Isolation of a vicilin gene from pea (pisum sativum l.), and nuclease sensitivity of seed storage protein genes in pea chromatin

Rosalind Mary Sawyer

ISOLATION OF A VICILIN GENE FROM PEA (*PISUM SATIVUM* L.), AND NUCLEASE SENSITIVITY OF SEED STORAGE PROTEIN GENES IN PEA CHROMATIN

A thesis submitted by:

ROSALIND MARY SAWYER, B.A.

In accordance with the requirements for the degree of Doctor of Philosophy in the University of Durham.

Department of Botany

January, 1986.

ISOLATION OF A VICILIN GENE FROM PEA (*PISUM SATIVUM* L.), AND NUCLEASE SENSITIVITY OF SEED STORAGE PROTEIN GENES IN PEA CHROMATIN.

by

ROSALIND MARY SAWYER

ABSTRACT

A library of pea genomic DNA in the bacteriophage vector EMBL3 was screened by hybridisation to cDNAs encoding vicilin, a major storage protein of pea (*Pisum sativum* L.) seeds. A vicilin gene, *vic A*, was isolated and characterised by restriction mapping and DNA sequencing. The nucleotide and predicted amino acid sequences of *vic A* were compared to those of vicilin cDNAs, and the gene was found to encode a 50,000 M_r non-glycosylated vicilin subunit that does not undergo post-translational proteolytic cleavage. The introns in *vic A* were typical of those in plant genes, being small and high in A+T content, and the nucleotide sequences at the splicesites showed good homology to the plant consensus. The positions of the introns in *vic A* were similar to those in a gene encoding a subunit of phaseolin, a related protein from French bean (*Phaseolus vulgaris*).

Methods were developed for the analysis of nuclease sensitivity of specific genes in pea chromatin. The DNAase I sensitivity of the seed storage protein genes was found to be greater in developing cotyledons, where the genes were transcriptionally active, than in leaves, where they were inactive. The pea ribosomal genes showed relative resistance to DNAase I in both tissues. The nucleosome repeat length, determined by digestion of chromatin with micrococcal nuclease, was similar in both tissues.

No evidence was obtained for DNAase I hypersensitive sites in pea chromatin. This result supports the findings of two other studies, and suggests that such sites are absent from plant chromatin.

ACKNOWLEDGEMENTS

I should like to thank my supervisor, Professor D. Boulter, for his advice and support during this work.

I am also very grateful to Dr. John Gatehouse for much encouragement and help whenever possible. Thanks are also due to Dr. Marta Evans, Dr. Anil Shirsat, who advised me during the screening of his gene library and demonstrated a number of techniques, and Dr. Ron Croy. Many of the technicians in the Botany Department were extremely helpful, in particular Mr. Russell Swinhoe, who showed me how to isolate plant nuclei, Mr. Paul Preston who introduced me to DNA sequencing and who sequenced part of the gene shown in Figure 13, and Mr. D. Bown. Mrs. Margaret Richards and Mrs. Christine Shaw were also very kind in introducing me to standard techniques in molecular biology.

A number of people helped in the preparation of this thesis; my thanks go to Professor Boulter and Dr. J. Gatehouse for critical reading of the manuscript, Mrs. Denise Hughes for figures, Dr. Andy Hepher for help with fluorescence microscopy, Mr. Paul Sidney for photography and Mrs. Marion Wilson for typing.

Finally, I should like to thank the S.E.R.C. for a Research Studentship, and my parents for their continued support of my education.

ABBREVIATIONS

All abbreviations used here are as in the 'Instructions to Authors', Biochemical Journal 225 (1985) 1-26, with the following exceptions:

bp = base pairs

kb = kilobase pairs

DNAase I = pancreatic deoxyribonuclease

RNAase = ribonuclease

DABA = 3′, 5′-diaminobenzoic acid

DAPI = 4,6-diamidino-2-phenylindole

EtBr = ethidium bromide

dNTP = deoxyribonucleoside triphosphate

ddNTP = dideoxyribonucleoside triphosphate

SDS = sodium dodecyl sulphate

SSC = saline sodium citrate

PEG = polyethylene glycol

A_{260} = absorbance at 260 nm

MNase = micrococcal nuclease

Xgal = 5 dibromo-4-chloro-3-indoylgalactoside

IPTG = isopropylthiogalactoside

5′ = 5′ terminal phosphate of a DNA or RNA molecule

3′ = 3′ terminal hydroxyl of a DNA or RNA molecule

PMSF = phenylmethylsulphonylfluoride

vic 50K = vicilin polypeptide of 50,000 M_r

vic 47K = vicilin polypeptide of 47,000 M_r

cDNA = complementary DNA

BSA = bovine serum albumin

c.p.m. = counts per minute

ABBREVIATIONS (contd.)

p.f.u. = plaque forming units

d.a.f. = days after flowering

poly(A^+)RNA = polyadenylated RNA

N-terminal = amino terminus of a peptide

C-terminal = carboxy terminus of a peptide.

DEP = diethylpyrocarbonate.

The one letter notation for amino-acids is given in the Biochemical Journal 219 (1984) 345-373.

CONTENTS

Page No.

FIGURES

FIGURES (contd.)

TABLES

CHAPTER ONE

INTRODUCTION

1.1 General Introduction

The elucidation of the regulatory mechanisms governing differential gene expression in eucaryotes is a major objective of modern molecular and developmental biology. The study of the control of gene expression in plants has been greatly facilitated by the advent of recombinant DNA technology, and the availability of plant transformation systems. The large genomes of most plants present difficulties in the isolation of specific genes, but molecular cloning has enabled the isolation of a number of these (Sorenson, 1984). The plant transformation system available using the Ti-plasmid of *Agrobacterium tumefaciens* (Barton and Chilton, 1983, Shaw *et al* 1983, Schell *et al*, 1984) has been used to study the regulatory sequences of plant promoters (Herrera-Estrella *et al*, 1984, Morelli *et al*, 1984, Broglie *et al*, 1984, Simpson *et al*, 1985), and to obtain developmentally regulated expression of plant genes (Sengupta-Gopalan *et al*, 1985). This system has also been proposed as a means of achieving the genetic manipulation of crop plants (Shewry *et al*, 1981, Miflin and Lea, 1984), to overcome some of the problems, such as inter-specific barriers and lack of genetic variation in inbred lines, which affect crop improvement by conventional plant-breeding methods (Shewry *et al*, 1981). It is envisaged however, that genetic manipulation will augment, rather than supersede, these methods (Miflin and Lea, 1984). One of the major drawbacks to the use of *Agrobacterium* plant transformation systems in crop improvement is that the host-range of the bacterium was, until recently, thought to be limited to dicotyledonous plants, while many of the important crop plants are

cereals. Evidence has been provided for the infection of monocotyledons by *Agrobacterium* (Drummond, 1984, Hooykaas-Van Slogteren *et al*, 1984), but it is not known whether the foreign DNA is integrated into the plant genome (Jones, 1985). However, the technique of direct gene transfer into protoplasts has proved successful (Jones, 1985) in two species of monocotyledonous plants, both of which are in the Gramineae, as are the cereal crop plants.

The problem of the introduction of foreign DNA is not the only one in achieving the transformation of plants. A suitable system for regeneration of whole, fertile plants, which have the foreign DNA stably inserted into the genome, is required. Such systems are available for some dicots, in particular tobacco, *Nicotiana tabacum*, and are being developed for monocots (Flavell and Mathias, 1984).

In order for genetic manipulation to achieve the desired results in crop improvement, it is important that not only the problems above be overcome but that

> 'the mechanisms employed should be based on good knowledge of underlying biochemical mechanisms and genetic controls'
>
> (Shewry *et al*, 1981).

Research into the molecular biology of plant genes, their structures, functions and regulation is therefore a prerequisite for the genetic engineering of plants, as well as providing an increased understanding of control mechanisms during plant growth and development.

1.2 Molecular biology of plant genes

Many of the plant genes which have been isolated by molecular cloning have been those encoding the seed storage proteins of cereals or legumes (Sorenson, 1984, Higgins, 1984). The seeds of these plants are an important source of dietary protein for human and animal populations (Shewry *et al*, 1981). The high levels of amide amino acids and arginine in these proteins, consistent with their role as a nitrogen store for the germinating seed (Spencer, 1984), results in a deficiency in other amino acids. The seed proteins of cereals usually have a low lysine content, although some may be deficient in threonine or tryptophan; while those of legumes are generally low in cysteine and methionine (Shewry *et al*, 1981, Higgins, 1984). Improvement in the nutritional quality of these proteins by genetic manipulation (Shewry *et al*, 1981, Croy and Gatehouse, 1985) would reduce the need for the other, more expensive, sources of essential amino acids (Shewry *et al*, 1981).

Economic and nutritional reasons are not however the only ones for studying seed storage protein genes. Storage protein synthesis occurs only in a particular tissue, the endosperm of cereal seeds (Miflin *et al*, 1981, Higgins, 1984) or the cotyledons and embryonic axis of legume seeds (Millerd, 1975), and is under strict developmental control. The storage proteins may represent as much as 70% of the total seed protein in legumes (Shewry *et al*, 1981) and the various fractions of these proteins accumulate at different, specific times during the cell expansion phase of development in these seeds (Boulter, 1981). Thus the synthesis of seed storage

proteins in seeds provides an example of developmentally regulated, tissue-specific gene expression in plants, and has been compared (Boulter, 1981) to analogous systems such as haemoglobin synthesis during erythropoiesis in animals (O'Malley *et al*, 1977).

1.3 Storage protein synthesis and regulation in pea (*Pisum sativum* L.)

1.3.1 Regulation of synthesis

As mentioned above, storage protein synthesis in seeds is under strict developmental control. In pea, as in other legumes, there is a rapid accumulation of protein during the cell expansion phase of seed development (Boulter, 1981).

The three major seed storage proteins of pea, vicilin, legumin and convicilin (Section 1.4), are synthesised in this order (Boulter, 1981), and protein accumulation is essentially complete by 22 d.a.f. (Boulter, 1981, Gatehouse *et al*, 1982a). During this cell expansion phase, the genomes of the cotyledon parenchyma cells undergo extensive endoreduplication, resulting in increases in nuclear DNA content of up to 5OC in *Pisum sativum* (Millerd and Spencer, 1974) or 64C in *Pisum arvense* (Smith, 1973). The additional DNA above the 2C level does not appear to act as a template for RNA synthesis (Millerd and Spencer, 1974).

Translation *in vitro* of polysomes and RNA isolated from pea cotyledons at various times after flowering

shows that the mRNA levels for the various polypeptides parallel the accumulation of these proteins (Evans *et al*, 1979, Gatehouse *et al*, 1982a). Isolation of total RNA and hybridisation to cDNAs encoding the seed storage proteins (Section 1.4) gives similar results (Gatehouse *et al*, 1982a), and also shows that the level of seed storage protein mRNA in leaf tissue is below the limits of detection of this assay, *i.e.* less than 0.01% of the level of total poly(A^+)RNA (Gatehouse *et al*, 1982a). 'Run-off' transcription from intact cotyledon nuclei (Evans *et al*, 1984), isolated at various stages of development, shows that the pattern of transcripts is similar to that of protein accumulation. No transcripts coding for these proteins were detected in leaf nuclei (Evans *et al*, 1984). These results suggest that the regulation of pea seed storage protein synthesis is primarily at the transcriptional level, although the relatively long half-life of these mRNAs (Gatehouse *et al*, 1982a) suggests that continued protein synthesis in the latter part of development is dependent upon mRNA stability rather than transcription (Gatehouse *et al*, 1984).

1.3.2 Intracellular sites of synthesis

The seed storage protein precursors of pea (Higgins and Spencer, 1980, and Section 1.4) are synthesised by polysomes on the endoplasmic reticulum (Chrispeels *et al*, 1982a) and are transported to the Golgi apparatus and thence to the protein bodies *via* electron-dense vesicles (reviewed in Chrispeels, 1984). The polypeptides are assembled into oligomers in the lumen of the endoplasmic reticulum and

glycosylation, where appropriate, also occurs here (Chrispeels *et al*, 1982a). The post-translational proteolysis of the precursors (Section 1.4) occurs in the protein-bodies (Chrispeels *et al*, 1982b). Storage proteins are thus usually regarded as those proteins which are synthesised to very high levels during seed development, deposited in protein bodies and hydrolysed rapidly on germination to provide a source of reduced nitrogen (Boulter, 1981, Spencer,1984). This definition excludes other seed proteins which may be synthesised at moderately high levels during embryogenesis, such as the major (Croy *et al*, 1984) and minor (Gatehouse *et al*, 1985) albumins of pea. These proteins are not deposited within protein bodies and are not hydrolysed rapidly during germination (Gatehouse *et al*, 1985), although previous reports showed a rapid decrease in the albumin proteins of pea during germination (Murray, 1979). Whether or not these proteins have a storage function, their amino acid compositions are often more balanced than those of the fractions generally regarded as seed storage proteins (Shewry *et al*, 1981). Thus, increases in the levels of these proteins may be important in improving the nutritional quality of the seed (Shewry *et al*, 1981).

1.4 Structure of pea seed storage proteins and their genes

1.4.1 Legumin

Legumin is the name given to the 11S storage protein fraction of peas and broad beans (Croy *et al*, 1979, Wright and Boulter, 1974), although these proteins occur in other legumes, *e.g.* glycinin in soyabean (Nielsen, 1984), and

in cereals, *e.g.* a glutelin of rice (Zhao *et al*, 1983). The homologies between these legumin-type proteins (Derbyshire *et al*, 1976, Gilroy *et al*, 1979) have been established, and have led to calls for an improved nomenclature to reflect these relationships (Croy and Gatehouse, 1985).

Legumin is a hexameric molecule of $M_r \simeq 360,000$. Each monomer consists of an acidic subunit ($M_r \simeq 40,000$) covalently linked *via* disulphide bonds to a basic subunit ($M_r \simeq 20,000$) (Wright and Boulter, 1976, Gilroy *et al*, 1979, Croy *et al*, 1979). In pea, both the acidic and basic subunits show heterogeneity in charge and molecular weight (Gatehouse *et al*, 1980). Legumin is initially synthesised in peas as a family of precursors of $M_r \simeq 60$-$65,000$ (Spencer and Higgins, 1980, Croy *et al*, 1980a, Matta *et al*, 1981), representing the monomeric form of the protein. The various subunit pairs have been divided into 'major' and 'minor' legumin species (Casey *et al*, 1981, Matta *et al*, 1981), with the latter sub-divided into 'big' and 'small' legumins (Matta *et al*, 1981). The $\simeq 60,000$ M_r precursors are cleaved post-translationally to form the acidic and basic subunits (Spencer and Higgins, 1981, Croy *et al*, 1980a). An $80,000$ M_r legumin precursor has also been identified (Domoney and Casey, 1984).

A number of cDNAs encoding legumin sequences have been made from poly(A^+) RNA isolated from developing pea cotyledons (Croy *et al*, 1982, Lycett *et al*, 1983b). The sequences of these cDNAs have confirmed that legumin is synthesised as a $\simeq 60,000$ M_r precursor containing both acidic and basic subunits (Croy *et al*, 1982). These cDNAs have been used to screen

libraries of pea genomic DNA in bacteriophage vectors, and a number of legumin genes have been isolated (Shirsat, 1984). One of these genes, *leg A*, has been fully sequenced (Lycett *et al*, 1984). The coding region is interrupted by 3 introns, each of which begin with GT and end with AG (see Breathnach *et al*, 1978), and whose boundaries are similar to the consensus sequence for plants (Slightom *et al*, 1983). The positions of the introns are similar to those of glycinin (Nielsen *et al*, 1984) but the introns are shorter (Lycett *et al*, 1984). The A+T content of the introns is high, as has been found for a number of plant genes (Slightom *et al*, 1983).

The 5´ region of the *leg A* gene shows that there is a sequence upstream from the N-terminus of the mature protein which encodes a 21 amino-acid peptide rich in hydrophobic residues (Lycett *et al*, 1984). This is believed to be a signal peptide (Blobel and Dobberstein, 1975), consistent with the synthesis of legumin on the rough endoplasmic reticulum (Chrispeels *et al* 1982a). A number of sequences implicated in transcriptional regulation are found upstream of the translation initiation codon (Lycett *et al*, 1984). A 'TATA box', shown to be important in determining the site of transcriptional initiation in many eucaryotic genes (Mathis and Chambon, 1981, Benoist and Chambon, 1981, Breathnach and Chambon, 1981), is present 66 bp upstream of this codon, and the start site of transcription has been shown, by S1 nuclease mapping, to be centred on a sequence CATC 25 bp downstream of the TATA box (Lycett *et al*, 1984). A 'CAAT box', found upstream of many eucaryotic genes (Benoist *et al*, 1980), is present in *leg A* at -126 (+1 is the start of translation) and

the sequences on the 5´ side of this box show partial homology to the 'AGGA' box found upstream of a number of plant genes (Messing *et al*, 1983). Two sequences showing partial homology to viral enhancer elements are also present in the 5´ flanking regions of *leg A* (Lycett *et al*, 1984).

The 3´ untranslated region of *leg A* shows good homology to the 3´ ends of previously isolated legumin cDNAs (Croy *et al*, 1982, Delauney, 1984). There are 3 polyadenylation signals within this region; in all the legumin cDNAs, polyadenylation occurs 19-20 bp downstream of the second site (Lycett *et al*, 1983b). The sequences around this site may be important in the potential secondary structure of the message (Lycett *et al*, 1984).

The presence of 5´ and 3´ regulatory sequences in *leg A*, and its identical sequence to that of the cDNA pDUB8, suggest that this gene is expressed *in vivo*. This has been confirmed by *in vitro* experiments, in which the *leg A* gene was transcribed, albeit poorly, in an extract of HeLa cell nuclei (Evans *et al*, 1985). Transcription of various restriction fragments showed that the TATAbox was required for initiation at the correct site, but sequences 5´ to -97 bp did not appear to be required (Evans *et al*, 1985), at least in this system.

A legumin pseudogene, ψ *leg D*, has also been characterised (Bown, *et al*, 1985). The sequence shows partial homology to *leg A*, and is located 1.3 kb from the 3´ end of *leg A* on the recombinant bacteriophage λleg 1 (Bown *et al*, 1985). Compared to *leg A*, there are a number of frame-shift errors and deletions in ψ *leg D*, and the

presence of stop-codons at codons 6 and 29 suggest that the gene is non-functional; this is supported by the alterations to the promoter sequences and absence of homology to *leg A* upstream of -145 bp (Bown *et al*, 1985).

Comparison of the 5' sequences of leg A with those of two other legumin genes, leg B and leg C (Lycett *et al*, 1985) shows that all three genes are homologous until about -300 bp, when *leg A* begins to diverge. (In this report the start of translation was taken as +1). A sequence present at -405 to -413 in all the genes shows close resemblance to a consensus sequence found in a number of seed specific genes (R. Goldberg and T. Sims, unpublished results, cited in Lycett *et al*, 1985). *Leg B* and *leg C* remain homologous for a further 550 bp after the divergence of *leg A*. At the point where these two genes begin to differ, there is a sequence in *leg C* which is homologous to a number of transposable elements. It has been suggested that insertion of such an element may be responsible for the difference and may play a role in generating sequence diversity (Lycett *et al*, 1985), in the legumin genes.

The number of legumin genes in pea has been variously estimated to be 4 (Croy *et al*, 1982), 7 (Shirsat, 1984) or 8 (Domoney and Casey, 1985). The major legumins are encoded by a small number of genes which are closely linked (Casey and Domoney, 1984), while the minor legumins (Matta *et al*, 1981, Casey *et al*, 1981) are encoded by other genes (Casey and Domoney, 1984). The occurrence of multigene families encoding legumin polypeptides, and the resultant charge and size heterogeneity, suggests that small changes

in primary structure can be tolerated without affecting protein function. However, the strong homology between 11S storage proteins in different species, *e.g.* glycinin of soyabean and legumin of pea (Casey *et al*, 1981, Negoro *et al*, 1985) suggests that these genes are derived from a common ancestral sequence, and that structural constraints in the proteins have prevented rapid divergence of sequences.

1.4.2 Vicilin

The vicilin (7S) fraction of pea seed storage proteins, like legumin, has been found to be heterogeneous (Thomson *et al*, 1978, Gatehouse *et al*, 1981). Vicilin isolated from mature pea cotyledons gives a range of polypeptides which show considerable size and charge heterogeneity (Gatehouse *et al*, 1981). *In vitro* translation of polysomal RNA from developing pea cotyledons gives only two vicilin polypeptides, one of 50,000 M_r and one of 47,000 M_r (Croy *et al*, 1980b). The 47,000 M_r polypeptide is not seen in vicilin isolated from mature seeds and has been suggested to be a precursor of the smaller vicilin polypeptides (Croy *et al*, 1980b, Gatehouse *et al*, 1981), which are produced from this subunit as a result of post-translational proteolysis. This model has been confirmed by comparison of the partial amino-acid sequences obtained for various vicilin polypeptides with those predicted by a number of vicilin cDNAs (Gatehouse *et al*, 1982b, 1983), as described later in this section.

Vicilin is a protein of M_r ≈150,000 and a trimeric structure, consisting of subunits of ≈50,000 M_r, has been proposed (Gatehouse *et al*, 1981). The smaller vicilin polypep-

tides produced as a result of endo-proteolytic cleavage of some of these precursors (see above) remain associated in ≈50,000 Mr subunits (Gatehouse *et al*, 1981). A similar trimeric structure has been described for the related 7S protein of *Phaseolus vulgaris*, phaseolin (Puztai and Stewart, 1980).

Vicilin is glycosylated (Derbyshire *et al*, 1976), unlike legumin which does not contain carbohydrate (Gatehouse *et al*, 1980). The site of glycosylation appears to be on the 12,500 Mr polypeptide (Davey *et al*, 1981), and glycosylation does not seem to be essential as its inhibition by tunicamycin did not prevent the synthesis and transport of vicilin (Chrispeels *et al*, 1982a, b). The appearance of polypeptides of the correct sizes for vicilin and convicilin (Section 1.4.3) when microsomal membranes from dog pancreas (Higgins and Spencer, 1981) or pea cotyledons (Gatehouse *et al*, 1981) were added to an *in vitro* translation system containing polysomal RNA from pea cotyledons confirms that these proteins are synthesised on membrane bound ribosomes (Section 1.3.2).

A number of cDNAs which encode pea vicilin sequences have been isolated (Croy *et al*, 1982, Lycett *et al*, 1983a, Delauney, 1984). These fall into two classes, those encoding a 47,000 M_r vicilin precursor, *e.g.* pDUB4(Croy *et al*, 1982) and pDUB7 (Lycett *et al*, 1983a, b), and those encoding 50,000 M_r precursors, *e.g.* pDUB2 (Croy *et al*, 1982) and pDUB9 (Delauney, 1984). These two classes of cDNAs show about 86% homology in their nucleotide sequences and most of the changes result in no difference in the predicted amino acid sequence,

or in conservative substitutions (Lycett *et al* 1983a). This reflects the relatively wide codon-usage in plant genes (Lycett *et al*, 1983b).

Comparison of the amino acid sequences predicted by the two classes of cDNAs with those of various vicilin polypeptides (Lycett *et al*, 1983a, Gatehouse *et al*, 1982b, 1983) confirmed that the smaller vicilin polypeptides were derived from the 47,000 M_r precursor. Cleavage of this subunit results in the 33,000 Mr ($\alpha+\beta$) and 12,500 M_r (or 16,000 if glycosylated, γ) polypeptides (Gatehouse *et al*, 1982b, Lycett *et al*, 1983a). The 50,000 M_r precursors do not undergo post-translational cleavage (Gatehouse *et al*, 1982b). A third class of polypeptides, cleaved at both the α/β and β/γ sites, results in the production of the α (19,000 M_r), β (13,500 M_r) and γ subunits (Gatehouse *et al*, 1982b, 1983) but no message corresponding to this polypeptide has been found (Lycett *et al*, 1983a). A fourth type of vicilin polypeptide, cleaved only at the α/β site, has also been proposed (Lycett *et al*, 1983a).

The amino acid sequences predicted from the cDNA sequences for the 50,000 M_r and 47,000 M_r polypeptides show considerable divergence around the β/γ cleavage site (Gatehouse *et al*, 1983, Lycett *et al*, 1983a). There are two deletions, one each side of this site, in the 47,000 M_r sequence (Lycett *et al*, 1983a). These sites are also regions of considerable divergence between the sequences of vicilin and those of phaseolin and α-conglycinin, a 7S protein of soyabean (Schuler *et al*, 1982a, b), although most of the remaining sequences show good homology for all three.

The two cDNAs encoding 50,000 M_r polypeptides, pDUB2 (Croy *et al*, 1982) and pDUB9 (Delauney, 1984) show some differences in the predicted amino acid sequences (Delauney, 1984) and in their restriction maps, suggesting that they may encode different polypeptides. This, together with the observations that at least 4 fragments in EcoR1 digests of pea genomic DNA hybridise to the insert of pDUB2, and pDUB4 hybridises to a different fragment at an intensity equivalent to 2-3 gene copies (Gatehouse *et al*, 1983) shows that each class of vicilin polypeptides is encoded by a number of genes (Lycett *et al*, 1983a). The various polypeptides of the related 7S proteins phaseolin and conglycinin have also been shown to be encoded by multigene families (Slightom *et al*, 1985, Schuler *et al*, 1982a, b).

A complete sequence for a vicilin precursor 'preprovicilin', has been compiled (Lycett *et al*, 1983a). The predicted amino acid sequence shows that there are 15 residues upstream from the N-terminus of the mature protein. The sequence is rich in hydrophobic residues, and is thought to be a signal sequence (Lycett *et al*, 1983a). The methionine codon (AUG) at the start of the signal sequence is thought to be the point of initiation of translation (Lycett *et al*, 1983a), although a 27 amino-acid signal sequence, containing a number of methionine residues, has been reported for a pea vicilin cDNA (Chandler, P.M., Arrifin, Z and Blagrove, R.J., 1984, cited in Spencer, 1984), and the complete sequence of a phaseolin gene suggests a signal peptide of 21-26 amino-acids in length (Slightom *et al*, 1983).

The C-terminus sequence predicted from the cDNAs did not agree with that determined for the mature protein,

(Lycett *et al*, 1983a) and the sequences predicted by the cDNAs pDUB2 and pDUB9 also differed (Delauney, 1984). This suggests the post-translational removal of a C-terminal peptide (Lycett *et al*, 1983a, Delauney, 1984). Despite the different C-termini, the two tandem stop codons in pDUB2 (Lycett *et al*, 1983a) are also present in pDUB9 (Delauney, 1984) although a single stop codon is found at the end of the coding sequence. The pDUB9 3´ non-coding sequence contains two overlapping polyadenylation signals, as found in a number of plant genes (Lycett *et al*, 1983b), while pDUB2 has a single polyadenylation signal, unusual in a plant gene (Lycett *et al* 1983b, Messing *et al*, 1983). However, as neither of these cDNAs contained poly(A) tails, there may have been additional polyadenylation signals downstream (Delauney, 1984).

As yet there have been no reports of genomic vicilin sequences for pea, although a complete phaseolin gene has been isolated and characterised (Slightom *et al*, 1983). A gene encoding a subunit of the 7S storage protein of soyabean has also been isolated (Schuler *et al*, 1982a). Isolation of a pea vicilin gene would extend the information available on the multigene families encoding vicilin polypeptides and provide more information about sequence relationships between pea vicilin and related proteins in other species.

1.4.3 Convicilin

When the vicilin fraction of peas is examined by SDS/PAGE, a ≃71,000 M_r polypeptide, immuno-precipitable by anti-vicilin antibodies is seen in addition to the vicilin polypeptides described in Section 1.4.2 (Gatehouse *et al*, 1981,

Croy *et al*, 1980c). This polypeptide is a subunit of convicilin, a tetrameric protein of about 280-290,000 M_r (Croy *et al*, 1980c), which, unlike vicilin, is not glycosylated and can be separated from vicilin by non-dissociating techniques (Croy *et al*, 1980c, Gatehouse *et al*, 1981). Convicilin accumulates later during seed development than vicilin (Croy *et al*, 1980c, Gatehouse *et al*, 1984). A cDNA encoding convicilin has been isolated (Domoney and Casey, 1983) and shown to hybrid-select convicilin mRNA, but not vicilin mRNA. However, the precipitation of convicilin with anti-vicilin antibodies suggests that there is sufficient sequence homology between the proteins for similar antigenic determinants to be present. Analysis of the nucleotide sequence of a convicilin cDNA and its predicted amino acid sequence shows it to be partly homologous to vicilin, phaseolin and conglycinin sequences (Casey *et al*, 1984) ; the homology is greater to vicilin than to the other two genes, suggesting that divergence of these from the pea 7S proteins occurred before the divergence of convicilin and vicilin (Casey *et al*, 1984).

1.4.4 Summary of features of pea seed storage proteins and their genes

The previous sections have shown that the major seed storage proteins of pea are synthesised as precursors which subsequently undergo a variety of post-translational cleavages (Chrispeels *et al*, 1982b, 1984, Gatehouse *et al*, 1984). The various storage proteins are synthesised at different times during the cell expansion phase of seed development (Boulter, 1981, Gatehouse *et al*, 1982a, Gatehouse *et al*, 1984) and the control of synthesis seems to be primarily at the level of transcription (Gatehouse *et al*, 1982a, Evans *et al*,

1984). The cDNAs and genes which have been isolated show that the precursors are encoded by multigene families and contain a signal sequence (Blobel and Dobberstein, 1975), consistent with their synthesis on membrane-bound ribosomes (Chrispeels *et al*, 1982a). The 5´ regions of the genes contain a number of regulatory sequences, including some which are plant-specific (Messing *et al*, 1983) while some of the 3´ regions show multiple polyadenylation sequences (Lycett *et al*, 1983b, 1984, Delauney, 1984) which are thought to be characteristic of plant messages (Messing *et al*, 1983). The introns in *leg A* are small and have a high A+T content (Lycett *et al*, 1984) as has been found for a number of plant introns (Slightom *et al*, 1983).

Thus the genes encoding pea seed storage proteins have a similar structure to those of most eucaryotes (Breathnach and Chambon, 1981), and also contain some plant specific sequences (Messing *et al*, 1983). As yet there is little information on the function of the various regulatory sequences; most of these are inferred by analogy with other genes. Studies of more sequences encoding pea seed storage proteins are therefore required in order to establish whether particular regions confer tissue-specificity or are otherwise involved in regulation.

1.5 Chromatin structure and the control of gene expression

1.5.1 Possible roles of chromatin structure in gene regulation

There is much evidence that the primary regulation of gene expression in plants, as in other eucaryotes, acts at the level of transcription. The presence of 5´ regulatory sequences implicated in the control of transcription (Messing *et al*, 1983) suggests that recognition of these sequences by particular factors, such as proteins, is important. As these sequences are present whatever the functional status of the gene, differential expression must depend on processes affecting this recognition. This might be achieved by the presence of different proteins which recognise these sequences in various tissues; those in the tissue(s) in which the genes are transcribed being activating proteins, and those in other tissues being repressors. The accessibility of the chromatin template to recognition factors might thus be a further important element in the regulation of gene expression.

The binding of factors involved in the formation of a potentially active transcription complex has also been suggested as a mechanism of 'marking' genes during cellular determination (Brown, 1984). It is proposed that once formed, such complexes can be stably propagated in the presence of the binding factors (Brown, 1984). If, however, DNA is replicated in the absence of such factors, a general repressor mechanism, probably involving histone H1 (Weintraub, 1984, 1985) serves to package the gene into

an inactive structure. Some aspects of chromatin structure have the property of self-propagation (Weintraub, 1985, Section 1.9) and are necessary, but not sufficient, for transcription. Such structures may therefore represent the stable complexes mentioned above. Investigations into the nature of transcriptionally active, or potentially active, chromatin are therefore an important part of studies on the control of gene expression.

1.5.2 Chromatin structure

The structure of chromatin has been the subject of recent reviews (Igo-Kemenes *et al*, 1982, Reeves, 1984, Conklin and Groudine, 1984) and therefore will be described here only briefly. The DNA helix is packaged into nucleosomes (Kornberg, 1974); octamers containing two each of the four 'core' histones H2A, H2B, H3 and H4, associated with 146 bp of DNA (Wang, 1982). The DNA is wrapped around the outside of the nucleosome in a left-handed superhelix with approximately 2 turns per nucleosome (Finch and Klug, 1977, Lutter, 1977). The length of DNA between the nucleosomes, the linker DNA, is variable, and is often associated with histone H1 (Noll and Kornberg, 1977). The fibre resulting from this packaging has a diameter of approximately 10nm (Reeves, 1984, Conklin and Groudine, 1984). A second level of packaging is achieved by the coiling of this fibre into a 30 nm fibre which is now believed to form a solenoid structure (Worcel, 1977, Thoma *et al*, 1979, McGhee *et al*, 1980, Butler, 1984) with about 6 nucleosomes/turn. A further hierarchical level of structure is the

formation of loops or domains of chromatin (Igo-Kemenes and Zachau, 1977, Vogelstein *et al*, 1980), attached at the base to a proteinaceous matrix (Lewis and Laemmli, 1982, reviewed in Reeves, 1984, Conklin and Groudine, 1984).

The initial models of chromatin structure were based on data obtained from animals; however, the available information (Spiker, 1984, 1985) suggests that the chromatin of plants shows a very similar structure. This is to be expected in view of the strong conservation of amino-acid sequences of histones (Patthy *et al*, 1973) and of their binding sites (Spiker and Isenberg, 1977) and the requirements for packaging of the large genomes of higher eucaryotes into discrete chromosomes within the nucleus. The packaging of DNA into chromatin does, however, present problems of accessibility to enzymes and other factors involved in replication and transcription. Many of the studies on transcriptionally active chromatin have used various nucleases to determine the accessibility of the DNA, and these experiments are described in the following sections.

1.6 Micrococcal nuclease sensitivity and nucleosomal structure of chromatin

Micrococcal nuclease cleaves the linker DNA between nucleosomes so that mild digestion of chromatin with this enzyme results in the generation of a mixture of cleavage products. When the DNA from these molecules is purified and electrophoresed, the lengths of the fragments are multiples of the length of DNA associated with a nucleosome

monomer. This enzyme has been used to show that many active genes in chromatin have a nucleosomal conformation (Reeves, 1984, Conklin and Groudine, 1984). A number of studies have also shown that mild digestion of chromatin with micrococcal nuclease results in the release of active sequences, although other results have shown no such preferential release (Conklin and Groudine, 1984). It has been suggested that the conflicting results were obtained because the sensitivity was dependent on a disruption of structure during transcription itself (Conklin and Groudine, 1984, Reeves, 1984). This suggestion is supported by the finding that the nucleosomal pattern of some genes becomes blurred or lost on activation, for example the α histone genes of sea-urchins (Spinelli *et al*, 1982) show a nucleosomal structure similar to that of bulk chromatin in sperm or mesenchyme bastula cells, which do not express these genes. In 32-64 cell embryos however, where these genes are active, the nucleosomal structure of these sequences is lost. The bulk chromatin in these embryos still shows a typical nucleosomal 'ladder', although with an altered repeat length (Spinelli *et al*, 1982). Taken together, the results of such studies show that most active genes remain associated with nucleosomes, but that the structure may be greatly altered during transcription itself (Reeves, 1984). In addition, small regions of chromatin may be nucleosome-free, as described later (Section 1.9, Conklin and Groudine, 1984).

The use of micrococcal nuclease to detect chromatin-specific structures has been somewhat limited by its sequence-specificity (Bernardi *et al*, 1975), which results in a preference for particular sites on protein-free DNA (Keene and

Elgin, 1981). However, digestions with this enzyme on protein-free DNA can be used to distinguish between sequence-specific and chromatin-specific cuts (Worcel *et al*, 1983) and micrococcal nuclease is still widely used to demonstrate alterations in chromatin structure on gene activation (*e.g.* Udvardy *et al*, 1985).

The problems of nucleosome phasing, *i.e.* the regular placement of nucleosomes relative to a repeating DNA sequence, and of positioning, *i.e.* the unique locations of nucleosomes on non-repetitive DNA (Reeves, 1984), are other features of chromatin structure for which micrococcal nuclease has been used as a probe. Chromatin is fractionated into nucleosomes by the enzyme and their location along a segment of DNA mapped by a variety of techniques (Igo-Kemenes *et al*, 1982). However, the sequence-specificity of the enzyme, as mentioned above, necessitates careful controls with protein-free DNA, which have not always been performed (Igo-Kemenes *et al*, 1982, Reeves, 1984). There are many conflicting results and much controversy over whether phasing and positioning are random or not (Igo-Kemenes *et al*, 1982, Reeves, 1984) and this feature of chromatin structure will not be further described here.

1.7 DNAase I sensitivity of transcriptionally active chromatin

A large number of studies have shown that active genes in chromatin are more sensitive to DNAase I than are inactive genes. The globin genes in chicken erythrocyte nuclei are digested more rapidly than are the inactive ovalbumin genes

(Weintraub and Groudine, 1976, Stalder *et al*, 1980a). This preferential sensitivity of the globin genes is not found in fibroblast nuclei, in which they are inactive (Weintraub and Groudine, 1976). Similar results have been obtained for the ovalbumin genes in oviduct nuclei (Garel and Axel, 1976), the heat-shock genes of *Drosophila melanogaster* (Wu *et al*, 1979b), the ribosomal genes of *Physarum polycephalum* (Stalder *et al*, 1978 and a number of other genes (Weisbrod, 1982, Igo-Kemenes *et al*, 1982, Conklin and Groudine, 1984, Reeves, 1984). Many of the experiments use solution hybridisation of DNAase-I resistant DNA or of total DNA to various cDNAs; the kinetics of reassociation indicate whether the sequences corresponding to the cDNA have been preferentially digested. A second method is to electrophorese the DNA through an agarose gel, blot it onto nitrocellulose or a nylon membrane, and hybridise to a specific probe. The differences in hybridisation at various levels of digestion between probes for active and inactive genes can be used to determine the sensitivity of the active sequences. There are few examples of the DNAase I sensitivity of specific genes in plants (Spiker, 1984, 1985). The renaturation of cDNA from wheat embryos with the resistant DNA purified from DNAase I treated wheat embryo nuclei is 40% slower than that with total DNA, indicating a preferential digestion of transcriptionally active sequences (Spiker *et al*, 1983). In a study of the gene encoding β-phaseolin, a subunit of the 7S seed storage protein of French bean, *Phaseolus vulgaris*, it was found that the active genes in cotyledons were more sensitive to DNAase I than were the inactive genes in leaves

(Murray and Kennard, 1984). Thus it appears that an increased DNAase I sensitivity is a general feature of transcriptionally active genes in animals and plants. However, in yeast there appears to be no difference in the DNAase I sensitivities of active and inactive genes (Lohr and Hereford, 1979, Lohr, 1983), although increased sensitivity on gene activation has also been reported (*e.g.* Bergman and Kramer, 1983).

The structural basis of this preferential sensitivity is usually regarded as an unfolding of higher-order chromatin structures, such as the 30 nm solenoid fibre (Section 1.5.2), in active genes (Conklin and Groudine, 1984, Kimura *et al*, 1983). However, changes in the structure or composition of the nucleosomes associated with active genes are suggested by the DNAase I sensitivity of the globin genes in nucleosome monomers from chicken erythrocyte nuclei (Weintraub and Groudine, 1976). This is supported by the restoration of DNAase I sensitivity to nucleosome monomers when HMG proteins 14 and 17, which are associated with active sequences and DNAase I sensitivity (Section 1.11, Weisbrod *et al*, 1980), are added. Thus the structural changes which result in the increased DNAase I sensitivity of active genes are not yet fully understood.

