thank the following people who have contributed towards the work described in this thesis:

Professor J.R. Coggins for his supervision, encouragement and financial support during the course of the project. The late Professor R.M.S. Smellie for making available the facilities of the Biochemistry Department. Dr D. P. Leader for his help and advice with computing. My friends in Lab B4a and D-floor, past and present, for making the lab an enjoyable place to work.

Most of all I thank my parents for their endless support and encouragement throughout my studies and my wife Karen, for her skill and patience in typing this thesis and for many other things.

Abbreviations

The abbreviations used in this thesis are as set out in 'Instructions to Authors', Biochemical Journal (1985) 225, 1-26, except the following:

a.a	amino acid
Amp	ampicillin
Ax	absorbance at x nm measured with a 1 cm path length
(k)bp	(kilo) base pairs
cpm	counts per minute
(k)Da	(kilo) daltons
DAHP	3-deoxy-D-arabino-heptulosonate-7-phosphate
DHQ	dehydroquinate
DHS	dehydroshikimate
DTT	dithiothreitol
DNase	deoxyribonuclease
E0	DAHP synthase
E1	DHQ synthase
E2	dehydroquinase
E3	shikimate dehydrogenase
E4	shikimate kinase
E5	EPSP synthase
E6	chorismate synthase
EPSP	5-enolpyruvylshikimate-3-phosphate
IPTG	isopropyl-B-D-thiogalactoside
* Mr	Molecular mass (relative)
MW	molecular weight

ORF	open reading frame
PEG	polyethylene glycol
Pi	inorganic phosphate
RF	replicative form
RNase	ribonuclease
S3P	shikimate-3-phosphate
SDS	sodium dodecyl sulphate
SDS-PAGE	polyacrylamide gel electrophoresis in the presence of SDS
TCA	trichloroacetic acid
u (units)	units of enzyme activity
X-gal	5-bromo-4-chloro-3-indolyl-β-galactoside
* l.m.p.	low melting point

Summary

Aromatic compounds are synthesised via the shikimate pathway. The fungal pentafunctional arom enzyme has five shikimate pathway activities on one polypeptide chain whereas in bacteria, all seven activities are separate enzymes. In all plants, which have been examined, including Pisum sativum, the shikimate pathway enzymes are separable except for 3-dehydroquinase and shikimate dehydrogenase. These activities occur on a single bifunctional polypeptide. The genes for the shikimate pathway enzymes have been isolated from a variety of microbial sources. This thesis is concerned with attempts to isolate the genes for the shikimate pathway enzymes from P. sativum.

A great deal of effort has been made to cross-hybridise the P. sativum or Arabidopsis thaliana genes for the shikimate pathway enzymes with either regions of the S. cerevisiae ARO1 gene which encodes S. cerevisiae arom or with the individual E. coli shikimate pathway genes. As part of this work a P. sativum genomic library has been constructed. However the desired genes have not been isolated using this strategy because there is insufficient homology between the microbial and plant genes.

A P. sativum cDNA library has been constructed from size fractionated poly A^+ mRNA and 21 cDNA clones which hybridise to a Petunia hybrida cDNA probe have been isolated and characterised. The complete amino acid sequence of the mature P. sativum EPSP synthase (E5) has been

deduced from the nucleotide sequence of one of the cDNA clones. The mature enzyme consists of 447 a.a. and has a calculated Mr of 48,716. The cDNA clone appears to encode part of a transit peptide which occurs at the N-terminus of the mature E5.

The P. sativum E5 sequence has been compared with other E5 sequences and a variety of conserved regions have been identified between plants, bacteria and fungi. The plant E5 sequences are highly homologous to one another. The P. sativum E5 amino acid sequence exhibits more homology to the bacterial sequences than to the fungal sequences.

Contents | Page No.

List of Figures

List of Tables

C H A P T E R 1

INTRODUCTION

1.1 The shikimate pathway

The biosynthesis of aromatic amino acids takes place in plants and microorganisms by way of the shikimate pathway (Fig 1.1). This pathway does not have a counterpart in animal systems and so the primary products of the pathway, phenylalanine, tyrosine and tryptophan are considered to be essential in the diet of animals. The seven steps of the pathway lead from erythrose-4-phosphate and phosphoenolypyruvate (PEP) to chorismate. Chorismate is a particularly versatile intermediate, (Fig 1.2) being the common precursor to the aromatic amino acids and to a variety of other primary and secondary metabolites including ubiquinone, plastoquinone, folic acid, tocopherol and vitamin K (Haslam, 1974; Weiss and Edwards, 1980; Gilchrist and Kosuga, 1980).

In higher plants the aromatic amino acids connect with pathways of secondary metabolism which synthesise alkaloids, coumarins, flavanoids, lignins, indole derivatives and other phenolic compounds (Conn, 1986). The metabolic impact of the shikimate pathway in plants is very great because of these connecting pathways and it has been estimated that between 10 and 35% of the dry weight of higher plants is composed of aromatic compounds which have at some point traversed the shikimate pathway (Boudet et al., 1985).

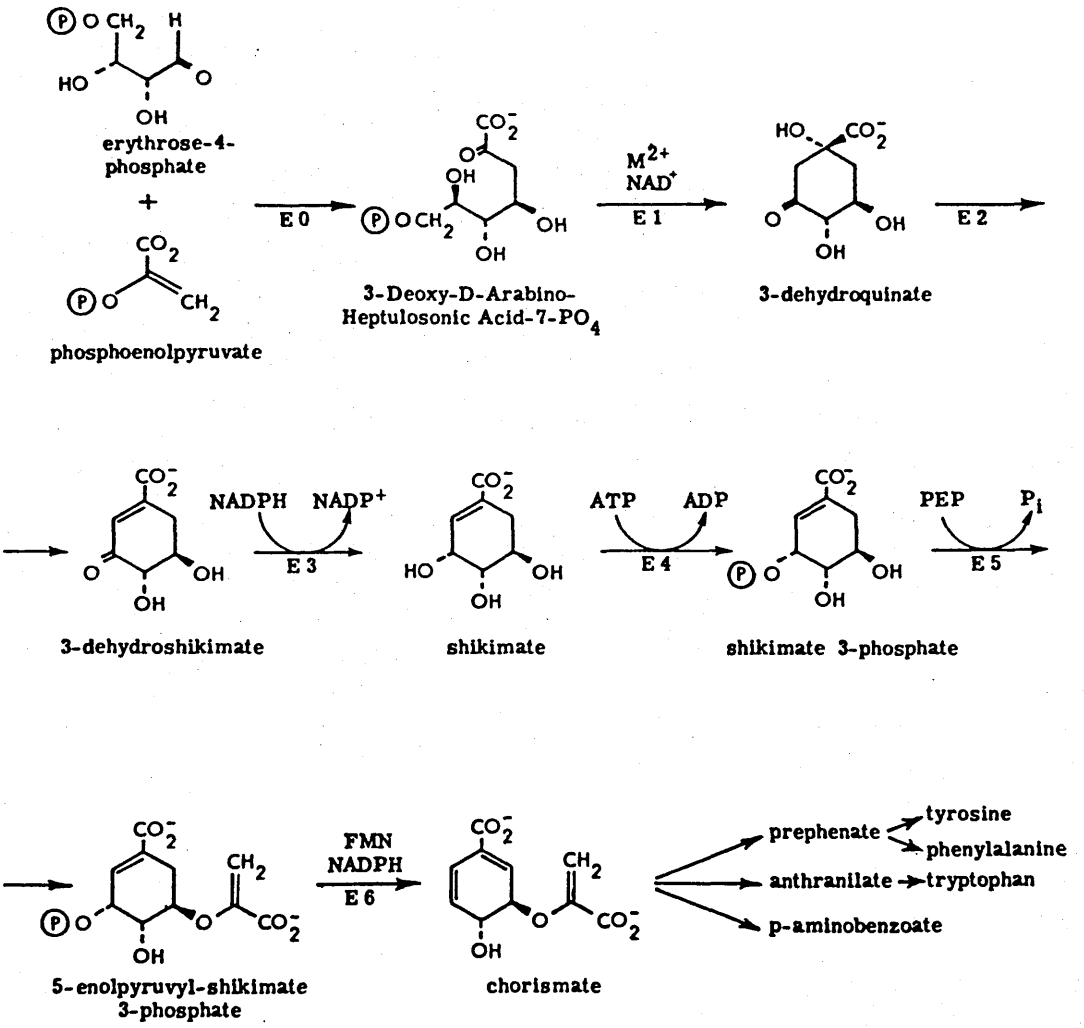


Figure 1.1 The shikimate pathway

Abbreviations used:

E0	DAHP synthase
E1	DHQ synthase
E2	dehydroquinase
E3	shikimate dehydrogenase
E4	skimimate kinase
E5	EPSP synthase
E6	chorismate synthase

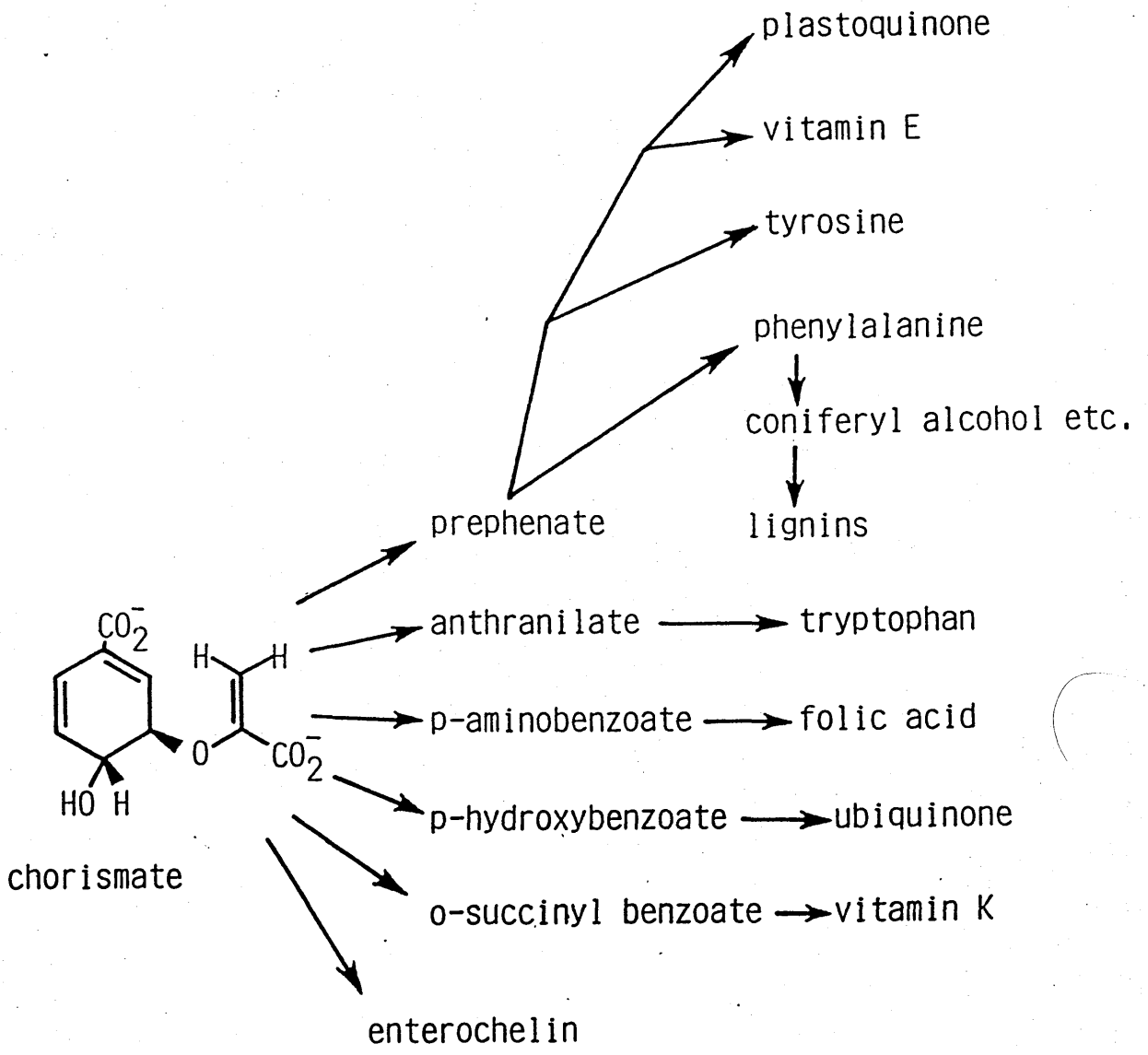


Figure 1.2 Chorismate is a versatile intermediate in plants and microorganisms

1.2 Organisation of the shikimate pathway

1.2.1 Organisation in prokaryotes

The seven enzyme-catalysed reactions that make up the shikimate pathway are the same in all organisms, but the structural organisation of the enzymes catalysing the reactions of the pathway varies. In Escherichia coli and most bacteria which have been studied (Berlyn and Giles, 1969; Coggins et al., 1985) the enzymes have been shown to be separable and monofunctional. An exception to this is Bacillus subtillus which has DAHP synthase (E0) and chorismate mutase activities on the one polypeptide (Nakatasukasa and Nester, 1972). The genes of the shikimate pathway enzymes in E. coli (Bachman, 1983) and in Salmonella typhimurium (Sanderson and Roth, 1983) were shown to be located at unlinked regions of the chromosome and those for the E. coli enzymes have been cloned and sequenced (Zurawski et al., 1978; Davis and Davidson, 1982; Shultz et al, 1984; Duncan, 1984; Duncan et al., 1984; 1986; Miller and Coggins, 1986; Anton and Coggins, 1988; White et al., 1988).

1.2.2 Organisation in fungi

In all fungi which have been examined (Giles et al., 1967; Ahmed and Giles, 1969) the five central steps of the shikimate pathway (numbered E1 to E5 in Fig 1.1) co-sediment on a sucrose gradient. It was thought that these enzymes formed a multienzyme complex, the arom complex, consisting of five types of polypeptide chain (Jacobson et

al., 1972; Gaertner, 1972; Gaertner and Cole, 1977). These enzyme activities have however subsequently been shown to occur on a single pentafunctional polypeptide in Neurospora crassa (Lumsden and Coggins, 1977, Schizosaccharomyces pombe (Nakanishi and Yamamoto, 1984), Aspergillus nidulans (Hawkins, 1985) and Saccharomyces cerevisiae (Duncan et al., 1987). The gene for arom has been cloned and sequenced from A. nidulans (Hawkins, 1985; Charles et al., 1986) and S. cerevisiae (Duncan et al., 1987). The remaining two activities, EO and E6, have been shown to be separable in N. crassa (Nimmo and Coggins, 1981; White et al., 1988).

The mechanism for coordinating the expression of the genes for the shikimate pathway enzymes in bacteria is complex and not yet fully understood (Herrman and Somerville, 1983). By having five of the enzymic functions catalysing the five central sequential steps of the shikimate pathway on a single multifunctional polypeptide chain, the problem of co-ordinating the expression of the five separate enzyme activities is avoided in fungi (Duncan et al., 1987; 1988).

1.2.3 Homology of the bacterial and fungal enzymes

The amino acid sequence of the five central shikimate pathway enzymes (E1 to E5) from E. coli were compared to the amino acid sequences of the S. cerevisiae and A. nidulans arom polypeptides (Duncan et al., 1987). Obvious regions of homology between the bacterial and fungal sequences were apparent confirming that the order of activities along the arom polypeptide was E1, E5, E4, E2, E3. The homologies break

down towards the end of the E. coli sequences indicating that the domains containing the shikimate pathway enzyme activities are separated by non-homologous linker regions. The most likely explanation for the origin of the pentafunctional fungal arom polypeptide is that it has arisen by the fusion of ancestral E. coli like genes (Duncan et al., 1987).

1.2.4 Organisation of the shikimate pathway in higher plants

In higher plants, five of the enzymes of the shikimate pathway are separable and therefore monofunctional (Mousdale and Coggins, 1984; 1985; 1986) whereas the 3-dehydroquinase (E2) and shikimate dehydrogenase (E3) co-purify (Koshiba, 1978; 1979; Mousdale et al., 1987).

The co-purification of E2 and E3 implied an enzyme complex where the two proteins were tightly bound or alternatively, a bifunctional enzyme where both activities are present on the one polypeptide chain. Demonstration by chromatographic techniques that the two activities co-purified on the one polypeptide in the moss Physcomitrella patens (Polley, 1978) and complete purification and electrophoretic analysis of the bifunctional polypeptide from Pisum sativum, Mr 59,000 (Mousdale et al., 1987) confirmed that the latter possibility was the case in these plants. Analysis of other plant species (see Mousdale et al., 1987) suggests the universal occurrence of a bifunctional polypeptide.

The catalytic sites of the two activities, although present on a single polypeptide, are apparently independent since the forward reaction of P. sativum E3 is not affected by the presence of dehydroquinate (DHQ), the substrate for the E2 activity (Mousdale et al., 1987).

DAHP synthase (E0) has been purified to homogeneity from carrot and potato and this will be discussed in detail in a later section. The gene for the potato E0 has been cloned and partially sequenced (Herrman, 1987) but the sequence data has as yet not been reported.

DHQ synthase (E1) and shikimate kinase (E4) have been detected in P. sativum (Mousdale and Coggins, 1985) but as yet have not been purified to homogeneity.

EPSP synthase (E5) is probably the most characterised enzyme of the shikimate pathway. This is because it is the target for the herbicide glyphosate. This enzyme will be discussed in detail in a later section as will the clones of the E5 genes which have been obtained from several higher plant species.

Chorismate synthase (E6) has been detected in P. sativum and found to be similar to the bacterial enzyme rather than that of N. crassa (Mousdale and Coggins, 1986). In the fungus, the E6 activity and a diaphorase activity, that reduces FMN in the presence of NADPH, occur on a bifunctional polypeptide. Both the plant and the bacterial E6 enzyme do not contain a diaphorase activity and therefore must be

assayed under reducing conditions in the presence of DTT. The P. sativum enzyme has not been purified to homogeneity.

1.3 Subcellular location of the shikimate pathway in plants.

1.3.1 The shikimate pathway of plants is predominantly chloroplastic

Isolated chloroplasts from spinach were shown to be capable of converting either labelled CO_2 or shikimate into aromatic amino acids (Bickel et al., 1978,; Bucholz et al., 1979). This indicated that a complete shikimate pathway, and subsequent steps leading to the formation of the aromatic amino acids, must be located within the chloroplast. All seven shikimate pathway enzymes were found to be present in the stroma of density gradient purified, washed chloroplasts from P. sativum seedlings (Mousdale and Coggins, 1985; 1986) and spinach (Coggins, 1986) confirming that chloroplasts are a major site for the biosynthesis of chorismate, the common precursor of the aromatic amino acids. However chloroplast-localised biosynthesis did not account for total aromatic amino acid biosynthesis (Buchholz et al, 1979) and it was proposed that a spatially separate pathway having its own complement of extrachloroplastic enzymes might exist. If so then the likely subcellular location would be the cytosol since the initial pathway precursors are formed in the cytosol (Ganson et al, 1986). This possibility predicts the existence of isozymic forms of shikimate pathway enzymes which have different sub-cellular locations (Jensen, 1985). Isozymes of several of the purified enzymes from a varity of plant species have been observed as will be described

in the following sections.

1.3.2 Evidence for two shikimate pathways in plants

The hypothesis that two pathways of aromatic amino acid biosynthesis exist in higher plants (Jensen, 1985; Morris *et al.*, 1989) was based mainly on the observation that higher plant species possess isozyme pairs of EO and chorismate mutase (CM). However isozymes of some of the intervening activities have also been demonstrated to be present in both the chloroplast and the cytosol and therefore the prospect of a dual shikimate pathway occurring in higher plants is quite attractive.

1.3.3 Isozymes of DAHP synthase (EO)

Two activities of EO have been identified in *Vigna radiata* (Rubin and Jensen, 1985) and *Nicotiana silvestris* by chromatography on an anion exchange column (Ganson *et al.*, 1986). These activities differ with respect to their metal ion requirements. One activity did not require Mn^{2+} but was stimulated by this cation (denoted EO-Mn) whereas the other activity required Co^{2+}, Mn^{2+} or Mg^{2+} for activity (denoted EO-Co). Conditions could be adjusted so that when one activity was assayed, the other activity was not detectable and using this differential assay the two activities were detected in other plants (Ganson *et al.*, 1986). Percoll density gradient purified chloroplasts were observed to contain EO-Mn exclusively in spinach and virtually exclusively in *N. silvestris* whereas crude plant extract contained

both activities. No EO activity was detected in fractions of the gradient containing microbodies or mitochondria. Therefore the EO-Mn was suggested to be the chloroplastic isozyme whereas EO-Co was the cytosolic isozyme (Ganson et al., 1986). Neither of these proposed isozymes have been substantially purified and so the molecular basis of their difference has not been investigated.

Three EO isozymes from carrot have been observed and the dominant isozyme III has been purified to homogeneity. It has Mr of 53,000 and exists as a dimer (Suzich et al., 1983; 1985). Potato tubers initially appeared to have only one EO isozyme which when purified was shown to resemble the carrot isozyme III in both size and quaternary structure (Pinto et al, 1986). However more recently, (Morris et al., 1989), potato tubers have been shown to contain the two activities of EO using the differential assay technique. It is not clear which isozyme was purified to homogeneity.

1.3.4 Isozymes of chorismate mutase

Two isozymes of chorismate mutase (CM) have been resolved by anion exchange chromatography from a number of plants and in depth studies have been reported from V. radiata, N. silvestris and Sorghum bicolor (Koruki and Conn, 1988a). Although CM is not an enzyme of the shikimate pathway it converts chorismate to prephenate and so is involved in the late stages of the aromatic biosynthetic pathways and therefore merits discussion here.

CM-1 is activated by tryptophan and inhibited by tyrosine and phenylalanine whereas CM-2 is not regulated by aromatic amino acids. Both CM-1 and CM-2 have been purified to homogeneity from sorghum and have Mr 56,000 and 48,000 respectively (Singh et al., 1985). It has been shown that in N. silvestris the CM-1 isozyme Mr 51,000 is located in the chloroplast whereas the CM-2 isozyme Mr 36,000 is located in the cytosol. A report that potato contains only one isozyme of CM (Koruki and Conn, 1988b) has been disputed (Morris et al., 1989).

1.3.5 Isozymes of 3-dehyroquinase/ shikimate dehydrogenase (E3/E3)

Electrophoretic isozymes of E2/E3 have been observed in a number of plants (see Mousdale and Coggins, 1985) including wheat (Neuman and Hart, 1983) and Phaseolus mungo (Koshiba, 1978) although whether the isozymes were present in separate cellular compartments was not shown. In addition, E2/E3 activity from P. sativum seedling extract was resolved into two unequally sized peaks of activity by high performance anion exchange chromatography. However E2/E3 activity from density gradient purified chloroplasts of P. sativum was resolved into a single peak of similar size to the large peak obtained from the whole seedling extract using similar techniques (Mousdale and Coggins, 1985). Thus a small proportion of E2/E3 appears to be present as distinct extra-chloroplastic isozymes (probably cytosolic) with the majority being located within the chloroplast (Mousdale and Coggins, 1985). Furthermore E2/E3 purified to electrophoretic homogeneity from whole P. sativum seedlings can be resolved into three peaks of activity by high performance chromatofocussing whereas E2/E3 purified

exclusively from isolated P. sativum chloroplasts is resolved into only two peaks of activity. Chloroplastic E2/E3 is therefore composed of at least two isozymes (Mousdale et al., 1987). Since the isozymes originate from a single protein band of Mr 59,000 then the differences between them are likely to be either slight changes in amino acid sequence or perhaps some sort of covalent modification. In either event, the change must allow copurification of the isozymes.

1.3.6 Isozymes of EPSP synthase (E5)

A small proportion of E5 activity has also been shown to be present outside the chloroplast in P. sativum suggesting the presence of a non-chloroplastic isozyme (Mousdale and Coggins, 1985). However in cell cultures of Corydalis sempervirens, no evidence for the existance of E5 outside the chloroplast was obtained. Sections of C. sempervirens cells were probed with antibodies against purified E5 and these antibodies were detected by protein A-gold immunochemistry (Smart and Amrhein, 1987). The situation in whole plants of C. sempervirens has not been determined.

1.3.7 Objections to the dual shikimate pathway hypothesis

The observation of non-chloroplastic isozymes of shikimate pathway enzymes is suggestive of a complete second shikimate pathway in higher plants. However until the remaining enzymes (DHQ synthase (E1), shikimate kinase (E4), chorismate synthase (E6)) have been demonstrated to be present outwith the chloroplast and the situation

with respect to E5 has been clarified then a complete non-chloroplastic pathway must still be considered as speculation. In addition, it cannot be ruled out that the cytosolic isozymes of EO and CM are precursor proteins which have not yet been translocated to the chloroplast (Kishore and Shah, 1988). The sub-unit Mr of CM-2 (cytosolic) is 51,000 whereas that of CM-1 (chloroplastic) in 36,000. The 15 kDa difference between the two isozymes could encode a transit peptide which targets CM-2 to the chloroplast and is cleaved off during tranlsocation producing CM-1. It has been demonstrated that a precursor enzyme of the shikimate pathway (pre-E5 from P. hybrida) is fully active (Della-Cioppa et al., 1986) and so CM-2 could equally be an active precursor of CM-1. A similar situation may exist for EO and because the two isozymic forms of EO have not been purified it is difficult to dispute this. On the other hand the isozymes of E2/E3 (Mousdale et al., 1987) arose from high performance chromatographic separation of an E2/E3 function which gave rise to a single band on SDS-PAGE. Both these isozymes therefore have the same subunit molecular weight and it is very unlikely that one isozyme is a precursor form of the other which has not yet been translocated.

1.4 A model for the differential control of two shikimate pathways

In anticipation of the second shikimate pathway being unambiguously identified, it has been proposed that the existance of two shikimate pathways in separate compartments of plant cells allows different forms of regulation of the pathway to occur in each compartment (Jensen, 1986).

In order to appreciate the proposed model of regulation it is necessary to understand the enzymological route for phenylalanine and tyrosine biosynthesis from chorismate in plants.

1.4.1 Arogenate is an intermediate in aromatic amino acid biosynthesis in higher plants.

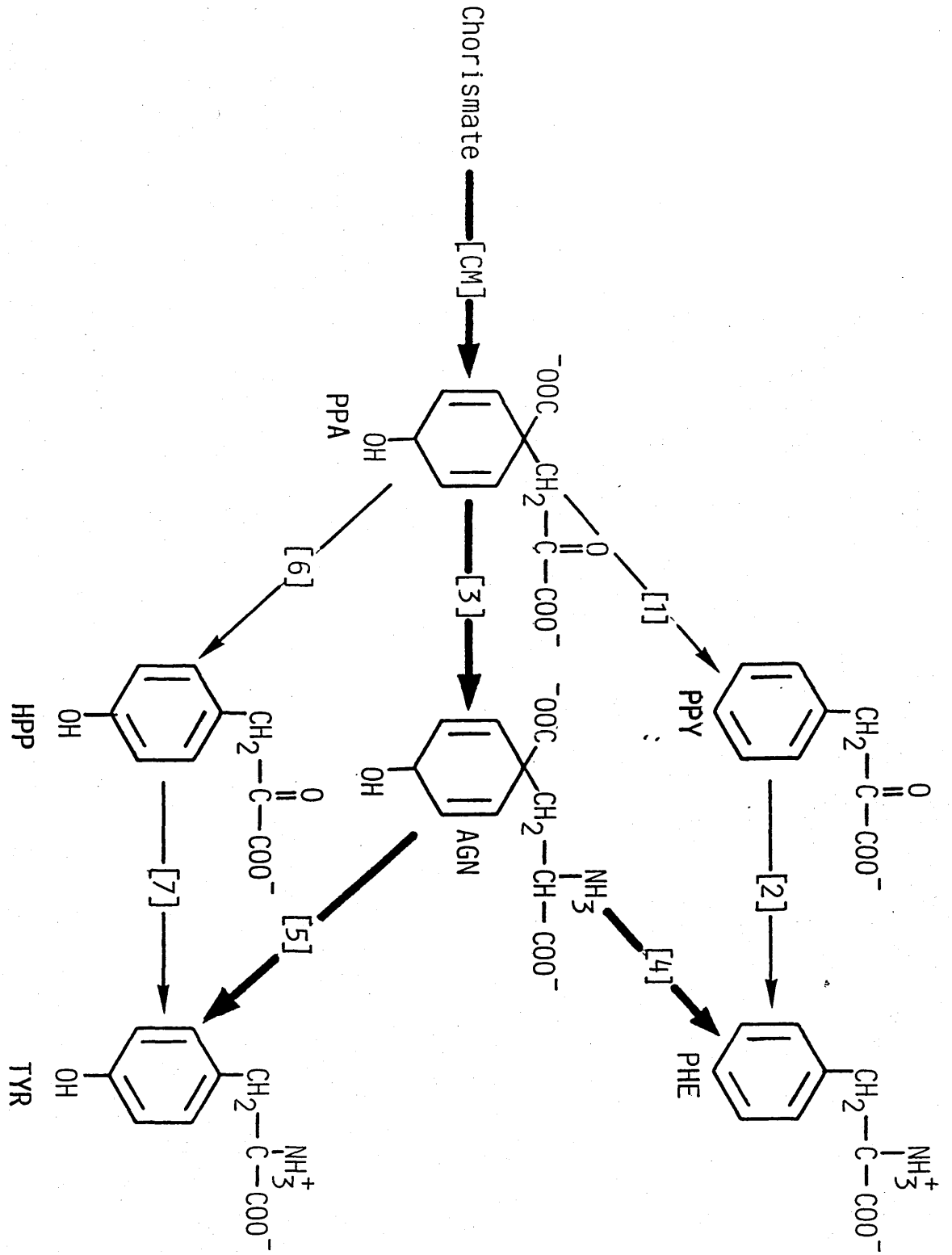
There are several ways in which tyrosine and phenylalanine can be synthesised (Fig 1.3). Some microbes e.g. E. coli, B. subtilis , S. cerevisiae and N. crassa do not use the arogenate route for biosynthesis while others e.g. Pseudomonads use all routes (Jensen, 1986). Prephenate aminotransferase, which converts prephenate to arogenate, and arogenate dehydrogenase, which converts arogenate to tyrosine, have been detected in all plants examined (Jensen, 1986). Arogenate dehydratase, which converts arogenate to phenylalanine, has been detected in N. silvestris and also in washed spinach chloroplasts (Jung et al., 1986). Prephenate dehydratase and prephenate dehydrogenase activities have never been demonstrated in any plant system (except in the developmental stage of seed germination in mung bean). Therefore it is likely that both phenylalanine and tyrosine are synthesised predominantly, if not exclusively, via arogenate in higher plants (Jung et al., 1986).

1.4.2 The model of regulation

Enzymes located at metabolic branch points of the shikimate pathway have been observed to be regulated by end products or intermediates of

Figure 1.3 Alternative pathways of phenylalanine and tyrosine biosynthesis in nature

Chorismate is converted to prephenate (PPA) by chorismate mutase [CM] . The flow routes through arogenate (AGN) are shown with bold arrows and consist of prephenate aminotransferase [3] , arogenate dehydratase [4] and arogenate dehydrogenase [5] . The enzymes catalysing the alternative routes are prephenate dehydratase [1] , phenylpyruvate (PPY) aminotransferase [2] , prephenate dehydrogenase [6] and hydroxyphenylpyruvate (HPP) aminotransferase [7] .


Chorismate
[CM]
PPA
OH
[6]
HPP
OH
[7]
TYR
OH
[5]
[3]
AGN
OH
[1]
PPY
[2]
PHE
[4]

the pathway. The allosteric regulation appears to be different between the enzymes present within the chloroplast and those present outwith the chloroplast. The proposed scheme of regulation is as follows (Fig 1.4; Jenson, 1987).

Within the chloroplast, phenylalanine, tyrosine and tryptophan inhibit their own synthesis by de-activating arogenate dehydratase, arogenate dehydrogenase and anthranilate synthase (Jung et al., 1986) respectively. Tryptophan on the other hand activates chorismate mutase thus there is a build up of arogenate when the three aromatic amino acids are in high physiological concentrations (Jensen, 1985). High arogenate concentration causes feedback inhibition of the plastidial EO-Mn (Rubin and Jensen, 1985). Therefore the pathway is closed down within the plastid when high aromatic amino acid concentration prevails.

The cytosolic enzymes EO-Co and CM-2 are insensitive to inhibition by the aromatic amino acids and arogenate and so perhaps the cytosolic pathway is not closed down when the aromatic amino acids are in excess (Jensen, 1986). Jensen further speculates that the cytosolic pathway supplies connecting networks of secondary metabolism and is active in tissues where demands for protein synthesis have declined.

1.4.3 Further features which support the dual shikimate pathway hypothesis

Several other features of the shikimate pathway are consistent with

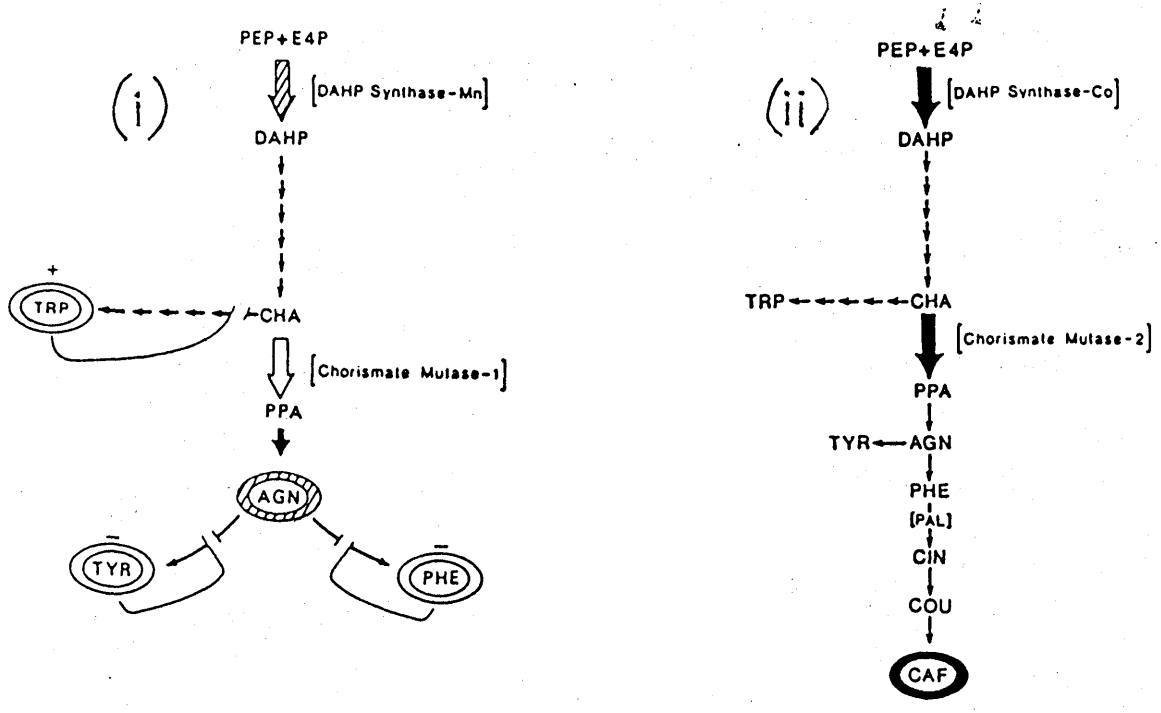


Figure 1.4 Model of regulation of the plant shikimate pathway

(i) The chloroplastic shikimate pathway. (ii) The cystolic shikimate pathway.

Abbreviations used: CHA=chorismate, TRP=tryptophan, PAL =phenylalanine ammonia-lyase, CIN=cinnamic acid, COU=coumaric acid, CAF=caffeic acid, other abbreviations are as defined in Fig 1.3.

the hypothesis of (i) a plastidic shikimate pathway concerned mainly with the biosynthesis of amino acids and probably plastidic aromatic compounds eg plastoquinone and phylloquinone (Rothe et al., 1983) and (ii) a cytosolic pathway concerned with the biosynthesis of polyphenolic compounds.

Chloroplasts are permeable to aromatic amino acids but impermeable to cinnamic acids (Rothe et al., 1983) which are the precursor for the polyphenolic compounds. Therefore if the polyphenols not destined for the chloroplast were synthesised within these organelles, perhaps they could not get through the double membrane into the cytosol.

The E2/E3 distribution of P. sativum has been shown to be age dependent. Up to 14 days the enzyme is predominantly located within the chloroplast but after this time the extra-chloroplastic activity increases whereas the chloroplastic isozyme activity remains relatively constant (Rothe et al, 1983). The activity of the extra chloroplastic isozyme of E2/E3 has been observed to increase as P. sativum seedlings and spinach leaves age (D.M. Mousdale, unpublished results). This perhaps reflects a tendency to increase the amount of secondary metabolite biosynthesis through the shikimate pathway as plants get older.

The purified carrot and potato E0 enzymes are activated by tryptophan (Suzich et al., 1985; Pinto et al., 1986). This tryptophan stimulation was viewed by the authors to be a mechanism by which polyphenol biosynthesis is activated in certain plant tissues. Potato

tubers are very rich in polyphenols. When the cell is energetically well off and has plenty tryptophan then this could be the signal for polyphenol biosynthesis. Activation of EO would ensure carbon flow through the shikimate pathway for the synthesis of the aromatic amino acids and subsequent polyphenol synthesis (Pinto et al., 1986). The EO enzymes which were shown to be tryptophan activated were not demonstrated to be either plastidic or cytosolic. Although the conclusions do not contradict the Jensen hypothesis, it would be appropriate to determine the location of the purified EO enzymes. If they turn out to be cytosolic then this would be consistent with the cytosolic pathway being geared for secondary metabolite biosynthesis.

1.5 Intermediates of the shikimate pathway are common to other metabolic pathways

The regulation of the shikimate pathway in higher plants is further complicated by the observation that dehydroshikimate (DHS), one of the central intermediates of the pathway, can also be converted in one step to either gallic acid or protocatachuic acid (Weiss and Edwards, 1980). Another pathway intermediate, (DHQ) can be converted to quinate by a reversible enzyme reaction catalysed by quinate dehydrogenase which has been detected in several plant species (Graziana et al., 1980; Ranjeva et al., 1983). Large endogenous pools of quinic, protocatechuic and gallic acids occur widely among plants although their function is not known. It is not clear if these aromatic compounds are synthesised via the early steps of the shikimate pathway or if a parallel pathway of aromatic biosynthesis

exists. The presence of both the E2 and E3 catalytic sites on the one bifuctional polypeptide perhaps protects the common intermediate (DHS) from the action of competing enzymes. This may be the reason for the occurence of a bifunctional E2/E3 enzyme in plants (Mousdale et al., 1987).

1.6 EPSP synthase (E5)

1.6.1 Characterisation of E5

EPSP synthase (E5) catalyses the sixth step of the shikimate pathway, converting S3P and PEP to 5-enolpyruvylshikimate-3-phosphate (EPSP) and Pi (Fig 1.1). E5 has been purified to homogeneity from a number of bacterial, fungal and plant sources. It has a Mr 44-55,000 in plants and bacteria (see Kishore and Shah, 1988) but is part of the arom multifunctional enzyme in fungi (Duncan et al., 1987). The genes for E5 from E. coli (Duncan et al., 1984), Salmonella typhimurium (Comai et al., 1983), A. nidulans (Hawkins, 1985; Charles et al., 1986) and S. cerevisiae (Duncan et al., 1987) have been cloned and sequenced. While the work reported in this thesis was being carried out, the E5 genes from Petunia hybrida, Lycopersicon esculentum (tomato) and Arabidopsis thaliana (Klee et al., 1987; Gasser et al., 1988) were cloned and sequenced. These are currently the only published sequences of any plant shikimate pathway gene.

There is great interest in E5 from plants because the herbicide glyphosate specifically inhibits this enzyme. Glyphosate inhibition

of E5 is not limited to plants but also occurs with the bacterial (Amrhein et al., 1980) and fungal (Boocock and Coggins, 1983) enzymes. An extensive array of glyphosate analogues are completely unable to inhibit E5 indicating that the herbicide is an isolated toxophore (Coggins, 1987; 1989).

1.6.2 The kinetic properties of E5

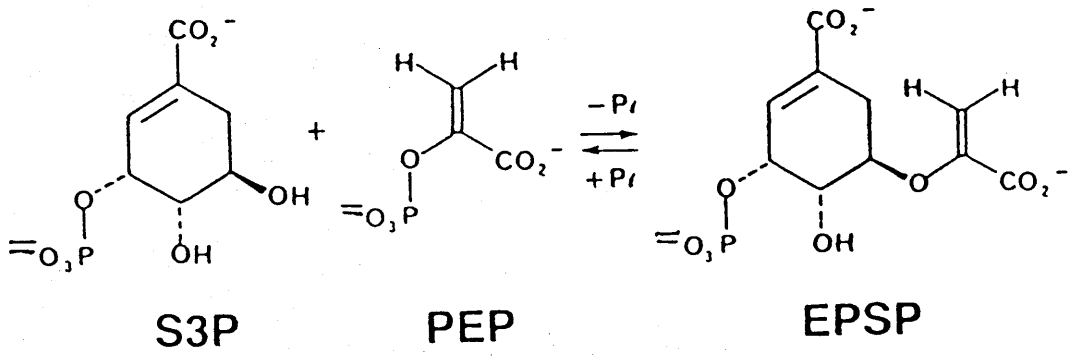
The kinetics of catalysis of E5 have been investigated thoroughly (reviewed by Kishore and Shah, 1988) and it appears that the mechanism proposed for the N. crassa enzyme (Boocock and Coggins, 1983; Boocock, 1983) is universal. E5 has a compulsory order sequential mechanism, where the enzyme firstly binds S3P and can then bind PEP. After catalysis, Pi and EPSP are liberated. Glyphosate acts as a competitive inhibitor with respect to PEP and an uncompetitive inhibitor with respect to S3P during the synthesis of EPSP.

It has been proposed that glyphosate may act as a transition state analogue of PEP with the positively charged nitrogen mimicking the carbonium C-2 of PEP (Steinrucken and Amrhein, 1984). However, since mutants of E5 which are fully active but glyphosate tolerant have been isolated (Stalker et al., 1985; Kishore et al., 1986) then this is now thought to be unlikely (Comai et al., 1985). Glyphosate is not a ground state analogue of PEP since it does not inhibit other PEP utilizing enzymes (Kishore and Shah, 1988).

1.7 Important residues within the sequence of E5

1.7.1 Identification of an active site lysine and arginine

In view of the anionic nature of glyphosate and the substrates which interact with E5 (Fig 1.5) it was anticipated that positively charged amino acid side chains would be present at the active site of E5 which would aid in the binding of these molecules. Chemical modification, affinity labelling and isolation of labelled peptides have indicated that Lys-22 of the E.coli enzyme and Arg-28 of the P. hybrida enzyme are involved at the active site of E5 since their labelling was prevented by the presence of S3P, EPSP or S3P and glyphosate. Arg-131 (P. hybrida) was also labelled however it was unclear whether modification could be prevented by the presence of substrates (Huynh et al., 1988a; Padgette et al., 1988a). All three of these residues are conserved in the E5 sequences which have been determined (Kishore and Shah, 1988) confirming that they are important residues. Lys-23 of P. hybrida E5 (corresponding to Lys-22 in the E. coli enzyme) was replaced using site directed mutagenesis, with three other amino acids: Ala, Glu and Arg. Of these replacements, only the Lys-23 to Arg mutant enzyme was active while the other two replacements led to complete inactivation of the enzyme. This suggested that it was the cationic nature of the residue at position 23 that plays an important role in the activity of E5 (Huynh et al., 1988). Using a method based on the fact that protein fluorescence of E5 is affected upon complex formation with S3P and glyphosate, it was shown that both the wild type and the Lys-23 to Arg mutants could bind S3P and glyphosate. The



GLYPHOSATE

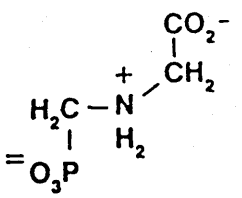


Figure 1.5 The ligands which interact with the active site of E5

Lys-23 to Ala mutant was unable to bind S3P or glyphosate. This indicated that the cationic groups at position 23 played a role in substrate binding (Huynh et al., 1988b).

1.7.2 Identification of other important residues

Similar studies have identified Glu-418 (in the E. coli enzyme) to be intimately involved at the active site. Its chemical modification is blocked by S3P and glyphosate and therefore its negatively charged carboxyl has been proposed to have a role in the binding of the positively charged imino group of glyphosate, (Huynh, 1988). The carboxyl character of this site is conserved in other E5 enzymes (Kishore and Shah, 1988).

Cys-408 (E. coli E5) was identified as being present at or near the active site because modification was blocked by S3P and glyphosate. When Cys-408 was modified with a bulky molecule 5'5'-dithiobis-(2-nitrobenzoic acid) the enzyme was inactivated however when this bulky molecule was replaced by the much smaller cyanide group, activity comparable to the wild type E5 was obtained. This strongly suggested that the Cys-408 was not involved in either binding of substrate or catalysis (Padgette et al., 1988). Interestingly a cysteine residue is present at this position in all E5 sequences (Kishore and Shah, 1988).

Diethylpyrocarbonate inactivation of E5 suggests the presence of an essential histidine residue. The modified residue has not been

identified but His-385 is thought to be the likely candidate since it is conserved in all E5 sequences (Kishore and Shah, 1988).

1.8 The genes of E5 from higher plants

1.8.1 Characterisation of plant E5 genes

The P. hybrida E5 gene was cloned from a cell line (MP4-G) which was induced to overproduce E5 by culturing the cells in increasing concentrations of glyphosate (Shah et al., 1986). E5 was purified from the cell line, its N-terminal sequence was determined and families of oligonucleotides were synthesised based on the deduced DNA sequence. Using these oligonucleotides a cDNA and genomic clone of E5 were obtained from libraries derived from the MP4-G cell line (Shah et al., 1986). It was subsequently shown by Southern blot analysis that the genomic clone of E5 had not been rearranged in the MP4-G cell line (Gasser et al., 1988). P. hybrida E5 is synthesised as a precursor consisting of an amino terminal transit peptide of 72 amino acids and a mature enzyme of 444 amino acids (Shah et al., 1986; Gasser et al., 1988). The P. hybrida E5 gene was used to isolate the L. esculentum E5 cDNA (Gasser et al., 1988) and A. thaliana E5 gene (Klee et al., 1987). The intron/exon boundaries of the E5 gene from P. hybrida and A. thaliana were completely conserved. Both genes contained seven introns (Klee et al., 1987; Gasser et al., 1988) although the total number of bases present as intron DNA in A. thaliana is 8 fold less than in the P. hybrida gene. This reflects the small size of the A. thaliana genome (discussed in Chapter 4).

When southern blots of P. hybrida and A. thaliana genomic DNA were probed with cDNA probes of the P. hybrida and A. thaliana gene respectively, bands which co-migrated with those of the genomic clones were observed, but additional bands were also seen. This is highly suggestive of more than one gene for E5 being present within both the P. hybrida and A. thaliana genome (Shah et al., 1986; Klee et al., 1987).

1.8.2 Overexpression of P. hybrida E5 in E. coli

In order to gain a greater understanding of the mechanism of E5 large quantities of the purified enzyme are essential. All of the plant shikimate pathway enzymes are present at extremely low levels in vivo and those which have been purified to homogeneity requred a 1000 (EO, Pinto et al., 1986), 3000 (E5, Mousdale and Coggins, 1984) and 6,700 (E2/E3, Mousdale et al., 1987) fold purification. The purification of P. hybrida E5 has been made considerably easier by over-expressing the plant cDNA clone in E. coli (Padgette et al., 1987). The cDNA coding for the mature (lacking the chloroplast transit sequence) P. hybrida E5 was cloned behind an E. coli promoter and Shine-Dalgarno sequence. An ATG start codon had to be introduced at the beginning of the mature E5, but otherwise the enzyme was identical in sequence. A 100 fold purification was required in order to obtain E5 which was electrophoretically and kinetically very similar to the mature native P. hybrida E5 (Padgette et al., 1987). This expression technique was utilised to provide sufficient quantities of the mutant P. hybrida E5

enzymes for characterisation.

1.8.3. Levels of expression of E5 in plants

The level of E5 mRNA in several organs of P. hybrida plants has been compared by northern blot analysis. Using the P. hybrida cDNA clone as a probe, expression of E5 was found to be lowest in leaves and pistils, higher in stems, roots and stamens and very much higher (50-fold over leaves) in petals (Gasser et al., 1988). This distribution of E5 mRNA was not found in L. esculentum which is another member of the family Solanaceae. Using the L. esculentum cDNA as a probe levels of E5 mRNA were found to be very similar in all floral parts and in leaves. In addition, they were substantially lower than the high levels found in P. hybrida petals (Gasser et al., 1988). Wild type P. hybrida flowers are purple in colour due to the accumulation of large quanatities of anthocyanin pigments. Tomato flowers do not accumulate these pigments. Phenyalaline is a precursor in the biosynthetic pathway of anthocyanin (Conn, 1986) therefore it was hypothesised that the high level of E5 expression seen in P. hybrida petals is necessary for the production of these aromatic pigments (Gasser et al., 1988). It would seem reasonable that in P. hybrida petals, the expression of other shikimate pathway enzymes would be correspondingly high although this has not been investigated.

The region of the P. hybrida E5 gene which confers tissue specificity to the expression of E5 in P. hybrida has been localised to the region situated between -1800 and -800 bp upstream of the ATG (+1)

translational start codon (Benfey and Chua, 1989). Deletions of the E5 gene between -30 and -800 were not required to give high expression of a reporter gene, chloramphenicol acetyl transferase (CAT), in the petals of regenerated P. hybrida which had been transformed with a variety of constructs containing the CAT gene. Expression of the CAT gene was clearly dependent on the developmental stage of the flower since CAT activity was low before the flower opened, and on the first day after opening but increased rapidly on subsequent days (Benfey and Chua, 1989). The authors indicate that studies on the E5 enzyme of P. hybrida show a similar pattern of developmental regulation.

1.9 Translocation of E5 into chloroplasts

The majority of chloroplast-localised proteins are encoded by nuclear genes and are synthesised as precursor proteins on free cytoplasmic ribosomes. The import of these precursors into chloroplasts has been shown to be mediated by amino-terminal sequences, termed transit peptides, which post-translationally direct the uptake process. After uptake into the chloroplasts, the transit peptide is removed by a proteolytic enzyme, termed stromal protease (reviewed by Keegstra, 1989). The transit peptide of P. hybrida E5 is essential for its translocation into chloroplasts where it is cleaved off to produce the 48,000 kDa mature E5 which is 7 kDa smaller than the precursor (Della-Ciopa et al., 1986). The precursor-E5 (pre-E5) was shown to be enzymatically active. Its activity could be inhibited by the presence of glyphosate (in the presence of S3P). In addition the translocation of pre-E5 to isolated chloroplasts was reduced by 75% in the presence

of glyphosate and S3P. This may reflect a reduction in the ability of the pre-E5 to unfold during translocation when these ligands are bound at its active site. Inhibition of translocation of E5 may be involved in the herbicidal action of glyphosate (Della-Ciopa and Kishore, 1988) although it is unclear whether sufficient S3P would be present in the cytosol of plant cells for this inhibition to occur *in vivo*.

1.10 Glyphosate tolerance in plants

There has been great interest in the engineering of glyphosate tolerance into plants. Two mechanisms by which this could be achieved have received the most attention. Firstly overproduction of E5 within plants and secondly the introduction of a glyphosate insensitive mutant E5 into plants. Both of these mechanisms can overcome concentrations of glyphosate which are sufficient to kill wild type plants.

1.10.1 Overproduction of E5 by natural selection

Plant cell cultures tolerant to glyphosate have been isolated by stepwise selection on increasing concentration of glyphosate (Nafziger *et al.*, 1984; Singer and McDaniel, 1985; Smart *et al.*, 1985, 1986). Using radioactively labelled glyphosate, tolerance of a carrot cell culture was shown not to be due to impaired uptake of glyphosate (Nafziger *et al.*, 1984) and in all cases, increased E5 activity was observed. Regenerated glyphosate tolerant tobacco plants maintained their tolerance in the short term, but this was lost over a 2.5 year

period (Singer and McDaniel, 1985) suggesting that continued selection pressure is required to maintain tolerance.

The mechanism of overproduction of E5 in P. hybrida was demonstrated to be gene amplification (Shah et al., 1986) because probing genomic DNA from P. hybrida glyphosate tolerant cell line with the cloned P. hybrida E5 cDNA indicated multiple copies of the E5 gene. In contrast, overexpression of E5 in a Corydalis sempervirens cell line is not due to gene amplification, since multiple copies of the E5 gene were not present (judged by probing genomic southern blots with an oligonucleotide corresponding to the N-terminus of the C. sepervirens E5 and with the P. hybrida E5 cDNA clone). A different mechanism is therefore responsible, which may involve a mutation in the DNA causing increased transcription of E5 mRNA and/or increased stability of the mRNA (Hollander-Cytko et al., 1988).

1.10.2 Overproduction of E5 by genetic engineering

The cDNA clone of the P. hybrida pre-E5 and the genomic clone of the A. thaliana E5 gene were separately cloned behind a high expression plant promoter (cauliflower mosaic virus 35S promoter) and their constucts were introduced, using Agrobacterium mediated transformation, into P. hybrida and A. thaliana respectively. This resulted in over expression of E5 in both plants and conferred the ability to grow on otherwise lethal concentrations of glyphosate (Shah et al., 1986; Klee et al., 1987). Although the herbicide treated plants which overproduced E5 grew to maturity, their growth rate in

the presence of glyphosate was reduced compared to untreated plants in the absence of glyphosate. Therefore the extent of overproduction of E5 was insufficient to confer complete resistance to glyphosate (Kishore and Shah, 1988).

1.10.3 Glyphosate resistant mutant E5 enzymes

Glyphosate resistant E5 enzymes have been isolated from Salmonella typhimurium (Stalker et al., 1985) and E. coli (Kishore et al., 1986). The former mutant E5 is the most characterised enzyme, with a single point mutation of Pro-101 to Ser being responsible for the resistance to the herbicide. The Pro-101 residue may be involved at the binding site of glyphosate (Kishore and Shah, 1988). The mutant S. typhimurium E5 was cloned between a high expression plant promoter and polyadenylation signal (from the octopine synthetase gene of Agrobacterium) and transformed into tobacco. Northern blot analysis of RNA from the transgenic tobacco indicated expression of the bacterial gene. Expression of the bacterial gene gave the plants moderate resistance to glyphosate (Comai et al., 1985). In view of the evidence that the shikimate pathway is predominantly located in the chloroplast, it is surprising that this bacterial enzyme provided any glyphosate tolerance at all because the mutant enzyme was not targetted to the tobacco chloroplast. This indicates that substrate exchange across the chloroplast membrane may be occuring, allowing the mutant enzyme to catalyse the conversion of S3P to EPSP in the presence of glyphosate. Alternatively it may lend weight to the hypothesis of a cytosolic shikimate pathway (Shah et al., 1986).

Both of the mutant bacterial E5 enzymes have been targetted to the chloroplast by linking the genes for these enzymes to genes encoding transit peptides and transforming the constructs into plants using Agrobacterium. The E. coli mutant E5 gene was linked to the P. hybrida transit peptide and upstream sequences (Della-Cioppa et al., 1987) whereas the S. typhimurium mutant gene was linked to the transit peptide sequence from various plant sources (Comai et al., 1988). The regenerated plants were found to be highly glyphosate tolerant indicating that mutant bacterial enzymes could be expressed in plants and would allow the transgenic plant to overcome inhibition by the herbicide.

The bacterial gene may not be expressed as efficiently within the plant as would a mutant heterologous or homologous plant gene. A mutant plant enzyme is likely to provide better glyphosate tolerance than the mutant bacterial enzyme. Mutations have been introduced into the plant E5 genes by site directed mutageneses producing tolerant enzymes (Kishore and Shah, 1988). The molecular basis for this tolerance has however not been revealed.

1.11 Effects of glyphosate on EO

The primary metabolic target for glyphosate is E5 but it has been reported that EO is also inhibited by the herbicide (Rubin et al., 1982; Ganson and Jensen, 1988). Inhibition of N. silvestris EO by glyphosate only occurs with the EO-Co isozyme and not with EO-Mn isozyme (see section 1.3.3). However glyphosate inhibition of the

EO-Co isozyme can be relieved by replacement of Co in the assay by Mg or Mn as the divalent-metal activator of EO-Co. This strongly suggests that the glyphosate inhibition of EO-Co activity is due to complex formation between glyphosate and Co and not due to the interaction of glyphosate with the polypeptide (Ganson and Jensen, 1988).

Although glyphosate does not cause inhibition of purified EO from Solanum tuberosum (potato), sub-lethal concentrations of the herbicide stimulate an increase in EO levels in cultures of S. tuberosum cells (Pinto et al., 1988). This has been shown to be an increase in EO protein as opposed to an activation of enzyme activity. The authors explain this result as follows. Glyphosate inhibits plant growth by reducing the production of EPSP and thus the production of aromatic amino acids and a host of aromatic secondary metabolites. If one of the aromatic amino acids or a metabolic derivative were to serve as a signal for the regulation of carbon flux through the shikimate pathway, and if the rate-limiting step to control this flux was catalysed by EO, then a modulation of EO activity would be expected following the presence of glyphosate (Pinto et al., 1988). It remains to be seen what mechanism of glyphosate-induction of EO expression is involved.

1.12 Intriguing aspects of the shikimate pathway of plants

The observation that some of the shikimate pathway enzymes exist as cytosolic isozymes raises the question of how this distribution could

arise. Presumably precursors of the chloroplastic enzymes exist which have transit peptides to enable their translocation into the chloroplast. Perhaps their cytosolic counterparts lack such a peptide and cannot be translocated. It is possible that there are different genes encoding proteins with or without transit sequences. Alternatively the transcripts from a single gene could be differentially spliced to create mRNA molecules which contain or lack the nucleotide sequence for the transit peptide.

The expression of the genes encoding the components of the pathway is likely to be tightly coordinated. It has been shown that aromatic compounds accumulate in response to wounding (Shirras and Northcote, 1984), infection by pathogens (Ebel and Grisebach, 1988; Habereder *et al.*, 1989) and flowering (Conn, 1987). It is likely that the flux of carbon through the shikimate pathway is increased at these times, which may imply a surge in the expression of shikimate pathway genes. Certainly the level of E5 transcript increases during maturation of *P. hybrida* petals (Gasser *et al.*, 1988).

A detailed study on the enzymology of the plant shikimate pathway is severely restricted because of the extremely low levels at which they are present *in vivo*. The overexpression of *P. hybrida* E5 in *E. coli* (Padgette *et al.*, 1987) has made the analysis of this enzyme an easier prospect. Overexpression, by similar means, of the other plant enzymes would greatly facilitate their study.

In order to investigate the mechanism for the generation of isozymes,

the coordination of gene expression and to aid enzymological work the genes for the shikimate pathway enzymes from higher plants require to be cloned. For the coordination of expression studies it is necessary to obtain these clones from a single species of plant.

1.13 Aims of this project

The aims of this project were as follows:

1. To initiate studies into the above aspect of the shikimate pathway (section 1.12) by cloning one or several of the genes for the shikimate pathway enzymes from Pisum sativum.
2. To analyse the DNA sequence of the cloned genes with a view to (i) further elucidate the evolutionary history of the pathway and (ii) identify particularly conserved regions of the deduced amino acid sequence which point to functionally critical regions of the protein.

P. sativum was chosen for study because a great deal of enzymology had already been carried out on the pathway of this plant which could be of benefit in the development of cloning strategies. At the outset of the project, the cloning effort was to be focussed on the gene for the bifunctional E2/E3 enzyme. It was recognised, however, that if the other genes for the pathway enzymes proved to be more amenable then their isolation would be pursued.

1.14 The cloning strategy

There were several strategies which could have been employed to clone the desired genes.

(i) Complementation. The E. coli and S. cerevisiae genes for the shikimate pathway enzymes had been cloned by complementation of the appropriate E. coli auxotrophic mutants. For this to be successful in the cloning of the corresponding plant genes, the eukaryotic gene (or a cDNA copy) would require to be expressed as a functional enzyme within its prokaryotic host. This is difficult enough to achieve when the eukaryotic gene has already been cloned although it has recently been shown that the P. hybrida E5 gene can complement an E. coli aroA$^-$ auxotroph (Padgette et al., 1987). The prospect of achieving complementation with the plant genes seemed daunting and therefore this strategy was not pursued.

(ii) Antibody screening. A cDNA copy of a eukaryotic transcript can be expressed as a fusion protein with a large proportion of the β-galactosidase enzyme (in the vector λgt11). The eukaryotic portion of the fusion protein may be sufficiently antigenic to be detected by appropriate antibody probes (Young and Davis, 1983). However no antibodies to any of the plant shikimate pathway enzymes were available and the antibodies to the microbial enzymes did not cross react with the purified plant enzymes (A. Likidlilid, unpublished work). Therefore this strategy could not easily be attempted.

(iii) Microprotein sequencing. The N-terminal sequence of the purified shikimate pathway enzymes could be determined and oligonucleotides prepared to the deduced DNA sequence. This would require tens of micrograms of purified enzyme which would be difficult to obtain since the enzymes are notoriously difficult to purify. At the time it seemed to be the second most obvious choice of strategies.

(iv) Heterologous probing. The genes for the E. coli enzymes and the S. cerevisiae arom had been cloned and sequenced in Glasgow (see section 1.2). There was considerable homology between the bacterial and fungal enzymes (Duncan et al., 1987) and it was anticipated that this homology would extend to the plant enzymes. It seemed likely that the plant enzymes would be more homologous to the fungal arom than to the bacterial enzymes because both fungi and plants are eukaryotic. Therefore the chosen strategy was heterologous probing using microbial genes with particular emphasis on using fungal genes as probes.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

General chemicals of analytical reagent grade were obtained from BDH Limited, Poole, UK; Boehringer Mannheim, Lewes, UK; Formachem Limited, Strathaven, UK; FSA Laboratory Supplies, Loughborough, UK; Koch-Light Limited, Haverhill, UK; May and Baker, Dagenham, UK.

ATP, anti-foam A, ampicillin, tetracycline, chloramphenicol, m-cresol, DTT, ethidium bromide, dextran sulphate, ficoll, MES buffer, MOPS buffer, Na salycilate, polyvinyl pyrrolidone, 8-hydroxyquinoline and tri-iso-propylnapthalenesulfonic acid were obtained from Sigma Chemical Co., Poole, UK. 4-aminosalycilic acid was obtained from Kodak, Liverpool, UK.

Bactotryptone, yeast extract and Bactoagar (agar) were obtained from Difco, Detroit, USA. Oxoid No. 1 agar was obtained from Oxoid Ltd, London, UK.

Amberlite MB3, acrylamide, NN'methylene bisacrylamide, bromophenol blue, caesium chloride, conc. hydrochloric acid, 30% hydrogen peroxide, 2-mercapto-ethanol, polyethylene glycol 6000, N, N, N', N'-tetraethylene diamine (TEMED), SDS and xylene cyanol were obtained from BDH Chemicals, Poole, UK.

Deionised formamide was obtained from Fluka, Glossop, UK. Agarose, low

melting point (l.m.p) agarose, Isopropyl-β-D-thiogalactoside (IPTG), phenol (ultrapure), and 5-bromo-4-chloro-3-indolyl-β-galactoside (X-gal) were obtained from BRL, Gibco Ltd, Paisley, UK. Glycogen (ultrapure) was obtained from Boehringer Corp., Lewes, UK, Sephadex-G50 (medium grade) and hexadeoxyribonucleotides (random primers) were obtained from Pharmacia (GB) Ltd., London, UK. Phenol was obtained from Formachem Limited, Strathaven, UK, and it was redistilled and stored in small aliquots at -20°C before use.

All solutions were prepared with glass-distilled water. Solutions for use in conjunction with enzymes and nucleic acids were prepared from glass-distilled, deionised water and were sterilised by autoclaving at 15 p.s.i.

Radiochemicals, namely [α-^{35}S]-dATP S (code SJ.304), [α-^{32}P]-dCTP (code PB.10205), [α-^{32}P]-dGTP (code PB.10206), ^{35}S-methionine (code SJ.204) and ^{14}C-methylated protein markers (MW 14,300 - 200,000; code CFA.626) were obtained from Amersham International plc., Amersham, UK.

Oligonucleotides were synthesised on an Applied Biosystems DNA synthesiser using phosphotriester chemistry courtesy of Dr V. Math (BIOMAC).

2.1.2 Enzymes and Proteins

Alkaline phosphatase from calf intestine and RNase-free DNase I (deoxyribonuclease I from bovine pancrease free from ribonuclease

activity) were obtained from Boehringer Corp., Lewes, UK.

Bovine serum albumin (BSA), DNase I, ribonuclease A (RNase A), lysozyme, proteinase K and zymolase (lyticase) were obtained from Sigma Chemical Co., Poole, UK.

Bacteriophage T4 DNA ligase, nuclease-free BSA, and all restriction enzymes were obtained from BRL, Gibco Ltd, Paisley, UK.

The Klenow fragment of E. coli DNA polymerase I was obtained from Amersham International plc., Amersham, UK.

2.2 Bacterial strains, plasmids, phage and phage libraries

The bacterial strains and the plasmids used during this study are shown in Tables 2.1 and 2.2 respectively.

λc857S7 DNA was obtained from BRL,Gibco Ltd, Paisley, UK. This DNA was digested with appropriate restriction enzymes and used exclusively as MW size markers. From this point onward λc857S7 DNA is referred to simply as λ-DNA.

Purified λL47/BamHI arms were a generous gift from Dr. M.A. Anderson.

The Arabidopsis thaliana genomic library was constructed by Dr. G. Gheysen and was screened in the Laboratorium Genetica, University of Gent, Belgium.

