A gene for legumin: a major storage protein of pisum sativum L.

Shirsat, Anil Harishchandra

To my parents

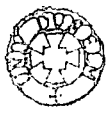
A Gene For Legumin : a Major Storage Protein of *Pisum sativum* L.

by

Anil Harishchandra Shirsat, B.Sc. (E. Anglia)

A thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy in the University of Durham.

March 1984



ABSTRACT

A gene coding for legumin, a major storage protein of pea, was isolated from a library of pea DNA, and partially sequenced. The library was constructed by partially digesting pea leaf DNA with Sau 3A, and ligating 15 - 20 Kb size fractions into the Bam H1 cleaved arms of the vector λ L47. Ribosomal clones were isolated from the library at a frequency which indicated that the sequence representation was near complete. Screening of the library for sequences encoding legumin was performed by using legumin complementary DNA.

The coding sequence of this gene was localised using restriction enzyme and hybridisation analysis, to a 2.7 Kb fragment within the 15 Kb genomic insert in the phage vector. This 2.7 Kb fragment was subcloned into the plasmid vector pUC 8, and analysed by restriction enzyme mapping and sequencing. This showed the presence of three small intervening sequences within the legumin coding region. Sequence data obtained for two of the intervening sequences indicated that they were rich in adenine and thymine nucleotides, and that the 5′ terminus of both introns began with the dinucleotide GT.

Hybridisation experiments using this gene to probe for homologous sequences in pea leaf DNA detected 7 hybridising fragments, suggesting the existence of more genes than were originally thought to comprise the legumin gene family.

ACKNOWLEDGEMENTS

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

I am extremely grateful to Dr. R.R.D. Croy for numerous helpful and stimulating discussions, and for his constant guidance throughout this course of study. Other members of the Botany Department, University of Durham, especially Dr. G.W. Lycett, Dr. C. Shaw, and Mr. A.J. Delauney, are to be thanked for introducing me to various techniques, and for critically reading parts of this manuscript. Thanks are also due to Dr. R.B. Flavell of the Plant Breeding Institute, Cambridge, in whose laboratory I learned to isolate pea DNA.

I should also like to thank Dr. J. Gatehouse and Dr. I.M. Evans for their help.

Finally, I am grateful to Mr. D. Bown, Mrs. P. Brown, and Mrs. D.M. Richards for their considerable laboratory skills, patience and good humour.

The Science and Engineering Research Council provided me with financial support to carry out this study.

Memorandum

Parts of the work presented in this thesis also appear in the following publications :

Croy, R.R.D., Shirsat, A., Boulter, D. (1984), Cloning and characterisation of genes for legumin from Pisum sativum L. Nucl. Acids. Res. (submitted)

Lycett, G.W., Croy, R.R.D., Shirsat, A., Boulter, D. (1984b), The complete nucleotide sequence of a legumin gene from pea (Pisum sativum L.). Nucl. Acids. Res. (submitted).

Lycett, G.W., Croy, R.R.D., Delauney, A.J., Shirsat, A., Boulter, D. (1984c) Molecular analysis of the gene families coding for the storage proteins of Pisum sativum L. Heredity (submitted).

CONTENTS

RESULTS

FIGURES

Figure No.

Figure No.

TABLES

ABBREVIATIONS

All abbreviations are as stated in the Biochemical Journal " Instructions to authors " (Biochemical Society 1975), apart from those detailed below.

Py - pyrimidine

Pu - purine

cDNA - complementary or copy DNA

dd NTP - dideoxynucleotide

N - terminal: Amino terminus of a protein

C - terminal: Carboxy terminal of a protein

bp - base pairs

kb - kilobasepairs

c.p.m. - counts per minute

SDS - sodium dodecyl sulphate

p.a.g.e. - polyacrylamide gel electrophoresis

5′ - 5′ terminal phosphate in a DNA or RNA molecule

3′ - 3′ terminal hydroxyl in a DNA or RNA molecule

INTRODUCTION

NUTRITIONAL VALUE OF SEED STORAGE PROTEINS

It has been estimated that the direct consumption of cereal and legume seeds accounts for 70 % of mankind's food (FAO 1970). The remaining 30 % comes from animals, some of which are fed on seed meals. The proteins present in seeds are therefore a major source of protein in the diet of both man and his animals, and as such, merit investigation from practical considerations alone. Cereals form the major source of plant protein followed by legumes. However, as expressed as a function of dry weight, the percentage of protein in legume seeds (20 - 50 %), is greater than that in cereals (8 - 15 % - Danielsson 1949), and legume crops are therefore especially important as a protein source in developing countries, where protein malnutrition is often very common (Shewry *et al* 1981). Legumes also play a key role in crop rotation as a break crop, and in improving soil fertility by their ability to fix atmospheric nitrogen.

Protein is required in the diet as a source of essential amino acids. Of the 20 amino acids commonly present in protein, nine are essential for mammalian nutrition. These are leucine, isoleucine, lysine, threonine, valine, methionine, tryptophan, tyrosine and phenylalanine. Despite the fact that legume seeds make important contributions to protein nutrition, they are limited by their content of the sulphur amino acids, cysteine (which spares methionine), and methionine (Croy *et al* 1980a, Casey and Short 1981). This important nutritional limitation offers scope for their improvement, either by conventional breeding techniques, or ultimately, by the recently developed techniques of genetic engineering. However, some of the many problems which have to be overcome before direct plant genetic

manipulation becomes a reality have been pointed out by Cocking (1981).

Storage proteins of seeds are those proteins used as a nitrogen source, to establish the seedling upon germination (Basha and Beevers 1975, Thompson et al 1978), and usually constitute 50 - 90 % of the seed protein (Shewry et al 1981). In keeping with their role as a nitrogen store, they normally contain a high proportion of the amides glutamine and/or arginine (Derbyshire et al 1976). According to the system of Osborne (1924), who classified proteins on the basis of their solubility in different solvents, legume seed storage proteins fall into the globulin class, i.e. those soluble in aqueous salt solutions. The legume globulins are of two main types, legumin and vicilin, and wide variation in the relative amounts of these types occurs within the Leguminosae. For example, Phaseolus vulgaris contains very little legumin, (Derbyshire et al 1976), while Vicia faba contains an excess of legumin over vicilin (Gatehouse et al 1980a), and Pisum sativum contains about equal amounts of both types (Schroeder 1983), but varies with genetic line. The globulin storage proteins are widely distributed in other legumes e.g. occurring in Glycine max, Vigna unguiculata, Vigna radiata, and Lupinus spp (Derbyshire et al 1976). The overall similarity of the legumin and vicilin proteins between closely related species of the Leguminosae has been demonstrated by Croy et al 1979. Besides legumin and vicilin, other proteins such as enzymes and lectins are present in lesser amounts within seeds. Lectins are proteins which have the ability of agglutinating human or animal red blood cells, and are synthesised during seed maturation along with the other seed storage proteins (Bollini and Chrispeels 1978).

Conventional plant breeding techniques involve the crossing of several variants followed by selection for the improved genotype. This approach can be extended by using mutagenesis to artificially extend the range of variability. An example of this is the series of mutant high lysine genes which have been reported in cereals (reviewed by Nelson 1980). However, although the mutants had increased lysine content, they were identified by mutant endosperm phenotypes caused by changes in carbohydrate metabolism, thus their designation as opaque, floury, sugary, shrunken and brittle (Nelson 1980). This disruption of carbohydrate metabolism resulted in decreased seed weight and yield. Although these problems appear in part to have been overcome (Shewry et al 1981), conventional breeding programmes involving the screening of hundreds of cultivars are tedious, and frequently result in the production of plants which might have the desired trait, but which also possess other undesirable characteristics such as decreased yield.

Genetic engineering methodologies, on the other hand, have the potential ability to alter the plant genotype in a defined manner. An essential prerequisite of this strategy is the isolation of protein coding genes from the plant genome, and their complete characterisation with respect to coding and regulatory sequences. However, even if cloned genes became available, their utilisation would depend upon understanding the biology of the system in which they act. Hence, the next section will give a brief outline of storage protein synthesis and pea seed development.

PEA SEED DEVELOPMENT AND STORAGE PROTEIN SYNTHESIS.

The stages involved in the seed development of pea have been reviewed by Boulter (1981), and will only be briefly summarised here. In higher plants, the seed is the organ of dispersal, and is normally produced by sexual reproduction. The end product of seed development is a mature seed, which is capable of surviving long periods under adverse conditions, and still being able to support the growth of a seedling upon germination. Usually, three phases can be identified with the production of a mature seed. During the first phase, 1 - 6 days after flowering*, cell division occurs and intermediates are built up, but the rate of storage protein synthesis is very low (Boulter 1981). By the beginning of the second phase, days 7 - 22, the structures of the seed, the cotyledons and embryo, have appeared, but DNA replication continues, leading to endoreduplication of the genome and chromosome values of up to 64C (Millerd and Spencer 1974, Spencer and Higgins 1979). Domoney *et al* (1980) have shown that legumin synthesis starts from the initial stages of embryo development, and that it can be detected in embryos weighing as little as 2×10^{-3} g. It is probable that vicilin is also similarly synthesised (Boulter 1981). From this initial, subliminal rate of synthesis, a sudden and massive increase in the amount of storage protein occurs, until by day 20, it comprises about 60 % of the total protein of the mature seed. This is accompanied both by an increase in the amount of rough endoplasmic reticulum (Boulter *et al* 1972, Chrispeels *et al* 1979), and an increase in the ratio of polysomes to monosomes in cotyledon cells (Beevers and Poulson 1972). At this stage, protein bodies which are 2 µm diameter single membrane bound organelles in which storage protein is deposited, start to appear (Baumgartner *et al* 1980). It is thought

* in *Pisum sativum* var Feltham First grown at 20°C and 12h days.

(Baumgartner et al 1980, Briarty et al 1969), that storage proteins are synthesised onto the rough endoplasmic reticulum (RER) and transported to the protein bodies via the smooth RER, in a manner analogous to that of animal secretory proteins (Caro and Palade 1964). The onset of the increased rates of synthesis and deposition of vicilin and legumin are sequential (Thompson et al 1979, Boulter 1981). Recently, it has been shown that vicilin is synthesised at 8 days, both vicilin and legumin at 13 days, and only legumin at 19 days after flowering (Evans pers. comm.). The second phase of seed development, involving the rapid synthesis and accumulation of storage protein, is followed by the third phase ; developmental arrest and dessication af the seed. Storage protein synthesis is not continued into this stage, and it has been shown (Gordon and Payne 1976), that storage protein mRNA does not appear to be present amongst the few mRNAs stored in the dry seed.