The increased DNAase I sensitivity may affect sequences outside of the transcription unit. The ovalbumin multigene family in oviduct nuclei is in a 100 kb DNAase I sensitive region, while the sensitivity of the chicken β-globin gene cluster extends about 7 kb 5′ and 8 kb 3′ to the genes (Stalder *et al*, 1980b). These large regions of sensitivity

are often referred to as domains, although this does not imply that they are equivalent to the domains or loops of chromatin previously mentioned (Section 1.5.2).

Unlike the alterations in chromatin structure detected by micrococcal nuclease (Section 1.6), those detected by DNAase I appear to be independent of transcription. The adult chicken β-globin gene is sensitive to DNAase I in embryonic red cells, in which it is not expressed (Stalder *et al*, 1980a); the gene becomes active in adult erythrocytes. The observation that frequently transcribed sequences show a similar level of DNAase I sensitivity to that of rarely transcribed genes (Garel *et al*, 1977) also demonstrates that the presence of transcriptional complexes is not required for sensitivity. Thus DNAase I sensitivity is a property of potentially, as well as actually, transcriptionally active genes.

1.8 DNAase I hypersensitive sites in chromatin

Specific fragments are observed when DNA purified from *Drosophila melanogaster* embryo nuclei, previously treated with DNAase I, is electrophoresed, blotted onto nitrocellulose and hybridised to ^{32}P-labelled probes (Wu, 1979a). A unique pattern of fragments is obtained for each probe, and is not observed with protein-free DNA. Restriction of the DNA from treated nuclei prior to electrophoreses, followed by hybridisation to a short probe abutting a chosen restriction site, allows mapping of the DNAase I cleavage sites (Wu, 1980).

The sites represent double-stranded cleavages at very low levels of DNAase I, and are known as DNAase I hypersensitive sites. The fragments resulting from such a cleavage within a restriction fragment appear as additional sub-bands (Wu, 1980) which are not present in protein-free DNA or DNA from control, undigested nuclei.

DNAase I hypersensitive sites have been found in a number of genes (Elgin, 1981), and are often, although not always, associated with the 5′ and 3′ flanking regions. In view of this, they have been implicated in the regulation of gene expression (Elgin, 1981). The majority of the experimental evidence available supports this hypothesis, although much of it shows only a correlation between presence of these sites and gene activation, potential or actual.

Hypersensitive sites are often present before gene activation (Weintraub and Groudine, 1982, Burch and Weintraub, 1983, Fritton *et al*, 1984), indicating that transcription is not required for their formation. This is also shown by the persistence of the sites in the absence of inducing factors (Groudine and Weintraub, 1981). These properties suggest (Elgin, 1981, Weintraub, 1985), that the sites may be involved in 'marking' potentially active genes during cellular differentiation (Section 1.5.1, Brown, 1984). Most studies on hypersensitive sites support this hypothesis as many have been shown to be tissue-specific (Stalder *et al*, 1980b, Shermoen and Beckendorf, 1982). However, the persistence of some sites, but loss of others, after the cessation of transcription (Burch and Weintraub, 1983) and the presence of subsets of hypersensitive sites depending on the functional

status of the gene (Fritton *et al*, 1984), have led to suggestions that there are arrays of such sites, each binding various combinations of *trans*-acting factors in different tissues (Weintraub, 1985).

Good evidence for a regulatory role for hypersensitive sites in gene expression is provided by experiments on the *Drosophila melanogaster* gene for salivary glue protein, *Sgs 4*. This protein is produced in the salivary glands of 3rd instar larvae of *Drosophila*, and a number of mutant strains are available in which the production of this protein is reduced, or absent altogether. There is a group of 5 DNAase I hypersensitive sites upstream of the gene in 3rd instar larvae, but not in embryos or tissue culture cells in which this gene is inactive (Shermoen and Beckendorf, 1982). Mutant strains producing 1-3% of the wild-type levels of glue protein mRNA have a 50 bp deletion 5´ to the gene which truncates one of the hypersensitive sites, while the strain BER I, in which there is no detectable glue protein mRNA, has a deletion covering the region of two of the hypersensitive sites. This strain however, lacks all five of the hypersensitive sites (Shermoen and Beckendorf, 1982). These observations, suggest that the sequences deleted in BER I are important for regulation of expression of *Sgs 4*; while those deleted in the other strains are important in regulating the levels of mRNA, but are not involved in temporal regulation as these mutants still produce the protein at the appropriate developmental stage (Shermoen and Beckendorf, 1982, Muskavitch and Hogness, 1982). The importance of the deleted regions

is supported by the effects of single-base changes near the hypersensitive regions in several glue protein underproducers (McGinnis *et al*, 1983a). The region appears to function independently of distance as when a 1.3 kb transposable element insertion just upstream to the TATA box is present, the upstream hypersensitive sites were still formed over the usual sequences (McGinnis *et al*, 1983b). mRNA production in this strain was decreased to 1-2% of wild-type levels, but the developmental specificity of expression was maintained.

The examples given above suggest that the chromatin structures perceived experimentally as DNAase I hypersensitive sites have a role in the control of expression of a number of genes, both developmentally regulated (*e.g.* chicken globin genes, *Sgs* 4) and inducible (*Drosophila* heat-shock genes). It is surprising, therefore, that in the only studies to date which have attempted to detect these sites in plant chromatin, none have been found (Murray and Kennard, 1984, Ferl, 1985). The study on a phaseolin gene in cotyledons and leaves of French bean showed no sub-bands representing DNAase I sites either within the gene, or in a 1.6 kb 5′ flanking region immediately adjacent to it (Murray and Kennard, 1984). General sensitivity, but no DNAase I hypersensitive sites, were reported for the maize *Adh-1* gene, (Ferl, 1985), although the pattern of accessibility of a number of restriction sites in the promoter of this gene was found to vary with its transcriptional status. The failure to detect hypersensitive sites in plant chromatin may reflect a genuine difference in chromatin structure between

animals and plants. Alternatively, the methods used may have prevented the detection of such sites, although this seems unlikely as the conditions permitted assessment of general nuclease sensitivity (Murray and Kennard, 1984). Further studies on these aspects of plant chromatin structure, using systems, such as legume seed storage protein synthesis, which are analogous to those studied in animals (Section 1.2), are needed to determine whether hypersensitive sites are present or absent from plant chromatin.

1.9 DNA structures associated with hypersensitive sites

The single-stranded nuclease S1 has been shown to cleave chromatin at or close to DNAase I hypersensitive sites (Larsen and Weintraub, 1981). Recombinant plasmids containing DNA from these regions were also cleaved at these sites by S1 nuclease when the plasmids were supercoiled (Larsen and Weintraub, 1981, Nickol and Felsenfeld, 1983). Supercoiling is known to favour the formation of alternative structures to B-DNA, such as cruciforms (Lilley and Markham, 1983, Lilley and Kemper, 1984), or Z-DNA (Nordheim *et al*, 1982, Nordheim and Rich, 1983, Singleton *et al*, 1982), and it has thus been suggested that hypersensitive sites contain DNA sequences which can adopt such conformations under appropriate conditions (Weintraub, 1983). However, it is unlikely that cruciform structures can be formed *in vivo* (Courey and Wang, 1983), and it is not known whether eucaryotic genomes are under torsional stress (Lilley, 1983). The binding of a reagent with differing affinities for relaxed and super-

coiled DNA suggests that the tension generated by negative supercoiling is constrained by nucleosomes in eucaryotic genomes (Sinden *et al*, 1980). Recent evidence suggests however that at least part of the genome may be under torsional tension (Lilley, 1983, Small and Vogelstein, 1985), and the energy of this tension has been implicated in the control of gene expression (Weintraub, 1985).

Although the cruciform structures mentioned above may not be formed *in vivo* (Courey and Wang, 1983), there is direct evidence that the regions around hypersensitive sites have an unusual DNA sequence and may lack nucleosomes. The 5′ end of the chicken adult β-globin gene, which has been shown to contain hypersensitive sites to DNAase I (Stalder *et al*, 1980b) and S1 nuclease (Larsen and Weintraub, 1982, Nickol and Felsenfeld, 1983), is excised from chromatin by the restriction endonuclease Msp 1 to yield a 115 bp fragment (McGhee *et al*, 1981). This result suggests that the region is nucleosome-free, as the presence of such a structure, associated with ≈146 bp of DNA, would be expected to block at least one of the enzyme recognition sites (McGhee *et al*, 1981). This region also contains a stretch of 16 consecutive guanosine residues (McGhee *et al*, 1981, Nickol and Felsenfeld, 1983). It has been suggested (Drew and Travers, 1984, 1985) that the rate of DNAase I cutting at a point where a sequence alters, *e.g.* from a (dA).(dT) stretch to a (dG).(dC) stretch, is determined by changes in the conformation of the minor groove, to which DNAase I is known to bind (Drew, 1984, Drew and Travers, 1984). Thus it may be that the junction between one type of sequence and another, and resulting changes in

the conformation of the sugar-phosphate backbone of the helix, are recognised, rather than the base sequence itself (Drew and Travers, 1985). In this context it is interesting to note that the junctions between B-DNA and Z-DNA in negatively supercoiled plasmids are sensitive to S1 nuclease (Singleton *et al*, 1984), again suggesting a structural rather than a sequence recognition. Thus sequences at or adjacent to DNAase I hypersensitive sites may be important in generating the altered structures at these sites (Elgin, 1984).

A regulatory role for left-handed Z-DNA (Wang *et al*, 1979) which usually consists of an alternating pyrimidine-purine sequence, has been proposed (Rich, 1982, Rich *et al*, 1984). Antibodies to Z-DNA bind to sites in eucaryotic chromosomes (Nordheim *et al*, 1981, Rich, 1982) and a sequence with the potential to form Z-DNA has been shown to be widely distributed in encaryotic genomes (Hamada *et al*, 1982). Methylation of cytosine residues can favour the B to Z-DNA transition (McIntosh *et al*, 1983, Rich *et al*, 1984). Under methylation of CpG dinucleotides is associated with transcriptional activation in many, but not all, eucaryotic genes (Doerfler, 1983), and clusters of such dinucleotides have been implicated in the nuclease sensitivity of certain genes (Wolf and Migeon, 1985). Thus models for negative regulation of gene expression by Z-DNA formation have been formulated (Rich, 1982, Reeves, 1984). In one model, the effects are local, *i.e.* affect only one or a few genes, and are mediated *via* changes such as methylation. The second model involves longer range effects, possibly mediated *via* supercoiling, within a loop or domain (Section 1.6) of chromatin (Rich, 1982).

Although most of the evidence implicating Z-DNA, or other perturbations of DNA structure, in the regulation of gene expression, is correlative, the experiments summarised in this section indicate that DNA secondary structure is probably involved in establishing different chromatin structures such as hypersensitive sites. Possible interactions of such sites with specific binding proteins, and evidence for the role of these proteins in gene expression, is presented below.

1.10 DNA-binding proteins and interactions with hypersensitive sites

The proteins described in this section are those which have been shown to bind in the vicinity of DNAase I hypersensitive sites and/or whose binding has been shown to be different to active and inactive genes. Thus the various histone modifications which have been correlated with active sequences (Reeves, 1984) and the associations of high mobility group (HMG) proteins, such as HMG 14 and 17, with active genes and DNAase I sensitivity (Weisbrod *et al*, 1980, Weisbrod and Weintraub, 1981) are not described further.

1.10.1 Proteins that bind to *Drosophila* heat-shock genes

Two DNA-binding proteins have been detected in *Drosophila* nuclei (Parker and Topol, 1984a,b). One of these, the B factor, binds to 65 bp around the start of transcription of the histone H3, H4 and the actin 5C genes. This region includes the TATA box (Section 1.4.1) and part of the leader sequence (Parker and Topol, 1984a). The second factor binds

to a 55 bp region 5´ to the TATA box in the heat-shock gene *hsp* 70 (Parker and Topol, 1984b); this region contains a consensus sequence common to *Drosophila* heat-shock genes (Pelham, 1982). This heat-shock transcription factor (HSTF) has been shown to be more active in heat-shocked than in control nuclei (Parker and Topol, 1984b). The binding activity of the B-factor is reduced in nuclei from heat-shocked cells; it has been suggested that this may be associated with a decrease in transcription of most sequences other than the heat-shock genes (Parker and Topol, 1984b).

Studies on DNA-binding proteins often involve the use of protection assays against various nucleases. The HSTF binding was shown by DNAase I footprint analysis (Galas and Schmitz, 1978). This assay reveals regions protected by protein factors as gaps in the pattern of nuclease-generated fragments compared to those which result from digests on DNA in the absence of protein factors. A second protection method involves digestion of DNA, in the presence of protein extracts, with exonuclease III, and analysis of the resulting fragments (Wu, 1984a). Use of this method to probe the structure of the *Drosophila hsp* 70 and *hsp* 82 genes has revealed a protected region in non-induced cells from -12 to -40 bp in *hsp* 70 and from -17 to -39 in *hsp* 82 (site I) (Wu *et al*, 1984a,b, 1985). These fragments are also protected in heat-shocked cells, as is another region upstream of each gene (site II). The protected regions map within the positions of DNAase I hypersensitive sites (Wu, 1980, 1984a). The initial studies involved mild digestion of nuclei with DNAase I, followed by exonuclease III digestion (Wu, 1984a). This work was extended using assays *in vitro* on chromatin (Wu,1984b)

and DNA (Wu, 1985) of *hsp* 82 or *hsp* 70 incubated with crude nuclear extracts from non-shocked or heat-shocked cells. Site I included the TATA box and the binding activity was present in both extracts. Site II, which included the consensus sequence (Pelham, 1982), was protected only in, or with extracts from, heat-shocked nuclei. These results suggest that the TATA-binding activity may be involved in establishing a chromatin structure, detected as a hypersensitive region, that allows binding by the heat-shock activating protein (HAP) on induction (Wu, 1985). The poor TATA-binding of *hsp* 82 compared with *hsp* 70 may be due to the high affinity of a 28 bp symmetrical dyad in *hsp* 82 for HAP (Wu, 1985). HAP and HSTF bind to similar regions of DNA 5´ to the heat-shock genes but their binding activities are different in non-induced cells (Wu, 1984b).

The binding of proteins within a hypersensitive region and the consequent protection of DNA may not be contradictory as it may be conformational changes due to the protein/DNA interactions which result in hypersensitivity. It has been shown (Udvardy *et al*, 1985) that two regions flanking a 12 kb region of the 87A7 locus in *Drosophila* show protection in footprinting assays. This locus contains two *hsp* 70 genes, transcribed in opposite directions. The protected sites are flanked by cutting sites and are longer than expected for a nucleosomal structure; extra cuts can be introduced into these regions after heat-shock. The structure gradually returns to the non-induced state after heat-shock, with a time course similar to that of the general recovery from heat-shock (Udvardy *et al*, 1985). These protected regions have therefore been

proposed to define a region which is inaccessible in non-induced cells, possibly due to interactions of proteins bound at the two sites, but which unfolds during heat-shock.

1.10.2 Proteins that bind to the chicken β-globin genes

Protected regions within the 5′ hypersensitive domain in the adult β-globin genes of chicken erythrocytes have also been found (Jackson and Felsenfeld, 1985). Two sites within the region are protected in adult erythrocytes, while the entire region in 5 day primitive red blood cells, in which these genes are inactive, is relatively resistant to DNAase I. Reconstitution of protein-free globin DNA with histones and an extract from cells expressing the gene restored DNAase I sensitivity (Emerson and Felsenfeld, 1984). Extracts from non-expressing cells did not restore the sensitivity. Footprinting experiments using partially purified preparations of the factor responsible for conferring sensitivity have shown two regions of protection (Emerson *et al*, 1985). The first contains palindromic sequences, and part of a consensus sequence found in many β-globin promoters, and the second contains the stretch of 16 guanosine residues (McGhee *et al*, 1981). The positions of protected regions in this *in vitro* assay corresponded to those *in vivo* (Jackson and Felsenfeld, 1985).

The correlation of the binding sites for the above proteins with regions of DNAase I hypersensitivity, and the ability of the factor from chicken erythrocyte nuclei to confer sensitivity, argues strongly for a role for chromatin

structure in the control of gene expression, *via* recognition by proteins involved in transcriptional regulation.

1.11 Summary of chromatin structure and gene regulation

The examples described (Sections 1.6 to 1.10) show that chromatin structure, both over large domains which show increased sensitivity to DNAase I, and at specific sites in the vicinity of active genes, is implicated in the control of gene expression. It is probable that both DNA secondary structure, and recognition by specific proteins are important in this regulation. The few studies which have been carried out on plant chromatin (Section 1.8) suggest that although the structure of bulk chromatin may be similar to that of animals, there are a number of differences (Murray and Kennard, 1984). In view of this, further studies on the chromatin structure of specific plant genes are required.

1.12 Aims and objectives of this work

In order to extend the studies of the molecular biology of pea seed storage protein genes, one of the aims of this work was to isolate a vicilin gene from a pea genomic library and to characterise it by restriction mapping and DNA sequencing.

An investigation of the nuclease sensitivity of specific genes in pea was seen as an important extension of the studies on plant gene regulation. At the time when this work was begun there was very little information available on the digestion of plant genes by nucleases. Thus it was

necessary to establish appropriate methodology for such an investigation. The general sensitivity of the pea seed storage protein genes was examined by comparing the rate of digestion of the active genes in developing cotyledons with that of the inactive genes in leaves. Another objective of this work was to establish if pea chromatin contained DNAase I hypersensitive sites, as at the time there was no information about the presence or absence of such sites in plant chromatin.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and biological reagents

All reagents, with the exception of those listed below, were from BDH Chemicals Ltd., Poole, Dorset, U.K., and were of analytical grade or the best available.

Acrylamide, bis-acrylamide, adenosine 5´-triphosphate (ATP), ampicillin, bovine serum albumin (BSA), 3´,5´-diaminobenzoic acid (DABA), DNAase I, dithiothreitol (DTT), ethidium bromide (EtBr), herring sperm DNA, isopropylthiogalactoside (IPTG), lysozyme, D-mannitol, maltose, MES (2[N-Morpholino]ethane sulphonic acid), pronase, PMSF, proteinase K, RNAase A, spermidine and spermine were from Sigma Chemical Co., Poole, Dorset, U.K.

Sephadex G-50, Ficoll 400 and Percoll were from Pharmacia Fine Chemicals, Uppsala, Sweden.

Caesium chloride (CsCl) and sodium chloride (NaCl) were from Koch-Light Ltd., Haverhill, Suffolk, U.K.

Nitrocellulose filters (BA85, 0.45μm) were from Schleicher and Schull, Anderman and Co. Ltd., Kingston-upon-Thames, Surrey, U.K.

3MM paper and GFC discs were from Whatman Ltd., Maidstone, Kent, U.K.

Bacto-Agar was from Difco Laboratories, Detroit, Michigan, U.S.A.

BBL trypticase peptone was from Becton Dickinson and Co., Cockeysville, M.D., U.S.A.

Yeast extract was from Sterilin Ltd., Teddington, U.K.

Restriction endonucleases were from Bethesda Research Laboratories (U.K.) Ltd., (BRL), Cambridge, U.K., The Boehringer Corporation (London) Ltd., Lewes, East Sussex, U.K., or New England Biolabs., CP. Laboratories Ltd., Bishops Stortford, Herts, U.K.

T4 DNA ligase, 5-dibromo-4-chloro-3-indoylgalactoside (Xgal) and Tris(hydroxy methyl)aminomethane were from The Boehringer Corporation (London) Ltd.

Micrococcal nuclease (MNase) was from Worthington Biochemicals, Millipore (U.K.) Ltd., London,, U.K.

Radiochemicals and nick-translation kit N.5000 were from Amersham International p.l.c., Amersham, Bucks, U.K.

M13 sequencing kit and agarose were from Bethesda Research Laboratories (U.K.) Ltd.

Plasmids and λNM258 used in the experiments described were supplied by Dr. R.R.D. Croy from communal stocks.

All solutions, with the exception of electrophoresis buffers and solutions for Southern blotting, were sterilised by filtration or by autoclaving.

2.1.2 Bacterial strains and plasmid and bacteriophage vectors

Bacterial strains were derivations of *E.coli* K12. The table below lists these strains, plasmids and bacteriophage used as vectors or probes, and the sources or references for each.

TABLE 1 *E.coli* strains, plasmids and bacteriophage

Bacterial Strains	Genetic Characters	Reference or Source
JM83	ara, Δ(lac-proAB) rpsL (=strA) φ80, lacZ Δ M15.	Bethesda Research Laboratories (BRL)
JM101	Δlacpro, supE, thi, F´traD36, proAB $lacI^q$ ZΔM15.	Dr. J. Messing or BRL.
K803	r_k^-, m_k^- gal^- met^-	Wood, W.B. (1966)
Q359(P2)	r_k^- m_k^+ 80^R su_{II}^+ with P2 lysogen.	Karn *et al* (1980)
Q358	r_k^- m_k^+ 80^R Su_{II}^+	Karn *et al* (1980)
Plasmids		
pDUB2	Ap^R vic 50K (in pBR322)	Croy *et al* (1982) Lycett *et al* (1983a)
pDUB9	vic 50K (in pBR322)	Delauney (1984)
pDUB7	vic 47K (in pBR322)	Lycett *et al* (1983a)
pDUB24	*leg. A.* (in pUC8)	Lycett *et al* (1984)
pUC8	Ap^R, lacZ	Vieira and Messing (1982)
pBR322	Ap^R Tc^R	Bolivar *et al* (1977)
Bacteriophages		
EMBL3	spi^-, trpE	Frischauf *et al* (1983)
M13mp18	multiple cloning site	BRL
M13mp19	multiple cloning site	BRL

KEY: Ap^R = ampicillin resistance, Tc^R = tetracycline resistance, vic 50K = 50K vicilin cDNA, vic 47K = 47K vicilin cDNA, *leg A* = legumin gene A.

2.1.3 Plant Material

Plants were grown as described by Evans *et al* (1979). Pea seeds (*cv*. Feltham First, obtained from Sutton Seeds Ltd., Reading, Berks, U.K.) were germinated on Alkathene polyethylene granules (I.C.I. Plastics Division, Welwyn Garden City, Herts, U.K.) in a dark spray room at 28°C, with water misting for 5 min. in every hour for 4-5d. Seedlings were transferred to 2ℓ water culture bottles of nutrient solution and grown under controlled conditions as follows: 16h total lighting, 28°C day temperature, 23°C night temperature, 75-80% relative humidity. Leaves were harvested from similar points on plants of the same age, and pods harvested at various times (9-22 days) after flowering. For nuclei preparations, tissue was used at once. For DNA preparations, tissue was frozen in liquid nitrogen and stored at -80°C until required.

2.1.4 Glassware and plasticware

Glassware and plasticware used in manipulations of DNA, bacterial cultures and for storage of sterile stock solutions and media, was autoclaved prior to use. Glassware, and plasticware used for very small amounts of DNA, or when very good recovery was required, was siliconised using 'Repelcote' (Hopkin and Williams, Romford, U.K.).

2.1.5 Growth media for bacteria and bacteriophage

The composition of the various media used for growth of micro-organisms is given below:

TABLE 2 Composition of growth media

Medium	Nutrients (1ℓ)
L broth	10g trypticase 5g yeast extract 10g NaCl.
BBL bottom layer agar	10g trypticase 5g NaCl 5g $MgSO_4 . 7H_2O$ 10g agar.
BBL top layer agar	As BBL agar but 6.5g agar.
YT broth	8g trypticase 5g yeast extract 5g NaCl
YT agar	As YT broth but 15g agar.
YT top layer agar	As YT agar but 7.5gagar
YT/Amp/Xgal	YT agar containing 50µg/ml ampicillin and 40µg/ml X-gal.
S-broth	33g trypticase 20g yeast extract 75g NaCl 1g D-glucose

For the gene library screens (Section 2.2.10), BBL top layer agarose was used instead of agar as agarose was less likely to slip or to stick to the filters during plaque lifts (Section 2.2.10.2). Agarose was also used in place of top layer agar for plate lysates (Section 2.2.11.1) because agar contains inhibitors of restriction endonucleases which are difficult to eliminate from DNA preparations.

2.2 Methods

2.2.1 Biochemical techniques

2.2.1.1 Phenol extraction of DNA samples

Samples were deproteinised by addition of 1 vol. of redistilled phenol equilibrated with T.E. buffer (10mM Tris, 1mℓ ethylenediaminetetracetic acid (EDTA)) pH8.0. Phases were mixed by vortexing for 30sec. and separated by centrifugation at 12000 *g* in a microfuge (MSE MicroCentaur) for 1 min., or, for larger samples, at 10,000 *g* in a Sorvall RC-5B centrifuge for 5 min. The aqueous phases were re-extracted with 1 vol. of phenol:chloroform;isoamylalcohol (25:24:1 by vol). In some cases, the organic phases were re-extracted with 0.5 vol of T.E. buffer and the aqueous phases combined. Phenol was removed either by 2-3 extractions with diethyl ether, or by 1 extraction with chloroform.

2.2.1.2 Precipitation of DNA with ethanol

0.1 vol of 3M sodium acetate pH 4.8 and 2.5 vol of ethanol were added to the DNA solution and mixed. Samples were kept at -80°C for 30 min or -20°C overnight. Precipitates were collected by centrifugation at 12000 *g* for 5 min in a microfuge, or at 12000 *g* for 20 min at 4°C in the Sorvall RC-5B centrifuge. Pellets were washed in 70% (v/v) ethanol, dried briefly under vacuum and resuspended in a small volume of distilled water or T.E. buffer.

2.2.1.3 Dialysis of DNA solutions

Dialysis tubing (Medicell International Ltd., London, U.K.) was prepared by boiling for 10 min. in 2%

(w/v) sodium carbonate solution, rinsing extensively in distilled water, boiling for 5 min in distilled water and rinsing several times in distilled water. One end of the tubing was secured and the DNA solution pipetted into the tubing. The other end was secured, leaving a space to allow for an increase in volume. The tubing was placed in 2ℓ of T.E. buffer and stirred at 4°C for 24h, with 2 further changes of buffer.

2.2.1.4 Preparation of denatured herring sperm DNA

500mg DNA was dissolved in 50ml of distilled water by stirring for 2-4 h at room temperature. DNA was sheared by passing through a 19gauge hypodermic needle. The solution was boiled for 10 min. and stored at -20°C in small aliquots. The DNA was boiled for 5 min prior to use.

2.2.1.5 Spectrophotometric quantitation of DNA solutions

The optical density (O.D.) of DNA solutions were determined by adding 1µl of DNA solution to a quartz curvette containing 1ml of distilled water. 1µl of the buffer in which the DNA was dissolved was added to 1 ml of water in a reference curvette, and the optical density from 320 to 230nm recorded using a Pye Unicam SP8-150 uv/vis spectrophotometer.

The $O.D._{260}$ of a DNA solution of 1µg/ml is 0.02. A pure DNA sample has an $O.D._{260}/O.D._{280}$ ratio of ∿1.8 and the $O.D._{260}/O.D._{235}$ ratio is greater than this. The $O.D._{320}$ is zero. Deviations from these values indicated the presence of contaminants such as phenol or proteins in the solution.

2.2.1.6 Storage of bacteria

Bacteria were stored for up to 6 weeks at 4°C on inverted agar plates tightly sealed with Nescofilm (Nippon Shoji Kaisha Ltd., Osaka, Japan). For longer storage, single colonies were streaked out onto selective agar plates and bacteria from the streak used to inoculate vials containing 1 ml of L broth (Section 2.1.5). Cells were resuspended and 1 ml 80% (v/v) glycerol added and mixed in. Vials were stored at -80°C. Bacteriophage were stored at 4°C in phage buffer (Section 2.2.10) saturated with chloroform.

2.2.1.7 Pre-treatment of ribonuclease

RNAase A was dissolved in water to 10mg/ml and boiled for 10 min. to inactivate DNAases. Small aliquots were stored at -80°C.

2.2.2 Enzymatic reactions used in manipulations of DNA

2.2.2.1 Restriction endonuclease digestions

The three-buffer system shown below was used (Maniatis *et al*, 1982).

TABLE 3. Restriction endonuclease buffers

Buffer	NaCl	Tris/HClpH7.4	$MgSO_4$	DTT
Low	0	10	10	1
Med	50	10	10	1
High	100	50	10	1
Sma I	20mM KCl	10 (pH8.0)	10	1

Concentrations are mM, and stocks of 10x these were used.

DNA was digested with a 2-3 fold excess of enzyme. 0.1 vol of a 10x stock of the appropriate buffer (Table 3) was added, and the volume made up with distilled water to 10x that of the enzyme. Plasmid and bacteriophage DNAs were incubated for 2-3 h at the appropriate temperature, genomic DNAs for at least 5 h. Reactions were terminated either by heating to 70°C for 5 min or by addition of agarose gel loading beads (20% (v/v) glycerol, 10mM Tris/HCl, pH8.0, 10mℓ EDTA pH8.0, 0.2% (w/v) agarose, 0.1% (w/v) each of bromophenol blue, xylene cyanol and orange G. The mixture was autoclaved and forced through a 19 gauge hypodermic needle to form the beads).

2.2.2.2 Ligation of DNA fragments

DNAs to be ligated were digested with appropriate restriction endonucleases, phenol extracted, ethanol precipitated and resuspended in a small volume of T.E. buffer pH8.0. Fragments were mixed to give a 3-fold molar excess of insert over vector. 0.1 vol of 10x ligation buffer (0.66M Tris.HCl pH7.5, 50mM $MgCl_2$, 50mM DTT, 10mM ATP) and 2-5 units of T4 DNA ligase (Weiss, 1966) were added, mixed in and the tubes incubated at 15°C overnight. 0.1 to 0.5 vol of the ligation mixes was used to transform competent *E.coli* cells, (Section 2.2.14.2).

2.2.3 Preparation of pea genomic DNA from leaves and cotyledons

A modification of the method of Graham (1978) was used. Tissue was weighed and ground in liquid nitrogen in a mortar and pestle at 4°C. Ground tissue was mixed into

5 vol of 0.1M NaCl, 0.025M EDTA pH8.0, 2% (w/v) SDS, and 0.1% diethylpyrocarbonate (DEP). 5M sodium perchlorate was added to 1M and the solution mixed well. 0.5 vol chloroform: octanol (99:1 (v/v)) and 0.5 vol phenol were added. The mixture was shaken on ice for 50 min, then centrifuged at 20,000*g* for 5 min at 4°C (Sorvall RC-5B centrifuge). Aqueous phases were pooled and extracted with an equal vol of chloroform: octanol by shaking as above. Phases were again separated by centrifugation. Aqueous phases were pooled, two vol ethanol added and tubes left at -20°C for 30 min.

Nucleic acids were spooled out using a siliconised glass rod, pooled and dried briefly under vacuum. 20 to 25 ml of resuspension buffer (50mM Tris, 10mM EDTA, pH8.0) were added and the solution shaken overnight on ice.

0.6 ml of a 2% (w/v) solution of Pronase, previously incubated at 37°C for 2h to self-digest, was added to the DNA solution and incubated at 37°C for 3h. The sample was weighed and solid caesium chloride added to 0.94g/g solution. EtBr solution (10mg/ml) was added to 100µg/g of solution, and the tube shaken gently until all the solid had dissolved. The solution was transferred to quick-seal tubes (Beckman) using sterile syringes and needles and centrifuged in a Beckman VTi-50 rotor in a Sorvall OTD-65 ultracentrifuge at 44000 rpm for 20 h at 15°C. DNA bands were removed and recentrifuged. Bands were again withdrawn, EtBr removed by extraction with isoamylalcohol saturated with resuspension buffer, and the DNA solution dialysed against resuspension buffer. DNA was ethanol precipitated, and dissolved in resuspension buffer.

DNA was stored in small aliquots at -80°C and the purity and concentration of the solution determined spectrophotometrically or by DABA fluorescence as described in Section 2.2.20.2.

2.2.4 Agarose gel electrophoresis

Agarose was dissolved in 200ml of Tris/acetate electrophoresis buffer (0.04M Tris-acetate, 0.001M EDTA pH7.7) by heating in a microwave oven for 4-5 min. EtBr solution (10mg/ml) was added to 0.75µg/ml and the gel poured in a horizontal 180 x 150mm perspex mould sealed to a glass plate with silicon grease. When set, the mould was removed and the gel placed in an electrophoresis tank containing Tris/acetate buffer. Samples were mixed with agarose gel beads (Section 2.2.2.1) and loaded into the wells. Electrophoresis was overnight at 30V, or for 4-5h at 100-120 V.

Agarose minigels were made using Tris/borate buffer (TBE, 0.089M Tris, 0.089M boric acid, 0.002M EDTA) and cast in a minigel electrophoresis apparatus (Cambridge Biotechnology Laboratories). Minigels were electrophoresed at 50V for 45-90 min.

Gels were visualised under short-wave ultraviolet illumination and photographed using a red filter and ASA 3000 film, Type 667 (Polaroid).

2.2.5 Transfer of DNA onto nitrocellulose filters

A modification of the method of Southern (1975) was used. Gels containing genomic DNA were soaked briefly in 0.1N HCl and rinsed in distilled water. This step was omitted for gels containing plasmid or bacteriophage DNA. Gels

were soaked 45 min in denaturing solution (1.5M NaCl, 0.5M NaOH), rinsed several times in distilled water and soaked for 45 min in neutralising solution (3M NaCl, 0.5M Tris/HCl, pH7.0). Gels were equilibrated in 20xSSC (3M NaCl, 0.3M sodium citrate) for 5 min and placed on top of a piece of Whatman 3MM paper on a glass plate, arranged so that the ends of the paper dipped into a reservoir of 20xSSC. A piece of nitrocellulose was soaked in distilled water for a few min, then in 20xSSC, and placed on top of the gel. A piece of Whatman 3MM paper soaked in 20xSSC was placed on top of the filter, followed by several dry pieces of paper and layers of absorbent material such as disposable nappies. The whole apparatus was weighed down with a glass plate and brick and transfer of DNA proceeded for 3h to overnight at room temperature.

The nitrocellulose filter was removed and baked at 80^{o}C under vacuum for 30-120 min. Filters were stored between sheets of 3MM paper until required.

2.2.6 Isolation of DNA fragments from agarose gels

A number of methods were tried; the two given here were those most commonly used. For both methods, the DNA was digested with an appropriate restriction endonuclease, electrophoresed through an agarose gel, and the required fragment cut out from the gel, trimming off excess agarose.

2.2.6.1 Isolation on glass-fibre discs

This was a modification of the method of Yang *et al* (1979), using sodium perchlorate to disrupt the gel. All perchlorate solutions were filtered on glass fibre discs

prior to use. The gel slice was weighed and 3 times this weight in volumes of 8M sodium perchlorate were added. The slice was allowed to dissolve for 10-30 min at room temperature.

A 6mm disc cut from a Whatman GFC glass-fibre disc was placed on a sheet of 3MM paper placed on top of several layers of absorbent tissue, and was washed with 5x30µl of 6M sodium perchlorate dissolved in T.E. buffer, pH7.5. The dissolved gel slice was dripped slowly onto the disc, followed by 50x30µl 6M sodium perchlorate in T.E. buffer. The disc was washed with 50x30µl of 95% ethanol, dried in air and placed in a 0.5ml eppendorf. 20µl of 0.1xT.E. buffer pH7.5 was added and the tube incubated at 37°C for 30 min. A hole was pierced in the base and the eluted DNA solution spun out into a 1.5ml eppendorf.

2.2.6.2 Electroelution onto dialysis membranes

This method was adapted from that of McDonnell *et al* (1977) by Dr. A. Shirsat. The gel slice was placed in a piece of dialysis tubing secured at one end. 0.5ml of Tris/acetate electrophoresis buffer (Section 2.2.4) was added, and the open end of the tubing closed, excluding air bubbles. The tubing was placed in an electrophoresis tank containing Tris/acetate buffer and electrophoresed at 100V for ∿1h until the DNA was visible, under u.v. illumination, as a thin line on the tubing. The current was reversed for about 30 sec. and the buffer removed from the tubing and placed in a 1.5ml eppendorf. The buffer was phenol extracted and the volume reduced by extraction with butanol until about 0.1ml remained.

Butanol was removed by 3-4 extractions with diethyl ether and DNA was precipitated with ethanol and resuspended in distilled water.

2.2.7 ^{32}P-labelling of DNA by nick-translation

The nick-translation kit (Section 2.1.1) was used according to the manufacturer's instructions for labelling to a specific activity of 10^8 c.p.m./µg, except that the volumes were adjusted so that 0.1 to 0.2µg of DNA was routinely labelled with 50 µCi of α[^{32}P]-dCTP in a volume of 25µl. Incubation at 15°C was for 2-2½h.

Labelled DNA was separated from unincorporated radionucleotide by passage through a 5-cm column of Sephadex G50 equilibrated in 150mM NaCl, 10mM EDTA, 50mM Tris/HCl pH7.5, and 0.1% SDS. 0.4ml fractions were collected and 1µl aliquots counted in a β-scintillation counter. Fractions corresponding to the 1st peak of radio-activity contained labelled DNA, and were pooled for use as a hybridisation probe. Specific activities of > $1x10^8$ c.p.m./µg were routinely obtained.

2.2.8 Hybridisation to nitrocellulose filters

Filters were prehybridised in the solutions indicated below for 1 to 4 h at 65°C.

The probe was denatured by boiling for 6-8 min, and was cooled rapidly on ice. The prehybridisation solution was removed from genomic filters and replaced with a similar one containing 1.5 x Denhardts solution, before addition of probe. Probe was added directly to the prehybridisation solutions of other

TABLE 4 Prehybridisation solutions

Type of filter	Solution
Genomic DNA	5xSSC 5xDenhardts' solution 100µg/ml denatured herring sperm DNA. 0.1% SDS
Plasmid or bacteriophage DNAS, colony hybridisations	As above but 3xSSC.
Plaque lifts	As above but 6xSSC.

1 x Denhardts' solution is 0.02% (w/v) of each of Ficoll 400, BSA and polyvinylpyrrolidone (PVP).

---

filters. Usually 1-5x10^7c.p.m.from a probe of specific activity >10^8c.p.m/µg were added to genomic DNA filters, and 5x10^6 to 1x10^7c.p.m.to other filters. Hybridisations were at 65°C overnight except where indicated otherwise.

Filters were washed to the stringencies indicated in the figure legends. Genomic DNA filters were typically washed 2x30 min at 50°C in 5xSSC, 0.1% SDS, 2x15 min at 65°C in 0.1xSSC, 0.1% SDS. Plaque lift filters were washed 2x30 min at 65°C in 3xSSC. Plasmid and bacteriophage filters, and colony hybridisations were washed 1x30min in 3xSSC, 0.1% SDS, 1x30min. in 2xSSC, 0.1% SDS and 2x15 min in 0.1xSSC, 0.1% SDS, all at 65°C.

Filters were dried in air, or under vacuum at 80°C for a few min.

2.2.9 Autoradiography

Filters were taped to a piece of Whatman 3MM paper on a glass plate, the origin and sides marked with radioactive ink, and the whole covered with cling film. Pre-flashed film (Fuji RX Safety) and an intensifying screen (Dupont Lightning Plus) were placed over the filter (Laskey and Mills, 1977), followed by a glass plate. The assembly was secured with tape or elastic bands, wrapped in black polyethene bags and placed inside a light-tight box. Autoradiography was at -80°C for 30 min to 3 weeks. Films were developed in Phenisol developer (Kodak) and fixed in Kodafix (Kodak).

2.2.10 Screening of a pea genomic library

The library of pea genomic DNA was constructed in the bacteriophage vector EMBL3 (Frischauf *et al*, 1983) by Dr. A. Shirsat of this department. The library was screened by the method of Benton and Davis (1977). An outline of the screening procedure is shown in Figure 1, and details of each screen are given in Section 2.2.10.4.