Table 2.1 Bacterial strains used during this study

strain	genotype	origin/reference
E. coli K12	wild type ATCC 14948, F^-, λ lysogenic	American Type Culture Collection (Rockvill, Maryland, U.S.A.)
E. coli HB101	F^-, *pro*, *leu*, *thi*, *lacY* *hsdR*, *endA*, *recA*, *rpsL20*, *ara-14*, *galK2*, *xyl5*, *mtl1*, *supE44*	Bolivar and Backman, 1979
E. coli TG1	Δ*lac pro*, *thi*, *SupE* [F' *traD36*, *proAB*, $lacI^qZ$ M15], *hsd*Δ5 (Ecok r^-m^-)	T. Gibson (unpublished)
E. coli JM109	*recAI*, *end AI*, *gyrA96*, *thi*, *hsd R17*, *supE44*, *relAI*, λ-, Δ*lac pro*, [F^-, *traD36*, *proAB*, $lacI^qZ$ ΔM15]	Yanisch-Perron *et al.*, 1985
E. coli LE392	F^-, *hsdR514* (Ecok r^-m^-), *supE44*, *supF58*, *lacYI* or (lacIZY)6, *galK2*, *galT22*, *metBI*, *trpR55*, λ-	Murray *et al.*, (1977)
E. coli C600	F^-, *thi-1*, *thr-1*, *leuB6*, *laeYI*, *tonA21*, *SupE44*, λ -	Appleyard, R.K. (1954)
E. coli C600 hfl	As C600 but also hfl^+	Stratagene
E. coli K803	*hsd* (Ecok r^-m^-) *gal*, *met*, *SupE*	Wood, W.B. (1966)
E. coli L87	*SupE*, *SupF*, *hsdR* (Ecok r^- m^-) *trp r*, *metD*, *tonA*	Amersham
E. coli NM154	*hsdR514* (Ecok r^-m^-) *ArgH*, *galE*, *galX*, *strA*, *lycB7* (hfl^+)	Amersham

Table 2.2 Plasmids used during this study

Plasmid	Markers	Source
pIA301	*amp*^{r}, *aroE*$^+$	Anton *et al.*, 1988
pKD201	*amp*^{r}, *aroD*$^+$	Duncan *et al.*, 1986
pGM602	*amp*^{r}, *aroC*$^+$	White *et al.*, 1988
pJB14	*amp*^{r}, *aroB*$^+$	Frost *et al.*, 1984
pME173	*amp*^{r}, *ARO1*$^+$	Duncan *et al.*, 1987
pMON9556	*amp*^{r}, (*P. hybrida* E5)$^+$	Gasser *et al.*, 1988
pSSU161	*amp*^{r}, (*P. sativum* rSSU)$^+$	Bedbrook *et al.*, 1980 (from Dr G I Jenkins)
pFab31	*amp*^{r}, (*P. sativum* cab)$^+$	Bennett *et al.*, 1984 (from Dr G I Jenkins)
pUC18	*amp*^{r}	Yanisch-Perron *et al.*, 1985
pUC19	*amp*^{r}	" " " " "

2.3 Growth media

Unless otherwise stated all media were sterilised by autoclaving at 15 p.s.i.

2.3.1 Rich media

The rich media used during the course of this work are listed in Table 2.3.

20% (w/v) glucose was autoclaved separately (5 p.s.i. for 50 minutes) and added where appropriate to prepared media. When L-broth (LB) was to be used for the titration of λ-phage, the glucose was replaced by 0.2% (w/v) maltose (final) along with 10 mM $MgSO_4$.

2.3.2 Minimal media

The minimal media used was based on the M9 salts medium and contained per litre of distilled water:

1g	NH_4Cl
0.13g	$MgSO_4.7\ H\ O$
3g	KH_2PO_4
6g	Na_2HPO_4

After autoclaving, vitamin B1 (thiamine hydrochloride; final concentration 2μg/ml), $CaCl_2$ (final concentration 0.1 mM), and glucose

Table 2.3 Rich Growth Media

Medium	Composition, per litre
L-broth (LB)	10 g bactotryptone 5 g yeast extract 10 g NaCl (+ 5 ml 20% glucose)
L-agar	as LB, + 15 g agar
λ-phage bottom agar	as LB, + 12 g agar (+10 ml 20% maltose) (+ 10ml 1M $MgSO_4$)
λ-phage top agar	as LB, + 7 g agar (+10 ml 20% maltose) (+10 ml 1M $MgSO_4$)
2 x TY	16 g bactotryptone 10 g yeast extract 5 g NaCl
H-agar	10 g bactotryptone 8 g NaCl 12 g agar
H-top agar	10 g bactotryptone 8 g NaCl 8 g agar
YPD	10 g yeast extract 20 g bactopeptone (+100 ml 20% glucose)

(final concentration 2 g/l) was added. The $CaCl_2$ and glucose were autoclaved together prior to the addition to the M9 salts while the vitamin B1 was filter sterilised by passage through a 0.22 μm filter (millipore).

For solid minimal media, the salts were autoclaved separately from the agar to avoid precipitation problems. Hence, the M9 salts were made up as a 2 times concentrated stock solution and autoclaved separately before being mixed with the agar solution. Oxoid No. 1 agar was used at 12.5 g/l (final). After mixing the two halves and cooling to 55°C, the various supplements were made.

When used to support growth of E. coli HB101 the above salts were supplemented prior to autoclaving with L-leucine (100 mg/l final) and L-proline (150 mg/l final). However for the growth of TG1 and JM109, the proline (and leucine) were omitted because these strains contain an unstable F'episome carrying the pro^+ marker in a pro^- background (Table 2.1) and the maintenance of the episome is selected on a true minimal medium.

2.3.4 Antibiotic supplements

Ampicillin (Amp) was used at a final concentration of 50 μg/ml. A stock solution of 25 mg/ml was filter sterilised and stored at -20 °C. Hot L-agar was cooled to 55°C before ampicillin was added. L-Amp plates were stable for at least 4 weeks if stored at 4°C.

Tetracycline (Tet) was used at a final concentration of 20 μg/ml. A stock solution of 10 mg/ml in ethanol was stored in the dark at -20°C. Hot agar was cooled to 55°C before tetracycline was added. L-Tet plates were used on the day of preparation.

Chloramphenicol, for plasmid amplification was used at a final concentration of 170 μg/ml. A stock of 34 mg/ml in ethanol was stored in the dark at -20°C

2.4 General Methods

2.4.1 pH measurement

The pH of solutions was measured using a Radiometer pH meter and combination electrode at room temperature.

2.4.2 Spectrophotometric determination of nucleic acid concentrations

Nucleic acid concentrations were determined spectrophotometrically at 260 nm, (Maniatus et al., 1982). In a 1 cm path length quartz cuvette an A_{260} value of 1.0 corresponds to 50 μg/ml for double stranded DNA, 40 μg/ml for RNA and 20 μg/ml for oligonucleotides.

2.4.3 General Microbiological techniques

(a) Bacterial growth

Bacterial growth was monitored at 600 nm when it was necessary to use cells at a particular stage of growth. Cell growth was at 37°C, shaking when liquid medium was used, and antibiotics were present if selection was desired.

(b) Bacterial strain storage

Several methods were employed for the preservation of strains. Long term storage was as a 40% (v/v) glycerol/LB solution at -80°C or as a stab (L-agar) at room temperature. Short term storage was on a suitable sealed plate at 4°C or as a 10 ml LB culture at 4°C. Long term storage of plasmids was usually in a *recA*$^-$ background eg *E. coli* JM109 or *E. coli* HB101.

2.5 Large scale preparation of plasmid DNA

The alkali lysis method of Birnboim & Doly,(1979), as modified by Maniatis *et al.*, (1982), was used routinely to prepare milligram quantities of pure plasmid DNA.

(a) 25 ml of LB medium, containing the appropriate antibiotic, was inoculated with 0.1 ml of an overnight culture of plasmid transformed cells. This culture was incubated at 37°C with vigorous shaking until the A_{600} reached 0.65 (late logarithmic phase). 500 ml of prewarmed LB medium (+ antibiotic) was inoculated with the entire late log phase culture (25 ml) and incubated at 37°C for exactly 2.5 hours. Chloramphenicol was then added to a final concentration of 170 μg/ml

and incubation was continued for a further 16 hours.

(b) The bacterial cells were harvested by centrifugation (MSE 18, 5000 rpm, 10 minutes, 4°C). The cell pellet was washed once with 100 ml ice-cold STE (0.1M-NaCl, 10mM-Tris/HCl, pH 7.5, 1mM-EDTA) and resuspended in 9 ml of 50mM-glucose, 25mM-Tris/HCl, pH8.0, 10mM-EDTA. Freshly prepared lysozyme, 1 ml of a 50 mg/ml solution, in the above glucose-containing buffer, was added and the suspension left at 25°C for 5 minutes. 20 ml of freshly prepared 0.2M-NaOH, 1% (w/v) SDS was added, the suspension was inverted gently to mix the contents and left on ice for 10 minutes. Lysis was achieved by the addition of 15 ml of 5M-potassium acetate pH 4.8, followed by mixing and incubation on ice for 10 minutes. Bacterial cell debris was removed by centrifugation (Beckman SW27 at 20,000 rpm, 30 minutes, 4°C) and nucleic acids precipitated from the resultant supernatant by addition of 0.6 volumes of isopropanol. After 15 minutes at 25°C the DNA was recovered by centrifugation (Sorvall HB-4, 12,000 rpm, 30 minutes, 25°C). The pellet was washed with 70% (v/v) ethanol, dried briefly under vacuum and resuspended in 4 mls of TE buffer. RNAse A (pretreated at 100°C for 10 minutes) was added to a final concentration of 100 μg/ml and the solution incubated at 37°C for 30 minutes.

(c) The volume of the RNAse-treated DNA solution was increased to 8 ml. Exactly 1g of solid CsCl was added for every millilitre of solution. The volume was remeasured and exactly 0.8 ml of a solution of ethidium bromide (10 mg/ml in H_2O) added for every 10 ml of CsCl solution. The final density of the solution was 1.55 g/ml. The

CsCl-ethidium bromide solution was transferred into two 10 ml polycarbonate tubes suitable for a Beckman Ti 70.1 rotor and centrifuged for at least 36 hours (Beckman Ti 70.1, 45,000 rpm, 20°C).

Only long-wave u.v. light was used to visualise the plasmid DNA band which migrated ahead of the host chromosomal DNA. Plasmid DNA was removed by inserting a needle into the top of the centrifuge tube and by pumping from below the desired DNA band. Ethidium bromide was removed by several extractions with TE saturated 1-butanol. The aqueous DNA solution was concentrated by extraction with non-saturated 2-butanol until the volume was approximately 1 ml. The solution was finally dialysed against multiple changes of 3 litres of TE (10 mM Tris/HCl pH8, 1 mM-EDTA) at 4°C. The DNA concentration was determined spectrophotometrically.

2.6 Small scale preparation of plasmid DNA

The method of Holmes and Quigley (1981) was used for small scale plasmid preparations. 1.5 ml of an overnight culture was harvested in a microfuge and the cell pellet resuspended in 350 µl of lysis buffer (10mM-Tris/HCl, pH 8.0, 50mM-EDTA, 8% (w/v) sucrose, 0.5% (v/v) Triton X-100). Lysozyme was added (25 µl of a 10 mg/ml solution) and the mixture boiled for 45 seconds followed by centrifugation in a microfuge. The supernatant was transferred to a fresh microfuge tube and plasmid DNA precipitated by addition of an equal volume of isopropanol. After chilling at -20°C for 30 minutes, plasmid DNA was recovered by centrifugation (10 minutes). The plasmid DNA pellet was

resuspended in 30 µl TE_2 (10mM-Tris/HCl pH 7.6, 2mm-EDTA). A 2 µl aliquot was sufficient for a single restriction enzyme digest.

2.7 Isolation of bacteriophage-λ DNA

2.7.1 Determination of phage titre

(a) Preparation of plating bacteria

10 ml (or 50 ml) of LB (with 0.2% maltose) was inoculated with a single colony of the appropriate E.coli strain eg LE392, C600 hfl, NM514, and grown overnight on an orbital shaker. The cells were pelleted by centrifugation (MSE18 7,000 rpm, 3 minutes, 4°C) and resuspended in 0.3 volumes of sterile, ice cold 10 mM-$MgSO_4$. The cell suspension was diluted until A_{600}=2 ie 1.5×10^9 cells/ml. The cells were stored at 4°C and usually used the same day as they were prepared, although they remained viable for up to 1 week.

(b) Plating the phage

Serial 10-fold dilutions of λ-phage stock were prepared in SM phage buffer (100 mM-NaCl, 8mM-$MgSO_4$, 50mM-Tris-HCl pH7.5, 0.01% (w/v) gelatin). Bacteriophage-λ infection was achieved by adding 10 µl and 100 µl aliquots of each dilution to be assayed to 100 µl (1.5×10^8 cells) of plating bacteria suspension. The samples were incubated at room temperature (20°C) for 10 minutes and then at 37°C for 20 minutes. 3 ml of λ-phage top agar at a temperature of 45°C was added

and the mixture was poured gently onto plates containing λ-phage bottom agar. After allowing the top agar to harden for 10 minutes at room temperature they were incubated at 37°C overnight. The plaques were counted and the titre determined for each dilution assayed.

2.7.2 Isolation of a single plaque

λ-phage were plated out at low density eg 100-200 plaques per 9 cm plate (see section 2.7.1). A single plaque (which may have been identified by screening procedures, see section 2.21.4) was cored out of the top agar into 1 ml of SM buffer and 10 µl of chloroform was added to kill the cells. After two hours incubation at room temperature the agar was centrifuged out of the buffer (optional) and the resultant λ-phage stock was stored at 4°C.

2.7.3 Preparation of λ-phage liquid lysates

Approximately 5 x 10^8 cells, of prepared bacterial suspension were infected with 10^6 plaque forming units (pfu) as described in section 2.7.1. These infected cells were used to inoculate 50 ml of prewarmed LB (containing 0.2% maltose, 10mM-$MgSO_4$) in an indented 250 ml conicle flask. 1 µl of anti-foam A (30% w/v emulsion) was added and the flask and was shaken vigourously overnight. The resultant lysates were cooled to room temperature and several drops of chloroform were added. After allowing to stand for 10 minutes the lysate was centrifuged (MSE18, 10,000 rpm, 10 minutes, 20°C). To the supernatants, 10 µl of 1 mg/ml DNase I and 10 µl 10 mg/ml RNase A was added and this was left

at room temperature for 1 hour. 100 µl of chloroform was added to the λ-phage lysates and they were stored at 4°C. These lysates acted as phage stocks for storage or for use in the isolation of λ-phage DNA.

It was realised that the optimum multiplicity of phage infection was almost certainly not being achieved, however it was impracticle to carry out multiplicity tests since many lysates were usually prepared simultaneously.

2.7.4 DNA preparation from λ-phage lysates

120 µl of 2.5% (w/v) SDS, 0.5M-Tris/HCl pH9, 0.25M-EDTA was added to 0.6 ml of λ-phage lysate and the mixture was incubated at 70°C for 15 minutes. After cooling to room temperature, 150 ml of 8M-potassium acetate was added. After gentle mixing to generate a precipitate, the mixture was vortexed and incubated on ice for 15 minutes. After centrifugation in a microfuge for 1 minute 700 µl of supernatant was decanted into 300 µl of TE and this was phenol/chloroform extracted (section 2.13.1) twice. 0.6 volumes of isopropanol was added to the aqueous phase and following incubation for 2 minutes at room temperature the λ-phage DNA was recovered by centrifugation in a microfuge for 1 minute. The DNA pellet was washed with 70% ethanol, dried briefly in a dessicator and rehydrated in 25 µl of TE for 30 minutes. This DNA was suitable for restriction analysis and southern blotting.

If the DNA was to be used for sub-cloning, it was phenol/chloroform

extracted once more and ethanol precipitated. It was stored for short periods as an ethanol precipitate.

2.8 Preparation of E. coli K12 genomic DNA

DNA was prepared from E. coli K12 as follows: A single colony of E. coli K12 was inoculated into 100 ml LB and shaken at 37°C till the A_{600} reached 0.7. The cells were harvested, washed with 50 ml TE, then resuspended in 2 ml 0.15M-NaCl, 0.1M-EDTA. 0.2 ml lysozyme (10 mg/ml) was added and the tube incubated at room temperature for 15 min. It was then frozen to -70° (dry ice/methanol). 12.5 ml of 0.1M-NaCl, 0.1M-Tris pH 8.0, 1% SDS was added and the tube placed in a 65°C water bath till the frozen material thawed. The tube was again frozen to -70° and thawed at 65°C. The sample was kept at 65° for several days to dissolve the DNA; very gentle inversion of the tube at intervals assisted this process.

The volume was estimated and 1.282 g CsCl/ml DNA solution added. Using a refractometer, the concentration of the solution was adjusted till the refractive index was 1.39970 (1.70 g/ml). The solution was centrifuged to equilibrium (Prepspin 50, 10 x 10 ml Rotor, 50,000 rpm, 64 hours, 20°C) and the gradient fractionated by taking samples down the tube with a wide-bore pipette. DNA containing fractions were pooled and dialysed against TE buffer.

2.9 Preparation of S. cerevisiae S228C genomic DNA

5 ml of YPD medium was inoculated with S. cerevisiae S228C and grown at 30°C for 48 hours, shaking vigourously. 50 µl of this stationary phase culture was used to inoculate 100 ml of YPD which was grown at 30°C for 16 hours, till mid log phase. The cells were harvested (MSE 18, 5,000 rpm, 5 minutes, 20°C) and washed in 5 ml of 1M-sorbitol, 20mM-EDTA, 20 mM-KH_2PO_4 pH7.5 then resuspended in 0.2 ml of 0.15M-NaCl, 5mM-KCl, 1mM-EDTA, 0.2%-Triton X-100, 1M sorbital, 20mM-NaHepes pH7.5, 2mM-DTT (buffer 2). Zymolase was added (dissolved in buffer 2 and at a final concentration of 10 mg/ml) and digestion continued at 37°C for 15-30 minutes until greater than 80% of the cells were non-refractable ghosts or spheroplasts (as judged by microscopy). When digestion was complete the volume was increased to 5 ml by making the mixture 1% SDS, 0.1M- EDTA, 0.1M—NaCl. Proteinase K was added at the concentration of 250 µg/ml and digestion took place for 90 minutes at 37°C. The sample was phenol/chloroform extracted (section 2.13.1) three times followed by a single chloroform extraction. Preboiled RNaseA was added to 100 µg/ml and incubated at 50°C for 15 minutes.

Two volumes of ethanol were added and the stringy DNA was spooled onto a sealed pasteur pipette. The pellet was dried briefly in a dessicator, rehydrated in 5 ml TE and ethanol precipitated as before. The dried pellet was rehydrated in 200 µl TE.

2.10 Preparation of P. sativum genomic DNA

2.10.1 Growth of P. sativum tissue

Peas were obtained from Clause (UK) Ltd, Reading, UK.

Moderately dwarf P. sativum (variety ONWARD) were soaked overnight in water. The swollen peas were spread thickly over a layer of Fisons Levington compost in a seed tray and were covered with another layer of compost. The peas were watered and grown in 12 hours light, 12 hours dark for about 10 days. Tissue was removed 7-10 days after germination, frozen in liquid nitrogen and stored frozen at -80°C. This tissue was suitable for DNA or RNA preparation.

2.10.2 DNA preparation

2 g of frozen seedling tissue was kept frozen by its addition to liquid nitrogen. It was ground to a fine powder with a mortar and pestle and transferred to a 100 ml Prepspin tube. 45 ml of 100mM-Tris/HCl pH8, 50mM-EDTA, 500mM-NaCl, 10mM-β-mercaptoethanol was added followed by 3 ml 20% SDS. The sample was thoroughly mixed by shaking vigourously then incubated at 65°C for 10 minutes. 15 ml of 5M-potassium acetate was added, the solution was mixed thoroughly and incubated on ice for 20 minutes. The sample was centrifuged (Prepspin, 16,000 rpm, 20 minutes, 40°C) and the supernatant was decanted off through muslin, into 30 ml of isopropanol. This mixture was incubated at -20°C for 30 minutes. The DNA was pelleted

(Prepspin, 13,000 rpm, 15 mins, 4°C) and dried by inversion of the tube on towels. The DNA was redissolved in 2.1 ml of 50mM-Tris/HCl, 10mM-EDTA then ethanol precipitated. The DNA was redissolved in 300 µl of 10mM-Tris/HCl, 1mM-EDTA at 4°C. Storage was at 4°C.

2.11 Digestion of DNA with restriction enzymes

The methods used were as described in Maniatis *et al*, (1982).

2.11.1 Complete digestion

Four buffers were used initially to completely digest DNA:

Low salt: 10mM-Tris/HCl pH 7.5, 10mM-$MgCl_2$
1mM-DTT, 0.1 mg/ml BSA

Medium salt: 50mM-NaCl, 10mM-Tris/HCl, pH 7.5,
10mM-$MgCl_2$, 1mM-DTT, 0.1 mg/ml BSA

High salt: 100mM-NaCl, 10mM-Tris/HCl, pH 7.5,
10mM-$MgCl_2$, 1mM-DTT, 0.1 mg/ml BSA

*Sma*I salt: 20mM-KCl, 10mM-Tris/HCl, pH 8.0,
10mM-$MgCl_2$, 1mM-DTT, 0.1 mg/ml BSA

Each buffer was prepared as a 10x stock solution (without BSA), sterile (nuclease-free) BSA at 1 mg/ml was added to each digest to the

correct final concentration. Analytical digests were carried in a volume of 10 or 20 µl, usually at 37°C unless otherwise recommended. Preparative digests were carried out in larger volumes. When DNA was digested with two restriction enzymes, the endonuclease requiring the lower salt buffer was used first. After the recommended duration of digestion the salt concentration was adjusted and the second enzyme added.

Latterly, restriction digests were carried out using the BRL React buffers which were provided with each batch of enzymes. There are ten different React buffers with a range of salt concentrations, each one suitable for a range of enzymes.

2.11.2 Establishing conditions for partial Sau3A digestion of high MW DNA

A 300 µl reaction mixture containing 20 µg of P. sativum DNA in the appropriate digestion buffer was thoroughly mixed. 30 µl was dispensed into the first tube and 15 µl into another 11 tubes. 4 units of enzyme was added to tube 1 (thus concentration was 2 u/µg of DNA), thoroughly mixed and 15 µl of the mixture was transferred to tube 2. This was repeated, continuing the twofold serial dilution to tube 11. Nothing was added to tube 12. Digestion was at 37°C for 1 hour and the reactions were stopped by heat inactivating the Sau3A at 65°C for 10 minutes.

2.12 Agarose gel electrophoresis of DNA

DNA was separated at room temperature on horizontal submerged agarose gels as described by Maniatis et al., (1982). The buffer system used was the Tris-borate (TBE) version. For accurately sizing restriction fragments of 0.3 - 3 kbp a 1% (w/v) agarose gel was used. Increased resolutions of DNA fragments 3 kbp or 0.5 kbp was achieved on 0.8% (w/v) or 2% (w/v) agarose gels respectively. Samples for agarose gels were prepared by addition of 0.15 volumes of 10mM-Tris/HCl pH 7.2, 20% ficoll, 0.5% bromophenol blue and 10 mg/ml ethidium bromide. Ethidium bromide (0.5 μg/ml) was added to both gel and buffer, and stained DNA bands visualised on a long wave u.v. transilluminator (U.V. Products Inc.).

Low melting point (l.m.p.) agarose gels (see Section 2.13.2) were run in an identical fashion except at 4°C.

Known restriction fragment markers, most commonly HindIII-digested λ-DNA (and HindIII+EcoRI-digested λ-DNA) were run alongside the unknown fragments.

2.13 Extraction and purification of DNA samples

2.13.1 Phenol/chloroform extraction and ethanol precipitation

Restriction digests and other solutions containing DNA were deproteinised by phenol/chloroform extraction. The volume was

adjusted to 100 µl (usually) and an equal volume of TE-saturated phenol/chloroform (1:1 v/v) added. This was mixed by vortexing and centrifuged in a microfuge for 3-5 minutes. The upper aqueous phase was removed to a fresh microfuge tube and the process was repeated. Traces of phenol were finally removed by extraction with an equal volume of chloroform in an identical manner. The final aqueous DNA pool was made 0.3M- Na acetate (or 3.75M-ammonium acetate) and 2.5 volumes of ice-cold ethanol was added. The mixture was cooled at -20°C overnight or at -70°C for 1 hour and the DNA recovered by centrifugation (12,000 rpm, 25°C, 10 minutes) in a microfuge.

2.13.2 Recovery of DNA from l.m.p. agarose

DNA was visualised by ethidium bromide staining and low energy long-wave u.v. transillumination. The desired DNA band was exised, placed in a 1.5 ml microfuge tube and 200 µl 20mM-Tris/HCl, pH 7.5, 1mM-EDTA added. The gel slice was melted at 65°C for 10 minutes and phenol (not phenol/chloroform) extracted. The first round extraction was back-extracted with 200 µl 20mM-Tris/HCl, pH 7.5, 1mM-EDTA and the aqueous samples pooled. The DNA was further purified by two rounds of phenol extraction (not phenol/chloroform extraction) followed by chloroform extraction. DNA was recovered by ethanol precipitation as detailed above (Section 2.13.1). DNA purified in this way was sufficiently pure for cloning and *in vitro* labelling by either nick-translation or random priming.

2.14 Size fractionation of DNA by sucrose density gradient centrifugation

The size of tube used to fractionate DNA molecules depended on the amount of DNA to be applied to the gradient. 200 µg of DNA was fractionated on a 14 ml gradient (Beckman SW40) whereas 20 µg of DNA was fractionated on a 4 ml gradient (Beckman SW60).

Two identical sucrose gradients were prepared in appropriate Beckman polycarbonate ultracentrifuge tubes, by layering equal volumes of 40-10% sucrose (at 5% increments), 1M-NaCl, 20mM-Tris/HCl pH8, 5mM-EDTA. The DNA was applied to the top of the gradient which was centrifuged at the equivalent of 92,000 g for 18 hours. Equal fractions were removed from the top of the gradient using an LKB pump. The volume of each fraction was increased to 3.5 times the original volume (to prevent sucrose and salt precipitation) and the DNA was precipitated with 2.5 volumes of ethanol.

2.15 Calf intestinal phosphatase treatment of DNA

To prevent self-ligation, the terminal 5' phosphates of plasmid/bacteriophage vector DNA were removed by calf intestinal alkaline phosphatase as described in Maniatis *et al.* (1982).

Up to 10 µg of phenol/chloroform purified DNA was ethanol precipitated and resuspended in 50 µl of 0.05M-Tris/HCl, pH 9.0, 1mM-$MgCl_2$, 0.1mM-$ZnCl_2$. Calf intestinal phosphatase was added (0.05 units) and

the reaction mixture incubated at 37°C for 30 minutes. A second aliquot of calf intestinal phosphatase was added and incubation continued for 30 minutes. 10 µl of 10 x STE and 5 µl of 10% (w/v) SDS was added to the reaction which was increased in volume to 100 µl making it 10mM-Tris/HCl, pH 8.0, 0.1mM-NaCl, 1mM-EDTA, 0.5% (w/v) SDS. The enzyme was inactivated by heating at 65°C for 15 minutes. Finally the dephosphorylated DNA was deproteinised by repeated phenol/chloroform extraction and ethanol precipitation.

2.16 Ligations

The vector and insert DNA were separately digested with the appropriate enzymes, phenol/chloroform extracted and ethanol precipitated. If appropriate the vector DNA was treated with calf intestinal phosphatase (section 2.15).

300 ng of DNA (vector + insert) was included in a 30 µl ligation (or sometimes this was scaled down to 50 ng per 5 µl ligation) but the proportion of vector and insert depended on the size of the DNA species and the number of clonable fragments. Ligations were set up where the molar ratios of total clonable ends of vector: insert were 3:1 since this reduces the number of recombinants with multiple inserts.

The ligations were performed overnight using bacteriophage T4 DNA ligase in 66mM-Tris/HCl pH 7.6, 6.6mM-$MgCl_2$, 0.5mM-ATP, 10mM-DTT. The amount of ligase and the temperature of the reaction depended upon the

nature of the ends of the DNA fragments. For sticky ends 0.5 units at 15°C, for blunt ends 1.0 unit at 4°C.

2.17 Transformation of E. coli with plasmid DNA

2.17.1 Preparation of competent cells

A single colony of E. coli JM109 (or TG1) from a stock minimal media plate, was used to inoculate 10 ml of 2xTY which was incubated overnight at 37°C. 2 ml of this fresh culture was used to inoculate 40 ml 2xTY which was incubated at 37°C for 3 hours (A_{600}=0.3). The cells were cooled on ice for 10 minutes and gently harvested (MSE18 7,000 rpm, 2 minutes). The cells were resuspended in 20 ml ice-cold sterile 50mM-$CaCl_2$ and left on ice for 20 minutes. The fragile competent cells were gently harvested and resuspended in 4 ml ice cold 50mM-$CaCl_2$.

When E. coli HB101 were being made competant a single colony from an L-agar plate was used to inoculate LB. Otherwise the steps were the same.

2.17.2 Transformation of competent cells

Transformations were carried out in sterile 5 ml bijoux tubes. Ligation mix refers to any DNA being used in the transforming process and usually was a ligation mixture. The ligation mix was added to 300 µl aliquots of competent cells and the mixture was incubated on

ice for at least 40 minutes. The DNA/cell mix was then heat shocked at 42°C for 3 minutes and then the tubes were returned to the ice bath.

When E. coli HB101 was being transformed, the mixtures were spread onto L-Amp plates at this stage, then incubated overnight.

When E. coli JM109 (or TG1) were being transformed, 40 µl of 100mM IPTG and 40 µl of 2% X-Gal (w/v, in dimethylformamide) was added to the heat shocked DNA/cell mix. This mixure was then spread onto L-Amp plates and incubated overnight.

2.17.3 Selection of pUC derived recombinant clones

The pUC plasmids have been constructed as cloning vectors using β-galactosidase activity as the basis of selection. The vector has a fragment of the E. coli lac operon containing the regulatory region and coding information for the first 146 amino acids of the β-galactosidase (Z) gene. This amino terminal peptide is able to complement the product of a defective β-galactosidase gene present on the F' episome in the host cell (eg E. coli JM109 or TGI). A 'polylinker' DNA fragment containing several unique restriction sites for cloning has been inserted, in phase, into the amino-terminal portion of the β-galactosidase gene. This insertion does not affect the complementation. However, insertion of additional DNA into the 'polylinker' region generally destroys the complementation.

The complementation produces active β-galactosidase which cleaves the chromogenic substrate X-Gal to produce a blue chromophore, when transformed cells are grown in the presence of the non-metabolizable inducer IPTG. However in recombinant plasmids the ability for complementation is lost since the enzyme is inactive and therefore the colonies appear white. False positive white colonies appear at low frequency, probably due to incorrect self ligation of the vector (Yanisch-Perron et al., 1985).

2.18 DNA transfer to nitrocellulose (Southern Blotting)

This method is based on that of Southern (1975). After electrophoresis the agarose gel was placed in 250 ml of denaturing solution (1.5M-NaCl, 0.5M-NaOH) and shaken for 30 minutes. The gel was then placed in 250 ml of neutralising solution (3M-NaCl, 0.5M-Tris/HCl pH 7.4) and shaken for a further 30 minutes. A depurination step was carried out before denaturation when the DNA being transferred was of high MW, especially in blots of genomic DNA. This step consisted of shaking the gel for 30 minutes in 0.25M-HCl. The DNA was blotted onto a nitrocellulose (Schleicher and Schuell) or nylon (Hybond-N, Amersham) membrane using 20 x SSC (SSC = 0.15M-NaCl, 15mM-Na citrate pH 7.3) as described in detail in Maniatus et al., (1982). The filters were then baked at 80°C for 2 hours (under vaccum for nitrocellulose).

2.19 ^{32}P labelling of DNA

Fragments of DNA less than 800 bp were labelled by nick translation whereas fragments greater than 800 bp in size were labelled by random priming.

2.19.1 Nick-translation

Purified DNA samples (50-100 ng in TE) were radioactively labelled *in vitro* using the Amersham Nick Translation Kit (Catalogue No. N.5000). The nick translation was carried out in a final volume of 50 µl containing 50mM-Tris/HCl pH 7.8, 10mM-MgCl , 0.1mM-DTT, 50 µg/ml BSA, 2.5 u DNA polymerase I, 50 pg DNase I, 20 µM each dATP, dGTP, dTTP and 50 µg of $[\alpha^{32}P]$-dCTP. The reaction was carried out at 15°C for 3 hours when it was stopped by the addition of EDTA to 20mM. Unincorporated label was removed using gel filtration (section 2.19.4). Nick translation provided labelled DNA with specific activity of 4 x 10^8 cpm/µg of DNA.

2.19.2 Random priming

This method is based on that of Feinberg and Vogelstein, (1983). 50-100 ng of purified DNA was present in approximately 30 µl of TE. The DNA was completely denatured by placing the tube in a boiling water bath for 5 minutes followed by rapid cooling on ice. The labelling reaction was carried out overnight at room temperature in a volume of 50 µl containing 200 mM-Na Hepes pH 6.6, 5mM-$MgCl_2$, 400

μg/ml BSA, 20 μM each of dATP, dGTP, dTTP, 100 μg/ml hexadeoxyribonucleotides (random primers), 30 μCi of [α^{32}P]- dCTP and 2-3 units of Klenow fragment of E. coli DNA polymerase I. The reaction was stopped by the addition of EDTA to 20 mM and unincorporated label was removed using gel filtration (section 2.19.4). Random priming provided labelled DNA with specific activity of 7 x 10 cpm/ μg of DNA.

2.19.3 Labelling flush end DNA with the Klenow fragment

1 μg of DNA was incubated at 37°C for 30 minutes in a volume of 10 μl containing 10mM-Tris/HCl pH 7.4, 1 μCi of the (appropriate) [α^{32}P]- dNTP and 1 u Klenow fragment. The enzyme was inactivated at 70°C for 10 minutes, then the reaction was phenol/chloroform extracted and ethanol precipitated (twice) in order to purify the DNA and remove unincorporated label. The DNA was resuspended in 5 μl of TE.

2.19.4 Removal of unincorporated radionucleotide

Unincorporated label from nick translation and random priming reactions was removed by gel filtration chromatography on a 10 cm x 1 cm Sephadex G-50 column. Sephadex G-50 was hydrated in 0.1 x SSC for several hours and stored at 4°C. The poured columns were equilibrated in 0.1 x SSC, 0.1% SDS at room temperature and the reaction mixture was loaded directly on top of the column. The radioctivity was monitored as it passed down the column and as it approached the bottom, fractions were collected manually. The labelled DNA was

eluted first, followed by a trough then a second peak of unincorporated label. A sample from each fraction was counted by liquid scintillation using 2 ml ECOSCINT (National, Diagnostics). The fractions containing the largest number of counts were pooled.

2.20 Hybridisation of filter bound nucleic acid

The temperature of hybridisation and the salt concentration and temperature of washing solutions was dependent on the particular experiment being carried out and exact details are given in the text. In general however, the filters were pre-wetted in 5 x SSPE (SSPE = 0.18M-NaCl, 10mM-Na H_2PO pH 7.4, 1mM EDTA). The pre-wetted filters (except screen filters, see 2.21.4) were placed into a polythene bag along with 1 ml/30 cm^2 (surface area of filter) of prehybridisation solution containing 5 x SSPE, 5 x Denhardt's solution (Denhardt's = 0.02% Ficoll, 0.02% polyvinyl-pyrolidine, 0.02% BSA, stored at -20°C), 0.1% SDS, 50 µg/ml denatured salmon sperm DNA. Deionised formamide was present at the appropriate concentration usually 50% (v/v) if required. The bag was prehybridised for at least 4 hours at the appropriate temperature.

The DNA probe was denatured by placing the tube containing the probe in a water bath at 100°C for 5 minutes followed by immediate cooling on ice.

After prehybridisation the bag was opened and the denatured probe was added at a concentration of 2 x 10^6 cpm/ml of solution. The bag was

resealed and the hybridisation carried out overnight in a shaking waterbath.

The hybridised filters were washed firstly at room temperature in 5 x SSC for 5 x 1 minute, followed by the appropriate washing conditions with several changes, for 2 hours. After washing, the filters were autoradiographed damp (under Saran wrap) and exposed to Fuji RX film using intensifying screens at -70°C. Films were developed by a Kodak X-OMAT processor.

2.21 Manipulation of bacteriophage-λ

2.21.1 In vitro packaging

(a) 'Homemade' packaging extracts

In vitro packaging extracts were termed 'homemade' when they had been prepared using the method of Maniatus et al., (1982). Such extracts were a generous gift from Dr M A M Anderson. The bacterial strains used were E. coli BHB 2688 [N205, recA$^-$(λimm434, cIts, b2, red3, Eam4, Sam7/λ] which accumulates heads - Extract A - and E. coli BHB 2690 [N205,recA$^-$,(imm434, cIts, b2, red3, Dam15, Sam7/λ] which accumulates packaging proteins - Extract B.

DNA to be packaged (usually from a ligation reaction) was ethanol precipitated and up to 1 μg was dissolved in 66mM-Tris/HCl pH 7.9, 10mM-$MgCl_2$. 15 μl aliquots of extract A and extract B were removed

from -70°C and allowed to thaw on ice. When just thawed the DNA to be packaged was added to extract A, quickly followed by 10 µl of extract B. After gentle mixing the reactions were incubated at room temperature for 1 hour. 1 ml of SM buffer was added to the reaction which was stored at 4°C. This was now suitable for titration (see section 2.7.1).

(b) Commercial in vitro packaging extracts

High efficiency in vitro packaging extracts were obtained from Amersham, catalogue No. N.334 (used in section 3.5.5 and 5.4.4) and Strategene, catalogue No. GP3-P (used in section 3.6.4) in the form of in vitro packaging kits. The methods given in their respective data sheets were very similar to those outlined above and they were strictly adhered to.

2.21.2 Precipitation of bacteriophage-λ with PEG

This method is based on one given in Yamamoto et al., (1970). The desired packaged phage, in SM buffer, are pooled and one tenth volume of cold 5M-NaCl was added. PEG-6000 was crushed to a fine powder in a mortar and pestle and this was added to a concentration of 110 mg/ml. The PEG was dissolved gently and the mixture was incubated on ice for at least 3 hours. After centrifugation (MSE18, 6,000 rpm, 15 minutes, 4°C) the pellet was resuspended in 0.5 ml of SM buffer.

2.21.3 Amplification of the P. sativum genomic library

5 ml of bacterial cell suspension (section 2.7.1) was infected with all of the PEG precipitated phage (ie 1.1 x 10^6 pfu) as described in section 2.7.1. The phage infected cells were plated out onto 23 cm x 23 cm petri dishes at a density of 10^5 pfu per plate using 35 ml 0.5% L-agar (0.2% maltose, 10mM-$MgSO_4$) per plate. These were incubated for 10 hours. The very sloppy top layers were removed into 250 ml centrifuge buckets. The plates were washed with 10 ml of SM buffer which was combined with the top agar. A few drops of chloroform were added to each bucket and after standing for 20 minutes, the buckets were centrifuged (MSE18, 6,000 rpm, 20 minutes, 4°C). The supernatant was removed and stored with 10 µl/ml chloroform, at 4°C

2.21.4 Screening a phage-λ library

(a) First round screening

Cells from a prepared bacterial suspension were infected (section 2.7.1) with phage from the appropriate bacteriophage-λ library at a ratio of 10^5 pfu/10^9 cells. 10^5 pfu were plated onto 23 x 23 cm petri dishes using 50 ml of 0.7% top agarose (0.2% maltose, 10 mM-$MgSO_4$). Agarose was used when plates were to be used for screening because it is more rigid than agar. LB was not present in the top agar since subsequent impressions from the plates became dirty and this trapped probe leading to a high background signal. The plates were incubated overnight at 37°C, then transferred to 4°C for at least 2 hrs. Cells

were killed by placing plates over Whatman 3MM paper soaked in chloroform for 5 minutes.

Up to six impressions could be taken from the one plate onto nitrocellulose or nylon membranes. Hybond-N (Amersham) nylon membrane was found to be most suitable. Alignment marks were made on the plate and the filter was removed and layered onto trays containing firstly 0.5M-NaOH, 1.5M-NaCl for 5 minutes secondly 0.5M Tris/HCl pH 7.7 for 10 seconds and thirdly 3M-NaCl, 0.5, Tris/HCl pH 7.4 for 5 minutes. The filter was allowed to air dry and was then baked at 80°C (under vacuum when nitrocellulose was used).

The filter was hybridised with a nucleic acid probe (section 2.19). The large filters were hybridised in large plastic boxes in a minimum of 100 ml hybridisation solution. Up to fourteen filters could be hybridised in 200 ml.

(b) Rescreening

Desired plaque regions or single plaques were isolated as described in section 2.7.2. These phage stocks were plated out at lower density on small (9cm diameter) petri dishes. Impressions were taken (2.21.4a) and these were hybridised as discussed in section 2.20. Usually many small filters were being screened therefore they were hybridised in plastic containers. This rescreening procedure was repeated at least once more until the hybridising plaque was purified to homogeneity.

2.22 Isolation of P. sativum mRNA

2.22.1 RNase free technique

The successful preparation of high molecular weight, biologically active RNA depends on the rapid inactivation of nucleases in the early steps of the isolation and the prevention of introduction of nucleases from glassware, plasticware and solutions. Gloves were therefore worn at all times when RNA was being dealt with. Glassware was baked at 200°C overnight before use. Plasticware was used from sterile unopened containers. Solutions were prepared from chemicals which were kept out of general circulation and all were autoclaved after preparation. In addition, solutions which did not contain chemicals with primary or secondary amines (Leonard et al., 1970) were treated with 0.1% (v/v) diethylpyrocarbonate (DEPC) shaken vigourously and allowed to stand overnight before autoclaving. Rubber tubing, quartz cuvettes, stirring bars were also treated with DEPC.

RNA samples were stored under ethanol at -80°C until they were required.

2.22.2 Isolation of P. sativum total RNA

The following RNA preparation provided 1 mg of total RNA but it was routinely scaled up to provide up to 10 times this amount. Pea tissue was obtained and stored as described in section 2.10.1 8-10 pea seedlings were transferred into liquid nitrogen and ground to a fine

powder using a mortar and pestle. The frozen powder was transferred into pre-cooled 30 ml corex tubes containing 10 ml freshly prepared Kirby reagent (10mM-Tris/HCl pH 7.6, 50mM-KCl, 6% (w/v) 4-aminosalycilic acid, 1% (w/v) tri-iso-propylnapthalenesulfonic acid, 8% phenol reagent) and 10 ml TE saturated phenol reagent (500 g phenol, 70 ml m-cresol (redistilled), 0.58 g 8-hydroxyquinoline). After vortexing, the mixture was centrifuged (Sorval HB-4 3,500 rpm, 7 minutes, 4°C). The aqueous layer was removed and added to 5 ml phenol reagent and 5 ml chloroform. This was extracted as before. The aqueous phase was removed and the nucleic acids were precipitated by adding 2.5 volumes of (-20°C) ethanol. The mixture was incubated at -20°C for at least 2 hours then the ethanol precipitate was pelleted by centrifugation (Sorval HB-4, 10,000 rpm, 10 minutes, 4°C). The pellet was washed in cold 70% ethanol, dried for 30 seconds in a dessicator and redissolved in 2 ml of DNaseI buffer (50 mM MES buffer, 2mM Mg acetate, pH7). After determining nucleic acid concentration by measuring A_{260}, 2 units of DNase I (RNase free) per 100 µg nucleic acid was added prior to incubation on ice for 30 minutes (samples were removed before and after digestion and examined by agarose gel electrophoresis). The digestion was extracted with phenol/chloroform, ethanol precipitated and stored under ethanol at -80°C.

2.22.3 Isolation of P. sativum poly A^+ mRNA

0.1-0.2 g of oligo-dT cellulose was equilibrated in loading buffer (20 mM-Tris/HCl pH 7.6, 0.5M-NaCl, 1mM-EDTA) in a sterile vessel. An oligo-dT column was poured in a 5 ml plastic syringe with siliconised,

baked, glass-wool as a stopper. A 100 µl capillary tube was passed through the rubber seal of the syringe and was connected to rubber tubing. The column was equilibrated by pumping loading buffer through at 0.5 ml/minute. The column was washed with H_2O (10 volumes), treated with 0.1M-NaOH, 5mM-EDTA and rewashed with H_2O until the pH was <pH8.0. The column was then washed with loading buffer (10 volumes).

Total RNA was ethanol precipitated and redissolved in 500 µl H_2O. A small sample was removed for later use in determining RNA concentration and the remainder was heated to 65°C for 5 minutes. After quick cooling on ice 500 µl of 2x loading buffer was added and the RNA solution was applied to the top of the column using a needle and syringe. The flow through was collected, reheated to 65°C and reapplied to the column. The column was washed with loading buffer until the A_{260} of the column eluant was zero. The column was then washed with 5 column volumes of washing buffer (20mM-Tris/HCl pH 7.6) or until the A_{260} equalled zero. The poly A^+ mRNA was eluted from the column with elution buffer (10mM-Tris/HCl pH7.5, 1mM-EDTA). 1ml fractions were collected and the A_{260} of these fractions was determined (using DEPC treated quartz cuvettes). The poly A^+ mRNA was then immediately ethanol precipitated and stored at -80°C.

2.22.4 Size fractionation of poly A^+ mRNA

4 ml denaturing sucrose gradients were prepared by layering equal volumes of 5%, 10%, 15% and 20% (w/v) sucrose in 3mM-Tris/HCl pH 7.9,

3mM-EDTA, 70% formamide. Up to 100 µg of poly A^+ mRNA was applied to the gradient and this was centrifuged at the equivalent of 92,000 g for 18 hours (Beckman SW60). 10-12 equal fractions were removed, the volume was increased to 350 µl and the concentration of poly A^+ in each fraction was determined before ethanol precipitation of the poly A^+ mRNA.

2.23 In vitro translation of RNA

2.23.1 The reaction

An Amersham *in vitro* translation kit (code No.90) was used for the translation of mRNA. RNA samples were precipitated from under ethanol and redissolved into H_2O. A 7:1 reaction mix of reticulocyte lysate: label was prepared and 8 µl was used for each translation reaction. RNA (in 2 µl) was added and the reactions were incubated at 30°C for 1 hour. 1 µl was removed from each translation for TCA precipitation (section 2.23.2) and 9 µl of sample buffer (2% SDS, 20% glycerol, 0.001% bromophenol blue, 60mM-Tris/HCl pH 6.8) was added to the remainder before storage at -20°C.

2.23.2 TCA precipitation of translation products

1 µl from the translation reaction was removed into 0.5 ml 1M-NaOH, 5% H_2O and incubated at 37°C for 10 minutes. 2 ml of 25% TCA (trichloroacetic acid), 1 mg/ml caseine hydrolysate (ice cold) was added and following incubation for at least 30 minutes the samples

were filtered through Whatman GF/A glass fibre discs. These were washed firstly with ice cold 8% TCA and then 95% ethanol before being dried under a lamp and counted by liquid scintillation in 4 ml of ECOSCINT.

2.23.3 Polyacrylamide gel electrophoresis in the presence of SDS (SDS-PAGE)

SDS-PAGE was performed by the method of Laemmli,(1970) in a slab gel apparatus. Separation gels were 10%, acrylamide (at an acrylamide:bis-acrylamide ratio of 30:0.8). The gel buffer was 375mM-Tris/HCl pH 8.8, containing 0.1% (w/v) SDS and polymerisation was initiated by adition of 0.033% (v/v) TEMED, 0.05% (w/v) ammonium persulphate (final concentrations). Stacking gels of 3% (w/v) acrylamide were overlaid upon the separation matrix. The stacking gel buffer was 125mM-Tris/HCl pH 6.8, containing 0.1% (w/v) SDS. Polymerisation was induced by addition of 0.067% TEMED and 0.1% (w/v) ammonium persulphate (final concentration).

SDS-PAGE was carried out at room temperature, and samples (in sample buffer, section 2.23.1) were made 1% (v/v) 2-mercaptoethanol followed by heating at 100°C for 3 minutes. The well buffer contained 3g/litre Tris base, 14.4g/litre glycine and 0.1% (w/v) SDS.

The gel was fixed in 10% (v/v) acetic acid, 10% (v/v) methanol for 1 hour and washed in H_2O (many changes) for 1 hour.

2.23.4. Fluorography of SDS-PAGE gels

The gels were submerged in 1M-Na salicylate for 1 hour with shaking. Following this they were dried onto Whatman 3mm paper and exposed to Fuji RX film.

2.24 Northern blot analysis of RNA

A 1% agarose gel was prepared in 1 x MOPS buffer (20mM-MOPS, 5mM-Na acetate pH 7.0, 1mM-EDTA), 6.8% (w/v) formaldehyde where the formaldehyde was added when the melted agarose had cooled to 65°C. The running buffer was 1 x MOPS buffer. Apparatus used for the northern blot (gel kit, blotting equipment) were pre-treated overnight with 3% H_2O_2 prior to use.

The RNA (in 4.8 ul H_2O) was added to 10 µl deionised formamide, 2 µl 10 x MOPS buffer and 3.2 µl 38% (w/w) formaldehyde. This mixture was heated at 65°C for 5 minutes, cooled on ice then made 12.5% (w/w) glycerol, 0.025 mg/ml bromophenol blue. This was loaded onto the gel, which was electrophoresed overnight. During electrophoresis the running buffer was recirculated. The gel was blotted overnight onto Hybond-N using (DEPC treated) 20 x SSC in a typical blot set up (Maniatus et al., 1982). The filter was baked for 2 hours at 80°C and hybridised as discussed previously (section 2.20).

2.25. Synthesis of P. sativum cDNA

2.25.1 The first and second strand synthesis reactions

The cDNA was synthesised using an Amersham cDNA synthesis kit (cat No. RPN12567). The standard 20 µl first strand synthesis reaction contained 1 µg of poly A^+ mRNA, 1mM-dATP,-dGTP and -dTTP, 0.5mM-dCTP (unlabelled), 5 µCi $[\alpha-^{32}P]$ – dCTP, 1 unit of human placental ribonuclease inhibitor, 20 units of reverse transcriptase and the reaction was primed with either 1.6 µg oligo-dT or 0.3 µg random hexanucleotide primers. The buffering constituents and concentration of Na pyrophosphate was not provided.

The mRNA (in H_2O) was heated to 70°C for 1 minute followed by rapid cooling on ice before addition to the reaction. Synthesis was carried out at 42°C for 90 minutes after which time it was placed on ice and a 1 µl sample removed for the analysis of cDNA products.

To the first strand cDNA mix was added 50 µCi $[\alpha^{32}P]$–dCTP, 0.8 units *E. coli* ribonuclease H, 23 units *E. coli* DNA polymerase I and the volume was increased to 100 µl. The buffer constituents were adjusted (details of constituents were not provided).

The second strand cDNA synthesis reaction was incubated at 12°C for 60 minutes and then the polymerase was inactivated at 70°C for 10 minutes. 2 units of T4 DNA polymerase I was added and the reaction was incubated at 37°C for 10 minutes. The reaction was stopped by the

addition of 4 µl 0.25M-EDTA. A sample was removed for analysis of cDNA products.

The cDNA was purified by phenol/chloroform extraction and ethanol precipitation with ammonium acetate in the presence of 20 µg glycogen.

2.25.2 Precipitation of nucleic acids with TCA

A known volume of the sample was spotted onto Whatman GF/A glass-fibre discs (2.4 cm diameter) using an Eppendorf Comforpipette 4700. An equal volume of sample was added to a tube containing 100 µl of 500 µg/ml salmon sperm DNA, 20mM-EDTA. To this was added 5 ml of ice cold 10% TCA which was left on ice for 30 minutes. The precipitate was filtered through another GF/A disc, washed with ice cold 10% TCA followed by 95% ethanol. Both filters were dried under a lamp and counted by liquid scintillation in ECOSCINT. The first filter measures total radioactivity in the sample; the second filter measures the radioactivity incorporated into nucleic acids.

2.25.3 Analysis of the cDNA synthesis products

The inclusion of $[\alpha\text{-}^{32}P]$-dCTP in the first and second strand cDNA synthesis enabled the efficiency of the synthesis to be monitored in terms of both the amount obtained and the size distribution of the products. An outline of the monitoring procedure is given below and details of both the protocol and the calculation of results are given in sections 2.25.4 and 2.25.5 respectively.

1. The percentage of labelled [α^{32}P]-dCTP incorporated into both first and second strand cDNA was determined by TCA precipitation followed by liquid scintillation.

2. The percentage incorporation values enabled the quantity of first and second strand cDNA to be calculated.

3. Knowing the quantity of cDNA obtained, allowed the calculation of (i) the percentage of mRNA transcribed into cDNA and (ii) the percentage of second strand cDNA transcribed from the first strand cDNA. These percentages (yields) could then be compared between various cDNA syntheses which gave an indication as to the relative efficiency of the synthesis.

4. The size of the cDNA products of both first and second strand synthesis was examined on an alkaline agarose gel (section 2.25.6)

2.25.4 Sequential labelling of first and second strand cDNA

5 μCi of [α^{32}P]-dCTP was added to the standard 20 μl first strand synthesis reaction mix. A 1 μl sample was removed into 19 μl of 0.2M-EDTA producing solution F. Using the remainder of the labelled first strand reaction, the standard 100 μl second strand synthesis reaction was set up where a further 50 μCi of [α^{32}P]-dCTP was included. A sample of 1 μl was removed into 19 μl of 0.2M-EDTA to produce solution S. 2 μl of solution F and 2 μl of solution S were separately spotted onto the centre of 2.4 cm discs of Whatman GF/C paper and these were

designated filters A and C respectively. 2 µl of solution F and 2 µl of solution S were separately added to 100 µl of 500 µg/ml salmon sperm DNA and these were designated B and D respectively.

The nucleic acids in B and D were TCA precipitated (Maniatus et al., 1982). 5 ml of ice cold 10% TCA was added to each and this mixture was incubated on ice for 15 minutes. Solutions B and D were filtered through Whatman GF/C paper discs, which were then washed extensively with firstly ice cold 10% TCA and then 95% ethanol.

Filters A, B, C and D were then dried under a lamp and counted by liquid scintillation. From the cpm obtained, the percentage incorporation was calculated (section 2.25.5).

Since the quantity of incorporated and unincorporated [α^{32}P]-dCTP in the reaction was being estimated using separate 2 µl samples from solution F (and solution S), it was essential that a reproducible volume (even if this was not exactly 2 µl) was removed from each of the diluted reactions. Gilson micropipettes were found to pipette irreproducible aliquots which resulted in uninterpretable yield values. The Eppendorf Comforpipette 4700 was found to give the most satisfactory results.

2.25.5 Calculation of efficiency of cDNA synthesis

filter A contains the total cpm in the first strand reaction = A cpm

filter B contains the cpm incorporated into first strand cDNA = B cpm

filter C contains the total cpm in the second strand reaction = C cpm

filter D contains the cpm incorporated into cDNA after second strand synthesis = D cpm

The percentage incorporation in the first strand reaction can be calculated as follows:

$$\text{percentage incorporation (first strand)} = \frac{B}{A} \times 100\%$$

However before calculating the percentage incorporation for the second strand reaction the amount of radioactivity in the first strand reaction which contributed to the cpm on filters C and D required to be subtracted. Since the first strand reaction mix was 1/5.5 of the volume of the stopped second strand mix (first strand reaction volume, 19 µl; second strand reaction volume, 104 µl), B must be divided by 5.5 to compensate for this and the resultant value subtracted from C and D.

The percentage incorporation in the second strand reaction was calculated as follows:

$$\text{percentage incorporation (second strand)} = \frac{D - \frac{B}{5.5}}{C - \frac{B}{5.5}} \times 100\%$$

The percentage incorporation values were used to calculate the amount

of cDNA synthesised for both the standard (20 µl) first strand and (100 µl) second strand reaction as follows:

(i) amount of cDNA synthesised = (% incorporation x 140) ng

(ii) total amount of double stranded cDNA = 2 x amount of second strand cDNA synthesised

The yield of the first and second strand synthesis was calculated from the following formulae:

$$\text{Yield 1} = \frac{\%\text{ of mRNA}}{\text{transcribed}} = \frac{\text{ng of first strand cDNA synthesised}}{\text{ng of input mRNA}} = 100\%$$

$$\text{Yield 2} = \frac{\%\text{ of second cDNA transcribed from first strand cDNA}}{} = \frac{\text{ng of second strand cDNA}}{\text{ng of first strand cDNA}} = 100\%$$

Yields 1 and 2 give an estimate of the success of the first and second strand syntheses. It is recommended (from Amersham) that using 1 µl of mRNA in the standard reaction the following yield values should be obtained:

Yield 1 = 15 - 30% Yield 2 = 90%

2.25.6 Alkaline agarose gel analysis of cDNA

The products of the cDNA synthesis reactions were analysed on alkaline agarose gels (McDonnell et al., 1977; Maniatus et al., 1982). Samples

containing 10-30,000 cpm of radioactivity incorporated into nucleic acid was taken from the cDNA synthesis reaction mix and placed in a microcentrifuge tube. An alkaline hydrolysis step was then carried out to remove any remaining RNA. 20 µl of 100 µg/ml salmon sperm DNA was added to the cDNA synthesis products followed by 1/3 of the combined volume of 1M-NaOH. After incubation at 46°C for 30 minutes a volume of 1M-HCl, equal to the volume of NaOH, was added. A volume of 1M-Tris pH 8, equal to that of the HCl was then added. The mixture was phenol/chloroform extracted and then ethanol precipitated. The precipitated cDNA was resuspended in 10 µl of 50mM-NaOH, 1mM-EDTA, 2.5% Ficoll, 0,025% bromophenol blue. This was loaded onto a 1% agarose gel prepared in 500mM-NaCl, 10mM-EDTA and electrophoresed in 300mM-NaOH, 10mM-EDTA. (The gel had been soaked in electrophoresis buffer for at least 30 minutes). After electrophoresis, the gel, it was soaked in 7% trichloroacetic acid (two changes) for 30 minutes. The gel was then dried onto 3MM paper and autoradiographed.

2.26 Generation of a P. sativum cDNA library

2.26.1 Cloning the P. sativum cDNA into λgt10

The cDNA synthesised from P. sativum poly A^+ mRNA was cloned into λgt10/EcoRI arms using the Amersham cDNA cloning kit (cat No. RPN 1257). Where known the reaction constituents are given.

Up to 1 µg of cDNA was methylated in a 20 µl buffered reaction containing S-adenosyl methionine and 20 units of EcoRI methylase which

was incubated at 37°C for 60 minutes. The methylase was heat inactivated at 70°C for 10 minutes. The methylated cDNA reaction was made up to 30 µl with the appropriate buffer, 1 µg of EcoRI linkers (GGAATTCC) and 5 units of T4 DNA ligase. The ligation took place overnight (16-20 hours) at 15°C, after which the enzyme was heat inactivated. The linkered cDNA reaction was adjusted to appropriate EcoRI digestion conditions, made up to 100 µl including 100 units of EcoRI and digestion occurred at 37°C for 5 hours. The enzyme was then heat inactivated.

The linkered cDNA was separated from the excess linkers by passage through a Sephadex-G50 gel filtration column. The column was rinsed with 3ml STE buffer and the 100 µl reaction (containing linkered cDNA) was applied. 200 µl of STE was applied to the column and allowed to flow through. 200 µl fractions were taken from the column and the cDNA was detected by Cerenkov counting.

2.26.2 Analysis of the phage titre

The titre of the cDNA-containing recombinants was determined on E. coli L87 (wild type) and E. coli NM514 (hfl). Several control ligations (i to iii) had been included in the ligation/packaging along with the cDNA reactions (iv).

(i) Control λgt10 DNA: the titre on the L87 host reflected the efficiency of the in vitro packaging reaction.

(ii) λgt10 arms only : this was the crucial control giving several items of information. The titre on L87 checked that the arms and T4 DNA ligase were working efficiently. Two other numbers could be obtained from this control. The ratio of titre on L87: NM514 (called the arms selective ratio) and the background titre (ie the number of non recombinants) on NM514.

(iii) λgt10 arms plus control blunt end DNA: this control monitored the performance of the whole cloning system.

(iv) cDNA reactions: when evaluating the results of the cDNA-containing tubes, the most important figure was the increase over the background level of plaques on NM514. A true background reading of the individual reactions was derived as follows: The arms selective ratio (from ii) resulted in an invariant property for the specific batch of λgt10 arms and host cells in combination and was directly applicable to the cDNA tubes since the same batches of arms and cells were used. The L87 titre represented the total products of the ligation/packaging reaction, which were mainly recreated parental λgt10. Dividing this figure by the arms selective ratio provided a value for the true background on NM514 if no recombinants had been formed. Subtracting this background figure from the actual titre obtained on NM514 cells by the cDNA reactions gave an accurate count of the number of recombinants obtained.

2.27 Cerenkov counting of ^{32}P labelled nucleic acid

In a non-destructive manner, the radioactivity present in a sample can be determined by Cerenkov counting (Berger, 1987). Each microcentrifuge tube was placed in a scintillation vial and counted with the ^{3}H channel of a scintillation counter.

2.28 Subcloning into M13

M13 is a male specific filamentous bacteriophage of E. coli. The virus particles contain single-stranded, circular DNA which following infection, serves as a template for the synthesis of the complementary strand. The double-stranded form of the viral DNA present in infected cells is called the replicative form (RF-form). Several modifications have been introduced to phage and host cells to produce an efficient vector/host system (Messing, 1983), making it an ideal source of single-stranded DNA to be sequenced by the dideoxy sequencing method (Sanger et al,, 1977). The double-stranded RF form was modified to serve as an efficient cloning vector, by introducing the same multiple cloning region (polylinker)and colour selection system for the identification of recombinants as in pUC vectors (section 2.17.3). All protocols for cloning are given in the "M13 Cloning and Sequencing Handbook" (Amersham International) and these were strictly adhered to.

2.28.1 Preparation of M13 DNA (RF)

Competant E. coli JM109 or TG1 (section 2.17.1) were separately

transformed with 10-20 ng of single-stranded DNA template (section 2.28.2). Single plaques were picked and used to inoculate 1 ml of a 1:100 dilution of an overnight culture of JM109 (in 2 x TY). This was shaken for 4-5 hours at 37°C. After removal of the cells by centrifugation, 1 ml of the supernatant was used to infect 1 ml of an overnight culture of JM109 in 500 ml of 2 x TY. This culture was shaken for 4-5 hours at 37°C and the RF DNA was isolated by a standard large scale plasmid preparation (section 2.5).

2.28.2 Ligation and transformation

Ligation of digested M13 RF DNA and insert DNA was exactly as described in section 2.16. Competant JM109 (or TG1) were prepared as in section 2.17.1. Tranformation of the cells with recombinant M13 DNA was as described in section 2.17.2 up to the plating out stage. After heat shock treatment 40 µl 100mM IPTG, 40 µl 2% w/v X-Gal (in dimethylformamide) and 200 µl fresh E. coli cells was added to the transformation. 3 ml of molten H-top agar was then added and this mixture was poured onto an H-plate, allowed to harden for 10 minutes then incubated at 37°C.

2.28.3 Preparation of single-stranded template DNA

A 10 ml 2xTY overnight culture of E. coli JM101 was used to inoculate fresh 2xTY medium at a ratio of 1 ml: 100 ml. A single plaque was lifted using a sterile Eppendorf pipette tip and inoculated into 1.5 ml of the low density E. coli JM101 culture. This culture was shaken

(37°C) for 5 hours and then centrifuged for 5 minutes in a microfuge. The supernatant was carfully transferred to a fresh Eppendorf tube and recentrifuged (5 minutes, microfuge) to remove any residual cells. The supernatant was added to 200 µl PEG/NaCl (20% (w/v) polyethylene glycol 6000/2.5M-NaCl), mixed thoroughly and left for 15 minutes at room temperature. Precipitated intact phage particles were harvested by centrifugation (5 minutes, microfuge), and residual PEG/NaCl removed by a second centrifugation step (2 minutes, microfuge). The viral pellet was resuspended in 100 µl TE buffer, 50 µl phenol/chloroform added, the mixture vortexed for 20 seconds (at least) and left at room temperature for 15 minutes. The aqueous layer was transferred to a fresh microfuge tube and extracted with 1 ml chloroform. 10 µl of 3M-sodium acetate pH6.0 and 250 µl ice-cold ethanol were added to the chloroform-extracted solution and the DNA precipitated at -20°C overnight. The DNA was recovered by centrifugation (10 minutes, microfuge) and the template DNA resuspended in 50 µl TE buffer. The integrity and purity of the template preparation was examined by running an aliquot on a 1% agarose gel. Template DNA was stored at -20°C.

2.29 DNA sequencing by the Sanger chain termination method

Sequencing described in Chapters 3 and 4 was carried out using Amersham M13 sequencing kit (cat No N40502) and protocols described in the 'M13 Cloning and Sequencing Handbook' (Amersham International) were strictly adhered to. Sequencing described in Chapter 5 was carried out using the United States Biochemical Corporation (USB)

sequencing kit and protocols described in 'DNA sequencing with Sequenase' (USB) were strictly adhered to.

The major difference between the two sequencing kits was the enzyme used to synthesise DNA complementary to the M13 template. The Amersham kit contained the Klenow fragment of DNA polymerase I whereas the USB kit contained Sequenase. Sequenase is a modified form of bacteriophage T7 DNA polymerase (Tabor and Richardson, 1987) and has a number of properties which make it very suitable for use in DNA sequencing, including low 3'-5' exonuclease activity and high rate of polymerisation of nucleotides. The USB DNA sequencing system was found to give better results than the Amersham sequencing system.

2.29.1. Sequencing using Klenow

(a) Annealings

5 µl of template DNA, prepared as described above, was annealed for 2 hours at 55°C in a mixture also containing:

1 µl M13 sequencing primer (see below)

1.5 µl 100mM-Tris/HCl, pH 8.5, 100mM-$MgCl_2$ (10 x Klenow Buffer)

2.5 µl distilled water

The M13 sequencing primer employed in this work was a 17-mer with the sequence 5'-GTA AAA CGA CGA CCA GT-3'. Annealings could be stored at -20°C for up to 1 month.

(b) Sequencing Reactions

The annealed primer/template mix was thawed and 1.5 µl (15µCi) [α^{35}S]-dATPαS at >600 Ci/mmol (Amersham SJ.304) and 1 unit of Klenow fragment added. After thorough mixing, 2.5 µl aliquots of this mix were spotted just inside the rim of four (A, C, G T) uncapped Eppendorf tubes. 2 µl of the relevant ddNTP/dNTP mix (see below) were added to the individual tubes and the contents mixed by a brief spin in the microfuge. The sequencing reactions were performed at ambient temperature and after twenty minutes 2 µl of a 0.5mM uniform chase solution of all four dNTP's added in a similar fashion. Following the 15 minute chase the sequencing reaction was stopped by addition of 4 µl of formamide dye mix (96% (v/v) deionised formamide, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanol, 20 mM-EDTA).

(c) Reaction mixes (composition)

(i) deoxyNTP mixes (A°, C°, T°) for [α-^{35}S]-dATPαS (volumes in µl)

	A°	C°	G°	T°
0.5mM dCTP	20	1	20	20
0.5mM dGTP	20	20	1	20
0.5mM dTTP	20	20	20	1
TE buffer	20	20	20	20

(ii) dideoxyNTP mixes

0.1mM - ddATP 0.1mM - ddCTP

0.3mM - ddGTP 0.5mM - ddTTP

(iii) dNTP/ddNTP mixes

To each dNTP mix (N°), an equal volume of the corresponding ddNTP mix was added.

2.29.2 Sequencing using Sequenase

(a) Annealings

The following mixture was heated to 65°C for 2 minutes and allowed to cool slowly to room temperature over a period of about 30 minutes. Cooling was carried out in a 100 ml glass beaker of 65°C H_2O.