Storage proteins, then, are laid down at a specific stage in seed development in large amounts, to be used as a nitrogen supply on germination. Since the synthesis of these storage proteins is specific to the seed (Thompson et al 1978, Dure 1975 Gatehouse et al 1982), the developing seed provides a good model system for studying developmentally regulated gene expression. Analogous animal systems are, amongst others, the egg white protein system in chicken oviducts (Palmiter 1972), the fibroin protein system in Bombyx mori (Suzuki et al 1972), and the haemoglobin system in erythroid cells (Ross et al 1974).

Legumin, and its mode of synthesis, has been investigated by studying the mRNAs responsible for its production. Morton et al

(1983) have shown that the sequence complexity of pea seed mRNA decreases dramatically during the period 15 to 21 days after flowering (when the synthesis of storage protein is at its greatest), from an initial complexity of 20,000 sequences at day 9 to approximately 200 sequences at day 19. Corresponding with this decrease in complexity, a limited number of mRNA sequences become superabundant. Studies by Croy et al (1980a) on in vivo ^{35}S labelled storage proteins showed that immunoprecipitable legumin was not detected until 11 to 13 days after flower opening. This coincidence between the superabundance of certain mRNA classes and the accumulation of legumin makes it likely that the increased rate of legumin synthesis is a direct result of increasing amounts of legumin mRNA. Confirmation of this was obtained by Croy et al (1980) who showed, using in vitro translation systems driven by mRNA extracted from 14 to 21 day cotyledons, that storage protein precursors were prominent components of the polypeptides being synthesised. A direct demonstration of the increase in storage protein mRNA species during development of the seed was provided by Gatehouse et al 1981. They showed, using the technique of 'Northern' hybridisation (Thomas 1980) that the levels of legumin and vicilin mRNAs increased and decreased in agreement with estimated rates of synthesis of the respective polypeptides. The tissue specific expression of legumin was also demonstrated, as legumin poly A+ mRNA was found to be present in leaves at less than one - thousandth the level in cotyledons. Casey (1979a), and Croy et al (1979) determined the molecular weight of pea legumin to be 395,000 and 390,000 respectively, by using gel filtration techniques. Studies on the subunit composition of legumins from pea have utilised polyacrylamide gel electrophoresis under reducing and non reducing conditions. Electrophoresis using non reducing denaturants like SDS, resulted in

the dissociation of native legumin into subunits of approximate Mr 60,000. On addition of reducing agents such as 2 - mercaptoethanol, the Mr 60,000 subunits dissociated into smaller polypeptides of Mr 40,000 and Mr 20,000 (Croy 1979). Similiar studies on the legumin from Vicia faba led Wright and Boulter (1974) to propose a model for the legumin molecule which was shown to be correct for both the Vicia faba and Pisum sativum legumins (Gatehouse et al 1980b). This model proposed that legumin consisted of a hexamer of Mr 60,000 subunits. Each subunit consisted of one large Mr 40,000 and one small Mr 20,000 subunit, held together by one or more disulphide bonds. In pea legumin, the larger subunit was shown to have a pI value of between 4.8 and 6.1 and was called the 'acidic' subunit, while the smaller Mr 20,000 subunit had a pI value of between 6.2 and 8.0 and was referred to as the 'basic' subunit (Krishna et al 1979, Matta et al 1981). The acidic subunits were richer in aspartic acid, glutamine, and arginine, while the basic subunits were richer in alanine, valine and leucine.

These acidic and basic subunits, could not, however, be regarded as single polypeptide species. Each of the subunits could be separated into a number of component polypeptides by a combination of SDS polyacrylamide gel electrophoresis and isoelectric focussing. Heterogeneity with respect to molecular weight and charge was shown by Croy et al 1979 and Thompson et al 1978, while heterogeneity with regard to the isoelectric points of the protein was shown by Krishna et al 1979 and Casey 1979. Matta et al 1981 extended these observations by showing that while Mr 54,000 disulphide bonded subunit pairs constituted a major part of total legumin, subunit pairs with Mr varying from 35,000 to 58,000 were also present. These associated in various ways to give at least three distinct molecular forms of

legumin. The main type (Mr 53,000 to 54,000) corresponded to the legumin subunit pair of the Wright and Boulter (1974) model, whilst the other two types were present in smaller amounts and were designated as big (Mr 55,000 to 58,000), and small (Mr 35,000) legumins. This heterogeneity was postulated to have arisen in three possible ways 1) single substitutions in the amino acid sequence of polypeptides, as has been reported for the primary structure of the basic pea legumin subunit by Casey et al 1981b), 2) extensive differences in sequences among polypeptides including differences in the total number of amino acids, and 3), formation of legumin molecules from different combinations of subunits. The existence of this heterogeneity, does not, however, challenge the original hexameric model for legumin structure proposed by Wright and Boulter (1974); i.e. the association of acidic and basic subunits into polypeptide pairs. A structural model for vicilin, envisaging the molecule as a trimer of Mr 50,000 subunits, has been put forward by Gatehouse et al 1981. Considerable heterogeneity in the Mr of the 50,000 subunit is also seen, a situation analogous to that in pea legumin.

An interesting insight into the synthesis and modification of legumin was gained by Croy et al (1980a), who used isolated poly A+ mRNA and in vitro protein synthesising systems. They showed that the in vitro synthesised Mr 60,000 legumin would not dissociate into the acidic and basic subunits under reducing conditions, while the native legumin, isolated from pea seeds, would easily do so. They concluded that legumin was synthesised as a Mr 60,000 polypeptide precursor, which was subsequently cleaved as a result of proteolysis in vivo into the observed Mr 40,000 and Mr 20,000 species. The

failure to observe this proteolysis using membrane bound polysomes *in vitro*, suggested that proteolysis occured after transport of the polypeptide *in vivo*, and thus probably after assembly of the subunit pairs into legumin molecules. Work by Spencer and Higgins (1980) using *in vitro* translation systems and high resolution polyacrylamide gels, has shown that pea legumin precursor polypeptides were initially assembled into a trimeric form of Mr 180,000 before complete assembly into the hexamer.

The inheritance of the storage protein genes in *Pisum sativum* L. was investigated by Matta and Gatehouse 1981, using SDS p.a.g.e. to study the proteins produced in crosses of various pea lines. They concluded that the legumin genes coding for the major Mr 53,000 to 54,000 subunits in pea were localised on the short arm of chromosome 7. As at least five different band patterns were observed for the major acidic subunits of legumin (Croy *et al* 1979, Casey 1979), it was concluded that multiple alleles were present at this locus. Data concerning the inheritance of the minor legumin genes suggested that while some were present at the same locus as the major legumin gene, others were not.

The elucidation of the structure of the legumin molecule, and the finding that high levels of storage protein mRNA were present at a specific stage in seed development, provided the impetus for the beginning of studies on the genes involved in storage protein synthesis. The prime requirement for isolating a gene from an eucaryotic genome is the availability of a suitable probe , complementary to the gene in question. At first sight, mRNAs isolated from pea seeds at the 14 to 21 day stage of development, appear to be

ideal probes for the isolation of genomic sequences. However, mRNA preparations are heterogenous, and therefore cannot be used as specific probes, for fear that undesirable genomic clones, corresponding to the contaminating mRNA may be selected. For this reason, it is advisable to synthesize cDNA copies of the mRNAs, and to throughly characterise these copies before attempting to use them as probes. In this way, one can be sure of having a pure probe, complementary to the gene under investigation.

Evans et al (1979), described the isolation of poly A+ mRNA from pea cotyledons harvested 14 days after flowering, when production of storage protein was at its greatest. Many eucaryotic mRNAs, including those of vicilin and legumin, have a terminal sequence of adenosine residues at their 3′ end, which is added post - transcriptionally (Brawerman 1974). This enables separation of the poly A+ fraction of polysomal mRNA on an oligo d - T cellulose column (Aviv and Leder 1972). Evans et al (1979) fractionated poly A+ RNA (which constitutes 1 % of total RNA), on sucrose gradients, and identified, using in - vitro translation techniques (Gordon and Payne 1976), fractions which showed the greatest synthesis of storage protein subunits. These partially purified mRNA fractions, enriched in the mRNAs coding for storage proteins, were used to synthesize complementary DNA′s (Evans et al 1980). The cloning of these cDNA′s into the plasmid vector pBR322 (Bolivar et al 1977) was reported by Croy et al (1982). Transformants obtained from this cloning experiment were selected on the basis of their hybridisation to poly A+ mRNA, and to cDNA. Individual clones were identified using the technique of hybrid selected translation (Smith et al 1979). Briefly, poly A+ mRNA was hybridised to the denatured clones, and specifically

bound mRNAs eluted and translated in a rabbit reticulolysate *in vitro* protein synthesising system. 2 clones coding for the Mr 60,000 precursor were selected. As a final means of identification, these cDNAs were completely sequenced. Agreement between the predicted amino acid composotion and the tryptic peptide and N terminal sequence data, unequivocably identified the clones as legumin cDNAs (Croy *et al* 1982). These clones provided further evidence for the synthesis of legumin as a Mr 60,000 precursor. The coding sequence of one of the clones was shown to extend unbroken upstream beyond the N terminus of the basic subunit, approaching or extending into the C terminus of the acidic subunit (Croy *et al* 1982). No termination or initiation codons were found in frame with the coding sequence in the pre - N terminal sequence, confirming the precursor theory. The synthesis of pea vicilin in a precursor form which was also proteolytically cleaved to give the observed subunits (Gatehouse *et al* 1981) was confirmed by an examination of the sequence of cloned vicilin cDNA's (Lycett *et al* 1983a). Sequences of the cDNAs coding for pea lectin (Higgins *et al* 1983a) showed that the characteristics of the synthesis of pea legumin and vicilin in a precursor form were also true for this protein. The large Mr 25,000 lectin precursor was found to give α and β subunits of Mr 17,000 and 6,000 upon post translational processing (Higgins *et al* 1983b). Similiar precursor - product relationships have been elucidated for the legumin like proteins of *Vicia faba* (Weber *et al* 1981), soybean (Sengupta *et al* 1981), a globulin protein of oat (Brinegar and Petersen 1982), and rice glutelin (Yamgata *et al* 1982). Analysis of cDNA clones coding for napin, the rape seed storage protein (Crouch *et al* 1983) showed that both subunits of napin were derived from a precursor polypeptide by proteolysis. It thus seems likely that the precursor - product relationship is a common one, at

least in relation to seed proteins.