2.2.10.1 Transfection of host cells with bacteriophage

E.coli host cells (Q359 or K803) were grown overnight in L broth supplemented with 0.4% (w/v) maltose. The culture was centrifuged 5 min at 4000 *g* and the pellet re-suspended in 0.5 vol of phage buffer (50mM $MgCl_2$, 10mM Tris/HCl, pH7.5). Phage at various dilutions were mixed with plating cells (2.5×10^5 p.f.u. and 3ml of cells in the first screen, 100µl of each phage dilution and 100µl of cells in subsequent

screens) and incubated at 37°C for 15 min. Top layer BBL agarose at 50°C was added to each (30 ml per plate in the first screen, 3ml for subsequent screens), mixed, and the contents poured onto BBL agar plates (20x20 cm square in the first screen, 9 or 15 cm diameter petri dishes in subsequent screens). When set, plates were inverted and placed at 37°C overnight, then stored at 4°C.

2.2.10.2 Transfer of plaques to nitrocellulose filters

Filters, marked with a hole punch to aid in orientation, were placed on the surface of the agar plates for 5 min. Holes were punched through into the agar *via* the holes in the filters. Filters were removed, placed plaque side up on blotting paper soaked in denaturing solution (Section 2.2.5), and left for 3-4 min. Filters were transferred to blotting paper soaked in neutralising solution (Section 2.2.5) for 3-4 min, and then to blotting paper soaked in 2xSSC (0.3M NaCl, 0.03M sodium citrate) for a further 3-4 min. Filters were baked at 80°C under vacuum for 1 h, hybridised (Section 2.2.8) to the ^{32}P-labelled insert of the vicilin cDNA, pDUB2 (Croy *et al*, 1982, and Fig.2), and autoradiographed at -80°C. In each screen, a duplicate set of filters was hybridised to ^{32}P-labelled pBR322.

2.2.10.3 Identification of recombinants

Plaques hybridising to pDUB2 were identified by aligning the autoradiographs with the holes punched in the agar, picked off the plates, and phage eluted in 1-2 ml of

FIGURE 1

Outline of protocol for screening of a pea genomic library.

Host cells (K803 or Q359) infected with phage at various dilutions.

Top layer agarose added, poured onto BBL agar plates. Incubated overnight at 37°C.

Plates transferred to fridge for 20 min. to stop further growth and harden top layer agarose.

Plaques transferred to nitrocellulose filters.

Filters baked under vacuum and hybridised to pDUB2. Autoradiographed at -80°C.

Plugs of agar removed from around plaques hybridising to the probe. Phage eluted into buffer and used in subsequent rounds of screening.

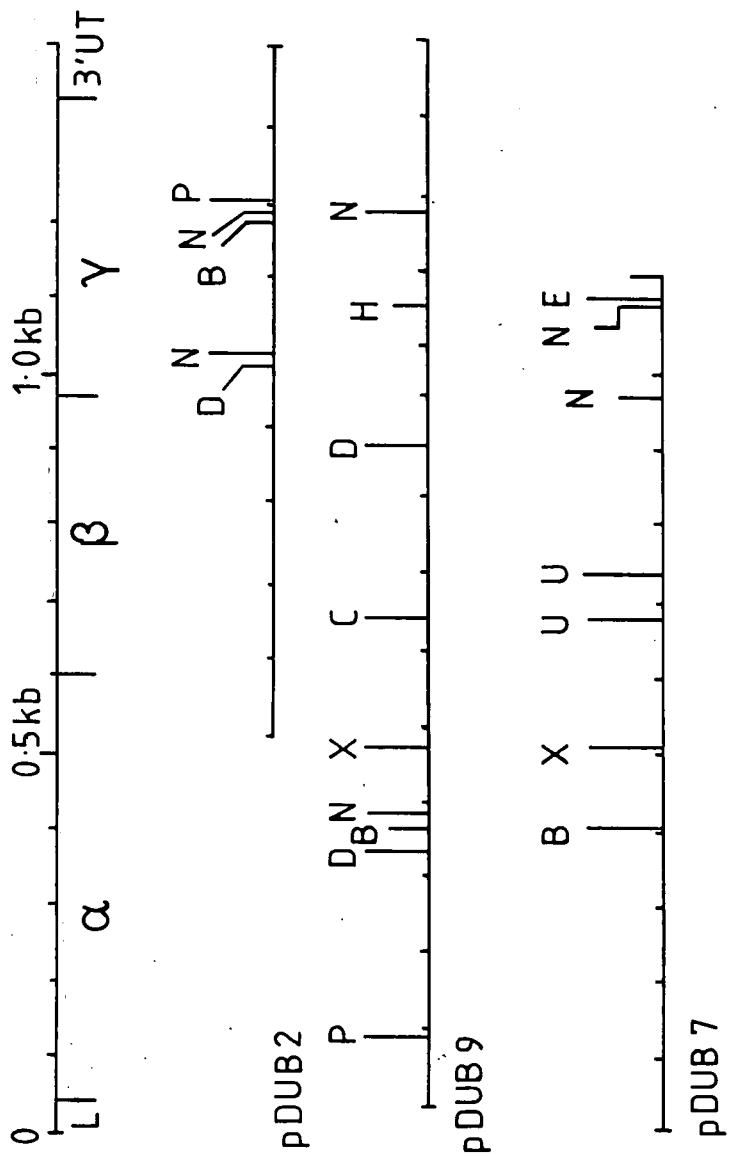

0
0·5kb
1·0kb
L1
α
β
γ
3'UT
pDUB2
pDUB9
pDUB7
P
D
B
N
X
C
D
H
N
D
N
B
N
P
B
X
U
U
N
N
E

Fig. 2

FIGURE 2

Restriction maps of vicilin cDNAs

The maps of the vicilin cDNAs pDUB2 (Croy *et al*, 1982, Lycett *et al*, 1983a), pDUB7 (Lycett *et al*, 1983a) and pDUB9 (Delaney, 1984 are aligned with an idealised vicilin message showing the α/β and β/γ potential processing sites.

Key - L = leader sequence, 3´UT = untranslated 3´ region, D = Hind II, N = Bst Nl, B = Bgl II, P = Pst 1, X = Xba I, C = Acc I, H = Hind III, U = Sau 96l, E = Bst EII.

phage buffer saturated with chloroform. In the first screen, a plug of agar was removed from around the positive plaque using a cork borer, in subsequent screens several plaques were picked off from around each area of hybridisation using pasteur pipettes. This was to allow for any misalignment of films with the holes in the agar. In the final screen, the number of plaques was such as to allow exact identification with hybridisation spots on the autoradiograph. Three individual plaques were picked off and eluted in 400µl of phage buffer saturated with chloroform. The eluates from the plaques picked off in each screen were used as sources of phage for the next screen.

2.2.10.4 Variations in screening procedure

In the first screen, 2.5×10^5 phage from the library were plated out. In the second screen, the phage from the agar plug removed from a 1st screen plate was used to infect both K803 and Q359 cells. In subsequent screens only K803 cells were used as hosts as more intense hybridisation signals appeared to be obtained with these cells. The plaques around 12 regions of hybridisation in the second screen were picked off and eluted. 5µl of each of these was spotted onto a lawn of plating cells in the third screen. Positives were not picked off, but were used to select which of the 12 clones should be screened further. 6 clones were screened a 4th time. Plaques were picked from 4 of these (7 in total) and screened a 5th time. Individual plaques were picked off and 3 of these screened a 6th time. Three plaques were picked from a plate on which >90% of the plaques hybridised (see previous section). One of these plaques was plated onto Q359 host cells and used

to prepare stocks of bacteriophage (Section 2.2.10.5).

2.2.10.5 Preparation of stocks of purified bacteriophage

Transfections were set up as in Section 2.2.10.1, using the eluate from one of the plaques picked off in the final screen. The dilutions used were such as to give confluent lysis (10^{-1} and 10^{-2}). After overnight incubation of the plates, 4 ml of phage buffer was added to each and the plates shaken at room temperature for 5h. Buffer was removed, pooled and centrifuged briefly at 8000*g* to remove bacterial and agar debris. A few drops of chloroform were added and the phage stored at 4°C. Plates were re-extracted overnight at 4°C with 3 ml of buffer.

Both extractions were titrated by plating out various dilutions. The total volume was less than the volume of phage buffer added to the plates due to some absorption by the agar.

2.2.11 Small-scale preparations of bacteriophage DNA

2.2.11.1 Plate-lysate method

The method was adapted from that in Maniatis *et al* (1982).

10^5 p.f.u. from the stocks were used to infect cells in plate lysates as above. Extractions were centrifuged 10 min. at 8000 *g* to remove debris and DNAase I and RNAase A added to the supernatant to 1µg/ml. The mixture was incubated at 37°C for 2 h and an equal vol. of 20% (w/v)

PEG, 2M NaCl in phage buffer was added. Tubes were left at 0°C for 2h and PEG precipitates collected by centrifugation at 10,000*g* for 20 min at 4°C. Precipitates were resuspended in 1.5 ml phage buffer, transferred to Eppendorf tubes and centrifuged briefly in the microfuge. 10µl 10% SDS and 20µl 250mM EDTA pH8.0 were added, and tubes incubated at 68°C for 15 min. 0.5ml aliquots were extracted with phenol, precipitated with ethanol and resuspended in 50µl of T.E. buffer.

2.2.11.2 Liquid lysate method

Transfections were set up using various numbers of phage (10^6 to $2x10^7$ p.f.u.). After 15 min at 37°C, transfections were added to 10ml of L broth containing 50mM $MgSO_4$ and 0.4% (w/v) maltose, and shaken vigorously at 37°C overnight. 0.2ml chloroform was added to complete the lysis, and cultures centrifuged for 10 min at 8000 *g* to remove bacterial debris. PEG precipitation of bacteriophage was as described above. Pellets were resuspended in 2ml of phage buffer and DNA purified by the method of Cameron *et al* (1977). 0.2ml aliquots were removed and mixed with 0.2ml phage buffer. 1µl of DEP was added and tubes vortexed. 10µl 10% SDS and 50µl of 2M Tris, 0.2M EDTA pH8.5 were added, tubes vortexed and incubated at 70°C for 5 min. 50µl potassium acetate was added and tubes cooled and left on ice for 2h. Protein debris was pelleted by centrifugation for 15 min. Supernatants were transferred to fresh tubes which were filled with ethanol at room temperature. DNA precipitates were collected by centrifugation for 4 min. and tubes drained onto absorbent tissue. Pellets were resuspended in T.E. buffer containing 50µg/ml RNAase A.

2.2.12 Large-scale preparation of bacteriophage DNA

Transfections were set up using 0.3 ml of plating cells and $5x10^8$ phage from the stock. 1.5ml of this was added to 250 ml of prewarmed S-broth, supplemented with 50mM $MgSO_4$, in a 2 ℓ flask. Cultures were shaken at 37°C overnight and chloroform added to complete the lysis. Cultures were centrifuged at 8000 *g* for 10 min and DNAase I and RNAase A added to the supernatant to a final concentration of 1µg/ml. PEG was added to 10% (w/v), and NaCl to 1M. Precipitation was overnight at 4°C. Precipitates were collected by centrifugation as for small-scale preparations and resuspended in 6-8ml phage buffer. The suspension was extracted once with chloroform, then twice with phenol using a rotary shaker at 4°C for 1 h. Phases were separated by centrifugation and the aqueous phase extracted twice with chloroform, and precipitated with ethanol at room temperature. DNA was pelleted by centrifugation at 12000 *g* for 10 min at 4°C, washed with 70% ethanol, dried and resuspended in 1 ml of T.E. containing 50µg/ml RNAase A.

2.2.13 Restriction mapping of genomic clones

The DNA obtained from large-scale preparations was digested with a number of restriction endonucleases in single and double digests (Section 2.2.2.1). Digests were electrophorised through 0.5% agarose gels with a variety of size markers. From the restriction sites in the vector DNA and the bands seen on EtBr-stained gels, a partial map of the clone was constructed. Gels were blotted onto nitrocellulose (Section 2.2.5) and hybridised to the insert of pDUB2. The

bands seen on the autoradiographs were used to predict the position of the vicilin gene within the insert (Section 3.1.4).

2.2.14 Sub-cloning into bacterial plasmids

2.2.14.1 Restriction and ligation of DNA

Two sets of sub-clones were made, one from a Sal I digest of the bacteriophage clone and one from an Eco RI/Sal I digest. Sal I-digested DNA was ligated into Sal I-digested pUC8, while Eco RI/Sal I digested DNA was ligated into either Eco RI, or Eco RI/Sal I digested pUC8 (Section 3.1.4.2).

2.2.14.2 Preparation of competent cells

This was a modification of the method of Mandel and Higa (1970). 100 ml of YT both was inoculated with 1 ml of an overnight culture of JM83 cells and shaken at 37°C until an $O.D._{600}$ of 0.3-0.4 was reached (approx.2½h). Cells were centrifuged at 4000*g* for 4 min at 4°C and the pellets resuspended in 0.5 volumes of 0.05M ice-cold $CaCl_2$. Tubes were left on ice for 10 min and centrifuged at 4000 *g* for 20 min at 4°C. Pellets were resuspended in $^1/_{15}$ of the original culture volume of ice-cold 0.05M $CaCl_2$ and left on ice for at least 30 min prior to use.

2.2.14.3 Transformation of host cells by plasmid DNA

2.5µl (0.25 vol.) of the ligation mix was added to 100µl of competent JM83 cells and incubated on ice for 20 min. Controls of the unrestricted vector, restricted and religated vector, and calcium chloride (0.05M) were included.

Cells were heat-shocked at 37°C for 5 min, and 1 ml YT broth added and left at 37°C for 1 h to allow expression of ampicillin resistance. 100µl of cells were plated out onto YT/Xgal/Amp plates (Section 2.1.5). For Eco RI/Sal I sub-clones 10^{-1} dilutions of cells were also plated out. Plates were incubated at 37°C overnight. Recombinants were identified as white colonies, this was because insertion of DNA onto the multiple cloning site of pUC8 prevents the action of the enzyme B-galactosidase on the chromogenic substrate Xgal (Vieira and Messing, 1982). Transformed white colonies were picked off onto fresh YT/Xgal/Amp plates.

2.2.15 Preparation of bacteria for colony hybridisations

A modification of the method of Grunstein and Hogness (1975) was used. Transformed colonies were picked off onto a grid drawn on a selective agar 'master' plate. Colonies were also streaked in the same positions onto a nitrocellulose filter placed on the surface of a fresh plate. The filter and plates were marked to aid orientation, and incubated at 37°C until streaks were visible. The 'master' plate was then placed at 4°C. The nitrocellulose filter was peeled off the plate and placed, colony side up, on filter paper soaked in denaturing solution (Section 2.2.5) and left for 5 min. The filter was placed on blotting paper soaked in neutralising solution, and on blotting paper soaked in 2xSSC, for 5 min each time. Filters were baked at -80°C under vacuum for 1 h, and hybridised to the ^{32}P-labelled insert of pDUB2. Positive colonies were picked from the 'master' plate and streaked onto fresh plates, or used to prepare plasmid DNA.

2.2.16 Small-scale preparations of plasmid DNA

This was essentially the method of Birnboim and Doly (1979). 2 ml aliquots of YT containing 50µg/ml ampicillin were inoculated with a single bacterial colony and grown overnight at 37°C. 1.5ml of cells were centrifuged for 30 sec in a microfuge, the supernatant removed and the tube drained onto absorbent tissue. 100µl of solution I (50mM glucose, 10mM EDTA, 25mM Tris/HCl pH8.0, 2mg/ml lysozyme) were added, tubes vortexed and kept on ice for 20 min. 20µl of solution II (0.2N NaOH, 1% SDS) were added, tubes vortexed gently, and kept on ice for 5 min. 150µl of 3M sodium acetate, pH4.8, were added, tubes inverted to mix and left on ice for 1h. Bacterial DNA and debris was pelleted by centrifugation for 5 min. 400µl of supernatant were removed and precipitated with 1 ml of ethanol at -20°C for 30 min. Precipitates were collected by centrifugation for 2 min and the supernatant removed by inverting tubes onto filter paper. Pellets were dissolved in 100µl of 0.1M sodium acetate/0.05M Tris/HCl pH6.0 and reprecipitated. Pellets were dried under vacuum and dissolved in 50µl of water or T.E. buffer.

2.2.17 Restriction mapping of plasmid sub-clones

Clones identified as containing pea genomic inserts were digested with a number of restriction endonucleases. Using known sites in the vector, the partial restriction map of the bacteriophage genomic clone, and the results of hybridisations to pDUB2 and pDUB9, maps were constructed of several plasmids (Section 3.1.4.2).

2.2.18 Construction of M13 clones

M13 cloning and dideoxysequencing were carried out by standard methods (Messing, 1983) using the BRL kit.

2.2.18.1 Restriction and ligation of DNA into M13 vectors

DNA from plasmid sub-clones containing vicilin sequences was cut with various restriction endonucleases (see sequencing strategy, Section 3.1.4.3). M13 vectors mp18 or mp19 were cut with appropriate restriction enzymes and the vicilin DNA fragments ligated into these vectors.

2.2.18.2 Transformation of *E.coli* cells by M13

JM101 cells, made competent as in Section 2.2.14.2, were used as the host. 0.1-0.5 vol of the ligation mixes were added to separate 0.3 or 0.2ml aliquots of cells, mixed, and left on ice for 40 min. Controls of unrestricted vector, restricted and religated vector and 0.05M $CaCl_2$ were included. 10µl of 100mM IPTG, 50µl of 2% (w/v) X-gal, 0.2ml of fresh exponentially growing JM101 cells and 3ml of YT top-layer agar at 50°C were added, tubes mixed and the contents poured onto YT plates. When set, plates were inverted and incubated at 37°C.

2.2.18.3 Selection of transformants

Colourless plaques, produced as a result of insertional inactivation of the β-galactosidase gene (Messing, 1983) were either picked off directly for template

preparation, or were picked off onto master-plates and nitrocellulose filters and treated as for colony hybridisation (Section 2.2.15). Plaques hybridising to the insert of pDUB2 were picked off for preparation of template DNA.

2.2.18.4 Preparation of single-stranded DNA from transformed cells

50ml of 2xYT broth was inoculated with 0.1 ml of an overnight culture of JM101 cells. 2ml aliquots were inoculated with a single plaque and grown at 37°C overnight. A blue plaque picked from a control plate (unrestricted vector) was also grown up. 1.5ml of each culture was centrifuged in a microcentrifuge for 3 min, the supernatants removed and the pellets stored at 4°C. Supernatants were re-centrifuged, and 1 ml of each transferred to a fresh tube. 200µl of 20% (w/v) PEG, 2.5M NaCl were added, mixed in and tubes left at room temperature for 20-40 min. Precipitates were collected by centrifugation for 3 min and the supernatants removed. Pellets were re-centrifuged and excess PEG removed. Pellets were resuspended in 100µl of T.E. buffer and extracted with 50µl of phenol, then with 50µl of chloroform. DNA was precipitated with 0.1 vol of 3M sodium acetate (pH4.8) and 3 vol of ethanol at -20°C for 1 h., pelleted by centrifugation for 7 min and washed with 1 ml of ethanol. Pellets were dried under vacuum and resuspended in 20µl of T.E. buffer. 2µl aliquots were subjected to gel electrophoresis for comparison of sizes with that of the control.

2.2.19 Sequencing of DNA by the di-deoxynucleotide chain termination method

2.2.19.1 Annealing and labelling reactions

DNA was sequenced according to the method of Sanger *et al* (1977) using the BRL kit and the manufacturers' instructions. The 10mM stocks of dNTPs were diluted to 0.5mM and used for the nucleotide mixes in Table 5.

TABLE 5 Solutions for dideoxy sequencing

Stock	G^o	A^o	T^o	C^o
10 x reaction buffer	20	20	20	20
dGTP)	1	20	20	20
dTTP) 0.5mM	20	20	1	20
dCTP)	20	20	20	1

ddNTP solutions

	vol.10mM stock	vol.H_2O	final concentration
ddGTP	1	19	0.5mM
ddATP	1	79	0.125mM
ddTTP	1	9	1.0mM
ddCTP	1	19	0.5mM

All dilutions were with distilled and deionised water. All volumes are in µl.

For the annealing reactions 5µl of single-stranded template DNA (Section 2.2.18.4) was mixed with 4.5µl water, 2µl M13 15 bp primer and 1µl 10 x polymerase reaction buffer, heated to 85-90°C for 5 min and allowed to cool slowly to room temperature.

1µl DNA polymerase (Klenow fragment, diluted to 1u/µl with the kit dilution buffer), 1µl 0.1M DTT and 1µl ^{35}S-dATP

were added to each template/primer mix. Equal volumes of the appropriate N_o mix and ddNTP were mixed and 2µl aliquots placed in the appropriate G,A,T or C sequencing tube for each template. 3µl of template/primer mix were added to each tube, mixed and tubes incubated at 30°C for 20 min. 1µl of dATP was added to each tube and tubes incubated for a further 15 min at 30°C. 3µl of formamide dyes (0.1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue, 10mM Na_2 EDTA, 95% (v/v) derionised formamide) were added, tubes incubated at 90°C for 5 min, and samples loaded directly onto DNA sequencing gels.

2.2.19.2 Preparation of polyacrylamide gels for DNA sequencing

A solution of 38% acrylamide, 2% bis-acrylamide was stirred for 2h with Amberlite Monobed resin, filtered and stored at 4°C until required. 15ml of this solution was added to 50g urea and 10ml 10x TBE buffer (Section 2.2.4), the volume made up to 100 ml and the mixture heated to dissolve urea. The solution was filtered through cellulose acetate ('Nuflow', 0.22µm, Oxoid Ltd.). 0.4ml 10% ammonium persulphate solution and 20µl of TEMED (N ,N,N´,N´-tetramethylethylene diamine) were added to 50ml of solution, mixed rapidly and poured between two glass plates separated by 0.4mm spacers and sealed by Whatman yellow tape. Combs were inserted and gels left to set for at least 1h at room temperature.

2.2.19.3 Polyacrylamide gel electrophoresis and autoradiography

Combs were removed from gels and tape from the base of the gel. Gels were set up vertically so that

the base dipped into a reservoir of TBE buffer. The top was sealed with silicon grease to a buffer reservoir. Wells were cleared with a syringe and needle, a metal plate clamped across one glass plate and gels electrophoresed at 50mA until the applied voltage reached 1500v (about 2-3h). Samples were loaded in the order G,A,T,C for each template, and were electrophoresed at 50mA, 1500v. One gel was run until the bromophenol blue had just reached the base (2h) and the other until 30 min after the xylene cyanol had reached the base (5h). Gels were removed from the electrophoresis apparatus, one plate lifted off and a piece of Whatman 3MM paper placed over the gel. The position and identification of samples was marked, the paper and gel turned over and the gel surface covered with a sheet of polyethene. Gels were dried under vacuum for several hours to overnight. Autoradiography was as described in Section 2.2.9 except that the film was placed directly in contact with the gel surface.

2.2.19.4 Analysis of DNA sequences

5´to 3´ sequences were read directly from the autoradiographs. The 3´ to 5´ sequence and amino acid sequences of each strand in all three reading frames were obtained using a programme 'Sequence Translate' on a BBC microcomputer. Sequences were compared to those of vicilin cDNAs, and to that of phaseolin, the 7S protein of *Phaseolus vulgaris*, French bean (Slightom *et al*, 1983).

2.2.20 Nuclease sensitivity of pea chromatin

An outline of the methods used to investigate the

nuclease sensitivity of specific genes in pea chromatin is shown in Figure 3.

2.2.20.1 Isolation of nuclei from pea leaves and cotyledons

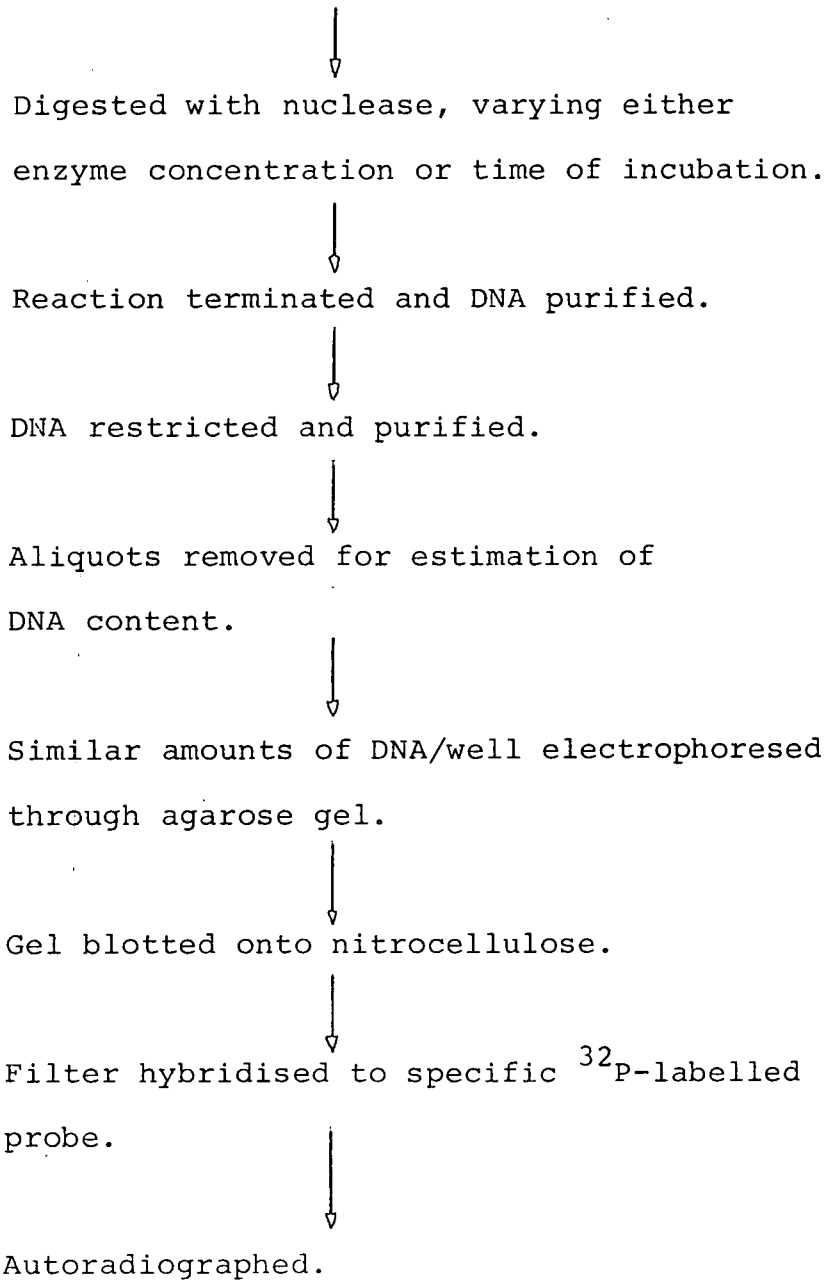
The method was modified from that of Willmitzer and Wagner (1981). A known weight of tissue, typically 4-5g, was used. Leaves were washed in sterile water, cotyledons were separated from the testa and embryonic axis. Tissue was chopped in 5 volumes of buffer A (0.7M mannitol, 10mM MES, 5mM EDTA, 0.1% BSA, 0.2mM PMSF, pH5.8). Chopping was performed on ice and all subsequent operations were carried out at 4°C or on ice except where stated. All centrifugation steps used pre-cooled swing-out rotors.

Chopped tissue was filtered on N125 cloth, ('Simonyl', Henry Simon, Stockport, U.K.), (boiled for 5 min prior to use and allowed to cool) and resuspended in 5-10 volumes of buffer B (buffer A supplemented with 0.05mg/ml purified pectinase, 0.1mg/ml cellulase) which had previously been incubated at 25°C for 15 min. Leaves were incubated for 2-3 h and cotyledons for 4-5h at 25°C.

The suspension was filtered on N53 cloth and washed with buffer C (0.25M sucrose, 10mM NaCl, 10mM MES pH6.0, 5mM EDTA, 0.15mM spermine, 0.5mM spermidine, 20mM mercapto-ethanol, 0.2mM PMSF, 0.6% Nonidet P40, 0.1% BSA). Tissue was resuspended in buffer C and a small volume homogenised in a ground glass homogenizer. 10 strokes were used for leaves and 15 for cotyledons. This was repeated until all tissue had been homogenised. The homogenate was filtered on N125

FIGURE 3

Outline of procedure used to investigate the nuclease sensitivity of genes in pea chromatin.



cloth, and the filtrate from this filtered on N53 cloth.*

The filtrate was centrifuged at 2000 *g* for 5 min at 4°C in the MSE Mistral centrifuge. The pellet was resuspended in a small volume of buffer C, adjusting to a volume equivalent to 3g after addition of 4.5g of buffer D (6g of 5x buffer C, with the Nonidet P40 concentration kept to 0.6%, added to 20.5ml Percoll and pH adjusted to 6.0 with 1N HCl). The solutions were mixed gently and thoroughly and filtered on N53 cloth. The filtrate was centrifuged at 1000 *g* for 5 min at 4°C. The pellet was resuspended in a small vol of buffer C and the supernatant recentrifuged after shaking to abolish density differences. This was repeated 2-3 times. The resuspended pellets were combined and centrifuged at 1000*g* for 5 min at 4°C, the resulting pellet resuspended in a small volume of buffer E (1.2g of 5x buffer C, with the Nonidet P40 concentration at 0.6%, added to 7.97 ml Percoll, pH adjusted to 6.0 with 1N HCl) and centrifuged at 5000 *g* for 5 min at 4°C using the Sorvall RC-5B centrifuge. The floating layer containing nuclei was removed, diluted with buffer C and centrifuged at 1000 *g* for 5 min at 4°C using the MSE Mistral 4L centrifuge. The pellet was washed several times with buffer C by centrifugation at 1000 *g* for 3-4 min at 4°C and the final pellet resuspended in 500µl of buffer C. A small amount of suspension (10µl) was removed, stained with DAPI and examined under the fluorescence microscope (Nikon Diaphot). Nuclei were counted using a haemocytometer (Hawksley, Gallenkamp and Co., Stockton-on-Tees, U.K.), 500µl sterile glycerol added to the remaining suspension and nuclei stored at -80°C. DNA content was determined using the fluoresence assay described below.

* the numbers of the membranes refer to the pore size in microns.

2.2.20.2 Fluorimetric estimations of DNA contents

It was necessary to know the DNA contents of nuclei and DNA preparations so that equal amounts could be used in each digestion. As detection of nuclease sensitivity depended on differences in intensities of bands or autoradiographs (Section 2.2.20.7) it was important that similar amounts of DNA were loaded into each well of the agarose gels. Therefore, DNA contents of nuclease-treated and restricted samples (Section 2.2.20.7) were measured.

DNA contents were estimated by a modification of the method of Thomas and Farquhar (1978). Solutions of calf thymus or herring sperm DNA were made in the range from 20-1000µg/ml. 10µl aliquots were placed in separate tubes. 10µl of nuclei and 90µl of buffer C and 50% (v/v) glycerol (Section 2.2.20.1), or 1µl of nuclease-treated samples and 9µl of water, were placed in separate tubes. All standards and samples, except for nuclei preparations when material was in short supply, were in duplicate. Two blanks containing no DNA were treated in the same way as the samples and standards.

10µl of 5mg/ml BSA were added to each tube and either NaCl added to 150mM (nuclei) or sodium acetate to 50mM (DNA). Samples were precipitated with 250µl ethanol for 30 min at -20°C. Pellets were collected by centrifugation for 10 min in a haematocrit, washed twice with 80% ethanol, dried *in vacuo* and resuspended by vortexing in 20µl of a 400mg/ml solution of DABA. Tubes were incubated at 60°C for 30 min, cooled and 1 ml 1N HCl added. Tubes were vortexed, contents pipetted into a 3ml quartz cuvette and the tube filled with a further 1 ml 1N HCl. This was added to the cuvette, the solution mixed and fluorescence intensity determined at an excitation wavelength of 405nm and an

emission wavelength of 505nm. 2ml of 1N HCl were used to zero the fluorimeter (Baird Atomic Fluoripoint), and the background fluorescence was obtained from the blanks.

A graph of fluorescence intensity against DNA content was plotted for the standards, and used to determine the DNA content of the samples.

2.2.20.3 Storage conditions of nucleases

The buffers used for storage of nucleases were those of Wu *et al* (1979a). DNAaseI was stored at a concentration of at least 1mg/ml in DNAase I digestion buffer (60mM KCl, 15mM NaCl, 15mM Tris/HCl, pH7.4, 0.5mM DTT, 0.25M sucrose, 0.05M $CaCl_2$, 3.0mM $MgCl_2$). MNase was stored at 1mg/ml in MNase digestion buffer (as above but 1.0mM $CaCl_2$ and no $MgCl_2$).

Small aliquots of the stocks were stored at -20°C and thawed as required. Aliquots were subjected to a maximum of 2 freeze-thaw cycles before discarding.

2.2.20.4 Activity of nucleases

The activity of nucleases was determined from the extent of digestion of various substrates. DNA or nuclei were digested for varying times, or with different concentrations of enzyme, the DNA purified and electrophoresed through agarose gels.

Activity was estimated from the extent of digestion of DNA compared to undigested controls, except for the digestion of nuclei by MNase, which was shown by the production of a 'ladder' of DNA fragments whose lengths were

multiples of the length of DNA associated with a nucleosome.

A spectrophotometric assay, which measured the rate of change in A_{260} as highly polymerised DNA was digested into smaller fragments, was also used to estimate DNAase I activity (Kunitz, 1950, Millipore Corporation, 1979). A range of DNAase I solutions, typically 20, 40, 60 and 100u/ml, were prepared from the stock by dilution with 0.15M NaCl. The spectrophotometer (Pye Unicam SP8-150 uv/vis) was adjusted to 25°C and 260nm. 2.5ml of substrate (10mg highly polymerised herring sperm DNA dissolved in 200ml of 6.25mM $MgSO_4$, 25ml of 1.0M acetate buffer added and made up to 250ml with sterile distilled water) was pipetted into each of 2 cuvettes and incubated in the spectrophotometer to establish the blank rate. 0.5ml of 0.15M NaCl was added to the reference cuvette, and 0.5ml of enzyme to the sample cuvette. The contents were mixed rapidly and the change in absorbance at 260nm monitored for 8-10min. The rate of change in A_{260} was calculated from the linear part of the trace. The rate of change/ml of assay mixture was calculated for each dilution and divided by the number of units/ml to obtain the activity (ΔA_{260}/min/unit). The mean activity was calculated for each stock, and the differences between the means were compared using a small-scale statistical test (Student t test). The test assumed that the samples were drawn from populations with a common variance; this assumption was justified for this assay as the main errors were in the dilutions and in the measurement of rates from the traces.

2.2.20.5 Digestion of DNA and nuclei with nucleases

Nuclei suspensions were washed in 5-10 vol of digestion buffer by centrifugation at 1000 *g* for 5 min at 4°C, and the pellet resuspended in this buffer to a DNA concentration of 0.5µg/ml. 15µg of DNA was routinely used for each digestion, in a final volume of 40-150µl (samples to be compared were always digested in similar volumes). 0.2 vol of 5x digestion buffer was added to each sample and the volume made up with water. Digestions were begun by addition of enzyme, and were incubated at 25°C (DNAase I and MNase) or 12-15°C (DNAase I) for the appropriate time. Reactions were terminated by addition of EDTA to 12mM and of SDS to 0.5%, the tubes placed on ice and DNA purified as described below.

Protein-free DNA was digested as for nuclei except that the initial washing step was omitted and usually replaced by ethanol precipitation and resuspension in distilled water.

For each set of digestions a control tube containing digestion buffer in place of enzyme was incubated for the same length of time as the longest digestion, then treated in the same way as the digestions.

2.2.20.6 Purification of DNA from nuclease-treated samples

A modification of the method of Stalder *et al* (1979) was used. RNAase A was added to the terminated digestions and controls to a concentration of 0.1mg/ml, and tubes incubated at 37°C for 30 min. SDS was added to 0.5%,

NaCl to 0.4M and proteinase K to 1mg/ml. Tubes were incubated at 37°C for at least 2h. 0.5 vol of water was added and samples were extracted with phenol (Section 2.2.1.1), re-extracting the organic phases with distilled water. DNA was precipitated with ethanol and re-dissolved in distilled water.

2.2.20.7 Restriction of samples and detection of nuclease sensitivity

DNA samples from nuclease-treatments and controls were restricted (Section 2.2.2.1) with appropriate enzymes. Reactions were terminated by phenol extraction and DNA precipitated with ethanol and dissolved in distilled water. Aliquots were taken for estimation of DNA content (Section 2.2.20.2) and similar amounts of DNA loaded into each well of an agarose gel, electrophoresed (Section 2.2.4) and blotted onto nitrocellulose (Section 2.2.5). Filters were hybridised to specific ^{32}P-labelled probes and autoradiographed as described in Sections 2.2.8 and 2.2.9. Nuclease sensitivity was determined by comparing the intensities of the bands for active sequences (*e.g.* seed storage protein genes in cotyledons) with those for inactive sequences (*e.g.* these genes in leaves). Hypersensitive sites, if present, were detected by the presence, at low levels of digestion, of bands in addition to the usual restriction fragments.

CHAPTER THREE

RESULTS

3.1 The isolation and characterisation of a pea vicilin gene

3.1.1 Screening of the pea genomic library

The library of pea genomic DNA in the vector EMBL3 was screened by hybridisation of plaques to the insert of pDUB2 as described in Section 2.2.10. A single plaque was observed to hybridise to pDUB2 on autoradiographs of plaque lifts from the 1st screen. Figure 4 shows an autoradiograph of a plaque-lift from the 6th and final screen. The phage at this stage were essentially pure, as greater than 90% of the plaques hybridised to the probe. As only one plaque hybridised in the 1st screen, all of the plaques hybridising in the final screen were derived from this plaque and therefore contained similar recombinant bacteriophage.

3.1.2 Titration of stocks of bacteriophage prepared from individual plaques

Stocks of recombinant bacteriophage made by the plate lysate method (Section 2.2.11.1) from the plaques isolated from the 6th screen had a titre of approximately 1.6×10^9 pfu/ml for each extraction. Sufficient bacteriophage were therefore available for the transfections in subsequent DNA preparations.

3.1.3 Preparation of DNA from recombinant bacteriophage

The recombinant bacteriophage which contained a pea genomic sequence hybridising to pDUB2 was designated λvic 1. In order to characterise the vicilin gene in λvic 1, it was necessary to obtain a restriction map of the clone, and

FIGURE 4

Autoradiograph of a plaque lift from the final screen of the genomic library

The filter was hybridised to the labelled insert of pDUB2, and washed to a higher stringency (1xSSC, 2x15 min. at 65^{o}C) than for other plaque lifts (Section 2.2.8). The film was exposed to the filter overnight at -80^{o}C.