7 µl template DNA (1-2 µg)

2 µl 200mM Tris/HCl pH 7.5; 100mM-$MgCl_2$; 250mM-NaCl

(5 x sequenase buffer)

1 µl M13 sequencing primer (see section 2.29.1)

(b) Sequencing Reaction

To the annealed primer/template was added 1 µl 0.1M DTT, 2 µl labelling mix (1.5 µM dGTP, 1.5 µM dCTP, 1.5 µM dTTP), 0.5 µl (5 µCi) [α^{35}S]-dATPαS, 1.5u Sequenase. After mixing this labelling reaction was incubated at room temperature for 5 minutes. During this incubation 3.5 µl of the labelling reaction was spotted just inside the rim of four (A, C, G, T) uncapped eppendorf tubes in which was 2.5 µl of the ddATP, ddCTP, ddGTP, ddTTP respectively termination mix (80 µM dATP, 80 µM dCTP, 80 µM dGTP, 80 µM dTTP, 8 µM ddNTP (appropriate

analogue), 50mM-NaCl). After the 5 minute incubation the tubes were incubated at 37°C for 1 minute, centrifuged in a microfuge (to mix the labelling reaction and termination mix) then returned to 37°C for a further 5 minutes. 4 µl of stop mix was finally added to the lip of each tube and after centrifugation the reactions were stored at -20°C.

2.29.3 Polyacrylamide gel electrophoresis

The nested set of primer extended fragments produced by either of the M13/dideoxy DNA sequencing protocols were resolved by electrophoresis on thin polyacrylamide gels.

Gels were either linear wedge or buffer gradient (Biggin et al., 1983), and were composed of the contituents detailed in Table 2.4. Linear gels were pre-run for at least 15 minutes at 30mA before electrophoresis. Samples were prepared by heating at 75-80°C for 2 minutes during which time the gel slots were thoroughly cleared of unpolymerised acrylamide and urea. The heat-denatured samples were loaded immediately on to the gel and electrophoresis carried out at 30mA (constant current). Aluminium sheets were used to distribute heat evenly across the gel plates thus preventing 'smiling' of the samples. The duration of electrophoresis varied with the amount of DNA sequence required. DNA sequence close to the primer extending to 250bp away from the primer was obtained from linear wedge or buffer gradient gels run for 2-5 hours (dye front reaches anode). DNA sequence from 200-400 bp away from the primer was obtained from linear wedge gels run for 7.5 hours.

	linear	buffer gradient	
		upper	lower
40% acrylamide	(ml) 10.5	6	1.12
10 x TBE	(ml) 7.0	2	1.88
sucrose	(g) -	-	0.75
urea	(g) 29.4	19.2	3.8
bromophenol blue (0.01g/ml)	(ml) -	-	0.08
distilled water	(ml) 52	31.8	4.4
TEMED	(µl) 70	80	15
AMPS (%)	(µl) 360 (10%)	80(25%)	15(25%)

10 x TBE: 108 g/l Tris base

(pH 8.3) 55 g/l boric acid

9.3 g/l EDTA. $2H_2O$

40% acrylamide: 38.2 acrylamide:bisacrylamide

TEMED: N, N, N', N'- tetramethylethylenediamine
AMPS: ammonium persulphate

Table 2.4 Sequencing acrylamide gel constituents

All gels were fixed in 10% (v/v) acetic acid, 10% (v/v) methanol for 30-45 minutes following electrophoresis and were then dried onto a sheet of Whatman 3MM paper using a Bio-rad model 1125 gel drier. The dried gel was autoradiographed using Fuji RX film at room temperature.

2.29.4 Treatment of oligonucleotides

Oligonucleotides for use as sequencing primers were provided as freeze dried DNA pellets. They were rehydrated in 133 µl of H_2O and 67 µl of 7.5M-ammonium acetate was added followed by 2.5 volumes of ethanol. The mixture was left at 4°C for 10 minutes (or overnight if 1 µg is present). DNA was recovered by centrifugation, washed with 70% ethanol dried in a dessicator for 10 seconds then rehydrated in 100 µl TE.

2.30 Computer programs for the analysis of DNA and amino acid sequences

A number of programs for the manipulation and analysis of DNA sequences of the UWGCG (University of Wisconsin Genetics Computer Group) package (Devereux *et al.*, 1984) were run on a VAX computer. This package contains programs for the comparison of DNA sequences with those in GenBank and EMBL (European Molecular Biology Laboratory, Heidelberg, Germany) databases.

2.30.1 UWGCG programs

BESTFIT: finds the best region of similarity between two sequences,

and inserts gaps if necessary to obtain optimal alignment.

CODONF: tabulates codon frequency from sequences.

GAP: finds an optimal alignment between two sequences by inserting gaps in either sequence. It considers all possible alignments and gap positions, and creates the alignment with the largest number of matched bases and fewest gaps.

LINEUP: is a screen editor for editing multiple sequence alignments.

MAP: displays both strands of a DNA sequence with a restriction map shown above the sequence and possible protein translations below.

SEQED: allows entering and modification of sequences.

STRINGS: finds sequences by searching sequence library documentation for character patterns eg EPSP.

TRANSLATE: translates nucleotide sequence into amino acid sequence.

WORDSEARCH: searches for similarity between a new sequence and any group of sequences.

2.30.2 File transfer

The output file from each UWGCG programme could be converted into a

text file on the Apple MacIntosh computer using the VersaTerm emulation program and the KERMIT-32 software mounted on the VAX computer. This file was then transferred to the format of the WRITE NOW wordprocessing program with the TRANSLATOR, and could then be edited for printing on the Laser Printer.

C H A P T E R 3

INITIAL ATTEMPTS TO CLONE THE 3-DEHYDROQUINASE/ SHIKIMATE DEHYDROGENASE GENE OF *Pisum sativum*

3.1 Introduction

As discussed in section 1.1.3 the most promising approach, when this work commenced, to clone the genes of the shikimate pathway enzymes from higher plants appeared to be use of microbial genes as heterologous hybridisation probes. This chapter describes investigations to determine the feasibility of this approach, in particular to isolate the E2/E3 gene of P. sativum, and goes on to describe attempts to clone the sequence which was detected.

The relationship between two nucleic acid sequences can be analysed using nucleic acid hybridisation. The process which underlies all of the methods based on nucleic acid hybridisation is the formation of the double helix from two complementary strands. There are two important features on which this process depends; the two sequences involved in duplex formation must have a degree of complementarity, and the stability of the duplex formed depends on the extent of the complementarity. The stability of nucleic acid duplexes can be related to one another by their melting temperature (Tm). This is the temperature at which the strands are half dissociated (Hames and Higgins, 1985).

The rate of a hybridisation reaction is affected by the incubation temperature (Marmur and Doty, 1961). For example at 0°C hybridisation proceeds very slowly but as the temperature is raised, the rate increases dramatically to reach a broad maximum at an optimum hybridisation temperature which is 20-25°C below the Tm for DNA-DNA

annealing. This relationship applies to the formation of completely homologous hybrids and also to the formation of poorly matched hybrids. Mismatching, however, has the effect of lowering the rate of hybridisation and therefore the Tm of hybrids (Bonner et al., 1973). As a result, the optimum incubation temperature for the hybridisation of heterologous DNA sequences is lower than the temperature required for perfect hybrid formation.

Formamide decreases the Tm of nucleic acid hybrids. This is a very useful property because by including this denaturing agent in hybridisation solutions, the incubation temperature (Ti) can be reduced (McConaughy et al., 1969; Howely et al., 1979).

At low ionic strengths, well or poorly matched nucleic acids hybridise very slowly, but as the ionic strength increases (up to 1.5M Na^+) the reaction rate increases (Wetmer and Davidson, 1968; Hames and Higgins, 1985). Heterologous and homologous hybridisations therefore are carried out at high ionic strengths. High salt concentrations stabilise mismatched duplexes and so to maintain cross-hybridised species, the salt concentration of washings must remain relatively high compared to those concentrations required to maintain homologous hybrids.

The stability (Tm) of nucleic acid hybrids is therefore dependent on the concentration of formamide and the ionic strength in the hybridisation reaction. Base composition of the DNA is also important since a high proportion of G and C will result in more hydrogen

bonding between the strands producing a more stable hybrid (Wetmur and Davidson, 1968). The following relationship (Hames and Higgins, 1985) has been derived from studies on hybridisation in solution but it is expected to be similar, qualitatively at least, for filter hybridisation:

$$Tm = 81.5 + 16.6(\log M) + 0.41(\%G+C) - 0.72(\%\text{formamide})$$

where M is the molarity of the monovalent cation. For a hybridisation reaction containing 50% formamide, 1M NaCl (equivalent to 6 x SSC) and assuming 50% G+C content then the calculated incubation temperature, bearing in mind that Ti = Tm - (20 to 25°C), is 41°C to 46°C. A typical hybridisation involving homologous DNA sequences using these conditions is carried out at 42°C.

The cross hybridisation of DNA from E. coli and S. cerevisiae with P. sativum DNA would involve mismatching between the sequences and so to obtain stable hybrids it was foreseen that hybridisation would have to be attempted under very permissive conditions. To achieve this low stringency, either the incubation temperature was kept low and/or the formamide concentration was kept as low as necessary. Washing of the hybrids was carried out at relatively low temperatures and at high ionic strength.

3.2 The DNA sequences used as probes

During the cloning and sequencing of the E. coli aroD (E2) and aroE (E3) genes and the S. cervisiae ARO1 gene, a variety of plasmids

derived from pAT153 and M13 single stranded templates, containing the whole genes or fragments of the genes, had been constructed (Duncan, 1984; Anton, 1985; Duncan et al.,1986, 1987; Anton & Coggins, 1988). These were available for use as probes. Convenient restriction enzyme sites were used to cut out appropriate fragments.

The complete coding regions, for the particular genes were used as probes, since using small pieces of each gene would have involved choosing the most homologous regions. This choice could only have been based upon the homologies found between the microbial sequences and may have emphasised regions which were not well conserved in the plant sequences. Also, if the plant genes were going to hybridise with the microbial probes, it would almost certainly be due to the cumulative effect of several regions of homology rather than just one. Furthermore the large size of the DNA containing the complete coding sequence would allow more labelled nucleotide to be incorporated into the probes with the result than any signal due to hybridisation would be amplified.

Plasmid pKD201 contains the complete E. coli aroD (E2) gene (Duncan et al., 1986). An aroD probe was isolated from pKD201 as a 1069 bp BglII-ClaI fragment (Fig 3.1a). This fragment contained 98% of the aroD coding region because BglII cuts 20 bp downstream of the AUG start codon. It also contained 350 bp of non-coding 3' DNA which could not conveniently be removed. The position of the BglII site meant that the upstream open reading frame, which may encode another polypeptide (Duncan et al., 1986) was excluded from the probe.

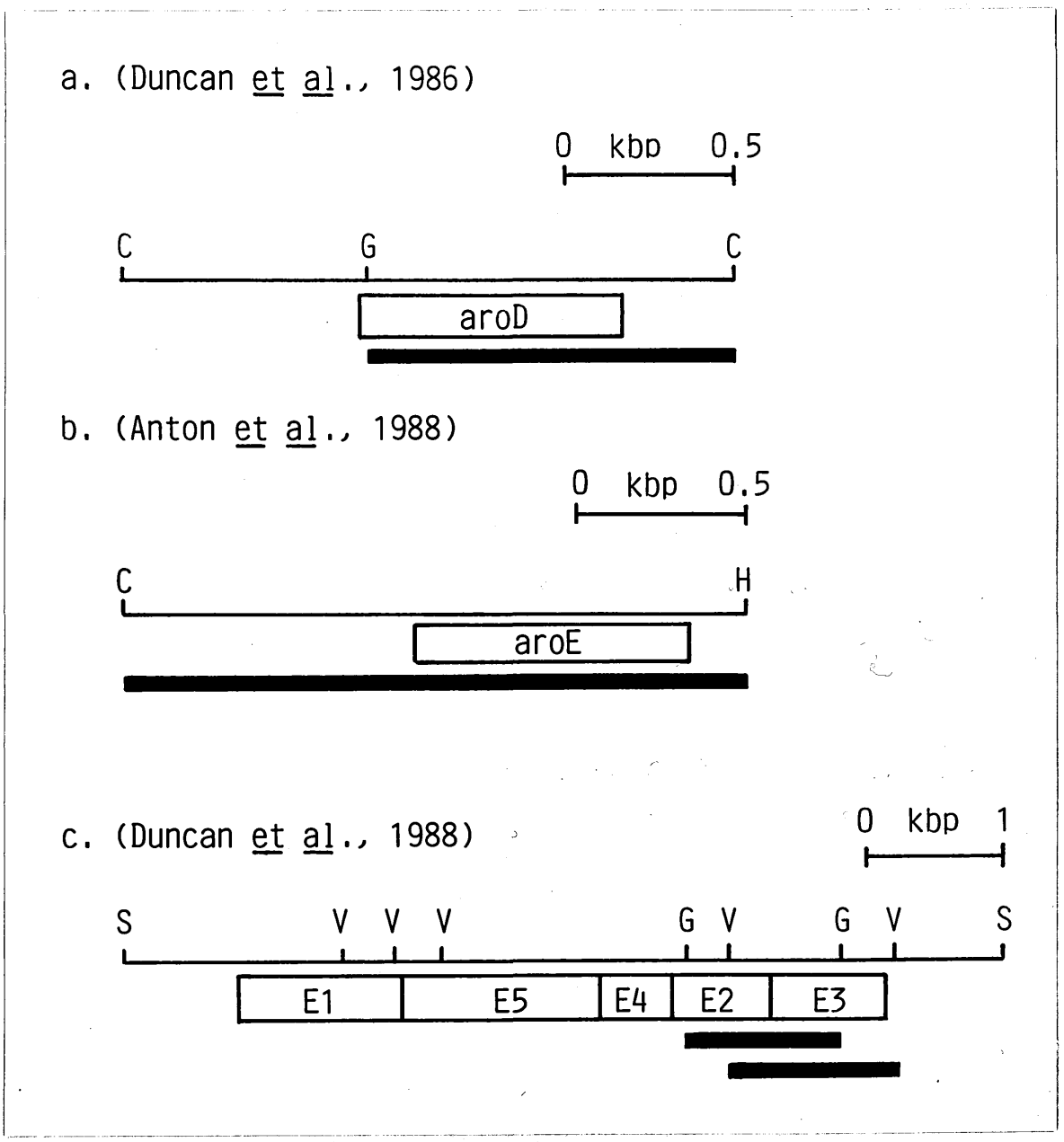


Fig 3.1 Fragments of DNA for use as probes

Key to restriction sites = C=ClaI, G=BglII, H=HindIII, V=EcoRV, S=Sau3A

The relevant restriction sites in the inserts of (a) pKD201 (b) pIA301 and (c) pME173 are shown. The open boxes indicate the location of the protein coding regions for the (a) aroD (b) aroE and (c) ARO1 gene products. The closed boxes indicate the fragments of DNA used as probes.

Plasmid pIA301 contains the complete E. coli aroE (E3) gene (Anton et al., 1988). An aroE probe was isolated as a 1820 bp ClaI-HindIII fragment from pIA301 (Fig 3.1b). The whole aroE gene was encoded in this fragment along with 840 bp of upstream and 170 bp of downstream non-coding DNA. There were no convenient restriction enzyme sites available which would have limited the probe to DNA containing only the aroE gene. This probe therefore contained a significant length of open reading frame within the 840 bp of upstream DNA (Anton, 1985), which may encode a polypeptide.

Plasmid pME173 encodes the complete S.cerevisiae ARO1 gene (Duncan et al., 1987, 1988). The DNA coding for the E2 and E3 domains of the arom polypeptide was isolated on two overlapping fragments: a 1132 bp BglII fragment and a 1277 bp EcoRV fragment (Fig 3.1c). These would be used together as a probe for the pea E2/E3 gene. 35 bp of DNA coding for the N-terminal region of the E2 domain was missing from the fragments which also contained 50 bp of non-coding DNA.

3.3 Southern blots of P. sativum genomic DNA

3.3.1 The integrity of isolated P. sativum genomic DNA

Genomic DNA was isolated from P. sativum, E. coli K12 and S. cerevisiae S228C. The latter two DNA stocks were used for hybridisation controls. Samples of the three DNA preparations were electrophoresed on an agarose gel (Fig 3.2). This indicated that the

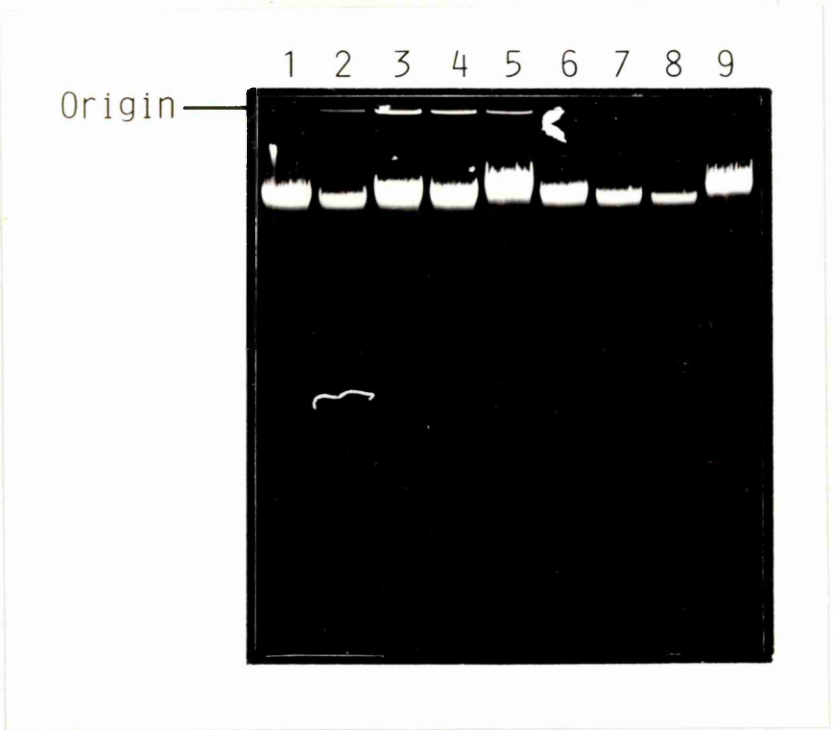


Figure 3.2 Agarose gel analysis of genomic DNA samples

lane:
1. 0.5 µg E. coli genomic DNA
2. 0.1 µg P. sativum genomic DNA
3. 0.5 µg P. sativum genomic DNA
4. 0.25 µg P. sativum genomic DNA
5. 1 µg S. cerevisiae genomic DNA
6. 0.5 µg S. cerevisiae genomic DNA
7. 0.25 µg S. cerevisiae genomic DNA
8. 0.05 µg S. cerevisiae genomic DNA
9. 0.6 µg λ- DNA

The DNA samples were electrophoresed on a 0.8% agarose gel.

DNA samples were of sufficiently high molecular weight, and at least equal in size to uncut λ-DNA. These DNA samples were therefore suitable for southern blot analysis. The P. sativum DNA contained considerable amounts of RNA and it was therefore routinely treated with RNaseA before use.

One way of testing the integrity of the P. sativum genomic DNA was to use a P. sativum gene to probe a genomic southern blot of the P. sativum DNA. P. sativum DNA was digested with BglII, EcoRI and EcoRV. The digests were electrophoresed on an agarose gel and southern blotted. The filter was probed at high stringency with the nick translated insert from pSSU161 (Bedbrook et al., 1980) which contains a partial cDNA clone for one of the genes encoding the small subunit of ribulose-bisphosphate carboxylase/oxygenase (rubisco SSU) of P. sativum. The autoradiograph (Fig 3.3) showed a variety of intense and not so intense bands which differed in size depending on the enzyme used to digest the P. sativum DNA. The pattern was consistent with the probe hybridising to DNA sequences from members of the rubisco SSU multigene family and the EcoRI and BglII digestions resembled published data (Corruzi et al., 1984).

3.3.2 P. sativum genomic southern blots probed with microbial DNA sequences

10 μg of total mammalian DNA (haploid genome size of 3×10^9 bp) is required to detect sequences which occur at the single copy level on a

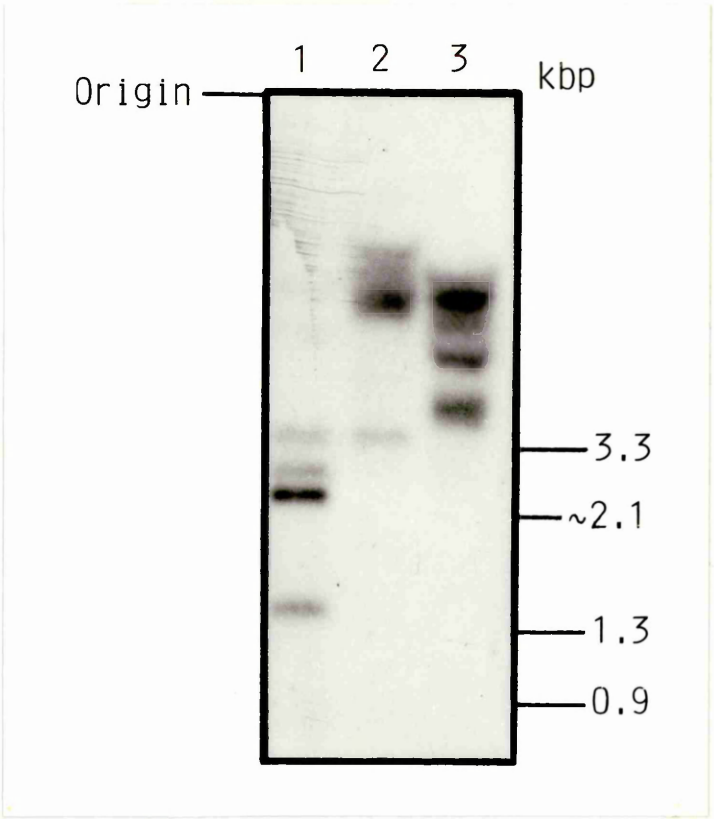


Figure 3.3 Southern analysis of P. sativum DNA hybridised with a rubisco SSU cDNA probe

lane: 1. EcoRV-digested P. sativum DNA
2. BglII- " " "
3. EcoRI- " " "

15 μg samples of digested P. sativum DNA were electrophoresed on a 0.8% agarose gel. The southern blot of this gel was hybridised at 42°C in hybridisation solution containing 50% formamide and washed at 50°C in 0.5 x SSC, 0.1% SDS for 2 hours.

southern blot (Maniatis et al., 1982). It was assumed that the P. sativum E2/E3 gene would be present in relatively few copies within the P. sativum genome which has a haploid genome size of 4.5×10^9 bp (Bennet and Smith, 1976). Therefore 15 µg of EcoRI-digested P. sativum DNA was used on southern blots. Six identical southern blots were prepared containing P. sativum, E. coli and S. cerevisiae DNA.

The E. coli aroD probe was labelled by nick translation and used to probe two of the southern blots. The first was hybridised under low stringency conditions (30% formamide, 42°C) and washed in 2xSSC at 50°C. The second was hybridised under high stringency conditions (50% formamide, 42°C) and washed in 0.1xSSC at 50°C . After 10 days autoradiography no bands were detected in the lanes containing digested P. sativum DNA at either low or high stringency. The probe did however identify the aroD gene in E. coli DNA on a high molecular weight band (>10 kbp). This observation is consistent with there being no EcoRI sites in close proximity to the aroD gene (Duncan, 1984).

The E. coli aroE probe was also labelled by nick translation and used to probe two of the southern blots under the same conditions as for the aroD probe. A similar result was obtained. The aroE probe hybridised only with the E. coli DNA and not with the P. sativum DNA.

Both fragments containing the DNA coding for the S. cerevisiae E2/E3 domains (from now on called the S. cerevisiae E2/E3 probe) were labelled by nick translation and used together to probe the remaining two filters under the same conditions employed for the aroD probe. The probe hybridised with 3 bands in the lane containing the EcoRI digested S. cerevisiae DNA (Fig 3.4); the sizes of the bands which hybridised were consistent with predictions from the restriction map of the ARO1 gene (Duncan et al., 1987). The E2/E3 probe also hybridised to two bands present in the lane containing EcoRI digested P. sativum DNA at both low and high stringency. This suggested a high degree of homology between the S. cerevisiae sequence and the P. sativum DNA. One of the hybridising bands, present in the lane containing the digested P. sativum DNA, was very similar in size to the 0.9 kbp band present in the lane containing digested S. cerevisiae DNA. This was thought to be coincidental especially when the 1.5 kbp band in the digested P. sativum DNA lane did not co-migrate with any of the other bands in the S. cerevisiae lane. A further band was identified by the E2/E3 probe in the lane containing the EcoRI digested E. coli DNA. Neither the E. coli aroD, nor the E. coli aroE probes had identified bands in lanes containing S. cerevisiae DNA. This potential discrepancy was assumed to be because there was contamination by P. sativum DNA in the gel track containing the E. coli DNA.

Other southern blots were prepared containing P. sativum DNA digested with AvaI, BamHI, BglII, EcoRI and HindIII and these were probed with the S. cervisiae E2/E3 probe at the high stringency conditions. The

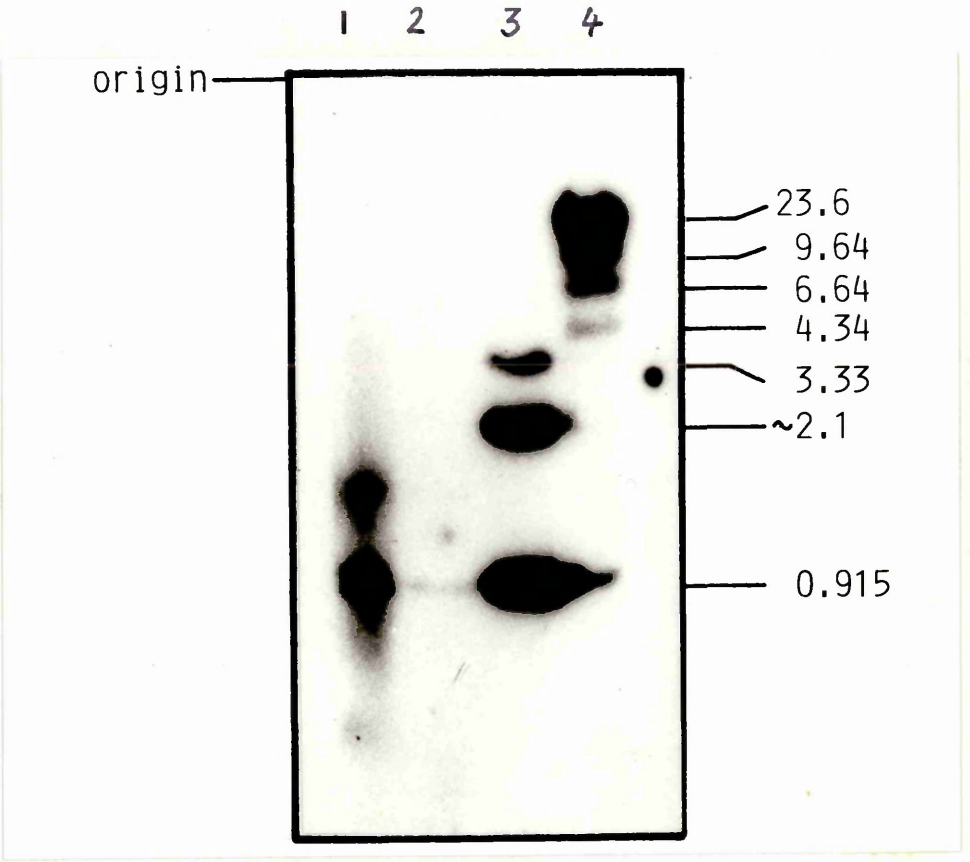


Figure 3.4 Southern analysis of P. sativum, E. coli and S. cerevisiae DNA hybridised with the S. cerevisiae E2/E3 probe

lane: 1. EcoRI-digested P. sativum DNA
2. " " E. coli DNA
3. " " S. cerevisiae DNA
4. HindIII- " λ-DNA

15 μg, 3 μg and 3 μg of digested DNA from P. sativum, E. coli and S. cerevisiae respectively, was electrophoresed on a 0.8% agarose gel prior to southern blotting. The southern blot was hybridised with the E2/E3 probe at 42°C in solution containing 50% formamide and washed at 50°C in 0.1 x SSC, 0.1% SDS. λ-DNA, labelled by nick translation, was included during hybridisation to show up the HindIII digested λ-DNA markers.

results (Fig 3.5) confirmed that this probe could pick up two bands in the lane containing EcoRI digested P. sativum DNA. Also a band of about 1.2 kbp could be detected in the lane containing BglII digested P. sativum DNA. However the probe did not detect any bands in the lanes containing AvaI, BamHI or HindIII digested P. sativum DNA. It was assumed that these enzymes produced fragments which were too large to be transferred efficiently even though all gels had been treated with HCl prior to blotting in order to depurinate the DNA and therefore cleave the larger DNA fragments into smaller, more easily transferred fragments.

The conclusion after this series of southern blotting experiments was that the E. coli aroD and aroE probes were unable to hybridise with the P. sativum DNA but that the S. cervisiae E2/E3 probe could reproducibly hybridise with P. sativum DNA, presumably with the P. sativum E2/E3 gene. The hybridisation of the S. cerevisiae E2/E3 probe with the digested P. sativum DNA was unlikely to be due to homologous repetitive DNA sequences (which will be discussed in section 3.7) because the stringency of probing was very high. It was hoped that this hybridisation would be useful for screening for the P. sativum E2/E3 gene.

3.4 Investigations involving P. sativum RNA

In order to investigate whether the transcript of the P. sativum E2/E3 gene could be detected by the S. cerevisiae E2/E3 probe, northern blot

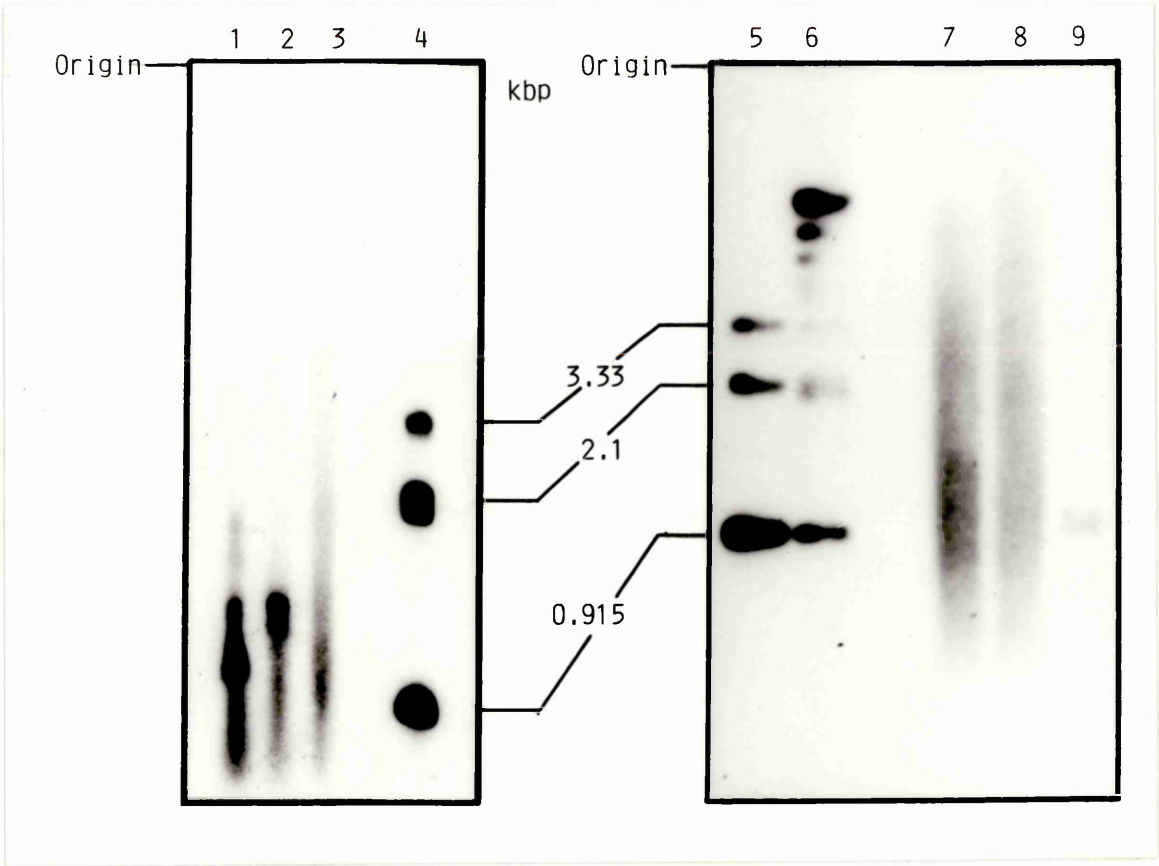


Figure 3.5 Southern analysis of P. sativum DNA hybridised with the S. cerevisiae E2/E3 probe

lane:
1. EcoRI-digested P. sativum DNA
2. BglII- " " "
3. AvaI - " " "
4. EcoRI- " S. cerevisiae DNA
5. " - " " "
6. HindIII- " λ-DNA
7. HindIII- " P. sativum DNA
8. BamHI - " " "
9. EcoRI - " " "

15 μg of digested P. sativum DNA and 3 μg of digested S cerevisiae DNA were electrophoresed on a 0.8% agarose gel prior to southern blotting. The southern blots were hybridised and washed as for Fig 3.4.

analysis of P. sativum RNA was carried out.

3.4.1 The integrity of the P. sativum RNA

Total RNA was isolated from 9 day old P. sativum seedlings and samples were examined on a non-denaturing agarose gel (Fig 3.6). From this gel it appeared that the DNaseI used during the RNA preparation was RNase free because the large DNA band (lane 1) in the non-DNaseI treated sample was removed without degrading the RNA (lane 2) to any great extent. The distinctive 28S, 18S and 5S rRNA bands were more obvious in lanes 3,4 and 5 where less RNA was loaded on the gel. There was a smear of RNA between these rRNA band which suggested that good quality RNA was being prepared.

However to ensure that the RNA smear was not due to RNA degradation, a northern blot, containing total RNA and poly A^+ mRNA, was probed at high stringency with the insert from pFab31, a cDNA clone for the P. sativum chlorophyll a/b binding (cab) protein (Bennet et al., 1984). A single band was present in the lane containing poly A^+ mRNA and a more intense band of similar size was present in the lane containing total RNA (Fig 3.7). Several other bands were also present in the latter lane. These were unlikely to be due to specifically sized degradation products of the cab mRNA but were possibly due to smaller RNA species which had homology to the cab cDNA. The cab protein is very abundant in green P. sativum tissue (Bennet et al., 1984) therefore it is not surprising that its message was present at such high levels in total RNA extracted from nine day old P. sativum seedlings. The signal was

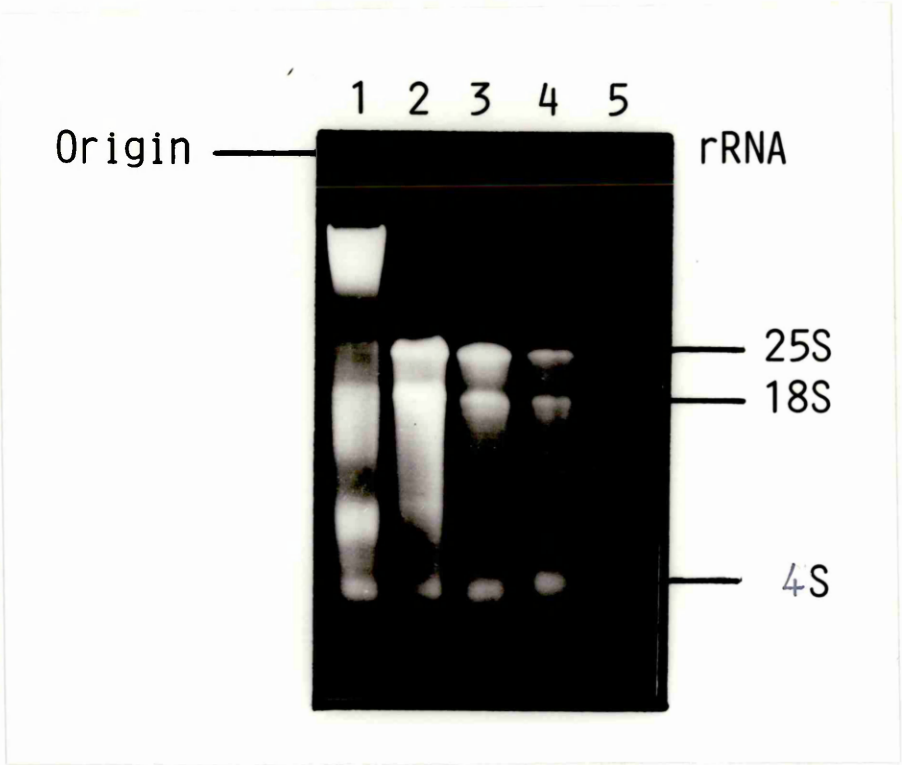


Figure 3.6 Agarose gel analysis of P. sativum RNA

lane: 1. 15 µg P. sativum RNA (crude preparation)
2. 15 µg DNase I treated P. sativum RNA
3. 10 µg " " " "
4. 5 µg " " " "
5. 1 µg " " " "

The RNA samples were electrophoresed on a 1% agarose gel. The rRNA bands (labelled appropriately) serve as size markers, as did the 4S tRNA.

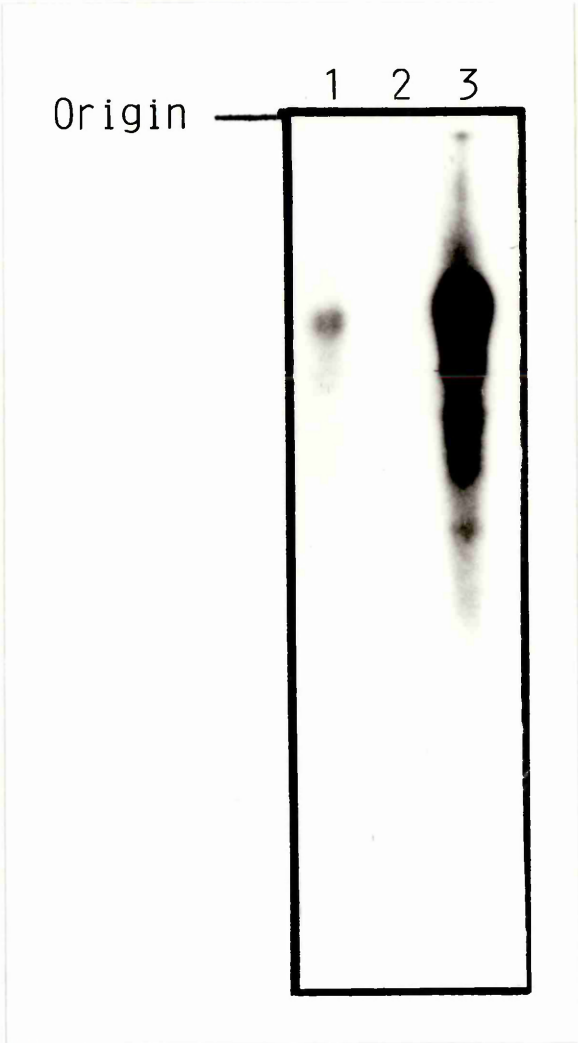


Figure 3.7 Northern blot of P. sativum RNA hybridised with a P. sativum cab protein cDNA probe

lane: 1. 2 µg P. sativum poly A^+ mRNA
2. 0.2 µg P. sativum poly A^+ mRNA
3. 60 µg " total "

The RNA samples were electrophoresed on a formaldehyde denatured 1.4% agarose gel. The northern blot was hybridised at 42°C in hybridisation solution containing 50% formamide and washed at 65°C in 0.1 x SSC, 0.1% SDS.

not so intense in the lane containing poly A^+ mRNA but this was because very little poly A^+ mRNA had been prepared and so it was difficult to quantify exactly how much had been applied to the gel.

The result did suggest that the P. sativum RNA was suitable for northern blot analysis and also that the northern blotting technique was satisfactory.

3.4.2. Northern blots of P. sativum RNA probed with the S. cerevisiae E2/E3 probe.

When northern blots containing similarly prepared P. sativum total RNA and poly A^+ mRNA were probed at low and high stringency with the S. cerevisiae E2/E3 probe no signal was obtained on autoradiographs, even after several months exposure. Similarly, RNA dot blots did not hybridise to the S. cerevisiae probe. Because of the success of the control northern blotting experiments described above this was thought unlikely to be due to either poor quality RNA or to unsatisfactory northern analysis technique. The most likely explanation was the low abundance of the P. sativum E2/E3 mRNA. Since there had been sufficient homology for the same probe to hybridise with presumably the P. sativum E2/E3 gene in P. sativum genomic DNA, (section 3.3.2) then the lack of hybridisation in northern blots of pea RNA was unlikely to be due to insufficient homology between the probe and the RNA. This reasoning strongly influenced the choice of cloning strategies designed to isolate the P. sativum E2/E3 gene.

3.5 Preparation of a P. sativum genomic library

3.5.1 The reason for generating a P. sativum genomic library

It would be preferable to obtain a cDNA clone of P. sativum E2/E3 initially rather than a genomic clone, because a cDNA clone would provide an amino acid sequence which could be compared with the bacterial aroD and aroE enzymes as well as with the S. cerevisiae E2 and E3 domains of the arom protein for homology, and so allow confirmation of the identify of the clone. In addition, with a cDNA clone, there was the possibility of expressing the plant enzyme within a bacterial host allowing complementation of auxotrophy thereby providing more evidence for the identity of the clone. The likelihood of the presence of introns within the pea E2/E3 gene would make identification of the coding sequence from a genomic clone very difficult and would certainly rule out bacterial expression.

However, because a signal was obtained from P. sativum genomic DNA and not from P. sativum RNA, using the S. cerevisiae E2/E3 probe, a genomic clone of P. sativum E2/E3 was pursued rather than a cDNA clone. It was envisaged that a genomic clone would then be used as a homologous probe to isolate a cDNA clone.

With hindsight, the most appropriate strategy would have been to pursue a cDNA clone of P. sativum E2/E3 rather than a genomic clone. Results described in Chapter 5 show that it is not necessary for a probe to identify a homologous mRNA species on a northern blot of

poly A^+ and total RNA, to be able to isolate a cDNA copy of the mRNA from a cDNA library, using the same probe. The absence of a signal on a northern blot, merely indicates that the message is present at low levels and that a larger than normal number of plaques require to be screened to find the particular cDNA clone.

3.5.2 The λL47 vector

A P. sativum genomic library was generated in the lambda replacement vector λL47 (Fig 3.8; Leonan and Brammar, 1980). The stuffer fragment can be removed from this vector by digestion with BamHI. Purification of the two phage arms from the stuffer prevents regeneration of viable phage DNA when the purified arms are ligated because the recombined DNA is too small to be packaged into phage heads. 4-17 kbp DNA fragments with the appropriate cohesive ends can be ligated between the phage arms, producing viable phage DNA.

3.5.3 Preparation of the insert DNA

In the generation of the P. sativum genomic library, the insert DNA consisted of 10-17 kbp DNA fragments isolated from a partial Sau3A digestion of P. sativum DNA. This size range of insert DNA reduced the possibility of multiple insertions of small fragments between the phage arms. Such juxtapositioning of DNA fragments which are normally non-contiguous in the genome would have created problems during the analysis of the recombinant DNA molecules.

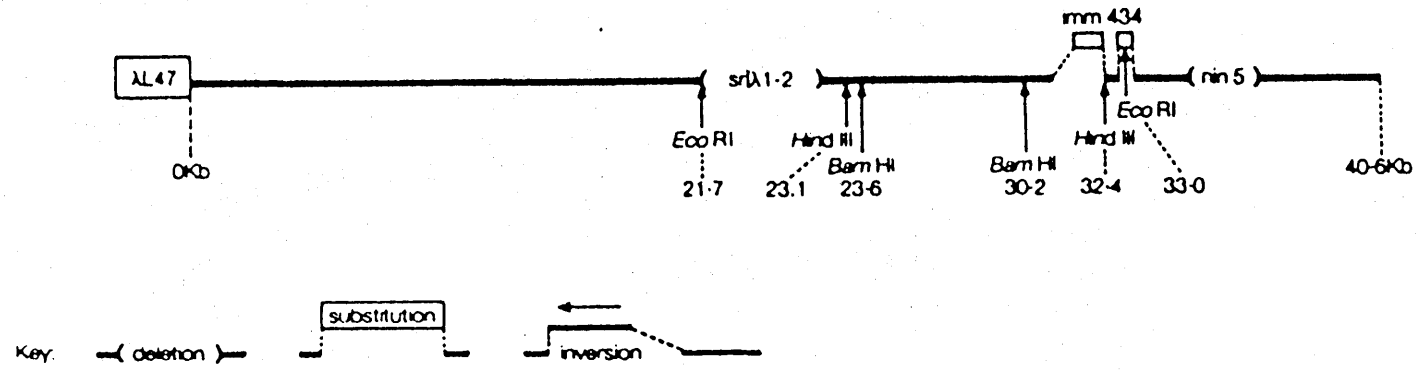


Figure 3.8 Restriction map of λL47

λL47 arms were prepared by digestion of the λL47 DNA with BamHI and separation of the left and right arms from the stuffer fragment (23.6 - 30.6). EcoRI or HindIII can also be used to allow larger inserts to be cloned (Leonan and Brammar, 1980).

Before the bulk partial Sau3A digestion of P. sativum DNA, it was necessary to establish the optimal ratio of DNA: Sau3A which would generate the largest proportion of 10-17 kbp DNA fragments. Samples of P. sativum DNA were digested with decreasing amounts of Sau3A and the partial digestion products were electrophoresed on an agarose gel (Fig 3.9). The largest proportion of 10-17 kbp fragments came from a one hour digestion of 2 μg P. sativum DNA with 0.125 units Sau3A.

The bulk partial Sau3A digestion consisted of 200 μg P. sativum DNA digested with 12.5 units Sau3A. Samples removed from the partial digestion after various times, were electrophoresed on an agarose gel (Fig 3.10). The gel was exposed to u.v. light for the minimum length of time possible in order to prevent u.v. induced mutations in the DNA. Although the test digestions had been for one hour, the optimal time at this DNA: Sau3A ratio for the bulk digestion appeared to be 30 minutes. A considerable amount of DNA had been digested to fragments smaller than 10 kbp. The 10-17 kbp DNA fragments were purified from the remainder of the partial digestion products by isolating the desired region from an l.m.p. agarose gel after electrophoresis of the partially digested DNA (Fig 3.11(i)). A sample of the isolated DNA was run on an agarose gel (Fig 3.11(ii)) which confirmed that the desired size range of fragments had been obtained.

3.5.4. Establishing optimal ligation conditions

A preparation of BamHI digested λL47 arms was available. This had been prepared by fractionating a BamHI digestion of λL47 on a sucrose

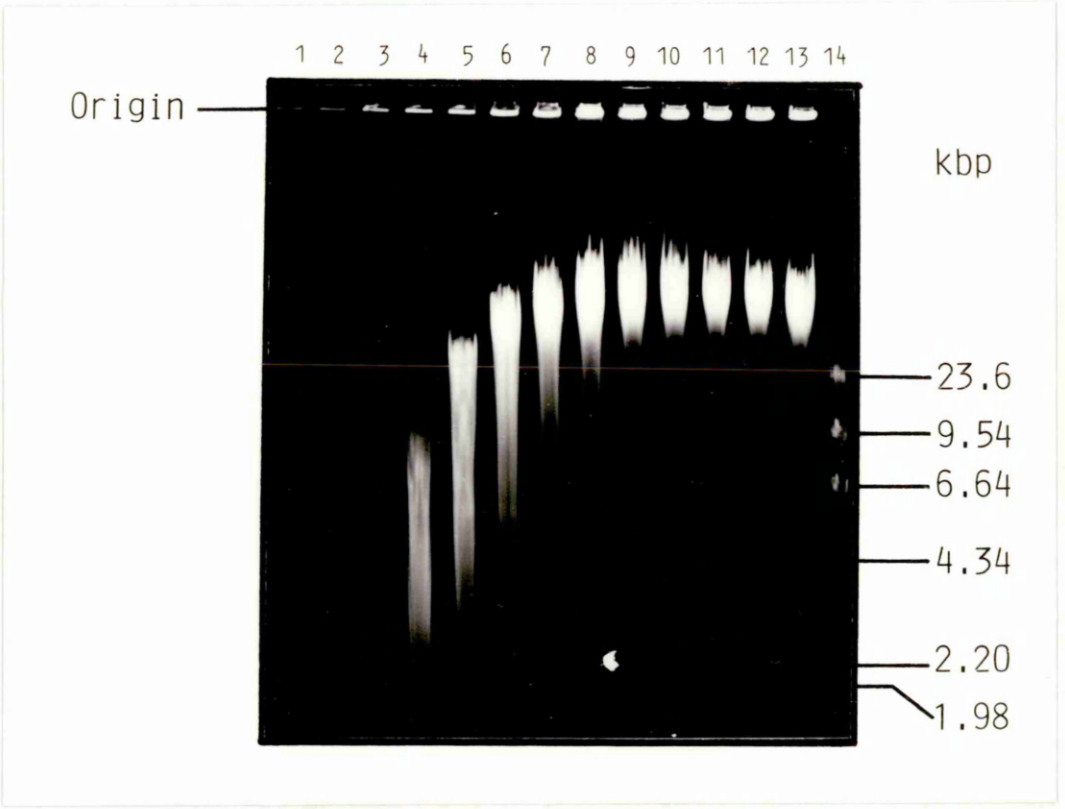


Figure 3.9 Optimisation of Sau3A partial digestion conditions

Lanes 1 to 13 contain 2 µg *P. sativum* DNA digested with the following amounts of Sau3A, for 1 hour (section 2.11b)

lane:
1. 2 units
2. 1 unit
3. 0.5 units
4. 0.25 units
5. 0.125 units
6. 0.0625 units
7. 0.0325 units
8. 0.0156 units
9. 0.0078 units
10. 0.0039 units
11. 0.00195 units
12. 0 units
13. 0 units

Lane 14 contains HindIII-digested λ-DNA

The samples were electrophoresed on a 0.4% agarose gel.

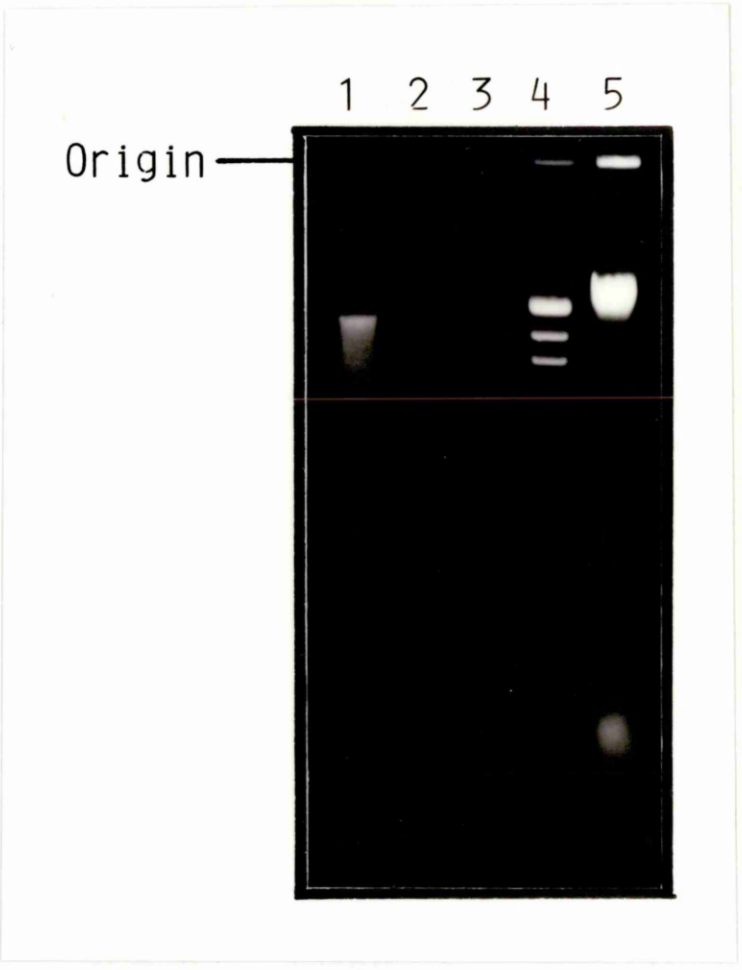
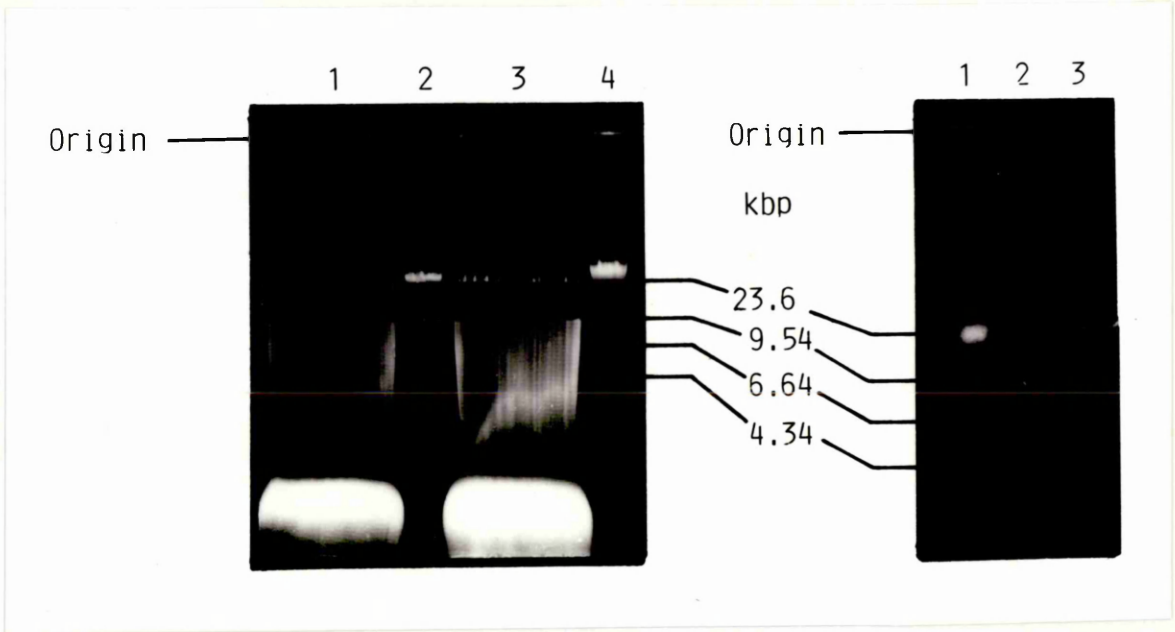


Figure 3.10 Analysis of a large scale partial Sau3A digestion of P. sativum DNA

lane: 1. Sau3A partially digested P. sativum DNA after 30 minutes
2. " " " " " " 60 minutes
3. " " " " " " 90 minutes
4. HindIII-digested λ-DNA
5. Undigested P. sativum DNA

One third of the bulk Sau3A partial digest was removed after 30 minutes, one third after 60 minutes and the remainder after 90 minutes. 1 μg samples from each sample were electrophoresed on a 0.8% agarose gel. The size marker bands (in kbp) in lane 4 are 23.6, 9.64, 6.64, 4.34, 2.26, 1.98.



Fraction 3.11 Isolation of 10-17 kbp fragments of Sau3A partially digested P. sativum DNA

(i) lane: 1. Sau3A partially digested P. sativum DNA
2. HindIII-digested λ-DNA
3. Sau3A partially digested P. sativum DNA
4. Undigested P. sativum DNA

The region containing the 10-17 kbp fragments of Sau3A partially digested P. sativum DNA was excised (lanes 1 and 3) from a 0.4% l.m.p. agarose gel. The photograph shows the gel after excision.

(ii) lane: 1. undigested P. sativum DNA
2. purified 10-17 kbp fragments of Sau3A partially digested P. sativum DNA
3. HindIII-digested λ-DNA

The samples were electrophoresed on a 0.4% agarose gel.

gradient thus separating the arms from the stuffer fragment.

It was now necessary to establish the optimal conditions from the ligation of the BamHI/ λL47 arms with the 10-17 kbp Sau3A-digested insert DNA, in other words to determine the optimal ratio of arms: insert in the ligation. Ligation conditions were required that would favour production of long concatenated DNA molecules which would in turn be packaged efficiently. If self ligation of the arms or insert occurred then fewer viable recombinant phage would be synthesised resulting in a decreased titre of the package reaction. The assay for the optimal arms: insert ratio was packaging of the ligated DNA followed by titration of the package reactions. When carrying out these optimal ratio test ligations, 'home-made' in vitro packaging mixes were used (see section 2.21.1).

A total amount of 2 μg of DNA was ligated at arms to insert ratios of 16:1, 8:1, 4:1, 2:1, 1:1, 1:2, 1:4. Samples were removed before and after ligation and analysed by electrophoresis on an agarose gel (Fig 3.12). This showed that the λL47 arms were present as two discrete bands in the unligated samples whereas the insert DNA was present as a smear which was positioned between the two bands. The insert DNA smear was most obvious at ratios of arms:inserts of 1:2 and 1:4. The insert DNA and the λL47 arms appeared to be able to ligate together. This was evident because the λL47 arms and the insert DNA smear were replaced by a larger molecular weight species after ligation had occurred. The occurrence of ligation was confirmed by ligating λ-DNA digested with Hind III (lane 13) to larger species (lane 14).

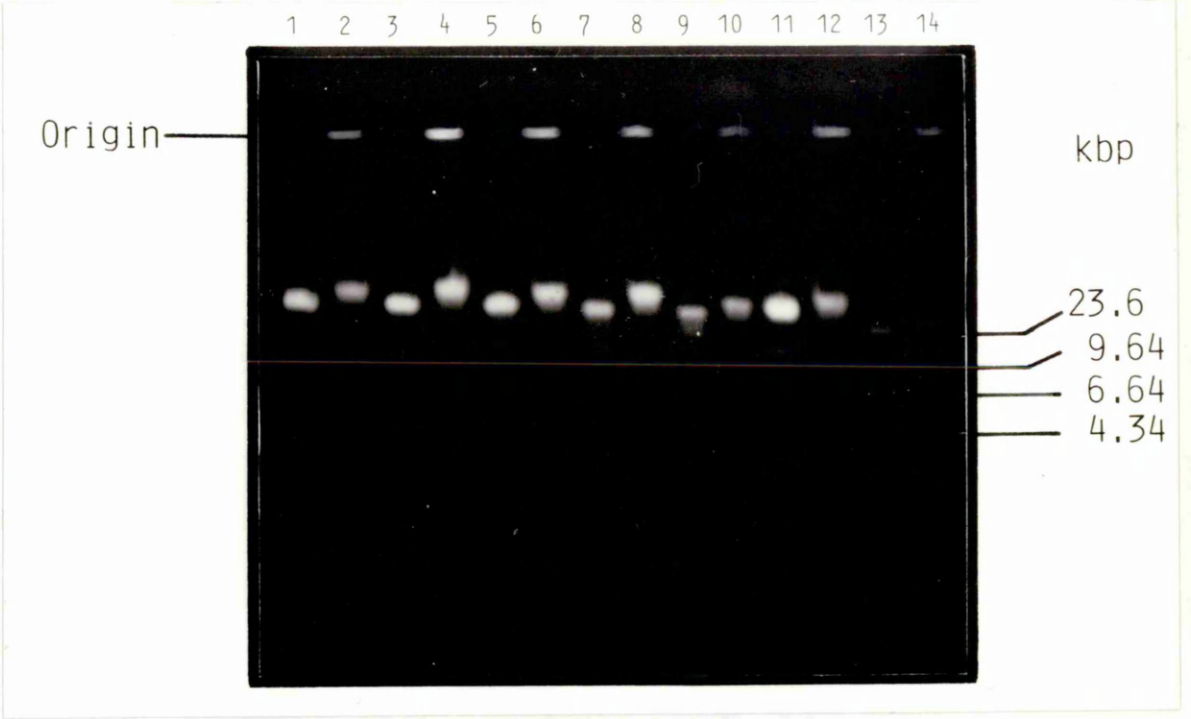


Figure 3.12 Agarose gel analysis of test ligation reactions

0.2 µg samples of DNA, at the following ratios of λL47 arms: P. sativum insert DNA, were electrophoresed before and after ligation on a 0.4% agarose gel.

lane:	1. 4:1 unligated	lane:	7. 1:2 unligated
	2. " ligated		8. " ligated
	3. 2:1 unligated		9. 1:4 unligated
	4. " ligated		10. " ligated
	5. 1:1 unligated		11. arms only unligated
	6. " ligated		12. " " ligated

Lane 13 contains HindIII-digested λ-DNA and lane 14 contains ligated HindIII-digested λ-DNA.

When these ligations were packaged and the packaging reactions were used to infect E. coli LE392, the titre results displayed in Table 3.1 were obtained. The homemade packaging mixes which were used exhibited extremely low efficiency in packaging the test ligations. Repetition of the experiment gave similar results and it was concluded that an arms:insert ratio of 4:1 was optimal. These tests also indicated the importance of (i) precipitating the DNA out of the ligation mix before in vitro packaging and (ii) using highly purified insert DNA. If either of these precautions were omitted, the titre dropped to zero.

3.5.5 Ligation and in vitro packaging

The number of recombinants required in order to have a single copy gene represented in a library at a particular probability can be calculated from the following formula:

$$N = \frac{\ln (1-P)}{\ln (1-\frac{x}{y})}$$ (Clarke and Carbon, 1976)

where P is the desired probability, x is the insert size, y is the haploid genome size and N is the necessary number of recombinants.

Since the insert sizes were 10-17 kbp, and P. sativum has a haploid genome size of 4.5×10^9 bp then 0.8×10^6 to 1.3×10^6 plaque forming units (pfu) were required in order to have a 95% probability of finding a particular gene represented in a P. sativum genomic library.

Table 3.1 Titre of the test ligations

The titre obtained from the in vitro packaging of ligated λL47 arms and P. sativum DNA insert at a variety of arms: insert ratios.

arms: insert ratio in ligations used for in vitro packaging	pfu per 100 µl of packaging reaction at the following dilutions 10^0 10^{-1} 10^{-3} 10^{-5}	titre (pfu/ml)
16:1	11 3	10^2
8:1	20 3	2×10^2
4:1	35 3	3×10^2
2:1	- -	-
1:1	- -	-
1:2	- -	-
1:4	- -	-
Controls		
No DNA	-	-
0.5 µg undigested λ-DNA	323	3×10^6 pfu/ml
λgt WES phage stock	437	4×10^8 pfu/ml

It was anticipated that using commercially available *in vitro* packaging reactions, would increase the efficiency of packaging dramatically, compared to the homemade packaging reactions and would provide a sufficient number of plaques to form a complete *P. sativum* genomic library. This turned out to be the case. Using five Amersham *in vitro* packaging reactions to package ligations containing 2 μg of DNA at a ratio of 4:1 (BamHI / λL47 arms:10-17 kbp Sau3A partially digested *P. sativum* DNA) a total of 1.2×10^6 pfu were obtained (Table 3.2). This was calculated to be sufficient to give a 95% probability of the library containing the *P. sativum* E2/E3 gene.

Inhibitory components present in the packaging reactions can adversely affect the titre. Therefore the recombinant phage were precipitated out of the packaging reactions using PEG with the intention of increasing the titre of the purified recombinant phage. However instead of increasing, the titre dropped to 1.1×10^6 pfu. Therefore, the only advantage gained by carrying out the PEG precipitation was the concentration of the recombinant phage.

The 1.1×10^6 pfu which had been generated were thought to be sufficient to give a 95% probability of obtaining the desired gene although the whole library would have had to be screened in one attempt.

To improve the chances of success it was decided to amplify the library. This would allow many screening attempts to be carried out. The titre of the resultant 140 ml of amplified library (Table 3.2),

Table 3.2 Generation of the P. sativum genomic library

The titre obtained from the five in vitro packaging reactions used to generate the P. sativum genomic library and also the titre of the amplified library.

package reaction number	pfu per 100 µl of packaging reaction at the following dilutions 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6}	titre (pfu/ml)
1	240 41	3×10^5
2	200 20	2×10^5
3	411 50	5×10^5
4	65 8	0.7×10^5
5	127 15	1.2×10^5
		Total 11.9×10^5
amplified library	4400 330	3.3×10^5

was 3.3 x 10^9 pfu per ml. This library should maintain a fairly high titre for several years and it was hoped that it would be a resource from which many genes could be obtained. One drawback of such an amplified library is that the sequence content of the amplified library will reflect the differential growth of recombinants during amplification. The number of clones to be screened in order to obtain a sequence that is under represented by such differential growth will therefore increase.

3.5.6 Screening with the S. cerevisiae E2/E3 probe

Aliquots of 10^6 pfu from the amplified library were screened with the S. cerevisiae E2/E3 probe. The stringency used each time varied as follows:

(a) hybridisation:- 30% formamide, 42°C
washing:- 2 x SSC, 50°C

(b) hybridisation:- 50% formamide, 42°C
washing:- 2 x SSC, 60°C

(c) hybridisation:- 50% formamide, 42°C
washing:- 2 x SSC, 37°C

(d) hybridisation:- 50% formamide, 30°C
washing:- 2 x SSC, 50°C

A high background was obtained using the low stringency screening conditions (a). Perhaps a signal was being masked behind this background so stringency (b) was used. At this stringency a signal had been obtained in the P. sativum genomic southern blots (section 3.3.2). However even after two screenings, no positive signal was obtained. Stringencies were reduced to (c) then ultimately to (d) but still no positive signals were obtained. Throughout the screenings, several false positive signals were obtained but on re-screening, these proved to be negative.

No positive clones were obtained from the P. sativum genomic library therefore it was assumed that the library was incomplete and that it did not contain the P. sativum E2/E3 gene.

3.6 Preparation of enriched P. sativum genomic libraries

3.6.1 Reasons for the preparation of these libraries

Further attempts to clone the whole P. sativum E2/E3 gene using the heterologous S. cerevisiae probe were discontinued and instead partial P. sativum genomic libraries were prepared. These were extensively enriched for the EcoRI band that hybridised with the S. cerevisiae E2/E3 probe in P. sativum genomic southern blots (section 3.3.2). The hope was that the resultant clone, although it would not contain the whole P. sativum E2/E3 gene, would be useful as a completely homologous probe in subsequent efforts to obtain either a full length genomic clone or a cDNA clone.

3.6.2. Preparation of the insert DNA

200µg of EcoRI-digested P. sativum DNA was fractionated on a sucrose gradient (10-40% sucrose, volume 14ml). Samples of DNA from each 1 ml fraction were electrophoresed on an agarose gel (Fig 3.13a). This confirmed that most of the higher molecular weight DNA was absent from the later fractions. The southern blot of this gel was probed with the S. cerevisiae E2/E3 probe at high stringency and the autoradiograph (Fig 3.13b), indicated that fraction-10 contained most of both the 0.9 kbp and 1.5 kbp bands. Therefore a significant enrichment of the E2/E3 bands had taken place.