The cloning, isolation, and successful identification of the pea legumin cDNAs provided a pure probe to isolate the homologous sequences from genomic pea DNA.

the next section will give a brief description of sequence organisation in the eucaryotic genome, with specific reference to pea.

DNA SEQUENCE ORGANISATION IN THE EUCARYOTIC GENOME WITH SPECIFIC REFERENCE TO PEA.

Eucaryotic genomes differ from procaryotic ones in two major aspects: 1) the amount of DNA per nucleus, and 2) the organisation of this DNA. The *E. coli* genome contains 4.6×10^6 bp (Klotz and Zimm 1972), while the haploid human genome contains 2.8×10^9 bp (Lawn *et al* 1978), which is an approximate increase of 600 fold. Plant genomes generally have an even higher DNA content, e.g. 4.8×10^9 bp, 5.6×10^9 bp and 8.6×10^9 bp for the genomes of pea, wheat and rye respectively (Thompson and Murray 1981).

Attempts to investigate the sequence of eucaryotic genomes have centered on DNA reassociation experiments. The genome of any organism can be characterised as to its DNA content, length and frequency of DNA repetition classes and sequence complexity, by following the reassociation of complementary DNA sequences over a range of Cot values. The theory and practice of DNA reassociation experiments has been reviewed by Britten *et al* (1974), and will not be mentioned here. The renaturation kinetics of procaryotic DNAs reveal that they are

composed of DNA sequences which are present only once per genome, i.e. very little sequence homology exists between different parts of the genome. Consequently, the number of genes in E. coli can be calculated with reasonable accuracy. Assuming an average gene size of 1500 bp (sufficent for a protein of Mr 55, 000), E. coli can be assumed to have 3000 genes. A substantial fraction of the E. coli genome has been genetically mapped, and estimates of the same order of magnitude as the above figure have been obtained (Bachmann and Low 1980). A similiar calculation for the numbers of genes in a human and a pea genome gives figures of 1.8×10^6 and 3×10^6 genes respectively. deVries et al (1983) have estimated, using reassociation studies, that approximately 20,000 genes are transcribed in pea. This agrees well with values reported for other higher plants - 35,000 in parsley root and callus, (Kiper et al 1981), 26,000 in cotton seedlings (Galau and Dure 1981), 31,000 in barley shoots (Heinze et al 1980) and 27,000 in tobacco leaves (Kamalay and Goldberg 1981). Assuming that the polysomal RNA's used in these studies were mRNA's, these values should represent the numbers of expressed genes. The discrepancy between the calculated value for the number of genes and the estimate obtained experimentally is so great, that the inescapable conclusion is that the vast majority of DNA in higher eucaryotes does not code for mRNA.

Reassociation experiments on eucaryotic genomes show that a variable and often large component reassociates relatively rapidly, indicative of a high degree of reiteration (Britten and Kohne 1968). These sequences may be genes for histones or ribosomal RNA which are reiterated several times in the genome, (haploid pea has 3900 rRNA genes), and some may be sequences which do not code for proteins.

The components of the genome which reassociate

slowly, consist of sequences which behave as if present only once per haploid genome and are termed single copy sequences. Sequences which reassociate at rates faster than those observed for single copy DNA are termed repetitive DNA sequences. Repetitive DNA sequences probably consist of two types of classes - those in which the repeating units are very similiar and organised in tandem arrays, and those which consist of dissimiliar repeating units organised in a complex manner (Flavell 1980).

Buoyant density centrifugation of DNA in caesium salt gradients has identified a class of nuclear DNA sequences, termed satellite DNAs in a wide variety of plants and animals (SKinner 1977). Satellite DNAs consist of very similiar repeating units tandemly arranged, of a length ranging from a few base pairs, to thousands of base pairs (Flavell 1980). Satellite DNA sequences can be separated away from the rest of the genome because of their differential base composition, or their ability to differently bind ligands e.g. Ag^{+} (Ingle et al 1973). Amongst plant species, Ingle et al (1973) found that cucumber had the largest amount of satellite DNA - it comprised 44 % of the nuclear genome.

Ribosomal cistrons, by virtue of the fact that they consist of thousands of copies arranged in tandem clusters (Flavell 1980), might be expected to form a dense satellite, if present in sufficent quantity. However, Ingle et al (1975) showed that rDNA in plants comprised only a small fraction of the satellite. The tomato satellite DNA has so far been the only one found in which ribosomal RNA cistrons constitute a significant fraction (50 %) of the satellite, - this comprises about 3 % of the genome (Chilton 1975).

Wall and Bryant (1981) have demonstrated the presense of satellite DNA's in the pea genome. Three satellites were isolated, which in total comprised 33 % of the genome. Hybridisation of ^{3}H labelled pea rRNA to these satellites showed that none of them contained ribosomal RNA cistrons. Tandem arrays of short repeated sequences such as are found in satellite DNAs, have been shown, using _in situ_ hybridisation studies (Bedbrook _et al_ 1980), to be distributed in a non random manner throughout the genome; localisation frequently appears to be at the ends of chromosomes.

Several animal and plant genomes have been investigated using reassociation techniques. Most animal genomes (Davidson _et al_ 1975, Crain _et al_ 1976), show what is termed a short period interspersion pattern in which a greater than 60 % fraction consists of repeated sequences, 200 - 400 bp long, interspersed with single copy DNA at intervals of less than 1000 to 4000 bp. Both short and long repetitive sequences have been found in plants, distributions between about 200 and several thousand bp have been found for repeats in DNA from wheat (Flavell and Smith 1976), cotton (Walbot and Dure 1976), and parsley (Kiper and Herzfeld 1978). Data presented by Murray _et al_ (1978) showed that the pea genome was characterised by a very short period interspersion pattern, in which single copy sequences 300 bp long were extensively interspersed with repetitive sequence elements also about 300 to 400 bp long. 70 % of the pea genome was found to consist of repeated sequences of the moderate complexity, middle repetitive type, in which 400 to 10,000 families had repetition frequencies of between 300 to 10,000 copies per haploid genome. The highly repeated fraction comprised 30 - 40 % of the genome, and corresponded well with the kinetics obtained for the satellite DNA of

pea by Wall and Bryant (1981).

Most structural genes in animal genomes are single copy sequences of about 1,200 to 1,500 bp (Britten and Davidson 1976). Similarly, Gray and Cashmore (1976) have found that polysomal mRNA molecules from peas frequently have lengths greater than 1000 bp. This is in contradiction to the results obtained by Murray et al (1978), indeed, these authors have stated that sequences longer than 1000 bp do not constitute a readily measurable fraction of the pea genome. DNA renaturation studies do not therefore appear to be able to identify this class of single copy sequences. Murray et al (1978), have pointed out that since only a very small fraction of the DNA in a typical eucaryote may be required to code for structural genes, it is possible that the coding sequences in the pea genome form so small a fraction as to be undetectable by renaturation experiments. If 3 % of the pea genome consisted of sequences longer than 1000 bp, this would provide coding capacity for 10^5 genes, which is well within the estimates for the number of expressed genes provided by Goldberg et al (1978).

There is a wide variation in the DNA content of various members of the legumes. There is a 7 fold variation in haploid DNA content among the Vicia genus (Bennett and Smith 1976); pea and mung bean genomes differ in size by 9 fold (Murray et al 1981); and pea has enough DNA for 1.5 x 10^5 genes, while Vicia has enough for 6.5 x 10^6(Bennet and Smith 1976). One would have expected various members of closely related species to have similiar genome sizes, but this is apparently not the case. Some evidence towards explaining this anomaly was provided by Murray et al (1981) who showed the existence

of fossil repeats in the pea and mung bean genomes. They found that essentially all the sequences in the pea genome which reassociated with single copy kinetics at standard criteria, formed mismatched duplexes, and exhibited repetitive reassociation kinetics at lower, less stringent temperatures. These were termed fossil repeats because they were postulated to represent members of ancient, repetitive families which had diverged to such an extent, that they no longer reassociated with one another under standard conditions. True single copy sequences were defined as those which reassociated with single copy kinetics at both stringent and more permissive criteria. Using this definition, the original value of 30 % for the amount of single copy sequence in the pea genome (Murray et al 1978) was reduced to 1 - 3 %. deVries et al (1983) found, in confirmation of the results of Murray et al (1981), that only 0.62 % of the haploid pea genome was transcribed into mRNA.

In the mung bean, however, true single copy sequences appeared to constitute 70 % of the genome (Murray et al 1981). However, taking into account that the mung bean genome is 9 times smaller than the pea, in absolute terms the fraction of true single copy DNA is only slightly more than in the pea. This implies that the amount of true single copy DNA in these two genomes is relatively uniform, while the difference in genome size can be accounted for by varying amounts of repetitive DNA sequences.

In order to investigate whether the pattern of extensive sequence interspersion seen throughout the bulk of the pea genome was also present in the vicinity of single copy genes, Murray and Thompson 1982 sheared pea DNA to an average length of 1300 bp, and separated these

into fractions which contained highly repeated elements and fractions which did not. The relative concentration of sequences complementary to mRNA in both fractions was studied, by using each fraction to drive the reassociation of a cDNA probe. It was shown that sequences which coded for mRNA were mainly found in the fraction of 1300 bp fragments which did not contain highly repeated sequences. Thus, it was concluded that the pattern of sequence organisation in regions of the pea genome coding for mRNA, differed significantly from that found in total DNA.

Several investigators have proposed theories to explain the very large amounts of non coding repetitive DNA in eucaryotic genomes. These can be divided into 2 groups, those which postulate a positive function for repetitive sequences (Britten and Davidson 1969, 1971, Jelinek et al 1980), and those which state that the only 'function' of repetitive sequences is that of survival within genomes (Doolittle and Sapienza 1980, Orgel and Crick 1980). At the present moment, none of these theories have been conclusively proved. Studies on the localisation of apparently non functional sequences, their association with flanking sequences of structural genes, and the similarity of homologous nonfunctional sequences in different species and different organisms , will have to be carried out before this problem can be adequately resolved.