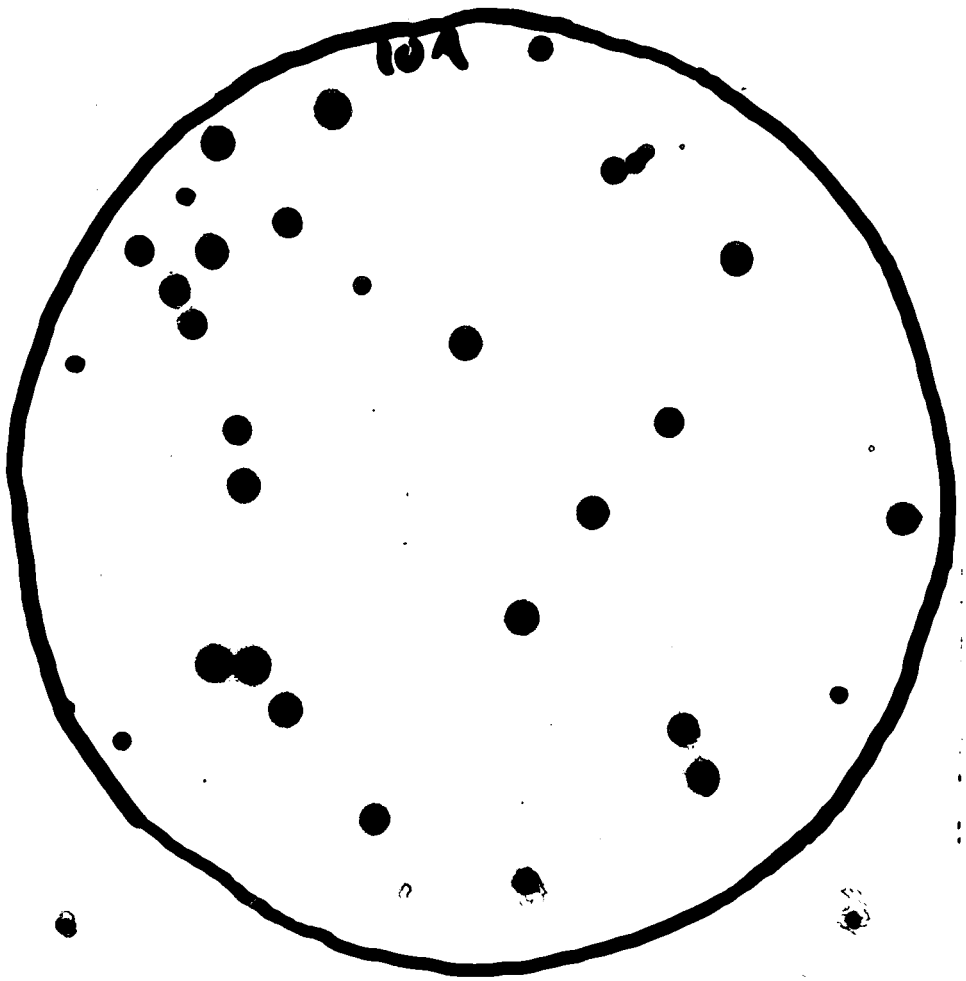


Fig. 4

to sub-clone fragments into bacterial plasmids. For these procedures a large quantity of DNA of sufficient purity to be cut by restriction endonucleases was required. A number of different methods of bacteriophage DNA preparation were therefore tested for yield and quality. The results of small-scale preparations are given in the following Table.

TABLE 6 Assessment of small-scale bacteriophage DNA preparations

Method	Yield of DNA	Restriction Endonuclease Digestion	Degradation
1. Small-scale plate lysates	2 µg/$2x10^5$ pfu	Poor	No
2. Small-scale liquid-lysates	10µg/$5x10^6$ pfu	Poor	*Yes

* Incubation of a tube without exogenous enzymes at 37°C for 2h resulted in degradation. Neither of the small-scale methods were therefore satisfactory.

Two methods of large-scale DNA preparation were used. When the method described in Section 2.2.12 was followed, the yield of DNA was 130µg from 5 ml of PEG precipitate, which contained $4x10^{11}$ pfu/ml. The DNA was cut into discrete fragments by restriction endonucleases (Figure 5) and there was no visible degradation. Omitting the chloroform extraction prior to the phenol extractions in this method resulted in a very poor yield of DNA, although titration of an aliquot of the PEG precipitate gave a value of $1.3x10^{13}$ pfu/l of culture. When aliquots of this PEG precipitate were purified as for the small-scale liquid lysates, approximately 15µg of DNA was obtained from $1.8x10^{11}$ pfu. The DNA was

FIGURE 5

Restriction digestions of recombinant phage DNA purified from large-scale PEG precipitates by chloroform and phenol extractions

Tracks (a) and (j), λNM258 restricted with Hind III/Eco RI; tracks (b) to (i) recombinant phage DNA restricted with (b) Eco RI/Sal I, (c) Hind III/Sal I, (d) Hind III, (e) Bam HI/Hind III, (f) Bam HI/Sal I, (g) Sal I/Bgl II, (h) Sal I, (i) Bgl II. Numbers on the right are size-markers (kb).

cut by restriction endonucleases as shown in Figure 6A. When the precipitates were purified in this way after 20 days of storage, the DNA was partially degraded during restriction (Figure 6B). The most reliable method of DNA preparation, with a good yield of purified DNA, was therefore that described in Section 2.2.12.

3.1.4 Characterisation of the vicilin gene in λvic 1.

3.1.4.1 Restriction mapping of λvic 1

A restriction map of the bacteriophage vector EMBL3 is shown in Figure 7. In order to map the pea sequence inserted into λvic 1, restriction digestions such as those in Figure 5 were performed on λvic 1 DNA. The restriction sites known to be present in the arms of the vector (Frischauf *et al*, 1983) were also used in the construction of the map from the restriction data. The location of the vicilin gene, referred to as *vic A* hereafter, was determined from the results of hybridisation of blots of gels such as that in Figure 5, to inserts of pDUB2 or pDUB9. The approximate sizes of fragments obtained from restriction digestions of λvic 1, and the results of hybridisation experiments, are summarised in the Table below.

The hybridisation of the probe to the large fragments (>20kb) extending from the left arm into the insert in several digests (*e.g.* Hind III, Bam HI, Eco RI) suggested that the vicilin gene was close to the left arm. This was confirmed by the loss of hybridisation to this fragment in double digests of these enzymes with Sal I, which has

FIGURE 6

Restriction digestions of recombinant phage DNA purified from a large-scale PEG precipitate by the potassium acetate / DEP method

A. Immediate purification of precipitate.

Track (a) λNM258 restricted with Hind III/ Eco RI.

Tracks (b) to (g) recombinant phage DNA restricted with (b) Xho I (c) Hind III (d) Bam HI/Hind III (e) Bam HI, (f) Eco RI/ Bam HI, (g) Eco RI.

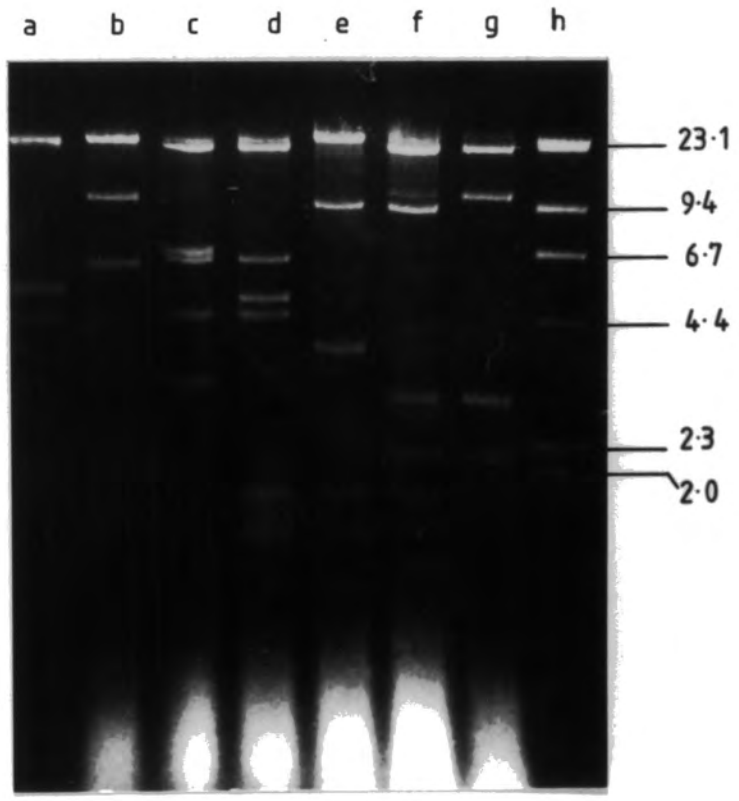
Track (h) λNM258 restricted with Hind III.

Numbers at the right-hand side of the photograph are size-markers in kb.

B. Purification of precipitate after 20 days' storage at $4^{o}C$.

Tracks (a) to (h) recombinant phage DNA restricted with (a) Eco RI/Sal I, (b) Bgl II, (c) Sal I/Bgl II, (d) Sal I, (e) Hind III/ Sal I, (f) Hind III, (g) Bam HI/Hind III, (h) Bam HI/Sal I.

A



B

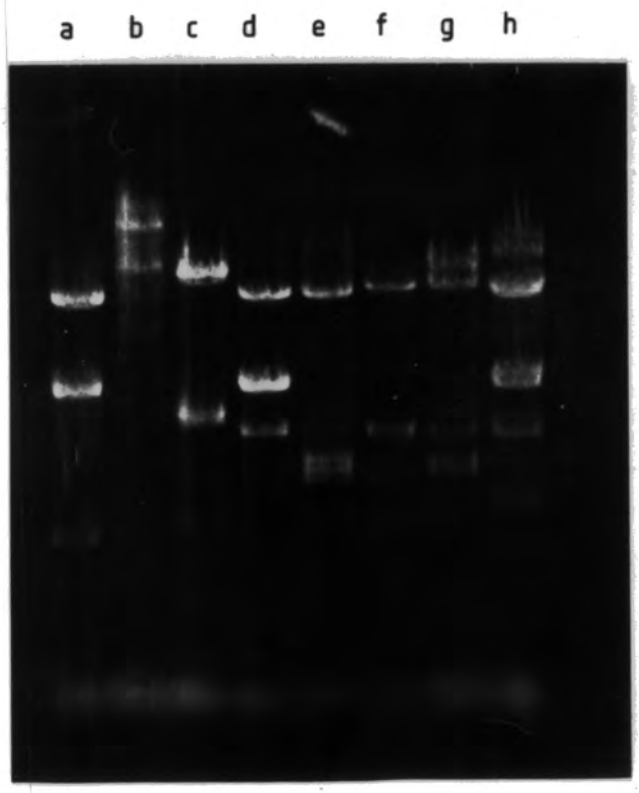


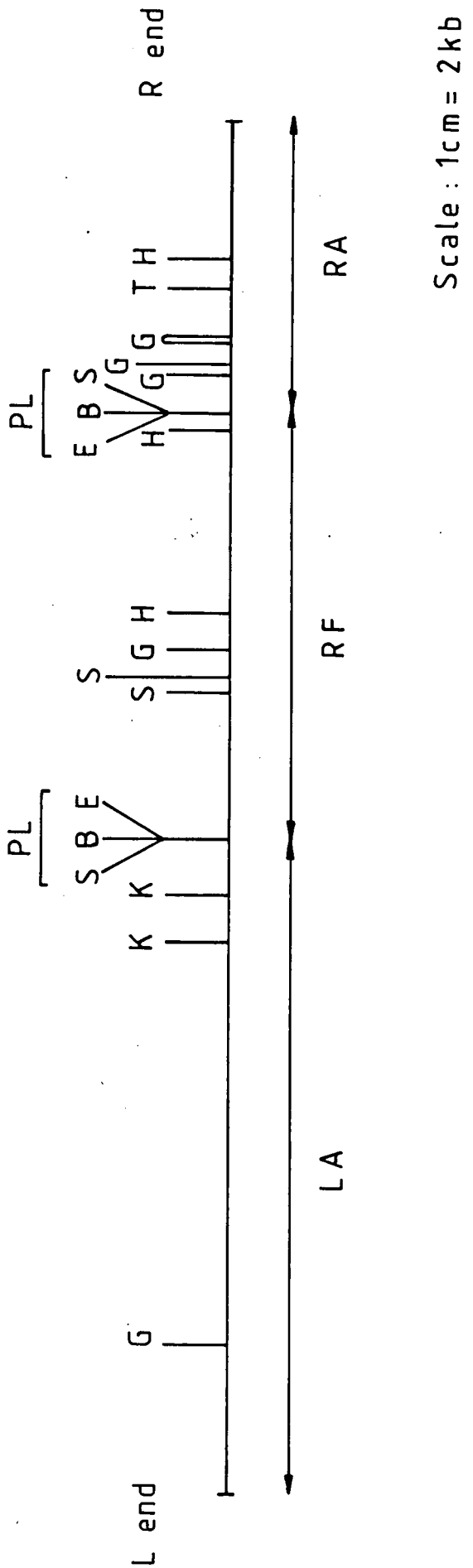
Fig. 6

FIGURE 7

Restriction map of the bacteriophage vector EMBL3 (Frischauf *et al*, 1983)

B = Bam HI, E = Eco RI, H = Hind III, G = Bgl II, K = Kpn I, S = Sal I, T = Sst I, LA = left arm of vector, RA = right arm, RF = replacement fragment, PL = polylinker.

In the pea genomic library, the fragment between the polylinkers was replaced by 15-20 kb fragments from a Sau 3A digest of pea DNA, ligated into the Bam HI sites of the polylinkers.



Scale : 1cm = 2kb

Fig. 7

TABLE 7 Restriction digestions of λvic 1

Digest	Band sizes (kb)
Xho I	24.5, 10.5, 5.8 (not hybridised)
Hind III	*20.5, 6.7, *6.2, 4.3, 2.9, 1.7, 1.45
Bam HI	*24.5, 9.4, 3.5, 1.95, 1.7
Eco RI	*20.5, 11.0, 2.85, 2.25, 1.7
Bgl II	*21.5, 7.2, 3.1, 2.6, 1.5
Sal I	20.5, *6.6, 9.5
Eco RI/Bam HI	*20.6, 9.4, 2.85, 2.25, 1.9, 1.7, 1.5
Eco RI/Sal I	20.5, 9.4, 2.65, *1.85 1.55, 1.35
Bam HI/Hind III	*20.5, *6.2, 4.8, 4.3, 1.75, 1.45
Hind III/Sal I	20.5, *5.2, 4.7, 4.2, 2.85, 1.71, *1.45
Bam HI/Sal I	20.5, 10.1, *6.3, 3.5, 2.0, 1.8, 1.36
Bgl II/Sal I	20.5, 7.2, *4.7, 3.1, 2.6, 1.5, 1.2

* Indicates hybridisation to the inserts of pDUB2 and to pDUB9 - no difference was observed in the patterns of hybridisation obtained with the two probes.

---

a site in the polylinker at either end of the insert (see Figure 7). When the size of the insert was calculated, considerable variation was seen. For example, the size of the insert calculated from a Hind III digest was 14.45 kb, that from the Bam HI digests 11.4 kb, while that from a Sal I digest was only 6.6 kb. Part of the difference could be accounted for by the obscuring of small fragments on the gels by RNA, and by the difficulty of determining precisely the sizes of the large (>20kb) fragments. However, there were too many inconsistencies for these to explain all the differences.

In order to construct a more complete restriction map, therefore, fragments of λvic 1 were sub-cloned into bacterial plasmids.

3.1.4.2 Sub-cloning of λvic 1 fragments into bacterial plasmids

The Sal I digest of λvic 1 suggested an insert size of only 6.6 kb. This size of insert would result in a phage genome of only 36.1 kb, which is not large enough to be packaged by the *in vitro* packaging system used in the construction of the gene library. A set of sub-clones of a Sal I digest of λvic 1 was therefore constructed in pUC8 (Vieira and Messing, 1982), as shown in Figure 8.

When small-scale preparations of plasmid DNA (Section 2.2.16) were made from the transformants (Section 2.2.14.3) and restricted with Sal I, two clones, S4 and S23, contained inserts. That of S4 was 6.4 kb in size and hybridised to pDUB2 and pDUB9 inserts. The insert of S23 was 9.2 kb and did not hybridise to either probe. The 6.4 kb and 9.2 kb fragments were assigned to the map positions in λvic 1 shown in Figure 8.

In order to locate the *vic A* gene more precisely, and to construct a more detailed restriction map of λvic 1, the two Sal I sub-clones S4 and S23 were digested with a variety of restriction endonucleases. The results of such digestions on the whole plasmids and on isolated inserts, and of hybridisations to inserts of pDUB2 or pDUB9 were used to construct the maps shown in Figure 9.

FIGURE 8

Construction of pUC8 sub-clones from a Sal I digest of λvic 1.

E = Eco RI, A = Ava I, M = Sma I, B = Bam HI,

I = Hind II, C = Acc I, S = Sal I, P = Pst I,

H = Hind III.

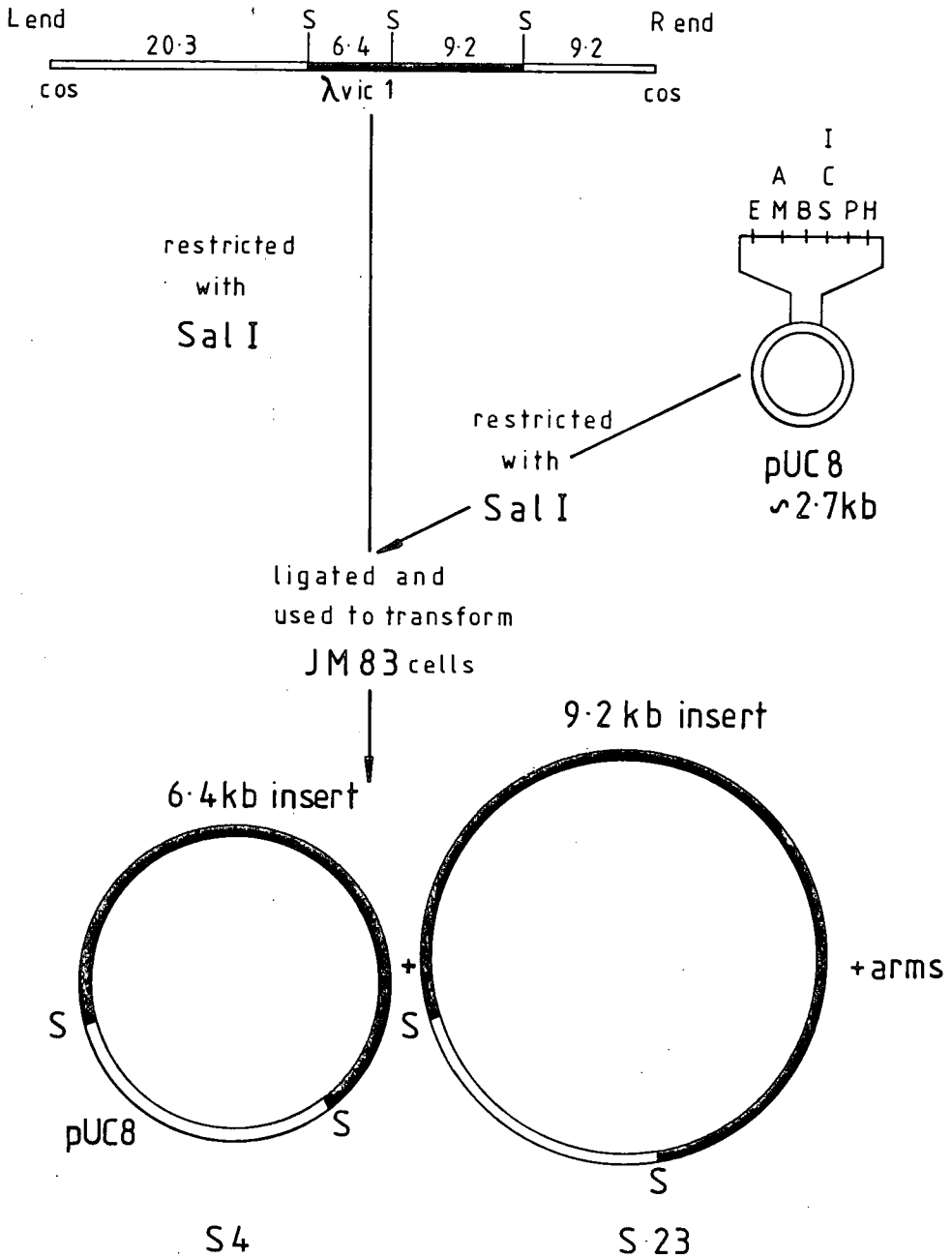


Fig. 8

Not to scale

FIGURE 9

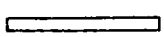
Restriction maps of Sal I sub-clones in pUC8

E = Eco RI, B = Bam HI, S = Sal I, X = Xba I,

P = Pst I, G = Bgl II, H = Hind III.

L = Sal I site at left-arm end of λvic I.

R = Sal I site at right-arm end of λvic I.

 = pUC 8  = pea DNA

 = region of hybridisation to vicilin cDNAs.

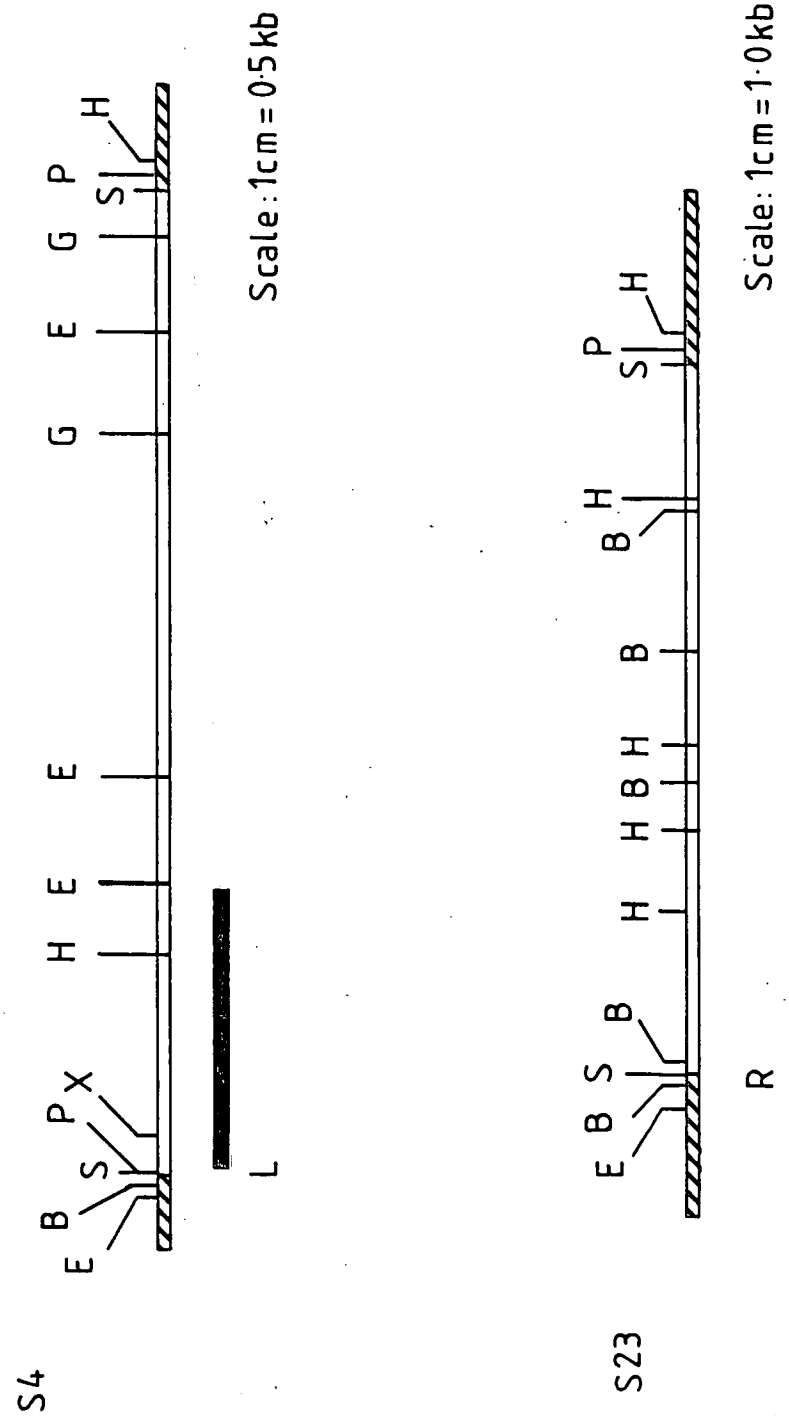

S4
E
B
S
P
X
H
E
E
G
E
G
P
S
H
L
Scale: 1cm = 0·5kb
S23
E
B
S
B
H
H
B
H
B
H
B
S
P
H
R
Scale: 1cm = 1·0kb

Fig. 9

The EcoRI/Sal I fragment of 1.85 kb present in S4 was the only EcoRI/Sal I fragment to hybridise to either of the vicilin cDNAs. In order to obtain a sub-clone containing this fragment, the products of an EcoRI/Sal I digestion of λvic 1 were ligated into pUC8 cut with EcoRI or EcoRI/Sal I, and used to transform competent cells (Figure 10). The transformed cells were screened by colony hybridisation (Section 2.2.15) to the insert of pDUB2. Several hybridising colonies were picked off the plates and used to prepare small amounts of plasmid DNA (Section 2.2.16). When this DNA was restricted with EcoRI/Sal I, 9 clones contained the 1.85 kb fragment which hybridised to the insert of pDUB 2. Several of the clones reverted to wild-type blue colonies (see Section 2.2.14.3). One sub-clone, ES 10, which remained white when streaked onto fresh plates, was chosen for further characterisation. This clone was derived from the ligation into EcoRI cut pUC8 and contained the 1.85 kb EcoRI/Sal I fragment ligated to a 1.3 kb EcoRI/Sal I fragment, as shown in Figure 11. The clone was mapped using restriction endonuclease digestions and hybridisation to pDUB2 and pDUB9 inserts, and the map is shown in Figure 11.

The final restriction map constructed for λvic 1 is shown in Figure 12.

3.1.4.3 DNA sequencing of *vic A*

As the EcoRI/Sal I 1.85 kb fragment of λvic 1 was the only fragment in such a digest to hybridise to the vicilin cDNAs it was chosen for DNA sequencing. A set of

FIGURE 10

Construction of sub-clones in pUC8 from an Eco RI/ Sal I digest of λvic I

Symbols are as in Figure 8.

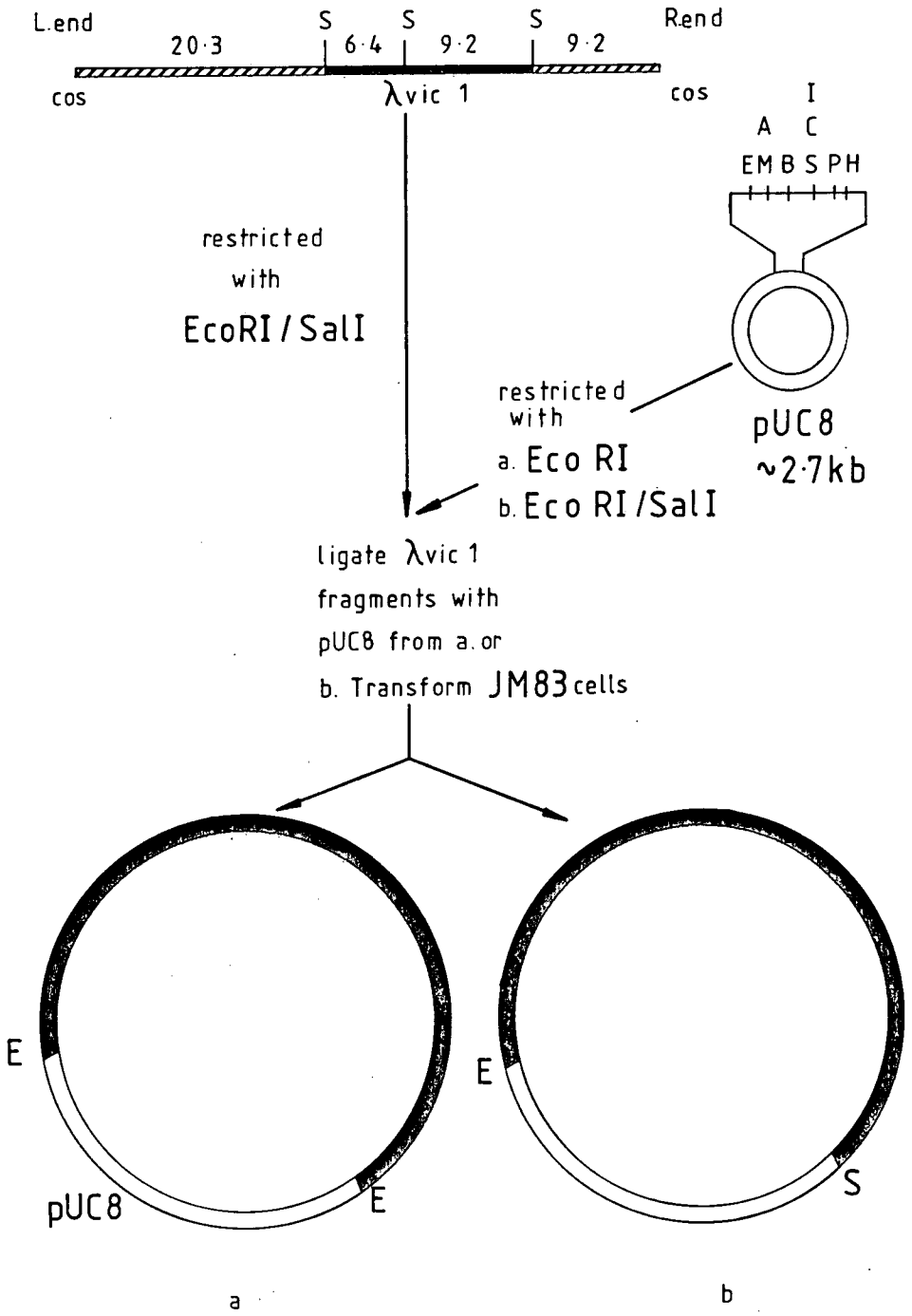


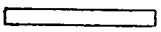
Fig. 10

Not to scale

FIGURE 11

Restriction map of the Eco RI/Sal I sub-clone, ES 10

Symbols are as in Figure 9.

 = pUC8  = pea DNA

 = region of hybridisation to vicilin cDNAs.

The Xba I site is now known to be ∿250 bp from the Sal I site

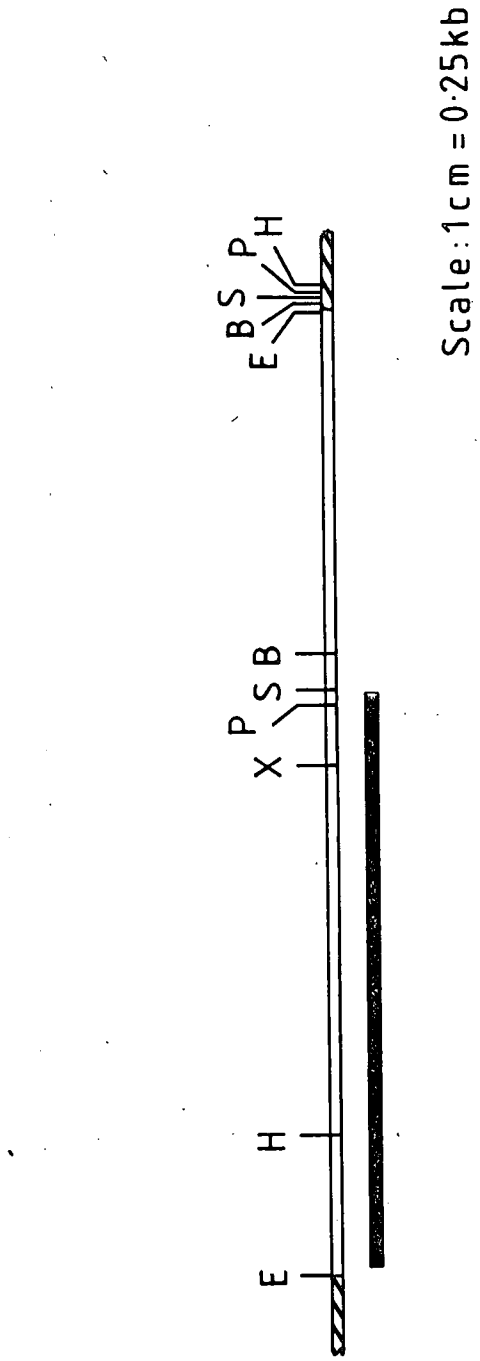


Fig. 11

FIGURE 12

Restriction map of λvic 1

The map was based on the details of the various sub-clones, and the restriction and hybridisation data for λvic 1.

Symbols are as in Figures 7 to 9.

 = EMBL3  = pea genomic DNA.

Expanded region is the 1.85 kb fragment subsequently used for DNA sequencing.

The 9.2 kb Sal I fragment has not been mapped in as great detail as the 6.6 kb fragment; thus only the Bam HI and Hind III sites are shown in this region of λvic 1. Scale of λvic 1 map is 1 cm ≈1 kb; scale of expanded region is 1 cm ≈100 bp.

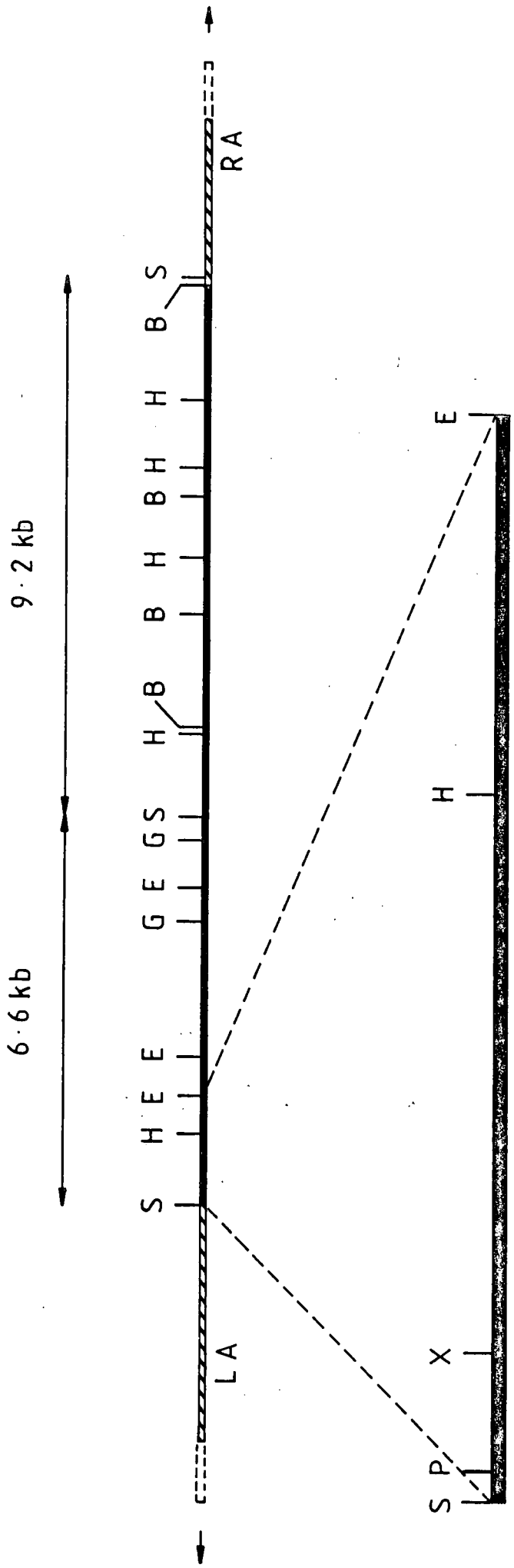


Fig. 12

sub-clones was constructed from the products of an EcoRI/Sal I/ Hind III digest of ES 10 (Section 3.14.2) ligated into either EcoRI/Hind III or SalI/Hind III cut M13 mp18 (Section 2.2.18). Transformants were selected by colony hybridisation to pDUB2 (Section 2.2.18.3). Two clones, 48 (Sal I/Hind III) and 15 (EcoRI/Hind III) hybridised very strongly under high stringency conditions (0.1xSSC, 0.1% SDS, 68°C). When single-stranded DNA prepared from these clones was electrophoresed through an agarose gel, 48 was considerably larger than the M13 control, and 15 was of a similar size to the control. Both of these clones were used as templates for DNA sequencing, as shown in Figure 13.

Further sets of sub-clones were made from Sau 3A digests of the 1.85 kb EcoRI/Sal I fragment, ligated into Bam HI cut mp19, and from XbaI/EcoRI digests of S4 (Section 3.1.4.2), ligated into XbaI/EcoRI cut mp18. DNA was purified and sequenced as before and the directions of sequencing from the various sites are shown in Figure 13, together with a partial map of the introns and exons of *vic A*. The partial sequence of *vic A* is shown in Figure 14, and is compared with the nucleotide and predicted amino acid sequences of pDUB9 in Figure 15. There is a close match over corresponding regions between the two sequences, confirming the hybridisation data. This shows that the *vic A* gene encodes a 50 K M_r vicilin polypeptide. The gene isolated is incomplete; comparison of the position of the Xba I site in the cDNA with that in the genomic clone shows that some of the 5′ coding sequence must be missing from the latter (Figure 13). The sequence is discussed in detail in Chapter Four.

FIGURE 13

Sequencing strategy for *vic A*

The restriction map of *vic A* shows the sites used in DNA sequencing, and the arrows below it show the directions of sequencing. The approximate positions of the introns and exons so far determined for *vic A* are also shown, as is a restriction map of pDUB9.

IVS = intervening sequence, S = Sau 3A,

X = Xba I, H = Hind III, G = Bgl II, E = Eco RI.

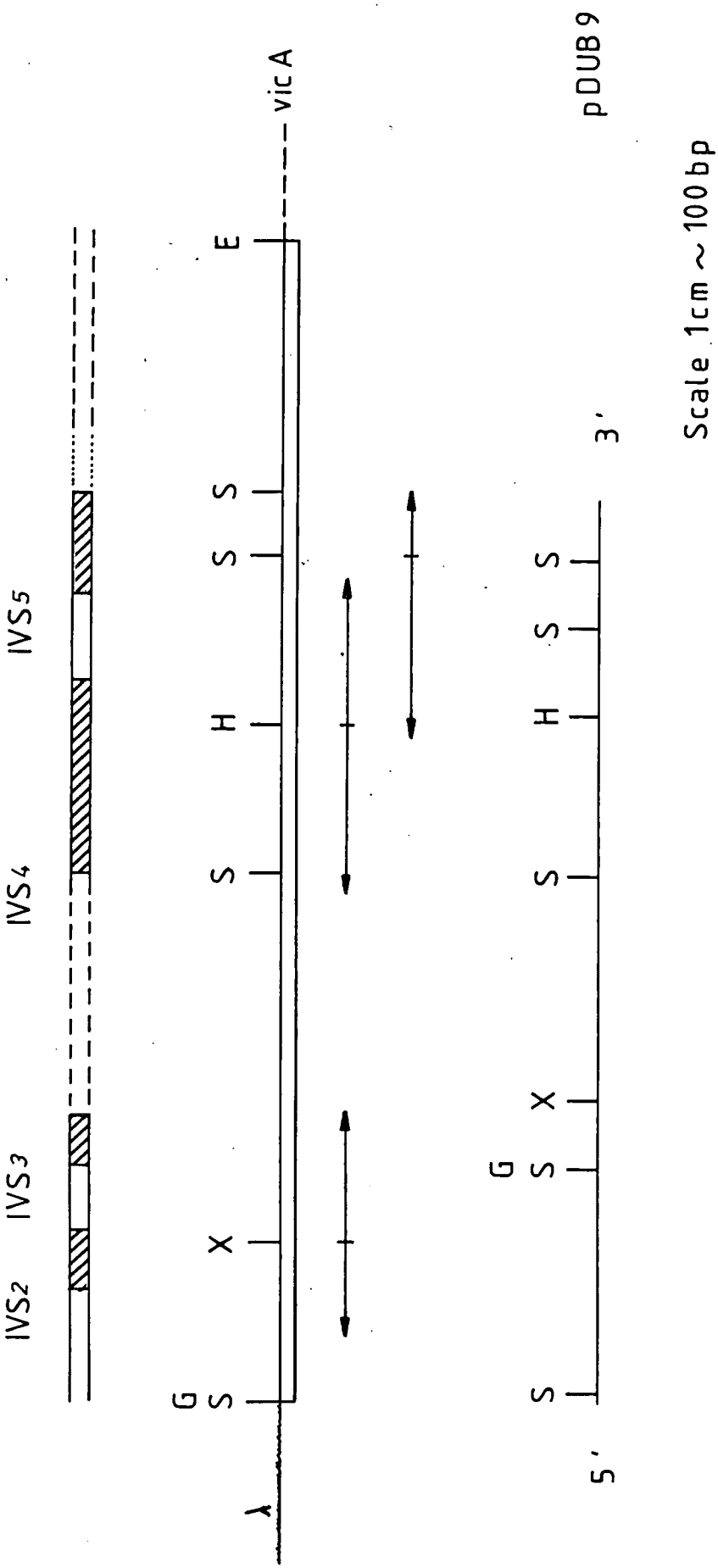

IVS2
IVS3
IVS4
IVS5
vicA
pDUB9
Scale 1cm ~ 100 bp
5'
3'
λ
G
S
X
S
H
S
S
E
G
S
X
S
H
S
S
S

Fig. 13

FIGURE 14

Partial nucleotide sequence of *vic A*

The sequences shown are those obtained as shown in Figure 13.