However fraction-10 still contained a large proportion of DNA outwith the desired size range (Fig 3.13a) and so the DNA from fraction-10 was fractionated on a second sucrose gradient (10-25% sucrose, volume 4ml). The maximum sucrose concentration in the second gradient was not as high as it had been in the first gradient because little useful DNA was present in the fractions of the first gradient where the sucrose concentration was above 25% (Fig 3.13a). Samples of DNA from each fraction of the second gradient were electrophoresed on an agarose gel (Fig 3.14a) which confirmed that fractionation had occurred. The southern blot of this gel was probed with the S. cerevisiae E2/E3 probe at high stringency and the autoradiograph (Fig 3.14b) indicated that fraction 10.05 contained both the 0.9 kbp and the 1.5 kbp bands whereas fraction 10.04 contained predominantly the 0.9 kbp band. Fraction-14 DNA, from the first gradient was included as a control to produce a definite signal from the 0.9 kbp band.

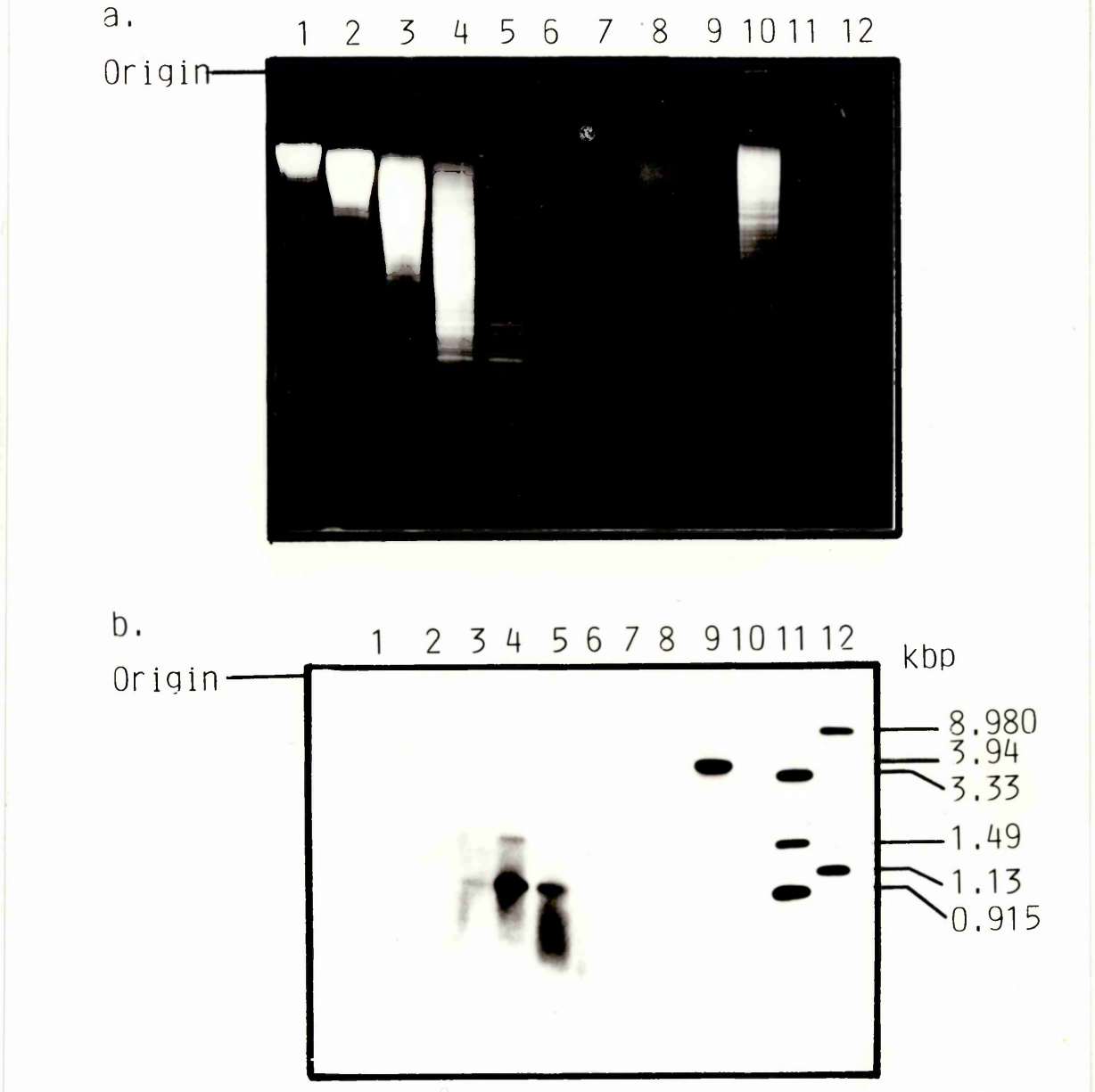


Figure 3.13 First sucrose gradient fractionation of P. sativum DNA

(a) Samples containing 2% of the EcoRI-digested P. sativum DNA present in fractions 7-14 from the 14 ml 10-40% sucrose gradient were electrophoresed on a 0.8% agarose gel.

lane:	1.	fraction- 7 DNA	
	2.	" - 8 "	
	3.	" - 9 "	
	4.	" -10 "	
	5.	" -11 "	
	6.	" -12 "	
lane:	7.	fraction-13 DNA	
	8.	" -14 "	
	9.	DraI-digested pME173	
	10.	EcoRI-digested total P. sativum DNA	
	11.	EcoRI-digested pME173	
	12.	BglII-digested pME173	

(b) Southern blot of gel (a) hybridised with the S. cerevisiae E2/E3 probe at 42°C in hybridisation solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

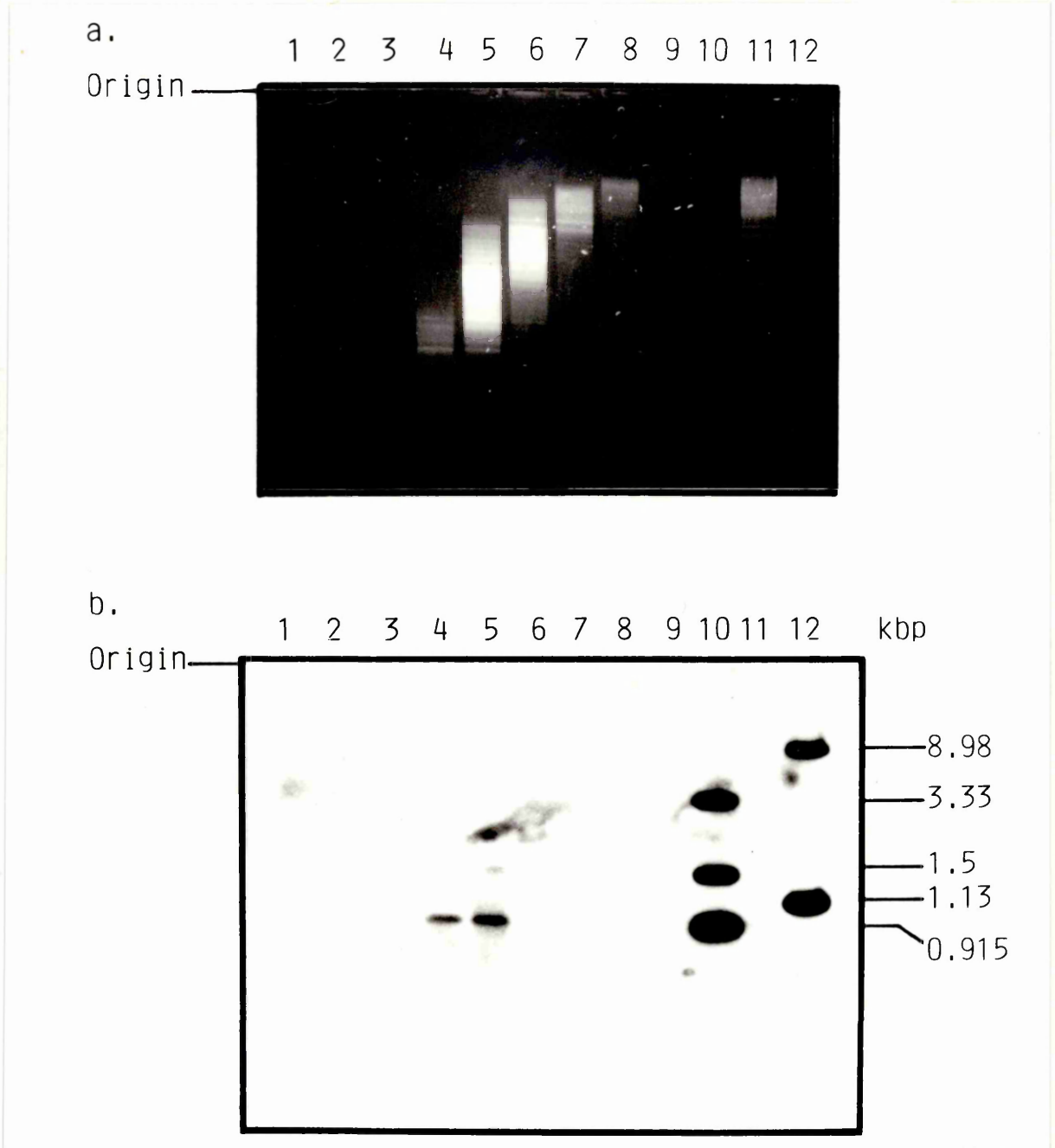


Figure 3.14 Second sucrose gradient fractionation of P. sativum DNA

(a) Samples containing 20% of the EcoRI digested P. sativum DNA present in fractions 10.01 to 10.09 from the 4 ml 10-25% sucrose gradient, were electrophoresed on a 0.8% agarose gel.

lane: 1.	fraction 10.01 DNA	lane: 7.	fraction 10.07 DNA
2.	" 10.02 "	8.	" 10.08 "
3.	" 10.03 "	9.	" 10.09 "
4.	" 10.04 "	10.	EcoRI-digested pME173
5.	" 10.05 "	11.	fraction ~ 14 DNA
6.	" 10.06 "	12.	BglII-digested pME173

(b) Southern blot of gel (a) hybridised with the S. cerevisiae E2/E3 probe at 42°C in hybridisation solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

There was very little extraneous DNA present in fraction 10.04 but there was still a significant amount present in fraction 10.05 (Fig 3.14a). In order to remove this extraneous DNA, the DNA from fraction 10.05 was electrophoresed on a 0.8% l.m.p agarose gel and fragments of around 0.9 kbp (named 10.05a) and fragments of around 1.5 kbp (named 10.05b) were isolated.

A total of 200 ng of DNA was present in fraction 10.04, and 20 ng in each of fractions 10.05a and 10.05b. These samples were used to construct the enriched genomic libraries.

3.6.3 The λgt10 vector

A lambda vector was used for the cloning of the fractionated *P. sativum* DNA because it was likely that high efficiency *in vitro* packaging would be required in order to ensure that the desired clones were present in the enriched libraries, even though the insert DNA had been enriched with the 0.9 kbp and 1.5 kbp DNA species.

The vector chosen was the insertion vector λgt10 because this contains a single EcoRI site, into which up to 7.6 kbp of DNA can be inserted (Fig 3.15; Huynh *et al.*, 1985). λgt10 also has an efficient biological selection for recombinants. When DNA is inserted into the EcoRI site, the phage repressor gene (cI) is inactivated. The inactivation of this repressor prevents lambda entering the lysogenic pathway and so forces it into the lytic pathway. Recombinant phage

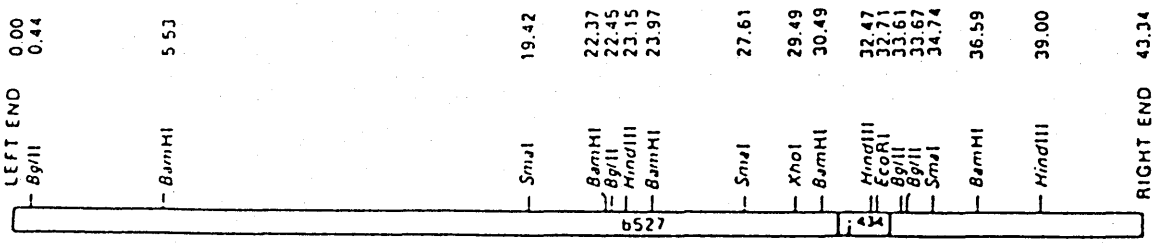


Figure 3.15 Restriction map of λgt10

Restriction endonuclease sites are designated in kbp from the left end. DNA is inserted at the unique EcoRI site at position 32.71 (Huynh et al., 1985).

(cI^-) therefore favour the lytic pathway and produce clear plagues on wild type E. coli (eg E. coli C600) whereas parental λgt10 (cI^+) can enter the lysogenic pathway and therefore produce turbid plaques on wild type E. coli. On E. coli strains with a high frequency lysogeny (hfl) mutation (eg E. coli C600 hfl) the parental λgt10 is forced into the lysogenic pathway and so does not cause cell lysis. On the other hand, infection of hfl mutants with recombinant λgt10 (cI^-) results in cell lysis since entry into the lysogenic pathway is prohibited. By plating λgt10 libraries on hfl strains, recombinants will form plaques but non-recombinants will not (Huynh et al., 1985).

3.6.4 Ligation and in vitro packaging

Only 10 μg of calf intestinal phosphatase-treated λgt10 arms was available and so tests to establish optimal arms : insert ratios were not carried out.

Equimolar amounts of DNA from fractions 10.04, 10.05a and 10.05b were ligated separately with λgt10 EcoRI arms. These ligations were then subjected to in vitro packaging using Stratagene packaging mixes. The titre of each packaging reaction was determined (Table 3.3).

If a complete genomic library had been synthesised using 1 kbp fragments instead of a partial library then 13.6×10^6 recombinants would need to be screened to give a 95% probability of finding the desired gene. The number of recombinants required to give a 95%

Table 3.3 Generation of the enriched P. sativum genomic libraries

The titre obtained from the in vitro packaging of ligated λgt10 arms and fractionated EcoRI-digested P. sativum DNA.

insert DNA present with λgt10 in ligations	pfu per 100 µl of packaging reaction at the following dilutions			titre (pfu/ml)
	10^{-3}	10^{-4}	10^{-5}	
10.04	700	70	9	3.5×10^6
10.05 a	96	9	-	4.5×10^5
10.05 b	281	18	2	1.4×10^6

probability of finding a gene in an enriched library depends on the enrichment factor. The quantitation of enrichment is difficult. However the ratio of DNA obtained after fractionation compared with the DNA present in the original digestion should give a rough estimate of enrichment. For the three enriched libraries prepared in λgt10 these ratios are for the library from fraction 10.04 (200 ng/200 μg) 1/1000, and for the libraries from both fractions 10.05a and 10.05b (20 ng/200 μg) 1/10,000. Thus the appropriate enrichment for the E2/E3 bands in the libraries from fraction 10.04 is 1000 and in the libraries from fractions 10.05a and 10.05b it is 10,000. The actual enrichment is likely to be lower than this because of DNA loss during fractionation.

It was estimated that the enrichment factor could be as low as one tenth of the approximate values indicated above. Assuming an enrichment factor of 100 for fraction 10.04, and 1000 for fractions 10.05a and 10.05b, then only 1.36×10^5 recombinants from the 10.04 library and 1.36×10^4 recombinants from the 10.05a and 10.05b libraries needed to be screened to have a 95% probability of finding the desired clones. In the event, a significantly larger number of plaques from each library were screened.

3.6.5. Screening the enriched libraries with the S.cerevisiae E2/E3 probe.

6×10^5 pfu, 3×10^5 pfu and 4×10^5 pfu from the 10.04, 10.05a and 10.05b libraries respectively, were screened at high stringency with the S. cerevisiae E2/E3 probe. A total of 24 positive signals were obtained from the 10.04 library, 21 from the 10.05a library and 3 from the 10.05b library (Fig 3.16).

The positive signals from the 10.04 and 10.05a libraries, were almost certainly due to the 0.9 kbp fragment because the DNA used in the synthesis of these libraries predominantly contained this band (section 3.6.3). Two positively hybridising plaques from the 10.04 library (named λ6 and λ7) and two positively hybridising plaques from the 10.05a library (named λ1 and λ2) were purified to homogeneity.

The positive signals from the 10.05b library were likely to be due to hybridisation to the 1.5 kbp fragment. Three positively hybridising plaques (named λ3, λ4 and λ5) were purified to homogeneity from this library.

DNA was isolated from each of the seven purified phage. It was digested with EcoRI and the digestion products were electrophoresed on an agarose gel (Fig 3.17a). The two high molecular weight DNA fragments present in each of the tracks containing phage DNA, are the λgt10 arms. Below them are the insert fragments. The southern blot of this gel was probed with the S. cerevisiae E2/E3 probe (Fig 3.17b) and

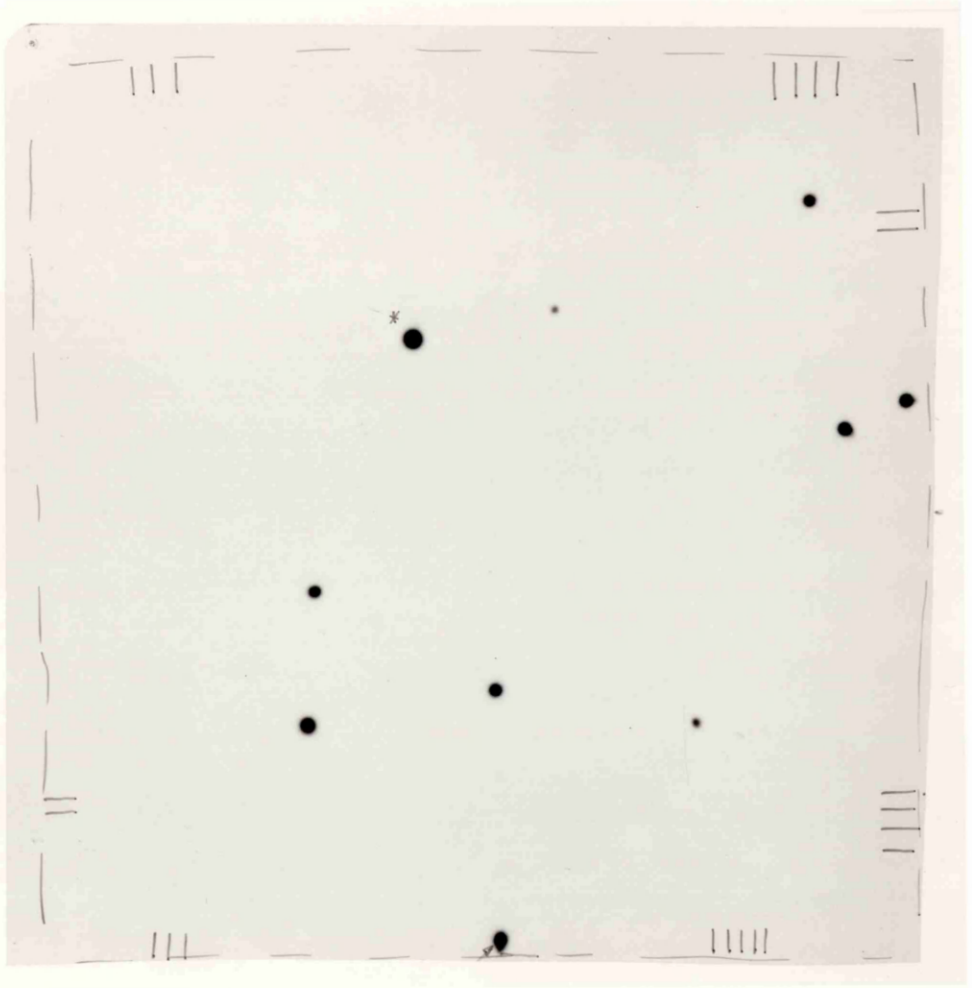


Figure 3.16 Autoradiograph of an E2/E3 screened filter

An autoradiograph of an impression from a 23 x 23 cm plate containing 10^5 plaques from the 10.05a library hybridised with the S. cerevisiae E2/E3 probe. The filter was hybridised at 42°C in solution containing 50% formamide and washed at 50 C in 2 x SSC, 0.1% SDS. The spot labelled with an asterisk was purified to homogeneity (section 2.21.4).

Figure 3.17 Analysis of recombinant phage DNA which hybridised with the E2/E3 probe

(a) Agarose gel analysis of EcoRI digested phage DNA

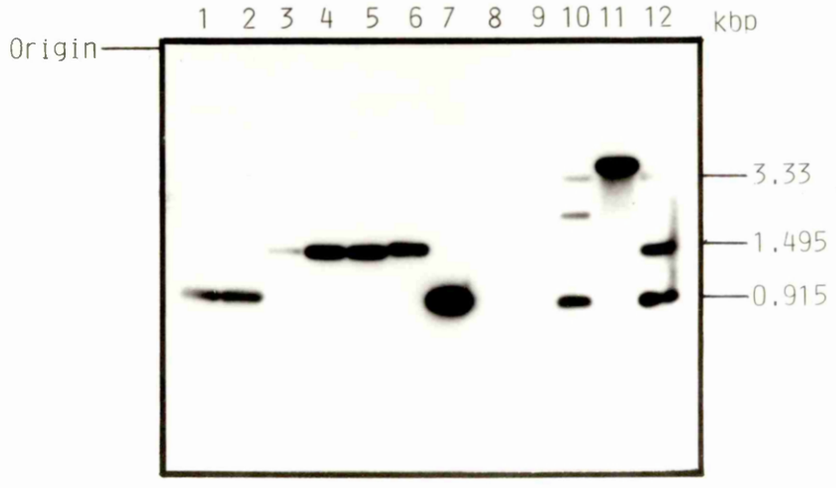
lane:
1. EcoRI-digested DNA from λ1 (isolated from 10.05a library)
2. " " " " λ2 " " " "
3. " " " " λ3 " " 10.05b "
4. " " " " λ4 " " " "
5. " " " " λ5 " " " "
6. " " " " λ6 " " 10.04 "
7. " " " " λ7 " " " "
8. HindIII-digested λ-DNA
9. -
10. 3 μg EcoRI-digested S. cerevisiae DNA
11. DraI-digested pME173
12. EcoRI-digested pME173

The samples were electrophoresed on a 0.8% agarose gel.

(b) Southern blot of gel (a) hybridised with the S. cerevisiae E2/E3 probe at 42°C in hybridisation solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

a.


1 2 3 4 5 6 7 8 9 10 11 12 kbp
Origin
3.33
1.495
0.915


1 2 3 4 5 6 7 8 9 10 11 12 kbp
Origin
3.33
1.495
0.915

it confirmed that the insert fragments were homologous to the probe. The fragments which hybridised were 0.9 kbp and 1.5 kbp in size and so it appeared that the intended fragments had indeed been cloned.

Interestingly, of the two clones chosen for analysis from the 10.04 library (Fig 3.17a) one contained the 0.9kbp fragment and the other contained the 1.5 kbp fragment. This would not have been predicted because the 10.04 fraction of DNA, from which the 10.04 library had been synthesised, had contained predominantly the 0.9 kbp band (Fig 3.14b). In addition, one of these clones (lane 6, Fig 3.17a) contained a second insert DNA fragment which did not hybridise with the S. cerevisiae E2/E3 probe. This fragment had obviously ligated into the EcoRI insertion site of λgt10 along with the 1.5 kbp fragment. This is an example of a multiple insertion (see Section 3.5.1).

Throughout the heterologous probing experiments using the S. cerevisiae E2/E3 probe, it had been evident that the 0.9kbp band from EcoRI-digested P. sativum DNA, co-migrated with the 0.9kbp band from EcoRI-digested S. cerevisiae DNA. This was thought to be a coincidence especially when there was no similar co-migration of the 1.5 kbp band from EcoRI-digested P. sativum DNA (see Fig 3.4). However EcoRI-digested pME173 had been included on the southern blot (Fig 3.17b). Unfortunately, both the 0.9kbp fragment and the 1.5 kbp fragment - both insert fragments from the EcoRI-digested phage DNA which hybridised with the S. cerevisiae probe - co-migrated with two of the EcoRI digested pME173 fragments which hybridised with the S.

cerevisiae E2/E3 probe. This introduced the worry that what had been cloned were two fragments of pME173 and not two fragments of P. sativum genomic DNA which were homologous to the S. cerevisiae E2/E3 probe.

To investigate the relationship of the cloned 0.9 kbp and 1.5 kbp fragments with pME173, each of the cloned fragments was sub-cloned into pUC18 creating pSGE2/E3-1 and pSGE2/E3-2 respectively. These plasmids were digested with appropriate restriction enzymes and the digests were electrophoresed on agarose gels (Fig 3.18). The restriction patterns on the gels, exactly matched the predicted digestion pattern of the 0.9 kbp and 1.5 kbp fragments from EcoRI digested pME173. Sequencing of the pSGE2/E3-1 and pSGE2/E3-2 inserts conclusively proved that the two bands which had been cloned from P. sativum DNA were EcoRI fragments of pME173.

At some point during the preparation of P. sativum DNA, the DNA must have become contaminated with pME173. This was disastrous because the S. cerevisiae E2/E3 probe had been derived from this plasmid and as a result hybridised strongly to the contaminating plasmid. The contamination was not limited to a single DNA preparation as signals had been obtained from southern blots containing P. sativum DNA from several different preparations. Therefore contamination was narrowed down to DNA preparation solutions. A southern blot, containing P. sativum DNA used to prepare the enriched libraries and also various samples of P. sativum DNA prepared using new solutions, was probed with the S. cerevisiae E2/E3 probe (Fig 3.19(a)). From this it was

Figure 3.18 Analysis of subclones from the recombinant phage

(a) Agarose gel analysis of restriction digests of pSGE2/E3-1 (0.9 kbp insert) and pSGE2/E3-2 (1.5 kbp insert)

gel 1 lane:
1. EcoRI + AccI-digested pSGE2/E3-1
2. " + BalI " "
3. " + HincII " "
4. " + XbaI " "
5. " + PstI " "
6. HindIII-digested λ-DNA
7. HindIII + EcoRI-digested λ-DNA
8. EcoRI-digested pSGE2/E3-1
9. AccI- " "
10. HincII- " "
11. XbaI- " "
12. PstI- " "

gel 2 lane:
1. EcoRI + AccI-digested pSGE2/E3-2
2. " + EcoRV- " "
3. " + BalI-digested "
4. HindIII-digested λ-DNA
5. HindIII + EcoRI cut λ-DNA
6. EcoRI-digested pSGE2/E3-2
7. HindIII- " "
8. AccI- " "
9. EcoRV- " "

For easy interpretation of these results it should be noted that all the restriction enzymes used, except BalI and EcoRV cut pUC18 only in the polylinker. There are no BalI or EcoRV sites in pUC18.

Fragments from HindIII-digested λ-DNA (kbp):- 23.6, 9.64, 6.64, 4.34 2.26, 1.98, 0.56, 0.14

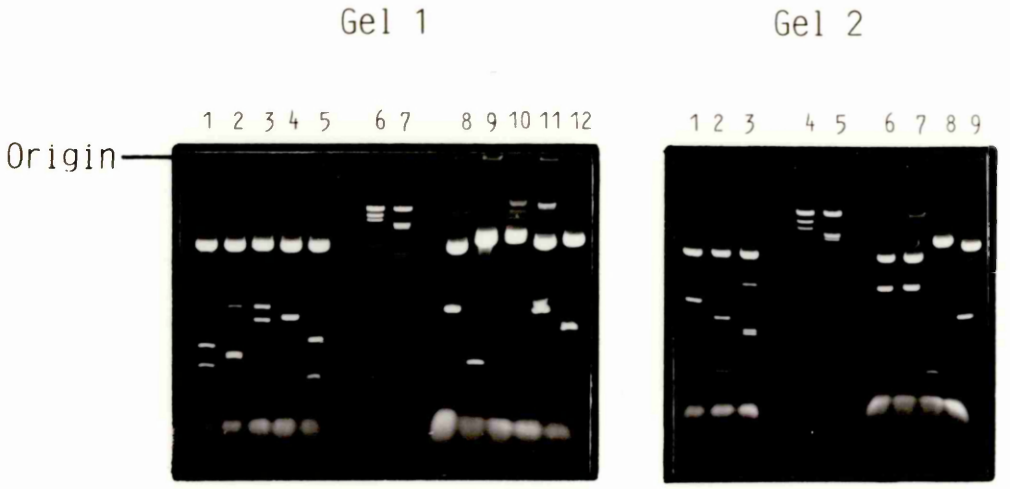
Fragments from HindIII + EcoRI-digested λ-DNA (kbp):- 21.7, 5.24, 5.05, 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58, 0.14

(b) Restriction map of 0.9 kbp and 1.5 kbp EcoRI fragments of pME173 (Duncan *et al.*, 1987; 1988)

Key to restriction sites: A=AccI, E=EcoRI, L=BalI, N=HincII, P=PstI, X=XbaI.

Relevant restriction sites in the (i) 0.9 kbp and (ii) 1.5 kbp EcoRI fragment of pME173. The 0.9 kbp fragment is composed solely of *arom* coding DNA. The 1.5 kbp fragment contains *arom* coding DNA, *S. cerevisiae* non-coding DNA and pAT153 DNA (from BamHI site to the EcoRI site).

a.

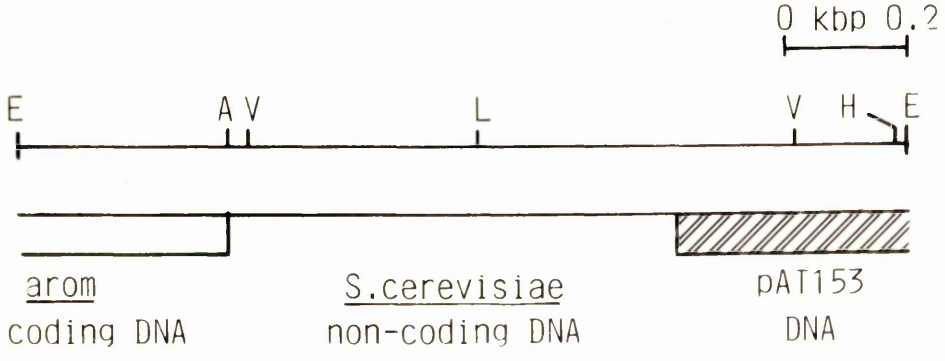

Gel 1
Gel 2
Origin
1 2 3 4 5 6 7 8 9 10 11 12
1 2 3 4 5 6 7 8 9

b.

(i)


0 kbp 0.2
E N A L P X E

(ii)


0 kbp 0.2
E A V L V H E
arom
coding DNA
S.cerevisiae
non-coding DNA
pAT153
DNA

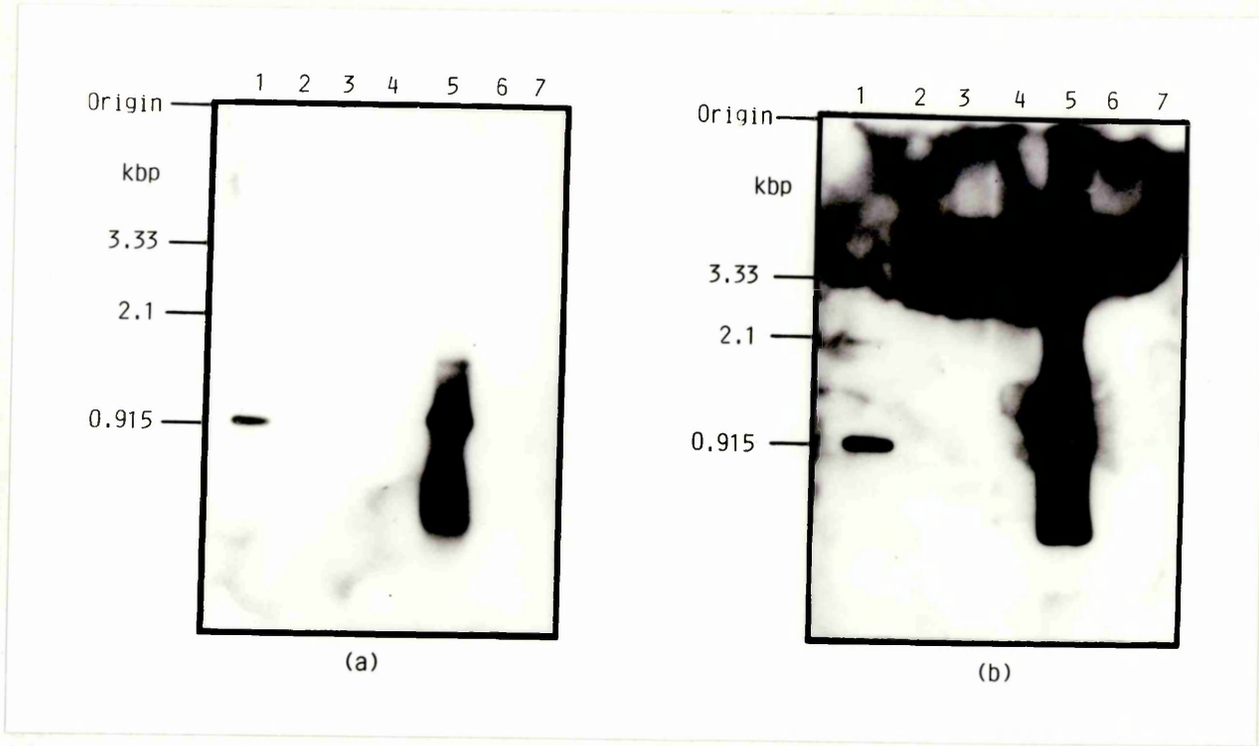


Figure 3.19 Analysis of hybridisation of various P. sativum DNA preparations with the S. cerevisiae E2/E3 probe

(a) Southern blot analysis under high stringency conditions of several P. sativum DNA preparations.

lane: 1. EcoRI-digested S. cerevisiae DNA
2. " " P. sativum DNA
3. " " " "
4. " " " "
5. " " " " (contaminated with pME173)
6. BglII- " " "
7. EcoRV- " " "

The P. sativum DNA present in lanes 2, 3, 4, 6 and 7 was prepared using fresh solution whereas the P. sativum DNA present in lane 5 was the older stock which had been used in the preparation of the enriched libraries.

15 µg samples of digested P. sativum DNA and 3 µg samples of digested S. cerevisiae DNA were electrophoresed on a 0.8% agarose gel and the southern blot was hybridised with the S. cerevisiae E2/E3 probe at 42°C in solution containing 50% formamide and washed at 65°C in 0.1 x SSC, 0.1% SDS.

(b) Southern blot analysis under low stringency conditions of several pea DNA preparations.

The southern blot of gel (a) was rehybridised with the S. cerevisiae E2/E3 probe at 42°C in solution containing 30% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

obvious that the 0.9 kbp and 1.5 kbp bands were present in the track containing older P. sativum DNA but not the newly prepared P. sativum DNA. In addition there was no hybridisation of uncontaminated P. sativum DNA with the probe at this high stringency.

The true hybridisation pattern at low stringency conditions of digested P. sativum DNA with the S. cerevisiae E2/E3 probe is shown in (Fig 3.19b). A smear was present in each of the lanes containing digested P. sativum DNA without any obvious banding pattern. This indicated that there was limited homology between the S. cerevisiae E2/E3 probe and the P. sativum DNA.

3.7 The hybridisation of other microbial probes with P. sativum DNA

3.7.1 DNA sequences used as probes

To investigate the homology of other microbial shikimate pathway genes with P. sativum DNA, several probes were derived. Naturally occurring, convenient restriction enzyme sites were utilized.

An S. cerevisiae E1 probe was isolated as an 852 bp DraI-EcoRI fragment from pME173. This fragment contained 70% of the DNA coding for the E1 domain of S. cerevisiae arom because it lacked 171 bp of upstream and 186 bp of downstream coding DNA.

An S. cerevisiae E5 probe was isolated as a 1084 bp NdeI fragment from pME173. This fragment contained 75% of the DNA coding for the E5

domain of S. cerevisiae arom because it lacked 154 bp of upstream and 211 bp of downstream coding DNA.

pGM602 contains the aroC gene from E. coli which encodes the shikimate pathway enzyme chorismate synthase (White et al., 1988). An aroC probe was isolated as a 1250 bp NruI-ClaI fragment from pGM602. The whole aroC gene was encoded in this fragment along with 50 bp of upstream and 130 bp of downstream non-coding DNA.

3.7.2 P. sativum genomic Southern blots probed with several microbial probes

P. sativum genomic DNA (which was uncontaminated with pME173) was digested with various restriction enzymes. The digests were electrophoresed on agarose gels and these gels were then subjected to Southern blot analysis.

One of the blots was probed with the S. cerevisiae E1 probe at low stringency. This resulted in a background smear in each track containing P. sativum DNA (Fig 3.20). Discrete bands were also observed in each of the tracks. Their positions were not the same when different enzymes were used to digest the P. sativum DNA.

A similar Southern blot was probed with the S. cerevisiae E5 probe at low stringency and again a background smear was observed in each track containing digested P. sativum DNA (Fig 3.21). Discrete bands were this time less obvious than in E1 probed blots but they were still

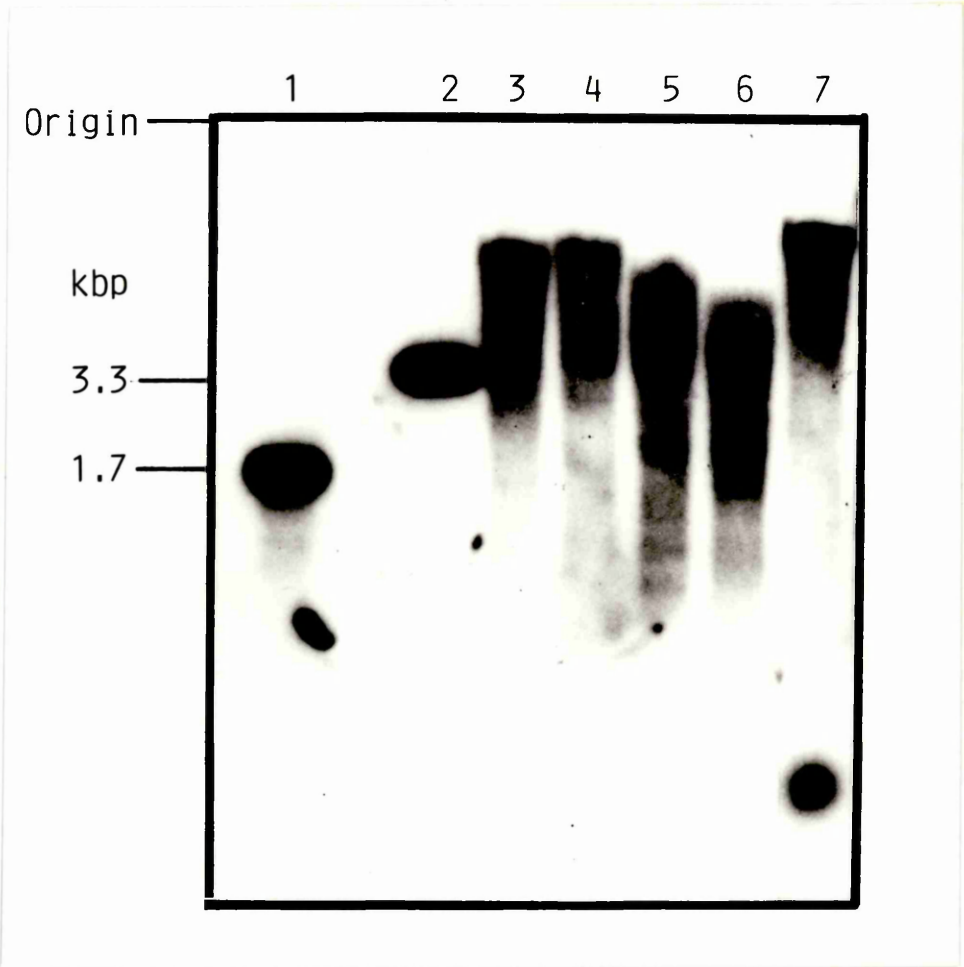


Figure 3.20 Southern analysis of P. sativum DNA hybridised with the S. cerevisiae E1 probe

lane: 1. DraI-digested pME173
2. EcoRI- " "
3. BglII- " "
4. BamHI- " P. sativum DNA
5. EcoRI- " " "
6. HindIII- " " "
7. SstI- " " "

15 µg samples of digested P. sativum DNA were electrophoresed on a 0.8% agarose gel and the southern blot was hybridised with the S. cerevisiae E1 probe at 42°C in hybridisation solution containing 30% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

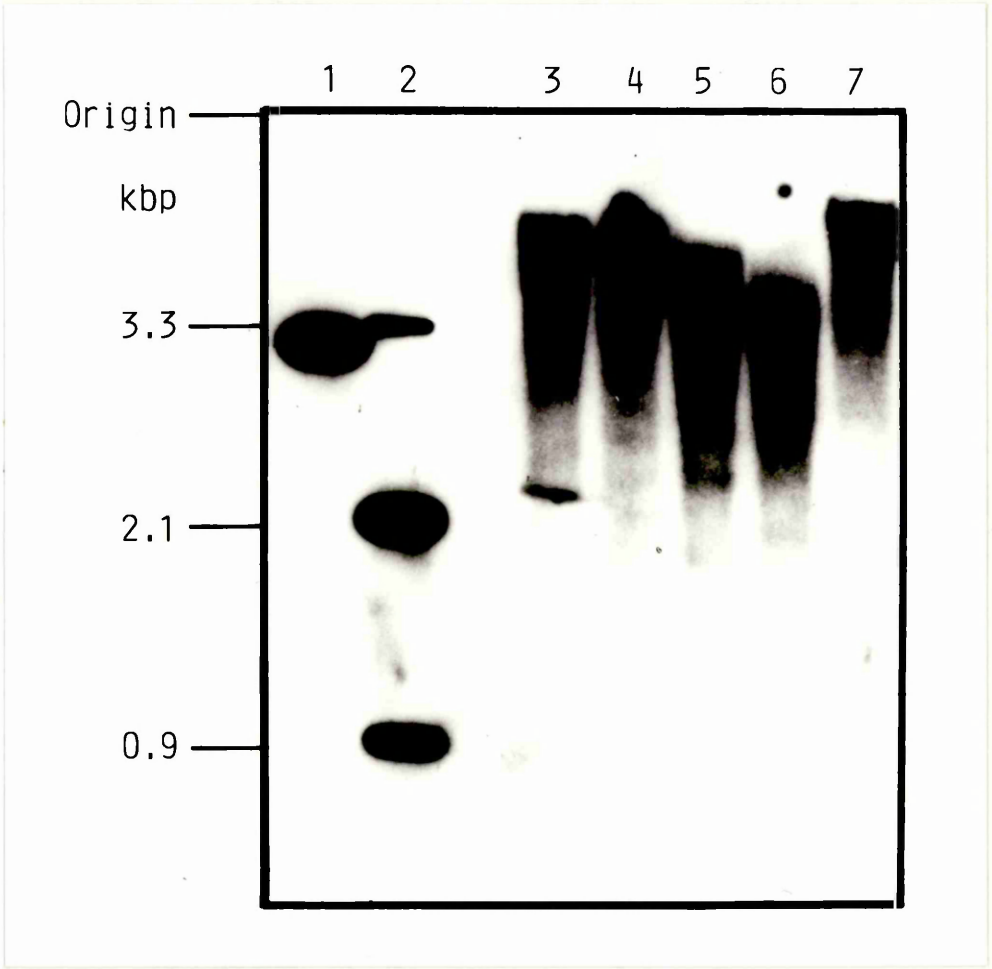


Figure 3.21 Southern analysis of P. sativum DNA hybridised with the S. cerevisiae E5 probe

lane:
1. EcoRI-digested pME173
2. DraI- " "
3. BglII- " P. sativum DNA
4. BamHI- " " "
5. EcoRI- " " "
6. HindIII- " " "
7. SstI- " " "

15 µg samples of digested P. sativum DNA were electrophoresed on a 0.8% agarose gel and the southern blot was hybridised with the S. cerevisiae E5 probe at 42°C in solution containing 30% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

present.

The discrete bands present in the digested P. sativum DNA tracks of the southern blots probed with S. cerevisiae E1 and to a lesser extent with the E5 probes were possibly due to hybridisation to the corresponding P. sativum genes. Alternatively the bands could represent repetitive elements within the P. sativum genome which were sufficiently homologous to the probe to give positive signals. Limited homology to a probe could result in a smear in a southern blot. On the other hand vaguely homologous repetitive elements would appear as discrete bands rather than a smear because the repetitive nature of the DNA would limit the restriction fragments to one specific size. The high concentration of these specific fragments would mean that even the slightest homology would be picked up as a single band in a southern blot probed at low stringency.

Another southern blot containing digested P. sativum DNA was probed with the E. coli aroC probe at low stringency (Fig 3.22). The digested P. sativum DNA did not hybridise at all with the aroC probe. However digested S. cerevisiae DNA had been included in this blot and discreet bands in tracks containing this DNA hybridised with the aroC probe. These bands changed position when digestion was with a different enzyme. Perhaps the E. coli aroC probe could have been used to isolate clones of the S. cerevisiae chorismate synthase gene however this was not pursued.

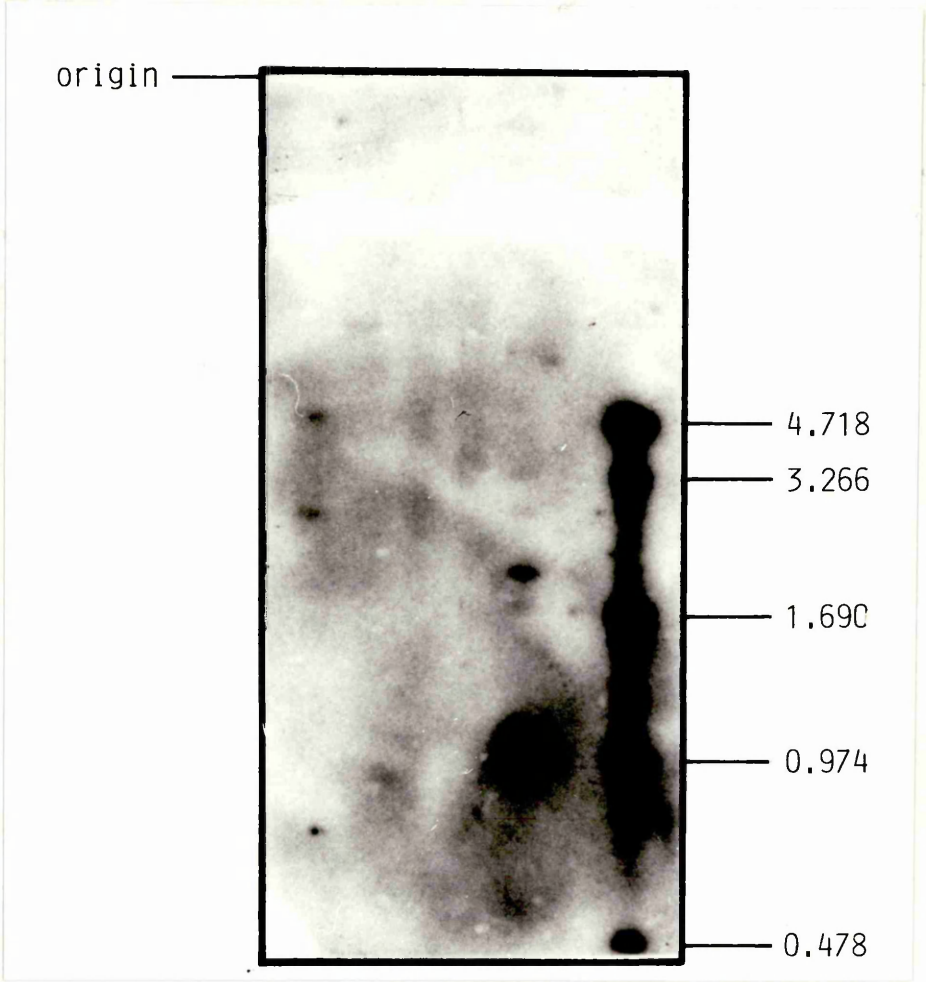


Figure 3.22 Southern analysis of P. sativum and S. cerevisiae DNA hybridised with the E. coli aroC probe.

lane: 1. EcoRV-digested S. cerevisiae DNA
2. EcoRI- " " "
3. EcoRV- " P. sativum "
4. EcoRI- " " "
5. EcoRV- " " "
6. Markers

The markers were generated by combining digests of pGM602/BamHI; pGM602/BamHI + EcoRV and pGM602/SalI + ClaI.

15 µg samples of digested P. sativum DNA, 3 µg samples of digested S. cerevisiae DNA and 0.3 µg of digested E. coli DNA were electrophoresed on a 0.8% agarose gel and the southern blot was hybridised with the E. coli aroC probe at 42°C in hybridisation solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

3.8 Conclusion

This chapter describes the attempts to cross-hybridise the P. sativum E2/E3 gene and its transcript with probes derived from corresponding microbial genes. Contamination of early P. sativum DNA preparations with pME173 influenced the strategy which was adopted and led to some false positive results being pursued. Despite the contamination problem some conclusions can be drawn from these experiments.

The results of the southern blots (containing P. sativum DNA contaminated with pME173) which had been probed with non-pME173 derived probes are certainly valid. The P. sativum rubisco SSU cDNA probe (section 3.3.1) did not hybridise to S. cerevisiae DNA which had been included on the same blots, therefore they were unable to hybridise to the pME173 DNA which contaminated the P. sativum genomic DNA. The rubisco SSU probe hybridised to several discrete bands in P. sativum genomic DNA, as would be expected, and so the southern hybridisation technique was satisfactory. The E. coli aroD and aroE probes did not hybridise to P. sativum genomic DNA therefore neither were of use in screening for the P. sativum E2/E3 gene.

The P. sativum RNA (section 3.4) was free of plasmid contamination because no hybridisation signal was obtained from northern blots probed with the S. cerevisiae E2/E3 probe. This lack of hybridisation was thought to be a consequence of the low abundance of P. sativum E2/E3 mRNA. Although this message would certainly be a low abundance message it now seems likely that the primary reason for no

hybridisation is a lack of homology between the P. sativum RNA and the S. cerevisiae E2/E3 probe.

No hybridisation was observed when the complete P. sativum genomic library (section 3.5) was screened with the S. cerevisiae E2/E3 probe. This again is almost certainly due to there being insufficient homology between the S. cerevisiae E2/E3 gene and the corresponding P. sativum gene. The genomic library apparently did not contain contaminating pME173 sequences as these would have been detected when the library was screened with pME173 derived probes. This genomic library could still be of use in the future if suitable probes become available.

When genomic blots containing uncontaminated P. sativum DNA were probed with the S. cerevisiae E2/E3 probe, no signals were obtained. This confirmed that the problem was one of lack of homology as opposed to under representation in the genomic library or low abundance of mRNA in northern blots and therefore strongly suggested that the S. cerevisiae E2/E3 gene could not be used to screen for the P. sativum E2/E3 gene.

On the other hand, the S. cerevisiae E1 and E5 probes appeared to hybridise to individual fragments of digested P. sativum DNA. Although it could not be ruled out that these hybridisation patterns were due to repetitive elements within P. sativum DNA it was possible that the probes were hybridising to the corresponding P. sativum genes.

The picture that emerged therefore was that the S. cerevisiae E1 and E5 genes were more homologous to the P. sativum DNA than the E2/E3 genes. This is perhaps not surprising because the homology between the S. cerevisiae multifunctional enzyme with the individual E. coli monofunctional enzymes is relatively high with respect to the E1 enzymes (36% homology) and E5 enzymes (38% homology) compared to the E2 enzymes (21% homology) and the E3 enzymes (25% homology). It appeared that using the S. cerevisiae E1 and E5 probes to isolate the corresponding plant genes possibly stood a better chance of success than using the S. cerevisiae E2/E3 gene to isolate the plant E2/E3 gene. Such attempts will be discussed in the next chapter.

The E. coli E6 (aroC) gene did not detect any signal from the P. sativum genomic DNA although it did detect discrete bands in southern blots of S. cerevisiae genomic DNA. Therefore, the E. coli aroC probe could be used to isolate the S. cerevisiae E6 gene for which there is as yet no reported clone. This was not part of the objectives of this project and so it was not pursued further.

CHAPTER 4

SCREENING AN *Arabidopsis thaliana* GENOMIC LIBRARY WITH HETEROLOGOUS PROBES

4.1 Introduction

The results described in chapter 3 did not conclusively show that the heterologous probes, derived from the microbial genes, would hybridise with individual fragments of digested P. sativum DNA. Two of the probes hybridised with multiple bands which were possibly repetitive elements within the P. sativum genome. By taking advantage of the low amount of repetative DNA within the genome of Arabidopsis thaliana the problem of hybridisation to repetitive sequences could be overcome to a large extent and signals which were obtained could be examined with some confidence.

The nuclear DNA content of higher plants ranges over three orders of magnitude, from 0.2 to over 200 picograms per cell, Table 4.1. Polyploidy can account for this in some cases, for example Fritillaria assyriaca has the largest documented nuclear DNA content however this plant is tetraploid. Another member of the same genus Fritillaria davisii has the largest genome size of a diploid plant (Bennet and Smith, 1976). Many plants have larger genomes than humans (see Table 4.1 for comparison) and since humans are generally regarded as being more complex than plants then it follows that not all of the plant genomic DNA codes for a product. In fact a large proportion of their DNA is composed of highly repetitive sequences which are unlikely to represent coding genes. The percentage of the nuclear genome present in repeated sequences varies from the small quantity found in A. thaliana (Leutwiler et al., 1984) to the 70-75% found in wheat, rye and P.sativum (Pruitt and Meyerowitz, 1986). Single copy sequences

Table 4.1 Nuclear DNA content of higher plants

The range of nuclear DNA content of a selection of higher plants compared to human, S. cerevisiae and E. coli

Organism	Nuclear DNA content in an unreplicated haploid genome		ploidy
	10^9 bp	picograms	
Allium cepa (onion)	16.2	16.2	2x
Arabidopsis thaliana (Arabidopsis)	0.07	0.07	2x
Avena sativa (oat)	13.2	13.7	6x
Fritillaria assyriaca (liliaceae)	122.0	127.0	4x
Fritillaria davisii (")	86.0	89.0	2x
Glycina max (soybean)	0.8	0.9	4x
Lycopersicon esculentum (tomato)	0.72	0.75	2x
Nicotiana tabacum (tobacco)	3.7	3.9	4x
Petunia hybrida (petunia)	1.5	1.6	2x
Phaseolus vulgaris (French bean)	1.7	1.8	2x
Pisum sativum (pea)	4.5	4.6	2x
Secale cereale (rye)	9.1	9.5	2x
Triticum aestivum (wheat)	16.7	17.3	6x
Triticum monococcum (")	6.0	6.2	2x
Homo sapiens (human)	3.0		2x
Escherichia coli	0.0039		
Saccharomyces cerevisiae (yeast)	0.02		

(Adapted from Bennet and Smith, 1976; Murray et al., 1978; Leutwiler et al., 1984)

are interrupted by repetitive sequences resulting in an interspersion pattern which can vary greatly even among closely related plants. This is demonstrated by the genome of P. sativum which contains no single copy sequences longer than 1000 bp (Murray et al., 1978) and the genome of Phaseolus mungo, a member of the same family, in which almost 50% of single copy sequences are present as blocks greater than 6,700 bp (Murray et al., 1979).

Genome size, repetitive sequence content and sequence interspersion are interrelated. Larger genomes tend to contain a greater percentage of repetitive sequences and have more of their single-copy sequences contained in regions of short period interspersion (Flavell, 1980).

The small mustard, A. thaliana, is a typical dicotyledonous plant in many respects but it has several properties which have made it a very useful tool for plant molecular biologists (Estelle & Somerville, 1986; Pang & Meyerowitz, 1987; Meyerowitz, 1987; 1989). The useful characteristic of A. thaliana relevant to this project is its relatively simple genome organisation. Its genome size of 7×10^7 kbp, is the smallest genome known in the plant kingdom (Table 4.1). Only 10-14% of its genome consists of repetitive sequences (Leutwiler et al., 1984) and in contrast to other plants the majority of the nuclear DNA of A. thaliana appears to be organised as extremely long blocks of unique sequences which are only interupted on average every 120 kbp by repetitive DNA.

As a consequence of this small genome, just over 20,000 λ-clones with

average inserts of 15 kbp need to be screened in order to achieve a 99% probability of isolating a specific unique sequence from an A. thaliana genomic library. This is 50 to 100 times less than the number of clones needed to be screened from a P. sativum genomic library and therefore makes screening an easier procedure. The low proportion of repetitive DNA present in the A. thaliana genome minimises the chances of obtaining hybridisation of the probe with repetitive DNA as was discussed in section 3.7.2. Therefore several problems involved in screening genomic libraries with heterologous probes could be eliminated by making use of an A. thaliana genomic library.

There was however, no greater likelihood that the shikimate pathway genes of A. thaliana were any more homologous to the microbial genes, than the genes of P. sativum. Therefore in attempting to clone the genes from A. thaliana using probes derived from the microbial genes, sufficient homology still had to exist between the genes of the micro-organisms and those of the higher plant.

Obtaining clones for the shikimate pathway genes from A. thaliana was not part of the original objective of obtaining the genes from P. sativum. However it seemed likely that the former could easily be used to cross-hybridise with the corresponding genes from other plants, including P. sativum since it was likely that the plant genes would exhibit considerable homology (Chang & Meyerowitz, 1986).

4.2 Screening an A. thaliana genomic library with several microbial probes.

4.2.1 The A. thaliana genomic library

A library of A. thaliana genomic DNA (Krebbers et al., 1988a), contained in the phage cloning vector λ Charon 35 (Leonan and Blattner, 1983) was available for screening. This library had been prepared by partially digesting A. thaliana nuclear DNA with Sau3A and isolating DNA fragments of approximately 14 kbp. The resultant DNA fragments were treated with calf intestinal phosphatase and ligated to sucrose gradient purified λ Charon 35 BamHI arms (Fig 4.1). Treatment of the insert DNA with calf intestinal phosphatase prevented multiple insertion of insert fragments. When the recombinant DNA was in vitro packaged, 7.5×10^5 independently arising pfu were obtained and these recombinant phage were amplified.

From this library clones of several A. thaliana genes have been obtained using other plant probes (Krebbers et al., 1988a; 1988b), therefore it was expected that the genes for the shikimate pathway enzymes would be represented within the library.

4.2.2. Screening the A. thaliana genomic library

Two samples of E. coli K803 were infected with 10^5 pfu (four genome equivalent) from the amplified A. thaliana genomic library. Duplicate impressions of the 10^5 plaques on each of the plates were made and

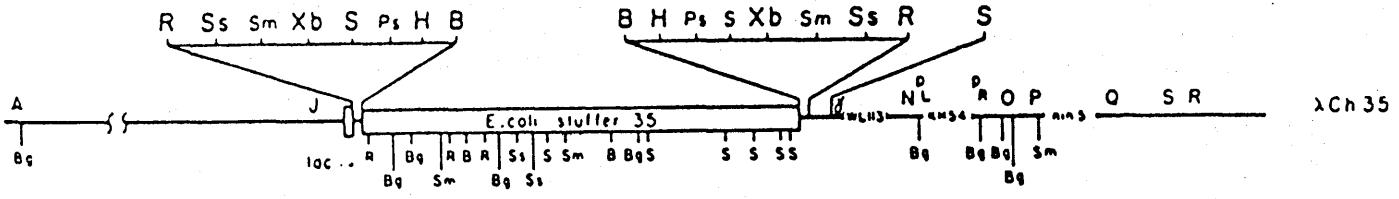


Figure 4.1 Restriction map of λCharon 35

The stuffer fragment of this substitution vector may be removed by digestion with any of the enzymes within the polylinker (Leonan and Blattner, 1983) except SmaI.

Key to restriction sites: B=BamHI, Bg=BglII, R=EcoRI, H=HindIII, S=SalI, Sm=SmaI, Ss=SstI, Ps=PstI, Xb=XbaI

each set of duplicate filters was screened with one of the following microbial probes: S. cerevisiae E1, E2/E3 and E5 probes and E. coli aroC probe. Hybridisation of the filters was at 30°C in solution containing 50% formamide and washing was at 50°C, 0.1% SDS.

After autoradiography for one week, no signals were obtained from the A. thaliana library screened with either the S. cerevisiae E2/E3 or E5 probes. A strip of nitrocellulose containing dotted samples of pME173 DNA had been included in each hybridisation and both probes produced a signal after autoradiography of the control strip. This verified that the probes could hybridise to homologous DNA and suggested that the lack of hybridisation during screening was not because of a problem with the probes.

Using the E. coli aroC probe, a very high background signal was obtained after autoradiography for three hours and the autoradiographs were uniformly black after overnight exposure. The signal did not appear to be due to the phage DNA present on the filter. Instead the lawn of E. coli host cells, or their genomic DNA, appeared to be hybridising to the E. coli aroC probe. This exhibited a problem which should be considered when screening libraries that are housed in E. coli, with a probe derived from an E. coli gene. The probe may, as it did in this case, hybridise with the E. coli genomic DNA rather than the plant DNA, if the homology to the latter is only slight and the stringency of screening is low. Use of other probes, derived from E. coli genes, to screen the A. thaliana library for shikimate pathway genes, was likely to give similar results and was therefore not

attempted.

Using the S. cerevisiae E1 probe, a speckled pattern was observed (in duplicate) on one week exposure autoradiographs (Fig 4.2). This result was obviously very different from the result obtained with either the S. cerevisiae E2/E3 or E5 probes. At the low stringency used, it was likely that any sequence remotely homologous to the S. cerevisiae E1 probe would hybridise to it. It was estimated that the S. cerevisiae E1 probe had hybridised with approximately 2,400 plaques out of a possible 100,000. This suggested that the majority of plaques were not hybridising to the probe and only 2.4% of four genomes of A. thaliana DNA, which is equivalent to 0.6% of one A. thaliana genome was hybridising with the S. cerevisiae E1 probe. It was conceivable that this 0.6% of the plaques, could contain a particular repetitive sequence, present within the A. thaliana genome, which was sufficiently homologous to the probe that it hybridised under the low stringency used.

If the A. thaliana E1 gene could also hybridise with the S. cerevisiae E1 probe then perhaps some of the spots on the autoradiograph may have arisen from the A. thaliana E1 gene. All the spots on the autoradiograph did not have the same intensity. There were many small faint spots and many intermediate dark spots but there were also approximately 20 large dark spots. These large dark spots occurred with a frequency of 5 per genome. Plaques occuring with this frequency potentially contained a sequence which was present in very few copies. The level of intensity of signal also suggested that these plaques were more homologous to the S. cerevisae E1 probe than

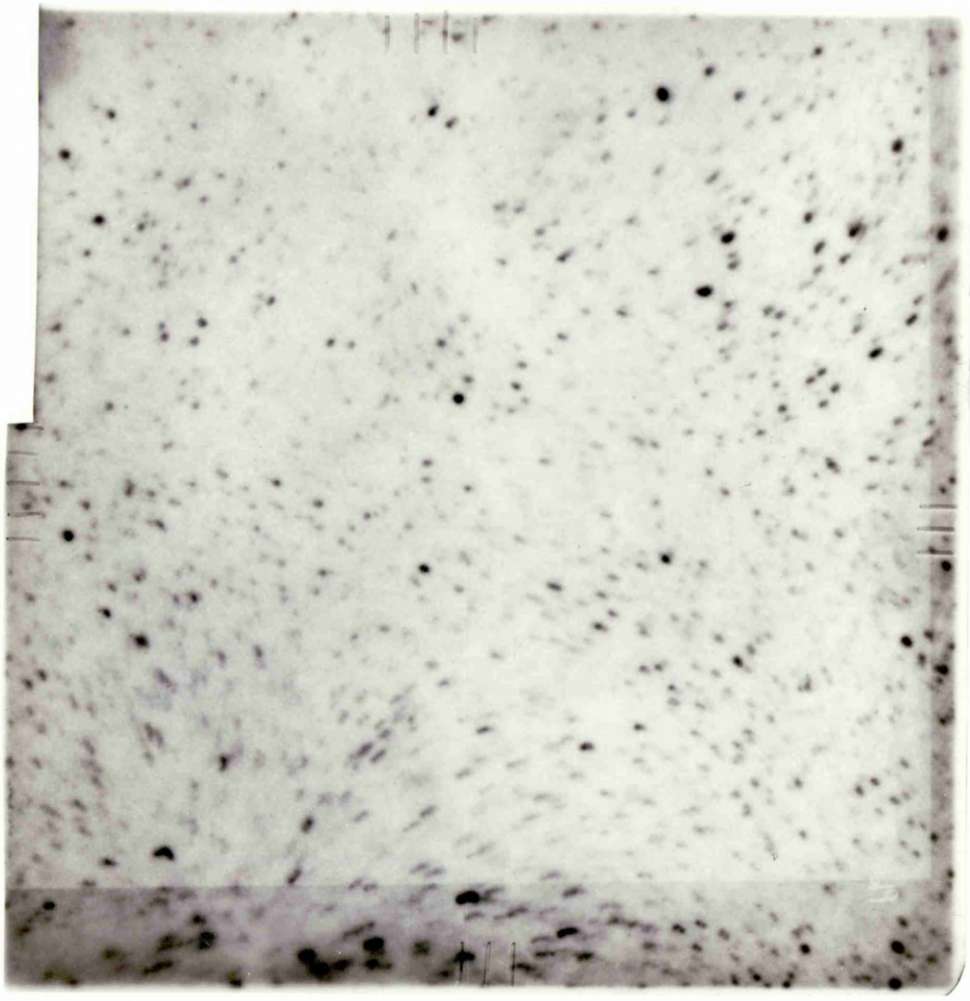


Figure 4.2 Autoradiograph of an E1 screened filter

Autoradiograph of an impression from a plate containing 10^5 plaques from the A. thaliana genomic library hybridised with the S. cerevisiae E1 probe. The filter was hybridised at 30°C in solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

the other plaques.

The difference in the character of the spots on the autoradiograph could have been due to differential viability of each recombinant phage. In other words, some recombinant phage were possibly more easily replicated and therefore able to produce larger plaques and bigger signals than other recombinant phage.

However, in the hope that the large dark spots, which occurred at the frequency of 5 per genome, corresponded to clones of the A. thaliana E1 gene, 18 were purified to homogeneity over several rounds of screening and further characterised.

4.3 Characterisation of the clones from the A. thaliana genomic library.

4.3.1 The phage DNA homologous to the S. cerevisiae E1 probe

The A. thaliana library (Krebbers et al., 1988a) had been constructed in a vector (λ charon 35) which allowed the convenient removal of insert DNA using one of several restriction enzymes (Fig 4.1). including EcoRI and HindIII.