Attempts to isolate single copy genes from eucaryotic genomes , have thus to take into account the fact that single copy DNA constitutes a very small fraction of the genome. However, it is possible to achieve this by constructing 'gene libraries' and screening them with appropriate probes.

GENE LIBRARIES

A gene library comprises a collection of recombinant clones which together represent the entire DNA complement of the organism under study. In principle, if a gene library is complete, one should be able to isolate single copy genes from it i.e. genes which are reiterated only a few times if at all. A gene library can be constructed to various levels of completeness; if a library is spoken of as 80 % complete, there is a 80 % chance of finding any given sequence within it. Clarke and Carbon (1976), have derived a formula which allows the calculation of the number of recombinants required to construct libraries at various levels of completeness. This number depends upon the size of the genome and the average size of the cloned insert. Thus, assuming a mean insert size of 16, 19 and 17 kb, and a haploid genome size of 1.6×10^8, 1×10^9, and 3×10^9 bp for the genomes of Drosophila, silkmoth and rabbit respectively, the number of recombinants required for a complete (99 %) library, are, respectively, 4.8×10^2, 4×10^5, and 8.1×10^5 (Clarke and Carbon 1976). Thus, it can be seen that as the size of the genome increases, the size of the recombinant collection required to represent it also increases.

The term recombinant refers to the hybrid formed between a piece of genomic DNA and a suitable vector molecule. The vector molecule is essentially a piece of DNA containing an origin of replication. Ligation between the vector molecule and the fragment of foreign DNA results in the formation of a recombinant, which, if introduced into a host cell, is capable of making multiple copies of itself. This process has been termed variously, genetic

engineering , molecular cloning , or gene cloning , because the end result is a line of genetically identical organisms, all of which contain the same hybrid molecule. These organisms can be propagated and grown in bulk, enabling the amplification of the hybrid molecule. The recombinant can then be separated from its host organism, resulting in the isolation of large amounts of the unique, foreign DNA fragment which was originally ligated into the vector.

The vectors available for the cloning and construction of gene libraries fall into three types - plasmid, cosmid and bacteriophage. Plasmids are extrachromosomal, self replicating, stably inherited molecules. The ancestors of most plasmids in current use have been isolated from E. coli, and have 2 main advantages 1) the plasmid is maintained in growing bacterial cells at a high copy number, and 2) this copy number can be amplified several fold by growth in cells inhibited for protein synthesis by chloramphenicol (Clewell 1972). Plasmids constructed for cloning purposes have unique restriction enzyme cleavage sites situated such that insertion of a foreign DNA fragment at this site interrupts an antibiotic resistance gene, resulting in a recombinant molecule which confers antibiotic sensitivity to its host cell (Bolivar and Backman 1979, Kahn et al 1979, Bernard and Helinsky 1981). Plasmid vectors are useful for cloning unique pieces of DNA and for preparing gene banks from small sized genomes such as E. coli, which has a genome size of 4.5×10^6 bp (Klotz and Zimm 1972). Clarke and Carbon (1976) have estimated that a bank size of 1400 colonies is sufficent to create a 99 % complete gene bank, using fragments 13 kb long cloned into the plasmid vector Col E1 (Hershfield et al 1974).

Sherrat et al (1979) have reported that the transformation frequency for plasmid DNA molecules of about 15kb is 10^5 per µg, so at first sight it appears feasible to clone an eucaryotic genome in a plasmid vector. For the pea genome, using fragments of 13kb cloned into the vector pBR322 (Bolivar et al 1977), 1.7 x 10^6 individual recombinant clones would be required to construct a 99 % complete library. Assuming a transformation frequency of 10^5transformants per µg of DNA, 17 µg of recombinant DNA would be needed, not an unreasonable figure. However, this calculation assumes that all the DNA molecules in the 17 µg will be of a uniform size i.e. 17 kb. Using conditions which optimised the formation of hybrid, closed covalent circular molecules, Gray (1979) and Clarke and Carbon (1976) reported that only 20 - 35 % of molecules in a ligation mixture would be the desired hybrids. The remainder would consist of a variety of forms varying from recircularised plasmids and inserts, to very long plasmids resulting from multiple ligation events Assuming that molecules of varying sizes in a ligation reaction all transform at the same rate, and that 30 % of the ligation reaction consists of the desired recombinant molecules, 56 µg of recombinant DNA would be required to construct the pea gene library (17 µg of DNA out of the 56 µg would be the desired 17 kb recombinants). Ignoring the fact that this is a relatively large amount of DNA to obtain easily, this calculation is flawed, because DNA molecules of varying sizes do not all transform at the same rate. Small molecules e.g. pBR322 (4.36 kb) transform at high frequencies (10^7 / µg), while molecules larger than 10 kb transform at frequencies of between 10^4 and 10^5 per µg (Clarke and Carbon 1976). Thus, in a transformation system, small molecules will be preferentially selected over larger ones. One way of overcoming this problem is to isolate recombinant molecules of the right size from a

ligation reaction, and to use these in a transformation, but the recovery of large fragments, especially closed circular ones, from agarose gels, is relatively poor (Dretzen et al 1981). Another way in which the ligation mixture could be enriched for for the desired recombinant sequences would be by phosphatasing the vector (Seeburg et al 1977), or by the addition of homopolymeric tails to the vector and the DNA to be cloned (Wensink et al 1974), both of which should result in the sole formation of recombinant molecules upon ligation, but the problem still remains.

The other two cloning vector systems, bacteriophage and cosmid, both rely on in vitro packaging of DNA in order to raise transformation frequencies to a level where construction of eucaryotic gene libraries becomes feasible. In vitro packaging, as the name suggests, is the encapsidation of a DNA molecule by a phage in vitro, to form a mature, infectious phage particle (Hohn 1979, Hohn and Hohn 1974). Phage DNA in concatemeric form, produced by a 'rolling circle' mode of replication, is the substrate in vivo, for the packaging reaction. Hohn and Katsura (1977) have reviewed lambda morphogenesis, and only brief details of packaging, as it relates to in vitro packaging, will be given here. In the presence of the phage head precursor (the product of λ gene E is the major capsid protein), and the product of λ gene A, the concatemeric DNA is cleaved into monomers at cos sites. Cos (cohesive) sites are 12 bp long single stranded complementary sequences present at each end of a monomeric phage DNA molecule, and it has been shown (Matsubara and Takeshi 1982), that sequences of 40 bp at the right and 45 bp at the left arm of the phage are essential for cleavage. The rolling circle mode of replication produces concatemers of DNA joined at their cos sites. After it is

cleaved into monomers, the DNA is encapsidated and the product of λ gene D incorporated into what becomes a completed phage head. The products of λ genes W and F11 amongst others, then unite the head with a separately assembled tail structure to form the mature phage particle.

The principle of *in vitro* packaging is to supply the phage DNA with high concentrations of phage head precursor, packaging proteins, and phage tails *in vitro*. It is important to note that the only requirements for packaging are that the DNA to be packaged be between 37 - 52 kb long, and bounded by cos sites (Feiss *et al* 1977, Hohn 1975). In the case of cosmid vectors (see below), this DNA need not have any lambda functions since a cosmid effectively behaves as a very large plasmid. In the case of bacteriophage vectors, the length of inserted foreign DNA rarely exceeds 20 kb, because a proportion of the DNA between the cos sites has to code for lambda lytic functions. An *in vitro* packaging system is prepared by growing, inducing, and concentrating two *E. coli* lambda lysogens, each of which has an amber mutation in different steps of the lambda morphogenetic pathway (Hohn 1979). When the 2 extracts are mixed, and DNA of the correct size bounded by cos sites supplied, the 2 extracts complement each other, and the DNA is packaged. The packaged phage are then allowed to infect a host *E. coli* strain, producing plaques if a bacteriophage vector has been used, and transformed bacterial colonies if a cosmid has been used. The advantages of an *in vitro* packaging system are that 1) it allows high transfection frequencies, typically values of 10^7 - 10^8 transformants per μg, using intact 49 kb long lambda DNA are obtained. In a gene cloning experiment, this figure is reduced to 10^3 - 10^4 transformants per μg of recombinant DNA. This is a consequence of

the random association of fragments in a ligation reaction, producing molecules with a variety of fragment combinations, some of which are inviable. A second advantage is that due to the nature of the packaging reaction, an automatic size selection of molecules is made, only molecules 73 to 105 % of the size of wild type lambda being packaged (Feiss et al 1977). This facilitates the preferential cloning of large fragments of DNA, leading to a reduction in the number of recombinants required to construct a library.

A plasmid which contains a cos site of bacteriophage lambda is termed a cosmid (Collins and Bruning 1978, Collins and Hohn 1979). The cosmid and the foreign DNA are cleaved with an appropriate restriction endonuclease and ligated together at high DNA concentrations, this tending to favour the formation of long concatenates of DNA which are the substrates for in vitro packaging (Feiss et al 1977). The DNA forms most likely to be packaged are those which have cos sites at both ends, and whose total DNA length is between 37 - 52 kb. Small vector plasmids do not constitute a large background in the transduced population, which is therefore markedly enriched in large, hybrid plasmids. Using the cosmid vector pHC 79 (Hohn and Collins 1980), and 40 kb long genomic fragments, only 5×10^5 recombinant clones are required to construct a complete pea gene library.