(1) Sequence from around the Xba I site.

(2) Sequence from around the Hind III and Sau 3A sites.

Figure 14

1.

```
VicA GTTTAGGGGCCTCTACCATAACATCACAACCAAAATTCGACTGTACATATTTTTTTTGTAATGTGTTCCTAATATTTTCTTATTTTTTTGTCCTTTTCATCAGTCTTTCTTATTGTCTGG
A.A. ..........................................................................IVS-2> S  F  L  L  S  G

                                                  Xba I
VicA AAATCAAAACCAACAATCCTACTTATCTGGGTTTAGTAAGAACATTCTAGAGGCTTCCTTCAATGTAAGTATAAAACACACAATTTTTTTTCATTTTTGTATGTATTAGTTTGGTATTGT
A.A.  N  Q  N  Q  Q  S  Y  L  S  G  F  S  K  N  I  L  E  A  S  F  N <IVS-3.....................................

VicA ATATGTTAATGCTCACTTTGTCAATGTATGTATTGTAAAAAAACACAGACTGATTATGAAGAGACAGAAAAGGTTCTTTTAGAAGAGCATGAGAAAGAGACACAACACAGAAGAAGCCTT
A.A. ..........................................IVS-3> T  D  Y  E  E  T  E  K  V  L  L  E  E  H  E  K  E  T  Q  H  R  R  S  L

VicA AAGGATAAGGA...
A.A.  K  D  K
```

2.

```
VicA ..................................................................................TCTTCAAATTAATATAGG
A.A. ..................................................................................IVS-4>

VicA GATCTATATTGTTGCCACACTACAATTCAAGGGCCATAGTAATAGTAACAGTTAACGAAGGAAAAGGAGATTTTGAACTTGTGGGTCAAAGGAATGAAAACCAACAAGAGCAGAGGAAAG
A.A. G  S  I  L  L  P  H  Y  N  S  R  A  I  V  I  V  T  V  N  E  G  K  G  D  F  E  L  V  G  Q  R  N  E  N  Q  Q  E  Q  R  K

VicA AAGTTGACGAGGAAGAAGAACAAGGAGAAGAGGAGATAAATAAACAAGTGCAAAATTACAAAGCTAAATTGTCTTCAGGAGATGTTTTTGTGATTCCAGCAGGCCATCCAGTTGCCGTAA
A.A. E  V  D  E  E  E  E  Q  G  E  E  E  I  N  K  Q  V  Q  N  Y  K  A  K  L  S  S  G  D  V  F  V  I  P  A  G  H  P  V  A  V

     Hind III
VicA AAGCTTCCTCAAATCTTGATTTGCTTGGGTTTGGTATTAATGCTGAGAACAATCAGAGGAACTTTCTTGCAGGTATATTATATTATCACCCAGTCTCTGTCACTATTTATTCATTTTAAG
A.A. K  A  S  S  N  L  D  L  L  G  F  G  I  N  A  E  N  N  Q  R  N  F  L  A <IVS-5.....................................

VicA TGTGTATTTTAAAAGCAACTTCTATTTAAATCAAGGGGAAAATATTAAGATATGCTTATTATTTTGGTGATTGAAAATTTGAAGGCGATGAGGATAATGTGATTAGTCAGATACAGCGAC
A.A. ..........................................................................IVS-5>G  D  E  D  N  V  I  S  Q  I  Q  R

VicA CAGTGAAAGAGCTTGCATTCCCTGGATCAGCTCAAGAGGTTGATAGGATACTAGAGAATCAGAAACAATCTCACTTTACAGATGCTCAACCTCAACAAAGGGAGAGAGGAAGTCGTGAAA
A.A. P  V  K  E  L  A  F  P  G  S  A  Q  E  V  D  R  I  L  E  N  Q  K  Q  S  H  F  A  D  A  Q  P  Q  Q  R  E  R  G  S  R  E

VicA CAAGAGATC.....................................................................................
A.A. T  R  D
```

FIGURE 15

Comparison of nucleotide sequences of *vic A* and pDUB9

The partial nucleotide sequence of *vic A* is shown above that of the cDNA pDUB9. The amino-acid sequence shown is that predicted from pDUB9; differences in the amino-acid sequence predicted by *vic A* are discussed in Chapter Four.

IVS = position of intervening sequence in *vic A*

The vertical arrow indicates the position of the β/γ cleavage site in a 47,000 M_r vicilin polypeptide.

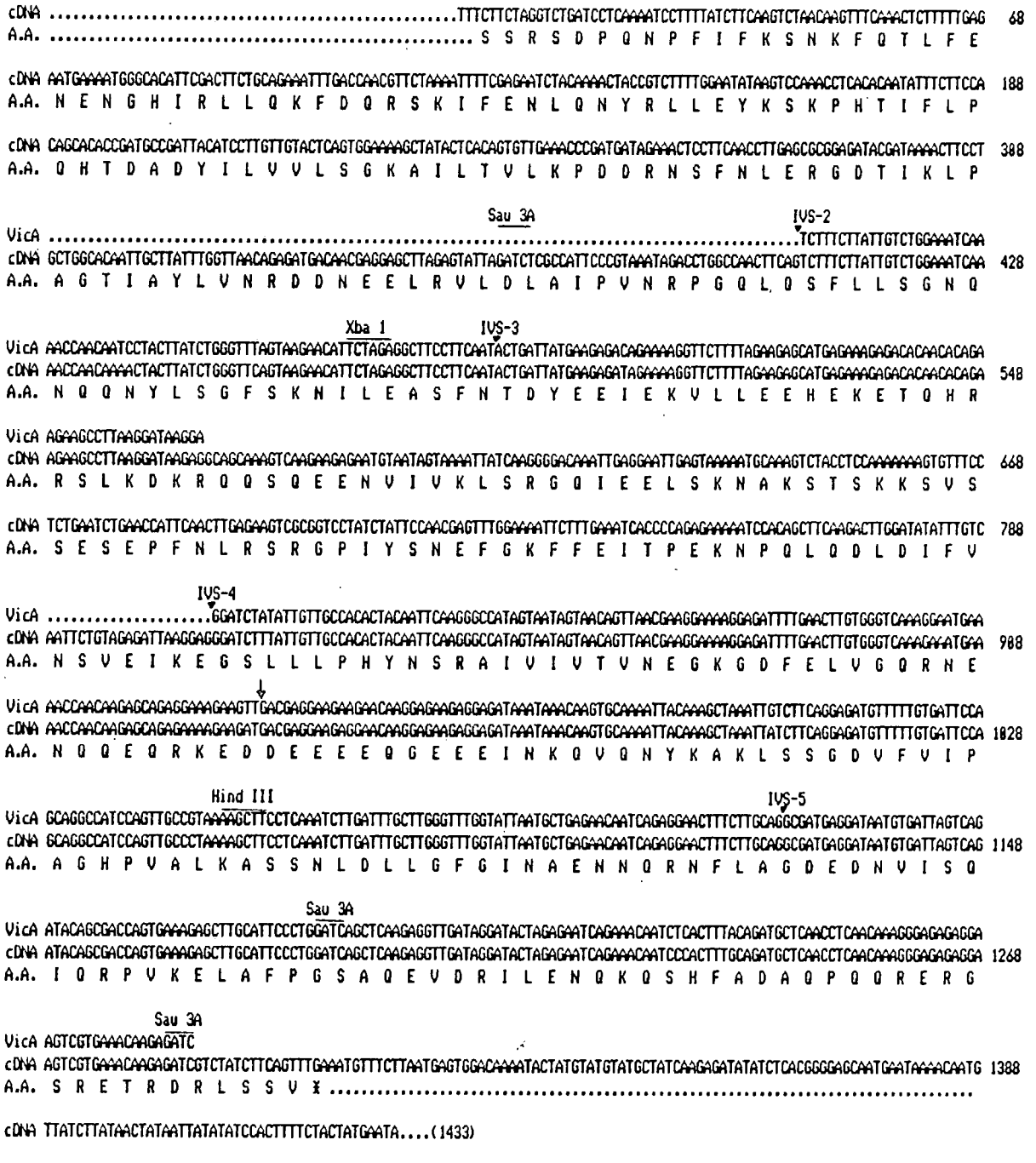

cDNA .....TTTCTTCTAGGTCTGATCCTCAAAATCCTTTTATCTTCAAGTCTAACAAGTTTCAAACTCTTTTTGAG 68
A.A. ..... S S R S D P Q N P F I F K S N K F Q T L F E
cDNA AATGAAAATGGGCACATTCGACTTCTGCAGAAATTTGACCAACGTTCTAAAATTTTCGAGAATCTACAAAACTACCGTCTTTTGGAATATAAGTCCAAACCTCACACAATATTTCTTCCA 188
A.A. N E N G H I R L L Q K F D Q R S K I F E N L Q N Y R L L E Y K S K P H T I F L P
cDNA CAGCACACCGATGCCGATTACATCCTTGTTGTACTCAGTGGAAAAGCTATACTCACAGTGTTGAAACCCGATGATAGAAACTCCTTCAACCTTGAGCGCGGAGATACGATAAAACTTCCT 388
A.A. Q H T D A D Y I L V V L S G K A I L T V L K P D D R N S F N L E R G D T I K L P
Sau 3A
IVS-2
VicA .....TCTTTCTTATTGTCTGGAAATCAA
cDNA GCTGGCACAATTGCTTATTTGGTTAACAGAGATGACAACGAGGAGCTTAGAGTATTAGATCTCGCCATTCCCGTAAATAGACCTGGCCAACTTCAGTCTTTCTTATTGTCTGGAAATCAA 428
A.A. A G T I A Y L V N R D D N E E L R V L D L A I P V N R P G Q L Q S F L L S G N Q
Xba 1
IVS-3
VicA AACCAACAATCCTACTTATCTGGGTTTAGTAAGAACATTCTAGAGGCTTCCTTCAATACTGATTATGAAGAGACAGAAAAGGTTCTTTTAGAAGAGCATGAGAAAGAGACACAACACAGA
cDNA AACCAACAAAACTACTTATCTGGGTTCAGTAAGAACATTCTAGAGGCTTCCTTCAATACTGATTATGAAGAGATAGAAAAGGTTCTTTTAGAAGAGCATGAGAAAGAGACACAACACAGA 548
A.A. N Q Q N Y L S G F S K N I L E A S F N T D Y E E I E K V L L E E H E K E T Q H R
VicA AGAAGCCTTAAGGATAAGGA
cDNA AGAAGCCTTAAGGATAAGAGGCAGCAAAGTCAAGAAGAGAATGTAATAGTAAAATTATCAAGGGGACAAATTGAGGAATTGAGTAAAAATGCAAAGTCTACCTCCAAAAAAAGTGTTTCC 668
A.A. R S L K D K R Q Q S Q E E N V I V K L S R G Q I E E L S K N A K S T S K K S V S
cDNA TCTGAATCTGAACCATTCAACTTGAGAAGTCGCGGTCCTATCTATTCCAACGAGTTTGGAAAATTCTTTGAAATCACCCCAGAGAAAAATCCAACAGCTTCAAGACTTGGATATATTTGTC 788
A.A. S E S E P F N L R S R G P I Y S N E F G K F F E I T P E K N P Q L Q D L D I F V
IVS-4
VicA .....GGATCTATATTGTTGCCACACTACAATTCAAGGGCCATAGTAATAGTAACAGTTAACGAAGGAAAAGGAGATTTTGAACTTGTGGGTCAAAGGAATGAA
cDNA AATTCTGTAGAGATTAAGGAGGGATCTTTATTGTTGCCACACTACAATTCAAGGGCCATAGTAATAGTAACAGTTAACGAAGGAAAAGGAGATTTTGAACTTGTGGGTCAAAGAAATGAA 908
A.A. N S V E I K E G S L L L P H Y N S R A I V I V T V N E G K G D F E L V G Q R N E
VicA AACCAACAAGAGCAGAGGAAAGAAGTTGACGAGGAAGAAGAACAAGGAGAAGAGGAGATAAATAAACAAGTGCAAAATTACAAAGCTAAATTGTCTTCAGGAGATGTTTTTGTGATTCCA
cDNA AACCAACAAGAGCAGAGAAAAGAAGATGACGAGGAAGAGGAACAAGGAGAAGAGGAGATAAATAAACAAGTGCAAAATTACAAAGCTAAATTATCTTCAGGAGATGTTTTTGTGATTCCA 1028
A.A. N Q Q E Q R K E D D E E E E Q G E E E I N K Q V Q N Y K A K L S S G D V F V I P
Hind III
IVS-5
VicA GCAGGCCATCCAGTTGCCGTAAAAGCTTCCTCAAATCTTGATTTGCTTGGGTTTGGTATTAATGCTGAGAACAATCAGAGGAACTTTCTTGCAGGCGATGAGGATAATGTGATTAGTCAG
cDNA GCAGGCCATCCAGTTGCCCTAAAAGCTTCCTCAAATCTTGATTTGCTTGGGTTTGGTATTAATGCTGAGAACAATCAGAGGAACTTTCTTGCAGGCGATGAGGATAATGTGATTAGTCAG 1148
A.A. A G H P V A L K A S S N L D L L G F G I N A E N N Q R N F L A G D E D N V I S Q
Sau 3A
VicA ATACAGCGACCAGTGAAAGAGCTTGCATTCCCTGGATCAGCTCAAGAGGTTGATAGGATACTAGAGAATCAGAAACAATCTCACTTTACAGATGCTCAACCTCAACAAAGGGAGAGAGGA
cDNA ATACAGCGACCAGTGAAAGAGCTTGCATTCCCTGGATCAGCTCAAGAGGTTGATAGGATACTAGAGAATCAGAAACAATCCCACTTTGCAGATGCTCAACCTCAACAAAGGGAGAGAGGA 1268
A.A. I Q R P V K E L A F P G S A Q E V D R I L E N Q K Q S H F A D A Q P Q Q R E R G
Sau 3A
VicA AGTCGTGAAACAAGAGATC
cDNA AGTCGTGAAACAAGAGATCGTCTATCTTCAGTTTGAAATGTTTCTTAATGAGTGGACAAAATACTATGTATGTATGCTATCAAGAGATATATCTCACGGGGAGCAATGAATAAAACAATG 1388
A.A. S R E T R D R L S S V X .....
cDNA TTATCTTATAACTATAATTATATATCCACTTTTCTACTATGAATA....(1433)

Figure 15

3.2 Nuclease sensitivity of genes in pea chromatin - preliminary results

Comparisons of nuclease sensitivity between various samples were made using standard conditions of digestion; the results of investigations into establishing these conditions are presented below.

3.2.1 Yield and DNA content of isolated nuclei preparations

Intact nuclei were used in digestions to ensure that chromatin was maintained, as far as possible, in its native state. Figure 16 shows isolated nuclei stained with DAPI and seen under the fluorescence microscope. Yields of nuclei were typically 10^6/g of leaf tissue, those from cotyledons were more variable, and were low from older (18-22 d.a.f.) cotyledons. The DNA contents of nuclei preparations, as measured by DABA fluorescence (Section 2.2.20.2), were 0.2-0.4µg/ml for leaf and 0.05 to 0.4µg/ml for cotyledon nuclei. Low concentrations of DNA did not always correlate with poor yields of nuclei from cotyledons; this may have been because of the increased amount of DNA/nucleus due to endomitosis and endoreduplication (Boulter, 1981). The nuclei in older cotyledons are larger because of this increase in DNA content, and may have been easily broken during isolation by the forces required to disrupt the tissue, so resulting in poor yields of intact nuclei. DNAase I digestions were carried out on similar amounts of DNA rather than numbers of nuclei.

3.2.2 Storage conditions and effects on activity of DNAase I

The activities of stocks of DNAase I stored in

FIGURE 16

Fluorescence micrograph of intact pea nuclei

Nuclei from 11 d.a.f. cotyledons were stained with the fluorescent dye DAP1 and examined as described in the text.

Magnification = approx. 400x

Figure 16

digestion buffer, as measured by the spectrophotometric assay described in Section 2.2.20.4, are given in Table 8. Differences between the mean activities were not found to be significant. This result suggested that DNAase I solutions stored under these conditions retained full activity. However, the activity obtained here, approximately 6.6×10^{-4}/min/u, was about two-thirds of that obtained under optimal conditions, *i.e.* under these conditions, one Kunitz unit gives a change in A_{260} of 1×10^{-3}/min. Storage of DNAase I at -20°C in 0.1 N HCl led to complete inactivity in this assay (data not shown).

TABLE 8 Mean* activities of DNAase I stocks as assayed spectrophotometrically

Stock concentration (mg/ml)	Period of storage at -20°C	A_{260}/min/unit of enzyme
1.66	none	6.6×10^{-4}
1.66	10 months	6.31×10^{-4}
1.0	none	6.63×10^{-4}

*Each figure is the mean of 8 determinations; two for each dilution of the stock. A small-sample statistical test (Student t test) was used to compare the means (Section 2.2.20.4).

3.2.3 Digestion of DNA and chromatin by nucleases

The electrophoresis assay for nuclease digestion was described in Section 2.2.20. Figure 17 shows that DNAase I stored at -20°C in 0.1 N HCl was inactive; no reduction was observed in the size of the DNA fragments to which enzyme was added (tracks (a) and (c) to (h)) compared to that of the control (track (b)).

FIGURE 17

Activity of DNAase I stock stored in 0.1N HCl

10µg of λNM258 were digested at 25°C with DNAase I at 5u/ml for (a) and (c), 0, (d) 2, (e) 5, (f) 10, (g) 15 and (h) 30 min. DNA was purified and electrophoresed through a 0.5% (w/v) agarose gel. (b) is a control containing buffer instead of nuclease, incubated for 30 min. (i) and (j) are digests of λNM258 with Eco RI and Hind III respectively.

Numbers on the right are size-markers in kb.

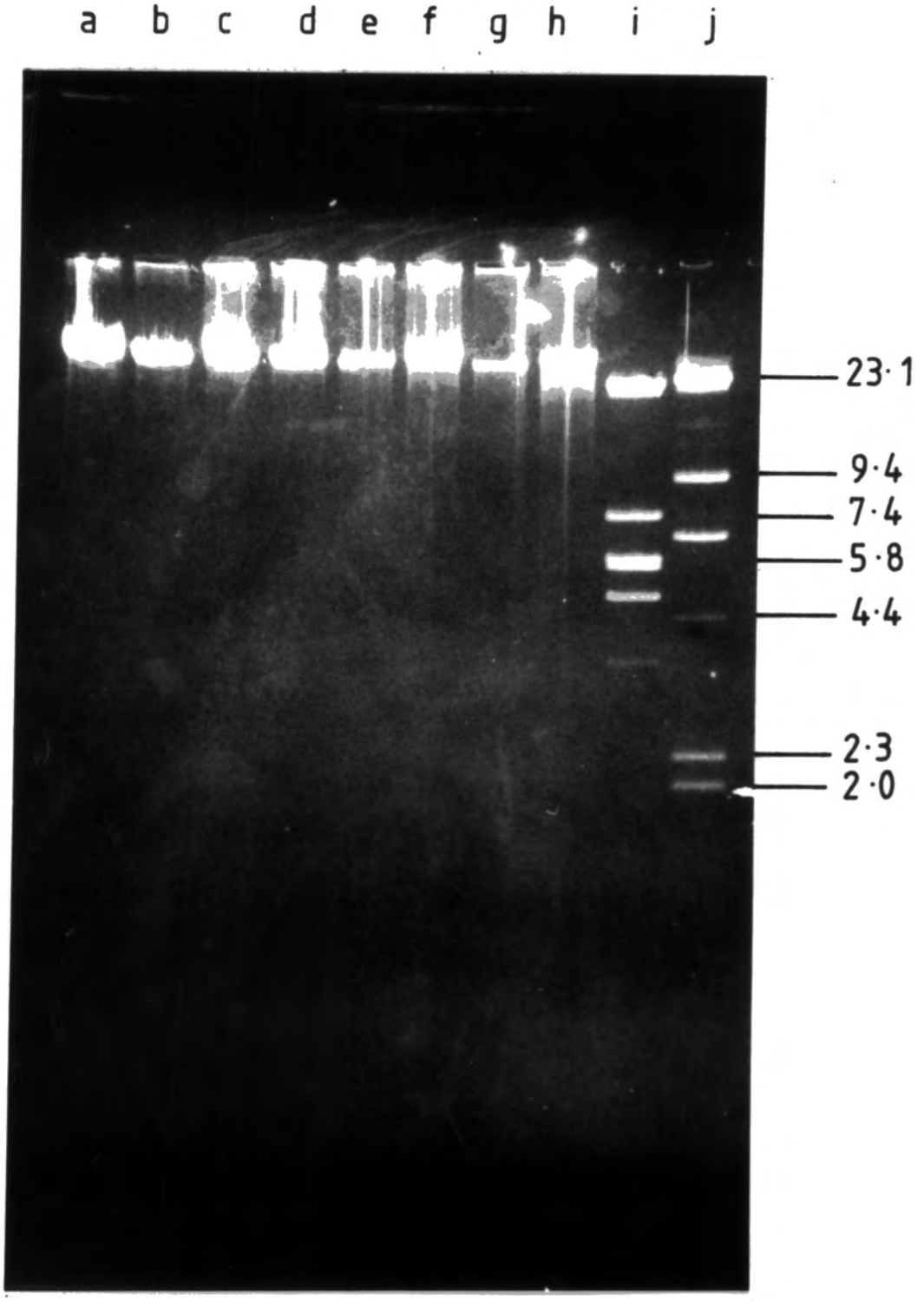


Fig. 17

The gels in Figure 18 show the activity of DNAase I stored in digestion buffer with λNM258 (Figure 18A) pea genomic DNA (Figure 18B) and pea chromatin (Figure 18C) as substrates. In each case, there was an increase in digestion, shown by a reduction in size compared to the control tracks, as the time of incubation (Figures 18A and 18C) or concentration of enzyme (Figure 18B) was increased. The lack of degradation in the control of 11 d.a.f. cotyledon nuclei (Figure 18C, track (a)), suggests that there was little endogenous nuclease activity within these nuclei. A similar result was found for leaf nuclei (data not shown).

The extent of digestion of pea DNA in track (c), Figure 18B, was similar to that of pea nuclei in track (d), Figure 18C. From the amount of DNA and enzyme in each digest (see legend to Figure 18) there were approximately 1.6×10^{-3} units of enzyme/μg DNA for pea DNA, and 0.67 units/μg for 11 d.a.f. cotyledon nuclei. If each substrate was equally accessible to the nuclease, then the digestion of pea leaf DNA at the 4-500 fold lower enzyme to DNA ratio might have been expected to take 4-500 times as long as that of the chromatin, *i.e.* 800 to 1000 min. As the DNA was digested to this extent within 10 min, it appeared to be 80 to 100 times as sensitive to DNAase I as was the chromatin in 11 d.a.f. cotyledon nuclei.

The results showed that the DNAase I stocks stored in digestion buffer digested linear DNA (λNM 258), genomic DNA and chromatin in isolated nuclei under the conditions described in Section 2.2.20.5.

The activity of MNase with DNA and chromatin as substrates is shown in Figure 19. Digestion of DNA (Fig.19A)

FIGURE 18

Activity of DNAase I stored in digestion buffer as assayed by gel electrophoresis

A. Bacteriophage DNA as substrate

2µg of λNM258 DNA were digested at 25°C with DNAase I at 10u/ml for (b) 0, (c) 1, (d) 2, (e) 5, (f) 10 and (g) 15 min. (a) is a control incubated for 15 min. with buffer instead of nuclease. DNA was purified and electrophoresed through a 0.5% (w/v) agarose gel.

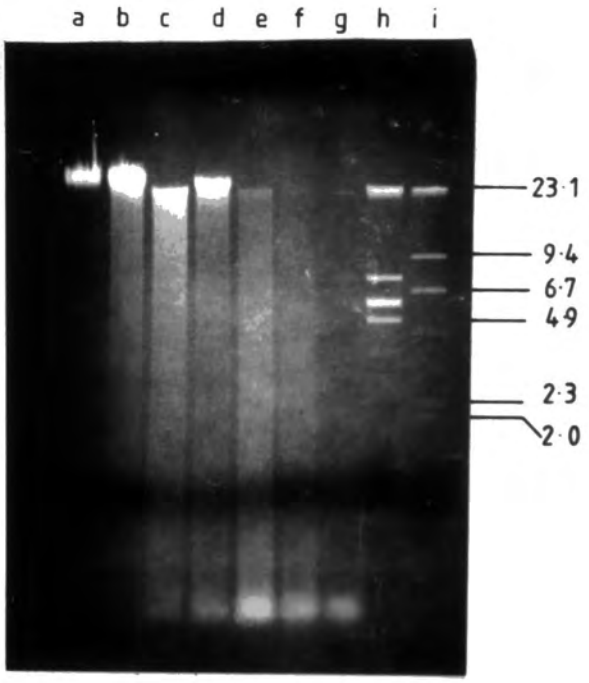
B. Pea genomic DNA as substrate

6µg of pea leaf DNA was digested at 25°C for 10 min. with DNAase I at (b) 0, (c) 0.1, (d) 0.05, (e) 0.2, (f) 0.5 and (g) 1.0u/ml. DNA was purified and electrophoresed through a 0.7% (w/v) agarose gel.

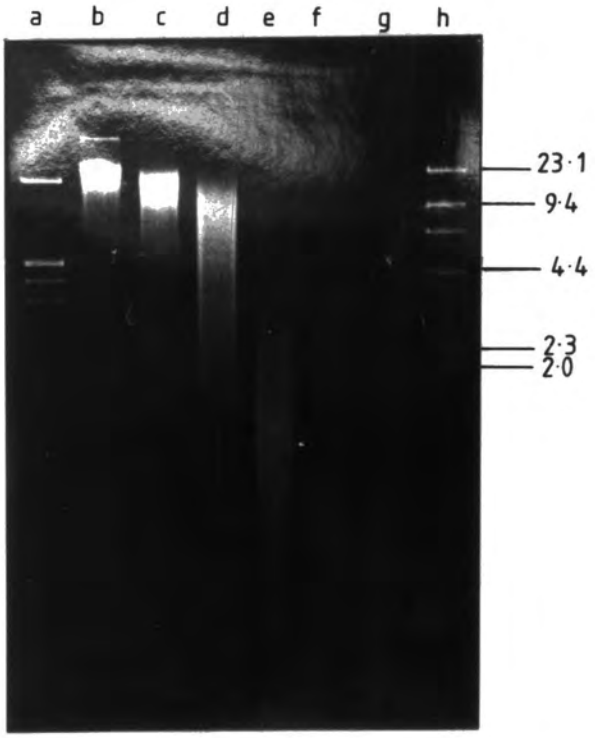
C. Pea chromatin as substrate

15µg of DNA in 11 d.a.f. cotyledon nuclei were digested at 25°C with DNAase I for (a) 0, (b) 0.5, (c) 1.0 and (d) 2.0 min, (e) is a control containing buffer instead of nuclease. DNA was purified and 1.5µg electrophoresed through a 1% (w/v) agarose minigel.

A



B



C

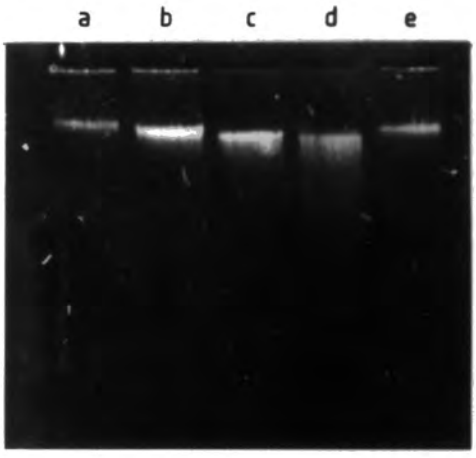


Fig. 18

FIGURE 19

Activity of micrococcal nuclease on DNA and chromatin

A. Digestion of pea leaf DNA

10µg of pea leaf DNA were incubated at 25°C with MNase at (b) and (c) 50, (d) 100, (e) and (f) 200u/ml. In (b) and (f) reactions were terminated immediately after addition of enzyme, other reactions were incubated for 6 min. (a) is a control containing buffer instead of nuclease. (g) and (h) are λNM258 restricted with Eco RI and Hind III respectively.

B. Digestion of (1) pea leaf and (2) pea 11 d.a.f. cotyledon nuclei

10µg of nuclei were incubated at 25°C for 6 min. with MNase at (a) 50, (b) 100 and (c) 200u/ml.
(a) is a control containing buffer instead of nuclease

Numbers at the right are size markers in kb.

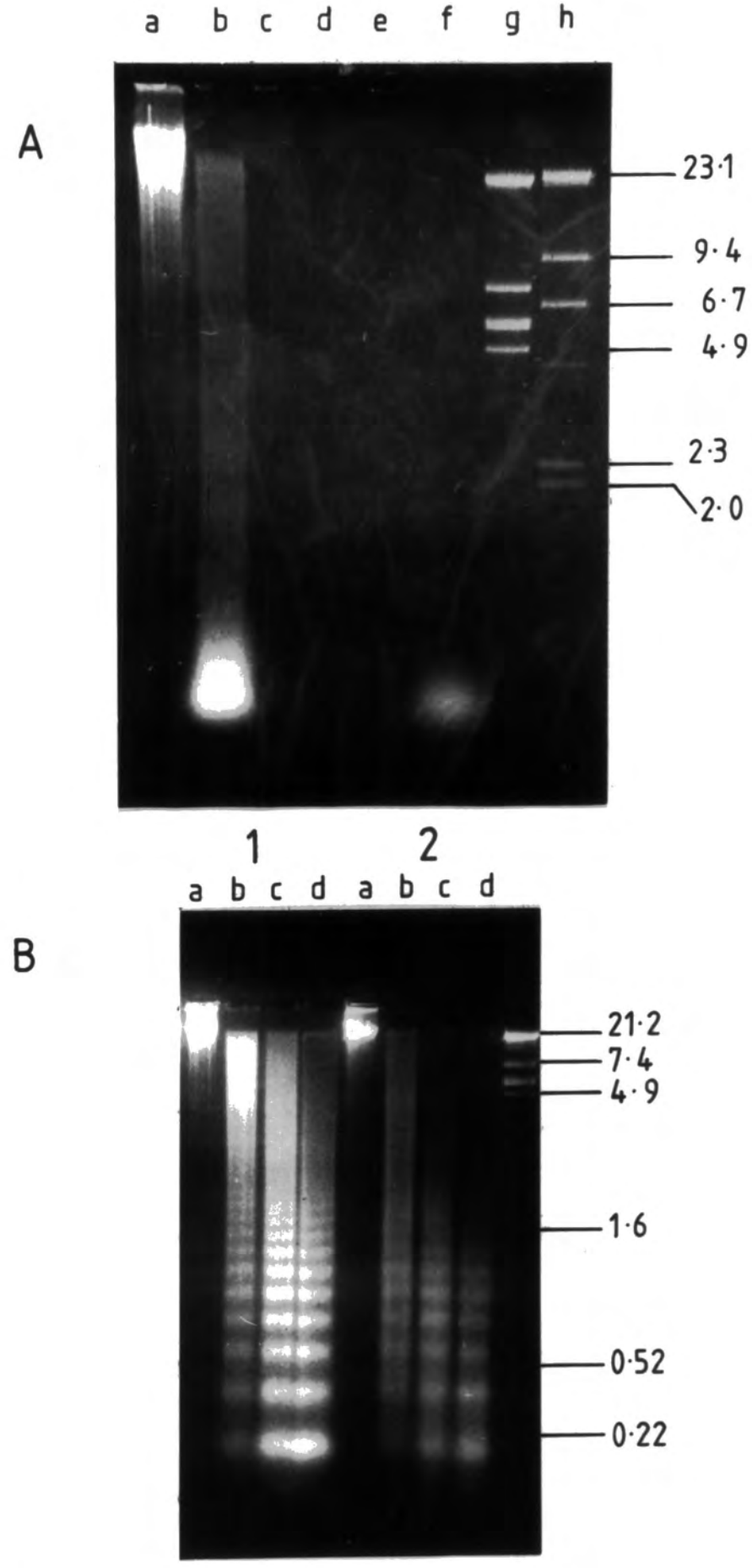


Fig. 19

was rapid; even the zero time track at 50u/ml MNase (track (b)) shows an appreciable decrease in size of DNA relative to the control (track (a)). Chromatin (Figure 19B) was more resistant to MNase digestion than was DNA, as was found in the DNAase I digestions. Discrete bands of defined sizes were present in digests of chromatin at each of the enzyme concentrations used (Figure 19B). The sizes of the bands were similar in both leaf and cotyledon nuclei. The mean size difference between bands was found to be 171 bp for both tissues. The large standard error, ±25 bp, was due to the difficulty of estimating the sizes of the bands, and the differing sizes within each band caused by the action of MNase at different sites in the linker DNA. Therefore the chromatin in nuclei isolated as described in Section 2.2.20.1 retained a nucleosomal conformation.

3.2.4 Activity of DNAase I from different sources, and effects of a proteinase inhibitor

The activities of DNAase I from Sigma Chemical Co. and from Worthington Biochemicals were compared using the electrophoresis assay. At similar concentrations, both enzymes digested pea DNA and nuclei, although the extent of degradation was slightly greater with the Worthington enzyme than with the Sigma enzyme (data not shown).

The effect of the proteinase inhibitor, PMSF, was investigated in order to determine whether a contaminating proteinase was present in the nuclei preparations. The presence of PMSF in digestions of pea leaf nuclei resulted in a slightly greater extent of digestion than in its absence (data

not shown). However, a similar result was observed when PMSF was present in digests of pea leaf DNA, suggesting that the inhibitor was affecting other processes besides proteinase activity.

3.2.5 Effect of temperature on DNAase I digestions

Detection of DNAase I hypersensitive sites, if present in pea chromatin, required very mild digestion of nuclei. Low levels of digestion were obtained either by reduction of the temperature of incubation or by the use of very low concentrations of enzyme, or by a combination of the two. The results of DNAase I digestions at 12°C of pea leaf DNA and nuclei are shown in Figure 20. As at 25°C, chromatin was more resistant to digestion than was DNA; all of the pea DNA was degraded to small fragments after 5 min, (Figure 17A, track (f)), whereas large-sized fragments remained in the chromatin after 7 min of digestion (Figure 17B, track (d)). For each substrate, digestion was slower at 12°C than at 25°C.

3.3 General DNAase I sensitivity of genes in pea chromatin

DNAase I sensitivity was investigated by gel electrophoresis and hybridisation to specific probes as previously described (Section 2.2.20, Figure 3); differential sensitivity was determined by comparisons of the intensities of bands seen on the autoradiographs.

3.3.1 DNAase I sensitivity of ribosomal genes

Detection of the multiple copy ribosomal genes was

FIGURE 20

DNAase I digestions at 12°C

A. 15µg of pea leaf DNA were digested at 12°C with DNAase I at 10u/ml for (b) 0, (c) 1, (d) 2, (e) 3 and (f) 5 min. (a) is a control containing buffer instead of nuclease.

B. 6µg of DNA in pea leaf nuclei were digested at 12°C with DNAase I at 4u/ml for (b) 0, (c) 1 and (d) 7 min. (a) is a control with buffer instead of nuclease.

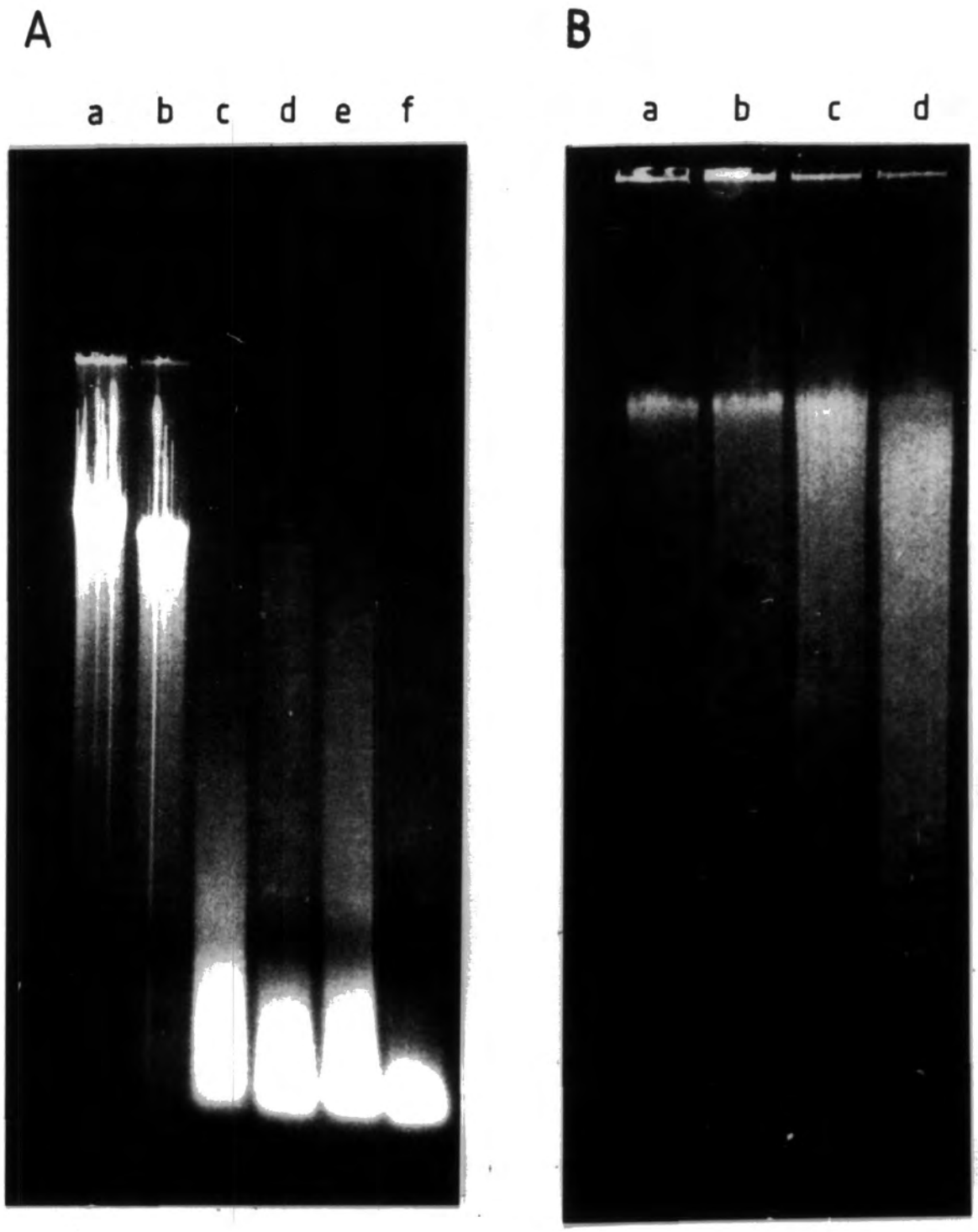


Fig. 20

readily obtained compared to that of the low copy number seed storage protein genes (Section 3.3.2). Figure 21 shows a map of pHA1 (Cuellar, 1982), a plasmid containing the pea ribosomal repeat inserted into the vector pACYC 184 (Chang and Cohen, 1978). This plasmid was labelled and used as a probe to assay the DNAase I sensitivity of ribosomal chromatin in pea leaves and cotyledons.