DNA was prepared from each of the 18 purified phage that were most homologous to the S. cerevisiae E1 probe. The DNA from each phage was digested singly with EcoRI and HindIII and the digestion products were electrophoresed on agarose gels (Figs 4.3a(i) and 4.3b(i)). The

Figure 4.3 Analysis of the recombinant phage DNA which hybridised with the E1 probe

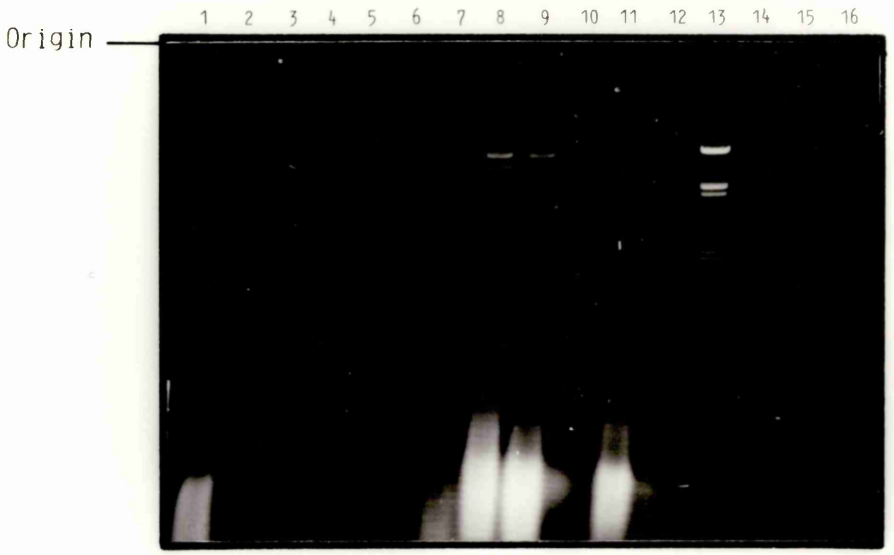
a (i) lane:

1. EcoRI-digested λE1-02
2. " " λE1-03
3. " " λE1-05
4. " " λE1-06
5. " " λE1-07
6. " " λE1-09
7. " " λE1-10
8. " " λE1-11
9. " " λE1-12
10. " " λE1-13
11. " " λE1-14
12. " " λE1-15
13. EcoRI + HindIII-digested λ-DNA
14. EcoRI-digested pME173
15. ———
16. DraI-digested pME173

The size marker bands (in kbp) in lane 13 are 21.7, 5.24 + 5.05 (doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58.

(ii) The southern blot of gel (i) was hybridised with the S. cerevisiae E1 probe at 30°C in 50% formamide and washed at 50°C in 2 x SSC. Lanes containing pME173 DNA were hybridised separately. The sizes of the bands which lit up are indicated.

(i)


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Origin

(ii)

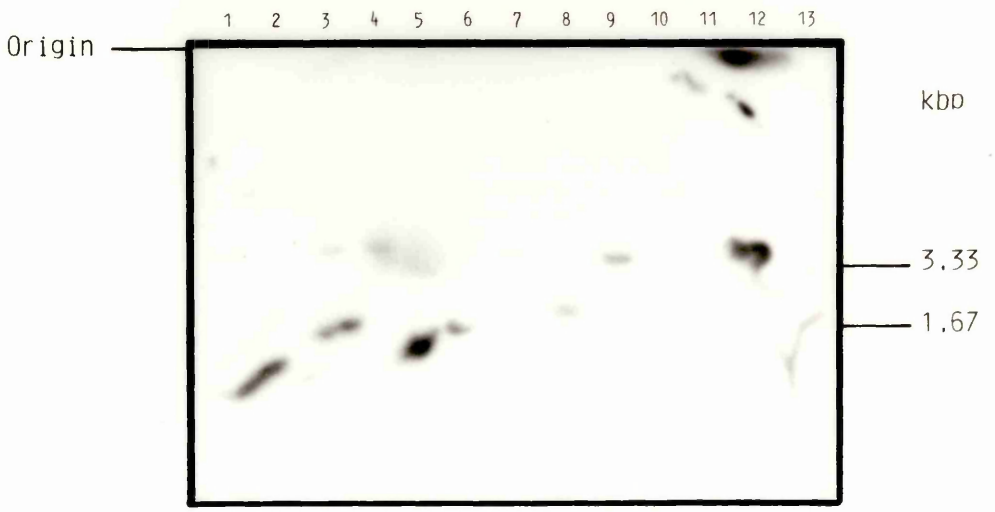

1 2 3 4 5 6 7 8 9 10 11 12 13
Origin
kbp
3.33
1.67

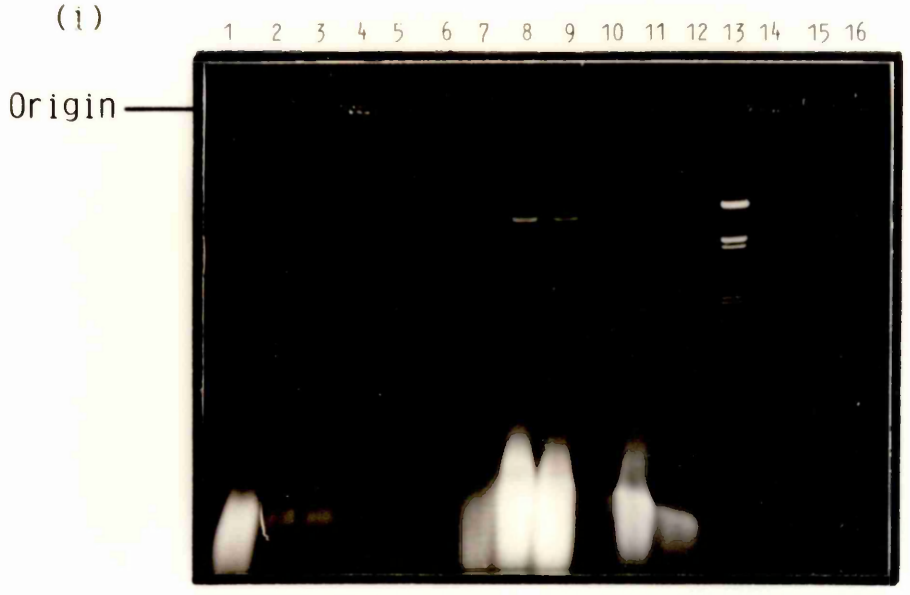
Figure 4.3b

(i) lane:
1. HindIII-digested λE1-02
2. " " λE1-03
3. " " λE1-05
4. " " λE1-06
5. " " λE1-07
6. " " λE1-09
7. " " λE1-10
8. " " λE1-11
9. " " λE1-12
10. " " λE1-13
11. " " λE1-14
12. " " λE1-15
13. EcoRI + HindIII-digested λ-DNA
14. DraI-digested pME173
15. ——
16. EcoRI-digested pME173

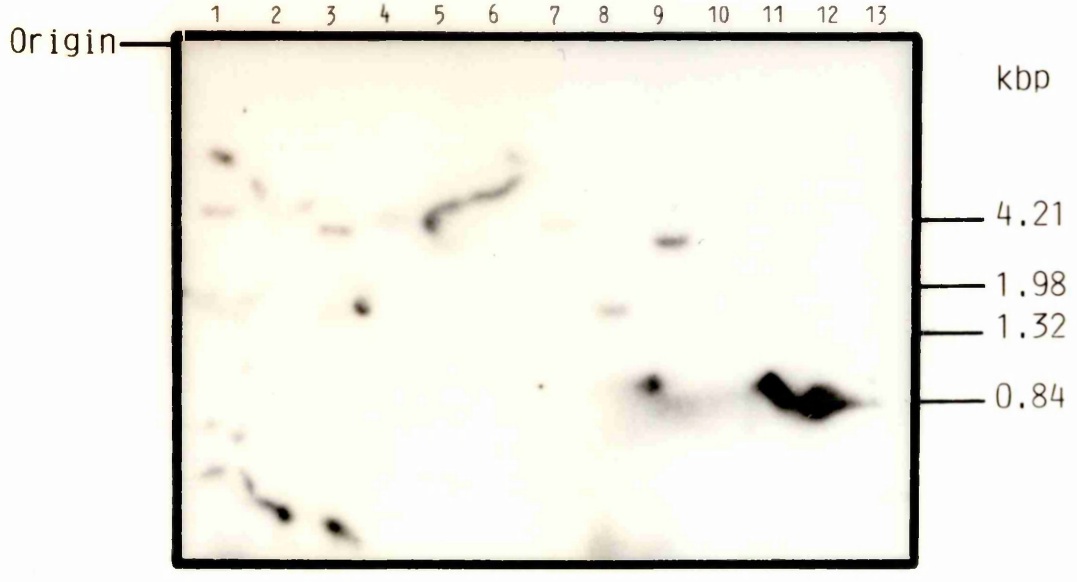
The size marker bands (in kbp) in lane 13 are 21.7, 5.24 + 5.05 (doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58.

(ii) The southern bot of gel (i) was treated as described in Figure 4.3a(ii) legend.

(i)


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Origin

(ii)


1 2 3 4 5 6 7 8 9 10 11 12 13
Origin
kbp
4.21
1.98
1.32
0.84

restriction patterns produced from each EcoRI digestion were not similar, and neither were the restriction patterns produced from each HindIII digestion. This suggested that the isolated phage contained different insert fragments of DNA which was worrying because it was hoped that each of these clones contained the A. thaliana E1 gene and therefore should have a similar restriction pattern. However it was assumed that the pattern obtained was a consequence of the partial digestion of Sau3A during the library construction. The enzyme would not have cut at the same site in each case and therefore each clone could contain a different region of the A. thaliana E1 gene. The southern blots of these gels, (Figs 4.3a(ii) and 4.3b(ii)) showed hybridising bands of similar, but not identical, sizes in both the EcoRI digestions and the HindIII digestions. This may indeed reflect a variety of Sau3A digestion products from one, or a few, DNA regions. The relationship of the clones to one another would only be resolved when an E1 homologous DNA fragment from one of the clones was used to probe a similar southern blot, to identify fragments containing similar sequences.

4.3.2 Generation of pSG101

λE1-12 was chosen to be completely characterised because this phage appeared to contain a large piece of A. thaliana insert DNA and the signal produced in the southern blot (Fig 4.3). from λE1-12 was the strongest.

λE1-12 DNA was digested with EcoRI and the total digestion products

were ligated with EcoRI-digested, calf intestinal phosphatase-treated pUC18. Small scale plasmid DNA preparations were carried out on the resultant white transformants and samples of the plasmids were digested with EcoRI. Agarose gel analysis of the digestions (Fig 4.4a), indicated that at least 3 fragments of λE1-12 insert DNA had been sub-cloned.

The southern blot of this gel was hybridised with the S. cerevisiae E1 probe. The autoradiograph (Fig 4.4b) showed that plasmids from recombinant 1 contained inserts with homology to the S. cerevisiae E1 probe. This plasmid was named pSG101.

The recombinant plasmids 2, 3 and 4 (named pSG102) and recombinant plasmid 6 (named pSG103) did not hybridise with the S. cerevisiae E1 probe. Further analysis of these plasmids would be carried out if pSG101 did not contain all of the desired information.

4.3.3 The restriction map of pSG101

A restriction map of pSG101 was required to allow a sub-cloning strategy to be worked out and also to allow further localisation of the homology to the S. cerevisiae E1 probe.

pSG101 was digested with a variety of restriction enzymes and the digestions were electrophoresed on agarose gels (Fig 4.5). This indicated that there was one BamHI site, one XhoI site, one AccI site, one SalI site, two BglII sites and three ClaI sites within the pSG101

Figure 4.4 Analysis of the subclones from the recombinant phage

a) Agarose gel analysis

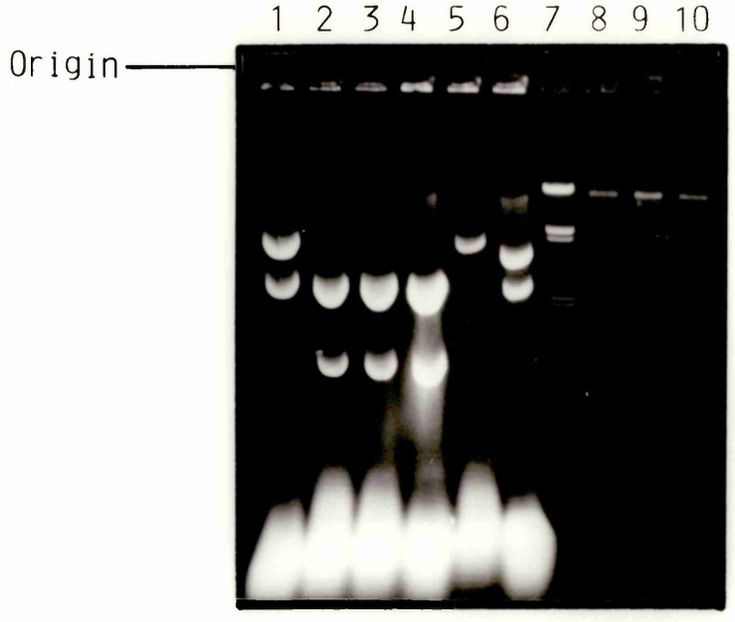
lane: 1. EcoRI-digested recombinant-1
2. " " " -2
3. " " " -3
4. " " " -4
5. " " " -5
6. " " " -6
7. EcoRI + HindIII-digested λ-DNA
8. EcoRI-digested λE1-12 DNA
9. HindIII-digested λE1-12 DNA
10. EcoRI + HindIII-digested λE1-12 DNA

The DNA samples were electorphoresed on a 1% agarose gel. The size marker bands (in kbp) in lane 7 are 21.7, 5.24 + 5.05 (doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58.

b) Southern blot analysis of gel (a)

The southern blot of gel (a) was hybridised with the S. cerevisiae E1 probe at 30°C in solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS. Notice that the hybridising band in lane 1 is of similar size to the hybridising band in lane 8.

(a)



(b)

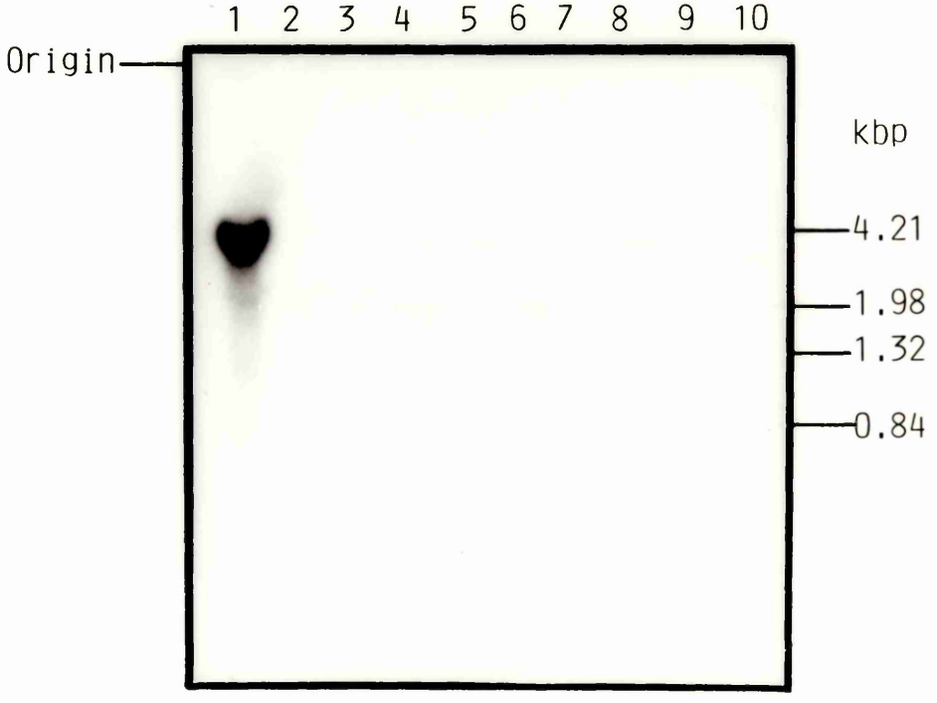


Figure 4.5 Restriction mapping of pSG101

gel (i)

lane 14 contains EcoRI + HindIII-digested λ-DNA with band sizes (in kbp) of 21.7, 5.24 + 5.05 (doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58. The other lanes contain pSG101 digested with the following restriction enzymes.

lane:				
	1.	AccI	8.	SalI
	2.	BamHI	9.	SmaI
	3.	HaeIII	10.	SphI
	4.	HindIII	11.	SstI
	5.	HincII	12.	XbaI
	6.	KpnI	13.	EcoRI
	7.	PstI		

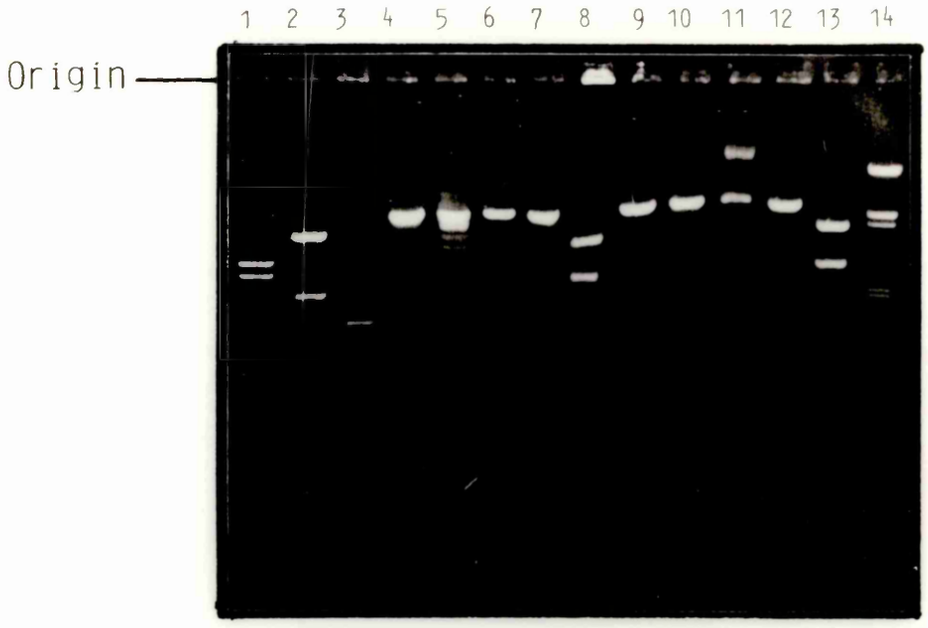
gel (ii)

lane 4 contains the same size marker DNA as described above. The other lanes contain pSG101 digested with the following restriction enzymes.

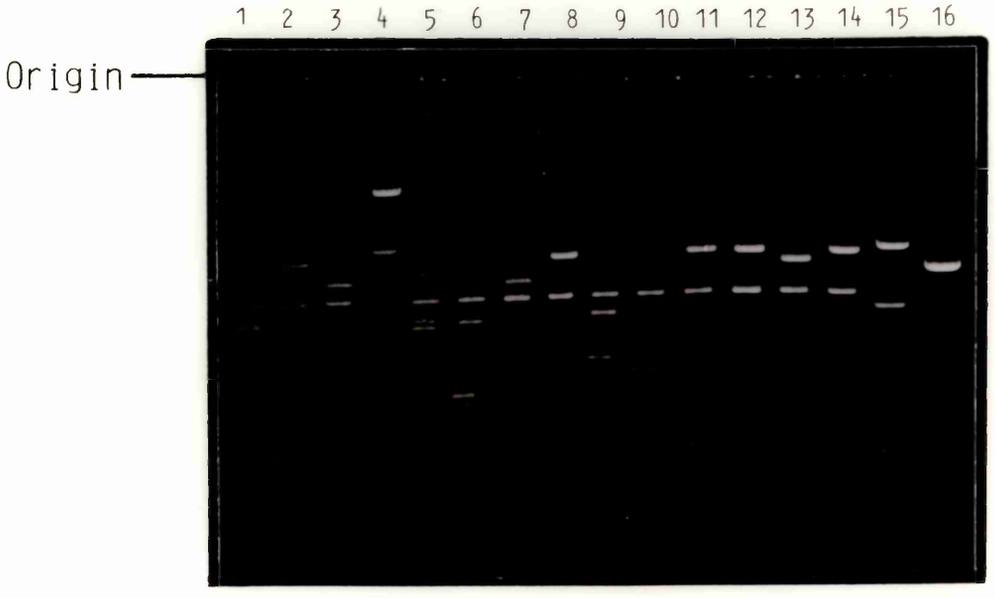
lane:			lane:		
	1.	EcoRI + BamHI		10.	EcoRI + ClaI
	2.	" + SalI		11.	" + EcoRV
	3.	" + AccI		12.	" + NruI
	5.	" + BamHI + SalI		13.	" + XhoI
	6.	" + BamHI + AccI		14.	EcoRI
	7.	" + AccI + SalI		15.	PvuII
	8.	" + BclI		16.	NdeI
	9.	" + BglII			

The DNA samples were electrophoresed on a 1% agarose gel. Enzymes used in gel (i) and lanes 1-6 of gel (ii) digest pUC18 within the polylinker. BclI, BglII, ClaI, EcoRV, NruI and XhoI do not cut pUC18. PvuII and NdeI digest pUC18 outwith the polylinker twice and once respectively.

(i)


1 2 3 4 5 6 7 8 9 10 11 12 13 14
Origin

(ii)


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Origin

insert.

Other appropriate double and triple digestions were carried out and these digestions were electrophoresed on an agarose gel (Fig 4.6a). The sizes of the various fragments in Figs 4.5 and 4.6a were estimated (Table 4.2) and a restriction map of the A. thaliana genomic DNA insert present in pSG101 was derived (Fig 4.7).

The SalI site which was observed in Fig 4.5 must have been very close to one end of the pSG101 insert because there was no detectable difference in pattern between EcoRI digested pSG101 and EcoRI+SalI -digested pSG101. The SalI site was therefore not useful for subcloning.

4.3.4 Further localisation of the homology to the S. cerevisiae E1 probe.

Gel(ii) in Fig 4.6a contained many fragments of the pSG101 insert and so to identify the fragments homologous to the S. cerevisiae E1 probe, the southern blot of this gel was hybridised with the E1 probe (Fig 4.6b). The fragments of pSG101 which hybridised to the probe are indicated underneath the restriction map in (Fig 4.7). It was difficult to discriminate between the EcoRI-ClaI fragment and the ClaI fragment (from the EcoRI+ClaI+AccI and EcoRI+ClaI+XhoI digests) as to which contained the region showing homology with the S. cerevisiae E1 probe, because both fragments were similarly sized.

Figure 4.6 Further restriction mapping of pSG101

a) Agarose gel analysis

lane:
1. EcoRI-digested λE1-12 DNA
2. EcoRI + HindIII-digested λ-DNA
3. " + XhoI + AccI-digested pSG101
4. " + BglII + AccI- " "
5. " + BglII + BamHI- " "
6. " + ClaI + AccI- " "
7. " + ClaI + XhoI- " "

The DNA samples were electrophoresed in a 1% agarose gel. The size marker bands (in kbp) in lane 2 are 21.7, 5.24 + 5.05 (doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58.

b) Southern blot analysis

The southern blot of gel (a) was hybridised with the S. cerevisiae E1 probe at 30°C in solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

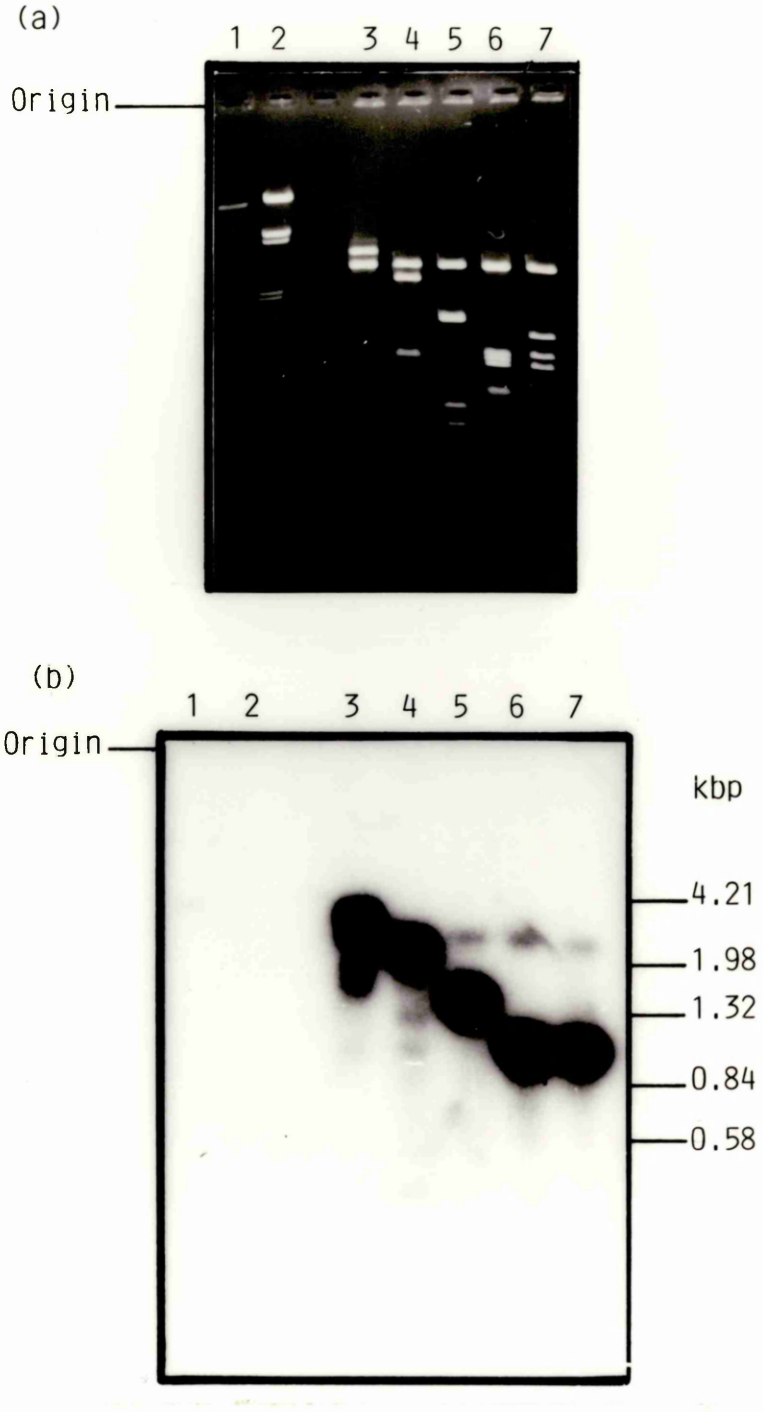

(a)
1 2 3 4 5 6 7
Origin
(b)
1 2 3 4 5 6 7
Origin
kbp
4.21
1.98
1.32
0.84
0.58

Table 4.2 Restriction mapping of pSG101

enzymes used to digest pSG101	sizes of DNA fragments produced (kbp)
EcoRI	vector (2.7), 4.2
" + BamHI	" , 2.1, 2.05
" + SalI	" , 4.2
" + AccI	" , 3.1, 1.1
" + BamHI + SalI	" , 2.1, 2.0
" + BamHI + AccI	" , 2.1, 1.0 (doublet)
" + AccI + SalI	" , 3.1, 1.1
" + BglII	" , 2.2, 1.5, 0.6
" + ClaI	" , 1.4, 1.1, 1.0, 0.8
" + XhoI	" , 3.7, 0.5
" + AccI + XhoI	" , 3.0, 0.6, 0.5
" + BglII + AccI	" , 2.2, 1.1, 0.6, 0.4
" + BglII + BamHI	" , 1.5 (doublet), 0.6, 0.7
" + ClaI + AccI	" , 1.1, 1.15, 1.0, 0.8
" + ClaI + XhoI	" , 1.3, 1.1, 1.0, 0.5

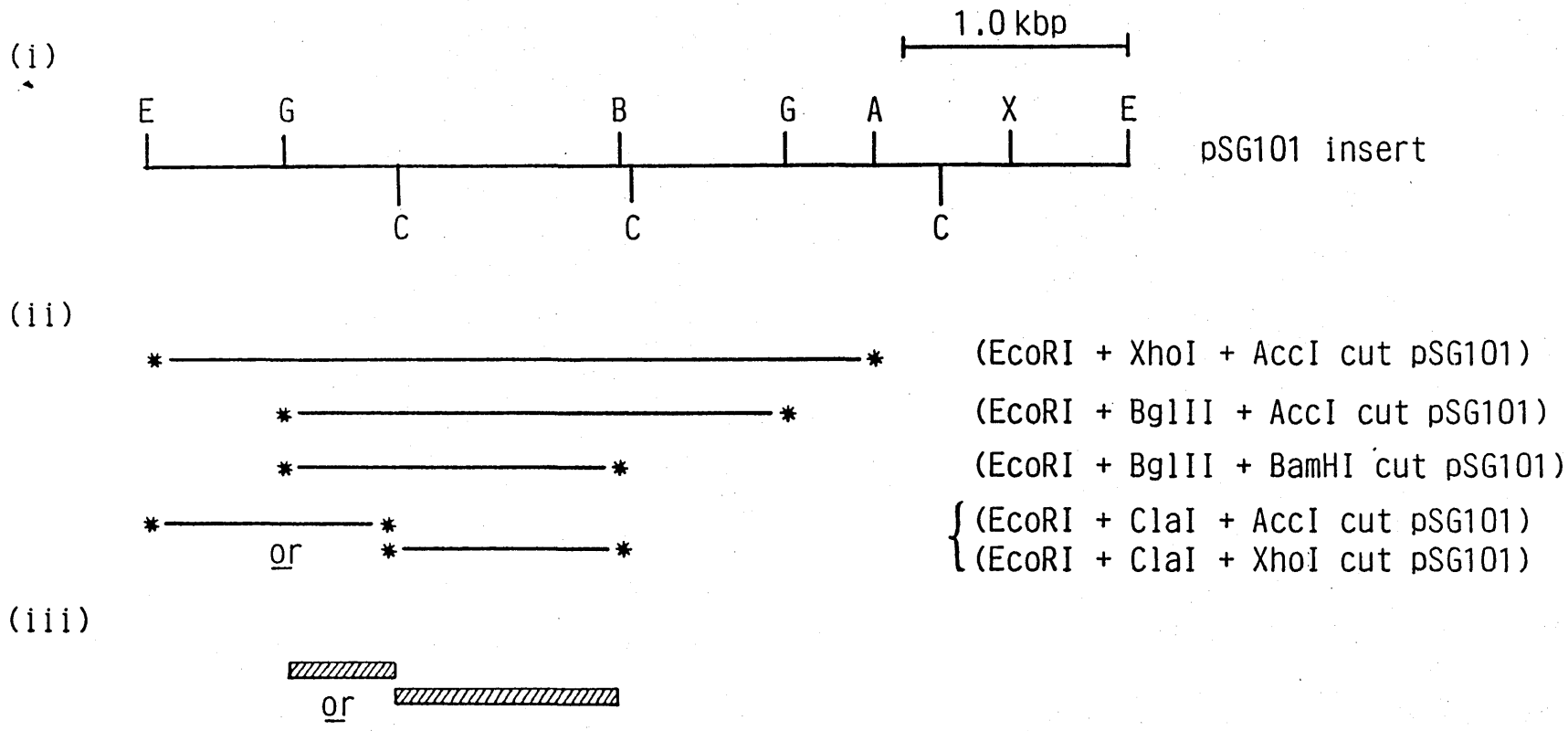


Figure 4.7 Restriction map of pSG101 and regions homologous to the S. cerevisiae E1 probe

Key to restriction sites: A=AccI, B=BamHI, C=ClaI, E=EcoRI, G=BglII, X=XhoI

(i) Restriction map of pSG101 and underneath (ii) the fragments which hybridised with the S. cerevisiae E1 probe in Fig 4.6b.(iii) The regions containing the homology to the S. cerevisiae E1 probe

As a result the DNA homologous to the S. cerevisiae E1 probe had been narrowed down to either the BglII-ClaI region or the ClaI-BamHI region of pSG101, (Fig 4.7). An EcoRI+ClaI+BglII-digest would have distinguished between these two possibilities. Similarly, both regions could be subcloned out of pSG101 and probed with the S. cerevisiae E1 probe to determine which contained the region showing homology. This latter strategy was adopted and will be described in the next section.

The southern blot in Fig 4.6b was rewashed at high stringency, ie 0.1 x SSC at 65°C. The signal from the various fragments of pSG101 was completely removed after the same length of exposure to autoradiography. This was encouraging because it confirmed that the hybridisation was occuring between heterologous sequences and could be prevented under high stringency conditions.

4.3.5 The generation of subclones from pSG101

Several regions of pSG101 were subcloned into pUC18 in order to help identify where the homology to the S. cerevisiae E1 probe occurred.

pSG101 was digested with BamHI+BglII+XhoI and the resultant 1.5 kbp BglII-BamHI fragment and 0.7 kbp BamHI-BglII fragment were isolated from an l.m.p. agarose gel. A sample of the 1.5 kbp BglII-BamHI fragment was digested with ClaI, resulting in a 0.5 kbp BglII-ClaI fragment and a 1.0 kbp ClaI-BamHI fragment. The latter two fragments were cloned into BamHI+AccI-digested pUC18, generating pSG105 and

pSG106 respectively, (Fig 4.8). The 1.5 kbp BamHI-BglII fragment and the 0.7 kbp BamHI-BglII fragments were cloned into BamHI-digested, calf intestinal phosphatase-treated pUC18 generating pSG104 and pSG107 respectively (Fig 4.8).

pSG104, pSG105 and pSG106 were digested with EcoRI+HindIII and the digestions were electrophoresed on an agarose gel (Fig 4.9a). The sourthern blot of this gel was probed with the S. cerevisiae E1 probe. The autoradiograph (Fig 4.9b) indicated that pSG105 contained the region of A. thaliana genomic DNA which was homologous to the probe.

It appeared therefore that the 0.5 kbp insert present within pSG105 was solely responsible for the hybridisation of λE1-12 with the S. cerevisiae E1 probe. The 0.5 kbp fragment of DNA was unlikely to encode the complete A. thaliana E1 gene, but it was possible that it contained the region of the gene which was most homologous to the S. cerevisiae E1 probe. To investigate this possiblity, the nucleotide sequence of the insert from pSG105 was determined.

4.4 Determination of the DNA sequence of the pSG105 insert.

4.4.1 Aspects of the sequencing strategy

The 0.5 kbp insert of pSG105 was sequenced by cleaving it, with restriction enzymes, into several smaller, more manageable sections which were individually sequenced using the M13/dideoxy method. The complete sequence was built up by overlapping the sequence from the

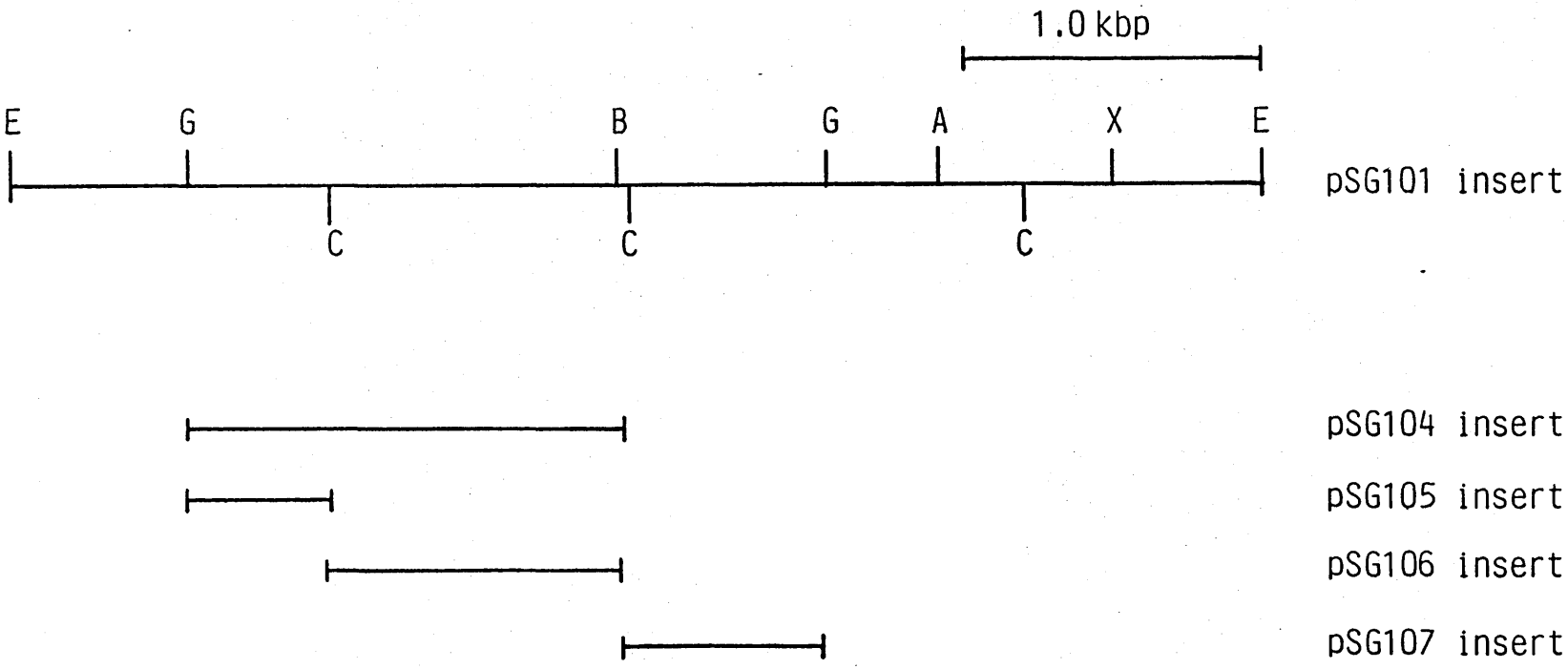


Figure 4.8 Subclones of pSG101

The regions of the pSG101 insert which were subcloned into pUC18 to produce pSG104, pSG105, pSG106 and pSG107.

Key to restriction sites: A=AccI, B=BamHI, C=ClaI, E=EcoRI, G=BglII, X=XhoI

Figure 4.9 Southern blot analysis of pSG105 and pSG106

a) Agarose gel analysis

lane: 1, 2, 3, 4, 6. EcoRI + HindIII-digested pSG106
5. EcoRI + HindIII + DraI-digested pSG106
7. EcoRI + BamHI + BglII-digested pSG101
8. EcoRI + HindIII-digested λ-DNA
9, 10, 11, 12. EcoRI + HindIII-digested pSG105
13. EcoRI + HindIII + DraI-digested pSG105
14. EcoRI + HindIII-digested pSG105
15, 16 EcoRI + HindIII-digested pSG104

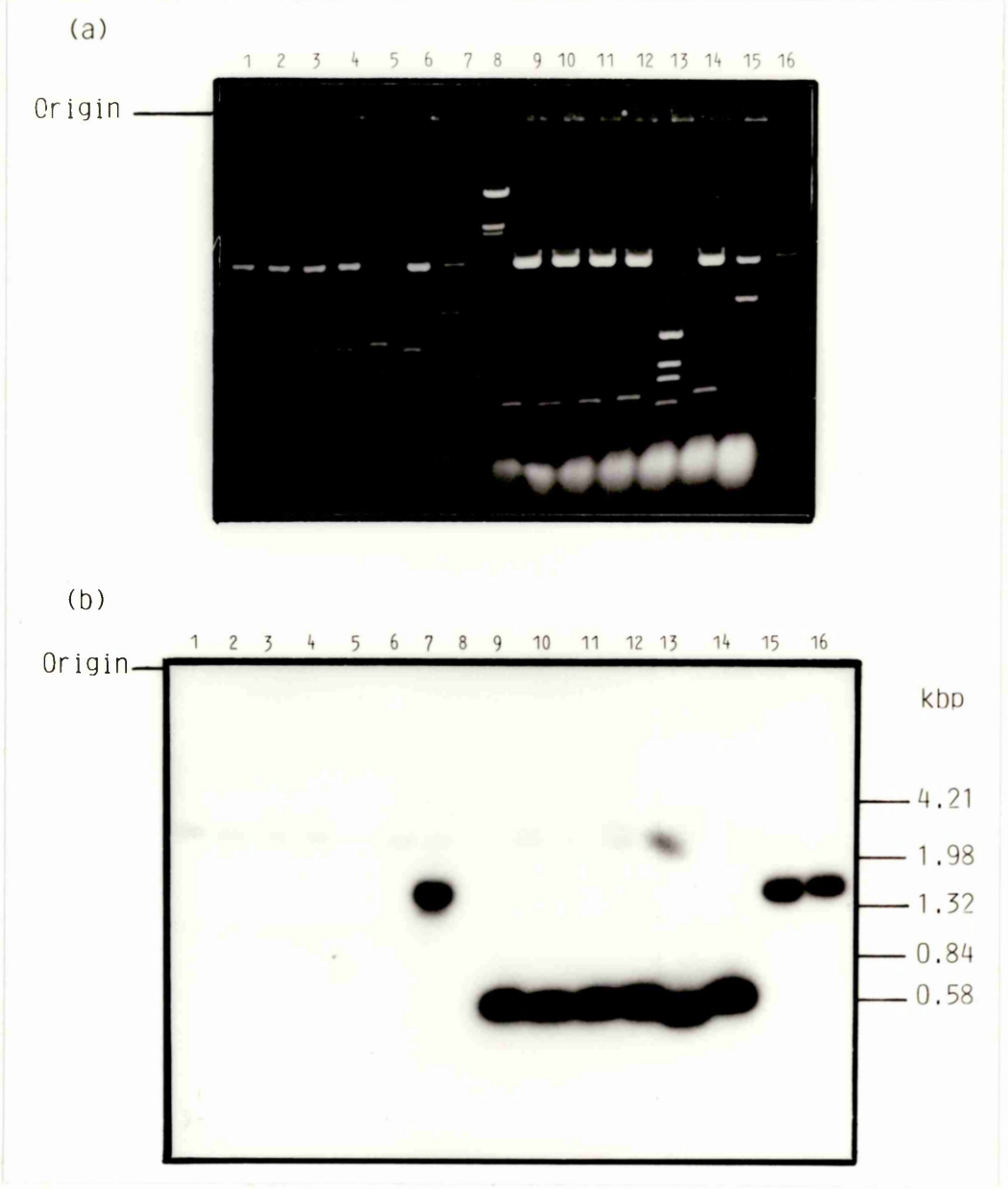
The DNA samples were electrophoresed on a 1% agarose gel. The size marker bands (in kbp) in lane 8 are 21.7, 5.24 + 5.05(doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58.

b) Southern blot analysis of gel (a)

The southern blot of gel (a) was hybridised with the S. cerevisiae E1 probe at 30°C in solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.

Note (i) The 1.5 kbp BglII-BamHI fragment of pSG101 and pSG104 hybridised with the E1 probe as did the 0.5 kbp Bgl-ClaI fragment of pSG105. None of the fragments derived from pSG106 showed homology to the E1 probe.

(ii) Gel (a) indicates that there is a DraI site within the pSG105 insert (lane 13).


(a)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Origin
(b)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Origin
kbp
4.21
1.98
1.32
0.84
0.58

smaller DNA fragments.

A prerequisite of any DNA sequencing project is that the DNA must be fully sequenced on both strands and all restriction sites used in the cloning must be overlapped to confirm the continuity of the final sequence.

4.4.2 Identification of convenient restriction sites

The pSG105 insert could not be removed from the plasmid by digestion with BamHI or ClaI because these sites had not been regenerated during the cloning of the insert into the BglII and AccI sites of pUC18, respectively (section 4.3.4). However, the whole insert and also the polylinker of pUC18 was removed from pSG105 by a HindIII+EcoRI digestion and the 0.5 kbp insert band was isolated from an l.m.p. agarose gel. This fragment was then analysed for convenient restriction sites that could readily be used in sub-cloning.

The 0.5 kbp insert from pSG105 was separately digested with Sau3A and HaeIII and these digests were electrophoresed on an agarose gel (Fig 4.10). This indicated that there was a Sau3A and HaeIII site within the pSG105 insert. Double digestion of the insert with these enzymes yielded three similarly sized fragments which were suitable for sequencing.

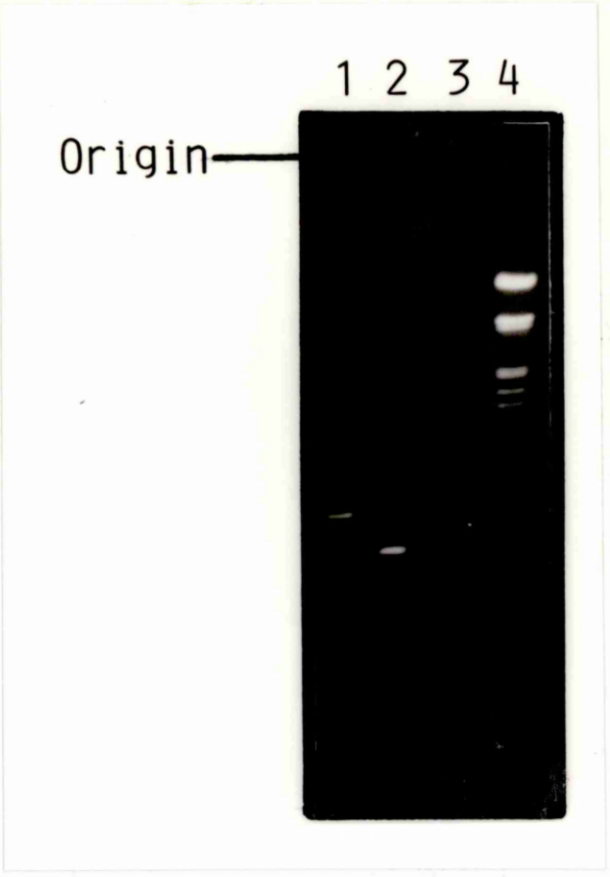


Figure 4.10 Restriction mapping of pSG105

lane: 1. Sau3A-digested pSG105 insert
2. HaeIII-digested pSG105 insert
3. Sau3A + HaeIII-digested pSG105 insert
4. EcoRI + HindIII-digested λ-DNA

Purified pSG105 insert was digested with the above enzymes. Notice that Sau3A did not digest the DNA to completion. In lane 1, a band corresponding to uncut pSG105 insert was present and in lane 3 a band of similar size to the HaeIII fragment of lane 2 was present.

The DNA samples were electrophoresed on a 1% agarose gel.

4.4.3 Sub-cloning and sequencing of the pSG105 insert

The purified 0.5 kbp HindIII-EcoRI fragment of pSG105, was cloned into HindIII+EcoRI-digested Mp18 and Mp19. The recombinant M13 provided the sequence of the fragment from both ends and this indicated the correct location of the HaeIII and Sau3A sites relative to either end of the fragment (Fig 4.11). The following subclones were then made. The Sau3A fragment of the pSG105 insert was cloned into BamHI-digested, calf intestinal phosphatase-treated Mp18. The Sau3A-HaeIII fragment was cloned into BamHI+SmaI-digested Mp18 and Mp19. The HaeIII-HindIII fragment was cloned into SmaI+HindIII-digested Mp18 and Mp19. The various recombinant M13 clones allowed the sequence of the pSG105 insert to be built up (Fig 4.11). The complete sequence is given in Fig 4.12 and this indicated that the pSG105 insert was 0.54 kbp in size.

The 0.54 kbp pSG105 insert was sequenced in both orientations, except for the region between the two Sau3A sites. The Sau3A fragment should have been cloned into Mp18 in either orientation however only clones containing the fragment in one orientation were ever obtained. Attempts firstly to reverse the orientation of the Sau3A fragment by excising it out of Mp18 using HindIII+EcoRI and recloning into Mp19 or secondly to clone the HaeIII-EcoRI fragment into Mp19, were not successful.

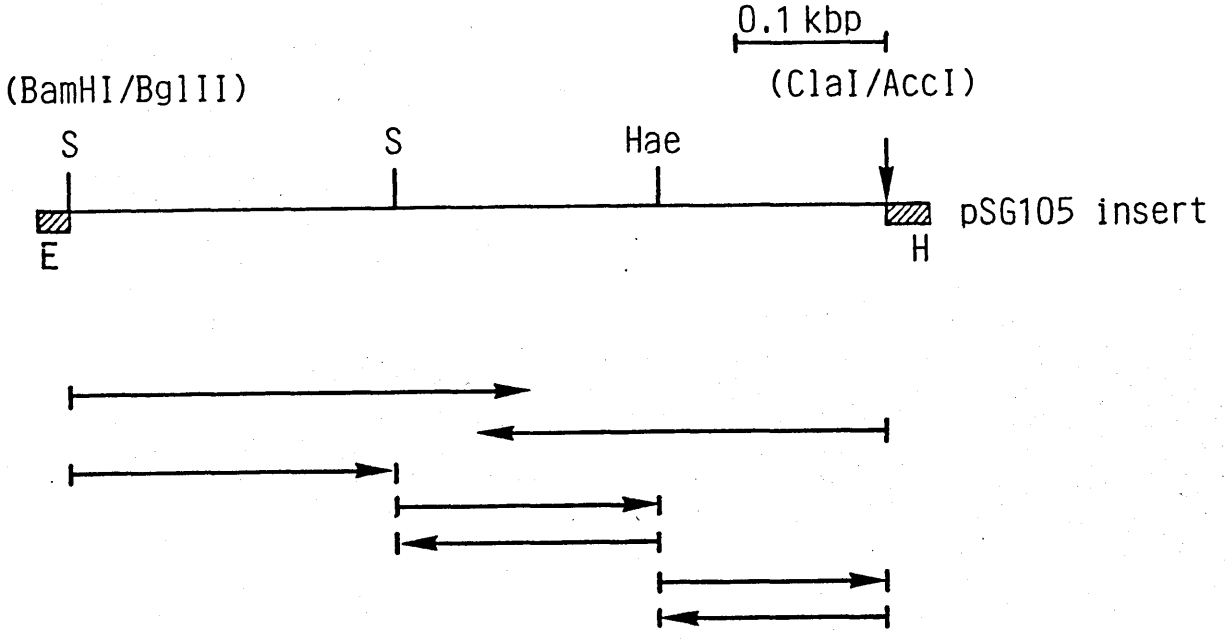


Figure 4.11 Restriction map and sequencing strategy of pSG105

Key to restriction sites: E=EcoRI, H=HindIII, Hae=HaeIII, S=Sau3A

The hatched boxes at the ends of the A. thaliana genomic DNA insert represents the pUC polylinker.

BglII

```
  1  AGATCTTAGG AGAAAAGATG GATTTATAAC CTACAAAAAT TTATAAACTA
 51  ACAAACTAAC CTATAAAAAC ACCTTTACAA TACAACATAT ATATAATATA
101  TATTATTATA ATTTTATGCT AACCACTAAC ACCACAACCA CCACCGCTAC
151  TGTCATCACC ACCACCGCTG CCGGAACTAT CACCACCACC GCTGCCGGAA
201  CTATCACCAC CATCAGATCC ACTGCCGTCA GTGTCACCAG ACCACCACCG
251  TTTTCGGTGC TACCGCTGTT GTAGGAGCAG CCGCAGTCGC AATGCCACTA
301  CCATCGCCGG AACCGCTCAT GTCACCGGAC CACCACCGTC TCTAGCGCTA
351  CCACTGTCGC CGTTGTCGGG CCGCCACAGT CGTCATAACC CCTACCCTCG
401  CCGGAACTGC CACGGTCGTC GGAAGTATGA CCAACTCCTT TAGTTATTTT
451  ATCCACCTAT TAAATTACTT AATTAATTGA TATAAAGTAT GGTTTTTAAA
501  GTTGTTAACT CATAAAATTG TGTCATATGT CAAATATATC GAT
                                                 ClaI
```

Figure 4.12 The DNA sequence of the pSG105 insert

4.4.4 Analysis of the DNA sequence of the pSG105 insert

When the 0.54 kbp insert of pSG105 was translated in all six reading frames there was no open reading frame which extended from one end of the fragment to the other. There were, however, several open reading frames of up to 106 amino acid residues in length and so it was possible that one of these corresponded to the coding DNA for an exon of the A. thaliana E1 gene. A BESTFIT comparison of these open reading frames with the peptide sequence of the E1 domain of S. cerevisiae arom did not identify a homologous sequence. This possibly indicated that the homologous sequence was relatively small compared to the whole open reading frame and that it was being masked by the non-homologous region.

The program BESTFIT was used to compare the S. cerevisiae E1 DNA sequence and the A. thaliana sequence (in both orientations). This allowed the identification of the particular DNA sequence within the pSG105 insert which was homologous to the S. cerevisiae E1 probe and indicated which open reading frames should be examined in detail. Two short sequences of A. thaliana DNA were most homologous to the S. cerevisiae E1 gene (Fig 4.13). One or other of these was apparently responsible for the hybridisation of pSG105 to the S. cerevisiae probe. Both of these sequences occurred within the Sau3A fragment of pSG105.

This result was readily testable. Several double stranded M13 clones, which were generated during sequencing and which contained small

(i)

```
 125  ACTAACACCACAACCACCAC  144
      |||||| |||||| ||  ||
1140  ACTAACCCCACAAGCATTAC  1159
```

(ii)

```
 441  ATATATATTATATATATGTTGTATTGTAAAGGTGTTTTTA  480
      | | |||||  |  ||||||| ||  || |        ||
 988  ACAGATATTGAAGCTATGTTGGATCATACA.....TA  1019

 481  TAGGTTAGTTTGTTAGTTTATAAA  504
      || |||||||  | ||  ||| ||
1020  TAAGTTAGTTCTTGAGAGTATTAA  1043
```

Figure 4.13 Homology of the pSG105 insert with the S. cerevisiae E1 probe

(i) BESTFIT comparison of the pSG105 insert (upper sequence - in the same orientation as Fig 4.13) with the S. cerevisiae E1 region of arom (lower sequence).

(ii) BESTFIT comparison of the pSG105 insert (upper sequence - in the opposite orientation to Fig 4.13) with the S. cerevisiae E1 region of arom (lower sequence).

The sequence of S. cerevisiae ARO1 gene is given in Duncan et al., (1987)and the numbering of the above regions is as in this reference.

regions of the pSG105 insert, were digested with HindIII + EcoRI and electrophoresed on an agarose gel (Fig 4.14a). The southern blot of this gel was probed with the S. cerevisiae E1 probe and the autoradiograph of the blot (Fig 4.14b) indicated that the probe was homologous only to the region of DNA within the pSG105 insert which was bounded by the Sau3A sites. The predictions of the BESTFIT comparison were therefore confirmed.

Because there was a very limited portion of homology of the two sequences at the DNA level, there was obviously correspondingly limited homology when these regions of DNA were translated into amino acid sequence. This homology did not extend outwith these small regions. In addition, the DNA sequence of homology 1 was in part repeated several times within the pSG105 insert (Fig 4.15 (i)). Similarly, regions of homology 2 were repeated either within homology 2 or elsewhere in pSG105 (Fig 4.15 (ii)).

The repetitive nature of the homologous DNA regions and the absence of a homologous amino acid sequence conclusively showed that the A. thaliana DNA within pSG105 did not code for any part of the A. thaliana E1 gene. Consistent with this conclusion was that no signal was obtained from a southern blot of digested pSG105 DNA, probed with the E1 gene of E. coli (aroB).

Since the A. thaliana genome is so small, there was a possibility that the pSG105 insert DNA, or part of it, could encode a product, although not the A. thaliana E1 gene. A WORDSEARCH comparison of the GENBANK

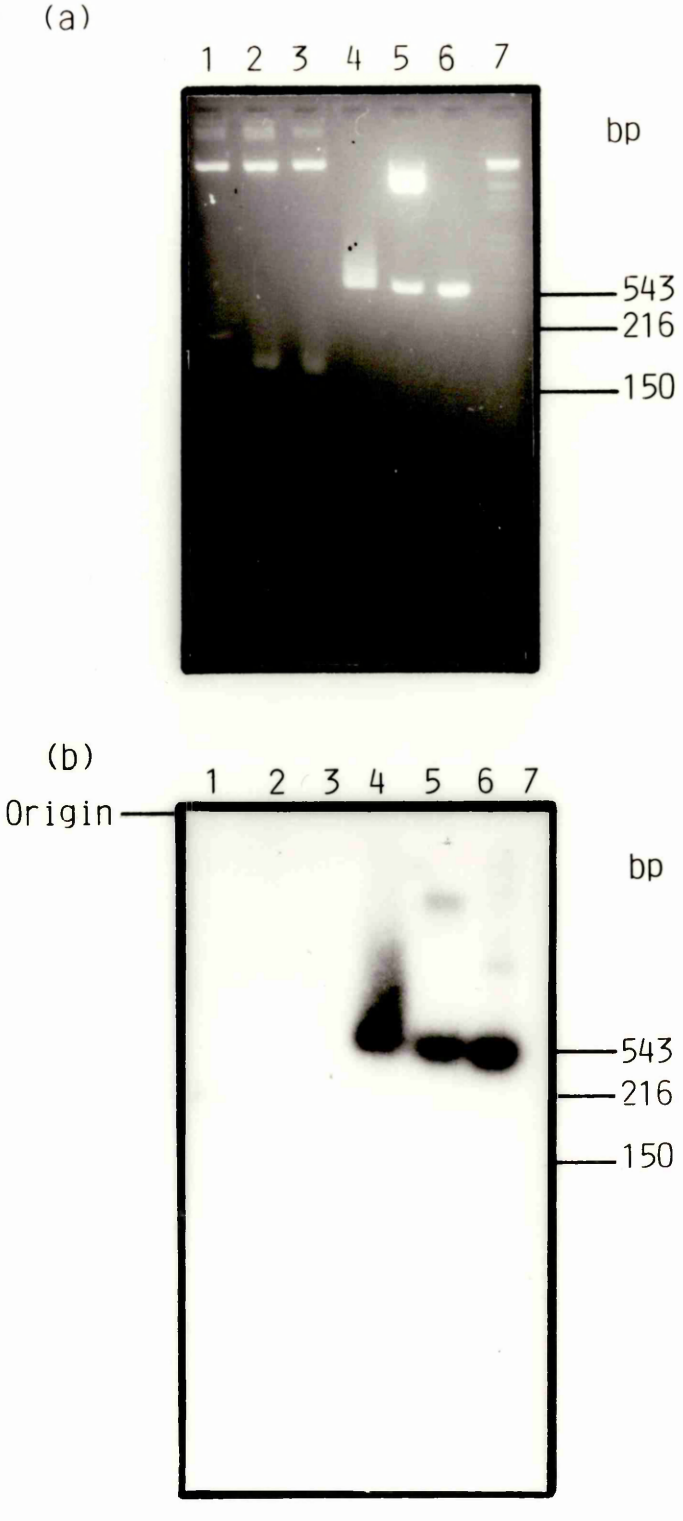
Figure 4.14 The region of pSG105 which hybridised with the S. cerevisiae E1 probe

a) Agarose gel analysis of double stranded M13 clones.

lane:
1. EcoRI + HindIII-digested Mp18 containing Sau3A fragment of pSG105 insert
2. EcoRI + HindIII-digested Mp18 containing Sau3A-HaeIII fragment of pSG105 insert
3. EcoRI + HindIII-digested Mp18 containing HaeIII-AccI/ClaI fragment of pSG105 insert
4. Isolated pSG105 insert
5. EcoRI + HindIII-digested Mp18 containing pSG105 insert
6. Isolated pSG105 insert
7. EcoRI + HindIII-digested λ-DNA

Samples were electrophoresed on a 2% agarose gel.

b) Southern blot of gel (a) hybridised with the S. cerevisiae E1 probe at 30°C in solution containing 50% formamide and washed at 50°C in 2 x SSC, 0.1% SDS.


(a)
1 2 3 4 5 6 7
bp
543
216
150
(b)
Origin
1 2 3 4 5 6 7
bp
543
216
150

(i)

```
125 ACTAACACCACAACC
    |||| |||||| |||
176 ACTATCACCACCACC

125 ACTAACACCAC
    |||| ||||||
200 ACTATCACCAC

130 CACCACAACC
    |||||| |||
157 CACCACCACC

133 CACAACCACC
    ||| ||||||
157 CACCACCACC

133 CACAACCACC
    ||| ||||||
181 CACCACCACC

137 ACCACCACCGCTAC
    |||||||||||| |
158 ACCACCACCGCTGC

137 ACCACCACCGCTAC
    |||||||||||| |
182 ACCACCACCGCTGC

137 ACCACCACCGCT
    |||||||||| |
241 ACCACCACCGTT

155 ATCACCACCACCGCTGCCGGAACTATCACCACCACC
    |||||||||||||||||||||||||||||||||||
179 ATCACCACCACCGCTGCCGGAACTATCACCACCACC
```

(ii)

```
413 TGTTAGT
    |||||||
491 TGTTAGT

441 ATATATAT
    ||||||||
450 ATATATAT

481 TAGGTTA
    |||||||
511 TAGGTTA

484 GTTAGTTT
    ||||||||
492 GTTAGTTT
```

Figure 4.15 Repeat sequences within homology 1 and homology 2

(i) Repeat sequences within all or part of homology 1 (numbering as in Fig 4.13)

(ii) Repeat sequences within all or part of homology 2 (opposite orientation to Fig 4.13)

The programme REPEAT was used to identify these repeat sequences

and EMBL databases with the pSG105 insert DNA sequence or a similar WORDSEARCH comparison of the NBRF proteins sequences databank with each of the open reading frames within the pSG105 insert found no significantly homologous DNA or protein sequence.

4.5 Analysis of the other recombinant phage clones

Only one of the eighteen recombinant phage clones which had been isolated from the A. thaliana genomic library was completely characterised. In order to determine if all of these recombinants were identical, DNA was isolated from each phage, digested with EcoRI and electrophoresed on an agarose gel. The southern blot of this gel was probed at high stringency with the pSG105 insert, (Fig 4.16), and this indicated that only λE1-05 and λE1-12 contained insert DNA which was very similar to that of pSG105. The other recombinants clearly contained other sequences of A. thaliana DNA and so there was still a possibility of one of these clones containing the A. thaliana E1 gene. They were much more likely, however, to contain A. thaliana DNA which had similarly limited homology to the E1 probe as had λE1-12. No further analysis of any of the recombinant phage was carried out.

4.6 Conclusion

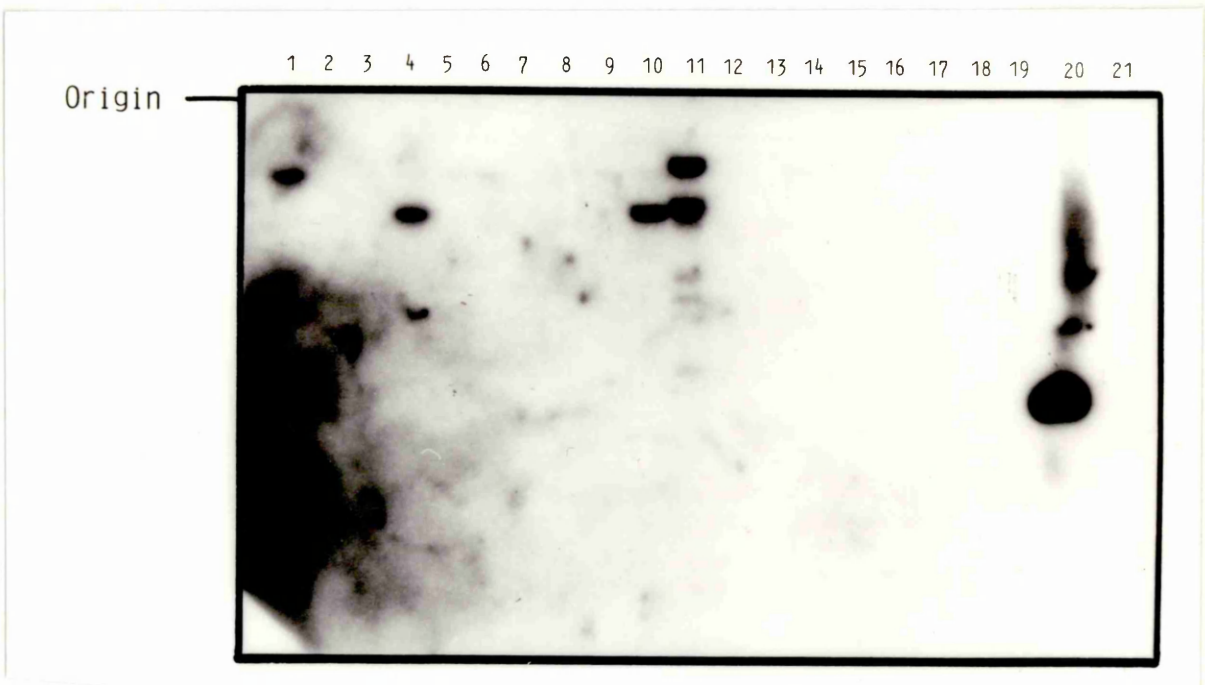
Of the probes used in the screening of the A. thaliana genomic library, only the E1 region of S. cerevisiae ARO1 gene hybridised with plaques. One of the eighteen most strongly hybridising plaques, λE1-12, was characterised and this allowed the identification of two

Figure 4.16 Southern blot containing digested recombinant phage DNA hybridised with the pSG105 insert

Each of the recombinant phage contain regions of A. thaliana genomic DNA.

lane:		lane:	
1.	EcoRI-digested λE1-02 DNA	11.	EcoRI + Hind III-digested λ-DNA
2.	" " λE1-03 "	12.	EcoRI-digested λE1-13 DNA
3.	———	13.	" " λE1-14 "
4.	EcoRI-digested λE1-05 DNA	14.	" " λE1-15 "
5.	" " λE1-06 "	15.	" " λE1-16 "
6.	" " λE1-07 "	16.	" " λE1-17 "
7.	" " λE1-09 "	17.	" " λE1-18 "
8.	" " λE1-10 "	18.	" " λE1-19 "
9.	" " λE1-11 "	19.	" " λE1-20 "
10.	" " λE1-12 "	20.	pSG105 insert
		21.	EcoRI-digested λE1-21 DNA

The samples were electrophoresed on a 1% agarose gel. The southern blot of this gel was hybridised with the pSG105 insert (labelled by nick translation) at 42°C in 50% formamide and washed at 65°C in 0.1 x SSC, 0.1% SDS.



very small regions of homology between the λE1-12 DNA and the probe. However it was shown that the DNA present in λE1-12 simply contained fortuitous homology with the S. cerevisiae E1 gene, which was sufficient for it to be isolated from the genomic library at the very low stringency conditions used for screening. λE1-12 did not contain the A. thaliana E1 gene nor was it likely that any of the other purified recombinant phage contained the E1 gene. Although there were a few short repeated sequences within the pSG105 insert, the fragment was still unlikely to be repetitive because firstly only two of the other eighteen strongly hybridising recombinant phage contained a similar insert and secondly the main reason for examining the eighteen strongly hybridising recombinants in the first place was because the pattern produced on the screening autoradiograph suggested that they did not contain repeated sequences. A BESTFIT comparison of the pSG105 insert with the five A. thaliana repetitive sequences (Martinez-Zapater et al., 1986; Simoens et al., 1988) present in the EMBL database did not show any homology between the pSG105 insert and the repetitive sequences which is also consistent with the pSG105 insert not being a repetitive sequence.

These results effectively ended the attempts to clone the shikimate pathway genes from higher plants using probes derived from the microbial genes.