The third type of cloning vector system uses bacteriophage lambda vectors. Because bacteriophage lambda has been the object of so much genetical research, it was natural that its genome would be modified to produce a wide variety of vectors. Wild type lambda DNA contains several sites for the commonly used restriction endonucleases

(Bam H1, Eco R1, and Hind 111), and so is not in itself useful as a vector. Derivatives of lambda have been produced which either have a single target site at which DNA can be introduced (insertion vectors), or have a pair of sites defining a fragment which can be removed, and replaced with foreign DNA (replacement vectors). The genome of lambda is so organised that the central one third contains genes which are entirely dispensable for lytic growth (Echols and Murialdo 1978). Using this fact, many groups have produced vector derivatives of both the insertional and replacement types (Blattner et al 1977, Murray and Murray 1975, Thomas et al 1974, Williams and Blattner 1979). Various methods have been employed in order to separate recombinant from non recombinant phage. In Charon 16A, the lac Z gene from E. coli has been incorporated into the lambda genome (Blattner et al 1977). A single Eco R1 site is located within this gene, and insertion of DNA at this site results in the inactivation of the gene. The resultant failure of the recombinants to produce β galactosidase is detected by their inability to hydrolyse a chromogenic substrate. Murray et al (1977) have constructed a vector in which insertion of foreign DNA within the immunity region of the phage destroys its ability to produce a functional repressor, with the result that recombinants give clear plaques, which can be distinguished from the turbid plaques of parental phage. Another important class of vectors developed by Karn et al (1980), and Loenen and Brammar (1980), employ the Spi⁻ phenotype as a selection process. Wild type lambda cannot grow on E. coli strains lysogenic for phage P2 - they display the Spi⁺ phenotype - Sensitivity to P2 Inhibition (Lindahl et al 1970). This sensitivity is due in part to the products of the lambda genes exo, bet, and gam which are situated within the replaceable region of the vector. Replacement of these genes by foreign DNA thus allows growth of the

recombinant in a P2 lysogen .

In a bacteriophage replacement vector, the background of non recombinant phage can be lowered by purification of the arms of the vector (Maniatis et al 1978). If the replaceable fragment which has been cleaved out of the vector is allowed to be present in a ligation reaction, it will compete with the foreign DNA for insertion into the vector, resulting in a high background of non recombinant phage. This can be avoided by first annealing the vector by its cos sites, restricting out the replaceable fragment, and separating out the two on a glycerol or sucrose gradient. The vector arms can then be purified from the gradient fractions and used in a ligation reaction.

One of the factors to be taken into account when constructing gene libraries, whether using bacteriophage or cosmid vectors, is that random cleavage of the eucaryotic DNA must be achieved. This is desirable so as not to exclude any particular class of fragment, which would then be under represented in the library. Large DNA fragments such as are required by bacteriophage and cosmid systems, can be obtained by non limit digestion of eucaryotic DNA, using a hexanucleotide sequence recognising enzyme such as Eco R1. However, there is always the possibility that a nonrandom distribution of Eco R1 sites near, or within the sequence of interest, might lead to its loss from the size selected population. A restriction endonuclease such as Sau 3A which recognises a tetra nucleotide sequence (Roberts 1981), is a better choice, since this sequence should be present, assuming a random distribution, every 256 (4^4) nucleotides. As the number of possible cleavage sites is very large, the number of ways of

generating a 20 Kb fragment by non limit digestion is also very large. This enzyme also has the advantage that its recognition sequence is contained within the Bam H1 recognition sequence, thus fragments obtained by cleavage with Sau 3A can be ligated into the Bam H1 site of vectors such as λ L47 (Loenen and Brammar 1980), or λ1059 (Karn et al 1980). Alternatively, randomly cut fragments can be obtained by mechanical shearing, and the appropriate linkers for cloning into a vector molecule added on, but this process is inefficent (Maniatis et al 1978). Once the eucaryotic DNA has been randomly cleaved with the appropriate restriction endonuclease, it must be size fractionated into molecules of the correct size for cloning into the relevant vector. This has to be done in order to prevent the coming together in one clone of fragments which were not initially adjacent in the genome. For instance, if an unsized fractionated randomly cleaved DNA population was ligated into a vector, the resultant clones might contain several small fragments from different chromosomes, which had come together in one clone due to multiple ligation events.

In contrast to cosmids, a much higher number of recombinant clones is needed to represent the genome if a bacteriophage vector is used. This is because lambda vectors can rarely accept DNA fragments longer than 20 Kb, while a cosmid vector can accomodate 40 Kb long fragments (Hohn and Collins 1980). Using the vector λ L47 (Loenen and Brammar 1980), 1.8×10^6 recombinant clones would be needed to represent the pea genome, contrasting with 5×10^5 clones needed using the cosmid pHC 79. However, this is not really a problem, since large numbers of bacteriophage can be easily screened for the sequence of interest, using the hybridisation technique of Benton and Davis (1977). The author has found low transformation frequencies using a

cosmid system to clone nuclear DNA (results not shown), and this, together with the problems associated with the stability of large cosmids in bacterial cells (Hohn and Hinnen 1980), make a bacteriophage system more desirable in gene library construction.

Few plant genes have been cloned to date, and most information has been derived from studies on animal genes. Before embarking on the isolation of the legumin gene from pea, it is useful to give a brief outline of the genes which have been cloned.

STRUCTURE AND REGULATORY ELEMENTS OF SOME EUCARYOTIC GENES

The transcription of mRNA from the genome by DNA dependent RNA polymerases is clearly an important stage in the regulation of gene expression. Of the sequences known to be important in transcription initiation, the TATA box or Goldberg Hogness box has been the best characterised (Breathnach and Chambon 1981). This sequence is located approximately 30 bp 5′ to the RNA transcription initiation site of many genes, and deletions at this box drastically reduced the rate of mRNA synthesis (Grosschedl *et al* 1981, Minty and Newmark 1980), and, in SV 40, resulted in the initiation of transcription at multiple sites instead of at a single, defined site (Mathis and Chambon 1981). All plant genes so far described have a sequence analagous to the TATA box (Messing *et al* 1983). Another sequence which may be involved in transcription initiation is the CAAT box, located some 80 - 100 bp upstream of the cap site of many genes (Benoist *et al* 1980). Messing *et al* (1983) have compared the 5′ flanking regions of several plant genes, and have found a conserved sequence showing analogy to the animal CAAT box, which they have named the AGGA box.

Recently, evidence has been found for the existence of elements involved in gene transcription which are located far upstream of the mRNA start site. Donahue et al(1983), found that deletion of a short nucleotide sequence in yeast (TGACTC), located between 138 to 194 bp upstream of the cap site of genes involved in amino acid biosynthesis, abolished the ability of the gene to be inductively regulated. Similiar sequence elements have been identified upstream of several co-ordinately expressed genes, e.g. the Drosophila heat shock, silkmoth chorion, chicken egg white protein and rat growth hormone genes (reviewed by Davidson et al 1983). Another type of upstream regulatory element, termed a transcriptional enhancer, was first identified in the genome of Simian virus 40 (Benoist and Chambon 1981). This enhancer had the property of greatly increasing gene expression in vivo, even when it was separated from its original site (70 - 155 bp upstream of the cap site), and reinserted several thousand base pairs away. Viral enhancers have been shown to increase expression of several eucaryotic genes, even when inserted at various points within, and downstream of the gene (reviewed by Khoury and Gruss 1983). Sequences similiar to the core sequence of viral enhancers proposed by Weiber et al (1983), have been found in an intron of the immunoglobulin heavy chain gene (Gillies et al 1983), and in the 5′ flanking region of human interferon genes (Ragg and Weissmann 1983). Conrad and Botchan (1982), isolated a human DNA sequence which hybridised to SV 40 DNA, and which functioned as an enhancer element in a thymidine kinase transformation system. This DNA sequence cross hybridised with many sequences in human DNA, suggesting that a family of such elements exists. No enhancer like element has been identified in plant genomes to date, but this may

solely be due to the paucity of experimental data on plant genes, rather than any major difference between plant and animal systems.

Most eucaryotic genes which have been cloned to date, have been found to contain intervening sequences (introns), which are spliced out of the primary transcript before the mRNA reaches the cytoplasm, where protein translation takes place (Breathnach and Chambon 1981). Comparison of the splice junctions of several eucaryotic genes has revealed that the intron junction begins with the dinucleotide GT at its 5′ end, and ends with the dinucleotide AG at its 3′ end (Breathnach et al 1978). Amongst plant genes, the soybean leghaemoglobin and actin genes each contain three introns (Brisson and Verma 1982, Shah et al 1982), while the french bean phaseollin and soybean glycinin genes contain five and three introns respectively (Fischer and Goldberg 1982 , Nielsen 1984, Slightom et al 1983). However, the eucaryotic split gene is not an universal phenomenom, and several exceptions have been found, notably the genes for histones (Kedes 1976), interferon (Nagata 1980), and, more recently, for zein, the maize storage protein (Pedersen et al 1982). More genes have to be analysed before it can be decided whether split genes are common, or rare, in plants.

In addition to the removal of introns, two other events take place before the mRNA is exported to the cytoplasm; 1) the addition of a methylated guanylate residue (cap) to the 5′ end, and 2) the addition of a poly A tail to the 3′ end. The cap structure has been found in most eucaryotic and viral mRNAs, thus arguing for a positive role in gene regulation (Shatkin 1976). Hentschel et al (1980) found that a short conserved sequence (5′ PyCATTCPu 3′) was located at the

5′ terminus of embryonic histone mRNAs, which they termed a 'cap box'. Evidence reviewed by Breathnach and Chambon (1981) indicates that transcription initiation occurs at the position corresponding to the first, capped nucleotide of the mature mRNA (cap site). A role in translation has been attributed to the cap site by Kozak (1980), who proposed a scanning model for the initiation of protein synthesis in eucaryotes, according to which ribosomes bind at the 5′ end of a message, and subsequently migrate along the RNA chain, searching for the first AUG translation initiation codon.

Most eucaryotic mRNAs, including those from higher plants, contain a terminal 3′ poly A sequence of 50 - 200 residues (Brawerman 1974), which is added post transcriptionally. Sequence analysis of cloned genes isolated using probes made from poly A^+ RNA, have revealed the existence of a conserved sequence (AAUAAA) 15 - 30 bases from the polyadenylation site, in the 3′ untranslated region of many animal messages (Benoist et al 1980). This sequence has also been found in thaumatin cDNA (Edens et al 1982), french bean phaseollin genes (Slightom et al 1983), soybean leghaemoglobin and actin genes (Brisson 1982, Shah et al 1982), maize zein and wheat gliadin genes (Bartels 1983, Hu et al 1982, Pedersen et al 1982), barley amylase cDNA (Rogers and Milliman 1983), and the pea legumin and vicilin cDNAs (Lycett et al 1983a). On the basis of this, the conserved sequence has been assumed to be a polyadenylation signal. However, not all mRNAs are poly adenylated, the most widely quoted exception being the genes for histones (Adesink and Darnell 1972). More recently, it has been found that a number of proteins appear to be coded for by mRNAs which are present in both the poly A^+ and poly A^- forms, e.g. protamine, caesin, ovalbumin, β actin, human β globin, and albumin

(reviewed by Katinakis et al 1980). Amongst plants, a similiar situation has been found for the mRNAs in cotton seeds (Galau et al 1981), and pea leaf mRNA (Gray and Cashmore 1976). The reasons for the existence of poly A+ and poly A- mRNAs remains unclear. However, as polyadenylation of mRNAs is known to increase their stability (Huez et al 1978), it is possible that poly A- mRNAs are required within tissues where translation can be very rapidly turned off by messenger degradation.