The gel in Figure 22A shows the Bam Hl digestions of DNA purified from DNAase I treatments of leaf DNA, leaf nuclei and 11 d.a.f. cotyledon nuclei. The restriction profiles of the samples were similar except for the controls (tracks (b), (f), (j)) which had restricted poorly, the 15u/ml tracks for leaf DNA (track (d)) and cotyledon nuclei (track (m)) wjich showed a reduction in fragment sizes, and the 5u/ml track for cotyledon nuclei (track (k)) which was incompletely restricted.

The result of hybridising a southern blot of this gel to pHA1 is shown in Figure 22B. The sizes of the fragments in genomic DNA which hybridised to the probe were 7.4, 4.9 and 4.5 (doublet), 2.8, 1.95, 1.18 and 0.63 kb (tracks (b) to (e)). Additional bands of 9.0 and 6.7 kb were seen in some tracks when the film was exposed for 30 min (data not shown). The larger bands were absent from tracks in which restriction was estimated to be complete (*e.g.* tracks (g) to (i) and (l)) and were therefore partial digestion products. Figure 23 shows the fragments expected if certain Bam Hl sites were not cleaved. The 4.5 kb band probably represented the distance from the last Bam Hl site in one repeat to the 1st site in the next (Fig.23).

FIGURE 21

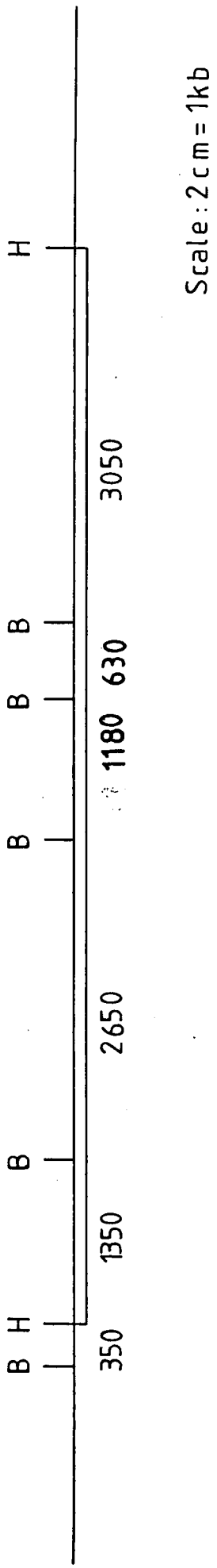
Map of Bam HI and Hind III sites in the plasmid pHA1

pHA1 (Cuellar, 1982) contains a Hind III fragment of pea DNA, carrying the ribosomal repeat, inserted into the vector pACYC184 (Chang and Cohen, 1978).

Sizes of fragments are in bp. B = Bam HI, H = Hind III, ——— = pACYC184 (3.75kb)

⊏———⊐ = ribosomal repeat of pea.

In addition to the internal fragments of 2.65, 1.18 and 0.63kb, a 1.7 kb fragment from the Bam H1 site in the vector to the first site in the insert, and a 6.4 kb fragment from the 4th site in the insert to the site in the vector, will result from a Bam HI digestion of this plasmid. All of these fragments can be seen in track (a), Figure 22A.



Scale: 2 cm = 1kb

Fig. 21

FIGURE 22

DNAase I sensitivity of the ribosomal genes in pea chromatin

A. DNAase I digestions and Bam HI restrictions of pea leaf DNA, nuclei and 11 d.a.f. cotyledon nuclei.

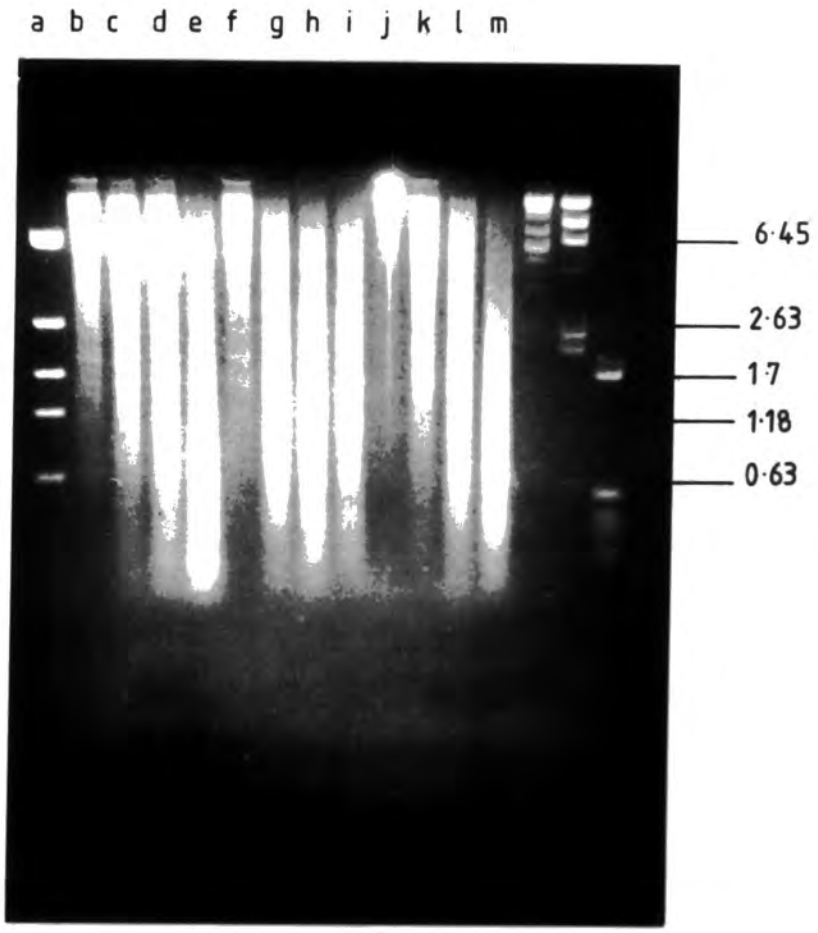
Track (a) Bam HI restriction of pHA1.

Tracks (b) to (e) pea leaf DNA, (f) to (i) leaf nuclei, (j) to (m) 11 d.a.f. cotyledon nuclei.

7µg of DNA were digested for 6 min. at 25°C with DNAase I at (c), (g), (k) 5u/ml, (d), (h), (l), 10u/ml and (e), (i), (m) 15u/ml. (b), (f) and (j) are controls containing buffer instead of nuclease. Other tracks are size markers.

B. Autoradiograph of a blot of the gel in (A) above, hybridised to pHA1. Tracks are as in (A). The filter was washed to high stringency (0.1 x SSC, 0.1% SDS, 65°C). Film was exposed for 3h at -80°C.

A



B

Fig. 22

FIGURE 23

Partial Bam HIcleavage of the ribosomal repeat in pea

The 9.0 kb band seen in Fig.23B may arise if only site is cut (predicted fragment length 8.74kb). The 7.0 and 6.7 kb fragments might arise if site 1 is not cut (predicted fragment length 6.93 kb); heterogeneity in the spacer length might result in production of two fragments. The 4.5 and 4.9 kb bands may represent the distance from site 4 to site 1 (4.3 kb); again, spacer length heterogeneity could account for the two different fragments. Alternatively, the 4.9 kb band might result from site 4 being blocked (predicted fragment length 4.93 kb). The 1.95 kb bond may result if site 3 is not cleared (predicted fragment length 1.81 kb). The 1.95 kb. band may result if site 3 is not cleaved (predicted fragment length 1.81 kb).

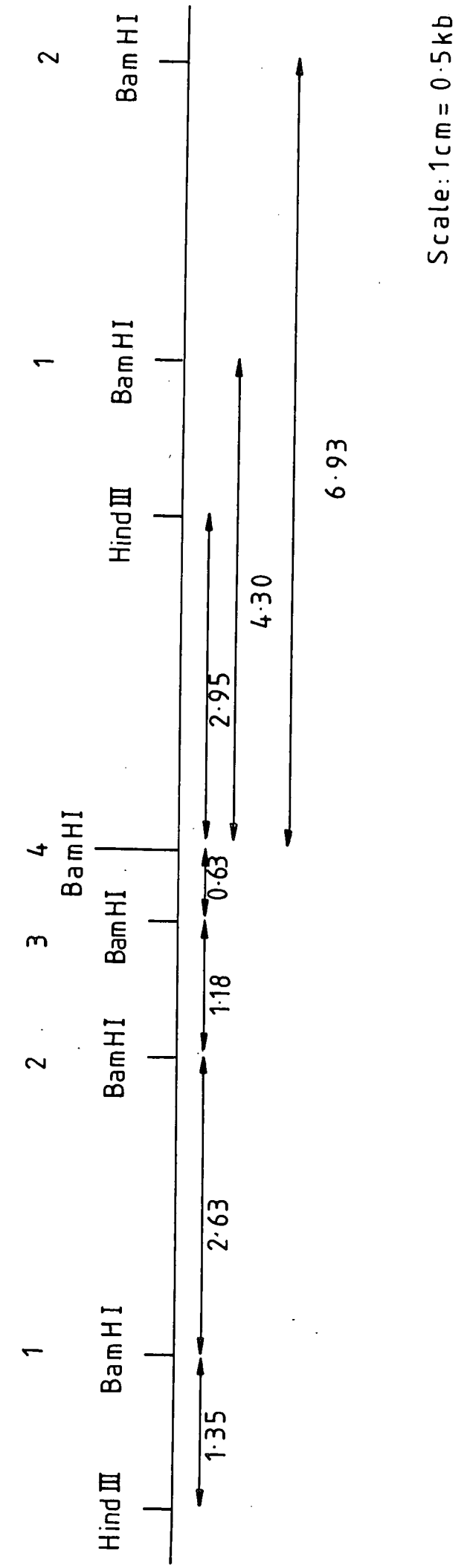

Scale: 1cm = 0·5kb
Hind III
Bam HI
Bam HI
Bam HI
Hind III
Bam HI
1
2
1
2
3
4
1·35
2·63
1·18
0·63
2·95
4·30
6·93

Fig. 23

The differences in the smears of fragments on the ethidium-bromide stained gel (Figure 22A) resulted in variations in intensities of the bands between tracks (Figure 22B). For example, the loss of bands from leaf DNA (track (e)) and 11 d.a.f. cotyledon nuclei (track (m)) was associated with the decrease in fragment size of total DNA in these samples (tracks (e) and (m), Figure 22A). For tracks in which the restriction profiles were similar (Figure 22A, tracks (g) to (i) and (l)), the intensities of the bands on the autoradiograph showed little difference between tracks (Figure 22B, tracks (g) to (i) and (l)). Thus the differences in band intensities could be attributed to variations in total DNA fragment sizes rather than to differences in nuclease sensitivity. The similar band intensities for leaf nuclei at 5u/ml (track (g)) and 15u/ml (track (i)), and for cotyledon nuclei at 10u/ml (track (l)), suggest that there is no difference in sensitivity of ribosomal chromatin in the two tissues, and that it is relatively resistant to DNAase I.

The genomic DNA (tracks (b) to (e)) in this experiment was similar in sensitivity to chromatin, in contrast to previous results (Section 3.2.3). This was probably because the DNA was resuspended in 50mM Tris, 10mM EDTA (Section 2.2.3) rather than distilled water (Section 2.2.20.5) and therefore the relatively high levels of EDTA may have chelated some of the Mg^{2+} ions required for DNAase I activity.

3.3.2 Legumin genes

A map of plasmid pDUB24, which has a Bam Hl insert containing the *leg A* gene from λleg 1 (Lycett *et al* (1984),

is shown in Figure 24. The insert was used to probe Hind III digests of pea leaf DNA in order to establish the usual restriction pattern for this sequence. The probe hybridised to fragments of 4.15, 3.15, 2.75, 2.4 and 1.6 kb in size (data not shown). The most intense was the 2.4 kb fragment. The 2.4 kb fragment corresponds to the Hind III fragments containing the coding regions and some 3′ flanking sequences of legumin genes A, B and C (Shirsat, 1984, Croy *et al*, in preparation). The 1.6 kb band probably represents the pseudogene, *ψ leg D* (Bown *et al*, 1985), while the other bands may represent hybridisation between the 5′ flanking sequences of *leg A* in the probe, and those of the various legumin genes. The distance from the Hind III site just within the *leg A* gene (see Figure 24) to the next site 5′ to the gene is unknown as the λ genomic clone does not include this site (Figure 24). Thus the 4.15 kb band may represent this fragment.

This probe was used to determine the DNAase I sensitivity of legumin genes in leaf and 11 d.a.f. cotyledon nuclei by hybridisation to a blot of the gel in Figure 25A. The resulting autoradiograph is shown in Figure 25B. The sizes of the fragments detected by the probe in the controls (tracks (a) and (f)) were 4.2 and 2.5 kb. A band of 1.7 kb was also visible in some tracks when the autoradiograph was exposed for 2 weeks. The intensity of the bands was 2.5>4.2>1.7, as for the genomic DNA restrictions.

The gel in Figure 25 shows that pea leaf DNA was digested to very small fragments (tracks (k) to (n)) by the levels of DNAase I used here. Digestion of leaf and 11 d.a.f. cotyledon nuclei with these levels of DNAase I did not markedly

FIGURE 24

Map of pDUB24, a plasmid containing legumin gene A

A. λleg 1 genomic clone showing legumin gene A and pseudogene ψ *leg D*, and the 3.4 kb Bam HI fragment that was sub-cloned into pDUB24.

B. Expanded map of insert of pDUB24. The mature message is shown below the map.

= Hind III, = Xho II, = Bam HI, = Acc I,

= Xho I, = Ava I, = Pst I, = Bst Nl,

= Bgl I, = Hinc II.

IVS = intervening sequences.

(Not to scale).

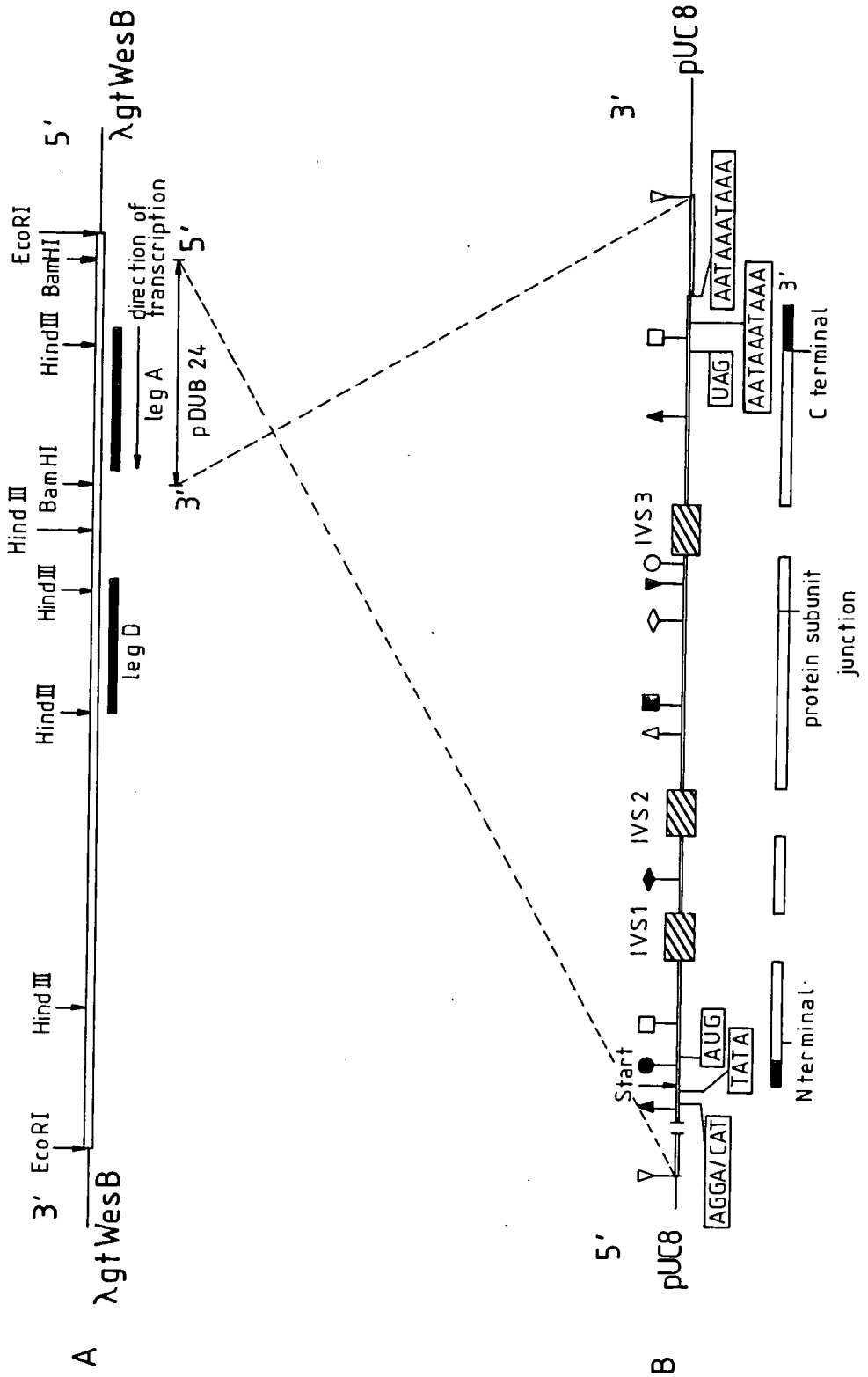

A
λgtWesB
5'
3'
EcoRI
Hind III
BamHI
Hind III
Hind III
Hind III
EcoRI
λgtWesB
leg D
leg A
direction of transcription
5'
3'
pDUB 24
B
pUC8
5'
3'
Start
AUG
TATA
AGGA/CAT
IVS1
IVS2
IVS3
UAG
AATAAATAAA
AATAAATAAA
3'
C terminal
protein subunit
junction
N terminal
pUC8

Fig. 24

affect the restriction profile (tracks (b) to (e) and (g) to (j)). No bands were observed on the autoradiograph for the protein-free DNA samples (data not shown).

Comparison of the intensities of the bands in leaf and cotyledon nuclei (Figure 25B) shows that the 2.51 kb fragment in leaf nuclei was present on the autoradiograph after 1 min of digestion (track (d)). At this level in 11 d.a.f. cotyledon nuclei, however, the band was only just visible (track (i)). After 2 min, this band was just visible in the leaf nuclei sample (track (e)) but was completely absent from the cotyledon nuclei (track (j)). The levels of hybridisation to the 4.2 and 2.5 kb bands in the control, zero-time and 30 second digests (tracks (a) to (c) and (f) to (h)) were similar for the two tissues, although restriction of the control track for cotyledon nuclei was incomplete (track (f)). The differences in intensity of hybridisation described above cannot be attributed to differences in DNA content. Thus examination of the gel photograph in Figure 25 shows that the 1 min digestion tracks for leaf (track (d)) and cotyledon (track (i)) nuclei show that each contained similar amounts of DNA. Therefore, the difference in intensity of the 2.5 kb band between leaves and cotyledons represented an increased sensitivity of the legumin genes in cotyledon nuclei.

Similar results were obtained using this probe in other experiments (data not shown) with leaf, 9 d.a.f., 10 and 11 d.a.f. cotyledon nuclei. The legumin genes in cotyledon nuclei always appeared to be 2-4 times as sensitive to DNAase I as those in leaf nuclei. The sensitivity of the legumin genes was similar in 9, 10 and 11 d.a.f. cotyledon nuclei.

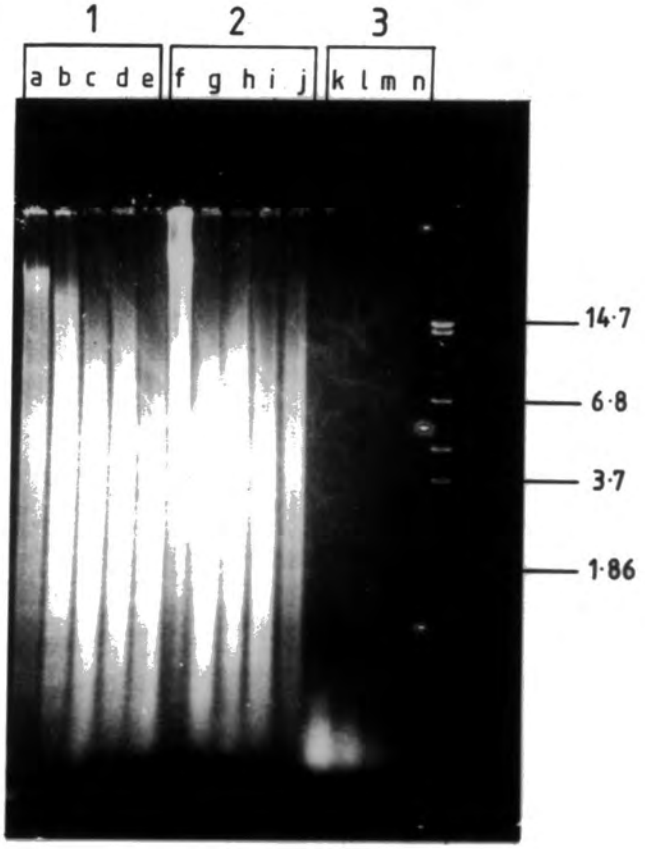
FIGURE 25

DNAase I sensitivity of legumin genes in pea chromatin

A. Ethidium-bromide stained gel of samples from (1) leaf nuclei, (2) 11 d.a.f. cotyledon nuclei and (3) protein-free DNA. 15µg of DNA were digested with DNAase I at 10u/ml for (b), (g), (k) 0, (c), (h), (l) 0.5, (d), (i), (m) 1.0 and (e), (j), (n) 2.0 min. at 25°C. Tracks (a) and (f) are controls containing buffer instead of nuclease, incubated for 2.0 min. Samples were purified, restricted with Hind III and electrophoresed through a 0.5% (w/v) gel. Numbers at the right are size-markers in kb.

B. Autoradiograph of a blot of the gel in (A), hybridised to the insert of pDUB24. (Stringency of washing was 0.1 x SSC, 0.1% SDS, 65°C). Sizes of fragments are kb. Film was exposed for 1 week at -80°C.

A



B

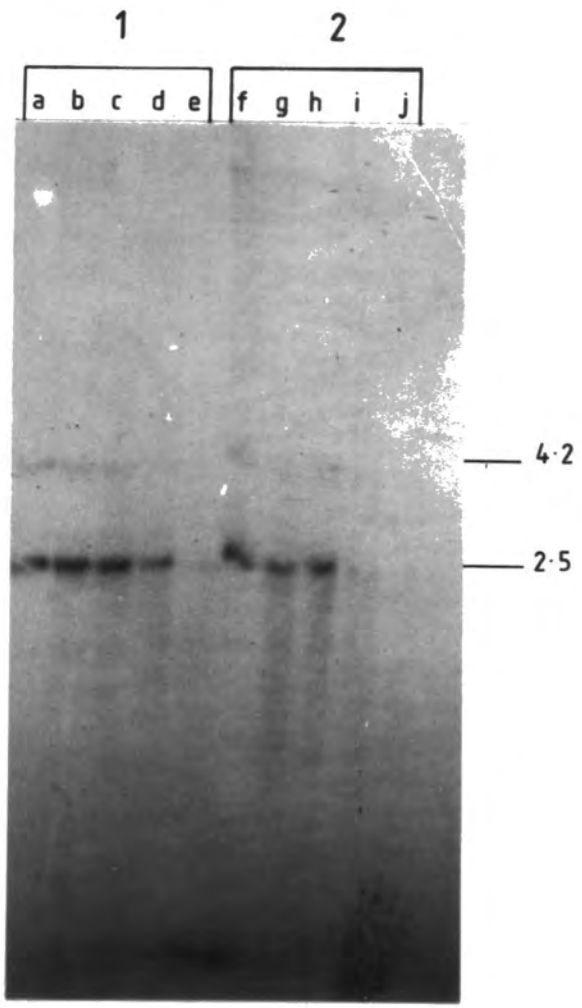


Fig. 25

3.3.3 Vicilin genes

The restriction maps of 3 vicilin cDNAs are shown in Figure 2, Methods. pDUB7 which encodes 47K M_r vicilin polypeptide, was used as a probe to assay the nuclease sensitivity of vicilin genes in chromatin.

The labelled insert of pDUB7 was hybridised to southern blots of gels of various restriction digests of pea DNA. The following bands were seen on the autoradiograph: Eco RI, 6.7 kb, Bam Hl, 6.5kb and Ava I, 4.8 and 3.0 kb. A similar experiment, in which a blot of DNA purified from DNAase I treated 11 d.a.f. cotyledon nuclei and restricted with Eco RI was hybridised to this probe, gave a band of 6.6 kb in the control (data not shown). This band was also present, but at a greatly reduced intensity, in tracks where nuclei were digested with 10 or 15u/ml DNAase I for 6 min at 25°C. This suggested that the vicilin genes in cotyledon nuclei are sensitive to DNAase I, but as the sensitivity in leaf nuclei was not assayed, no comparisons can be made between the active and inactive vicilin genes.

3.4 DNAase I hypersensitive sites in pea chromatin

The levels of enzyme used in Section 3.3.2 (10u/ml) may have resulted in considerable digestion of total chromatin even during the short incubation times (0.5-2 min). Hypersensitive sites, if present, might therefore have been masked by the degradation of bulk chromatin. In order to reduce the extent of digestion, the temperature and/or enzyme concentraction was reduced (Section 3.2.4).

The gel photograph in Figure 26 shows the extent of digestion of chromatin in leaf and 13 d.a.f. cotyledon nuclei by very low levels of DNAase I for 30 min at 15°C. Over a 50-fold increase in enzyme concentration, there was no size reduction in the restriction profiles of digested samples as compared with the controls (tracks (a) and (b)). However, there were some differences in the amounts of DNA; this was due to the difficulty in determining accurately the small amounts of DNA in the aliquots taken for fluorescence analysis (Section 2.2.20.2).

A blot of this gel was hybridised to the labelled plasmid pDUB9, containing a vicilin 50K M_r cDNA insert (Figure 2, Methods). Although some bands were detected (data not shown) no conclusions about any putative DNAase I hypersensitive sites could be drawn from this experiment as hybridisation was poor and the background high due to prolonged exposure of the film (3 weeks).

Further experiments, employing these very mild digestion conditions and using probes of high specific activity complementary to the 5′ and 3′ flanking sequences of the genes, are required to establish the presence or absence of hypersensitive sites in pea chromatin (see Chapter Four).

FIGURE 26

DNAase I digestions of leaf and 13 d.a.f. cotyledon nuclei at 15°C

15µg of DNA were digested at 15°C for 30 min, the DNA purified and restricted with Bam HI, and electrophoresed through a 0.8% (w/v) gel.
L = leaf nuclei, C = 13 d.a.f. cotyledon nuclei.
Numbers above each pair of tracks are the DNAase I concentrations, in u/ml, and those on the right are size-markers in kb.

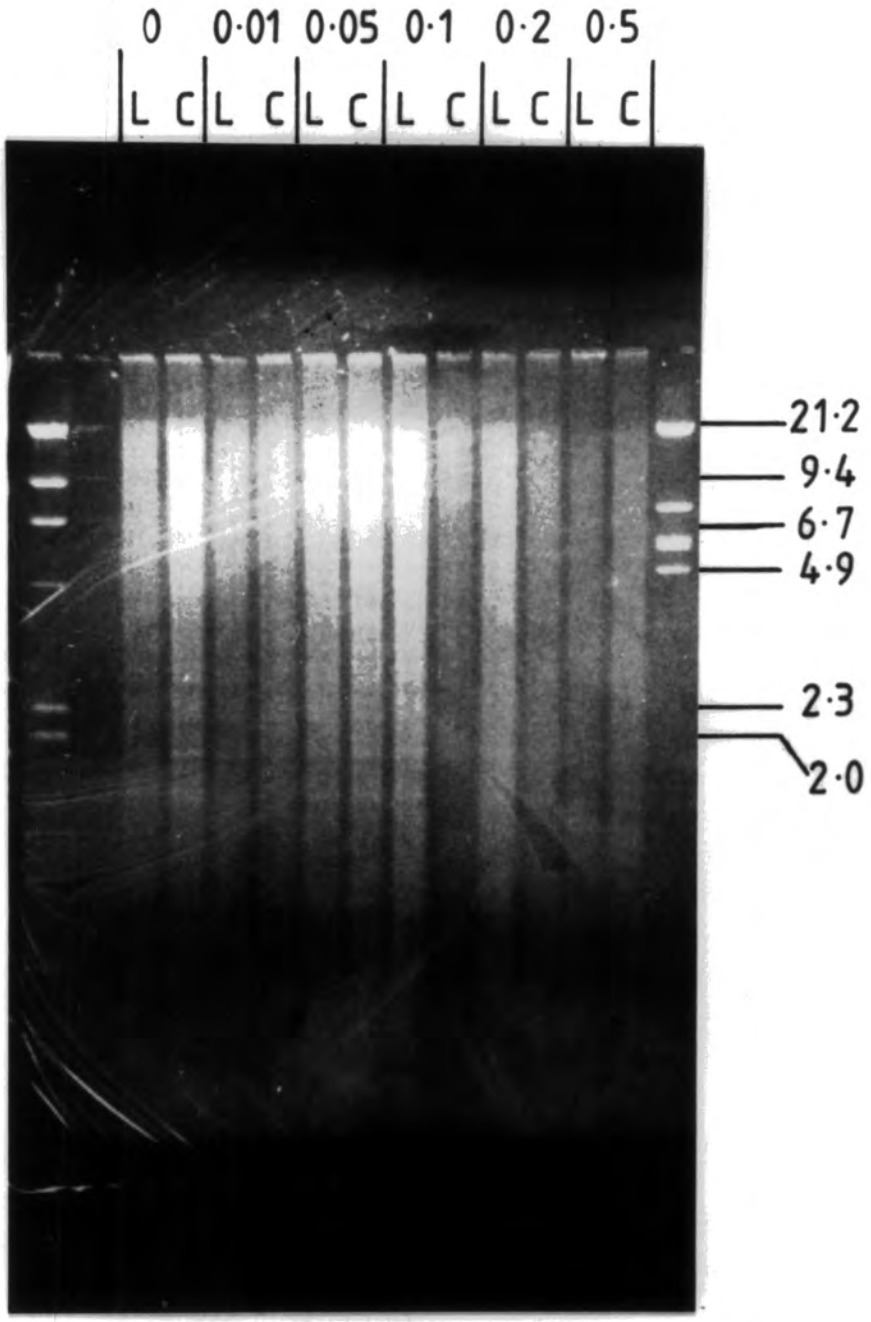


Fig. 26

CHAPTER FOUR

DISCUSSION

4.1 Isolation and characterisation of a vicilin gene from pea

4.1.1 General assessment

The *vic A* gene (Section 3.1.4.1) was shown to have a restriction map similar to that of pDUB9 (Figure 13) and its sequence shows over 98% homology with that of the cDNA (Figure 15). The gene therefore encodes a 50K M_r vicilin polypeptide that does not undergo post-translational cleavage to give the smaller vicilin polypeptides (Gatehouse *et al*, 1982b, Lycett *et al*, 1983a).

There is approximately 450 bp of coding sequence 5′ to the Xba I site in pDUB9 (Figures 13 and 15) but there is only about 250 bp 5′ to this site in *vic A* (Figure 13). The gene contains an intron, IVS2, upstream of this site (Figure 13) and thus a considerable proportion of the coding region, the 5′ untranslated sequence and promoter and other regulatory signals, are missing from this gene (Section 3.1.4.3, Figure 13).

The gene library was screened (Section 2.2.20) by hybridisation to a short vicilin cDNA, pDUB2 (Figure 2, Methods) which is complementary to approximately two-thirds of the coding region (Lycett *et al*, 1983a). This cDNA should have hybridised to all complete vicilin gene sequences, as well as to the 3′ part of the sequence. The isolation of only one gene from a screen of 2.5×10^5 phage from the library (Section 2.2.10.1) suggests that no complete vicilin gene sequences complementary to pDUB2 were present. Screening of this number of recombinant phage gives a probability of only 58% of detecting a single-copy gene (Maniatis *et al*,

1982); the screen was originally intended only as a test for a subsequent larger one, but the hybridisation of a plaque in this test screen meant that a larger screen was not necessary. When approximately 10^6 phage from the library were screened with a variety of probes, about three vicilin genes were obtained (I.M. Evans and R.R.D. Croy, personal communcation), while approximately 30 legumin clones were isolated. The number of vicilin genes in the pea genome has been variously estimated to be 5-6 (Gatehouse *et al*, 1983), 7 (Delauney, 1984) and 11 (Domoney and Casey, 1985), while the estimates of legumin gene numbers are about 4 (Croy *et al*, 1982), 7 (Shirsat, 1984) or 8 (Domoney and Casey, 1985). Thus the isolation of far fewer vicilin than legumin genes suggests that the library may be deficient in vicilin genes. The library was constructed by ligating 15-20 kb fragments from a partial Sau 3A digest of pea DNA into the Bam Hl sites of EMBL3 (A.H. Shirsat, personal communication). If the vicilin genes were cut very frequently by Sau 3A, the small fragments containing the genes would have been excluded from this fraction of the DNA. Alternatively, if the genes were in regions of DNA in which cutting by Sau 3A was infrequent, the recombinant phage genomes containing these sequences might have been too large to be packaged into bateriophage particles. Although it is not possible to determine whether either of these explanations is correct, there is a Sau 3A site in the cDNA 37bp 5′ to the position of IVS2 in *vic A* (Figure 15). Further sequencing of *vic A* will determine whether this is the site at which the pea genomic fragment was cut by Sau 3A and ligated into the vector. The explanation of the apparent deficiency of this library in vicilin

genes as being due to the restriction enzymes used in its construction is supported by the isolation of several such genes from a library constructed from a partial Eco R1 digest of pea DNA (N. Ellis, unpublished results). Sequencing of these genes is in progress in this laboratory.

4.1.2 Comparison of nucleotide and predicted amino acid sequences of *vic A* and vicilin cDNAs

Comparison of the partial nucleotide sequence of *vic A* with the corresponding sequence of pDUB9 (Figure 15) shows that there are 11 base differences. Taking the first arginine residue of the sequence predicted from pDUB9 as codon 1, the N-terminus of the mature vicilin polypeptide (Lycett *et al*, 1983a), the amino acid changes in the sequence predicted from *vic A* which result from these base differences are as follows: asparagine to serine at codon 144, isoleucine to threonine at 165, leucine to isoleucine at 270, aspartic acid to valine at 309 and leucine to valine at 347 (Figures 14 and 15). At codon 187, there are only 2 bases available from the *vic A* sequence. These are GA_ instead of AGG found in the cDNA and suggest the replacement of arginine with aspartic or glutamic acid. The amino acid changes, with the exception of leucine to isoleucine and leucine to valine, lead to changes in charge or in number of polar groups. Charge heterogeneity in the vicilin fraction of peas has been shown by ion-exchange chromatography, two dimensional gel electrophoresis and iso-electric focussing (Gatehouse *et al*, 1981) and it has been suggested that each class of vicilin precursors (Section 1.4.2, Gatehouse

et al, 1981, Lycett *et al*, 1983a) contains a number of polypeptides, each being the product of a different gene. Two cDNAs, pDUB2 and pDUB9, which encode vicilin 50K M_r polypeptides that differ slightly in sequence (Delauney, 1984) represent one class of genes, type A, while the cDNAs pDUB4 (Croy *et al*, 1982) and pDUB7 (Lycett *et al*, 1983a) represent a second class, type B, encoding the 47K M_r precursor (Section 1.4.2). Two other classes have also been proposed (Lycett *et al*, 1983a). Thus the differences between the *vic A* gene and the cDNA pDUB9 suggest that the gene encodes a 50K M_r vicilin polypeptide of type A, which may be different to those encoded by pDUB9 or pDUB2. The detection of several 50,000 M_r type vicilin genes in Southern blots of pea genomic DNA (Gatehouse *et al*, 1983, Delauney, 1984, Section 1.4.2) shows that each class of vicilin polypeptides is encoded by a number of genes, and the isolation of three such sequences, *vic A*, pDUB2 and pDUB9, confirms this.

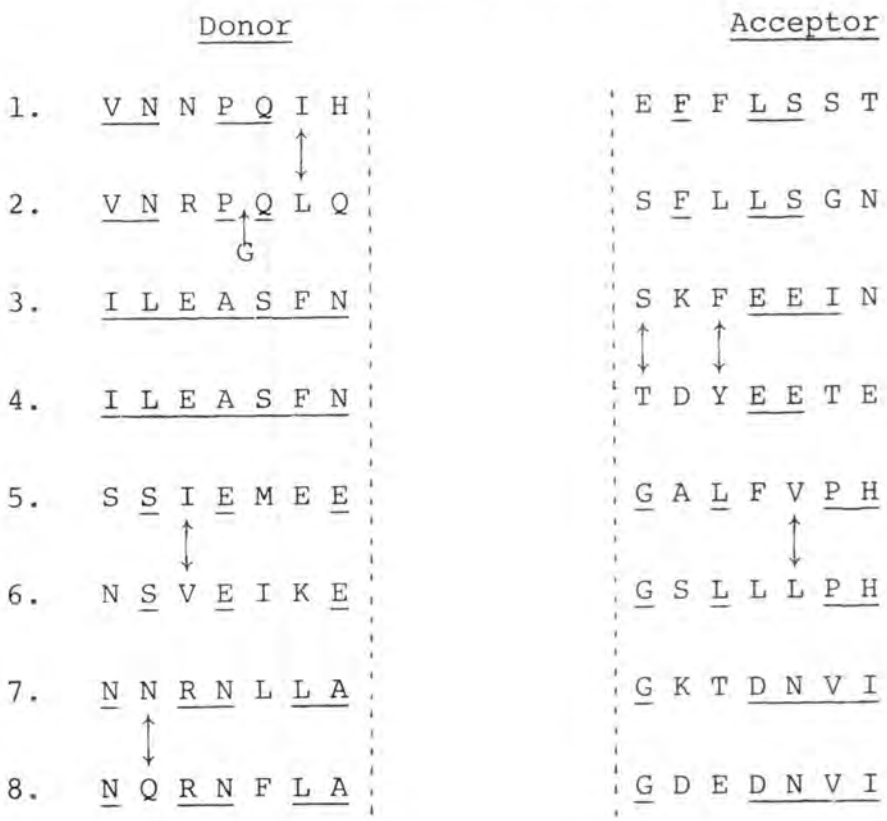
The sequence identity of *leg A* and the cDNA pDUB8 (Lycett *et al*, 1984, Section 1.4.1) showed that this gene was expressed in pea cotyledons. The absence of a cDNA identical to *vic A* prevents assessment of its functional status. However, the evidence for multigene families encoding the pea vicilin fraction, and the charge heterogeneity of the vicilin polypeptides (see above and Section 1.4.2) suggests that a number of vicilin genes are transcriptionally active. The good homology of *vic A* with pDUB9, and the absence of deletions, frameshift errors or stop codons within the coding region, which have been found in a legumin pseudo-

gene (Bown *et al*, 1985), suggest that this gene is expressed in pea cotyledons.