CHAPTER 5

THE ISOLATION OF A cDNA CLONE FOR EPSP SYNTHASE FROM Pisum sativum

5.1 Introduction

Since the heterologous probing approach had failed (see chapters 3 and 4) the next most promising cloning strategy for the genes of the shikimate pathway enzymes was to use information from protein sequence determination to design appropriate oligonucleotide probes. These would be used to screen a suitable *P. sativum* cDNA library. The obvious candidates for initial attempts at this were EPSP synthase (E5), 3-dehydroquinase/shikimate dehydrogenase (E2/E3) and DAHP synthase (E0) since these were the most characterised enzymes of the shikimate pathway in higher plants. Our principle objective was still to obtain the E2/E3 gene and so our efforts were focussed mainly on this gene. A prerequisite for isolating a *P. sativum* cDNA clone was a suitable cDNA library and the preparation of such a library is described in the following sections.

During the course of the work described in this thesis, the Monsanto Group reported the successful cloning of a cDNA for *P. hybrida* E5 using this strategy (Shah *et al.*, 1986); Gasser *et al.*, 1988). This reinforced our view that microprotein sequencing and probe design was the correct direction to move in. A cDNA clone for the *P. hybrida* E5 was given to us by Dr D. Shah and this was used to probe a *P. sativum* cDNA library. This chapter describes the isolation of a cDNA for *P. sativum* E5 and the analysis of its DNA sequence.

5.2 The RNA used for the generation of a P. sativum cDNA library

5.2.1 The P. sativum tissue used for RNA preparation

The only expression studies on shikimate pathway genes which had been reported were on E5 genes. There is high level transcription of the E5 gene in the purple petals of P. hybrida whereas the yellow flowers of L. esculentum do not exhibit these elevated levels of mRNA (see section 1.8.3). It seems likely that high levels of shikimate pathway enzymes may be a characteristic of purple flowers which contain the aromatic pigment anthocyanin (Gasser et al., 1988). The flowers of P. sativum are white in colour, presumably because they lack this pigment, and so it was anticipated that expression of the shikimate pathway enzymes in P. sativum flowers would not be significantly higher than in other parts of the plant.

Mousdale and Coggins have observed that relatively high shikimate pathway enzyme activities are obtained in newly emerged leaves of 7 to 10 day pea seedlings compared to roots and stems . They have also noticed a relationship between the greenness of the tissue and the amount of enzyme activity present. This is probably due to a higher number of mature chloroplasts within older pea seedlings and therefore a corresponding increase in shikimate pathway enzyme activity since the enzymes are predominantly located in the chloroplast, (Mousdale and Coggins, 1985). The mRNA which was used to synthesise the P. sativum cDNA library was isolated from 7 to 10 day old pea seedlings, at a stage where the seedlings had become a deep green colour but

where the leaves had not completely opened out. Older seedlings that had fully opened leaves were avoided because the increase in size of older seedlings is predominantly due to cell expansion as opposed to cell proliferation.

5.2.2 Isolation of poly A^+ mRNA

Small quantities of poly A^+ mRNA had previously been isolated from P. sativum and this poly A^+ mRNA was shown to hybridise with a probe for a highly expressed plant gene (section 3.4.1). Larger amounts of poly A^+ mRNA were required for the preparation of a cDNA library and so the same method of RNA preparation was employed but RNA was prepared on a larger scale.

For optimal cDNA synthesis it is important to have a pure and undegraded mRNA preparation. Therefore the precautions described in section 2.22.1 were strictly adhered to. Several RNA preparations were carried out in order to optimise the isolation conditions. The poly A^+ mRNA preparation which was used in subsequent cDNA syntheses is described below.

9,380 µg of total RNA was isolated from approximately 80 nine day old pea seedlings which had been stored at -80°C for several weeks. An aliquot containing 100 µg of total RNA was removed and stored before applying the remainder to an oligo-dT column. After two passages through the column, a total of 170 µg of poly A^+ mRNA, which represented approximately 2% of the total RNA, was eluted in a volume

of 8 ml. The yield of poly A^+ mRNA was within the recommended range of 2-5% (Jacobson, 1987) suggesting that the separation had been successful.

5.2.3 The integrity of the poly A^+ mRNA

Examination of the translational activity in *in vitro* translation systems can act as an indicator of the purity and integrity of the mRNA.

There are two types of *in vitro* translation systems in common use; those derived from rabbit reticulocytes and those from wheat germ extracts. Messages which can be translated well in one system may be poorly translated in the other. It has been suggested that this is because the systems are unequally sensitive to inhibitors (Hames and Higgins, 1984). Since the integrity of the *P. sativum* poly A^+ mRNA preparation was under examination it may have seemed obvious to use the translation system derived from a plant, ie the wheatgerm system. The reticulocyte system was favoured however because it has been reported that this system is usually better for the synthesis of large proteins (Boulnois, 1987). *In vitro* translation was being utilised to demonstrate that large polypeptides could be synthesised from the *P. sativum* poly A^+ mRNA preparation and therefore use of a system which was most efficient at doing this seemed sensible.

Samples of total RNA and poly A^+ mRNA were translated using the Amersham rabbit reticulocyte (nuclease treated) *in vitro* translation

system. ^{35}S labelled methionine was incorporated into the translation products and volumes of each translation reaction were examined by SDS-PAGE. The autoradiograph of the dried gel indicated that polypeptides with sizes at least up to 100 kDa were being synthesised using both the total RNA and the poly A^+ mRNA (Fig 5.1). There were more products formed from 100 ng of poly A^+ mRNA than from 250 ng or 500 ng of poly A^+ mRNA. This indicated that the translating ability of the lysate was adversely affected by concentrations of poly A^+ mRNA which were higher than optimal, which probably reflected the presence of inhibitory components in the poly A^+ mRNA preparation. The optimal amount of poly A^+ mRNA in the reaction could have been above or below 100 ng but further investigation into its value was not carried out.

The translation of 2 μg of total RNA produced more polypeptide products than 5 μg of total RNA. This suggested that the optimal amount of total RNA in the translation should be less than 5 μg.

More polypeptide products were produced from 100 ng of poly A^+ mRNA than from 2 μg of total RNA and so an enrichment for coding potential of more than 20 fold was being achieved by separating the poly A^+ mRNA from the poly A^- RNA. It had been demonstrated that only 2% of the total RNA bound to the oligo-dT column (section 5.2.2) which represented an enrichment of about 50 fold. The estimations of the enrichment of coding potential were fairly consistent. It was concluded that the poly A^+ mRNA which had been isolated contained a large proportion of the coding potential of the total RNA. In addition, the poly A^+ mRNA was able to code for large polypeptides

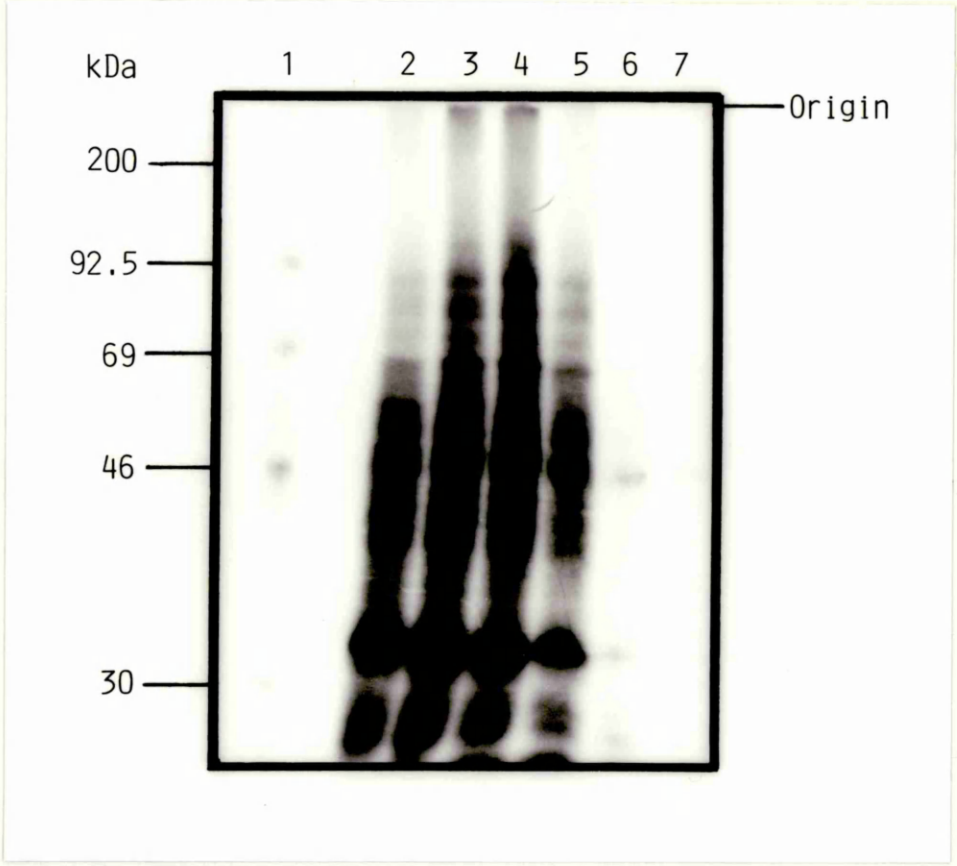


Figure 5.1 In vitro translations of P. sativum RNA

The RNA samples listed below were translated with a rabbit reticulocyte lysate system and samples of the translation products were electrophoresed on a 10% polyacrylamide gel. The dried gel was fluorographed.

lane:
1. ^{14}C labelled protein markers
2. 500 ng P. sativum poly A^+ mRNA
3. 250 ng " " "
4. 100 ng " " "
5. 2 µg " total RNA
6. 500 ng " " "
7. No RNA

which suggested that it was of sufficiently good quality to be used for the construction of a cDNA library.

5.2.4 Fractionation of the poly A^+ mRNA

The majority of the translation products (Fig 5.1) from the P. sativum poly A^+ mRNA were less than 50 kDa in size. Although some of these products were probably truncated versions of larger polypeptides a large proportion of the small polypeptides would be encoded by correspondingly small mRNA molecules present within the sample of poly A^+ mRNA used for translation.

When constructing a cDNA library with the aim of obtaining clones for P. sativum E2/E3 and E5, anything which can be carried out to enrich the mRNA population for the transcripts encoding these enzymes, would make the identification of the appropriate cDNA clones much easier. The tissue from which the mRNA was isolated was thought to contain the highest possible amount of the transcripts. Before the poly A^+ mRNA was used to synthesise cDNA it was fractionated according to its size in the hope that a further enrichment of the E2/E3 transcripts would be achieved. It was also hoped that this fractionation would remove small mRNA molecules which would produce small, easily clonable cDNA.

P. sativum E2/E3 and E5 have been shown to be monomeric with Mr 59,000 and 50,000 respectively, as judged by gel filtration and SDS-PAGE (Mousdale and Coggins, 1984; Mousdale et al., 1987). These are the sizes for the native chloroplastic enzymes. There are presumably

precursor E2/E3 and E5 polypeptides which have a larger Mr because they contain a transit peptide (perhaps up to 10 kDa in size), to enable translocation of the precursor enzyme to the chloroplast.

The isolation of the mRNA population of P. sativum which encodes polypeptides within the size range 46-75 kDa, would enrich for mRNA molecules encoding the E2/E3 and E5 enzymes. This would also allow for any error in the estimation of Mr of E2/E3 and E5.

100 µg of P. sativum poly A^+ mRNA was completely denatured and applied to a sucrose density gradient (5-20% sucrose, volume 4 ml). This gradient contained 70% formamide to prevent the mRNA population from aggregating during sedimentation through the sucrose. After centrifugation the gradient was fractionated into 12 approximately equal fractions. The amount of poly A^+ mRNA present in each fraction is indicated in Table 5.1.

The RNA was ethanol precipitated out of each fraction of the gradient and 0.5 µg samples were translated using the rabbit reticulocyte lysate. 0.5 µg of total poly A^+ mRNA, and 2 µg of both poly A^- RNA and total RNA were also translated. One tenth of the samples were removed from each reaction and the ^{35}S labelled products were TCA precipitated and counted by liquid scintillation (Table 5.2). Although equal amounts of poly A^+ mRNA from each fraction were added to each translation reaction, unequal amounts of labelled products were produced from the various translations as judged by the TCA precipitation results. This probably reflected differences in the

Fraction	poly A^+ mRNA (μg)
1	1.9
2	2.4
3	3.0
4	3.2
5	16.9
6	21.1
7	28.0
8	8.5
9	6.7
10	1.1
11	-
12	-
	Total 92.8 μg

Table 5.1 Fractionation of the poly A^+ mRNA

The quantity of poly A^+ mRNA present in each of the 12 fractions from the denaturing sucrose gradient. Each fraction was increased in volume to 350 μl using TE and the A_{260} determined in a 0.2 cm path length quartz cuvette.

RNA present in translation	TCA precipitable cpm	RNA present in translation	TCA precipitable cpm
fraction - 1 poly A^+	16787	fraction - 9 poly A^+	32776
" - 2 "	13192	" - 10 "	33574
" - 3 "	18473	" - 11 "	19266
" - 4 "	27046	No RNA	13330
" - 5 "	199517	poly A^-	21029
" - 6 "	240123	poly A^+	139416
" - 7 "	45005	Total RNA	68833
" - 8 "	83263		

Table 5.2 TCA precipitable cpm present in translations of poly A^+ mRNA fractions.

The values in the table represent TCA precipitable cpm from 0.05 μg of poly A^+ mRNA from various fractions, 0.2 μg of poly A^- and 0.2 μg of total RNA.

translatability of each poly A^+ mRNA fraction. mRNA molecules have sedimentation coefficient ranging from 6S to 35S or more, with the vast majority in the 16S-20S range (Hames and Higgins, 1984). Sucrose density gradient centifugation is therefore particularly useful for purifying very large or very small mRNA molecules. Contaminating tRNA, 5S rRNA and 25S rRNA, which would otherwise be present even after passage through the oligo-dT column, would have sedimented in regions of the gradient which were above or below the majority of the mRNA. (Contaminating 18S rRNA would sediment in the same region as the majority of mRNA molecules). Therefore fractions towards the top or the bottom of the gradient would contain less translatable RNA than fractions from the middle, which was consistent with the TCA precipitation results.

20,000 cpm from each translation reaction was subjected to SDS-PAGE and the dried gel was fluorographed (Fig 5.2 (i)). There were a few discrete products present in the Poly A^- RNA track which were not accounted for by the water blank. However when this translation pattern was compared to that of the total poly A^+ RNA it was clear that very little translatable material had been lost during isolation of the poly A^+ mRNA confirming that the separation had been successful.

Fractions from nearer the top of the gradient were translated into relatively low molecular weight products whereas fractions from nearer the bottom of the gradient were translated into higher molecular weight products. This was more obvious if each lane was compared to

Figure 5.2 In vitro translation of fractionated P. sativum poly A^+ mRNA

P. sativum RNA samples, listed below, were translated using a rabbit reticulocyte lysate system and the products were electrophoresed on 10% polyacrylamide gels. The dried gels were fluorographed.

Fraction-1 is from the top of the denaturing sucrose gradient whereas fraction-11 is from the bottom of the gradient.

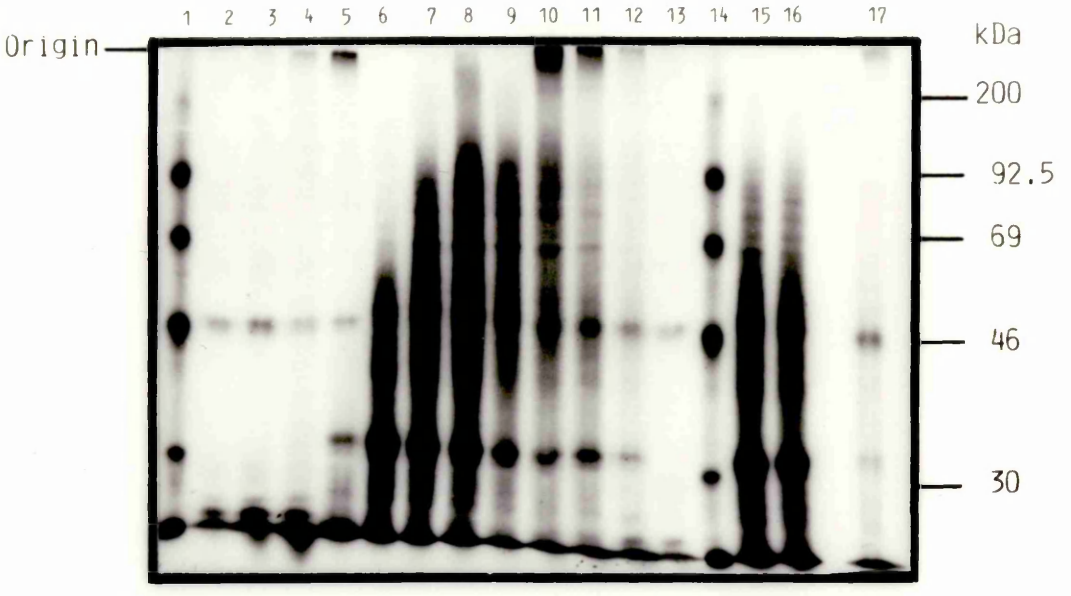
(i) lane:

1. ^{14}C labelled protein markers
2. fraction-1 poly A^+ mRNA
3. " -2 " "
4. " -3 " "
5. " -4 " "
6. " -5 " "
7. " -6 " "
8. " -7 " "
9. " -8 " "
10. " -9 " "
11. " -10 " "
12. " -11 " "
13. No RNA (H_2O blank)
14. ^{14}C labelled protein markers
15. total poly A^+ mRNA
16. total RNA
17. poly A^- RNA

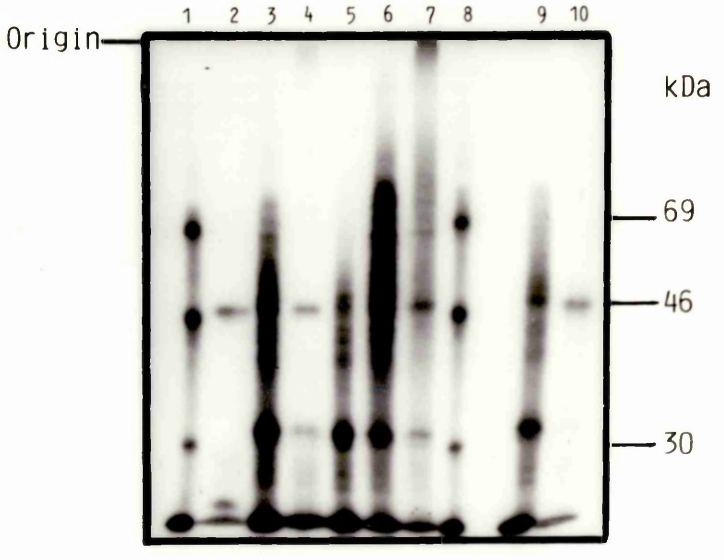
(ii) lane:

1. ^{14}C labelled protein markers
2. No RNA (H_2O blank)
3. total poly A^+ mRNA
4. fraction-4 poly A^+ mRNA
5. " -5 " "
6. " -7 " "
7. " -9 " "
8. ^{14}C labelled protein markers
9. total RNA
10. poly A^- RNA

(i)


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Origin
kDa
200
92.5
69
46
30

(ii)


1 2 3 4 5 6 7 8 9 10
Origin
kDa
69
46
30

the lane containing the translation products of total poly A^+ mRNA. The extent of fractionation was more obvious when a second gel, containing every second fraction was run and fluorographed (Fig 5.2ii). It is likely that a significant proportion of the smaller molecular weight products present in lanes containing translations of larger poly A^+ mRNA species, were a result of truncated polypeptides which migrated faster in the gel than fully formed products.

The fractions of poly A^+ mRNA which encoded the majority of polypeptides within the size range of 46-70K were fractions 7, 8 and 9. These fractions contained 28, 8.5 and 6.7 μg of poly A^+ mRNA respectively. The latter two fractions contained so little poly A^+ mRNA that subsequent optimisation of cDNA synthesis procedures would be limited by the quantity of RNA. Rather than pool the three fractions, fraction-7 was chosen to prepare the cDNA library.

It would have been very appropriate at this stage to attempt to further identify which fraction of poly A^+ mRNA contained the mRNA coding for the E2/E3 or E5 polypeptide by screening translation products with antibodies or screening Northern blots with an appropriate nucleic acid probe. However neither of these probes, antibody or nucleic acid, were yet available and so the enrichment of the mRNA was based solely on the size of the translation products

5.3 cDNA synthesis

The starting point of any cDNA cloning procedure is the efficient

synthesis of double stranded cDNA. This is dependent on the quality of the isolated mRNA. It appeared, from its high translating capacity, that the P. sativum poly A^+ mRNA (fraction-7) was of sufficiently high quality (section 5.2.4) for its use in the synthesis of cDNA. cDNA synthesis is also dependent on the efficiency of the enzymes and the optimisation of reaction conditions during the cDNA synthesis. A commercially available cDNA synthesis system was utilized to minimise these potential difficulties.

The strategy for detecting the clone of interest must be considered prior to a cDNA cloning programme because this will strongly influence the strategy of cDNA synthesis. The Amersham cDNA synthesis system was chosen because it offered two appropriate methods of generating cDNA from the poly A^+ mRNA both of which appeared to offer an advantage for clone detection.

5.3.1 Rational of the cDNA synthesis

The basis of the Amersham cDNA synthesis system is outlined in Fig 5.3. First strand cDNA synthesis is primed with either oligo-dT or random primers and synthesised using reverse transcriptase. Second strand synthesis is performed using the mRNA/cDNA hybrid as a substrate for E. coli ribonuclease H, an endonuclease which digests RNA in an RNA/DNA hybrid only and is used to produce nicks and gaps in the mRNA strand. This provides RNA primers with 3'OH for E. coli DNA polymerase I to replace the mRNA strand with DNA in a reaction similar to nick translation. T4 DNA polymerase is used to remove any small

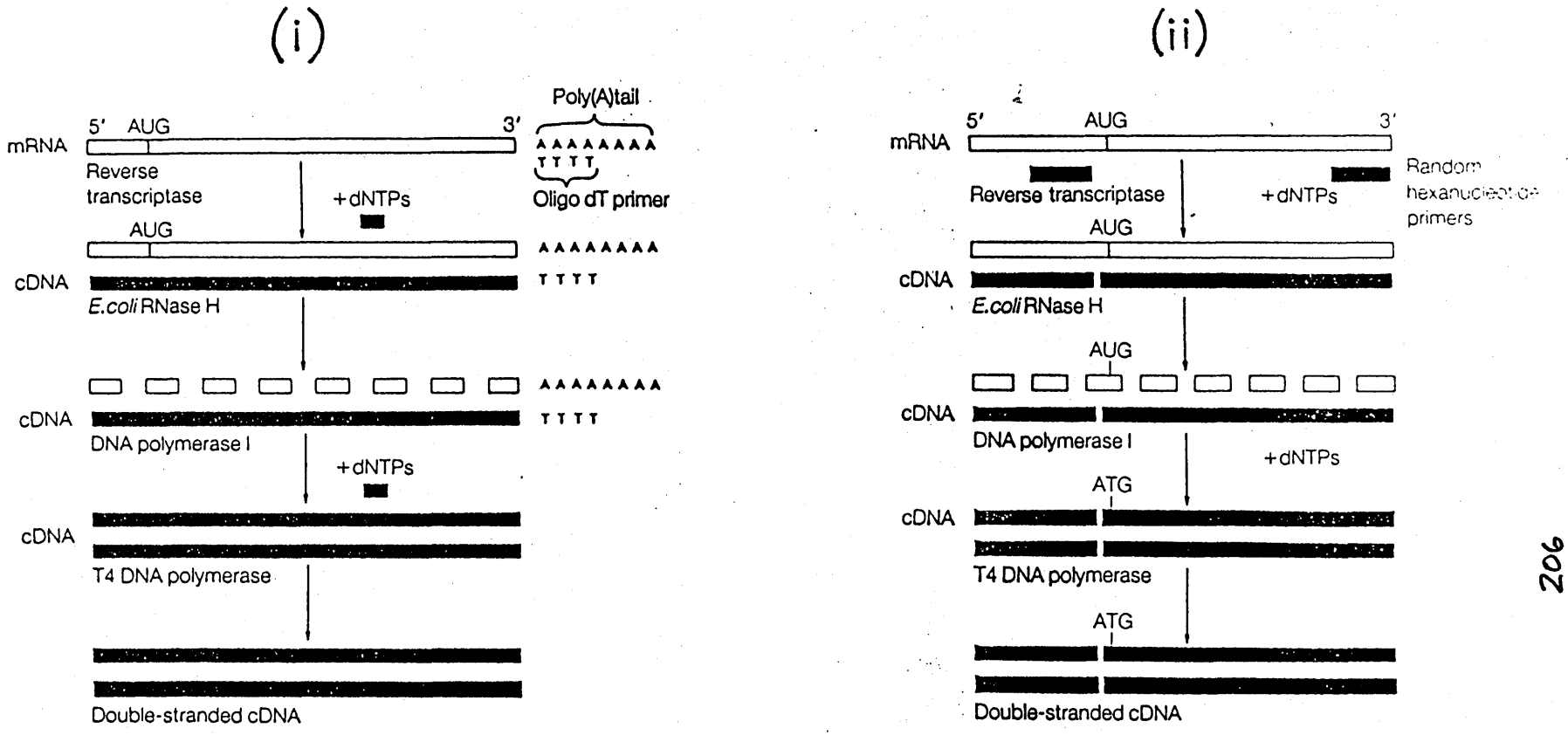


Figure 5.3 The cDNA synthesis procedure

First strand cDNA synthesis was primed with either (i) oligo-dT or (ii) random hexanucleotide primers. Second strand synthesis and subsequent steps were similar.

3'-overhangs on the first strand cDNA.

This procedure does not depend on a hairpin loop priming second strand synthesis and avoids the S1 nuclease digestion of the hairpin loop prior to the cloning of the cDNA. These latter steps are widely used in many other cDNA preparation protocols (Kimmel and Berger, 1987). The hairpin loop is a relatively inefficient primer which makes this method of second strand synthesis a poorly controlled step. The SI nuclease step is a major disadvantage because it can result in a significant loss of important cDNA sequences and produce libraries with low yields of recombinant DNA's that contain full length cDNA sequences (Okayama and Berg, 1982).

In addition to these advantages, the Amersham system offers two alternative methods for priming first strand cDNA synthesis (Fig 5.3). Oligo-dT will bind to the poly A tail of mRNA and act as a primer. It enables the production of cDNA libraries from which there is a chance of obtaining full length cDNA clones. However the libraries are particularly rich in sequences near the 3'-end of mRNA molecules. The alternative primers for first strand cDNA synthesis are random hexanucleotides, which prime randomly from internal positions within the mRNA molecules. Libraries generated using these primers are unlikely to contain full length cDNA clones but they will not have a predominance of sequences from the 3'-end of the mRNA.

It was planned that the cDNA library would eventually be screened with an oligonucleotide probe derived from the N-terminal sequence of the

P. sativum E2/E3 enzyme. Only cDNA clones containing the sequences from the 5'-end of the E2/E3 mRNA would be detected with this probe. If for some reason the reverse transcriptase does not generate full length cDNA molecules during first strand synthesis primed with oligo-dT then the E2/E3 cDNA clones would not be detected. A cDNA library was therefore also derived from E2/E3 and E5 enriched P. sativum poly A^+ mRNA (fraction-7 from the sucrose gradient), where first strand cDNA synthesis was primed with random primers. The cDNA used to produce this library was designated random fraction-7 cDNA.

The P. hybrida E5 cDNA clone was also to be used to screen a cDNA library. The successful isolation of the P. sativum cDNA clone would not be dependent in this case on the presence of sequences from the 5'-end of E5 mRNA since the probe is a full length clone. A second cDNA library was therefore constructed using fraction-7 poly A^+ mRNA, where first strand synthesis was primed with oligo-dT. The cDNA used to produce this library was designated oligo-dT fraction-7 cDNA. This library could very well provide a full length cDNA clone of E5. Full length cDNA clones of E2/E3 may also be present within this library and so it could be screened with any partial E2/E3 clone isolated from the first library, or alternatively could be screened directly with the E2/E3 oligonucleotide.

A third cDNA library was also constructed. Total P. sativum poly A^+ mRNA was used and first strand synthesis was primed with oligo-dT. The cDNA used to produce this library was called oligo-dT total cDNA. This library was generated as a precaution in case either of the E2/E3 or

E5 messages were under-represented in fraction-7 of the sucrose gradient. This may have occurred if the mRNA molecules coding for either enzyme were exceedingly large ie, if they contained long 5' or 3'-untranslated regions.

5.3.2 Synthesis of random fraction-7 cDNA

The poly A^+ mRNA which had been prepared was not available in unlimited quantities. Rather than carry out the cDNA synthesis reactions simultaneously, they were carried out separately, with random fraction-7 cDNA being prepared first. This allowed any problems, which arose during the synthesis to be overcome without wasting a large proportion of the poly A^+ mRNA. The Amersham cDNA synthesis system included a sample of rabbit globin mRNA from which cDNA was prepared. This control synthesis was intended to check the performance of the whole cDNA synthesis process.

cDNA synthesis was carried out using 1 μg of fraction-7 poly A^+ mRNA and 500 ng of control globin mRNA where first strand synthesis was primed using random hexanucleotides. Samples were removed from the first and second strand reactions and the remainder was stored at -20°C. The efficiency of cDNA synthesis was examined by methods outlined in section 2.25 and results are given in Table 5.3.

The percentage of *P. sativum* fraction-7 poly A^+ mRNA reverse transcribed into cDNA (yield 1) was significantly greater than the suggested value (of 15-30%) indicating that efficient first strand

cDNA synthesis	filter	mean cpm	% first strand incorporation	% second strand incorporation	Yield 1	Yield 2	Quantity of double stranded cDNA
1	A B C D	54473 2388 95962 4100	4.4%	3.8%	61.6%	86.4%	1060 ng
2	A B C D	104103 3541 136168 3602	3.4%	2.2%	41.6%	74%	616 ng
3	A B C D	114962 3997 149873 5018	3.5%	2.9%	49%	82.8%	812 ng

Table 5.3 Efficiency of the cDNA syntheses from fraction - 7 poly A^+ mRNA

Filters were prepared in duplicate. The figure used in the calculation was the mean of the two values obtained. Yield 1 represents the percentage of mRNA transcribed into cDNA and Yield 2 represents the percentage of first strand cDNA transcribed into second strand cDNA. The total amount of random fraction - 7 cDNA synthesised is given.

synthesis had taken place. Yield 2 approached the suggested value of 90% which showed that a large proportion of first strand cDNA had been transcribed into the second strand cDNA. The efficiency of both first and second strand synthesis from globin mRNA was low compared to that from both P. sativum fraction-7 poly A^+ mRNA and to the suggested values. This was consistently found to be the case for the globin mRNA, over several rounds of cDNA synthesis.

The first and second strand cDNA products were analysed on an alkaline agarose gel (Fig 5.4). First strand products gave a less intense signal than second second strand products from P sativum poly A^+ mRNA because there was ten fold less label in the first strand reaction and therefore the products were less radioactive. The globin cDNA products were difficult to visualise until the gel was exposed to autoradiography for several days. The products from the P. sativum random fraction-7 cDNA synthesis had sizes up to 4 kb although most of the products were less than 2.26 kb. This was very encouraging because it suggested that large cDNA molecules were being synthesised even although internal priming was (presumably) occurring within the mRNA. The efficiency of cDNA synthesis provided further evidence for fraction-7 mRNA being of sufficiently good quality to pursue cDNA cloning procedures.

It was anticipated that a proportion of the cDNA would be lost during the purification by phenol extraction and ethanol precipitation. To accommodate such losses, further synthesis of random fraction-7 cDNA was carried out using another 2 µg of fraction-7 poly A^+ mRNA in two

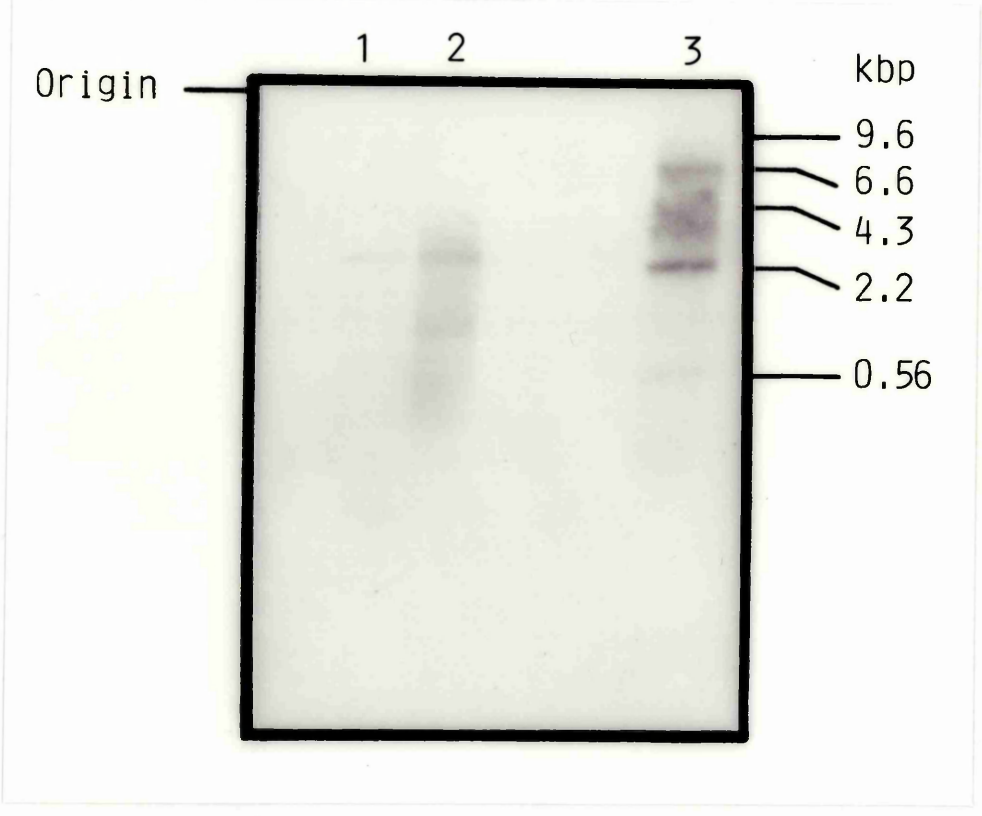


Figure 5.4 Analysis of the random fraction-7 cDNA

lane: 1. random fraction-7 first strand cDNA
2. " " second " "
3. HindIII-digested λ-DNA (single stranded)

The samples were electrophoresed on a 1.4% alkaline agarose gel

standard cDNA synthesis reactions. The results (Table 5.3) indicated that the cDNA synthesis had been efficient and analysis of the products on an alkaline agarose gel (Fig 5.5) confirmed that high molecular weight cDNA had been prepared.

5.3.3 Synthesis of oligo-dT primed cDNA

P.sativum oligo-dT fraction-7 cDNA and oligo-dT total cDNA was synthesised from 3 μg of both fraction-7 poly A^+mRNA and total poly A^+ mRNA respectively. The efficiency of the syntheses was investigated as detailed in section 2.25 and the results are given in Table 5.4.

Yields 1 and 2 of both syntheses were very similar and were in the range of the suggested values, indicating that the cDNA synthesis from both mRNA population had been very efficient. When yield 1 of the oligo-dT fraction-7 cDNA was compared to yield 1 of the random fraction-7 cDNA it was clear that just over half as much fraction-7 mRNA was being transcribed to first strand cDNA using oligo-dT as a primer compared with the first strand cDNA produced using random hexanucleotide primers.

Samples of the products from second strand synthesis of the 2 types of P. sativum oligo-dT primed cDNA were analysed on an alkaline agarose gel alongside products from the second strand synthesis of P. sativum random fraction-7 cDNA and globin random cDNA (Fig 5.5). The difference in the size range of products from random primed and oligo-dT primed first strand synthesis was very obvious. cDNA up to at

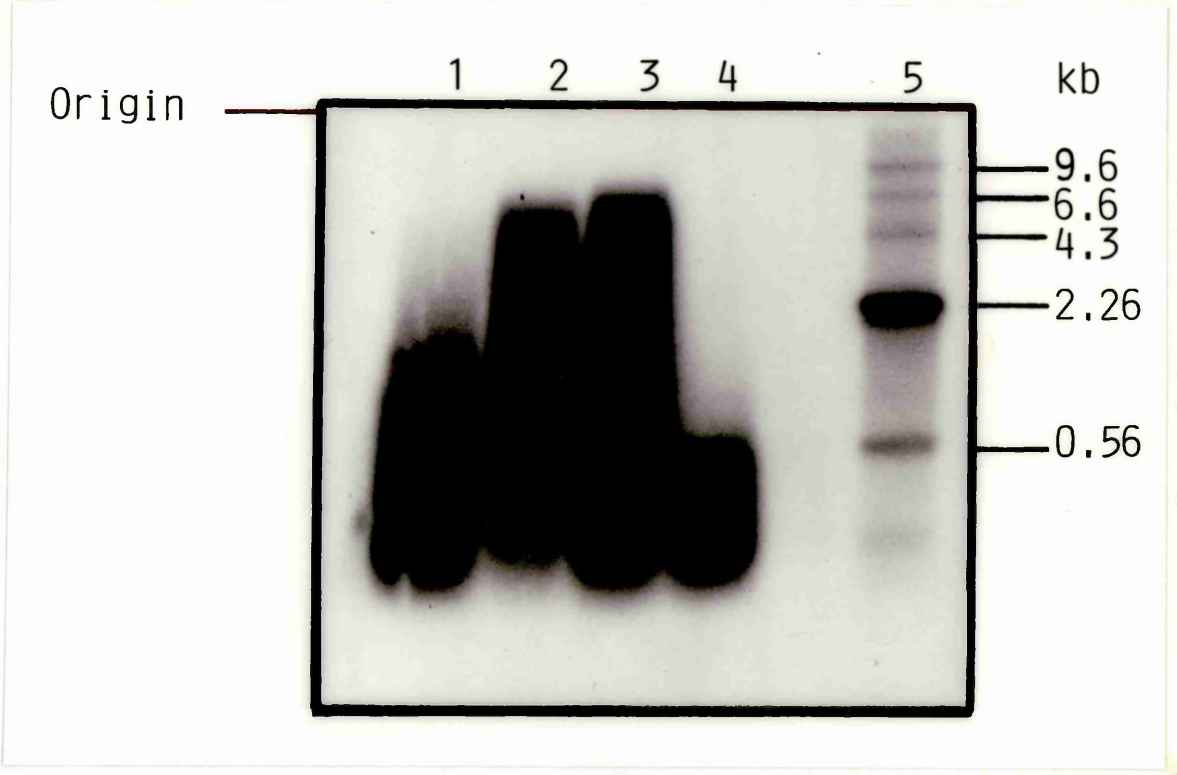


Figure 5.5 Analysis of the three cDNA populations

lane: 1. random fraction-7 cDNA
2. oligo-dT " "
3. " total poly A^+ cDNA
4. random control globin cDNA
5. HindIII digested λ-DNA (single stranded)

The samples were electrophoresed on a 1.4% alkaline agarose gel.

cDNA synthesised	filter	mean cpm	% first strand incorporation	% second strand incorporation	Yield 1	Yield 2	Quantity of double stranded cDNA
oligo-dT fraction - 7 cDNA	A B C D	78188 1994 183244 4768	2.6%	2.4%	35.7%	94%	2016 ng
oligo-dT total cDNA	A B C D	80306 2053 164508 4005	2.6%	2.2%	35.8%	86.4%	1854 ng

Table 5.4 Efficiency of the cDNA syntheses primed with oligo-dT

3 cDNA synthesis reactions using fraction - 7 poly A^+ mRNA were pooled and samples removed at the appropriate times (section 2.25). 3 cDNA synthesis reactions using total poly A^+ mRNA were similarly treated. Filters were prepared in duplicate. The figure used in the calculation was the mean of the two values obtained. Yield 1 represents the percentage mRNA transcribed into cDNA and Yield 2 represents the percentage of first strand cDNA transcribed into second strand cDNA. The total amount of oligo-dT fraction - 7 cDNA and oligo-dT total cDNA synthesised is given.

least 6.6 kb in length was being synthesised using oligo-dT as first strand primer from both the total and the enriched fraction-7 poly A^+ mRNA whereas virtually all of the cDNA from random primed first strand synthesis was less than 2.26 kb in size. This was a result of the random primers priming the cDNA synthesis from internal sites within the mRNA and therefore producing shorter cDNA molecules compared to oligo-dT which primed cDNA synthesis from the 3'-end of the mRNA only and therefore allowed synthesis of full length cDNA molecules. The cDNA products which were smaller than 0.56 kb were less abundant in the oligo-dT fraction-7 cDNA than in the oligo-dT total cDNA, reflecting the reduction in the amount of very small mRNA molecules in the fractionated poly A^+ mRNA.

From the sizes of the polypeptides (see section 5.2.4) it was expected that a mRNA of at least 2 kb would encode the E2/E3 protein and a mRNA which was just marginally smaller would encode the E5 protein. The large size of both oligo-dT primed cDNA populations was very encouraging since it made the prospect of obtaining a full length cDNA clone of P. sativum E2/E3 and E5 more likely.

5.3.4 Purification of the cDNA

The three P. sativum cDNA populations were phenol extracted and ethanol precipitated in the presence of high quality glycogen. The glycogen acted as a carrier for the precipitation of the cDNA in a similar way to tRNA or sonicated DNA.

Because the cDNA was radioactively labelled it could easily be followed during the purification. It was found that very few cpm were present in the organic phase after phenol extraction, but that a significant proportion was present outwith the cDNA pellet after ethanol precipitation. The cpm remaining in the supernatant increased dramatically when glycogen was not included, as was observed during trial cDNA synthesis reactions.

It was essential to determine how much cDNA had been recovered after purification and this was easily done by comparing the cpm of the cDNA pellet, with the total cpm before cDNA purification (Table 5.5). It was recommended (Amersham cDNA cloning booklet) that between 0.5 and 1 μg of purified cDNA was required and the results given in Table 5.5 suggested that sufficient cDNA from the three populations had been purified.

5.4 Cloning the cDNA into λgt10

5.4.1 The cloning rational

The P. sativum cDNA libraries were to be screened using nucleic acid probes. The vector λgt10 was chosen as the cloning vector because of the advantages that were discussed in section 3.6.3. Firstly the vector could accommodate the size range of oligo-dT fracton-7 cDNA and oligo-dT total cDNA (ie at least 6.6 kbp) and secondly recombinants could be distinguished by the efficient biological selection (Huynh et al., 1985; Jendrisak et al., 1987). In addition commercially

cDNA synthesised	Total cpm before cDNA purification	Total cpm in cDNA pellet	% of total cDNA recovered	ng cDNA recovered
random fraction - 7 cDNA	11232000	5569000	49.6%	1240 ng
oligo-dT fraction - 7 cDNA	13281000	4532000	34.1%	690 ng
oligo-dT total cDNA	11154000	5525000	49.5%	920 ng

Table 5.5 Recovery of the purified cDNA species

Total cpm was measured by Cerenkov counting. The quantity of cDNA synthesised (values in Table 5.3 and Table 5.4) were used to determine the quantity of cDNA recovered.

available λgt10 cloning systems are available which minimise optimisation of reaction condition and allow easy cloning of the cDNA. An Amersham λgt10 cDNA cloning system was used to clone the cDNA.

The purified, double stranded, flush ended, radioactively labelled cDNA molecules which had been synthesised (section 5.3) were suitable for use with the cDNA cloning system. The radioactivity was used to follow the cDNA through the cloning stages and to calculate the amounts needed in the ligation reactions.

The number of independently arising plaques required to cover a complete mRNA population of P. sativum seedlings is difficult to determine. The information required is how many RNA molecules there are in the mRNA population. The abundance and complexity of the mRNA population of a typical eukaryotic cell has been estimated (Williams, 1981). The estimate is 36,790 individual mRNA molecules with about 29% of these being low abundance mRNA molecules (ie ~14 copies/cell). Using the formula of Clark and Carbon,(1976) (see section 3.5.5) when n=36,790 then for a 99% probability of obtaining a given clone 1.69×10^5 independent recombinants should be screened. It was assumed that the mRNA molecules coding for P. sativum E2/E3 and E5 would be low abundance messages and that the figure of 1.69×10^5 was the minimum number of plaques required from the library generated from total poly A^+ mRNA. Fewer recombinants were presumably required from the enriched libraries.

5.4.2 Preparation of suitable cDNA suitable for cloning

An outline of the procedures leading to preparation of a cDNA library using the Amersham cDNA cloning system is given in Fig 5.6.

Supplied with the cloning system was flush ended DNA generated from a HaeIII digest of Mp18. This was end labelled with [α^{32}P]-dGTP and the purified labelled DNA acted as a control to monitor the whole process.

Any internal EcoRI restriction sites that may be present in the cDNA were protected from subsequent digestion by EcoRI using EcoRI methylase. This methylates the A in the EcoRI recognition site (GAA*TTC) which prevents EcoRI from digesting at this site (Greene et al., 1975) therefore ensuring the full cDNA sequence could be cloned. 1 µg, 690 ng and 920 ng of random fraction-7, oligo-dT fraction-7 and oligo-dT total cDNA were methylated in this way.

Phorphorylated EcoRI linkers (GGAATTCC) were then ligated onto both ends of the cDNA in each of the three populations using T4 DNA ligase. This resulted in cDNA molecules with multiple linkers ligated to both ends. Digestion of the linkered cDNA with EcoRI produced single EcoRI 'sticky' ends on each terminus of the cDNA.

The excess linker molecules were removed by passing the complete reaction products from each linkered cDNA population through separate Sephadex G50 gel filtration columns. The total cpm in each sample was determined (before the samples were applied to the columns) and the

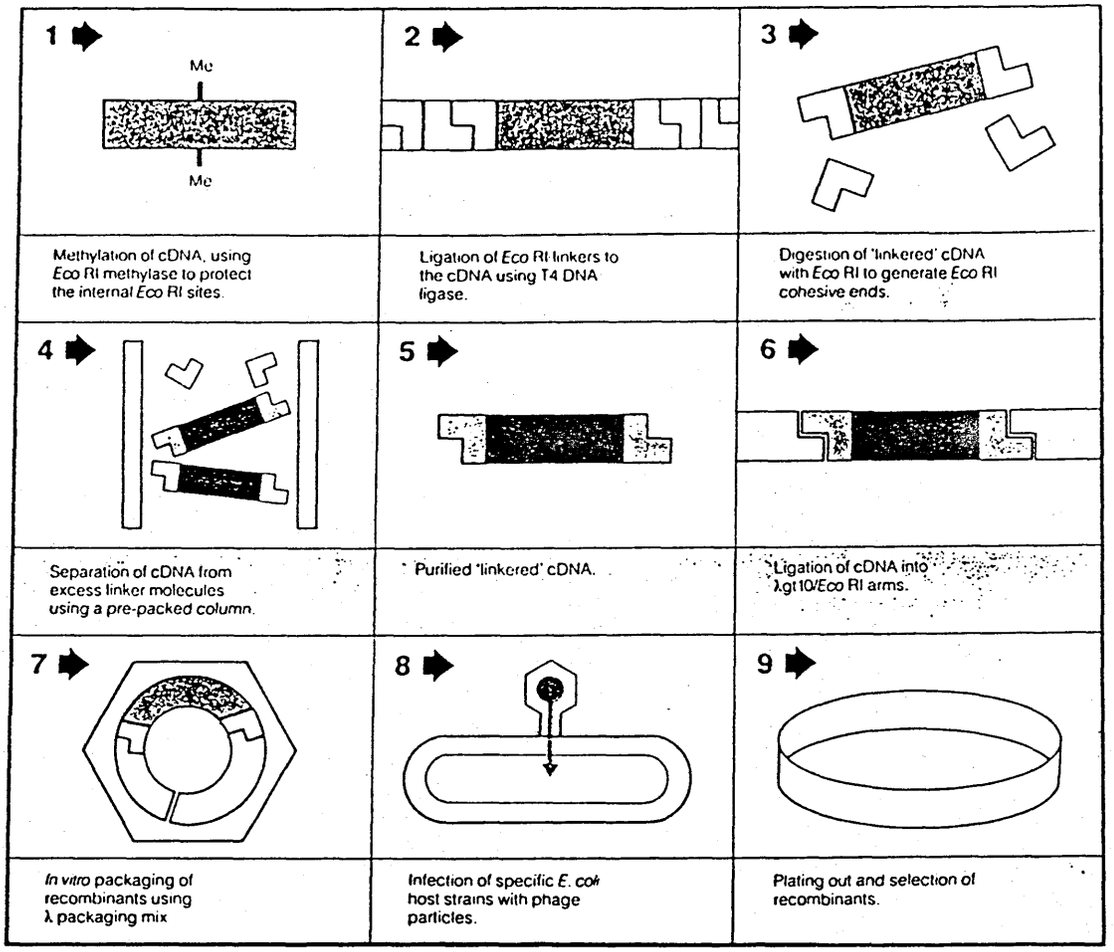


Figure 5.6 The cDNA cloning procedure

cpm in the 10 fractions obtained from each column was also determined (Table 5.6). From these results it was clear that fractions-3 and 4 in each case contained the highest cpm. These fractions were therefore pooled and the cDNA present in each was ethanol precipitated in the presence of glycogen. There was complete recovery of the labelled P. sativum cDNA after ethanol precipitiation however only 4% of the labelled control DNA was recovered (Table 5.7). It was unlikely that the fragments produced from the control DNA (fragments from a HaeIII digestion of M13 Mp18 DNA ranging in size from 69 to 2587 bp) were too small to be efficiently precipitated, but no other explanation could be found.

5.4.3 Ligation of linkered cDNA

It was important to determine the exact quantity of EcoRI linkered cDNA which had been recovered, in order to set up optimal ligation reactions. This estimation could have been done by comparing the cpm of the cDNA after passage through the column with the total cpm of cDNA initially treated with EcoRI methylase taking into account the decay of the radioactive nucleotide while carrying out the procedures. However all of the cDNA manipulations up until the passage of the linkered cDNA through the column had occurred in a single tube. If it was assumed that no DNA had been lost until passage through the column then the percentage recovery of cDNA would be equivalent to the ratio of cpm in the ethanol precipitated cDNA pellet compared to the cpm in the reaction before application to the column. The percentage recovery was calculated by this latter procedure and the quantity of

Fraction / cDNA	1	2	3	4	5	6	7	8	9	10
random fraction - 7	15	35	1304277	1431735	78579	26583	50611	42334	2221	339
oligo-dT fraction - 7	41	76	1811222	1274319	83039	22777	52471	49032	2867	442
oligo-dT total	29	140	2054434	1060014	66885	26575	93492	39671	2025	300
control DNA	28	223	1257	1183	126	44	88	74	41	30

Table 5.6 Elution profile of the cDNA populations from gel filtration columns

The cpm (measured by Cerenkov counting) for each of the 10 fractions from the four columns are tabulated. The major peak was in fractions 3 and 4 in each case. A minor peak (fraction - 7) was probably due to unincorporated radioactive nucleotide which had been carried over from cDNA synthesis.

cDNA	cpm before ethanol precipitation	cpm in cDNA pellet	% recovery
random fraction - 7 cDNA	2736012	2770266	100%
oligo-dT fraction - 7 cDNA	3085541	3204932	100%
oligo-dT total cDNA	3114448	3186239	100%
control	2440	180	5%
blank	-	71	-

Table 5.7 Recovery of cDNA after ethanol precipitation

Total cpm was determined by Cerenkov counting.

cDNA remaining was determined using these figures (Table 5.8)

It was not feasible to determine the optimal cDNA to λgt10 arms ratio using test ligations simply because there was insufficient DNA to do so. Instead three separate ligations containing 50, 100 and 150 ng of cDNA from each of the three cDNA populations mixed with 1 μg of λgt10 arms (a total of 9 ligation reactions) were necessary to attempt to encounter the optimal cloning ratio. A number of control ligations were set up to test various stages of the cloning process and to provide data necessary for the analysis of the results (section 2.26.2).

Whole λgt10 vector DNA (0.5 μg) was included to monitor the overall efficiency of the *in vitro* packaging reaction and the biological selection of the two host cell types. λgt10 EcoRI arms (1μg) alone monitored the efficiency of the ligation reaction and helped determine the selection ratio of the two cell lines and the background plating efficiency on the hfl strain of *E. coli*. λgt10 arms (1 μg) and linkered control DNA (40 ng) were also included as control DNA samples to check the performance of the whole cloning system.

The DNA was ethanol precipitated out of the ligation reactions since previous experiments (section 3.5.4) had established that inhibitory components in the ligation reaction adversely affected *in vitro* packaging. Liquid scintillation counting before and after ethanol precipitation indicated that more than 90% of the cDNA had been recovered.

cDNA	A (cpm)	B (cpm)	% recovery	Quantity of cDNA used in cloning (ng)	Quantity of cDNA recovered (ng)
random fraction - 7 cDNA	3067026	2770266	90%	1000	900
oligo-dT fraction - 7 cDNA	3419488	3204932	94%	680	640
oligo-dT total cDNA	3497363	3186239	91%	930	840
control DNA	3119	110	4%	1000	40

Table 5.8 Recovery of cDNA throughout the cloning procedure

A = Total cpm added to the cloning reactions
B = Total cpm in cDNA pellet after cloning reactions

Total cpm was determined by Cerenkov counting

5.4.4 In vitro packaging of the recombinant λ-phage

All twelve ligations were packaged using the Amersham *in vitro* packaging system and the packaged phage were stored at 4°C in 0.5 ml of SM buffer.

The titre of the 12 packaging reactions was determined on *E. coli* L87 (wild type), and *E. coli* NM514 (hfl) (Table 5.9). From these results the arms selective ratio and the background titre were calculated (section 2.26.2) as well as the total number of recombinant phage in each reaction (Table 5.10).

The titre of the packaged whole λgt10 DNA was very much less than the expected value (2×10^8 pfu/μg on L87). This suggested that the packaging reaction had not worked well. However the results of the other controls were more encouraging. They confirmed that *in vitro* packaging had been efficient. Perhaps much less λgt10 DNA had been present in the reaction than was intended and this had reduced the number of plaques obtained.

The recommended number of plaques per μg of DNA from the arms-only control on L87 was 10^7 pfu/μg. The observed titre was 2.2×10^7 which showed that efficient ligation and packaging had occurred. The arms selective ratio exceeded 100 which was what had been recommended. The control, which consisted of blunt-ended DNA ligated into λgt10, gave a high titre of recombinants which indicated that the cloning system had been successful.

ligation being packaged	ng cDNA in ligation	titre on L87 of 100 µl of following dilutions		titre on NM514 of 100 µl of following dilutions	
		10^{-4}	10^{-5}	10^{-2}	10^{-3}
whole λgt10	-	87	8	50	-
λgt10/EcoRI arms only	-	433	43	126	19
control DNA	40	76	10	1184	182
random fraction - 7 cDNA + λgt10/EcoRI arms	150	346	36	810	92
	100	435	44	878	92
	50	636	91	972	148
oligo-dT fraction - 7 cDNA + λgt10/EcoRI arms	150	280	23	1268	103
	100	196	21	1520	172
	50	340	28	1304	150
oligo-dt total cDNA + λgt10/EcoRI arms	150	147	15	378	35
	100	335	33	711	84
	50	308	32	540	67

Table 5.9 Titre of the in vitro packaged phage

The titre of 100 µl of the diluted package reactions allowed calculation of the pfu/µg of λgt10/EcoRI arms (given in Table 5.10).

package reaction		L87 titre (pfu/μg arms)	NM514 titre (pfu/μg arms)	L87 : NM514 ratio	Adjusted Background	Total recombinants	% recombinants	recombinants per μg cDNA
whole λgt10		8.7×10^6	5×10^4	174				
arms only		2.2×10^7	6.5×10^4	338				
control DNA		3.8×10^6	7.5×10^5	5	1.1×10^4	7.4×10^5	98%	1.8×10^7
random fraction - 7 cDNA	150 ng	1.8×10^7	4×10^5	45	5.3×10^4	3.5×10^5	87%	2.3×10^6
	100 ng	2.2×10^7	4.3×10^5	51	6.5×10^4	3.6×10^5	84%	3.6×10^6
	50 ng	3.2×10^7	5×10^5	64	9.5×10^4	4×10^5	80%	8×10^6
						1.1×10^6		
oligo-dT fraction - 7 cDNA	150 ng	1.4×10^7	5×10^5	28	4.1×10^4	4.6×10^5	92%	3×10^6
	100 ng	1.0×10^7	7.5×10^5	13	3×10^4	7.2×10^5	96%	7.2×10^6
	50 ng	1.5×10^7	6×10^5	25	4.4×10^4	5.6×10^5	93%	1.1×10^7
						1.7×10^6		
oligo-dT total cDNA	150 ng	0.7×10^7	1.9×10^5	37	2×10^4	1.7×10^5	89%	1.1×10^6
	100 ng	1.6×10^7	4×10^5	40	4.7×10^4	3.5×10^5	87.5%	3.5×10^6
	50 ng	1.5×10^7	3×10^5	50	4.4×10^4	2.6×10^5	86.6%	5.2×10^6
						0.8×10^6		

Table 5.10 Number of recombinant phage present in each package reaction

The total number of recombinants from the three packaging reactions involving random fraction-7 cDNA, the three reactions involving oligo-dT fraction-7 cDNA and the three involving oligo-dT total cDNA were 1.6×10^6, 1.8×10^6 and 0.8×10^6 respectively. Repetition of the titration of the package reaction produced very similar results. Therefore sufficient recombinant phage had been produced to ensure that the P. sativum E2/E3 and E5 mRNA were represented in the three cDNA libraries.

5.5 Isolation of a cDNA clone for P. sativum E5

5.5.1. Sequences used as probes

It was the intention that the three cDNA libraries prepared should be screened with an oligonucleotide that corresponded to the P. sativum E2/E3 enzyme and with the Monsanto cDNA clone cDNA of P. hybrida E5.

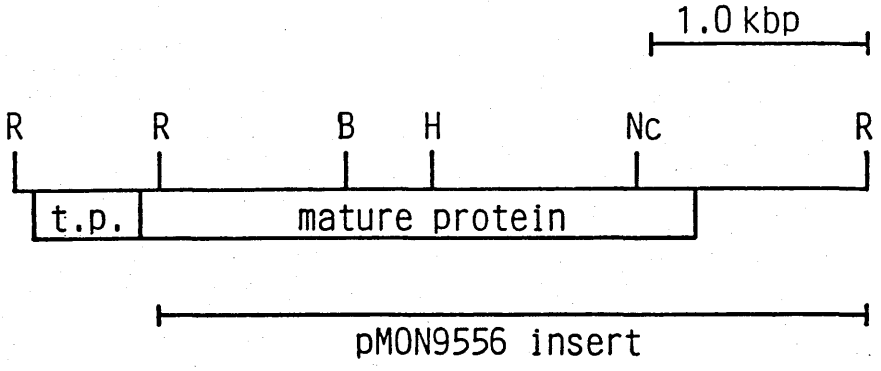
During the construction of the cDNA libraries, a large scale purification of the P. sativum bifunctional E2/E3 enzyme was carried out (D.M. Mousdale, unpublished work). The purified protein was sequenced on an Applied Biosystems gas phase sequencer but no N-terminal sequence was obtained. Complete amino acid analysis of a sample of the protein yielded an amino acid composition which confirmed that protein was present but also indicated that there was very little material (approximately 2 nmoles). The inability to sequence may have reflected insolubility of the E2/E3 enzyme in the solvents used during sequencing or alternatively the N-terminal

sequence of the enzyme may have been blocked. Coincidently, the E0 enzyme from S. tuberosum (potato) was also found to have a blocked N-terminus (Pinto et al., 1986). There was still therefore no sequence information to allow design of an oligonucleotide probe for the P. sativum E2/E3 gene. All effort was therefore concentrated on obtaining a P. sativum E5 clone.

Plasmid pMON9556 is a cDNA clone for the P. hybrida E5 enzyme (Gasser et al., 1988). A small amount of this plasmid was obtained and more was subsequently generated by carrying out a large scale plasmid preparation. A P. hybrida E5 probe was isolated from pMON9556 as a 1.65 kbp EcoRI fragment (Fig 5.7). This fragment contained 94% of the cDNA coding for the mature enzyme but the transit peptide was not encoded. In addition 400 bp of downstream sequence was also present which corresponded to the untranslated region of the E5 mRNA.

5.5.2 Northern blot analysis using the P. hybrida E5 probe

To investigate whether the transcript of the P. sativum E5 gene could be detected by the P. hybrida E5 probe, northern blot analysis was carried out. A blot containing 50 μg of total P. sativum RNA, 2 μg of P. sativum total poly A^+ mRNA and 2 μg of P. sativum poly A^+ mRNA from fractions 5, 6, 7 and 8 of the sucrose gradient (section 5.2.4) was probed with the P. hybrida E5 probe at the following low stringency conditions: aqueous hybridisation mix at 60°C, washing in 2 x SSC, 60°C. However no signal was obtained even after two months exposure to autoradiography. A control southern blot on plasmid pMON9556 gave



(Gasser et al., 1988)

Figure 5.7 Restriction map of pMON9556

Key to restriction sites: R=EcoRI, B=BamHI, H=HindIII, Nc=NcoI

The structure of the P. hybrida E5 cDNA clone is given and the region present in pMON9556, which was used as the P. hybrida E5 probe, is indicated.

a good signal and indicated that the probe was satisfactory. It appeared that the P. sativum mRNA was present at such low levels that it could not be detected directly by the P. hybrida E5 probe. It was still hoped that although undetectable under the conditions for northern analysis, the E5 mRNA was represented in the poly A^+ mRNA used to synthesise the cDNA libraries.

5.5.3 Screening the P. sativum cDNA library with the E5 probe

Since the P. hybrida cDNA clone was to be used for screening, there was no necessity to screen the P. sativum random fraction-7 library (see section 5.3.1). 6×10^6 pfu from the P. sativum oligo-dT fraction-7 cDNA library were screened at moderately high stringency with the P. hybrida E5 probe. Hybridisation was in aqueous hybridisation solution at 60°C. The stringency of hybridisation used was very similar to that used to isolate the A. thaliana E5 gene using the P. hybrida E5 probe (Klee et al., 1987). The filters were washed for short times at gradually increasing stringency until there was very little radioactivity remaining. Ultimately the filters were washed in 1 x SSC at 60°C for 30 minutes.

The six screen filters were exposed to autoradiography for four days. The developed autoradiographs exhibited two different types of spot, (Fig 5.8). There were 21 large strongly hybridising spots and 40 small strongly hybridising spots, all visible against a very low background. These spots were coincidental on duplicate filters indicating that they were not a result of spurious signals.

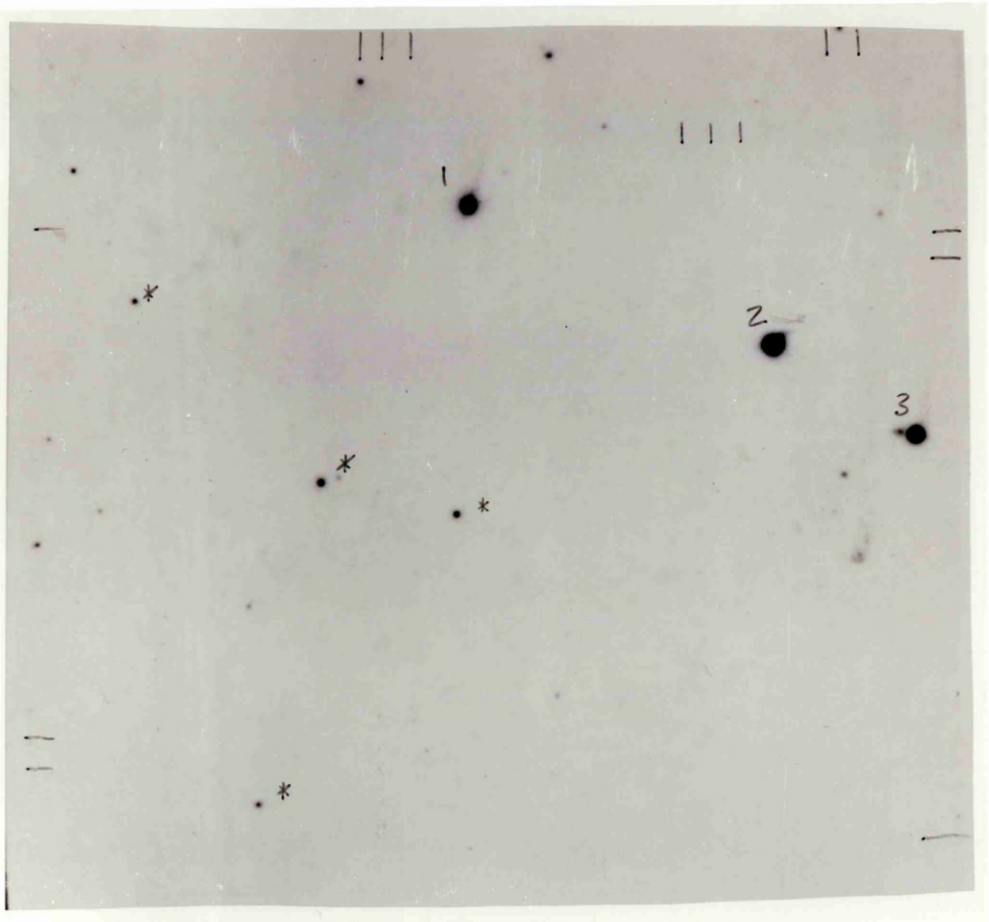


Figure 5.8 Autoradiograph of an E5 screened filter

An autoradiograph obtained after screening the enriched P. sativum cDNA library with the P. hybrida E5 cDNA probe. The large strongly hybridising plaques (numbered 1, 2 and 3) were among the plaques purified to homogeneity and further characterised. The small strongly hybridising plaques (*) were not further characterised. Other spots were spurious, i.e. they did not occur on a duplicate filter.

There were several possible explanations for the occurrence of the two types of spot. The large spots could have been due to clones which contained DNA which was more homologous to the P. hybrida E5 probe than the small spots. Alternatively the large spots could contain clones with smaller inserts than the smaller spots. These phage would be able to replicate more efficiently and therefore produce larger plaques and consequently larger spots than phage with larger inserts.

The large spots were chosen to be examined in detail.

5.5.4 Characterisation of the hybridising plaques

Eighteen of the plaques which gave the large spots on autoradiography were purified to homogeneity. DNA was isolated from each of the 18 purified phage, digested with EcoRI, and the digestion products were electrophoresed on an agarose gel (Fig 5.9a). Two bands corresponding to the λgt10 arms were present in each of the lanes containing recombinant phage DNA, along with one or two cDNA insert bands. All of the recombinant phage, except λP508, contained a similar cDNA insert band of approximately 1.4 kbp in size. In addition to this band, λP516 and λP520 contained second bands of approximately 1 kbp and 0.9 kbp in size respectively. The insert band of λP508 was considerably smaller than the other inserts with a size of approximately 0.5 kbp.