CONTROL OF STORAGE PROTEIN SYNTHESIS

Control of the process of storage protein synthesis could occur at the transcriptional, post - transcriptional, translational, or post - translational levels. Transcriptional control could be exercised by switching 'on' previously inactive storage protein genes during embryogenesis, leading to the accumulation of large amounts of mRNA. A demonstration of this type of control was shown amongst others, by Korn and Gurdon (1981), who described how previously inactive rRNA genes were reactivated by treatment with various deprotenising agents. Similar conclusions on the transcription of globin genes were reached by Ross et al 1974, who were able to induce synthesis of haemoglobin in previously inactive erythroleukaemic cells, by treatment with dimethylsulfoxide. Using Southern blot analysis and DNA from both cotyledons and leaves, Croy et al (1982) demonstrated that only about 4 legumin genes were present per haploid genome, arguing against selective amplification of legumin genes during development. Using similar techniques, 3 - 7 genes have been found for the pea vicilins (Gatehouse et al 1983), 4 for phaseollin (Hall et al 1983), and 4 - 5 for the soybean glycinins (Fischer and Goldberg 1982, Goldberg et al 1981). This lends further support to the

hypothesis that increased transcription of storage protein genes is probably a major control process. Other systems may also be involved in regulating the expression of seed protein genes. It is possible that after synthesis, seed protein hnRNA is selectively chosen over others for processing into mRNA by the processes of splicing, polyadenylation, methylation, or capping (Lewin 1975, Abelson 1979, Mans 1967). This is termed a process vs discard decision (Darnell 1979, Korn and Gurdon 1981), but no detailed examples have been described to date. Another possibility is that seed protein mRNAs have a longer life in the cytoplasm, and are thus translated more than other mRNAs, which are quickly degraded. This type of control system has been shown to operate during erythroblast diffrentiation in mammals, where specific preservation of globin mRNAs over the cellular mRNA population is apparent (Volloch and Housman 1981). This system also operates in oestrogenised roosters, where hormone withdrawl leads to a decrease in the stability of vitellogenin mRNA (Wiskocil et al 1980). Morton et al(1983) have shown that some seed mRNAs have a relatively long half life (greater than 10 h), and so this is a possible additional control mechanism. Post - translational modifications involving phosphorylation, proteolytic processing or glycosylation (Badenoch - Jones et al 1981), are all well established events in the production of proteins and these processes may also play a part in the control of the final production of the protein. However, transcriptional control is probably the primary mechanism, as many studies (Galau and Dure 1981, Goldberg et al 1981, Hall 1979, Meinke et al 1981), have shown that a few mRNAs become superabundant when an increase in a specific set of proteins occurs. As stated above (Morton et al 1983), the pea seed protein system also follows this trend.

This study aimed to isolate members of the family of genes coding for pea legumin. The availability of such genes is important for several reasons. The pea seed provides an excellent system for studying developmentally regulated gene expression, and the isolation of seed protein genes will make it easier to understand the precise nature of this regulation, by a study of the control elements governing transcription of the gene. This can be done, for example, by deleting or adding sequences to the gene, and investigating the effect of this on gene transcription. The availability of cloned genes also affords the opportunity to investigate DNA sequences adjacent to the genes, and thus to map the region of the chromosome where the storage protein genes are located. This should help to confirm the clustering of the major legumin genes at a single locus on chromosome 7 , which was found by classical mapping experiments (Matta and Gatehouse 1982). It should also be possible to study the evolution of storage protein genes throughout the Leguminosae, and thus to deduce whether or not they are all descended from a common ancestral gene. Finally, the in vitro modification of these genes to produce proteins of better nutritional quality is possible, though major advances in gene transfer and expression systems have to be made before this can be done.

In this work, genomic pea DNA was purified, and ligated into the isolated arms of the bacteriophage vector λ L47, to create a gene library of pea DNA. A clone containing the legumin gene was selected from this library by hybridisation with various cDNA probes. The clone was characterised by restriction mapping, hybridisation to legumin cDNAs, and by sequencing.

MATERIALS

Seeds of Pisum sativum L. (var. Feltham first) were obtained from Sutton Seeds Co, Torquay, Devon, U.K.

Chemicals and reagents, apart from those listed below, were obtained from BDH PLC, Poole, Dorset, and were of analytical grade or the best available.

Sigma Chemical Co. PLC, London, U.K.

Caesium chloride, Ethidium Bromide, Antibiotics, DNase, RNase, Lysozyme, Putrecine, Spermidine trihydrochloride.

Pharmacia Fine Chemicals, Uppsala, Sweden.

Sephadex G50 Superfine.

Bethesda Research Laboratories (UK) PLC. Cambridge, CB4 4BE, U.K.

Agarose - electrophoresis grade, Restriction enzymes, Alkaline phosphatase, polynucleotide kinase, T4 ligase.

Boehringer Mannheim, Bell Lane, Lewes, East Sussex. BN7 1LG.

Restriction enzymes, T4 ligase.

Boots (UK) PLC. Nottingham, U.K.

Disposable nappies.

Amersham International. White Lion Road, Amersham, U.K.

Nick translation kit, ^{32}P radio labelled nucleotides.

Whatman PLC. Springfield Mill, Maidstone, Kent.

3MM paper, DEAE cellulose paper.

Fisons Horticultural Division, Bramford, Ipswich.

Fisons Levington Universal Compost.

P.L. Biochemicals, PLC. P.O. Box 98, Northampton NN3 1AW.

Deoxy and Dideoxynucleotides.

Schleicher and Schull, D.3354, Dassel, W. Germany.

BA85 Nitrocellulose paper.

More Disposable, Dockfield Road, Shipley, W. Yorkshire, U.K.

1.5 ml and 0.5 ml disposable, plastic centrifuge tubes.

E.I. DuPont de Nemours and Co (Inc), Photo Products Dept. Wilmington, D.E. U.S.A.

Intensifying screens - "Lightning Plus Cronex".

Fuji Photo Co. Japan.

Fuji RX 100 X-ray film.

Difco laboratories, Detroit, Michigan, U.S.A.

Bacto tryptone, Bacto agar, Yeast extract.

Bectom, Dickinson and Co. Cockeysville, Maryland 21030. U.S.A.

BBL Trypticase peptone.

STRAIN LIST

In vitro packaging λ lysogens.

BHB 2688 : N205 rec A, gal$^+$, (λ imm 434, cIts, b 2$^\Delta$, red 3, Eam 4, Sam 7) /λ.

BHB 2690 : N205 rec A, gal$^+$, (λ imm 434, cIts, b 2$^\Delta$, red 3, Dam 15, Sam 7) / λ.

These packaging strains were a gift from Dr. D. Lonsdale, Plant Breeding Institute, Cambridge, England.

Phage λ cloning vectors and their host strains.

λ L47 - (srI λ 1-2)$^\Delta$, imm 434 cI, NIN 5$^\Delta$, chi A131.

E. coli 5K - sup E, rK$^-$, mK$^+$, thr$^-$, leu$^-$, B1, ton A.

E. coli L95 - sup E, sup F, rK$^-$, mK$^+$, ton A, (P2).

These strains were a gift from Professor W.J. Brammar, Dept. of Biochemistry, University of Leicester.

Recombinant clones

pHA 1 - a plasmid clone containing a pea ribosomal gene was a gift from Dr. R.E. Cuellar of the Plant Breeding Institute, Cambridge.

Plasmid vectors used in subcloning

pUC 8, pBR 322 derivative, amp resistant, lac po z$^+$.

E. coli JM 103 - lac pro, sup E, thi$^-$,str A, sbc B15, end A, hsp R4, F′ tra D 36, pro AB, lac I Z M15.

These strains were obtained from Bethesda Research Laboratories, Cambridge, CB4, 4BE.

METHODS

PREPARATION OF PEA DNA.

The method used for the extraction of high mol.wt. DNA was based on that of Graham (1977). Pea seeds (var. Feltham first) were sown in sterilised compost at a density of 50 seeds per tray, and were germinated in a temperature controlled (25°C) spray room for 4 days, after which they were transferred to an artificially illuminated growth cabinet and allowed to grow for a further 9 days. Plants were then transferred to a 4°C cold room and 40 g of leaves removed. Sufficent liquid nitrogen was added to freeze the leaves, which were ground in a mortar to a size small enough to fit in an electric coffee grinder, in which they were ground at full speed until a fine powder resulted. All steps from this point onwards were done as quickly as possible, in order to keep nuclease activity at a minimum.

40 ml of homogenising buffer were mixed in with the ground tissue, quickly followed by 10 ml of 5M sodium perchlorate and 40 ml of redistilled phenol. 40 ml of a 1 % v/v octanol / chloroform solution was added and the whole suspension gently shaken on a rotary shaker for 1 h at 4°C to ensure good mixing. The suspension was then centrifuged at 8000 rpm for 5 min in an MSE High Speed 18 centrifuge using a 6 x 100 ml rotor, in order to sediment leaf debris, and the supernatant removed and extracted with an equal volume of chloroform / octanol. After a further 1 min centrifugation at 8000 rpm to separate the phases, the aqueous phase was collected and the nucleic acids precipitated out by the addition of 3 volumes of ethanol at - 20°C. The nucleic acids were spooled out on a spatula, and redissolved by shaking very gently for 17 h, in order to minimise shearing of the high mol.wt. nucleic acids. Pronase at a concentration of 500µg / ml

was then added, and the solution gently shaken at 37°C until dissolution of the nucleic acids was complete. The nucleic acid solution was made up to 8.3 ml with resuspension buffer, and 8 g of CsCl added, along with ethidium bromide to a final concentration of 300 µg / ml. The well mixed solution was centrifuged at 40,000 rpm in a MSE Superspeed ultracentrifuge using a 10 x 10 ml rotor at 15°C for 36 h. The DNA bands were collected from the tube by side puncture using a wide bore hypodermic needle, pooled, at recentrifuged in fresh CsCl solutions as before. The DNA bands were collected, and the ethidium bromide removed by extensive partitioning against iso amyl alcohol. The DNA solution was then extensively dialysed against several changes of T.E. buffer over a 12 h period. DNA was then alcohol precipitated the pellet vacuum dried, and redissolved in a minimal volume of T.E. buffer.