4.1.3 Analysis of predicted amino acid residues adjacent to the intron positions in *vic A*

The sites of the four introns which have been found to interrupt the coding sequence of *vic A* are shown in Figure 15, and their sequences in Figure 14. The numbering of the introns was by analogy with the positions of those in phaseolin (Slightom *et al*, 1983). A comparison of the amino acid sequences at the boundaries of the introns in vicilin with those of the corresponding introns in phaseolin is shown in Figure 27. Although the sequences of vicilin and phaseolin show some regions of divergence (Lycett *et al*, 1983a), there is considerable conservation of amino acid sequences around the intron/exon boundaries (Figure 27). The average homology at these boundaries is 48.5% but for IVS 5 it is over 70% at both donor and acceptor boundaries, while at the donor boundary of IVS 3 the homology is 100%. Where there are differences, the residues in the two sequences are often similar types of amino acids (see Figure 27). There would appear to be no functional reason for this conservation, as the introns are spliced out of the primary transcript before the mature message is transported out of the nucleus (Breathnach and Chambon, 1981). The consensus nucleotide sequences at the splice sites in plant (Slightom *et al*, 1983) and animal (Lewin, 1980), Breathnach and Chambon, 1981) genes include only two nucleotides in the coding region on either side of the splice site. The con-

FIGURE 27 Amino acid sequences at splice sites in *vic A* and phaseolin



1. Phaseolin IVS 2
1. *Vic A* IVS 2
3. Phaseolin IVS 3
4. *Vic A* IVS 3
5. Phaseolin IVS 4
6. *Vic A* IVS 4
7. Phaseolin IVS 5
8. *Vic A* IVS 5

Underlining indicates similar sequences in both genes. Dotted line indicates intron/exon (acceptor) or exon/intron (donor) boundary. The donor boundary of *vic A* IVS 2 has been aligned to give maximum homology to that of phaseolin; there is however an extra G residue in the former sequence. Vertical arrows between residues indicate conservative amino acid changes.

servation of the amino-acid sequences is thus unlikely to be related to the requirements of a splicing system. It is probable that the introns were at these positions in an ancestral gene from which both the vicilin and phaseolin genes were derived. Divergence of these regions may have been limited by constraints on the structure of the proteins, related to their role as a store of reduced nitrogen. The requirements of synthesis of the polypeptides on the rough endoplasmic reticulum, assembly into oligomers and post-translational processing (Chrispeels, 1984) also place restrictions on the secondary, and therefore the primary, structures of these proteins. The occurrence of multigene families encoding seed storage proteins (Higgins, 1984, Croy and Gatehouse, 1985), and the resulting charge and size heterogeneity within each fraction (Section 1.4), indicates that some changes in primary structure are not deleterious to function. However, the good homology between the sequences of pea vicilin, phaseolin and conglycinin, (Lycett *et al*, 1983a), except at the regions of the vicilin processing sites, and the sequence conservation observed in seed proteins from a wide range of species (Spencer, 1984), suggests that functional constraints on protein structure have limited their divergence (Spencer, 1984).

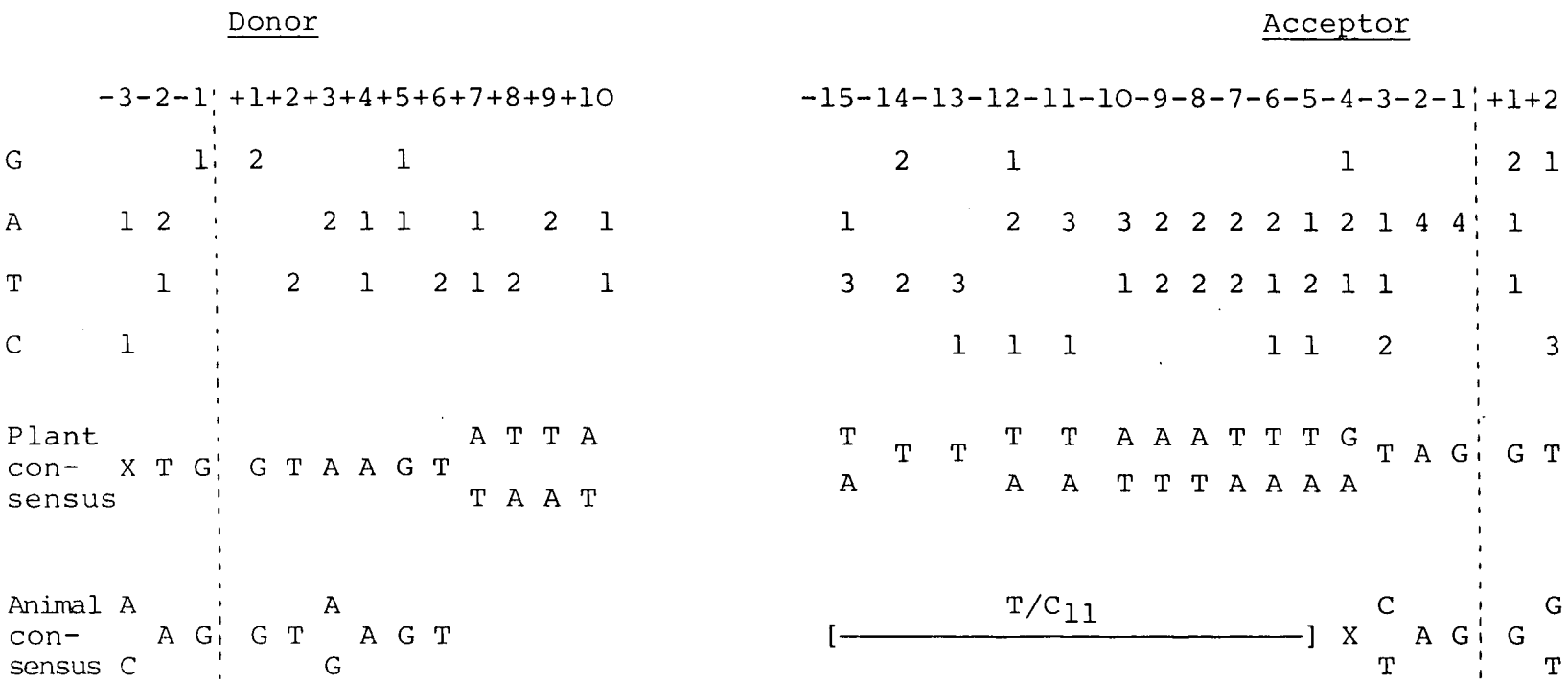
4.1.4 Analysis of nucleotide sequences at the splice sites in *vic A*

The nucleotides present at the donor and acceptor boundaries of the introns in *vic A* have been determined (Section 3.1.4.3, Figures 14 and 15) and the nucleotide

preferences at each position are shown in Figure 28, together with the consensus sequences derived from 20 plant splice sites (Slightom *et al*, 1983) and from other split genes (Lewin, 1980, Breathnach and Chambon, 1981). All of the introns in *vic A* for which the sequences are available begin with GT and end with AG (Breathnach *et al*, 1978). The sequences around the splice sites in *vic A* show greater homology to the plant than to the animal consensus sequences, with a few exceptions (Figure 28). The two donor sites both have an A at -2 instead of a T, which is consistent with the plant rather than the animal consensus. The preference for C at position +2 of the acceptor site (Figure 28) does not fit with either of the consensus sequences. Comparison of the intron/exon boundary sequences of individual introns of *vic A* with those of the corresponding introns in phaseolin (Slightom *et al*, 1983) shows that the differences from the plant consensus sequences are often common to both genes (Figure 29), *e.g.* the presence of G instead of T at position +2 of the acceptor site in IVS 4, and A instead of G at position +1 of this site in IVS 3. However, the preference for C instead of T at the +2 position in the acceptor sites of 3 of the *vic A* introns (see above) is not shown by the corresponding phaseolin sequences.

Homologies at the exon sides of the boundaries are unsurprising, in view of the conservation of intron position with respect to amino acid sequence between *vic A* and phaseolin (Section 4.1.3). Homologies at the intron side of the boundaries may reflect possible requirements of a splicing mechanism; these are likely to involve those nucleo-

FIGURE 28 Donor and acceptor splice site sequences for vicilin gene A



The total number of times a particular nucleotide occurs in a given position is scored. The data for the donor sites comes from IVS 3 and 5, that for the acceptor sites from all 4 introns so far identified (Figures 14 and 15, Results).

FIGURE 29 Comparison of vicilin and phaseolin intron boundary sequences

DONOR			ACCEPTOR	
-3-2-1	+1+2+3+4+5+6+7+8+9+10	IVS 2	-15-14-13-12-11-10-9-8-7-6-5-4-3-2-1	+1+2
		vicilin	T G T C C T T T T C A T C A G	T C
		phaseolin	A T G T T T G T C C T G T A G	G A
		IVS 3		
		vicilin	T G T A A A A A A A C A C A G	A C
		phaseolin	G C A T G A T T T T T A T A G	A G
		IVS 4		
A A T	G T A A G T A T A A	vicilin	T T C A A A T T A A T A T A G	G G
A A T	G T A A G A A A A G A	phaseolin	T G T T A A C A A A T T T A G	G G
		IVS 5		
C A G	G T A T A T T T A T	vicilin	A T T G A A A A T T T G A A G	G C
C A G	G T A T A T A T A T	phaseolin	A T T G T A A A T A T G A A G	G T

The phaseolin sequences are from Slightom *et al* (1983).

tides which are similar to the plant consensus sequence. The homologies between *vic A* and phaseolin sequences at the splice sites are overall less than those between *vic A* and the plant consensus sequence (Figure 28), however, the homology for IVS 5 is 82% at the acceptor site and 92% at the donor site (Figure 29). This supports the hypothesis that both genes have diverged from a common ancestor, as suggested by the conservation of intron positions, the sequence homology (Lycett *et al*, 1983a, Spencer, 1984) and the wide distribution of 7S proteins amongst the legume family (Derbyshire *et al*, 1976). The conservation of intron positions and close homology of nucleotide sequences at the splice sites suggests that the introns were present in the ancestral gene at these positions.

4.1.5 Size and A+T content of introns in *vic A*

The sizes of the completely sequenced introns in *vic A* (Figure 14) are, IVS 3, 104 bp and IVS 5, 131 bp, compared with 124 bp and 103 bp for the corresponding introns in phaseolin (Slightom *et al*, 1983). These introns are small compared to many from non-plant species (Breachnach and Chambon, 1981) and those of glycinin (Nielsen, 1984) or soybean leghaemoglobin (Jensen *et al*, 1981), but are similar in size to those of legumin (Lycett *et al*, 1984) and soybean actin (Shah *et al*, 1982).

It has been noted (Slightom *et al*, 1983, Lycett *et al*, 1984) that introns of higher plants tend to be A+T rich. The two introns of *vic A* for which complete sequences are available, IVS 3 and IVS 5 (Figure 14), have A+T contents

of 74.5% and 75% respectively. The corresponding figure for the coding sequence, taken from that of pDUB9, is 60.5% and so the average difference in A+T content between these introns and the exons is 14.25%. This is not such a large difference as observed for phaseolin, where the A+T content of the exons was 55.3% and the difference between this value and that of the introns was 17.1%. However, the *vic A* sequence is not yet complete and the final difference may be greater. The reason for the high A+T content of plant introns is not yet apparent (Slightom *et al*, 1983).

4.1.6 Other sequences of interest in *vic A*

The divergence of amino acid sequences between vicilin 47,000 M_r and 50,000 M_r polypeptides around the potential β/γ processing site (Gatehouse *et al*, 1983, Lycett *et al*, 1983a) has led to suggestions that precursors with the sequence neutral amino acid residue-hydrophobic amino acid residue - basic amino-acid residue at the cleavage site do not undergo proteolysis (Gatehouse *et al*, 1983). However, this sequence is not found in this region in the amino-acid sequence predicted by pDUB9 (Delauney, 1984), nor is it present in that predicted by *vic A* (Figure 15), and is therefore not universal in uncleaved precursors (Delauney, 1984).

An analysis of the sequences in the vicinity of cleavage sites of a number of seed proteins (Delauney, 1984) showed that cleavage occurs within highly hydrophilic regions, suggesting that the sites are on the outside of the protein molecule (Gatehouse *et al*, 1983). Cleavage occurs on the C-terminus of an asparagine residue (Delauney, 1984), and a

consensus sequence N↓-X_1-X_2-D/E-E at the cleavage site has been determined (Delauney, 1984) where X_1 is often a small neutral residue and X_2 a hydrophobic residue. The presence of this sequence does not always imply cleavage, which may be determined by the accessibility of the site to proteases (Delauney, 1984). The *vic A* amino acid sequence, like that predicted from pDUB9, lacks these residues in this position.

The functions of the proteolytic cleavage of vicilin precursors is not known, although it has been suggested (Lycett *et al*, 1983a; Delauney, 1984), that it may make particular sub-units more accessible to hydrolysis at different times during germination. The abundance of uncleaved 50,000 M_r precursors in vicilin isolated from mature pea seeds (Gatehouse *et al*, 1981, 1982a) shows that post-translational proteolysis is not essential for protein assembly, transport and function. Pulse-chase experiments using radioactive sugars have shown that glycosylation of vicilin occurs initially on a vicilin sub-unit of ≃50,000 M_r (Chrispeels *et al*, 1982a), The label subsequently accumulates in the ≃16,000 M_r (γ) vicilin polypeptide, indicating proteolytic cleavage of the precursor (Chrispeels *et al*, 1982a, 1984). A potential glycosylation site, N-X-T/S,has been identified in the ≃47,000 precursor amino acid sequence. This sequence is absent from the non-glycosylated 50,000 M_r vicilin polypeptide (Lycett *et al*, 1983a). The amino-acid sequence at this site predicted from the *vic A* gene, is K-A-S, as is that predicted from pDUB9 (Delauney, 1984), corresponding to a sequence R/K-A-S in the polypeptide (Lycett *et al*, 1983a).

This shows that the *vic A* gene encodes a non-glycosylated vicilin precursor. Glycosylation does not seem to be essential for synthesis, assembly and transport of vicilin (Section 1.4.2, Chrispeels *et al*, 1982a, b, Spencer, 1984), and neither legumin nor convicilin contains carbohydrate. The presence of a mixture of glycosylated and non-glycosylated peptides in the vicilin fraction of pea may therefore be associated with differential breakdown of the various components during germination, as suggested for the post-translational proteolysis of vicilin (see above).

4.2 Assessments of methods used to determine the nuclease sensitivity of genes in pea chromatin

The methods used to investigate the nuclease sensitivity of genes in pea chromatin were based on those described for other systems, *e.g.* globin genes in chicken erythrocyte nuclei (Stalder *et al*, 1980a, b), various genes in *Drosophila* embryonic nuclei (Wu *et al*, 1979a) and ribosomal genes in *Physarum polycephalum* (Stalder *et al*, 1978, 1979).

4.2.1 Nuclei isolations

Digestions were carried out on intact nuclei rather than isolated chromatin in order to maintain the native chromatin structure. It has been shown (Prentice and Gurley, 1983, Prentice *et al*, 1983) that nuclei isolation procedures can affect the kinetics of subsequent DNAase I digestions. The most important parameter was the cell concentration; the level of divalent cations, *e.g.* Mg^{2+} and

Ca^{2+} also had a small effect (Prentice and Gurley, 1983). The nuclei isolated in these experiments were from cultured Chinese hamster cells; it is therefore difficult to predict whether similar parameters are important during the isolation of nuclei from pea tissues. The number and concentration of the plant cells, for example, will depend not only on the weight of tissue and volume of buffer, but on the extent of disruption of tissue after cutting with razor blades, and the extent of digestion of the cell walls by pectinase and cellulase (Section 2.2.20.1). Cell division in pea cotyledons ceases by about 7 d.a.f., and seed development enters the cell expansion phase (Boulter, 1981). Thus the 9-18 d.a.f. cotyledons used here contained similar numbers of cells, but as similar weights of tissue were used in each isolation, the cell concentrations were variable. High cell concentrations during isolation of nuclei resulted in a decrease in the DNAase I sensitivity of chromatin in Chinese hamster cells (Prentice and Gurley, 1983) and it has been shown that diffusable cytoplasmic and nuclear factors can affect DNAase I sensitivity (Prentice *et al*, 1983). If the effects of high cell concentration are due to release of these diffusable factors, then a similar effect might be expected for the large cells of older cotyledons, which would contain more of such factors than smaller cells. Thus the lower concentrations of cells in the isolations from older cotyledon nuclei probably reduced the effective concentration of any such factors.

Chromatin structure in Chinese hamster cells was shown to be affected by swelling of cells in buffer prior to

disruption (Prentice and Gurley, 1983). This corresponds to the incubation of pea tissue with cellulase and pectinase prior to homogenisation (Section 2.2.20.1); this period was however, considerably longer than the 5 min. incubation for Chinese hamster cells (Prentice and Gurley, 1983). Rearrangements or alterations in chromatin structure may therefore have occurred during this period. However, the isolation method used here (Section 2.2.20.1) has been shown to produce intact, functional nuclei (Willmitzer and Wagner, 1981, Evans *et al*, 1984), and the results of MNase digestions (Section 3.2.3) showed that the nucleosomal structure of chromatin was maintained in these nuclei. As both leaf and cotyledon nuclei were isolated by the same procedure, and the results were reproducible (Section 3.3.2), it is unlikely that the observed differences in sensitivity of legumin genes in the two tissues (Section 3.2.3) were due to the nuclei isolation methods. The results obtained with the ribosomal genes (Section 3.3.1), which are similarly active in the leaves and cotyledons and were relatively resistant to DNAase I in both tissues (Section 4.4), also suggest that DNAase I sensitivity was unaffected by nuclei isolation procedures.

4.2.2 Digestions of nuclei

4.2.2.1 Use of similar amounts of DNA per digestion

The use of similar amounts of DNA in each digestion, rather than numbers of nuclei (Section 3.2.1) was justified here because of the large increase in DNA content of cotyledon nuclei during seed maturation (Millerd and Spencer, 1974, Smith, 1973, Boulter, 1981 and Section 1.3.1).

If similar numbers of nuclei were used, up to 50-64 times more DNA (Millerd and Spencer, 1974, Smith, 1973) would be present in cotyledon rather than leaf nuclei. As there is no evidence that DNA above the 2C level acts as a template for RNA synthesis (Millerd and Spencer, 1974), the excess of DNA might mask any sensitivity of the fraction of DNA which does act as a template. The large amount of DNA substrate in the older cotyledon nuclei might also have caused the enzyme concentration to become limiting, resulting in a false estimation of the rate of digestion. Similar amounts of DNA, rather than numbers of nuclei, were also used in an investigation of the nuclease sensitivity of the β-phaseolin gene in cotyledons and leaves of *Phaseolus vulgaris* (Murray and Kennard, 1984).

4.2.2.2 Activities of DNAase I solutions

The results of the spectrophotometric assay for DNAase I activity (Section 3.2.2) showed that it was approximately 2/3rds of optimal activity. The reasons for this may have been partly due to the dilutions; the DNAase I solid contained 2000 or 2670u/mg and thus the 1mg/ml (or greater concentration) stocks had to be diluted extensively to obtain the range from 20-100u/ml. In addition, the degree of polymerisation of high molecular weight DNA cannot be determined once it is in solution (Millipore Corporation, 1979) and thus in order to determine precisely the activity of an unknown DNAase I solution it is usual to compare the activity with that of a standard solution (Millipore Corporation, 1979). This was not performed for the solutions here as only the relative activities, before and after storage,

were required. The retention of activity during storage in digestion buffer (Table 8) showed that comparisons of nuclease sensitivity between samples digested with enzyme stored for different times were still valid; in practice, however, samples to be compared were usually digested at the same time with the same aliquot of enzyme. The DNAase I stock which was stored in 0.1N HCl was discarded as it was shown to be inactive in both the spectrophotometric (Section 3.2.2) and electrophoresis assays (Section 3.2.3).

The various investigations into the establishment of digestion conditions showed that both DNAase I and MNase digested DNA and chromatin substrates under the conditions used (Section 3.2.3). The 80 to 100-fold greater DNAase I sensitivity of DNA compared to that of chromatin has been found previously (Kunnath and Locker, 1985). Although preliminary investigations showed that a slightly greater extent of digestion was obtained with the Worthington enzyme (Section 3.2.4), the DNAase I obtained from Sigma was used throughout; this was because limited, rather than extensive digestion was required. The use of PMSF, which appeared to increase the rate of digestion, was also avoided because of this requirement. The effects of PMSF on pea DNA as well as on pea nuclei suggest that it affects other processes besides acting as a proteinase inhibitor. A phosphatase inhibitor has been found to reduce the rate of DNAase I digestion (Prentice and Gurley, 1983); it thus seems advisable to omit all types of inhibitors from buffers used for nuclease digestions.

4.2.2.3 Comparison of DNAase I digestion conditions with those used in other studies

The DNAase I and MNase I digestion buffers used here (Section 2.2.20.3) were similar to those used for the detection of altered chromatin structures in *Drosophila* embryo nuclei (Wu *et al*, 1979a). The method of purification of DNA from nuclei was based on that used for *Physarum polycephalum* (Stalder *et al*, 1978, 1979). The purified DNA was cleaved by restriction endonucleases, as shown by the control tracks in the autoradiographs (Figure 22B, tracks f, j, Figure 25B, tracks a, f). The methods subsequently described for nuclei isolations, digestions with nucleases and DNA purifications in plants (Spiker, 1983, Murray and Kennard, 1984, Ferl, 1985) are similar to those used in this work.

The levels of enzyme used in DNAase I digestions in the literature vary over a wide range from less than 1u/ml (Murray and Kennard, 1984) to 16u/ml (Wu *et al*, 1979a). Many of the concentrations are given as µg of DNAase I/ml, and so the activity cannot be compared with those used here unless the units/mg of solid enzyme are given. The range of concentrations used in this work was initially from 2 to 15u/ml, usually 10u/ml, although later a range of 0.01 to 0.5 u/ml was used (Section 3.4).

The amounts of DNA per digestion also vary from 1mg/ml (Wu *et al*, 1979a, Stalder *et al*, 1980a, b) to 0.3-0.4mg/ml (Stalder *et al*, 1978) or 0.1mg/ml (Murray and Kennard, 1984). The concentrations used here were from 0.1mg/ml to 0.38mg/ml, *i.e.* in the lower end of the range.

The two other major variables in digestions are the times and temperatures of incubation. The digestions in animal systems are often at 37°C for 10 min. (Stalder *et al*, 1980 a, b, Burch and Weintraub, 1983) although some were digested for shorter times at lower temperatures, *e.g.* 25°C for 3 min (Wu *et al*, 1979a). In plants digestions ranged from 37°C for 10 sec. to 10 min. (Leber and Hemleben, 1979), to 15°C for 30 min. (Murray and Kennard, 1984). In the present work, the temperatures of digestion varied from 25°C to 12°C and the times of incubation from 30 sec. to 30 min. The conditions required for the digestion of plant chromatin were milder than those for animal chromatin (Murray and Kennard, 1984); this was supported by the rapid loss of even the inactive seed storage protein genes in pea leaf nuclei (Figure 25B and Section 4.5), when digested with DNAase I at 10u/ml for short periods of time.

Overall, the methods described here (Section 2.2.20) were such as to permit detection of nuclease sensitivity of specific genes in pea. The conditions employed to obtain mild levels of digestion (Section 3.4) were similar to those used for French bean cotyledons and leaves (Murray and Kennard, 1984). This suggests that such conditions may be the most appropriate for nuclease digestions of chromatin from leguminous plants.

4.3 Micrococcal nuclease digestions of plant chromatin

The nucleosomal repeat length for pea chromatin was 171±25 bp (Section 3.2.3) and was similar in both cotyledon

and leaf nuclei. A similar repeat length, 170±30 bp, has been obtained for chromatin from pea seedings (McGhee and Engel, 1975). These results contrast with those obtained for the chromatin of *Phaseolus vulgaris* (Murray and Kennard, 1984); the repeat length was 191±6 bp in nuclei from leaves and mature cotyledons, and 177±7 bp in nuclei from developing cotyledons. This difference was observed both in total chromatin on an EtBr-stained gel and when probes for ribosomal or phaseolin gene sequences were hybridised to a blot of the gel, (Murray and Kennard, 1984). A similar shortening of the nucleosomal repeat length has been observed during sea-urchin embryogenesis (Spinelli *et al*, 1982, Section 1.6).

It has been suggested that the change in repeat length is associated with DNA replication rather than transcription (Murray and Kennard, 1984), as both young leaves and developing cotyledons are highly active transcriptionally. If this is correct, then the results for pea are surprising, as the cotyledon cells, like those of bean, undergo extensive replication of DNA (Millerd and Spencer, 1979, Smith, 1973) during seed development. The width of some of the bands in the nucleosomal ladder (Figure 19, Section 3.2.3) complicates the measurement of the size of the fragments within the band, resulting in the large standard error in repeat length (Section 3.2.3). The data for French bean chromatin were obtained by scanning negatives or autoradiographs with a densitometer. A similar approach is required in order to establish whether or not the repeat-length in pea chromatin changes during seed maturation.

4.4 DNAase I sensitivity of ribosomal genes in pea chromatin

The results of experiments designed to test the nuclease sensitivity of ribosomal genes in leaf and cotyledon nuclei (Section 3.3.1) also showed that the conditions of hybridisation allowed the detection of specific fragments within the smear of DNA (Figure 22B, Section 3.3.1). Therefore the methods were suitable for the study of nuclease sensitivity in pea chromatin, at least for multiple copy genes.

The presence of bands in the genomic DNA and nuclei which hybridised to pHA1 but were not detected in the Bam HI digest of pHA1 (track (a), Figure 22B) and could not be explained from the map of the ribosomal repeat (Figure 23), has been explained as due to partial restriction (Section 3.3.1). This is supported by the DNA restriction profiles on the EtBr-stained gel (tracks (b), (f), (j), Figure 22B). The poor restriction may have been due to insufficient enzyme or too short a time of digestion, however, it is possible that sites may not have been cut (Figure 23) because they were modified in some way which prevented recognition by the enzyme. There is evidence that the ribosomal genes in both leaves and cotyledons of pea are heavily methylated (Waterhouse, 1985). Thus at least some of the bands may be due to methylation of CpG dinucleotides at the Bam Hl sites, rendering them refractory to cleavage by this enzyme. The loss of these bands in some tracks may therefore have been due to DNAase I digestion, rather than restriction. All the bands present in the DNAase I digests of cotyledon and leaf nuclei were also found in at least one digest of protein-free DNA (Figure 22B), so there was no evid-

ence for chromatin-specific cleavages by DNAase I, *i.e.* hypersensitive sites, in the ribosomal genes.

Differences in the nuclease sensitivity between active and inactive genes are usually detected by the relative rates of disappearance of the restriction fragments corresponding to the genes (Sections 1.7 and 2.2.20.7). The differences in intensities of the bands hybridising to the ribosomal genes can be largely explained by the variations in the extent of restriction and by the size distributions of DNA fragments (Figure 22B and Section 3.3.1). It thus appears that the ribosomal genes in pea leaf and cotyledon nuclei show similar extents of degradation by DNAase I, and are in fact relatively resistant to digestion by this enzyme as fragments still persist in the 15u/ml tracks (Figure 22B, tracks (i), (m)).

Comparison of the rate of DNAase I digestion of ribosomal genes in chromatin of *Brassica pekinensis* and *Matthiola incana* to that of total chromatin showed that there was no preferential digestion of ribosomal sequences (Leber and Hemleben, 1979). The initial rate of digestion of ribosomal genes was slightly slower than that of total chromatin, suggesting that they were more resistant to low levels of enzyme than was bulk chromatin. A similar result was found for ribosomal genes in leaf and cotyledon nuclei of *Phaseolus vulgaris* (Murray and Kennard, 1984). When the blot which had been used to demonstrate the sensitivity of the active phaseolin gene (Section 1.7) was washed to remove the phaseolin probe, and rehybridised to pHA1, no loss of fragments hybridising to the latter probe was detected across an 8-fold range of DNAase I concentrations (Murray and Kennard, 1984).

This apparent resistance of plant ribosomal chromatin to DNAase I digestion contrasts with results obtained in a number of other organisms. The ribosomal genes of the slime mould *Physarum polycephalum* have been shown to be preferentially sensitive to DNAase I (Stalder *et al*, 1978) and DNAase I hypersensitive sites have been detected in the ribosomal chromatin of *Tetrahymena*, a protozoan (Bonven and Westergaard, 1982). DNAase I sensitivity is correlated with the expression of ribosomal genes of only one species in hybrids between *Xenopus laevis* and *X. borealis*; the genes from the other species are transcriptionally inactive and resistant to DNAase 1 (Macleod and Bird, 1982). The conflicting results may be reconciled by the suggestion (Murray and Kennard, 1984) that most of the ribosomal genes in plants are inactive even in tissues where the transcription rate is high, and that the resistance of these inactive genes masks the preferential sensitivity of the few active ones. This is supported by the observation (Garel *et al*, 1977) that both rarely and frequently transcribed genes show similar sensitivities to DNAase I; thus it might be predicted that the sensitivity of active ribosomal genes would be similar to that of the legumin genes in cotyledons (Figure 25, Section 3.3.2). Despite the differences in copy number, *i.e.* ∿8 genes for legumin (Domoney and Casey, 1985) and 7,800 genes for ribosomal RNA (Ingle, 1979) the persistence of the ribosomal bands after 6 min. of digestion with 15u/ml DNAase 1 (Figure 22B, tracks (i), (m)) compared to the loss of legumin genes in cotyledons after 1 min. of digestion with 10u/ml DNAase 1 (Figure 25B, track (i)) suggests that it is the inactive ribosomal genes which are being detected.

4.5 DNAase I sensitivity of pea seed storage protein genes

The different rates of digestion by DNAase I of the legumin genes in pea cotyledons and leaves (Figure 25B, and Section 3.3.2) correlate with the transcriptional activity of these genes in the two tissues (Gatehouse *et al*, 1982, Evans *et al*, 1984). The bands in cotyledons disappear at about 1 min. of digestion (Figure 25B, track (i)) whereas bands are still just visible in the leaf samples after 2 min. of digestion (Figure 25B, track (e)). These results suggest that the genes in cotyledons are 2-4 times as sensitive to DNAase I as those in leaves; this is similar to the 5-fold difference reported for the active and inactive ovalbumin genes (Garel and Axel, 1976) but is somewhat lower than the order of magnitude difference seen in most examples (Reeves, 1984). Further experiments involving mild digestion conditions over a wide range of enzyme concentrations, such as that in Section 3.4, combined with densitometric analysis of autoradiographs, should allow a more precise determination of the relative sensitivities of the active and inactive genes.

The demonstration that active pea seed storage protein genes are more sensitive than inactive genes is similar to the results obtained for the β-phaseolin gene in French bean (Murray and Kennard, 1984), the *Adh-1* gene promoter in maize (Ferl, 1985) and transcriptionally active chromatin in wheat embryos (Spiker *et al*, 1983). Thus it appears that, for class II genes at least, the structure of transcriptionally active plant chromatin is altered such that it is more accessible to nucleases (Section 1.7).

The expression of legumin and β-phaseolin is developmentally regulated (Section 1.2). Analogous systems such as globin synthesis during erythropoiesis (Stalder *et al*, 1980a, b), were among the first to be studied with respect to DNAase I sensitivity of animal chromatin (Weintraub and Groudine, 1976, Garel and Axel, 1976, Stalder *et al*, 1980a,b, Shermoen and Beckdorf, 1982). However, a number of inducible genes have also been investigated such as the heat-shock genes in *Drosophila melanogaster* (Wu *et al*, 1979 a, b, Wu, 1980), a gluco-corticoid-dependent viral enhancer element (Zaret and Yamamoto, 1984) and induced globin synthesis in murine erythroleukaemia cells (Sheffery *et al*, 1984, Salditt-Georgieff *et al*, 1984). The promoter of the maize *Adh-1* gene, which is induced in anaerobic roots, has been shown to be DNAase I sensitive (Ferl, 1985), although it was not stated whether this sensitivity was associated with induction. The accessibility of various restriction sites within this promoter does however change on the induction of this gene in anaerobic roots (Ferl, 1985) and is discussed in more detail in Section 4.6. A number of plant genes, such as the small sub-unit of ribulose bisphosphate carboxylase, are regulated by light (Tobin and Silverthorne, 1985). Therefore, studies of nuclease sensitivity of light-inducible genes in plants are possible, provided that suitable methods for the isolation of nuclei from dark-grown and light-grown tissues, (*e.g.* Gallagher and Ellis, 1982), and appropriate hybridisation probes, are available. Such an analysis of the chromatin structure of light-regulated and other inducible genes in plants will determine whether the preferential DNAase I

sensitivity of active genes in plants is a general phenomenon, as appears to be the case in animal chromatin.

4.6 DNAase I hypersensitive sites - are these structures present in plant chromatin?

The rapid digestion of the inactive seed storage protein genes in leaf nuclei (Section 3.3.2, Figure 25B, tracks (a) to (e)) suggests that the overall degradation of chromatin was fairly extensive, and thus may have masked any hypersensitive sites (Section 3.4) as these are often two orders of magnitude more sensitive than bulk chromatin (Reeves, 1984). However, had such sites been present, sub-bands in addition to those detected in the controls (Figure 25B, tracks (b) and (f)) might have been expected in the 0.5 min. or zero-time samples of cotyledon nuclei (tracks (k) and (l)). No such bands were observed in this experiment, and the hybridisation of a vicilin cDNA to a blot of very mild DNAase I digestions (Figure 26) also suggested that no sub-bands were present, although the results here were difficult to interpret (Section 3.3.4). Thus the data available for pea suggests that there are no DNAase I hypersensitive sites in chromatin.

This result is similar to that obtained for the β-phaseolin gene in *Phaseolus vulgaris* (Murray and Kennard, 1984) in which general DNAase I sensitivity (see previous section) was detected but DNAase I hypersensitive sites were not. It is unlikely that conditions which permitted the detection of general DNAase I sensitivity would not have detected hypersensitive sites if they were present (Murray and Kennard,1984).

However, it is possible that the probe and restriction endonucleases used were inappropriate in both that study and the present work. The probe used to assay for hypersensitive sites in phaseolin was a 1.6 kb fragment extending from 110 bp within the gene to 1471 bp 5′ to the start of transcription (Murray and Kennard, 1984). The DNA purified from French bean nuclei was restricted so that the gene was cut into two fragments, one corresponding to the probe and the other from the site 110 bp within the gene to a site well past that of poly(A) addition (Murray and Kennard, 1984). In the experiments on legumin genes in pea, digestion of the DNA with Hind III gave two fragments, one of which (∿2.5 kb, Section 3.3.2) contained all of the leg A gene downstream of the site just after the translational start point (see Figure 24 and Lycett *et al*, 1984). The second fragment, thought to be that of 4.15 kb (Section 3.3.2), contains all of the 5′ promoter and flanking sequences of leg A (Figure 24). The probe used to assay for hypersensitive sites was the insert of pDUB24 (Figure 24, Section 3.3.2). If there had been any 5′ hypersensitive sites in either gene, then the restriction fragments corresponding to the flanking regions would have been cut into two or more bands detected by the probes. The absence of such bands (Figure 25B, and Murray and Kennard, 1984) suggests that there are no such sites 5′ to the gene, however, if the sites were very close to the start of the gene it is possible that the smaller fragment generated by a double-stranded DNAase I cleavage, and the resulting change in size of the larger fragment, would not be detected. Hypersensitive sites are detected close to the

cap site in the heat-shock genes of *Drosophila* (Wu, 1980, Wu, 1984a) as well as further upstream. As many genes which have been studied have more than one 5´ hypersensitive site, *e.g. Drosophila* heat-shock genes (Wu, 1980, Wu, 1984a), *Drosophila Sgs* 4 genes (Shermoen and Beckendorf, 1982), chicken β-globin genes (Stalder *et al*, 1980b), it would be expected that even if sites close to the cap site were not detected in legumin and phaseolin because of the probe and restriction enzyme combination used, other sites further upstream would have been detected if present. However, further experiments using probe/restriction enzyme combinations that will allow the detection of sites close to the start of the gene, are necessary.

When the β-phaseolin gene was ligated into pBR322 and tested for S1 sensitivity (see Section 1.9), it was found that this enzyme cut at a number of sites 5´ and 3´ to the gene when the plasmid was supercoiled (Murray and Kennard, 1984). A correlation could be made between sequences with the potential for cruciform formation, and S1-sensitive sites; however, the absence of both DNAase I and S1-hypersensitive sites *in vivo* suggests that such studies on supercoiled plasmids are of limited value in predicting the chromatin structure of these sequences (Murray and Kennard, 1984). Similar experiments were therefore not performed for the pea seed storage protein genes.

The lack of any DNAase I hypersensitive sites in the maize *Adh-1* gene (Ferl, 1985) supports the results from the legumin and phaseolin genes. However, examination of the accessibility of three restriction sites upstream of the

promoter of this gene suggests that the chromatin structure of this region changes on induction of the gene. The order of the sites is Pst I, Xba I and Bam Hl, with Pst I being nearest to the gene. Comparison of the cutting by each enzyme in chromatin from anaerobic roots, in which the gene is induced, with that in aerobic roots, in which the gene is inactive, showed the Pst I and Xba I sites were much more accessible in chromatin from anaerobic roots. This shows that the chromatin was in a more open conformation in the active gene. In nuclei from leaves, a tissue in which the *Adh-1* gene is not expressed, the Bam Hl and Xba I sites were inaccessible, but the Pst I site was cut to the same extent as in anaerobic roots. It has therefore been suggested that on induction during anaerobiosis the entire promoter region becomes more accessible than in aerobic roots, but that in leaves, specific binding of a protein rather than general chromatin condensation represses the gene (Ferl, 1985). At first, these results seem at variance with the lack of DNAase I hypersensitive sites in plant chromatin. However, the patterns of accessibility to restriction endonucleases, which are by definition site-specific, can be explained in terms of a general opening up of chromatin on induction (Ferl, 1985), as suggested by the general increase in DNAase I sensitivity of the active pea legumin and French bean phaseolin genes. The pattern of cutting in leaves can be explained if the putative repressor protein (Ferl, 1985) binds to a region covering the Bam Hl and Xba I sites, but not the Pst I site.

The information so far available for specific genes in plant chromatin suggests that, unlike animal genes, they lack

DNAase I hypersensitive sites. However, the presence of these sites in animal chromatin is only indicative of a particular structure which results in very high sensitivity of the DNA to this enzyme, and there may be many other chromatin structural alterations which cannot be detected by this nuclease. More experiments on a range of genes in different plant species, using various nucleases, and other probes of chromatin structure such as copper phenanthroline (Drew, 1984, Cartwright and Elgin, 1982), are required to establish the detailed chromatin structure of plant genes. The mild conditions required for digestion of plant chromatin (Murray and Kennard, 1984) suggest that it may be generally more sensitive to nucleases than animal chromatin, and this high level of general sensitivity may mask any specific hypersensitive regions. Experiments involving one of the footprinting assays used to detect protection by DNA-binding proteins (Section 1.10) may therefore be more useful than nuclease digestions in probing the details of plant chromatin structure, and protein/DNA interactions. Such an assay may also be useful in determining whether there is a protein in pea cotyledons which binds to the putative tissue-specific sequence in the legumin genes (Section 1.4.1).