The southern blot of this gel (Fig 5.9b) was hybridised with the P. hybrida E5 probe and this confirmed that the homology to the probe lay

Figure 5.9 Analysis of putative P. sativum E5 recombinant phage

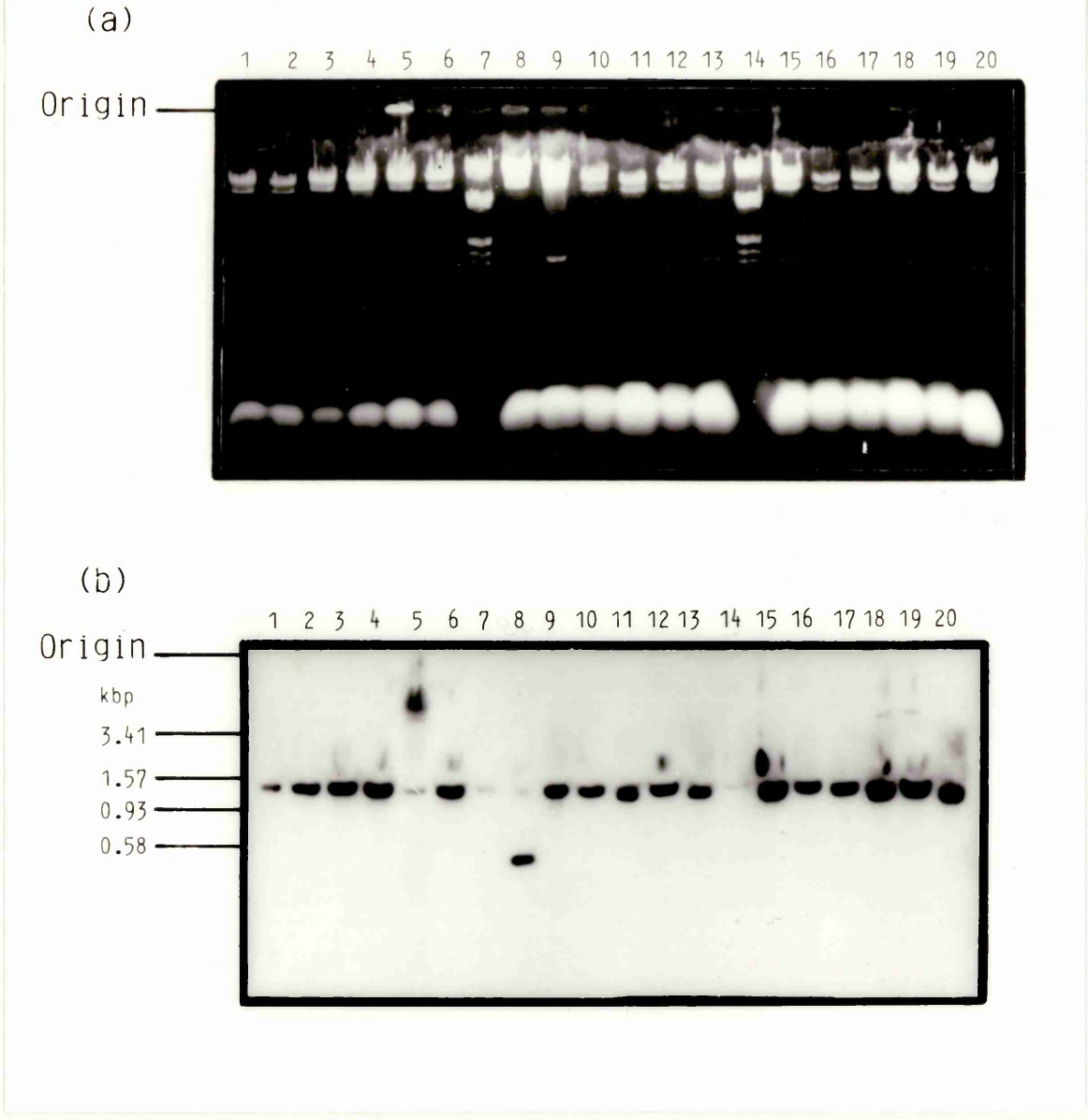
(a) Agarose gel analysis of digested phage DNA

lane:					
1.	EcoRI digested- λP501	11.	EcoRI digested- λP511
2.	" " - λP502	12.	" " - λP512
3.	" " - λP503	13.	" " - λP513
4.	" " - λP504	14.	Marker DNA
5.	" " - λP505	15.	EcoRI digested- λP514
6.	" " - λP506	16.	" " - λP515
7.	Marker DNA	17.	" " - λP516
8.	EcoRI digested- λP508	18.	" " - λP518
9.	" " - λP509	19.	" " - λP520
10.	" " - λP510	20.	" " - λP521

The samples were electrophoresed on a 0.8% agarose gel. The marker DNA was HindIII + EcoRI-digested λ-DNA with fragment sizes (kbp) 21.7, 5.24 + 5.05 (doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58, 0.14.

(b) Southern blot of gel (a)

The blot was hybridised with the P. hybrida E5 probe at 60°C in aqueous solution and washed in 1 x SSC, 60°C.


(a)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Origin
(b)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Origin
kbp
3.41
1.57
0.93
0.58

within the insert bands. The 1.4 kbp band present in 17 out of 18 of the recombinant phage hybridised with the probe as did the small 0.5 kbp insert of λP508. This strongly suggested that the 1.4 kbp fragments from each of the clones were extremely similar, if not identical and that perhaps the 0.5 kbp fragment was a truncated version of the 1.4 kbp fragment. Neither of the second insert bands present in λP516 or λP520 had any homology to the P. hybrida E5 probe.

5.5.5 Subcloning of several cDNA inserts

To confirm that the 1.4 kbp fragments from each clone were the same, several of the inserts were subcloned and examined by restriction enzyme digestion.

Samples of λP501, λP516 and λP520 DNA were digested with EcoRI and the EcoRI cDNA inserts were subcloned into EcoRI-digested, calf intestinal phosphatase-treated pUC18. The resultant plasmids pSG501, pSG5161 and pSG5201 contained the 1.4 kbp inserts of λP501, λP516 and λP520 respectively. Samples of each of these plasmids were digested with several restriction enzymes that cut within the pUC18 polylinker and the products of digestion were electrophoresed on agarose gels, (Fig 5.10). The restriction patterns of pSG501, pSG5161 and pSG5201 were identical for all the enzymes which did and did not digest the insert DNA. This was further evidence that the 1.4 kbp cDNA inserts were very similar if not identical.

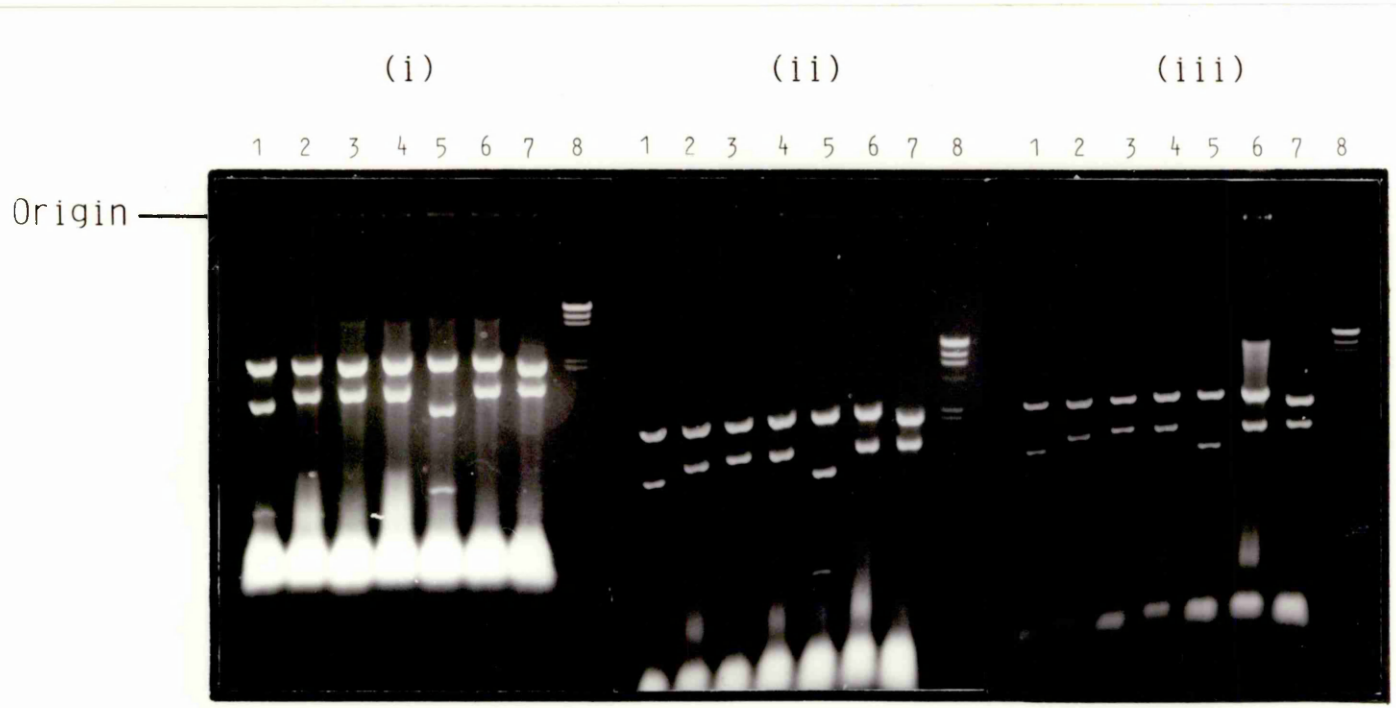


Figure 5.10 Restriction analysis of pSG501, pSG5161 and pSG5201

Plasmids pSG501 (on gel i), pSG5161 (on gel ii) and pSG5201 (on gel iii) were digested with the following enzymes and electrophoresed on 1% agarose gels. Marker DNA in lane 8 on each gel was HindIII-digested λ-DNA with sizes (in kbp) 23.6, 9.64, 6.64, 4.34, 2.26, 1.98, 0.56.

lane: 1. BamHI 2. SalI 3. HindIII 4. ClaI
5. PstI 6. BglII 7. PvuII

5.5.6. Restriction mapping of pSG5161

The insert of plasmid pSG5161 was chosen to be completely characterised by DNA sequence analysis. A restriction map of pSG5161 was required to allow subcloning and sequencing strategies to be worked out. It had been observed that a BamHI and PstI site were present within the pSG5161 insert (Fig 5.10). Double digestion of pSG5161 with BamHI and PstI, (Fig 5.11) indicated that these sites were situated at either end of the 1.4 kbp fragment. Further mapping of pSG5161 with several other restriction enzymes which cut within the pUC18 polylinker detected a KpnI site in the insert which was very close to one end of the fragment, (Fig 5.11). Restriction mapping of the 1.4 kbp insert from pSG5161, which had been isolated from an l.m.p. agarose gel, indicated that a DraI site was present towards the middle of the 1.4 kbp fragment, (Fig 5.11). Double digestion with DraI and either BamHI or PstI of the purified insert helped to map the DraI site more specifically.

The sizes of the various fragments of the pSG5161 insert in Fig 5.11 were estimated and a restriction map of the P. sativum cDNA insert present in pSG5161 was derived (Fig 5.12).

5.6 Determination of the DNA sequence of the pSG5161 cDNA insert

5.6.1 The sequencing strategy

The restriction enzyme sites which had been mapped to the pSG5161

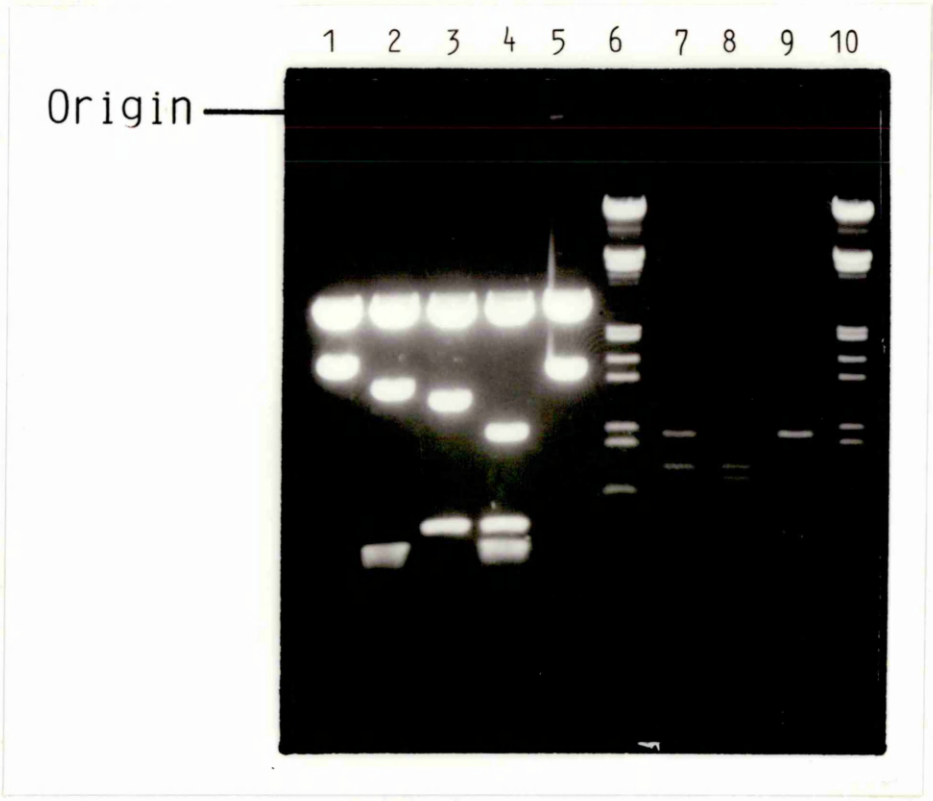


Figure 5.11 Restriction mapping of pSG5161

lane: 1. EcoRI-digested pSG5161
2. " + BamHI-digested pSG5161
3. " + PstI- " "
4. " + " + BamHI-digested pSG5161
5. KpnI digested pSG5161
6. EcoRI + HindIII-digested λ-DNA
7. DraI-digested pSG5161 insert
8. " + BamHI digested pSG5161 insert
9. " + PstI " " "
10. EcoRI + HindIII digested λ-DNA

Samples were electrophoresed on a 1% agarose gel. Size marker bands (in kbp) in lane 6 are 21.7, 5.24 + 5.05 (doublet), 4.21, 3.41, 1.98, 1.90, 1.57, 1.32, 0.93, 0.84, 0.58.

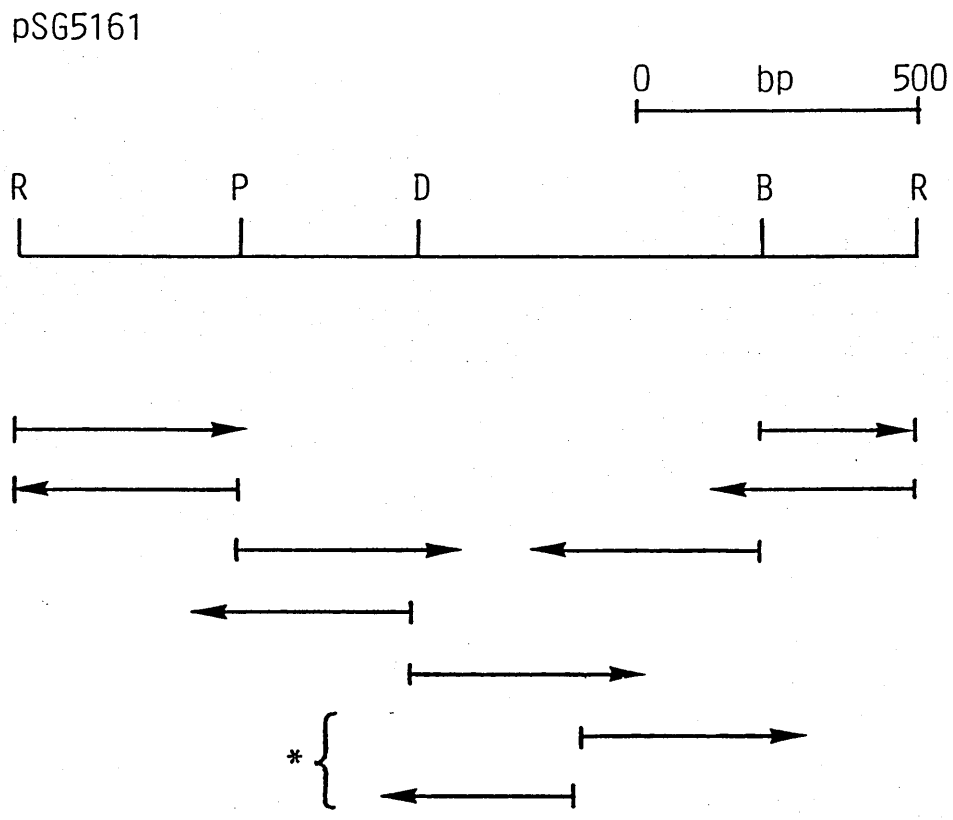


Figure 5.12 Restriction map and sequencing strategy of pSG5161

Key to restriction sites: R=EcoRI, P=PstI, D=DraI, B=BamHI

* sequenced by oligonucleotide primers

insert were very conveniently situated. Generation of subclones of the pSG5161 insert using these sites and the subsequent sequencing of the subclones was likely to provide the complete sequence of the cDNA insert. This therefore was the strategy which was adopted.

5.6.2 Sub-cloning and sequencing of the pSG5161 insert

An EcoRI digestion of pSG5161 was electrophoresed on an l.m.p. agarose gel and the complete cDNA insert was isolated. This fragment was cloned into EcoRI-digested, calf intestinal phosphatase-treated Mp18 and a selection of recombinant M13 were sequenced. This provided the DNA sequence for both ends of the fragment. The sequence overlapped the KpnI and PstI site at one end of the fragment and the BamHI site at the opposite end thus determining the orientation of these sites with respect to either end of the fragment.

The following subclones of the pSG5161 cDNA insert were then made. The 400 bp PstI-EcoRI fragment and the 275bp BamHI-EcoRI fragment were cloned into appropriately digested Mp18. A DraI digestion of the EcoRI insert from pSG5161 allowed the cloning of two fragments of the insert into SmaI+EcoRI-digested Mp18. The PstI-BamHI fragment was cloned in both orientations by ligation into PstI+BamHI-digested Mp18 and Mp19.

Sequencing of these sublones allowed the complete DNA sequence of the pSG5161 insert to be built up (Fig 5.12).

This sequencing strategy provided DNA sequence through all of the restriction enzyme sites used during the cloning. As mentioned in section 4.4.1, this was particularly important since it ensured that the DNA sequence obtained was continuous. Discontinuity of the DNA sequence could have occurred if small fragments of DNA were overlooked and not sequenced. Such fragments might have occurred if two sites for a restriction enzyme used in the sequencing strategy were in close proximity to each other and had not been observed during mapping. This problem is more likely to occur when sequencing strategies involve the shotgun cloning approach of small fragments using enzymes which cut the DNA frequently. The strategy used to sequence the pSG5161 insert involved the isolation of recognised fragments generated by enzymes which cut DNA relatively infrequently.

At this stage, although all the restriction sites had been overlapped, the DNA sequence of the pSG5161 cDNA insert had still not been determined completely on both strands. There were two regions of DNA, situated between the DraI and BamHI sites which had only been sequenced on one strand. The sequence of the middle of this region had however unambiguously been determined on both strands. This latter sequence was used to design two oligonucleotides which were in turn used to prime the DNA sequencing reactions on appropriate single stranded DNA templates. Oligonucleotide-410 (5'-ACACTCGCTGTTGTTGC-3') mapped to positions 1058-1074 (see Fig 5.13) and primed DNA sequencing towards the BamHI site. Oligonucleotide-411 (5'-ATCTTGTTCATATTGAC-3') was complementary to the DNA at position 1041-1025 (see Fig 5.13) and primed DNA synthesis towards the DraI site.

```
     EcoRI                                                    |mature
   1 GAATTCCTTT GTGATGAAGG CTTCTGCTTC GGTTGCCGCC GCGGAGAAGC
  51 CGTCGACGGC GCCGGAGATT GTGCTGGAAC CGATTAGTGA AATTTCTGGT
 101 ACCATTACTT TGCCTGGATC AAAGTCTCTA TCCAATCGGA TTCTGCTTCT
 151 AGCTGCTCTC TCTGAGGGAA CAACTGTTGT AGAGAACTTG TTAGATAGTG
 201 AGGATATTCA TTACATGCTC GAGGCATTGA AGACCCTTGG ACTTCGAGTG
 251 GAAGACGACA AAACAACCAA ACAAGCAGTT GTGGAAGGTA GTGGGGGATT
 301 GTTTCCCACC GGTAGAGAAT CTAAAGATGA AGTTAATTTA TTCCTTGGAA
 351 ATGCCGGTAC GGCAATGCGT CCTTTGACGG CAGCCTTGGT TGCTGCAGGT
 401 GGAAATACAA GATACATACT TGATGGAGTT CCCCGAATGA GAGAAAGACC
 451 AATTGGAGAT TTGGTTTCCG GTCTCAAGCA ACTTGGTGCT GATGTTGATT
 501 GTTTTCTTGG CACAAATTGT CCACCTGTTC GTATAATTGG GAAGGGAGGA
 551 CTTCCAGGGG GAAAGGTGAA ACTGTCCGGA TCTATTAGCA GTCAGTACCT
 601 AACTGCTTTG CTTATGGCAG CACCGTTGGC TCTTGGCGAC GTTGAGATTG
 651 AGATTATTGA TAAGCTGATT TCTGTTCCGT ATGTTGAAAT GACTTTAAAG
 701 TTGATGGAGC GCTTTGGAGT CTCTGTCGAA CACAGTGATA ATTGGGATAG
 751 ATTTTTGGTC CACGGAGGTC AAAAGTACAA GTCTCCCGGA AATGCTTTCG
 801 TTGAAGGTGA TGCTTCTAGC GCCAGTTACT TCCTAGCCGG TGCAGCAGTT
 851 ACTGGTGGAA CTATCACGGT TATAGGCTGC GGGACAAGCA GTTTACAGGG
 901 AGATGTAAAA TTTGCTGAAG TTCTTGAGAA GATGGGGGCT AAAGTTACGT
 951 GGACAGAAAA CAGCGTCACA GTTACCGGGC CTCCACGAGA TTCTTCTGGT
1001 CGGAAAGTGT TGCAAGGCAT TGATGTCAAT ATGAACAAGA TGCCAGATGT
1051 TGCCATGACA CTCGCTGTTG TTGCGCTATT TGCTAATGGT CCCACTGCTA
1101 TTAGAGACGT GGCAAGTTGG AGAGTTAAAG AGACAGAAAG AATGATAGCA
1151 ATCTGCACAG AACTCAGAAA GCTAGGAGCA ACAGTTGAAG AAGGTCCCGA
1201 TTACTGCGTG ATAACACCAC CTGAGAAACT GAATGTCACG TCAATAGACA
1251 CATACGACGA TCATAGAATG GCTATGGCAT TTTCTCTTGC TGCTTGCGGC
1301 GATGTTCCGG TCACAATCAA GGATCCTGGT TGCACAAGGA AAACCTTCCC
1351 TGACTACTTT CAAGTCCTTG AGAGGTTCAC AAAGCATTAA GCATTCTTGT
1401 ACATTATTAT GTAGGCAAGA GAGGTAAAAT GCTCACACAA AGTTGTGAGT
1451 ATTTTTGATT GAGATGATTT TGTTTCTGAA TGTTCCCTTG TAAGGAGGAT
1501 TTGAGATGAT TTGTATTGTT GTTGAAATGA GGTAGTGGTT TTATTGTATG
1551 AATGCAACTT AATTATCAAT AATGACTTAG GGTCTGATTG GAAGGAATTC
                                                        EcoRI
```

Figure 5.13 The DNA sequence of the pSG5161 insert

Positions corresponding to oligonucleotides-410 and -411 are underlined

The sequence data obtained using these oligonucleotides as sequencing primers completed the sequencing of the P. sativum cDNA insert present in pSG5161 on both strands. The sequence of this cDNA insert is given in Fig 5.13.

5.7 Analysis of the cDNA sequence of the pSG5161 insert

5.7.1 Characterisation of the cDNA insert of pSG5161

The sequence of the P. sativum cDNA insert of pSG5161 was translated in all six reading frames. There was a large ORF of 462 amino acids, which extended from the EcoRI site (closest to the PstI site in Fig 5.12) almost to the opposite end of the fragment (Fig 5.14). A BESTFIT comparison (modified to look for identities rather than similarities) of the amino acid sequence of this ORF with the amino acid sequence of P. hybrida E5 showed up the extensive homology between the two sequences (Fig 5.15). This suggested that the DNA coding for the large ORF was the coding cDNA for P. sativum E5.

The homology between the two sequences extended into transit peptide sequence of the P. hybrida E5 enzyme which strongly suggested that the P. sativum E5 contains a transit peptide. The sequence was however disrupted near the junction of the P. hybrida transit peptide and the beginning of the mature P. hybrida E5 enzyme. An insertion of three amino acid was present in the P. sativum sequence at this point (Fig 5.15). Elsewhere, no gaps had to be introduced into either sequence for maximum alignment to occur between them.

```
991
TTCTTCTGGTCGGAAAGTGTTGCAAGGCATTGATGTCAATATGAACAAGATGCCAGATGTTGCCATGACACTCGCTGTTGTTGCGCTATTTGCTAATGGTCCCACTGCTA
 SerSerGlyArgLysValLeuGlnGlyIleAspValAsnMetAsnLysMetProAspValAlaMetThrLeuAlaValValAlaLeuPheAlaAsnGlyProThrAlaIle

1101
TTAGAGACGTGGCAAGTTGGAGAGTTAAAGAGACAGAAAGAATGATAGCAATCTGCACAGAACTCAGAAAGCTAGGAGCAACAGTTGAAGAAGGTCCCGATTACTGCGTG
  ArgAspValAlaSerTrpArgValLysGluThrGluArgMetIleAlaIleCysThrGluLeuArgLysLeuGlyAlaThrValGluGluGlyProAspTyrCysVal
1211
ATAACACCACCTGAGAAACTGAATGTCACGTCAATAGACACATACGACGATCATAGAATGGCTATGGCATTTTCTCTTGCTGCTTGCGGCGATGTTCCGGTCACAATCAA
IleThrProProGluLysLeuAsnValThrSerIleAspThrTyrAspAspHisArgMetAlaMetAlaPheSerLeuAlaAlaCysGlyAspValProValThrIleLys
1321
GGATCCTGGTTGCACAAGGAAAACCTTCCCTGACTACTTTCAAGTCCTTGAGAGGTTCACAAAGCATTAAGCATTCTTGTACATTATTATGTAGGCAAGAGAGGTAAAAT
 AspProGlyCysThrArgLysThrPheProAspTyrPheGlnValLeuGluArgPheThrLysHis***
1431
GCTCACACAAAGTTGTGAGTATTTTTGATTGAGATGATTTTGTTTCTGAATGTTCCCTTGTAAGGAGGATTTGAGATGATTTGTATTGTTGTTGAAATGAGGTAGTGGTT
1541                                                    1600
TTATTGTATGAATGCAACTTAATTATCAATAATGACTTAGGGTCTGATTGGAAGGAATTC
```

Figure 5.14 Translation of pSG5161 showing the largest ORF

1

GAATTCCTTTGTGATGAAGGCTTCTGCTTCGGTTGCCGCCGCGGAGAAGCCGTCGACGGCGCCGGAGATTGTGCTGGAACCGATTAGTGAAATTTCTGGTACCATTACTT

AsnSerPheValMetLysAlaSerAlaSerValAlaAlaAlaGluLysProSerThrAlaProGluIleValLeuGluProIleSerGluIleSerGlyThrIleThrLeu

111

TGCCTGGATCAAAGTCTCTATCCAATCGGATTCTGCTTCTAGCTGCTCTCTCTGAGGGAACAACTGTTGTAGAGAACTTGTTAGATAGTGAGGATATTCATTACATGCTC

ProGlySerLysSerLeuSerAsnArgIleLeuLeuLeuAlaAlaLeuSerGluGlyThrThrValValGluAsnLeuLeuAspSerGluAspIleHisTyrMetLeu

221

GAGGCATTGAAGACCCTTGGACTTCGAGTGGAAGACGACAAAACAACCAAACAAGCAGTTGTGGAAGGTAGTGGGGGATTGTTTCCCACCGGTAGAGAATCTAAAGATGA

GluAlaLeuLysThrLeuGlyLeuArgValGluAspAspLysThrThrLysGlnAlaValValGluGlySerGlyGlyLeuPheProThrGlyArgGluSerLysAspGlu

331

AGTTAATTTATTCCTTGGAAATGCCGTACGGCAATGCGTCCTTTGACGGCAGCCTTGGTTGCTGCAGGTGGAAATACAAGATACATACTTGATGGAGTTCCCCGAATGA

ValAsnLeuPheLeuGlyAsnAlaGlyThrAlaMetArgProLeuThrAlaAlaLeuValAlaAlaGlyGlyAsnThrArgTyrIleLeuAspGlyValProArgMetArg

441

GAGAAAGACCAATTGGAGATTTGGTTTCCGGTCTCAAGCAACTTGGTGCTGATGTTGATTGTTTTCTTGGCACAAATTGTCCACCTGTTCGTATAATTGGGAAGGGAGGA

GluArgProIleGlyAspLeuValSerGlyLeuLysGlnLeuGlyAlaAspValAspCysPheLeuGlyThrAsnCysProProValArgIleIleGlyLysGlyGly

551

CTTCCAGGGGGAAAGGTGAAACTGTCCGGATCTATTAGCAGTCAGTACCTAACTGCTTTGCTTATGGCAGCACCGTTGGCTCTTGGCGACGTTGAGATTGAGATTATTGA

LeuProGlyGlyLysValLysLeuSerGlySerIleSerSerGlnTyrLeuThrAlaLeuLeuMetAlaAlaProLeuAlaLeuGlyAspValGluIleGluIleIleAsp

661

TAAGCTGATTTCTGTTCCGTATGTTGAAATGACTTTAAAGTTGATGGAGCGCTTTGGAGTCTCTGTCGAACACAGTGATAATTGGGATAGATTTTTGGTCCACGGAGGTC

LysLeuIleSerValProTyrValGluMetThrLeuLysLeuMetGluArgPheGlyValSerValGluHisSerAspAsnTrpAspArgPheLeuValHisGlyGlyGln

771

AAAAGTACAAGTCTCCCGGAAATGCTTTCGTTGAAGGTGATGCTTCTAGCGCCAGTTACTTCCTAGCCGGTGCAGCAGTTACTGGTGGAACTATCACGGTTATAGGCTGC

LysTyrLysSerProGlyAsnAlaPheValGluGlyAspAlaSerSerAlaSerTyrPheLeuAlaGlyAlaAlaValThrGlyGlyThrIleThrValIleGlyCys

881

GGGACAAGCAGTTTACAGGGAGATGTAAAATTTGCTGAAGTTCTTGAGAAGATGGGGGCTAAAGTTACGTGGACAGAAAACAGCGTCACAGTTACCGGGCCTCCACGAGA

GlyThrSerSerLeuGlnGlyAspValLysPheAlaGluValLeuGluLysMetGlyAlaLysValThrTrpThrGluAsnSerValThrValThrGlyProProArgAsp

Figure 5.15 Homology of the putative P. sativum E5 and the P. hybrida E5 amino acid sequences

The largest ORF of pSG5161 was compared with the P. hybrida E5 amino acid sequence using a BESTFIT comparison modified to look for identities rather than similarities.

```
          Imature    .         .         .         .         .         .         .         .
   8 SASVAAAEKPSTAPEIVLEPISEISGTITLPGSKSLSNRILLLAALSEGTTVVENLLDSEDIHYMLEALKTLGLRVEDDKTTKQAVVEGSGGLFPTGRES
     ||||| | |||   |||| || |||||  |||||||||||||||||||||||| ||| | |||||| ||||||| || |    ||||| ||||| | ||
  65 SASVATAQKPS...EIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLSSDDIHYMLGALKTLGLHVEEDSANQRAVVEGCGGLFPVGKES
             .         .         .         .         .         .         .         .         .
 108 KDEVNLFLGNAGTAMRPLTAALVAAGGNTRYILDGVPRMRERPIGDLVSGLKQLGADVDCFLGTNCPPVRIIGKGGLPGGKVKLSGSISSQYLTALLMAA
     | |  ||||||||||||||||   |||| || ||||||||||| ||| ||||||| ||||||| ||||||  |||||||||||||||||||||||||
 162 KEEIQLFLGNAGTAMRPLTAAVTVAGGNSRYVLDGVPRMRERPISDLVDGLKQLGAEVDCFLGTKCPPVRIVSKGGLPGGKVKLSGSISSQYLTALLMAA
             .         .         .         .         .         .         .         .         .
 208 PLALGDVEIEIIDKLISVPYVEMTLKLMERFGVSVEHSDNWDRFLVHGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGTITVIGCGTSSLQGDVKFAEV
     |||||||||||||||||||||||||||||||| |||||  |||| | |||||||| ||||||||||||||||||||||| |||| |||| ||||||
 262 PLALGDVEIEIIDKLISVPYVEMTLKLMERFGISVEHSSSWDRFFVRGGQKYKSPGKAFVEGDASSASYFLAGAAVTGGTITVEGCGTNSLQGEVKFAEV
             .         .         .         .         .         .         .         .         .
 308 LEKMGAKVTWTENSVTVTGPPRDSSGRKVLQGIDVNMNKMPDVAMTLAVVALFANGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCVITPP
     ||||||  |||||||||  ||||  |||||  |  |||||||||||||||||  | |||||||||||||||||||||||||||||||||||||||||||||  ||||
 362 LEKMGAEVTWTENSVTVKGPPRSSSGRKHLRAIDVNMNKMPDVAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCIITPP
             .         .         .         .         .
 408 EKLNVTSIDTYDDHRMAMAFSLAACGDVPVTIKDPGCTRKTFPDYFQVLERFTKH 462
     ||||||  |||||||||||||||||  ||||||  |||||||||  ||  ||    ||
 462 EKLNVTDIDTYDDHRMAMAFSLAACADVPVTINDPGCTRKTFPNYFDVLQQYSKH 516
```

Despite the insertion, it seemed reasonable to assign Lys-16 (numbered with respect to the first amino acid of the large ORF in pSG5161) as the N-terminal residue of the mature P. sativum E5 since this would preserve the sequence around the transit peptide cleavage site.

The mature P. sativum E5, consisting of the 447 amino acids from lys-16 to the C-terminal histidine (Fig 5.15), had a calculated Mr of 48,716. This value was consistant with the Mr of 50,000 which had been estimated from SDS-PAGE and gel filtration chromatography of purified E5 from P. sativum (Mousdale and Coggins, 1984).

From these two pieces of evidence, the homology of the sequences and the calculated size of the mature polypeptide, it was concluded that the DNA coding for the larger ORF in pSG5161 was the coding cDNA for P. sativum E5.

It was unlikely that all of the transit peptide of the P. sativum E5 was encoded by the insert of pSG5161 since there were only 15 amino acids present upstream of the presumed N-terminal residue of the P. sativum mature E5 sequence. The three previously characterised plant E5 enzymes have transit peptides of 72-76 amino acids therefore it was not unreasonable to expect a similarly sized transit peptide as part of the P. sativum E5. Also, there was no sequence near the beginning of the coding cDNA which closely matched the consensus sequence (AACAATGG) for the translation start site of dicotyledenous plants (Scioli and Zilinskas, 1988). This implied that coding cDNA was missing from upstream of the EcoRI site which interupted the DNA

coding for the transit peptide. Apparently therefore, this EcoRI site was a naturally occurring site which had not been artificially derived during the cDNA cloning procedure.

The absence of the upstream cDNA in the pSG5161 insert meant that (i) the P. sativum E5 transit peptide could not be completely characterised (ii) the translational start site of the P. sativum precursor E5 could not be determined and (iii) there was no information as to possible control sequences in the 5' untranslated region of the P. sativum E5 mRNA since this sequence would have been present upstream of the EcoRI site.

There was no cDNA sequence, downstream of the stop codon of the P. sativum E5 cDNA, which corresponded to the Poly A tail of the E5 mRNA (Fig 5.14). A run of A residues was certainly expected since the cDNA used to generate the library, from which the pSG5161 insert was isolated, had been synthesised using oligo-dT as a first strand cDNA synthesis primer. First strand priming would only have been successful on poly A^+ mRNA. This suggested that there was a second naturally occurring EcoRI site present within the full length cDNA. Its position must be between the 3'-end of the coding cDNA and the poly A tail.

Inefficiency of the EcoRI methylase during cDNA cloning is the most likely explanation for the absence of both upstream and downstream cDNA. If methylation of the cDNA by the EcoRI methylase had failed then the two internal, naturally occurring EcoRI sites would not have

been protected from digestion with EcoRI. On digestion of the cDNA with this restriction enzyme, cleavage of the full length cDNA molecules at the two internal EcoRI sites would have occurred yielding 1.6 kbp EcoRI fragments. This would explain why all but one of the cDNA fragments which hybridised with the P. hybrida E5 probe were this size (Fig 5.9). A proposed structure of the deduced full length cDNA clone of P. sativum E5 is given in Fig 5.16.

The only fragments of cDNA from the recombinant phage analysed which could conceivably have been the missing upstream or downstream DNA were the second EcoRI inserts present within the two phage clones λP516 and λP520 (Fig 5.9). These fragments did not hybridise to the P. hybrida E5 probe which suggested that they did not contain coding DNA. Although not unprecedented (Kingsman and Kingsman, 1987) untranslated regions which are the size of these fragments, approximately 1 kbp, are unusual and so it seemed unlikely that these clones or any others isolated from the P. sativum cDNA library contained the missing 5' or 3'-cDNA.

5.7.2 Homology of the P. sativum E5 sequence with other E5 sequences

The mature P. sativum and P. hybrida enzymes are very similar (Fig 5.15). Not including the three amino acid insertion present in the P. sativum sequence, they differ in only 59 positions along their sequence which represents an 87% identity. It is not therefore surprising that the P. hybrida E5 probe was able to hybridise with the P. sativum cDNA clone under the relatively stringent conditions used

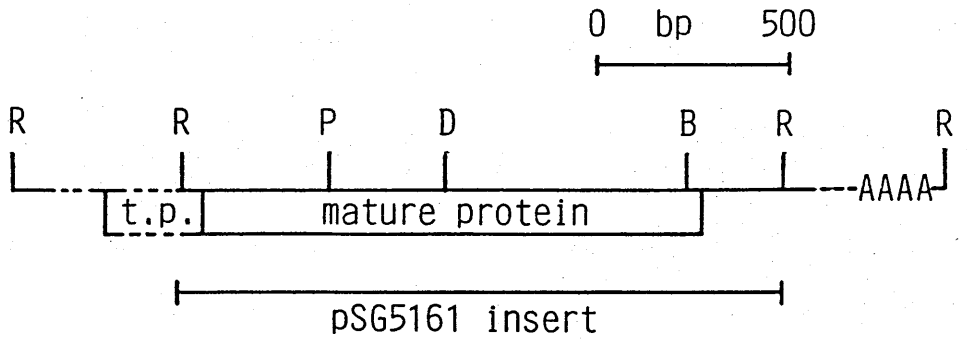


Figure 5.16 The proposed structure of a full length cDNA clone of P. sativum E5

The region of the P. sativum cDNA present in pSG5161 is indicated. The broken lines in the DNA and the box containing transit peptide (t.p.) sequence are of unknown length.

Key to restriction sites: R=EcoRI, P=PstI, D=DraI, B=BamHI

AAAA=polyA tail

during screening (see section 5.5.3).

There is extensive homology between all four known plant E5 amino acid sequences (Fig 5.17). The three previously characterised plant enzymes are between 84 to 93% homologous to one another with the P. hybrida and L. esculentum enzymes being the most homologous since they are both members of the family Solanaceae (Klee et al., 1987; Gasser et al., 1988) Based on the alignment given in Fig 5.17, the P. sativum and L. esculentum enzymes are 86% identical and the P. sativum and A. thaliana enzymes are 83% identical. These homologies are therefore within the same range as has been observed for the other plant E5 enzymes.

In view of the absence of gaps in the alignment of the P. hybrida, L. esculentum and A. thaliana amino acid sequences, it was unexpected when the P. sativum sequence broke this trend by having a three amino acid insertion (i.e. -T-A-P-) close to its N-terminus. However the N-terminal sequence of purified E5 from Corydalis sempervirens has been determined (see section 1.10.1; Hollander-Cytko et al., 1988). Alignment of this sequence with the other plant E5 sequences (Fig 5.17) suggests that C. sempervirens E5 also has this three amino acid insertion. This allayed fears that the DNA sequence of the P. sativum E5 cDNA clone in this region had been incorrectly determined. It is interesting that the C. sempervirens E5 appears to have two additional amino acids at its immediate N-terminus compared to the other plant E5 sequences.

Figure 5.17 Alignment of the mature plant E5 amino acid sequences

The N-terminal sequence of the C. sempervirens mature E5 is aligned with the four mature E5 amino acid sequences.

```
Corydalis TEKPSTAPEIL LQPIKEISGT RKLRG                                                                    99
      Pea   KPSTAPEIV LEPISEISGT ITLPGSKSLS NRILLLAALS EGTTVVENLL DSEDIHYMLE ALKTLGLRVE DDKTTKQAVV EGSGGLFPTG RESKDEVNLF
   Petunia   KPS...EIV LQPIKEISGT VKLPGSKSLS NRILLLAALS EGTTVVDNLL SSDDIHYMLG ALKTLGLHVE EDSANQRAVV EGCGGLFPVG KESKEEIQLF
   Tomato    KPH...EIV LXPIKDISGT VKLPGSKSLS NRILLLAALS EGRTVVDNLL SSDDIHYMLG ALKTLGLHVE DDNENQRAIV EGCGGQFPVG KKSEEEIQLF
Arabidopsis  KAS...EIV LQPIREISGL IKLPGSKSLS NRILLLAALS EGTTVVDNLL NSDDINYMLD ALKRLGLNVE TDSENNRAVV EGCGGIFPAS IDSKSDIELY
                                                                                                         199
      Pea  LGNAGTAMRP LTAALVAAGG NTRYILDGVP RMRERPIGDL VSGLKQLGAD VDCFLGTNCP PVRIIGKGGL PGGKVKLSGS ISSQYLTALL MAAPLALGDV
   Petunia  LGNAGTAMRP LTAAVTVAGG NSRYVLDGVP RMRERPISDL VDGLKQLGAE VDCFLGTKCP PVRIVSKGGL PGGKVKLSGS ISSQYLTALL MAAPLALGDV
   Tomatoe  LGNAGTAMRP LTAAVTVAGG HSRYVLDGVP RMRERPIGDL VDGLKQLGAE VDCSLGTNCP PVRIVSKGGL PGGKVKLSGS ISSQYLTALL MAAPLALGDV
Arabidopsis  LGNAGTAMRP LTAAVTAAGG NARYVLDGVP RMRERPIGDL VVGLKQLGAD VECTLGTNCP PVRVNANGGL PGGKVKLSGS ISSQYLTALL MSAPLALGDV
                                                                                                         299
      Pea  EIEIIDKLIS VPYVEMTLKL MERFGVSVEH SDNWDRFLVH GGQKYKSPGN AFVEGDASSA SYFLAGAAVT GGTITVIGCG TSSLQGDVKF AEVLEKMGAK
   Petunia  EIEIIDKLIS VPYVEMTLKL MERFGISVEH SSSWDRFFVR GGQKYKSPGK AFVEGDASSA SYFLAGAAVT GGTITVEGCG TNSLQGDVKF AEVLEKMGAE
   Tomatoe  EIEIIDKLIS VPYVEMTLKL MERFGVFVEH SSGWDRFLVK GGQKYKSPGK AFVEGDASSA SYFLAGAAVT GGTVTVEGCG TSSLQGDVKF AEVLEKMGAE
Arabidopsis  EIEIVDKLIS VPYVEMTLKL MERFGVSVEH SDSWDRFFVK GGQKYKSPGN AYVEGDASSA SYFLAGAAIT GETVTVEGCG TTSLQGDVKF AEVLEKMGCK
                                                                                                         399
      Pea  VTWTENSVTV TGPPRDSSGR KVLQGIDVNM NKMPDVAMTL AVVALFANGP TAIRDVASWR VKETERMIAI CTELRKLGAT VEEGPDYCVI TPPEKLNVTS
   Petunia  VTWTENSVTV KGPPRSSSGR KHLRAIDVNM NKMPDVAMTL AVVALYADGP TAIRDVASWR VKETERMIAI CTELRKLGAT VEEGPDYCII TPPEKLNVTD
   Tomatoe  VTWTENSVTV KGPPRNSSGM KHLRAIDVNM NKMPDVAMTL AVVALFADGP TTIRDVATWR VKETERMIAI CTELRKLGAT VVEGSDYCII TPPEKLNVTE
Arabidopsis  VSWTENSVTV TGPPRDAFGM RHLRAIDVNM NKMPDVAMTL AVVALFADGP TTIRDVASWR VKETERMIAI CTELRKLGAT VEEGSDYCVI TPPKKVKTAE
                                                     447
      Pea  IDTYDDHRMA MAFSLAACGD VPVTIKDPGC TRKTFPDYFQ VLERFTKH
   Petunia  IDTYDDHRMA MAFSLAACAD VPVTINDPGC TRKTFPNYFD VLQQYSKH
   Tomatoe  IDTYDDHRMA MAFSLAACAD VPVTIKNPGC TRKTFPDYFE VLQKYSKH
Arabidopsis  IDTYDDHRMA MAFSLAACAD VPITINDSGC TRKTFPDYFQ VLERITKH
```

Overall 75% of the amino acids present in the complete sequence of each of the four plant mature enzymes are identical. Some of the non-identical amino acids exhibit an obvious conservation of chemical character for example the acidic residues D/E (at position 15, 56, 59, 95, 149, 151); the small aliphatic residues I/V/L (at postions 20, 78, 96, 114, 124, 204, 225, 268, 273, 388, 394, 422); the hydroxyl containing residues T/S (at positions 301, 357, 445); the aromatic residues F/Y (at positions 251, 345); the basic residues R/K (at position 320). Position numbers are given with respect to the P. sativum sequence. More subtle similarities perhaps involving the allowable size of the amino acid side chains which could be accommodated within the secondary structure of the enzyme (Risler et al., 1988) with less regard to chemical nature are perhaps among the other non-identical residues. Identification of the latter conservations will become possible when a crystal structure of a plant E5 becomes available.

The transit peptides of the three previously characterised plant E5 enzymes are quite diverged with the L. esculentum having four insertions relative to the P. hybrida sequence (Klee et al., 1987; (Gasser et al., 1988). However the amino acid sequence close to the site of transit peptide cleavage is highly conserved between the known E5 sequences from the three plant species. The small amount of amino acid sequence deduced from the pSG5161 DNA sequence, which was homologous to the P. hybrida E5 transit peptide sequence exhibited considerable homology to the other transit peptide sequences. Ten of the 15 amino acids adjacent to the proposed cleavage site, were

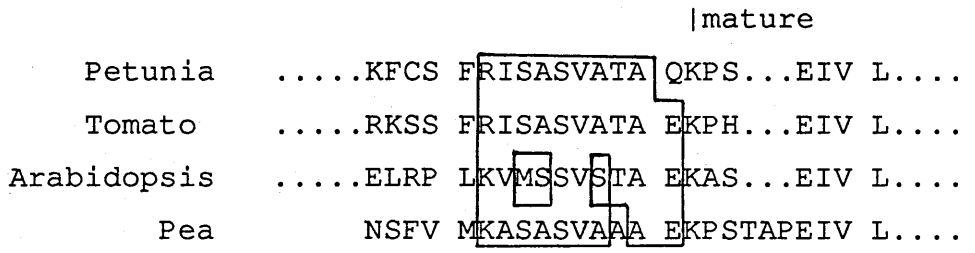


Figure 5.18 Alignment of part of the transit peptide sequences of the four plant E5 enzymes

The 15 amino acids, of the petunia, tomatoe and Arabidopsis transit peptides, which are closest to the transit peptide cleavage site are shown.

homologous in all four sequences (Fig 5.18). It remains to be seen whether the complete P. sativum transit peptide sequence is of similar length to the others and whether the homology pattern remains similar along the length of the sequence.

A comparison of the three previously characterised plant E5 enzymes with the available sequences of those from other organisms has been carried out (Kishore and Shah, 1988) and this revealed higher homology to the bacterial enzymes than to the fungal enzyme. This situation was also observed when the P. sativum mature enzyme was compared with the bacterial enzymes (Fig 5.19) and the fungal enzymes (Fig 5.20). P. sativum E5 had 52% homology with the E. coli E5 and 51% homology with the S. typhimurium E5 whereas it had only 34% homology with the S. cerevisiae E5 and 36% homology with the A. nidulans E5.

Table 5.11 summarises the number of positional identities and percentage homologies (identicalities) for pairwise combinations of the mature P. sativum E5 enzyme and all the other plant, fungal and bacterial sequences available. The homology values are based on the alignment derived by Kishore and Shah, 1988 since it was felt that the P. sativum sequence data fitted well with their scheme. Their alignment is an intuitive adjustment of BESTFIT alignments between pairwise combinations of all the E5 enzymes in an attempt to obtain the optimal alignment.

A comparison of all of the available E5 sequences is given in Fig 5.21. There are 116 residues which are identical in all of the E5

Figure 5.19 Alignment of the P. sativum mature E5 with the known bacterial E5 amino acid sequences

The alignment is based on that of Kishore and Shah, (1988). Identical and chemically similar amino acids are boxed.

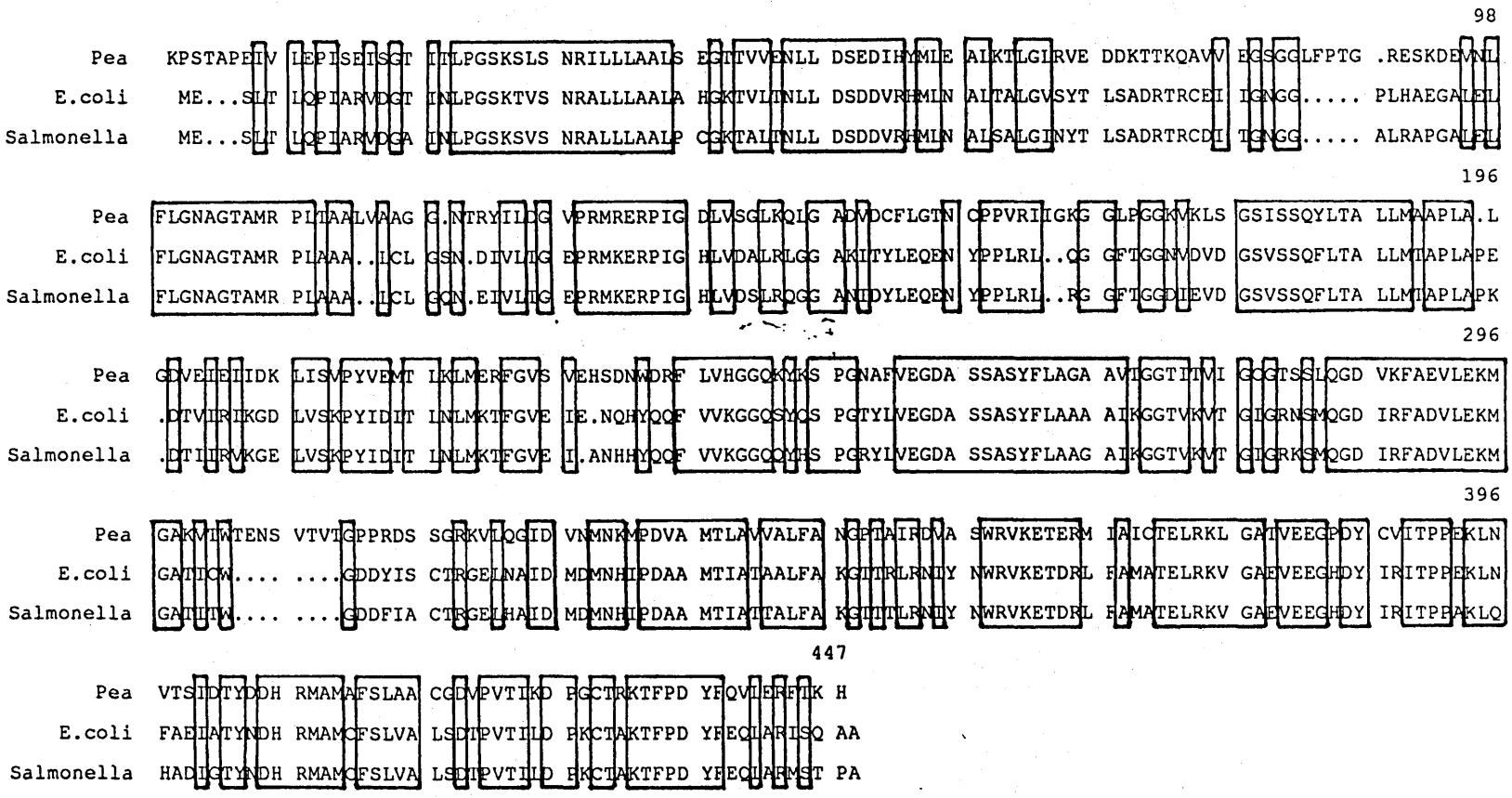
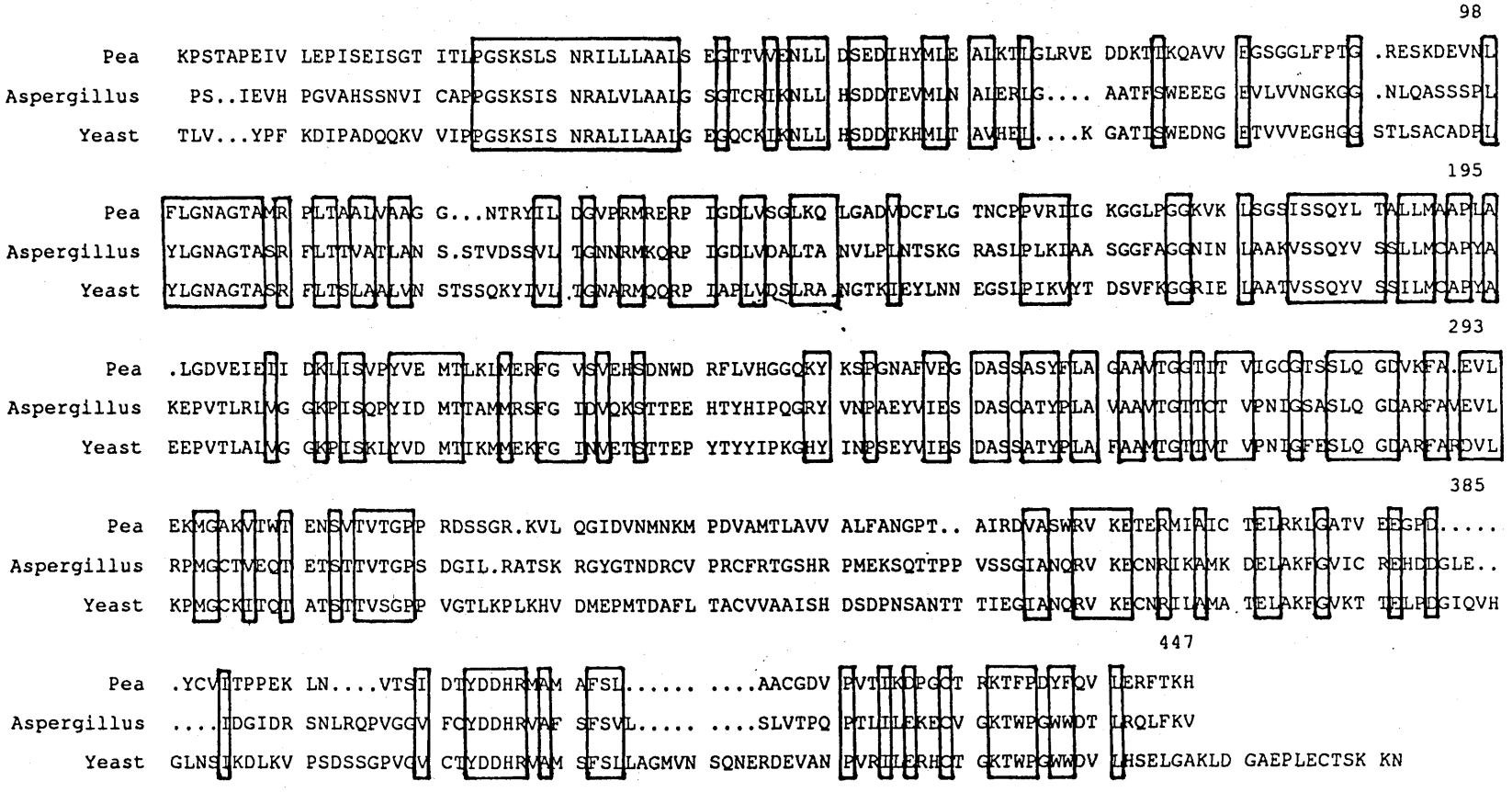


Figure 5.20 Alignment of the mature P. sativum E5 with the known fungal E5 amino acid sequences.

The alignment is based on that of Kishore and Shah (1988). Identical and chemically similar amino acids are boxed.



Organism	Length of E5 polypeptide (aa)	No. of conserved residues between P. sativum and other E5 enzymes
P. hybrida	444	385 (87%)
L. esculentum	444	382 (86%)
A. thaliana	444	370 (83%)
E. coli	427	231 (52%)
S. typhimurium	427	226 (51%)
A. nidulans	451	161 (36%)
S. cerevisiae	469	152 (34%)

Table 5.11 Homologies (identities) between P. sativum E5 and other E5 a.a. sequences

These identities are based on the alignments given in Fig 5.21

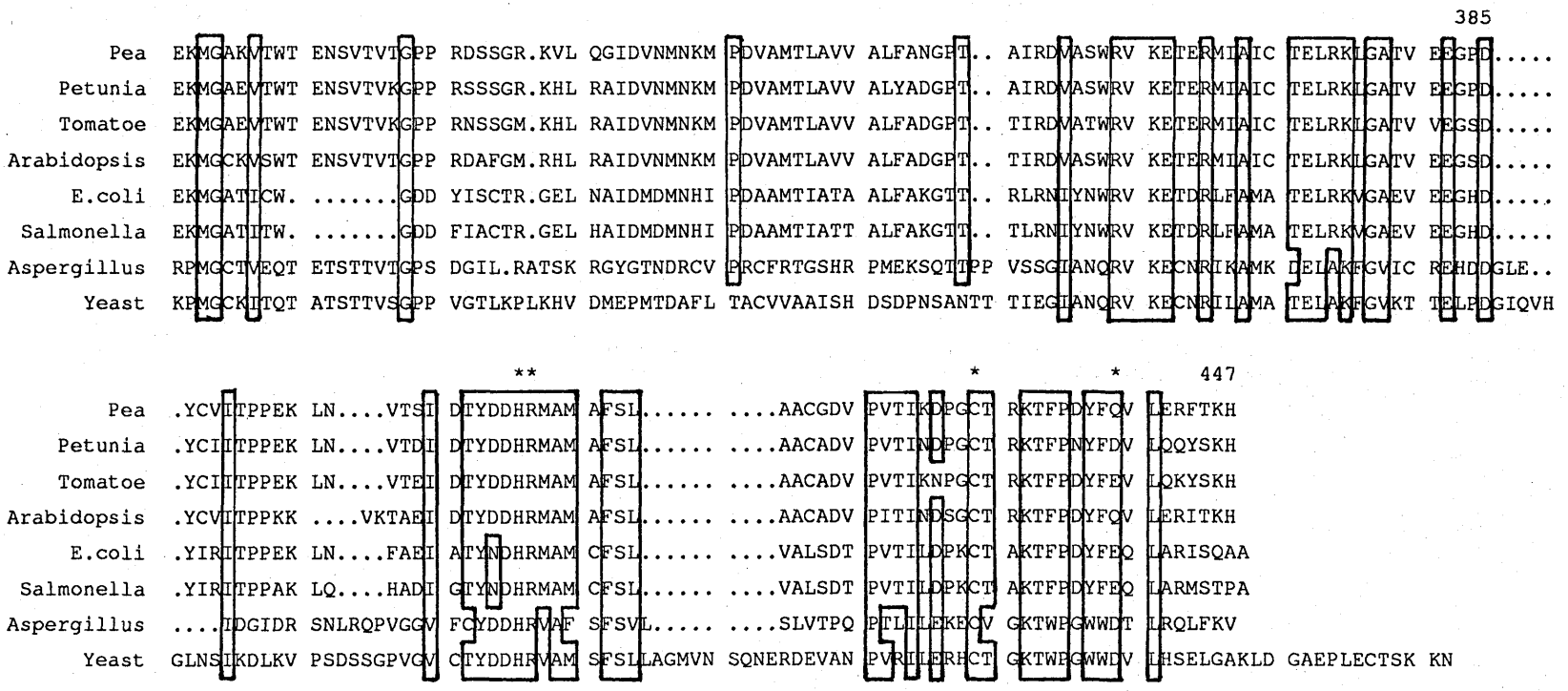
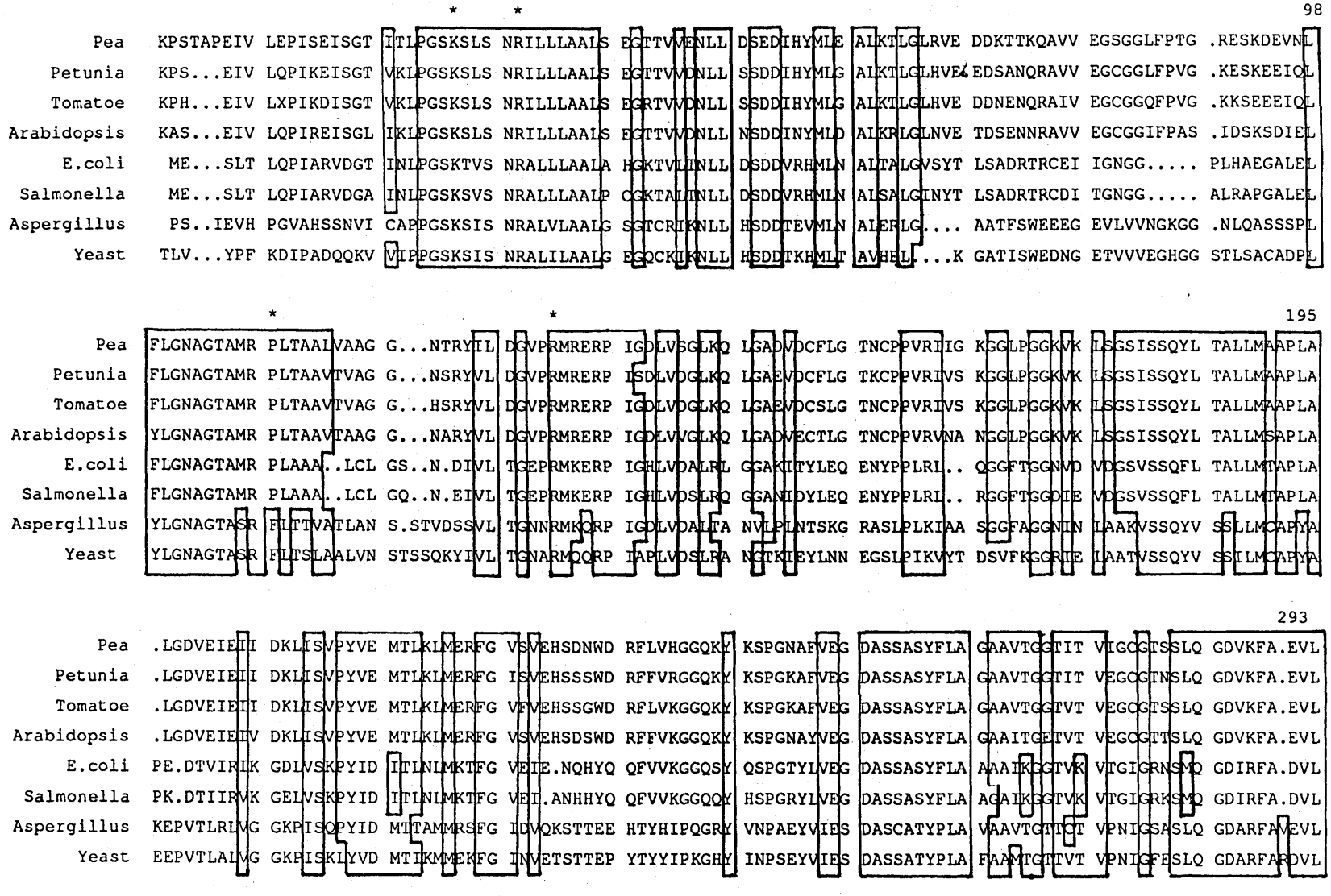


Figure 5.21 Alignment of all known E5 amino acid sequences

The alignment is based on that of Kishore and Shah (1988). Previously identified important residues are asterisked. Identical and chemically similar amino acids are boxed. Numbering is with respect to the P. sativum E5 sequence.