Homogenising buffer	Resuspension Buffer	T.E. buffer
0.1M NaCl	50 mM Tris - HCl	10 mM Tris - HCl pH 8.0
0.025 M EDTA	10 mM EDTA pH 8.0	1 mM EDTA
2 % SDS		

PREPARATION OF PLASMID DNA.

Bacterial cultures were tested for the presence of the required plasmid by streaking out for single colonies on L agar plates supplemented with 20µg / ml of the appropriate antibiotics. Single colonies taken from these plates were used to make mini preparations of plasmid DNA, which were analysed by restriction enzyme cleavage and agarose gel electrophoresis, in order to ensure that the correct restriction enzyme cleavage sites were present. 25 ml of antibiotic

supplemented L broth was inoculated with the characterised single colonies, and grown to saturation in an overnight incubation. This overnight culture was used to inoculate 2 litres of antibiotic supplemented L broth at a 1 / 500 dilution, and grown to a A_{600} of 0.6, when chloramphenicol (150 µg / ml) was added, and the culture incubated for 14 - 16 h to amplify the plasmid (Clewell 1972). Routinely, a cleared lysate procedure based on that of Katz (1977) was employed . Cells were harvested by centrifugation at 10,000 rpm at 4°C for 10 min in a MSE 18 centrifuge using a 6 x 250 ml rotor, and resuspended in a total volume of 25 ml iced 25 % sucrose in 0.05 M Tris - HCl pH 8.0. 10 ml of a 10 mg / ml solution of lysozyme in 0.25 M Tris - HCl pH 8.0 was then added to the resuspended cells, and the whole shaken gently in a 37°C water bath for 30 s in order to promote sphaeroplast formation. Lysis was achieved by adding 17.5 ml of sterile 0.2 M EDTA and, after a further 5 min at 0°C, 27 ml of Triton X100 solution (2 % Triton in 0.05 M Tris - HCl, 0.0625 M EDTA pH 8.0). After good mixing, the resultant viscous solution was left at 0°C for 20 min, whirlimixed for 10 s, stood on ice, and again whirlimixed for a further 10 s. The solution was then centrifuged at 18,000 rpm at 4°C for 1 h in a MSE 18 centrifuge using a 6 x 100 ml rotor, and the supernatant carefully decanted off and saved. CsCl gradients were prepared using the saved supernatant to give a refractive index of between 1.3900 - 1.3905, and processed as for the pea DNA preparation. Precipitated DNA was redissolved in T.E. buffer and stored in small aliquots at - 80°C.

PREPARATION OF PHAGE DNA (MAXIPREP).

The method used for the extraction of phage DNA was based on that of Blattner (1977). 5 x 10^8 pfu of phage were preadsorbed to 0.3 ml of a fresh overnight bacterial culture (grown in L broth + 10 mM MgSO , 0.4 % maltose), for 15 min in a water bath at 37°C. This was diluted into 500 ml of prewarmed L broth (+10 mM $MgSO_4$), and shaken vigorously for 16 h at 37°C. Bacterial debris was seen at this stage, and chloroform at 2 ml / litre was added to promote lysis. This unclarified phage lysate was concentrated by precipitation with polyethyleneglycol according to the method of Yamamoto (1970). 40 g / litre of NaCl and 1 µg / ml of DNase and RNase were added, dissolved, and left at room temperature for at least 1 h. The phage lysate was then clarified by centrifugation at 10,000 rpm for 10 min in the MSE 18 centrifuge, using a 6 x 250 ml rotor. Supernatant was carefully decanted, and the phage pellet resuspended in phage buffer by gentle rotary shaking in the cold. The volume of phage buffer was chosen so as to concentrate the phage 50 fold.

This crude phage preparation was concentrated and purified by centrifugation through a three step CsCl gradient with steps of 1.3, 1.5, and 1.7 g / ml of CsCl. Centrifugation was at 35,000 rpm at 20°C for 2 h, using a MSE Prepspin ultracentrifuge and a 3 x 25 ml rotor. The phage band was seen at the interface between the 1.3 and the 1.5 step gradients, and was withdrawn by side puncture using a hypodermic needle. Overhead illumination of the centrifuge tubes using a collimated light source was found to be ideal for visualising the phage bands.

CsCl was dialysed out several changes of T.E. buffer and the phage solution extracted with an equal volume of phenol equilibrated

in 0.5 M Tris - HCl pH 8.0. The extraction was performed by gentle inversion and rolling of the tube, never by shaking, in order to minimise shearing of the high mol.wt. DNA. After 4 phenol extractions, the aqueous phase was removed and extensively partitioned against chloroform in order to remove residual traces of phenol. The solution was then dialysed extensively against T.E. buffer to remove any remaining organic solvents, and the A_{260} measured. After alcohol precipitation, the phage DNA was redissolved in a minimal volume of T.E. buffer, distributed into eppendorf tubes, and stored at - 80°C.

SMALL SCALE PREPARATION OF PLASMID DNA.

The method used was based on that of Birnboim and Doly (1979). A 10 ml culture of antibiotic supplemented L broth (20 µg / ml) was inoculated with the appropriate bacterial strain and grown to saturation overnight. The cells were pelleted by centrifuging at 4,000 rpm for 10 min in a MSE bench centrifuge, after which they were resuspended in 200 µl of a 4 mg / ml lysozyme solution and kept on ice for 30 min. 600 µl of a NaOH - SDS solution (0.2 N NaOH, 1 % SDS) were then added. After 5 min, 450 µl of a 3 M sodium acetate solution pH 4.8 was added and the tube shaken every 5 min until a large clot of DNA and protein had formed. After a 15 min centrifugation in a benchtop Polska microcentrifuge, 1.1 ml of the supernatant was taken and alcohol precipitated. The DNA pellet was redissolved in 100 µl of T.E. buffer + 1 µg / ml boiled RNase, and 10 µl of this was used for restriction enzyme analysis.

SMALL SCALE PREPARATION OF PHAGE DNA.

The method used was based on that of Leder et al (1977). A single, well isolated phage plaque was picked and put into 1 ml of phage buffer (10 mM $MgSO_4$, 10 mM Tris - HCl pH 7.5) with chloroform. 0.5 ml of this bacteriophage suspension was mixed with 1.6 x 10^8 bacterial cells of the appropriate host strain, and incubated at 37°C for 15 min. 4 ml of L broth with maltose (0.4 %) was then added, the broth transferred to a sterile 50 ml conical flask, and shaken vigorously at 37°C for 9 h. 0.1 ml of chloroform was then added, and the culture shaken for a further 15 min to aid lysis. Lysate was then centrifuged at 4°C for 10 min in a MSE 18 centrifuge, using a 6 x 100 ml rotor. To the supernatant, 1µg / ml of RNase and DNase 1 were added and incubated at 37°C for 15 min. An equal volume of a solution containing 20 % w / v of polyethylene glycol and 2 M NaCl was added and incubated at 0°C for at least 1 h. The precipitated bacteriophage particles were recovered by centrifugation at 10,000 rpm for 20 min at 4°C in the Sorvall RC5B centrifuge, using a 4 x 25 ml rotor. The supernatant was totally removed and the phage pellet resuspended in 0.5 ml of phage buffer. To the resuspended pellet, 5 µl of 10 % SDS and 5 µl of 0.5 M EDTA pH 8.0 were added, and incubated at 68°C for 10 min. This solution was extracted once with phenol, once with phenol / chloroform, and once with chloroform. To the final aqueous phase, an equal volume of isopropanol was added, and the tube stored at - 70°C for 20 min, after which it was centrifuged in a Polska microcentrifuge for 15 min. The pellet was washed once with 70 % ethanol, dried, and resuspended in 50 µl of T.E. buffer. 10 µl of this solution was used in restriction enzyme analysis.

PREPARATION OF 15 - 20 Kb FRAGMENTS OF Sau 3A CLEAVED PEA DNA.

The procedure used was essentially that described by Maniatis et al (1978). Restriction enzyme conditions were established which yielded the maximum amount of DNA in the required size range. Large scale digestions were then performed, using 100 µg of pea DNA per digestion in a total volume of 500 µl, and 0.5, 1, and 2 times the amount of Sau 3A which yielded the maximum amount of 15 - 20 Kb fragments. After cleavage, the reactions were terminated by adding EDTA to a concentration of 20 mM. Before loading onto the gradient, the restriction enzyme digests were heated to 70°C in a water bath for 10 min, and rapidly cooled, in order to separate annealed cohesive ends. Glycerol gradients were chosen instead of the sucrose ones used by Maniatis, because glycerol was easily autoclavable. 100 µg were carefully layered onto the top of a 25 ml 10 - 60 % glycerol gradient (loading of more DNA onto a single gradient caused poor separation of size fractions), and centrifuged at 25,000 rpm for 14 h at 10°C, in a MSE Prepspin 65 ultracentrifuge using a 3 x 25 ml rotor. Gradients were unloaded using an ISCO fractionator attached to an ISCO u.v. absorbance cell. Alternate fractions from each gradient were analysed by agarose gel electrophoresis and the appropriate size fractions pooled. High concentration of glycerol were found to result in poor recovery of DNA after alcohol precipitation, so the pooled fractions were dialysed against T.E. buffer at 4°C for 15 h, with several changes of buffer. In order to further enhance DNA recovery, the volume of the aqueous solution was reduced by extraction with butanol before alcohol precipitation, and the precipitated DNA redissolved in a minimal volume of T.E. buffer.

GRADIENT SOLUTIONS

Glycerol 10 % and 60 % v/v

25 mM NaCl

10 mM Tris - HCl pH 7.5.

1 mM EDTA

H_2O to volume

PREPARATION OF L47 PHAGE ARMS.