A number of differences at the sequence level have been detected between animal and plant genes (Sections 1.4 and 4.1.3 to 4.1.5, Lycett *et al*, 1983b, Messing *et al*, 1983) and it would not be surprising if differences existed at higher levels of structure. Investigations into chromatin structure should therefore be combined with those on DNA sequences in efforts to understand the mechanisms involved in regulating differential gene expression in plants.

REFERENCES

Barton, K.A. and Chilton, M-D. (1983), *Agrobacterium* T_i plasmids as vestors for plant genetic engineering. Methods Enzymol. 101, 527-539.

Benoist, C. and Chambon, P. (1981), *In vivo* sequence requirements of the SV40 early promoter region. Nature 290, 304-310.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P., (1980), The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. 8, 127-142.

Benton, W.D. and Davis, R.W. (1977), Screening λ gt recombinant clones by hybridization to single plaques *in situ* Science 196, 180-182.

Bergman, L.W. and Kramer, R.A. (1983), Modulation of chromatin structure associated with derepression of the acid phosphatase gene of *Saccharomyces cerevisiae*. J.Biol. Chem. 258, 7223-7227.

Bernardi, A., Gaillard, C. and Bernardi, G. (1975), The specificity of five DNAases as studied by the analysis of 5´-terminal doublets. Eur.J.Biochem. 52, 451-457.

Bethesda Research Laboratories, Inc. (1982), M13 cloning/ dideoxy sequencing. BRL instruction manual.

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

Blobel, G. and Dobberstein, B. (1975), Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin. Light chains on membrane-bound ribosomes of murine myeloma. J.Cell Biol. 67, 835-851.

Bolivar, F., Rodriguez, R.L., Green, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977), Construction and characterisation of new cloning vehicles. II. A multi-purpose cloning system. Gene 2, 95-113.

Bonven, B. and Westergaard, O. (1982), DNAase I hypersensitive regions correlate with a site-specific endogenous nuclease activity on the r-chromatin of *Tetrahymena*. Nucleic Acids Res. 10, 7593-7608.

Boulter, D. (1981), Biochemistry of storage protein synthesis and deposition in the developing legume seed. Adv. Bot.Res. 9, 1-31.

Bown, D., Levasseur, M., Croy, R.R.D., Boulter, D. and Gatehouse, J.A. (1985), Sequence of a pseudogene in the legumin gene family of pea (*Pisum sativum* L.). Nucleic Acids Res. 13, 4527-4538.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978), Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. U.S.A., 75, 4853-4857.

Breathnach, R. and Chambon, P. (1981), Organisation and expression of eucaryotic split genes coding for proteins. Ann.Rev. Biochem., 50, 349-383.

Broglie, R., Coruzzi, G., Fraley, R.T., Rogers, S.G., Horsch, R.B., Niedermeyer, J.G., Fink, C.L., Flick, J.S. and Chua, N.-H (1984)., Light-regulated expression of a pea ribulose -1,5-bisphosphate carboxylase small subunit gene in transformed plant cells. Science 224, 838-843.

Brown, D.D. (1984), The role of stable complexes that repress and activate eucaryotic genes. Cell 37, 359-365.

Burch, J.B.E. and Weintraub, H. (1983), Temporal order of chromatin structural changes associated with activation of the major chicken vitellogenin gene. Cell 33, 65-76.

Butler, P.J.G. (1984), A defined structure of the 30 nm chromatin fibre which accommodates different nucleosomal repeat lengths. EMBO J. 3, 2599-2604.

Cameron, J.R., Philippsen, P. and Davis, R.W. (1977), Analysis of chromosomal integration and deletions of yeast plasmids. Nucleic Acids Res. 4, 1429-1448.

Cartwright, I.L. and Elgin, S.C.R. (1982), Analysis of chromatin structure and DNA sequence organisation: use of the 1,10-phenanthroline-cuprous complex. Nucleic Acids Res. 10, 5835-5852.

Casey, R. and Domoney, C. (1984) , The genetics of legume storage proteins. Phil.Trans.Roy.Soc. Lond.B., 304, 349-358.

Casey, R., Domoney, C., and Stanley, J. (1984), Convicilin mRNA from pea (*Pisum sativum* L.) has sequence homology with other legume 7S storage protein mRNA species. Biochem.J., 224, 661-666.

Casey, R., March, J.F., Sharman, J.E. and Short, M.N., (1981). The purification, N-terminal amino acid sequence and some other properties of an α^{M}-subunit of legumin from the pea. (*Pisum sativum* L.), Biochim.Biophys.Acta. 670, 428-432.

Casey, R. and Short, M.N., (1981), Variation in amino acid composition of legumin from *Pisum*. Phytochem. 20,21-23.

Chang, A.C.Y. and Cohen, S.N. (1978), Construction and characterisation of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J.Bacteriol. 134, 1141-1156.

Chrispeels, M.J. (1984), Biosynthesis, processing and transport of storage proteins and lectins in cotyledons of developing legume seeds. Phil.Trans.Roy.Soc. Lond. B 304, 309-322.

Chrispeels, M.J., Higgins, T.J.V., Craig, S. and Spencer, D. (1982a), Role of the endoplasmic reticulum in the synthesis of reserve proteins and the kinetics of their transport to protein bodies in developing pea cotyledons. J.Cell.Biol. 93, 5-14.

Chrispeels, M.J., Higgins, T.J.V. and Spencer, D. (1982b), Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. J.Cell.Biol. 93, 306-313.

Conklin, K.F. and Groudine, M. (1984), Chromatin structure and gene expression. In 'DNA Methylation' (Razin, A., Cedar, H., Riggs,A.D., eds.), Springer-Verlag, New York, pp.293-351.

Courey, A.J. and Wang, J.C. (1983), Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. Cell 33, 817-829.

Croy, R.R.D., Derbyshire, E., Krishna, T.G. and Boulter, D. (1979, Legumin of *Pisum sativum* and *Vicia faba*, New Phytol. 83, 29-35.

Croy, R.R.D. and Gatehouse, J.A. (1985), Genetic engineering of seed proteins - current and potential applications. In 'Plant genetic engineering' (Dodds, J.H., ed.), Cambridge University Press, pp.143-268.

Croy, R.R.D., Gatehouse, J.A., Evans, I.M. and Boulter, D. (1980a), Characterisation of the storage protein subunits synthesised *in vitro* by polyribosomes and RNA from developing pea (*Pisum sativum* L.), I. Legumin. Planta 148, 49-56.

Croy, R.R.D., Gatehouse, J.A., Evans, I.M. and Boulter, D. (1980b), Characterisation of the storage protein subunits synthesised *in vitro* by polyribosomes and RNA from developing pea (*Pisum sativum* L.). II Vicilin, Planta 148, 57-63.

Croy, R.R.D., Gatehouse, J.A., Tyler, M. and Boulter, D. (1980c), The purification and characterisation of a third storage protein (convicilin) from the seeds of pea (*Pisum sativum* L.), Biochem. J. 191, 509-516.

Croy, R.R.D., Hoque, M.S., Gatehouse, J.A., and Boulter, D.(1984), The major albumin proteins from pea (*Pisum sativum* L.). Purification and some properties. Biochem.J. 218, 795-803.

Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N. and Boulter, D. (1982), Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature, 295, 76-79.

Croy, R.R.D., Shirsat, A.H. and Boulter, D., (1985), manuscript in preparation.

Cuellar, R.E., (1982), Ph.D. thesis ,Stanford University.

Davey, R.A., Higgins, T.J.V. and Spencer, D. (1981), Homologies between two small subunits of vicilin from *Pisum sativum* Biochem. Int. 3, 595-602.

Delauney, A.J. (1984), Cloning and characterisation of cDNAs encoding the major pea storage proteins, and expression of vicilin in *E. coli*. Ph.D. thesis, University of Durham.

Derbyshire, E., Wright, D.J. and Boulter, D. (1976), Legumin and vicilin, storage proteins of legume seeds. Phytochem.. 15, 3-24.

Doerfler, W. (1983), DNA methylation and gene activity. Ann.Rev.Biochem. 52, 93-124.

Domoney, C. and Casey, R. (1983), Cloning and characterisation of complementary DNA for convicilin, a major seed storage protein in *Pisum sativum* L. Planta 159, 446-453.

Domoney, C. and Casey R. (1984), Storage protein precursor polypeptides in cotyledons of *Pisum sativum* L. Identification of, and isolation of a cDNA clone for, an 80,000 -M_r legumin-related polypeptide. Eur.J.Biochem.139 321-327.

Domoney, C. and Casey, R. (1985), Measurement of gene number for seed storage proteins in *Pisum*. Nucleic Acids Res. 13, 687-699.

Drew, H.R. (1984), Structural specificities of five commonly used DNA nucleases. J.Mol.Biol. 176, 535-557.

Drew, H.R. and Travers, A.A. (1984), DNA structural variations in the *E.coli tyrT* promoter. Cell 37, 491-502.

Drew, H.R. and Travers, A.A. (1985), Structural junctions in DNA: the influence of flanking sequence on nuclease digestion specificities. Nucleic Acids.Res. 13, 4445-4467.

Drummond, M.H. (1984), More prospects for T_i plasmids. Nature 311, 706.

Elgin, S.C.R. (1981), DNAase I-hypersensitive sites of chromatin, Cell 27, 413-415.

Elgin, S.C.R. (1984), Anatomy of hypersensitive sites, Nature 309, 213-214.

Emerson, B.M. and Felsenfeld, G. (1984), Specific factor conferring nuclease hypersensitivity at the 5′ end of the chicken adult β-globin gene. Proc.Natl.Acad.Sci. U.S.A. 81, 95-99.

Emerson, B.M., Lewis, C.D. and Felsenfeld, G. (1985), Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin gene: nature of the binding domain. Cell 41, 21-30.

Evans, I.M., Bown, D., Lycett, G.W., Croy, R.R.D., Boulter, D. and Gatehouse, J.A. (1985), Transcription of a legumin gene from pea (*Pisum sativum* L.) *in vitro*. Planta 165, 554-560.

Evans, I.M., Croy, R.R.D., Hutchinson, P., Boulter, D., Payne, P.I. and Gordon, M.E. (1979), Cell-free synthesis of some storage protein subunits by polyribosomes and RNA isolated from developing seeds of pea (*Pisum sativum* L.), Planta 144, 455-462.

Evans, I.M., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984), Regulation of the transcription of storage-protein mRNA in nuclei isolated from developing pea (*Pisum sativum* L) cotyledons. Planta 160, 559-568.

Ferl, R.J. (1985), Modulation of chromatin structure in the regulation of the maize *Adh 1* gene. Mol.Gen.Genet.200, 207-210.

Finch, J.T. and Klug, A. (1977), X-ray and electron microscope analyses of crystals of nucleosome cores. Cold Spring Harbor Symp. Quant. Biol. 42, 1-9.

Flavell, R. and Mathias, R. (1984), Prospects for transforming monocot crop plants. Nature 307, 108-109.

Frischauf, A.-M, Lehrach, H., Poustka, A. and Murray, N. (1983), Lambda replacement vectors carrying polylinker sequences. J.Mol.Biol. 170, 827-842.

Fritton, H.P., Igo-Kemenes, T., Nowock, J., Strech-Jurk, U, Theisen, M. and Sippel, A.E. (1984), Alternative sets of DNAase I - hypersensitive sites characterize the various functional states of the chicken lysozyme gene. Nature 311, 163-165.

Galas, D. and Schmitz, A. (1978), DNAase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. 5, 3157-3170.

Gallagher, T.F. and Ellis, R.J. (1982), Light-stimulated transcription of genes for two chloroplast polypeptides in isolated pea leaf nuclei. EMBO J. 1, 1493-1498.

Garel, A. and Axel, R. (1976), Selective digestion of transcriptionally active ovalbumin genes from oviduct nuclei. Proc.Natl.Acad.Sci. U.S.A. 73, 3966-3970.

Garel, A. Zolan, M. and Axel, R. (1977), Genes transcribed at diverse rates have a similar conformation in chromatin. Proc.Natl.Acad.Sci. U.S.A. 74, 4867-4871.

Gatehouse, J.A, Croy, R.R.D. and Boulter, D. (1980), Isoelectric-focussing properties and carbohydrate content of pea (*Pisum sativum*) legumin. Biochem.J. 185, 497-503.

Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984), The synthesis and structure of pea storage proteins. CRC Crit.Rev. Plant Sci. 1, 287-314.

Gatehouse, J.A., Croy, R.R.D., Morton, H., Tyler, M. and Boulter, D. (1981), Characterisation and subunit structures of the vicilin storage proteins of pea (*Pisum sativum* L), Eur.J. Biochem. 118, 627-633.

Gatehouse, J.A, Evans, I.M., Bown, D., Croy, R.R.D., and Boulter, D. (1982a), Control of storage-protein synthesis during seed development in pea (*Pisum sativum* L.), Biochem.J. 208, 119-127.

Gatehouse, J.A., Gilroy, J., Hoque, M.S. and Croy, R.R.D. (1985), Purification, properties and amino acid sequence of a low-M_r abundant seed protein from pea (*Pisum sativum* L.), Biochem. J. 225, 239-247.

Gatehouse, J.A., Lycett, G.W., Croy, R.R.D. and Boulter, D. (1982b), The post-translational proteolysis of the subunits of vicilin from pea (*Pisum sativum* L.), Biochem.J. 207, 629-632.

Gatehouse, J.A., Lycett, G.W., Delauney, A.J., Croy, R.R.D. and Boulter, D. (1983), Sequence specificity of the post-translational proteolytic cleavage of vicilin, a seed storage protein of pea (*Pisum sativum* L.), Biochem, J. 212, 427-432.

Gilroy, J., Wright, D.J. and Boulter, D. (1979), Homology of basic subunits of legumin from *Glycine max* and *Vicia faba*. Phytochem. 18, 315-316.

Graham, D.E. (1978), The isolation of high molecular weight DNA from whole organisms or large tissue masses. Anal.Biochem. 85, 609-613.

Groudine, M. and Weintraub, H. (1982), Propagation of globin DNAase I - hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. Cell 30, 131-139.

Grunstein, M. and Hogness, D.S. (1975), Colony hybridisation: a method for the isolation of cloned DNAs that contain a specific gene. Proc.Natl.Acad.Sci.U.S.A. 72, 3961-3965.

Hamada, H., Petrino, M.G. and Kakunaga, T., (1982), A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc.Natl.Acad.Sci. U.S.A. 79, 6465-6469.

Herrera-Estrella, L., Van Den Broeck, G., Maenhaut, R., Van Montagu, M. and Schell, J., Timko, M. and Cashmore, A. (1974), Light-inducible and chloroplast-associated expression of a chimaeric gene introduced into *Nicotiana tabacum* using a T_i plasmid vector. Nature 310, 115-120.

Higgins, T.J.V. (1984), Synthesis and regulation of major proteins in seeds. Ann.Rev.Plant Physiol. 35, 191-221.

Higgins, T.J.V. and Spencer, D. (1980), Biosynthesis of pea seed proteins: evidence for precursor forms from *in vivo* and *in vitro* studies. In 'Genome Organisation and Expression in Plants' (Leaver, C.J. ed.), N.A.T.O. Advanced Study Institute Series, Plenum Press, New York and London, pp.245-253.

Higgins, T.J.V. and Spencer, D. (1981), Precursor forms of pea vicilin subunits. Modification by microsomal membrances during cell-free translation. Plant Physiol. 67, 205-211.

Hirano, H., Gatehouse, J.A. and Boulter, D. (1982), The complete amino acid sequence of a subunit of the vicilin seed storage protein of pea (*Pisum sativum* L.) FEBS. Lett. 145, 99-102.

Hooykaas-Van Slogteren, G.M.S., Hooykaas, P.J.J. and Schilperoort, R.A. (1984), Expression of T_i plasmid genes in monocotyledonous plants infected with *Agrobacterium turnefaciens*. Nature 311, 763-764.

Hyldig-Nielsen, J.J., Jensen, E.Ø., Paludan, K. Wiborg, O., Garrett, R., Jørgensen, P. and Marcker, K.A. (1982), The primary structures of two leghaemoglobin genes from soybean. Nucleic Acids Res. 10, 689-701.

Igo-Kemenes, T., Hörz, W. and Zachau, H.G. (1982), Chromatin. Ann.Rev.Biochem. 51, 89-121.

Igo-Kemenes, T. and Zachau, H.G. (1977), Domains in Chromatin Structure. Cold Spring Harbor Symp. Quant, Biol. 42, 109-118.

Ingle, J. (1979), Plant ribosomal RNA genes - a dynamic system? In 'Molecular Biology of Plants' (Rubenstein, I., Phillips, R.L., Green, C.E. and Gengenbach, B.G.), Academic Press, New York, pp.139-164.

Jackson, P.D. and Felsenfeld, G. (1985), A method for mapping intranuclear protein-DNA interactions and its application to a nuclease hupersensitive site. Proc.Natl.Acad.Sci. U.S.A. 82, 2296-2300.

Jensen, E.Ø., Paludan, K., Hyldig-Nielsen, J.J., Jørgesen, P. and Marcker, K.A. (1981), The structure of a chromosomal leghaemoglobin gene from soybean. Nature 291, 677-679.

Jones, M.G.K. (1985), Transformation of cereal crops by direct gene transfer. Nature 317, 579-580.

Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980), Novel bacteriophage λ cloning vector. Proc.Natl.Acad. Sci. U.S.A. 77, 5172-5176.

Keene, M.A. and Elgin, S.C.R. (1981), Micrococcal nuclease as a probe of DNA sequence organisation and chromatin structure. Cell 27, 57-64.

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

Kornberg, R.D. (1974), Chromatin structure: a repeating unit of histones and DNA. Science 184, 868-871.

Kunitz, M. (1950), Crystalline desoxribonuclease. I. Isolation and general properties. Spectrophotometric method for the measurement of desoxyribonuclease activity. J.Gen.Physiol. 33, 349-362.

Kunnath, L. and Locker, J. (1985), DNAase I sensitivity of the rat albumin and α-fetoprotein genes. Nucleic Acids Res. 13, 115-129.

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

Laskey, R.A. and Mills, A.D. (1977), Enhanced autoradiographic detection of ^{32}P and ^{125}I using intensifying screens and hypersensitized film. FEBS Lett. 82, 314-316.

Lawson, G.M., Knoll, B.J., March, C.J., Woo, S.L.C., Tsai, M.-J and O'Malley, B.W. (1982), Definition of 5′ and 3′ structural boundaries of the chromatin domain containing the ovalbumin multigene family. J.Biol.Chem. 257, 1501-1507.

Leber, B. and Hemleben, V. (1979), Structure of plant nuclear and ribosomal DNA containing chromatin. Nucleic Acids Res. 7, 1263-1281.

Lewin, B. (1980), Alternatives for splicing: recognising the ends of introns. Cell 22, 324-326.

Lilley, D.M.J. and Markham, A.F. (1983), Dynamics of cruciform extrusion in supercoiled DNA: use of a synthetic inverted repeat to study conformational populations. EMBO J. 2, 527-533.

Lilley, D.M.J. (1983), Eukaryotic genes - are they under torsional stress? Nature 305, 276-277.

Lohr, D. (1983), The chromatin structure of an actively expressed, single copy yeast gene. Nucleic Acids Res. 11, 6755-6773.

Lohr, D. and Hereford, L. (1979), Yeast chromatin is uniformly digested by DNAase I. Proc.Natl.Acad.Sci. U.S.A. 76, 4285-4288.

Lutter, L.C. (1977), Characterisation of DNAase I cleavage sites in the nucleosome. Cold Spring Harbor Symp. Quant. Biol. 42, 137-147.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984), The complete nucleotide sequence of a legumin gene from pea (*Pisum sativum* L.), Nucleic Acids Res. 12, 4493-4506.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H., Richards, D.M. and Boulter, D. (1985), The 5′-flanking regions of three pea legumin genes; comparison of the DNA sequences. Nucleic Acids Res. 13, 6733-6743.

Lycett, G.W., Delauney, A.J. and Croy, R.R.D., (1983b), Are plant genes different? FEBS Lett. 153, 43-46.

Lycett, G.W., Delauney, A.J., Gatehouse, J.A., Gilroy, J., Croy, R.R.D. and Boulter, D. (1983a), The vicilin gene family of pea (*Pisum sativum* L.): a complete cDNA coding sequence for preprovicilin. Nucleic Acids Res. 11, 2367-2380.

Macleod, D. and Bird, A. (1982), DNAase I sensitivity and methylation of active *versus* inactive rRNA genes in *Xenopus* species hybrids. Cell 29, 211-218.

Mandel, M. and Higa, A. (1970), Calcium-dependent bacteriophage DNA infection. J.Mol.Biol. 53, 159-162.

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

Mathis, D.J. and Chambon, P. (1981), The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature 290, 310-315.

Matta, N.K., Gatehouse, J.A. and Boulter, D. (1981), Molecular and subunit heterogeneity of legumin of *Pisum satinum* L. (garden pea) - a multi-dimensional gel electrophoretic study. J.Exp.Bot. 32, 1295-1307.

McDonnell, M.W., Simon, M.N. and Studier, F.W. (1977), Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J.Mol.Biol. 110, 119-146.

McGhee, J.D. and Engel, J.D. (1975), Subunit structure of chromatin is the same in plants and animals. Nature 254, 449-450.

McGhee, J.D., Rau, D.C., Charney, E. and Felsenfeld, G. (1980), Orientation of the nucleosome within the higher order structure of chromatin. Cell 22, 87-96.

McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. and Felsenfeld, G. (1981), A 200 base pair region at the 5′ end of the chicken adult β-globin gene is accessible to nuclease digestion. Cell 27, 45-55.

McGinnis, W., Shermoen, A.W. and Beckendorf, S.K. (1983b), A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. Cell 34, 75-84.

McGinnis, W., Shermoen, A.W., Heemskerk, J. and Beckendorf, S.K. (1983a), DNA sequence changes in an upstream DNAase I-hypersensitive region are correlated with reduced gene expression. Proc.Natl.Acad.Sci. U.S.A. 80, 1063-1067.

McIntosh, L.P., Grieger, I., Eckstein, F., Zarling, D.A., Van de Sande, J.H. and Jovin, T.M. (1983), Left-handed helical conformation of poly[d(A-m^5C)-d(G-T)]. Nature 304, 83-86.

Messing J. (1983a), New M13 vectors for cloning. Methods. Enzymol. 101, 20-78.

Messing, J., Geraghty, D., Heidecker, G., Hu, N., Kridl, J. and Rubenstein, I. (1983). Plant gene structure. In 'Genetic engineering of plants' (Kosuge, T., Meredith, C.P. and Hollaender, A., eds.), Academic Press, New York, pp.211-227.

Miflin, B.J., Burgess, S.R. and Shewry, P.R. (1981), The development of protein bodies in the storage tissues of seeds: subcellular separations of homogenates of barley, maize and wheat endosperms and of pea cotyledons. J.Exp.Bot. 32, 199-219.

Miflin, B. and Lea, P.J. (1984), The genetic manipulation of crop plants. Nature 308, 498-499.

Millerd, A. (1975), Biochemistry of legume seed proteins, Ann.Rev.Plant Physiol. 26, 53-72.

Millerd, A. and Spencer, D. (1974), Changes in RNA-synthesising activity and template activity in nuclei from cotyledons of developing pea seeds. Aust.J. Plant Physiol. 1, 331-341.

Millipore Corporation (1979), Worthington enzymes and related biochemicals. Catalogue, Millipore Corporation, Bedford, Mass., U.S.A.

Morelli, G., Nagy, F., Fraley, R.T., Rogers, S.G. and Chua, N.-H (1985), A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose 1,5-bisphosphate carboxylase small subinit of pea. Nature 315, 200-204.

Murai, N., Sutton, D.W., Murray, M.G., Slightom, J.L., Merlo, D.J., Reichert, N.A., Sengupta-Gopalan, C., Stock, C.A., Barker, R.F., Kemp, J.D. and Hall, T.C. (1983), Phaseolin gene from bean is expressed after transfer to sunflower *via* tumour-inducing plasmid vectors. Science 222, 476-482.

Murray, D.R. (1979), A storage role for albumins in pea cotyledons. Plant Cell Envt. 2, 221-226.

Murray, M.G. and Kennard, W.C. (1984), Altered chromatin conformation of the higher plant gene phaseolin. Biochemistry 23, 4225-4232.

Muskavitch, M.A.T. and Hogness, D.S. (1982), An expandable gene that encodes a *Drosophila* glue protein is not expressed in variants lacking remote upstream sequences. Cell 29, 1041-1051.

Negoro, T., Momma, T. and Fukazawa, C. (1985), A cDNA clone encoding a glycinin A_{1a} subunit precursor of soybean. Nucleic Acids Res. 13, 6719-6731.

Nickol, J.M. and Felsenfeld, G. (1983), DNA conformation at the 5´ end of the chicken adult ß-globin gene. Cell 35, 467-477.

Nielsen, N.C. (1984), The chemistry of legume storage proteins. Phil.Trans.Roy.Soc.Lond.B. 304, 287-296.

Noll, M. and Kornberg, R.D. (1977), Action of micrococcal nuclease on chromatin and the location of histone H1. J.Mol.Biol. 109, 393-404.

Nordheim, A,, Lafer, E.M., Peck, L.J., Wang, J.C., Stollar, B.D. and Rich, A. (1982)., Negatively supercoiled plasmids contain left-handed Z-DNA segments as detected by specific antibody binding. Cell 31, 309-318.

Nordheim, A., Pardue, M.L., Lafer, E.M., Möller, A., Stollar, B.D. and Rich, A. (1981), Antibodies to left-handed Z-DNA bind to interband regions of *Drosophila* polythene chromosomes. Nature 294, 417-422.

Nordheim, A. and Rich, A. (1983), The sequence $(dC.dA)_n.(dG.dT)_n$ forms left-handed Z-DNA in negatively supercoiled plasmids. Proc.Natl.Acad.Sci. U.S.A. 80, 1821-1825.

O'Malley, B.W., Towle, H.C. and Schwartz, R.J. (1977), Regulation of gene expression in eucaryotes. Ann.Rev. Genet. 11, 239-275.

Parker, C.S. and Topol, J. (1984b), A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an *hsp* 70 gene. Cell 37, 273-283.

Parker, C.S. and Topol, J. (1984a), A *Drosophila* RNA polymerase II transcription factor contains a promoter-region specific DNA-binding activity. Cell 36, 357-369.

Patthy, L. and Smith, E.L., Johnson, J. (1973), Histone III. V. The amino acid sequence of pea embryo histone III. J.Biol.Chem. 248, 6834-6840.

Pelham, H.R.B. L982), A regulatory upstream promoter element in the *Drosophila hsp* 70 heat-shock gene. Cell 30, 517-528.

Prentice, D.A. and Gurley, L.R. (1983), Nuclease digestibility of chromatin is affected by nuclei isolation procedures. Biochim.Biophys.Acta. 740, 134-144.

Prentice, D.A., Tobey, R.A. and Gurley, L.R. (1983), DNAase I and cellular factors that affect chromatin structure. Biochim.Biophys.Acta. 741, 288-296.

Pusztai, A. and Stewart, J.C. (1980), Molecular size, subunit structure and microheterogeneity of glycoprotein II from the seeds of kidney bean (*Phaseolus vulgaris* L.), Biochim.Biophys.Acta. 623, 418-428.

Reeves, R. (1984), Transcriptionally active chromatin. Biochim.Biophys.Acta. 782, 343-393.

Rich, A. (1982), Right-handed and left-handed DNA: conformational information in genetic material. Cold Spring Harbor Symp.Quant.Biol. 47, 1-12.

Rich, A., Nordheim, A. and Wang, A.H.-J. (1984), The chemistry and biology of left-handed Z-DNA. Ann.Rev.Biochem. 53, 791-846.

Salditt-Georgieff, M., Sheffery, M., Krauter, K., Darnell,J.E., Jr., Rifkind, R. and Marks, P.A. (1984), Induced transcription of the mouse β-globin transcription unit in erythroleukemia cells. Time-course of induction and of changes in chromatin structure. J.Mol.Biol. 172,437-450.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977), DNA sequencing with chain-terminating inhibitors. Proc.Natl. Acad.Sci. U.S.A. 74, 5463-5467.

Schell, J., Van Montagu, M., Willmitzer, L., Leemans, J., Deblaere, R., Joos, H., Inze, D., Wöstemeyer, A., Otten, L. and Zambryski, P. (1984), Transfer of foregin genes to plants and its use to study developmental processes. In 'Cell fusion: gene transfer and transformation' (Beers, R.F., Jr. and Basset, E.G., eds.), Raven Press, New York, pp.113-128.

Schuler, M.A., Schmidtt, E.S. and Beachy, R.N. (1982a), Closely related families of genes code for the α and α´ subunits of the soybean 7S storage protein complex. Nucleic Acids Res. 10, 8225-8244.

Schuler, M.A., Ladin, B.F., Pollaco, J.C., Freyer, G. and Beachy, R.N. (1982b), Structural sequences are conserved in the genes coding for the α, α´ and β-subunits of the soybean 7S seed storage protein. Nucleic Acids Res. 10, 8245-8261.

Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., Hall, T.C., and Kemp, J.D. (1985), Developmentally regulated expression of the bean β-phaseolin gene in tobacco seed. Proc.Natl.Acad.Sci. U.S.A. 82, 3320-3324.

Shah, D.M., Hightower, R.C. and Meagher, R.B. (1982), Complete nucleotide sequence of a soybean actin gene. Proc.Natl.Acad.Sci. U.S.A. 79, 1022-1026.

Shaw, C.H., Leemans, J., Shaw, C.H., Van Montagu, M. and Schell, J. (1983), A general method for the transfer of cloned genes to plant cells. Gene 23, 315-330.

Sheffery, M., Marks, P.A. and Rifkind, R.A. (1984), Gene expression in murine erythroleukemia cells. Transcriptional control and chromatin structure of the α_1-globin gene. J.Mol.Biol. 172, 417-436.

Shermoen, A.W. and Beckendorf, S.K. (1982), A complex of interacting DNAase I-hypersensitive sites near the *Drosophila* glue protein gene, *Sgs* 4. Cell 29, 601-607.

Shewry, P.R., Miflin, B.J., Forde, B.G. and Bright, S.W.J. (1981), Conventional and novel approaches to the improvement of the nutritional quality of cereal and legume seeds. Sci.Prog.(Oxford), 67, 575-600.

Shirsat, A.H. (1984), A gene for legumin: a major storage protein of *Pisum sativum* L. Ph.D. Thesis, University of Durham.

Simpson, J., Timko, M.P., Cashmore, A.R., Schell, J., Van Montagu, M. and Herrera-Estrella, L. (1985), Light-inducible and tissue-specific expression of a chimaeric gene under control of the 5´-flanking sequence of a pea chlorophyll *a/b*-binding protein gene. EMBO J. 4, 2723-2729.

Sinden, R.R., Carlson, J.O. and Pettijohn, D.E. (1980), Torsional tension in the DNA double helix measured with trimethylpsoralen in living *E.coli* cells: analogous measurements in insect and human cells. Cell 21, 773-783.

Singleton, C.K., Kilpatrick, M.W. and Wells, R.D. (1984), S1 nuclease recognises DNA conformational junctions between left-handed helical $(dT\text{-}dG)_n.(dC\text{-}dA)_n$ and contiguous right-handed sequences. J. Biol.Chem. 259, 1963-1967.

Singleton, C.K., Klysik, J., Stirdivant, S.M. and Wells, R.D. (1982), Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. Nature 299, 312-316.

Slightom, J.L., Drong, R.F., Klassy, R.C. and Hoffman, L.M. (1985), Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucleic Acids Res. 13, 6483-6498.

Slightom, J.L., Sun, S.M. and Hall, T.C. (1983), Complete nucleotide sequence of a French bean storage protein gene: phaseolin. Proc.Natl.Acad.Sci.,U.S.A. 80, 1897-1901.

Small, D. and Vogelstein, B. (1985), The anatomy of supercoiled loops in the *Drosophila* 7F locus. Nucleic Acids Res. 13. 7703-7713.

Smith, D.L. (1973), Nucleic acid, protein, and starch synthesis in developing cotyledons of *Pisum arvense* L. Ann.Bot. 37, 795-804.

Sorenson, J.C. (1984), The structure and expression of nuclear genes in higher plants. Adv.Genetic. 22, 109-144.

Southern, E.M. (1975), Detection of specific sequences among DNA fragments separated by gel electrophoresis. J.Mol. Biol. 98, 503-517.

Spencer, D. (1984), The physiological role of storage proteins in seeds. Phil.Trans.Roy.Soc.Lond. B 304, 275-285.

Spencer, D. and Higgins, T.J.V. (1980), The biosynthesis of legumin in maturing pea seeds. Biochem.Int. 1, 502-509.

Spiker, S. and Isenberg, I. (1977), Evolutionary conservation of histone-histone binding sites: evidence from interkingdom complex formation. Cold Spring Harbor Symp. Quant.Biol. 42, 157-163.

Spiker, S. (1984), Chromatin structure and gene regulation in higher plants. Adv.Genetic. 22, 145-208.

Spiker, S. (1985), Plant chromatin structure. Ann.Rev.Plant Physiol. 36, 235-253.

Spiker, S., Murray, M.G. and Thompson, W.F. (1983), DNAase I sensitivity of transcriptionally active genes in intact nuclei and isolated chromatin of plants. Proc.Natl. Acad.Sci. U.S.A., 80, 815-819.

Spinelli, G., Albanese, I., Anello, L., Ciarcio, M. and Di Liegro, I. (1982), Chromatin structure of histone genes in sea urchin sperms and embryos. Nucleic Acids Res. 10, 7977-7991.

Stalder, J., Groudine, M., Dodgson, J.B., Engel, J.D. and Weintraub, H. (1980a), Hb switching in chickens. Cell 19, 973-980.

Stalder, J., Larsen, A., Engel, J.D., Dolan, M. Groudine, M. and Weintraub, H. (1980b), Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell 20, 451-460.

Stalder, J., Seebeck, T. and Braun, R. (1978), Degradation of the ribosomal genes by DNAase I in *Physarum polycephalum*, Eur.J.Biochem. 90, 391-395.

Stalder, J., Seebeck, T. and Braun, R. (1979), Accessibility of the ribosomal genes to micrococcal nuclease in *Physarum polycephalum*, Biochim.Biophys.Acta 561, 452-463.

Sun, S.M., Slightom, J.L. and Hall, T.C. (1981), Intervening sequences in a plant gene - comparison of the partial sequence of cDNA and genomic DNA of French bean phaseolin. Nature 289, 37-41.

Thoma, F., Koller, Th. and Klug, A.L. (1979), Involvement of histone H1 in the organisation of the nucleosome and of the salt-dependent superstructures of chromatin. J.Cell Biol. 83, 403-427.

Thomas, P.S. and Farquhar, M.N. (1978), Specific measurement of DNA in nuclei and nucleic acids using diaminobenzoic acid. Anal.Biochem. 89, 35-44.

Tobin, E.M. and Silverthorne, J. (1985), Light regulation of gene expression in higher plants. Ann.Rev.Plant Physiol. 36, 569-593.

Vieira, J. and Messing, J. (1982), The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268.

Udvardy, A., Maine, E. and Schedl, P. (1985), The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. J.Mol.Biol. 185, 341-358.

Vogelstein, B., Pardoll, D.M. and Coffey, D.S. (1980), Supercoiled loops and eucaryotic DNA replication. Cell 22, 79-85.

Wang, J.C. (1982), The path of DNA in the nucleosome. Cell 29, 724-726.

Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., Van Boom, J.H., Van Der Marel, G. and Rich, A. (1979), Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature 282, 680-686.

Waterhouse, R.N. (1985), Ph.D. Thesis, University of Durham.

Weintraub, H. (1983), A dominant role for DNA secondary structure in forming hypersensitive structures in chromatin. Cell 32, 1191-1203.

Weintraub, H. (1984), Histone-H1-dependent chromatin superstructures and the suppression of gene activity. Cell 38, 17-27.

Weintraub, H. (1985), Assembly and propagation of repressed and derepressed chromatin states. Cell 42, 705-711.

Weintraub, H., Beug, H., Groudine, M. and Graf, T. (1985), Temperature-sensitive changes in the structure of globin chromatin in lines of red cell precursors transformed by ts-AEV. Cell 28, 931-940.

Weintraub, H. and Groudine, M. (1976), Chromosomal subunits in active genes have an altered conformation. Science 193, 848-856.

Weisbrod, S. (1982), Active chromatin. Nature 297, 289-295.

Weisbrod, S., Groudine, M. and Weintraub, H. (1980), Interaction of HMG 14 and 17 with actively transcribed genes. Cell 19. 289-301.

Weisbrod, S. and Weintraub, H. (1981), Isolation of actively transcribed nucleosomes using immobilised HMG 14 and 17, and an analysis of α-globin chromatin. Cell 23, 391-400.

Weiss, B., Jacquemin-Sablon, A., Live, T.R., Fareed, G.C. and Richardson, C.C. (1968), Enzymatic breakage and joining of deoxyribonucleic acid. VI. Further purification and properties of polynucleotide ligase from *Escherichia coli* infected with bacteriophage T4. J.Biol.Chem.243, 4543-4555.

Willmitzer, L. and Wagner, K.G. (1981), The isolation of nuclei from tissue-cultured plant cells. Exp.Cell Res. 135, 69-77.

Wolf, S.F., and Migeon, B.R. (1985), Clusters of CpG dinucleotides implicated by nuclease hypersensitivity as control elements in housekeeping genes. Nature 314, 467-469.

Wood, W.B. (1966), Host specificity of DNA produced by *Escherichia coli*: bacterial nutations affecting the restriction and modification of DNA. J.Mol.Biol. 16. 118-133.

Worcel, A. (1977), Molecular architecture of the chromatin fibre. Cold Spring Harbor Symp.Quant.Biol. 42. 313-324.

Worcel, A., Gargiulo, G., Jessee, B., Udvardy, A., Louis, C. and Schedl, P. (1983), Chromatin fine-structure of the histone gene complex. Nucleic Acids Res. 11, 421-439.

Wright, D.J. and Boulter, D. (1974), Purification and subunit structure of legumin of *Vicia faba* L. (Broad bean), Biochem.J. 141, 413-418.

Wu, C. (1980), The 5′ ends of *Drosophila* heat-shock genes in chromatin are hypersensitive to DNAase I. Nature 286, 854-860.

Wu, C. (1984a), Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. Nature 309, 229-234.

Wu, C. (1984b), Activating protein factor binds *in vitro* to upstream control sequences in heat shock gene chromatin. Nature 311, 81-84.

Wu, C. (1985), An exonuclease protection assay reveals heat-shock element and TATA box DNA-binding proteins in crude nuclear extracts. Nature 317, 84-87.

Wu, C., Bingham, P.M., Livak, K.J., Holmgren, R. and Elgin, S.C.R. (1979a), The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. Cell 16, 797-806.

Wu, C., Wong, Y.-C. and Elgin, S.C.R. (1979b), The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. Cell 16, 807-814.

Yang, R. C.-A., Lis, J. and Wu, R. (1979), Elution of DNA from agarose gels after electrophoresis. Methods Enzymol. 68, 176-182.

Zaret, K.S. and Yamamoto, K.R. (1984), Reversible and persistent changes in chromatin structure accompany activation of a glucocorticoid-dependent enhancer element. Cell 38, 29-38.

Zhao, W.-M., Gatehouse, J.A. and Boulter, D. (1983), The purification and partial amino acid sequence of a polypeptide from the glutelin fraction of rice grains; homology to pea legumin. FEBS Lett. 162, 96-102.

Zimmerman, S.B. (1982), The three-dimensional structure of dna. Ann.Rev.Biochem. 51, 395-427.