98
Pea KPSTAPEIV LEPISEISGT ITLPGSKSLS NRILLLAALS EGTTVVENLL DSEDIHYMLE AIKTLGLRVE DDKTTKQAVV EGSGGLFPTG .RESKDEVNL
Petunia KPS...EIV LQPIKEISGT VKLPGSKSLS NRILLLAALS EGTTVVDNLL SSDDIHYMLG AIKTLGLHVE EDSANQRAVV EGCGGLFPVG .KESKEEIQL
Tomatoe KPH...EIV LXPIKDISGT VKLPGSKSLS NRILLLAALS EGRTVVDNLL SSDDIHYMLG AIKTLGLHVE DDNENQRAIV EGCGGQFPVG .KKSEEEIQL
Arabidopsis KAS...EIV LQPIREISGL IKLPGSKSLS NRILLLAALS EGTTVVDNLL NSDDINYMLD AIKRLGLNVE TDSENNRAVV EGCGGIFPAS .IDSKSDIEL
E.coli ME...SLT LQPIARVDGT INLPGSKTVS NRALLLAALA HGKTVLTNLL DSDDVRHMLN AITALGVSYT LSADRTRCEI IGNGG..... PLHAEGALEL
Salmonella ME...SLT LQPIARVDGA INLPGSKSVS NRALLLAALP CGKTALTNLL DSDDVRHMLN AISALGINYT LSADRTRCDI TGNGG..... ALRAPGALEL
Aspergillus PS..IEVH PGVAHSSNVI CAPPGSKSIS NRALVLAALG SGTCRIKNLL HSDDTEVMLN AIERLG.... AATFSWEEEG EVLVVNGKGG .NLQASSSPL
Yeast TLV...YPF KDIPADQQKV VIPPGSKSIS NRALILAALG EGQCKIKNLL HSDDTKHMLT AVHELK....K GATISWEDNG ETVVVEGHGG STLSACADPL
195
Pea FLGNAGTAMR PLTAALVAAG G...NTRYIL DGVPRMRERP IGDLVSGLKQ LGADVDCFLG TNCPPVRIIG KGGLPGGKVK LSGSISSQYL TALLMAAPLA
Petunia FLGNAGTAMR PLTAAVTVAG G...NSRYVL DGVPRMRERP ISDLVDGLKQ LGAEVDCFLG TKCPPVRIVS KGGLPGGKVK LSGSISSQYL TALLMAAPLA
Tomatoe FLGNAGTAMR PLTAAVTVAG G...HSRYVL DGVPRMRERP IGDLVDGLKQ LGAEVDCSLG TNCPPVRIVS KGGLPGGKVK LSGSISSQYL TALLMAAPLA
Arabidopsis YLGNAGTAMR PLTAAVTAAG G...NARYVL DGVPRMRERP IGDLVVGLKQ LGADVECTLG TNCPPVRVNA NGGLPGGKVK LSGSISSQYL TALLMSAPLA
E.coli FLGNAGTAMR PLAAA..LCL GS..N.DIVL TGEPRMKERP IGHLVDALRL GGAKITYLEQ ENYPPLRL.. QGGFTGGNVD VDGSVSSQFL TALLMTAPLA
Salmonella FLGNAGTAMR PLAAA..LCL GQ..N.EIVL TGEPRMKERP IGHLVDSLRQ GGANIDYLEQ ENYPPLRL.. RGGFTGGDIE VDGSVSSQFL TALLMTAPLA
Aspergillus YLGNAGTASR PLTTVATLAN S.STVDSSVL TGNNRMKQRP IGDLVDALTA NVLPLNTSKG RASLPLKIAA SGGFAGGNIN IAAKVSSQYV SSLLMCAPYA
Yeast YLGNAGTASR PLTSLAALVN STSSQKYIVL TGNARMQQRP IAPLVDSLRA NGTKIEYLNN EGSLPIKVYT DSVFKGGRIE IAATVSSQYV SSILMCAPYA
293
Pea .LGDVEIEII DKLISVPYVE MTLKLMERFG VSVEHSDNWD RFLVHGGQKY KSPGNAFVEG DASSASYFLA GAAVTGGTIT VIGCGTSSLQ GDVKFA.EVL
Petunia .LGDVEIEII DKLISVPYVE MTLKLMERFG ISVEHSSSWD RFFVRGGQKY KSPGKAFVEG DASSASYFLA GAAVTGGTIT VEGCGTNSLQ GDVKFA.EVL
Tomatoe .LGDVEIEII DKLISVPYVE MTLKLMERFG VFVEHSSGWD RFLVKGGQKY KSPGKAFVEG DASSASYFLA GAAVTGGTVT VEGCGTSSLQ GDVKFA.EVL
Arabidopsis .LGDVEIEIV DKLISVPYVE MTLKLMERFG VSVEHSDSWD RFFVKGGQKY KSPGNAYVEG DASSASYFLA GAAITGETVT VEGCGTTSLQ GDVKFA.EVL
E.coli PE.DTVIRIK GDLVSKPYID ITLNLMKTFG VEIE.NQHYQ QFVVKGGQSY QSPGTYLVEG DASSASYFLA AAAIKGGTVK VTGIGRNSMQ GDIRFA.DVL
Salmonella PK.DTIIRVK GELVSKPYID ITLNLMKTFG VEI.ANHHYQ QFVVKGGQQY HSPGRYLVEG DASSASYFLA AGAIKGGTVK VTGIGRKSMQ GDIRFA.DVL
Aspergillus KEPVTLRLVG GKPISQPYID MTTAMMRSFG IDVQKSTTEE HTYHIPQGRY VNPAEYVIES DASCATYPLA VAAVTGTTVT VPNIGSASLQ GDARFAVEVL
Yeast EEPVTLALVG GKPISKLYVD MTIKMMEKFG INVETSTTEP YTYYIPKGHY INPSEYVIES DASSATYPLA FAAMTGTTVT VPNIGFESLQ GDARFAQDVL

260

amino acid sequences. Thus 26% of the P. sativum (and the other plant) E5 residues have been completely conserved. There are many positions where there is identicality between the majority of the aligned sequences but where the non-identical residues are chemically very similar. When the identical amino acids and chemically similar amino acids are boxed in together (Fig 5.21), then obvious blocks of homology between the plant, bacterial and fungal sequences emerge. These blocks of homology are likely to indicate functionally critical regions of E5 since it appears that they have evolved more slowly.

As discussed in section 1.7 several amino acids have been identified to be present at or near the active site of E5, since their modification is blocked by the presence of S3P and glyphosate. The numbers of the following residues refer to the position in the sequence of the organism in parenthesis. Residues which have been identified: Lys-23 (P. hybrida); Arg-28 (P. hybrida); Arg-131 (P. hybrida); Cys-408 (E. coli); His-385 (E. coli), (Huynh, 1987, 1988; Huynh et al., 1988a, 1988b; Padgette et al., 1988a, 1988b) are completely conserved in all the sequences including that of P. sativum E5. Pro-101 (S. typhimurium) which when mutated to a Ser produces a glyphosate resistant E5 (Comai et al., 1985) is conserved in all plant and bacterial sequences but not in the fungal sequences.

Glu-418 (E. coli) has been identified as being situated at or near the active site of the E. coli enzyme. Alignment of the other sequences suggested that perhaps a negative charge was what was important at this position since it was replaced by an aspartate residue in several

cases (Fig 5.21). However the sequence of both the A. thaliana and the P. sativum E5 enzyme reveals that a glutamine residue can be present at this position in place of a carboxyl containing residue and presumably still give rise to a functional enzyme. This is certainly the case with the A. thaliana E5 (Klee et al., 1987). It may be the case therefore, that it is the carbonyl part of the carboxyl group that is important and that its function can be accommodated by the carbonyl part of an amide group.

These residues are also situated within the blocks of homology which were identified by the alignment of the E5 sequences (Fig 5.21). Thus it would seem that these residues are not only situated at or near the active site of E5 but that they are essential to the structure or function of the enzyme since they are so highly conserved. In addition they are situated in regions where neighbouring amino acids are also very important because they too have been highly conserved during evolution.

5.7.3 DNA homologies of the P. sativum, S. cerevisiae and E. coli E5 sequences

A LINEUP comparison of the cDNA sequences of P. sativum and P. hybrida E5 (Fig 5.22) indicated that there was 77% identity between the DNA sequences of the mature peptides but only 52% identity between the DNA sequences of the small region of transit peptide. This latter figure will probably differ considerably when a comparison of the two full length transit peptides can be carried out.

```
           801                                                                                  900
      Pea  TTGAAGGTGA TGCTTCTAGC GCCAGTTACT TCCTAGCCGG TGCAGCAGTT ACTGGTGGAA CTATCACGGT TATAGGCTGC GGGACAAGCA GTTTACAGGG
   Petunia  TCGAAGGTGA TGCTTCAAGT GCTAGCTACT TCTTGGCTGG TGCAGCAGTC ACAGGTGGAA CTATCACTGT TGAAGGGTGT GGGACAAACA GTTTACAGGG
 Consensus  t.gaaggtga tgcttc.ag. gc.ag.tact tc.t.gc.gg tgcagcagt. ac.ggtggaa ctatcac.gt t..agg.tg. gggacaa.ca gtttacaggg
           901                                                                                 1000
      Pea  AGATGTAAAA TTTGCTGAAG TTCTTGAGAA GATGGGGGCT AAAGTTACGT GGACAGAAAA CAGCGTCACA GTTACCGGGC CTCCACGAGA TTCTTCTGGT
   Petunia  GGAGGTCAAA TTTGCTGAGG TACTTGAAAA AATGGGAGCT GAAGTTACGT GGACAGAGAA CAGTGTCACA GTCAAAGGAC CTCCAAGGAG TTCTTCTGGG
 Consensus  .ga.gt.aaa tttgctga.g t.cttga.aa .atggg.gct .aagttacgt ggacaga.aa cag.gtcaca gt.a..gg.c ctcca.g... ttcttctgg.
           1001                                                                                1100
      Pea  CGGAAAGTGT TGCAAGGCAT TGATGTCAAT ATGAACAAGA TGCCAGATGT TGCCATGACA CTCGCTGTTG TTGCGCTATT TGCTAATGGT CCCACTGCTA
   Petunia  AGGAAGCATT TGCGTGCCAT TGATGTGAAC ATGAATAAAA TGCCTGATGT TGCCATGACA CTTGCTGTTG TTGCACTTTA TGCTGATGGT CCCACAGCTA
 Consensus  .ggaa....t tgc..g.cat tgatgt.aa. atgaa.aa.a tgcc.gatgt tgccatgaca ct.gctgttg ttgc.ct.t. tgct.atggt cccac.gcta
           1101                                                                                1200
      Pea  TTAGAGACGT GGCAAGTTGG AGAGTTAAAG AGACAGAAAG AATGATAGCA ATCTGCACAG AACTCAGAAA GCTAGGAGCA ACAGTTGAAG AAGGTCCCGA
   Petunia  TAAGAGATGT TGCTAGCTGG AGAGTCAAGG AAACTGAGCG CATGATCGCC ATATGCACAG AACTTAGGAA GTTAGGAGCA ACCGTTGAAG AAGGACCAGA
 Consensus  t.agaga.gt .gc.ag.tgg agagt.aa.g a.ac.ga..g .atgat.gc. at.tgcacag aact.ag.aa g.taggagca ac.gttgaag aagg.cc.ga
           1201                                                                                1300
      Pea  TTACTGCGTG ATAACACCAC CTGAGAAACT GAATGTCACG TCAATAGACA CATACGACGA TCATAGAATG GCTATGGCAT TTTCTCTTGC TGCTTGCGGC
   Petunia  CTACTGCATA ATCACCCCAC CGGAGAAACT AAATGTGACC GATATTGATA CATACGATGA TCACAGGATG GCCATGGCTT TTTCTCTTGC TGCTTGTGCA
 Consensus  .tactgc.t. at.ac.ccac c.gagaaact .aatgt.ac. ...at.ga.a catacga.ga tca.ag.atg gc.atggc.t tttctcttgc tgcttg.g..
           1301                                                                          1390
      Pea  GATGTTCCGG TCACAATCAA GGATCCTGGT TGCACAAGGA AAACCTTCCC TGACTACTTT CAAGTCCTTG AGAGGTTCAC AAAGCATTAA
   Petunia  GATGTTCCCG TCACCATCAA TGACCCTGGC TGCACGCGGA AAACCTTCCC TAACTACTTT GATGTACTTC AGCAGTACTC CAAGCATTGA
 Consensus  gatgttcc.g tcac.atcaa .ga.cctgg. tgcac..gga aaaccttccc t.actacttt .a.gt.ctt. ag..gt.c.c .aagcatt.a
```

Figure 5.22 Alignment of the P. sativum E5 cDNA sequence with that of P. hybrida

Only the cDNA sequence coding for the P. sativum E5 protein (including the small amount of transit peptide sequence) and the corresponding P. hybrida cDNA sequence are aligned. The consensus indicates the exact matches between the two sequences.

```
           1                                                                                                  100
       Pea GAATTCCTTT GTGATGAAGG CTTCTGCTTC GGTTGCCGCC GCGGAGAAGC CGTCGACGGC GCCGGAGATT GTGCTGGAAC CGATTAGTGA AATTTCTGGT
    Petunia AAAGTTTTGT TCCTTTAGGA TTTCAGCATC AGTGGCTACA GCACAGAAGC CTTCT..... ....GAGATA GTGTTGCAAC CCATTAAAGA GATTTCAGGC
 Consensus .aa.t..t.t ....t.a.g. .ttc.gc.tc .gt.gc..c. gc..agaagc c.tc..... ....gagat. gtg.tg.aac c.atta..ga .atttc.gg.
           101                                                                                                  200
       Pea ACCATTACTT TGCCTGGATC AAAGTCTCTA TCCAATCGGA TTCTGCTTCT AGCTGCTCTC TCTGAGGGAA CAACTGTTGT AGAGAACTTG TTAGATAGTG
    Petunia ACTGTTAAAT TGCCTGGCTC TAAATCATTA TCTAATAGAA TTCTCCTTCT TGCTGCCTTA TCTGAAGGAA CAACTGTGGT TGACAATTTA CTAAGTAGTG
 Consensus ac..tta..t tgcctgg.tc .aa.tc..ta tc.aat.g.a ttct.cttct .gctgc..t. tctga.ggaa caactgt.gt .ga.aa.tt. .ta..tagtg
           201                                                                                                  300
       Pea AGGATATTCA TTACATGCTC GAGGCATTGA AGACCCTTGG ACTTCGAGTG GAAGACGACA AAACAACCAA ACAAGCAGTT GTGGAAGGTA GTGGGGGATT
    Petunia ATGATATTCA TTACATGCTT GGTGCCTTGA AAACACTTGG ACTGCATGTA GAAGAAGATA GTGCAAACCA ACGAGCTGTT GTTGAAGGTT GTGGTGGGCT
 Consensus a.gatattca ttacatgct. g..gc.ttga a.ac.cttgg act.c..gt. gaaga.ga.a ...caa.c.a ac.agc.gtt gt.gaaggt. gtgg.gg..t
           301                                                                                                  400
       Pea GTTTCCCACC GGTAGAGAAT CTAAAGATGA AGTTAATTTA TTCCTTGGAA ATGCCGGTAC GGCAATGCGT CCTTTGACGG CAGCCTTGGT TGCTGCAGGT
    Petunia TTTCCCTGTT GGTAAAGAGT CCAAGGAAGA AATTCAACTG TTCCTTGGAA ATGCAGGAAC AGCAATGCGG CCACTAACAG CAGCAGTTAC TGTAGCTGGT
 Consensus .tt.cc.... ggta.aga.t c.aa.ga.ga a.tt.a..t. ttccttggaa atgc.gg.ac .gcaatgcg. cc..t.ac.g cagc..t... tg..gc.ggt
           401                                                                                                  500
       Pea GGAAATACAA GATACATACT TGATGGAGTT CCCCGAATGA GAGAAAGACC AATTGGAGAT TTGGTTTCCG GTCTCAAGCA ACTTGGTGCT GATGTTGATT
    Petunia GGAAATTCAA GGTATGTACT TGATGGAGTT CCTCGAATGA GAGAGAGACC AATTAGTGAT TTGGTTGATG GTCTTAAACA GCTTGGTGCA GAGGTTGATT
 Consensus ggaaat.caa g.ta..tact tgatggagtt cc.cgaatga gaga.agacc aatt.g.gat ttggtt...g gtct.aa.ca .cttggtgc. ga.gttgatt
           501                                                                                                  600
       Pea GTTTTCTTGG CACAAATTGT CCACCTGTTC GTATAATTGG GAAGGGAGGA CTTCCAGGGG GAAAGGTGAA ACTGTCCGGA TCTATTAGCA GTCAGTACCT
    Petunia GTTTCCTTGG TACGAAATGT CCTCCTGTTC GAATTGTCAG CAAGGGAGGT CTTCCTGGAG GGAAGGTCAA GCTCTCTGGA TCCATTAGCA GCCAATACTT
 Consensus gttt.cttgg .ac.aa.tgt cc.cctgttc g.at..t..g .aagggagg. cttcc.gg.g g.aaggt.aa .ct.tc.gga tc.attagca g.ca.tac.t
          601                                                                                                  700
       Pea AACTGCTTTG CTTATGGCAG CACCGTTGGC TCTTGGCGAC GTTGAGATTG AGATTATTGA TAAGCTGATT TCTGTTCCGT ATGTTGAAAT GACTTTAAAG
    Petunia GACTGCTCTG CTTATGGCTG CTCCACTGGC TTTAGGAGAT GTGGAGATTG AAATCATTGA CAAACTAATT AGCGTACCTT ATGTCGAGAT GACATTGAAG
 Consensus .actgct.tg cttatggc.g c.cc..tggc t.t.gg.ga. gt.gagattg a.at.attga .aa.ct.att ...gt.cc.t atgt.ga.at gac.tt.aag
           701                                                                                                  800
       Pea TTGATGGAGC GCTTTGGAGT CTCTGTCGAA CACAGTGATA ATTGGGATAG ATTTTTGGTC CACGGAGGTC AAAAGTACAA GTCTCCCGGA AATGCTTTCG
    Petunia TTGATGGAGC GATTTGGTAT TTCTGTGGAG CACAGTAGTA GCTGGGACAG GTTCTTTGTC CGAGGAGGTC AGAAATACAA GTCTCCTGGA AAAGCTTTTG
 Consensus ttgatggagc g.tttgg..t .tctgt.ga. cacagt..ta ..tggga.ag .tt.tt.gtc c..ggaggtc a.aa.tacaa gtctcc.gga aa.gcttt.g
```

The most highly conserved regions between the P. sativum amino acid sequence and either the E. coli or S. cerevisiae sequence were identified from Fig 5.19 and Fig 5.20 and the DNA sequences of the P. sativum and microbial genes corresponding to these regions were compared. The four most homologous E. coli regions of sequence were 60-83% indentical to the P. sativum cDNA sequence (Fig 5.23) whereas the four most homologous S. cerevisiae regions of sequence were 64-72% identical (Fig 5.24).

There is therefore a high percentage DNA sequence identicality between the two plant cDNA sequences which extend along the whole length of the cDNA. This is not the situation when either the E. coli or S. cerevisiae E5 sequences are compared to the P. sativum E5 sequence. In this comparison there are only small dispersed regions of homology. However of the two microbial sequences, that of E. coli is the most homologous with a stretch of 47 bp exhibiting 83% identicality to the P. sativum sequence (Fig 5.23 (ii)). The regions of the microbial sequences showing most homology do not extend far.

It has been shown that there is insufficient homology between the S. cerevisiae and P. sativum E5 gene for significant hybridisation to occur between it and P. sativum genomic DNA (Chapter 3) or with A. thaliana genomic DNA (Chapter 4). The degree of homology between regions of the E. coli and P. sativum E5 gene suggests that the E. coli gene is perhaps more likely to hybridise with the P. sativum E5 gene than is the S. cerevisiae E5 gene. However the E. coli E5 gene was never used to probe (i) genomic southern blots of P. sativum DNA

Figure 5.23 Alignment of the most homologous regions of the E5 DNA sequences of P. sativum and E.coli

The four most homologous regions are shown. In each case the P. sativum sequence is uppermost. The corresponding amino acid sequence is given alongside the DNA sequence. For ease of identification of the homologous region within the whole amino acid sequences, the position of the first amino acid of the P. sativum sequence is given in parenthesis.

```
    (18)      G  T  I  T  L  P  G  S  K  S  L  S  N  R  I  L  L  L  A  A  L  S  E  G  T  T  V  V  E  N  L  L  D  S  E  D

         98 GGTACCATTACTTTGCCTGGATCAAAGTCTCTATCCAATCGGATTCTGCTTCTAGCTGCTCTCTCTGAGGGAACAACTGTTGTAGAGAACTTGTTAGATAGTGAGGA 204
(i)        || || |||| | |||| || || ||| |  | || || ||  | |  | || || ||  |  |  | || | ||| ||  ||  ||  || | ||||| || ||
         40 GGCACTATTAATCTGCCCGGTTCCAAGACCGTTTCTAACCGCGCTTTATTGCTGGCGGCATTAGCACACGGCAAAACAGTATTAACCAATCTGCTGGATAGCGATGA 146

            G  T  I  N  L  P  G  S  K  T  V  S  N  R  A  L  L  L  A  A  L  A  H  G  K  T  V  L  T  N  L  L  D  S  D  D

    (98)      L  F  L  G  N  A  G  T  A  M  R  P  L  T  A  A

        338 TTATTCCTTGGAAATGCCGGTACGGCAATGCGTCCTTTGACGGCAGC 384
(ii)        || ||||| || || ||||| |||||||||||||  || |||||||
        268 TTGTTCCTCGGTAACGCCGGAACGGCAATGCGTCCGCTGGCGGCAGC 314

            L  F  L  G  N  A  G  T  A  M  R  P  L  A  A  A

    (252)     V  E  G  D  A  S  S  A  S  Y  F  L  A  G  A  A

        800 GTTGAAGGTGATGCTTCTAGCGCCAGTTACTTCCTAGCCGGTGCAGCA 847
(111)       || ||||| ||||| |||   ||   |||||| || || |  ||||||
        715 GTCGAAGGCGATGCATCTTCGGCTTCTTACTTTCTGGCAGCAGCAGCA 762

            V  E  G  D  A  S  S  A  S  Y  F  L  A  A  A  A

    (402)     T  Y  D  D  H  R  M  A  M  A  F  S  L  A  A  C  G  D  V  P  V  T  I  K  D  P  G  C  T  R  K  T  F  P  D  Y  F

       1250 ACATACGACGATCATAGAATGGCTATGGCATTTTCTCTTGCTGCTTGCGGCGATGTTCCGGTCACAATCAAGGATCCTGGTTGCACAAGGAAAACCTTCCCTGACTACTT
(iv)        |||||| | |||||  | |||||  |||  || || || |  || |    |||  || || || ||   |||||   |||||   ||||| || || || || ||
       1141 ACATACAATGATCACCGGATGGCGATGTGTTTCTCGCTGGTGGCGTTGTCAGATACACCAGTGACGATTCTTGATCCCAAATGCACGGCCAAAACATTTCCGGATTATTT

            T  Y  N  D  H  R  M  A  M  C  F  S  L  V  A  L  S  D  T  P  V  T  I  L  D  P  K  C  T  A  K  T  F  P  D  Y  F
```

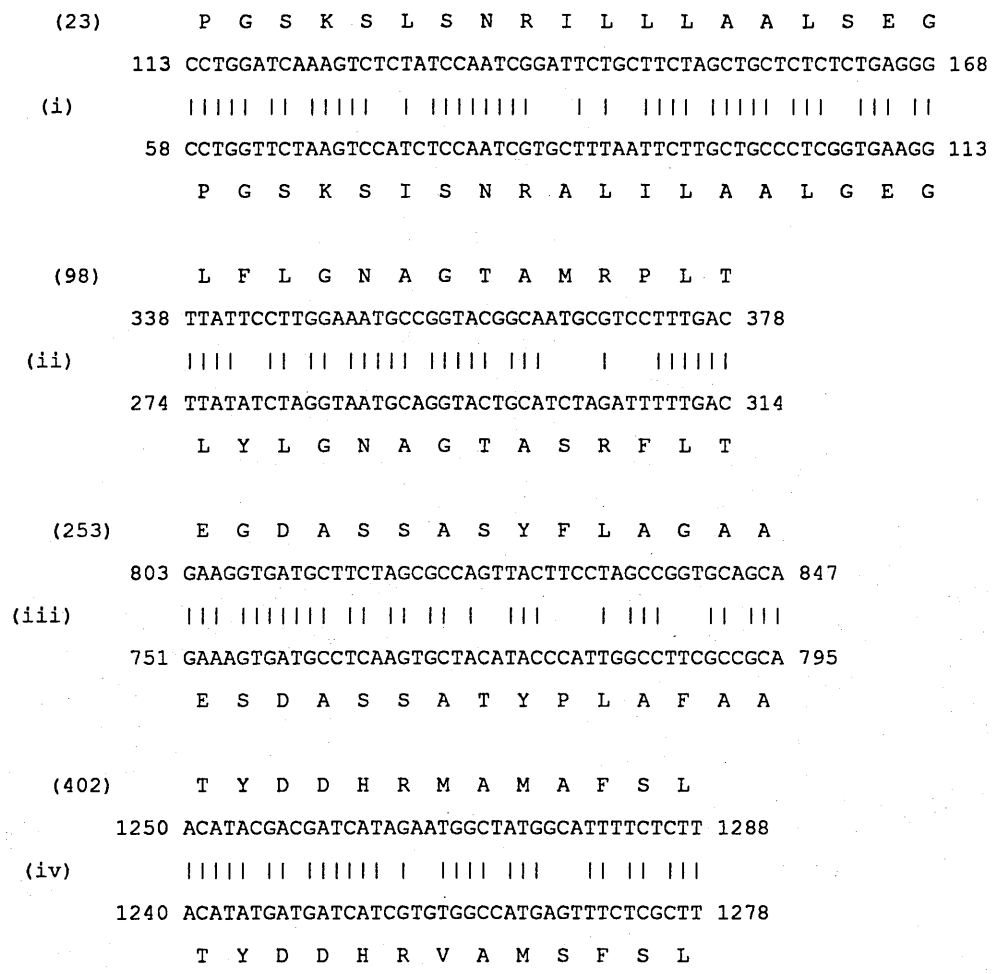


Figure 5.24 Alignment of the most homologous regions of the E5 DNA sequences of P. sativum and S. cerevisiae

The four most homologous regions are shown. In each case the P. sativum sequence is uppermost. The corresponding amino acid sequence in given alongside the DNA sequences. For ease of identification of the homologous region within the whole amino acid sequences, the position of the first amino acid of the P. sativum sequence is given in parenthesis.

or (ii) the P. sativum genomic library or (iii) the A. thaliana genomic library. Therefore it is not known if sufficient homology exists between the E. coli E5 gene and the P. sativum gene for it to have been used as a heterologous probe to identify the P. sativum E5 gene.

5.7.4 Patterns of codon utilisation of P. sativum E5

Many mRNA-coding genes show a bias in the choice of which of several degenerate codons are used to code for a particular amino acid. The abundance of different charged tRNA molecules often reflects this codon bias, i.e. a tRNA molecule for a rarely used codon is correspondingly rare. Any mRNA that contains many rare codons may therefore be translated poorly because of a limiting concentration of tRNA molecules and conversely, genes which are expressed efficiently contain codons for which the tRNA molecules are abundant (Bennetzen and Hall, 1982; Kingsman and Kingsman, 1988). This has been shown to be the case in yeast (Sharp et al., 1986) and in E. coli (Gouy and Gautier, 1982; Grosjean and Fiers, 1982) although the bias is not the same in both organisms.

The profile of the P. sativum E5 codon usage may reflect the codon bias of a gene which is expressed at low levels in P. sativum. It may be useful for the design of oligonucleotide probes from sequence data of other P. sativum enzymes, especially for those of shikimate pathway enzymes.

The codon usage of the available P. sativum E5 sequence, ie the mature enzyme and the small amount of transit peptide is given in Table 5.12. There is perhaps a predominance of A or T nucleotides in the third position of some of the codons, namely those of alanine glycine, valine, isoleucine, serine, threonine, aspartate and asparagine. Usage of the other codons appeared to be random with the possible exception of the codon TAC, which encoded eight of the nine tyrosines.

The relevence of these observations will be ascertained when sequence data from other low expression P. sativum genes, including the other shikimate pathway genes, becomes available.

5.8 Conclusion

P. sativum cDNA libraries were prepared with the plan of screening them with nucleic acid probes for the genes of several of the shikimate pathway enzymes. The failure to obtain any sequence data from the bifunctional E2/E3 enzyme of P. sativum meant that oligonucleotide probes, based on the deduced DNA sequence of the E2/E3 gene, could not be synthesised. A cDNA clone for the P. hybrida E5 gene was however used to screen one of the cDNA libraries which had been prepared from sucrose gradient fractionated poly A^+ mRNA.

Two types of hybridising plaques were obtained. They differed in the size of the signal produced on the autoradiograph but the reason for this difference was not investigated. Eighteen large strongly hybridising plaques were analysed and found to contain very similar

Table 5.12 Codon utilisation of P. sativum E5

The codons for the P. sativum mature E5 enzyme and also the 15 codons for the known region of the E5 transit peptide are included in the table.

LEU:	TTA	4 (9)	SER:	AGT	8 (24)	ARG:	ALA	12 (20)
	TTG	12 (27)		AGC	4 (12)		AGG	2 (10)
	CTA	5 (11)		TCA	2 (6)		CGA	3 (15)
	CTG	6 (13)		TCG	2 (6)		CGG	2 (10)
	CTT	13 (29)		TCT	13 (39)		CGT	2 (10)
	CTC	5 (11)		TCC	4 (12)		CGC	1 (5)
ALA:	GCA	13 (32)	GLY:	GGA	18 (41)	VAL:	GTA	2 (5)
	GCG	7 (17)		GGG	6 (14)		CTG	8 (19)
	GCT	18 (44)		GGT	15 (34)		GTT	24 (57)
	GCC	3 (7)		GGC	5 (11)		GTC	8 (19)
PRO:	CCA	6 (25)	THR:	ACA	16 (46)	ILE:	ATA	6 (25)
	CCG	6 (25)		ACG	6 (17)			
	CCT	7 (29)		ACT	7 (20)		ATT	15 (63)
	CCC	5 (21)		ACC	6 (17)		ATC	3 (13)
PHE:	TTT	9 (64)	CYS:	TGT	2 (29)	HIS:	CAT	3 (60)
	TTC	5 (36)		TGC	5 (71)		CAC	2 (40)
GLN:	CAA	5 (71)	ASN:	ATT	11 (79)	ASP:	GAT	19 (73)
	CAG	2 (29)		AAC	3 (21)		GAC	7 (27)
GLU:	GAA	16 (55)	TYP:	TAT	1 (11)	LYS:	AAA	10 (38)
	GAG	13 (45)		TAC	8 (89)		AAG	16 (62)
MET:	ATG	14 (100)	TRP:	TGG	3 (100)	STOP:	TAA	1
							TAG	0
							TGA	0

cDNA inserts. One of these clones was completely characterised and because it showed striking homology to the P. hybrida E5 cDNA clone it was concluded that it was a cDNA clone of P. sativum E5. The P. sativum mature E5 was also shown to be very homologous to the other plant E5 enzymes and to have considerable homology to the bacterial E5 enzymes but less homology to the fungal E5 enzymes. Previously identified important active site amino acid residues of E5 enzymes were conserved in the P. sativum sequence. The side chain of one of these residues previously found to be a carboxylate group was replaced by an amide group in the P. sativum E5 sequence.

All of the coding cDNA for the P. sativum mature E5 was isolated but a portion of the transit peptide coding cDNA was apparently absent from the clone. It was deduced that none of the other 17 cDNA clones were likely to contain the complete full length cDNA. However it was not ruled out that the second type of plaque which hybridised with the P. hybrida E5 probe during screening of the cDNA library contained the missing cDNA.

Poor efficiency of the EcoRI methylase reaction during the cloning of the cDNA was assumed to be solely responsible for the non-isolation of full length cDNA clones of P. sativum E5. The other cloning steps were thought to have been very successful. With hindsight, it would have been appropriate to have carried out test methylation reactions in order to determine the efficiency of the enzyme.

Full length cDNA clones, or at least cDNA molecules long enough to

pass the EcoRI site within the transit peptide coding sequence, had probably been synthesised during the cDNA synthesis reactions. This indicated that (i) cDNA synthesis and (ii) the RNA preparation and fractionation of the poly A^+ mRNA had been successful. Since 20 clones were obtained after screening 600,000 plaques, then the ratio of 1 clone: 30,000 plaques suggests that enrichment of the poly A^+ mRNA, which was achieved, was approximately 5 fold, assuming that there are about 150,000 mRNA molecules per eukaryotic cell (Williams, 1981).

Northern blot analysis of the mRNA populations used to synthesise the cDNA libraries did not reveal mRNA sequences which were complementary to the P. hybrida E5 probe. Thus it was concluded that the P. sativum E5 mRNA molecules were present in very low abundance. However these low abundance sequences were still represented within the enriched cDNA library and were identified readily with the P. hybrida E5 probe. Should appropriate nucleic acid probes become available then the enriched cDNA library and probably the other two cDNA libraries which were prepared have every chance of providing cDNA clones of the other shikimate pathway genes.

CHAPTER 6

GENERAL DISCUSSION AND FUTURE PROSPECTS

6.1 The achievements of this thesis

The aim of the work presented in this thesis was to isolate one or more of the *P. sativum* genes for the shikimate pathway enzymes, determine the DNA sequence, deduce the corresponding amino acid sequence and compare the peptide sequences with those microbial sequences which have previously been obtained. This objective has been achieved with respect to the *P. sativum* E5 gene. In addition it has been shown that heterologous probes derived from the *S. cerevisiae* ARO1 gene were unable to cross-hybridise with the *P. sativum* or *A. thaliana* genes for the shikimate pathway enzymes. The studies suggested that the *E. coli* genes for these enzymes were unlikely to do so either.

6.2 Heterologous probing with the S. cerevisiae probes

6.2.1 Heterologous probing experiments were justifiable

At the outset of the project there were four obvious ways in which the genes for the *P. sativum* shikimate pathway enzymes might be isolated (section 1.14). The use of heterologous probes, derived from the microbial genes, to screen libraries of *P. sativum* (or any other higher plant) cDNA or genomic DNA appeared to be the strategy which could be attempted most readily. In particular it seemed likely that the *S. cerevisiae* genes would be most homologous to the *P. sativum* genes because both organisms are eukaryotic and are therefore presumably more closely related than *E. coli* and *P. sativum*.

Although heterologous hybridisations have been employed to isolate genes from related species, the technique had not been used extensively at that time (or even since) to isolate genes from evolutionary diverse sources, as judged by the lack of such reports in the literature. It was realised that the attempts to isolate the P. sativum genes using this approach were ambitious and would probably set a precedent.

While this work was being carried out, such a precedent was set by workers at DuPont, who isolated the acetolactate synthase (ALS) gene from genomic libraries of two dicotyledenous plants, A. thaliana and Nicotiana tabacum using the ALS gene of S. cerevisiae (Mazur et al., 1987). Apparantly no signal was obtained when the microbial probe was used to probe genomic blots of the plant DNA. However the DNA amplification in the phage plaques was sufficient for a hybridisation signal to be detected when the probe was used against both plant genomic libraries. The plant sequences were highly homologous to one another (85% identicality) whereas the plant and S. cerevisiae sequences were only 40% identical at the amino acid level. This report provided great encouragement to our pursuit of the shikimate pathway genes using permissive hybridisation with heterologous probes.

6.2.2 The S. cerevisiae and the plant genes were unable to cross-hybridise

As detailed in chapters 3 and 4, the probes derived from the S. cerevisiae ARO1 gene were unable to detect the corresponding genes

from P. sativum on genomic southern blots. With hindsight, the multiple banding pattern obtained on the southern blots hybridised with the S. cerevisiae E1 and E5 probes was probably a result of slight homology of the probes to repeated elements within the P. sativum genome. The S. cerevisiae E2/E3 probe failed to isolate the desired gene from a P. sativum genomic library. In addition the E1, E2/E3 or E5 probes could not detect the corresponding genes within an A. thaliana genomic library. These findings indicated that insufficient homology existed between the S. cerevisiae ARO1 gene and the genes for the P. sativum or A. thaliana shikimate pathway enzymes.

6.2.3 Explanation for non-hybridisation

The homology between the S. cerevisiae ALS genes and the plant ALS genes was only 40% and yet the fungal gene could hybridise with the plant gene (Mazur et al., 1987). The homology between the S. cerevisiae E5 gene and the cDNA clone for the P. sativum E5 (this work) is 34%. There is therefore a difference of 6% between the homology of the shikimate pathway gene (E5) from P. sativum and S. cerevisiae compared to the ALS gene from A. thaliana (or N. tabacum) and S. cerevisiae. This difference must be significant enough for genes of E5 to be too distantly related to allow hybridisation between them. Mazur et al., (1987) point out that the homology between the S. cerevisiae and plant ALS protein sequences is restricted, to a large degree, to three highly conserved regions which are separated by regions of divergence. They suggest that it is the concentration of homology at these functionally critical regions which was sufficient

to allow hybridisation of the fungal gene to the plant genes. Although there are similar regions of highly conserved peptide sequence among the plant and fungal E5 sequences, the homology does not extend to the DNA level (Fig 5.24).

The homology between the S. cerevisiae and E. coli E5 amino acid sequences is 38% which is the highest percentage homology between the S. cerevisiae and E. coli shikimate pathway genes (Duncan et al., 1987). The E1 sequences have the next highest homology (36%) while those of E2 (21%) and E3 (25%) are both only weakly homologous. Presumably therefore, it is more important that the amino acid sequence of E1 and E5 is conserved than it is for E2 and E3. It is reasonable to expect that the microbial E5 and E1 sequences will be more homologous to the plant sequences than will E2 and E3. Now that the P. sativum and other plant E5 sequences have been compared to that of S. cerevisiae and the low amount of homology observed (34%), it is likely that the homology between the other sequences of plant and S. cerevisiae shikimate pathway enzymes will not be higher than 34%. This explains why the S. cerevisiae genes were unable to detect the plant shikimate pathway genes.

6.3 Plant E5 sequences are very homologous

A P. hybrida E5 cDNA clone was sufficiently homologous to a P. sativum E5 cDNA clone to enable the isolation of the latter from a P. sativum cDNA library. From DNA sequence analysis of the P. sativum E5 clone it was deduced that the mature enzyme was 447 amino acids in length

and had Mr 48,716. There was evidence that the enzyme contained a transit peptide, although the transit peptide was not completely characterised.

There was extensive homology between the E5 cDNA clones of P. hybrida and P. sativum which had enabled hybridisation of the two sequences to occur. However it is likely that any of the previously characterised plant E5 genes (from L. esculentum or A. thaliana) could have been used to isolate the P. sativum E5 cDNA clone because the homology between all four sequences was so extensive. This is a consoling observation because it vindicates the strategy described in section 4.1, where it was suggested that a clone for any of the A. thaliana shikimate pathway enzymes would have been able to hybridise with the corresponding P. sativum gene.

The sequence at the transit peptide cleavage site appears to be highly conserved between all four plant E5 enzymes. It is not unusual among transit peptides of homologous proteins from related species to exhibit small regions of homology (Shmidt and Mishkind, 1986). However this homology rarely extends along the length of the transit peptide and it is unlikely that the conservation will be upheld throughout the full length of the P. sativum E5 transit peptide since the other E5 transit peptide sequences diverge substantially. Also the complete transit peptide sequence of P. sativum E5 is unlikely to be similar to those peptides of other translocated proteins since transit peptides are known to exhibit a high degree of diversity (Keegstra, 1989).

Amino acids are conserved all the way along the E5 sequences from the four species of plant (P. sativum, P. hybrida, L. esculentum and A. thaliana) and many of the residues which do differ appear to be conservative substitutions. However it is when the bacterial and fungal sequences are aligned alongside the plant sequences (Fig 5.21) that regions which are highly conserved across the evolutionary divide become apparent. Within these regions are residues which have been identified as being functionally critical, i.e. residues either involved in catalysis or binding of substrates at the active site of E5. When a crystal structure for E5 becomes available it will be easier to identify why these regions and particular residues are important.

6.4 A fourth plant E5 sequence is useful.

There are now four plant E5 sequences which have been determined. The P. hybrida E5 gene has been expressed within a bacterial host (Padgette et al., 1987) and so presumably the other plant E5 genes could be expressed in a similar manner. As yet it has proved difficult to obtain good crystals of purified plant E5 enzyme. The reasons for this are not known but the difficulty in crystallisation is presumably a characteristic of the amino acid sequence of the E5 enzyme. Having a fourth plant E5 gene available (i.e. that of P. sativum), which could be expressed within a bacterial host, provides another opportunity for attempts to crystallise the enzyme in order to obtain structural information about E5.

The isolation of a clone for E5 from P. sativum is also part of the long term aims of the group at Glasgow as is described in section 1.12.

6.5 The origin of plant E5 genes

6.5.1 The origin of photosynthetic organelles

It is now widely believed that plastids are descended from a prokaryotic cell, a cyanobacterium, that took up a symbiotic residence within the cytosol of some ancient eukaryote. This hypothesis was initially based on striking similarities between chloroplasts and cyanobacteria e.g. both contain photosynthetic pigments, both contain a chromosome and both have a similar protein synthesising system (Doolittle, 1980). More recently, significant homologies between cyanobacterial genes and genes present within the chloroplast genome of higher plants have been observed (Wolfe, 1989). Studies involving the comparison of the 16S rRNA sequences from cyanobacteria and chloroplasts confirm that they originate from a common ancestor. Not only that but the progenitor of the cyanobacteria family had already diverged into a variety of species before one of these entered the precursor of plant cells (Giovannoni et al., 1988). There is currently disagreement over which member of the cyanobacterial family is the progenitor of chloroplasts (Turner et al., 1989; Morden and Golden, 1989). However, once the endosymbiotic event had taken place, a significant amount of gene transfer probably occurred from the prokaryotic genome to the eukaryotic genome since the majority of

chloroplast proteins require to be imported from the cytosol.

6.5.2 Plant E5 sequences are similar to their bacterial counterparts

The comparison of the E5 sequences from plants, fungi and bacteria clearly indicates that the bacterial and plant enzymes (52% identicality between the E. coli and P. sativum sequences) exhibit a greater degree of similarity than the fungal and plant enzymes (34% identicality between the S. cerevisiae and P. sativum sequences). This was a very unexpected finding. These homologies suggest that prokaryotes are more closely related to higher plants than fungi, which is contrary to traditional evolutionary theories and is therefore unlikely. The alternative, more plausable explanation is that since the event of plant and fungal divergence, the E5 gene of an ancestral prokaryote has migrated from the prokaryotic genome to the nucleus of an ancestral plant. This hypothesis is in strong agreement with the proposed model of the endosymbiotic origin of plastids and lends weight to the hypothesis.

6.6 Use of the E.coli genes as heterologous probes

The P. sativum genes for the other shikimate pathway enzymes are also likely to exhibit more similarity to the prokaryotic genes than the fungal genes and so perhaps with hindsight it may have been wise to pursue these plant genes more extensively using probes derived from the bacterial genes.

It was shown conclusively that the E. coli E1, E2, E3 and E6 (E5 was never used) would not cross hybridise with genomic southern blots of P. sativum DNA. However perhaps amplification of the P. sativum genes within a cDNA or genomic library would have enabled the appropriate clones to be identified. This was certainly the situation when the plant ALS genes were isolated using the S. cerevisiae probe (Mazur et al., 1987). Only the E. coli E6 probe was used to screen a plant library. The result of screening the A. thaliana genomic library was a background signal which obliterated any signal that may have been produced between the plaques and the E6 probe. Perhaps this signal could have been eliminated by housing the plant genomic library within a strain of E. coli which had the aroC (E6) gene completely deleted from its genome. In order to screen with the other E. coli derived probes, similar deletion auxotrophs would also be required. These were not available during the course of this work.

6.7 Future prospects

6.7.1 Studies with the E5 cDNA clone

The only basis on which the P. sativum E5 cDNA clone has been identified is that of sequence homology to the E5 sequences from other sources. An unequivocal test as to the identity of the cDNA clone would be its complementation of an E. coli aroA$^-$ auxotroph e.g. E. coli AB2829. For this to be achieved the cDNA clone for (presumably) E5 would require manipulation at the 5' and 3'- end of the mature enzyme, by site directed mutagenesis, to allow expression of the

eukaryotic enzyme within the prokaryotic host. Expression of the E5 cDNA clone would also enable enzymological studies to be carried out on the plant enzyme.

The unfortunate natural occurence of an EcoRI site within the transit peptide coding sequence of P. sativum E5 along with the poor EcoRI methylation of the cDNA, prevented the isolation of the full transit peptide sequence at the same time as the cDNA coding for the mature E5 enzyme. However the size of the cDNA products, after cDNA synthesis, (Fig 5.5) combined with the regularity in size of the E5 cDNA clones (only one clone was smaller than 1.6 kbp, Fig 5.9) strongly suggests that cDNA of sufficient length to encode the E5 transit peptide would have been present within at least the size enriched cDNA library. With this in mind it would be interesting to examine the small strongly hybridising spots which occurred on screening the cDNA library with the P. hybrida E5 probe (Fig 5.8). These plaques may contain a larger cDNA fragment making them less able to replicate and therefore smaller in size. Providing the EcoRI methylase was functioning to some extent then the inserts of these plaques may contain, in addition to the mature E5 coding cDNA, the transit peptide coding cDNA. It is unlikely that the plaques would contain the transit peptide coding cDNA on its own because the P. hybrida probe did not contain any transit peptide cDNA (Fig 5.7).

A more certain way of obtaining the transit peptide coding sequence would be to generate more P. sativum cDNA (not all the cDNA was used in the construction of the libraries) and assume the cDNA to be full

length. Using an oligonucleotide designed specifically to hybridise to the E5 cDNA clone, the E5 cDNA could be amplified using the polymerase chain reaction. Instead of using EcoRI linkers, BamHI linkers could be used (or linkers for any other restriction enzyme) and the digested amplified cDNA could be cloned into an appropriate host. Perhaps amplification would be sufficient for a plasmid host (or M13) to be used. Screening the resultant clones with the insert from pSG5161 (Fig 5.16) should yield the desired sequence.

The most direct way of obtaining the transit peptide would be to isolate a genomic clone for the P. sativum E5 gene (perhaps from the P. sativum genomic library generated during this work) and completely characterise the clone by restiction mapping and sequencing. This should provide the transit peptide coding sequence. Alternatively the appropriate region of the clone could be used to screen the enriched P. sativum cDNA library generated during this study for the desired sequence.

Once obtained, investigations into the targeting capacity of the P. sativum E5 transit peptide could be carried out. Products from in vitro translated mRNA of constructs with/without the transit peptide could be used in conjunction with isolated P. sativum chloroplasts for such investigations.

The thorough characterisation of an E5 genomic clone would enable the sensible interpretation of genomic southern blots of P. sativum DNA probed with the insert from pSG5161. Prediction of the band sizes

would be possible and any deviation from the predicted pattern would suggest that there was more than one copy of the E5 gene. In view of the interest in isozymes of the shikimate pathway enzymes it would be interesting to investigate such bands by attempting to clone them.

The cDNA insert of pSG5161 could be used to study expression of the E5 gene. RNA could be isolated from various regions of pea plants, blotted onto nitrocellulose and probed with the insert from pSG5161. This would indicate the regions of the pea plant where the E5 gene was being expressed most highly.

6.7.2 Isolating the other shikimate pathway genes

The most obvious strategy for the continued pursuit of the *P. sativum* shikimate pathway genes is to purify the enzymes from the plant. This is notoriously difficult since they are present at such low levels however it has been achieved for E2/E3 and E5. It was disconcerting and unpredicted when the purified E2/E3 failed to sequence. The explanation for this was thought to be blocking of the N-terminus of the enzyme. E2/E3 is presumably translocated to the chloroplast via a transit peptide mechanism and therefore blocking of the N-terminus of the mature protein must have occurred within the chloroplast. It has also been reported that EO from *S. tuberosum* is blocked at its N-terminus and so perhaps N-terminal blocking is a characteristic of some of the enzymes of this pathway. It cannot be a completely general feature since it is known that the mature choroplastic E5 of *P. hybrida* is not blocked (Shah *et al.*, 1986). To obtain sequence

information for the blocked enzymes of the pathway will require their purification followed by digestion into peptide fragments. One or more of the purified peptides would then be sequenced and from this sequence oligonucleotides could be prepared to screen P. sativum cDNA libraries. The codon utilisation table (Table 5.12) may be of assistance in this respect.

Alternatively the E. coli genes for the shikimate pathway enzymes could be used to screen P. sativum cDNA libraries housed in appropriate auxotorphic strains as discussed in section 6.6.

These genes may indeed ultimately be cloned via A. thaliana. Recently, mutants of A. thaliana requiring tryptophan to survive, have been isolated (Meyerowitz, 1989). One of these has been characterised and been found to contain a mutation in the anthranilate phosphoribosyl transferase gene. Perhaps other mutants exist, or can be generated, where the mutation is in the genes for the shikimate pathway enzymes.

REFERENCES

Ahmed, S. I. and Giles, N.H. (1969) J. Bacteriol. 99, 213-237.

Anton, I. A. (1985), PhD Thesis, University of Glasgow.

Anton, I.A. and Coggins, J.R. (1988) Biochem. J. 249, 319-326.

Appleyard, R.K. (1954) Genetics 39, 440-452.

Amrhein, N., Schab, J. and Steinrucken, H.D. (1980) Naturwissenschaften 67, 356-357.

Bachman, B. (1983) Microbiol Rev. 44, 180-230.

Bedbrook, J. R., Smith,S.M. and Ellis, J.R. (1980) Nature 257, 692-697.

Benfey, P.N. and Chua, N-H. (1989) Science 244, 174-181.

Bennett, J. (1981) Eur. J. Biochem. 118, 61-70.

Bennett, J., Jenkins, G.I. and Hartley, M.R. (1984) J. Cell. Biochem. 25, 1-13.

Bennett, M.D. and Smith, J.B. (1976) Philos. Trans. R. Soc. London, Ser. B. 274, 227-274.

Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3018-3025.

Berger., S. L. Methods in Enzymology 152, 49-54.

Berlyn, M.B. and Giles, N.H. (1969) J. Bacteriol. 99, 220-230.

Bickel, H., Plame, L. and Shultz, G. (1978) Phytochemistry 17, 119-124.

Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Nat. Acad. Sci. U.S.A. 80, 3963-3965.

Birnboim, H.C. and Doly, J. (1979) Nuc. Acid Res. 7, 1513-1523

Bolivar, F. and Backman. K. (1979) Methods in Enzymology 68, 245-267.

Bonner, T.I., Brenner, D.J. Neufield, B.R. and Britten, R.J. (1973) J. Mol. Biol. 81, 123-135.

Boocock, M.R. (1983) PhD Thesis, Glasgow University.

Boocock, M.R. and Coggins, J.R. (1983) FEBS Letts. 154 (1), 127-133.

Boudet, A. M., Graziana, A. and Ranjeva, R. (1985) in 'The Biochemistry of Plant Phenolics' (Von Sumere, C.F. and Lea, P.J., eds.) Ann. Proc. Phytochem. Soc. Eur. 25, Claredon Press, Oxford.
Boulnois, G.T. (1987) in 'Gene Cloning and Analysis - a laboratory guide' (G.T. Boulnois, ed) Blackwell Scientific Press .

Bucholz, B., Reupke, B., Bickel, H. and Shultz, G. (1979) Phytochemistry 18, 1109-1111.

Chang, C. and Meyerowitz, E. M. (1986) Proc. Nat. Acad. Sci, U.S.A. 83, 1408-1412

Charles, G., Keyte, J.M., Brammer, W.J., Smith, M. and Hawkins, A.R. (1986) Nuc. Acid. Res. 14 (5), 2201-2213.

Clark, L. and Carbon, J. (1976) Cell 9, 91-99.

Coggins, J.R., Boocock, M.R., Campbell, M.S., Chaudhuri, S., Lambert, J.M., Lewendon, A., Mousdale, D.M. and Smith, D.D.S. (1985) Biochem. Soc. Trans. 13, 299-303.

Coggins, J.R. (1986) in Biotechnology and Crop Improvement and Protection (Day, P.R., ed.), 101-110, British Crop Protection Council publications, Croydon.

Coggins, J.R. (1989) in 'The Enzyme Catalysis Process: Energetics, Mechanisms and Dynamics' (A. Cooper and J.L. Houben, eds) Plenum in press.

Comai, L., Sen, L. and Stalker, D. (1983) Science 221, 370-371.

Comai, L., Faccioti, W.R., Hiatt, G., Thompson, G., Rose, R.E. and Stalker, D.M. (1985) Nature 317, 741-744.

Comai, L., Kelly, N.L., Kiser, J., Mau, C.J.D., Pokalsky, A.R., Shewmaker, C.K., McBride, K., Jones, A. and Stalker, D.M. (1988) J. Biol. Chem. 263 (29), 15104-15109.

Conn, E.E. (Ed.) (1986) Recent Advances in Phytochemistry - Vol 20. 'The Shikimic Acid Pathway' Plenum Press, New York.

Corruzi, G., Broglie, R., Edwards, C. and Chua, N. (1984) EMBO J. 3 (8), 1671-1679.

Davis, W.D. and Davidson, B.E. (1982) Nuc. Acid Res. 10, 4045-4058

Della-Ciopa, G., Bauer, S.C., Klein, B.K., Shah, D.M., Fraley, R.T. and Kishore, G.M. (1986) Proc. Nat. Acad. Sci. U.S.A. 83, 6873-6877.

Della-Cioppa, G., Bauer, S.C., Taylor, M.L., Rochester, D.E., Klein, B.K., Shah, D.M., Fraley, R.T. and Kishore, G.M. (1987) Bio/Technology 5, 579-584.

Della-Cioppa, G. and Kishore, G.M. (1988) EMBO J. 7 (5), 1299-1305.

Devereux, J., Haeberli, P. and Smithies, O. (1984). Nuc. Acid Res. 12, 387-395.

Doolittle, W.F. (1980) Trends Biochem. Sci. 5, 146-149.

Duncan, K. (1984) PhD Thesis, Glasgow University.

Duncan, K., Lewendon, A. and Coggins. J.R. (1984) FEBS Letts. 170, 59-63.

Duncan, K., Chaudhuri, S., Campbell, M.S. and Coggins, J.R. (1986) Biochem. J. 238, 475-483.

Duncan, K., Edwards, R.M. and Coggins, J.R. (1987) Biochem. J. 246, 375-386.

Duncan, K., Edwards, R.M. and Coggins, J.R. (1988) FEBS Letts. 241 (1), 83-88.

Ebel, J. and Grisbach, H. (1988) Trends Biochem. Sci. 13, 23-27.

Estelle, M.A. and Somerville C.R. (1986) Trends Genet. 2, 89-93.

Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.

Flavell, R. (1980) Ann. Rev. Plant Physiol 31 569-596.

Frost, J., Bender, J., Kadonoga, J.T. and Knowles, J.R. (1984) Biochemistry 23, 4470-4475.

Gaertner, F.H. (1972) Arch. Biochem. Biophys. 151, 277-284.

Gaertner, F.H. and Cole, K.W. (1977) Biochem. Biophys. Res. Commun. 75, 259-264.

Ganson, R.J., D'Amato, T.A. and Jensen, R.A. (1986) Plant Physiol. 82, 203-210.

Gansen, R.J. and Jensen, R.A. (1988) Arch. Biochem. Biophys. 260 (1), 85-93.

Gasser, C.S., Winter, J.A., Hironaka, C.M. and Shah, D.M. (1988) J. Biol. Chem. 263 (9), 4280-4289.

Gilchrist, D.G. and Kosuge, T. (1980) in 'The Biochemistry of Plants' (Miflin, B.T. ed.) 5, 507-531, Academic Press, New York.

Giles, N. H., Case, M.E., Partridge, C.W.H and Ahmed, S.I. (1967) Proc. Nat. Acad. Sci. U.S.A. 58, 1453-1460.

Giovannoni, S.J., Turner, S., Olsen, G.J., Barns, S., Lane, D.J. and Pace, N.R. (1988) j. Bacteriol. 170 (8), 3584-3592.

Gouy, M and Gautier, C. (1982) Nuc. Acids. Res. 10, 7055-7074.

Graziana, A., Boudet, A. and Boudet, A.M. (1980) Plant Cell. Physiol. 21 (7), 1163-1174.

Greene, P.J., Poonian, M.S., Nussbaum, A.L., Tobias, L., Garfin D.F., Boyer, H.W., and Goodman, H.M. (1975) J. Mol. Biol. 99, 237-261.

Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209.
Habereder, H., Schroder, G. and Ebel,J. (1989) Planta 177. (1), 58-65.

Hames, B.D. and Higgins, S.J. (1984) in 'Transcription and Translation - a practical approach' IRL Press, Oxford.

Hames, B.D. and Higgins, S.J. (1985) 'Nucleic Acid Hybridisation - a practical approach' IRL Press, Oxford.

Haslam, E. (1974) 'The Shikimate Pathway' Butterworths, London.

Hawkins, A.R. (1985) Nuc. Acid. Res. 13 (22), 8119-8128.

Herrman, K.M. and Somerville, R.L. (1983) in'Amino acids: biosynthesis and genetic regulation' London, Addison-Wesley.

Herrman, K.M., (1987, July 19-23) Abstract No. 864 in Annual Meeting Am. Soc. Plant Physiol. St. Louis. U.S.A.

Hollander-Cytko, H., Johanning, D., Meyer, H.E. and Amrhein, N. (1988) Plant Mol. Biol. 11, 215-220.

Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193-197.

Howley, P.M., Isreal, M.F., Law, M.F. and Martin, M.A. (1979) J. Biol. Chem. 254, 4876-4883.

Huynh, Q.K. (1987) Arch. Biochem Biophys. 258, 233-239.

Huynh, Q.K. (1988) J. Biol. Chem., 263 (24), 11631-11635.

Huyng, Q.K., Kishhore, G.M. and Bild, G.S. (1988a) J. Biol. Chem. 263 (2), 735-739.

Huyng, Q.K., Bauer, S.C., Bild, G.S., Kishore, G.M. and Borgmeyer, J.R. (1988b) J. Biol. Chem. 263 (24), 11636-11639.

Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in DNA Cloning Techniques, A Practical Approach (D. Glover, ed) 1, 49, IRL Press, Oxford.

Jacobson, A. (1987) Methods in Enzymology 152, 254-261.

Jacobson, J.W., Hart, B.A., Doy, C.H. and Giles, N.H. (1972) Biochim. Biophys. Acta. 289, 1-12.

Jung, E., Zamir, L.O. and Jensen, R.A. (1986) Proc. Nat. Acad. Sci. U.S.A. 83, 7231-7235.

Jendrisak, J., Young, R.A. and Engel, J.D. (1987) Methods in Enzymology 152, 359-371.

Jensen, R.A. (1985) Physiol. Plant. 66, 164-168.

Jensen, R.A. (1986) in 'Recent Advances in Phytochemistry - The Shikimic Acid Pathway' (E.E. Conn, ed) 20, 57-81 Plenum Press, New York.

Keegstra, K (1989) Cell 56, 247-253.

Kingsman, S.M. and Kingsman, A. J. (1987) in 'Genetic Engineering' 8, Blackwell Scientific Publications.

Kishore, G.M., Brundage, L. Kolk, K., Padgette, S.R. and Rochester, D (1986) Fed. Proc. Am. Soc. Expt. Biol. 45 , 1506-1509.

Kishore, G.M. and Shah, D.M. (1988) Annu. Rev. Biochem. 57, 627-663.

Kimmel, A.R. and Berger, S.B. (1987) Methods in Enzymology 152, 307-316.

Klee, H T., Muskopf, Y.M. and Gasser, C.S. (1987) Mol. Gen. Genet. 210, 437-442.

Koshiba, T. (1978) Biochim. Biophys. Acta. 522, 10-18.

Koshiba, T. (1979) Plant Cell. Physiol. 20 (3), 667-670.

Krebbers, E., Herdies, L., De Clerque, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu , M. and Van-dekerckhove, J. (1988a) Plant Physiol. 87, 859-866.

Krebbers, E., Seurinck, J., Herdies, L., Cashmore, A.R. and Timko, M.P. (1988b) Plant Mol. Biol. 11 (6), 745-750.

Kuroki, G.W. and Conn, E.E. (1988a) Arch. Biochem. Biophys. 260 (2), 616-621.

Kuroki, G.W. and Conn, E.E. (1988b) Plant Physiol. 86, 895-898.

Laemmli, U.K. (1970) Nature 227, 680-685.

Leonard, N.J., McDonald, J.J. and Rēichman, M.E. (1970) Proc. Nat. Acad. Sci. U.S.A. 67 (1), 93-98.

Leonen, W.A.M. and Brammar, W.J. Gene 26, 171-179.

Leonan, W.A.M. and Blattner, F.R. (1983) Gene 26, 171-179.

Leutwiler, L.S., Hough-Evans, B and Meyerowitz, E.M. (1984) Mol Gen Genet. 194, 15-23.

Lumsden, J. and Coggins J.R. (1977) Biochem J. 161, 599-607.
Maniatis T., Fritsch, E.F. and Sambrook, J. (1982) 'Molecular Cloning: A laboratory Manual', New York: Cold Spring Harbor Publications.

Marmur, J.G. and Doty, P. (1961) J. Mol. Biol. 3, 585-594.

Martinez-Zapater, J.M., Estelle, M.A., Somerville, C.C. (1986) Mol. Gen. Genet. 204, 417-423.

Mazur, B.J., Chui, C-F. and Smith, J.K. (1987) Plant Physiol. 85, 1110-1117.

McDonnell, M.W., Simon, M.V. and Studier, F.W. (1977). J. Mol. Biol. 110, 119-146.

McConaughy, B.L., Laird, C.D. and McCarthy, B.J. (1969) Biochemistry 8, 3289-3295.

Messing, J. (1983) Methods in Enzymol. 101, 20-78.

Meyerowitz, E.M. (1987) Annu. Rev. Genet. 21, 93-111.

Meyerowitz, E.M. (1989) Cell 56, 263-269.

Millar, G. and Coggins, J.R. (1986) FEBS Letts. 200, 11-17.

Morden, C.W. and Golden, S.S. (1989) Nature. 337, 382-385.

Morris, P.E., Doong, R-L. and Jensen, R.A. (1989) Plant Physiol. 89, 10-14.

Mousdale, D.M. and Coggins, J.R. (1984) Planta 160, 78-83.

Mousdale, D.M. and Coggins, J.R. (1985) Planta 163, 241-249.

Mousdale, D.M. and Coggins, J.R. (1986) FEBS Letts. 205 (2), 328-332.

Mousdale, D.M., Campbell, M.S. and Coggins, J.R. (1987) Phytochemistry 26, 2665-2670.

Murray, M.G., Cuellar, R.E. and Thompson, W.F. (1978) Biochemistry 17, 5781-5790.

Murray, M.G., Cuellar, R.E. and Thompson, W.F. (1979) Biochemistry 18, 5259-5266.

Murray, N.E., Brammer, W.J. and Murray, K. (1977) Mol. Gen. Genet. 150, 53-61.

Nafziger, E.D., Widholm, J.M., Steinrucken, H.C. and Killmer, J.L. (1984) Plant Physiol. 76, 571-574.

Nakanishi, N. and Yamamoto, M (1984) Mol. Gen. Genet. 195, 164-169.

Nakatasukasa, W.M. and Nester, E.W. (1972) J. Biol. Chem. 247, 5972-5979.

Neuman, P.R. and Hart, G.E. (1983) Biochem. Genet. 21, 963-968.

Nimmo, G.A. and Coggins, J.R. (1981) Biochem. J. 197, 427-436.
Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161-170.

Padgette, S.R., Huynh, Q.K., Borgmeyer, J., Shah, D.M., Brand, L.A., Biestre, D., Bishop, B.F., Rogers, S.G., Fraley, R.T. and Kishore, G.M. (1987) Arch. Biochem. Biophys. 258 (2), 564-578.

Padgette, S.R., Smith, C.E., Huynh, Q.K. and Kishore, G.M. (1988a) Arch. Biochem. Biophysics. 266 (1), 254-262.

Padgette, S.R., Huynh, Q.K., Ayken, S. Sammons, R.D., Sikorski, J.A. and Kishore, G.M. (1988b) J. Biol. Chem. 263 (4), 1798-1802.

Pang, P.P. and Meyerowitz, E.M. (1987) Bio/Technology 5, 1177-1181.

Pinto, J.E.B.P., Suzick, J.A. and Herrman, K.M. (1986) Plant Physiol. 82, 1040-1044.

Pinto, J.E.B.P., Dyer, W.E., Weller, S.G. and Herrmann, K.M. (1988) Plant Physiol. 87, 891-893.

Polley, L.D. (1978) Biochim. Biophys. Acta. 526, 259-266.

Pruitt, R.E. and Meyerowitz. E.M. (1986) J. Mol. Biol. 187, 169-183.

Ranjeva, R., Rafeno, G., Boudet, A.M. and Marme, D. (1983) Proc. Nat. Acad. Sci. U.S.A. 80, 5222-5224.

Risler, J.L., Delamore, M.O., Delacroix, H. and Henaut, A. (1988) J. Mol. Biol. 204, 1019-1029.

Rothe, G.M., Hengst, G., Mildenberger, I., Sharer, H. and Utesch. D. (1983) Planta 157, 358-366.

Rubin, J.L., Gaines, C.G. and Jensen, R.A. (1982) Plant. Physiol. 72, 833-839.

Rubin, J.L. and Jensen, R.A. (1985) Plant Physiol. 79, 711-718.

Sanderson, K.E. and Rothe, J.R. (1983) Microbial. Rev. 47, 410-533.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 5463-5467.

Sciolo, J.R. and Zilinskas, B.A. (1988) Proc. Nat. Acad. Sci. U.S.A. 85, 7661-7665.

Schmidt, G.W. and Mishkind, M.L. (1986) Annu. Rev. Biochem. 55, 879-912.

Shah, D.M., Horsch, R.B., Klee, H.J., Kishore, G.M., Winter, J.A., Tumer, N.E., Hironaka, C.M., Sanders, P.R., Gasser, C.S., Ayken, S.,

Siegel, N.R., Rogers, S.G. and Fraley, R.T. (1986) Science 233, 478-481.

Sharp, P.M., Tushy, T,M.F. and Mosurski, K.R. (1986) Nuc. Acid Res. 14, 5125-5143.

Shirras, A.D. and Northcote, D.H. (1984) Planta 162, 353-360.

Shultz, J., Hermodson, M.A., Garner, C.G. and Herrman, K.M. (1984) J. Biol. Chem. 259, 9655-9661.

Simoens, C.R., Gielen, J., Van Montagu , M. and Inze, D. (1988) Nuc. Acid Res. 16 (14), 6753-6766.

Singer, S.R. and McDaniel, C.N. (1985) Plant Physiol. 78, 411-416.

Singh, B.K., Connelly, J.A. and Conn, E.E. (1985) Arch. Biochem. Biophys. 243, 374-384.

Smart, C.C., Johanning, D., Muller, G. and Amrhein, N. (1985) J. Biol. Chem. 260 (30), 16338-16346.

Smart, C.C. and Amrhein, N. (1987) Planta 170, 1-6.

Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.

Stalker, D.M., Hiatt, W.R. and Comai, L. (1985) J. Biol. Chem. 260 (8), 4724-4728.

Steinruken, H.C. and Amrhein, N. (1984) Eur. J. Biochem. 143, 351-357.

Suzich, J.A. Ranjeva, R., Hasegawa, P.M. and Herrman, K.M. (1983) Plant Physiol. 75, 369-371.

Suzich, J.A., Dean, J.F.D. and Herrman, K.M. (1985) Plant Physiol. 79, 765-770.

Tabor, S. and Richardson, C.C. (1987) Proc. Nat. Acad. Sci. U.S.A. 84, 4767-4771.

Turner, S., Burger-Wiersma, T., Giovannoni, S.J., Mur, L.R. and Pace, N.R. (1989) Nature 337, 380-382.

Weiss, U. and Edwards, J.M. (1980) 'The Biosynthesis of Aromatic Compounds' Wiley, New York.

Wetmur, J.G. and Davidson. N. (1968) J. Mol. Biol. 31, 349-370.

White. P.J., Miller, G. and Coggins J.R. (1988) Biochem. J. 251, 313-322.

Williams, J.G. (1981) in 'Genetic Engineering' 1, (R. Williamson, ed) Academic Press, New York, U.S.A.

Wolfe, K.H. (1989) Plant Mol. Biol. Rep. 7 (1), 30-48.

Wood, W.B. (1966) J. Mol. Biol. 16, 118-133.

Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawherne, L. and Trieber, G. (1970) Virology 40, 734-744.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.

Young, R.A. and Davis, R.W. (1983) Proc. Nat. Acad. Sci. U.S.A. 80, 1194-1198.

Zurawski, G., Brown, K., Killingly, D. and Yanofsky, C. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 4271-4275.