The procedure used was a modification of that described by Maniatis et al(1978). The phage was annealed, the annealed arms ligated, and the 'stuffer' fragment removed by Bam H1 digestion, as described below. 114 µg of λ L47 were annealed for 2 h in annealing buffer, cooled to room temperature, and put on ice. Ligation was accomplished by the addition of 10X ligase buffer, ATP to 1 mM, and 11.4 WU (Weiss units) of T4 ligase. The total volume of the ligation was 500 µl, and the ligation was continued at 25°C for 7 h. The reaction was then extracted twice with phenol / chloroform and twice with ether. It was then alcohol precipitated, redissolved in 200 µl of T.E. pH 7.5, and cleaved with 500 U of Bam H1 for 4 h. Three 10 - 60 % 25 ml glycerol gradients were made, and 38 µg of annealed, ligated and restricted DNA layered onto each. Centrifugation was at 25,000 rpm for 14 h at 10°C using the MSE Prepspin 65 ultracentrifuge with the 3 x 25 ml rotor. Gradients were unloaded, and the appropriate fractions pooled, processed and stored as in the section on preparation of 15 - 20 Kb fragments of restricted pea DNA.

5X ANNEALING BUFFER

500 µl 1M Tris - HCl pH 8.0

50 µl 1 M MgCl

450 µl H_2O

PREPARATION OF AN IN VITRO PACKAGING SYSTEM.

This packaging system was prepared from two E. coli lysogens - BHB 2690 (Dam) and BHB 2688 (Eam). These lysogens are described by Hohn (1977) and Hohn (1979). Before starting a preparation, the presence of the cI mutation was checked by streaking single colonies of both lysogens at 32°C and 42°C. Growth should only be observed at 32°C. Extracts of the two lysogens were prepared in two ways - BHB 2688 was prepared using the freeze thaw lysate (FTL) method, while BHB 2690 was prepared using the sonicated extract (SE) method.

PREPARATION OF FREEZE THAW LYSATE

60 ml of L broth were inoculated with a single colony of BHB 2688 and grown to a A_{600} of 0.64. 3 x 500 ml of L broth were then inoculated with 16 ml of the culture and grown with vigorous aeration at 32°C until the A_{580} reached 0.32. The cultures were then induced, by placing them in a 70°C water bath until the culture temperature reached 45°C, and then shifted to a 45°C water bath for 20 min. Cultures were equally distributed amongst 6 x 1000 ml flasks, shaken at 37°C for 1 h, and tested for the successful induction of lysogens by shaking 2 ml of the culture with 2 drops of chloroform. Induction

was shown by complete clearing of the culture and the presence of lysed bacterial debris. Cultures were cooled on ice for 20 min, harvested by centrifugation in a 6 x 250 ml rotor in a MSE 18 centrifuge at 10,000 rpm for 10 min, and the pellets resuspended in 0.5 ml cold 10 % sucrose, 50 mM Tris - HCl pH 7.5. The pooled, resuspended pellets were dispensed in 3 ml aliquots into ultracentrifuge tubes, to which 75 µl of a fresh lysozyme solution (2 mg/ml in 0.25 M Tris - HCl pH 7.5) was added and mixed. the tubes were then quick frozen in liquid nitrogen, stored at - 80°C for 15 min, and thawed out at room temperature. If the resuspended pellets at this stage did not become extremely viscous, the freeze thaw cycle was repeated. To each aliquot, 75 µl of buffer M1 was added, mixed, and the aliquot centrifuged at 35,000 rpm for 1 h at 10°C using the MSE Prepspin ultracentrifuge with the 10 x 10 ml rotor. Supernatant was removed, dispensed in 100 µl aliquots in eppendorf tubes, and stored at - 80°C as the FTL extract.

PREPARATION OF SONICATED EXTRACT

Growth, induction, and harvesting of cells of BHB 2690 was as for BHB 2688. Each pellet after centrifugation was resuspended in 0.5 ml of Buffer A. Pellets were then pooled, diluted with 2.6 ml of Buffer A, and dispensed into 1ml aliquots for sonication. Sonication was performed using a MSE Soniprep sonicator at a power setting of 6. 10 x 3 s bursts were used for each aliquot. After 7 bursts, the cultures became extremely viscous. Sonication was continued until the viscosity just disappeared. The lysed, sonicated cultures were centrifuged at full speed in the Hawksley Haematocrit microcentrifuge for 10 min, and the supernatant stored in 100 µl amounts at - 80°C.

PACKAGING PROTOCOL FOR λ DNA.

7 µl Buffer A
2 µl DNA (0.5 µg)
1 µl Buffer M1
10 µl S.E.
10 µl F.T.L.

After incubation at 25°C for 60 min, 500 µl of phage buffer was added. The packaging was treated as a phage suspension and titered on the appropriate bacterial host.

Buffer A

110 µl H_2O
6 µl 0.5 M Tris - HCl pH 7.5
300 µl spermidine / putrescine soln.
9 µl 1M $MgCl_2$
75 µls 0.1M ATP pH 7.0
1 µl β - mercaptoethanol

Buffer M1

20 mM Tris - HCl pH 8.0
3 mM $MgCl_2$
0.05 % v/v β mercaptoethanol
1 mM EDTA pH 7.0

Spermidine / Putrescine solution

0.05 M spermidine (neutralised in Tris)
0.1 M putrescine

SCREENING OF BACTERIOPHAGE LAMBDA LIBRARIES BY HYBRIDISATION

The procedure used was essentially that described by Benton and Davis (1977), but screening of the library was carried out using cafeteria trays as described by Blattner et al 1978. Aliquots of a packaging mixture containing 5 x 10^5 bacteriophage in a volume of 3 ml of phage buffer, were mixed with 3 ml of plating bacteria and incubated at 37°C for 15 min. A 30 cm x 40 cm plastic cafeteria tray was ethanol sterilised, and 1.5 l of molten BBL bottom agar poured into it. When the surface had completely dried,(a laminar flow

cabinet was used to maintain sterile conditions), the transfected cells were mixed with 50 ml of molten top layer agarose at 50°C and carefully poured over the surface of the cafeteria dish. Top layer agarose was used rather than top layer agar, because it was found that an agarose top layer was stronger than agar. When the top layer agar had set, another sterile cafeteria tray was taped into place to act as a lid, and the megaplate incubated at 37°C. Incubation was continued for 10 h, until the phage plaques were a diameter of 1 mm, and just beginning to come into contact with each other. At intervals during the incubation period, the top lid of the cafeteria tray was carefully removed, and the condensed moisture swabbed off with sterile pads. At the end of the incubation period, the megaplates were placed at 4°C to harden the top layer agar, as this aided subsequent manipulations.

A filter which would easily fit over the surface of the megaplate was cut from a roll of nitrocellulose paper, and holes punched around its periphery to act as orientation marks. The filter was then carefully layered over the surface of the megaplate, being careful not to trap any air bubbles, and the agar marked with ink at positions corresponding to the holes punched in the filter. The filter was left in contact with the agar surface for 5 to 10 min, after which it was carefully peeled off and another substituted in its place. Thus, duplicate lifts were taken from each megaplate. The filters were then sequentially layered plaque side up, on pads of blotting paper soaked in the following solutions.

SOLUTIONS	TIME
0.5 M NaOH	5 min
0.1 M NaOH, 1.5 M NaCl	5 min
0.5 M Tris - HCl pH 7.5, 1.5 M NaCl	5 min
2 x SSCP	1 min

Filters were then blotted dry between pads of Whatman 3MM paper and baked at 80°C for 2 h, after which they were ready to be hybridised with an appropriate probe. Hybridisation of the nitrocellulose filter replicas was carried out in heat sealed plastic bags at 65°C. Filters were prehybridised sequentially in 200 ml of the following solutions:
(1) 3 x SSC, (2) 3 x SSC, 10 x Denhardts solution (Denhardt 1966), (3) 3 x SSC, 10 x Denhardts solution, 100 µg / ml denatured salmon sperm DNA. The probes used were labelled by nick translation to a specific activity of at least 10^8 / µg, and 10^7 cpm of this was usually used per megafilter. The probe was denatured before use by placing in a 98°C water bath for 10 min, and was immediately added to 50 ml of pre heated pre hybridisation solution (3). The probe was then added to the megafilter, the bag heat sealed, and hybridised in a 65°C water bath for 10 - 12 h, with shaking. The probe was then poured off, and the filters sequentially washed in 200 ml of the following solutions (1) 3 x SSC (2) 1 x SSC (3) 0.1 x SSC. The filter was then dried in a 80°C oven for 15 min, and prepared for autoradiography. Autoradiography was carried out at - 80°C for 3 days, using a flashed film and one intensifying screen. After the film was developed, hybridising signals which appeared in the same place on duplicate filters were noted, and their position marked on a piece of tracing paper placed over the films. The position of the filter orientation marks was also

marked on the tracing paper. The approximate position of the positive hybridising plaque was identified, by aligning the filter orientation marks on the tracing paper and on the megaplate, and a 0.7 mm plug of agar was cut from the megaplate. The agar plug was stored in 2 ml of phage buffer, and an aliquot replated so as to obtain approximately 500 plaques on a 85 mm agar petri dish. These plaques were screened a second time by hybridisation, and a single, well isolated positive plaque picked, and used to make a phage stock .

Phage Buffer	20 x SSPE	50 x Denhardt's
10 mM MgSO	3.6 M NaCl	Ficoll - 5g.
10 mM Tris HCl pH 7.5	200 mM NaH PO pH 7.4	PVP - 5g.
	20 mM EDTA pH 7.4	BSA - 5g.
		H_2O to 1litre

Denatured salmon sperm DNA was prepared according to the method of Maniatis (1982).

SCREENING SMALL NUMBERS OF BACTERIAL COLONIES BY HYBRIDISATION

This procedure, (Grunstein and Hogness 1975), was used whenever it was necessary to screen small numbers of colonies, e.g. when subcloning a lambda fragment into a plasmid vector. The plasmid vector used for all subcloning experiments was pUC 8 (Viera and Messing 1982). The transformation mixture was plated out on selective media plates and grown overnight at 37°C.The following morning, a nitrocellulose filter with an imprinted grid pattern was placed on to a selective agar plate and up to 100 transformants picked onto the filter in an ordered pattern. The same transformants were also picked onto a selective master plate in the same pattern. The plates were inverted and incubated at 37°C until the colonies had grown to 2mm in diameter. The master plate was stored at 4°C at this point. The

filter was then transferred colony side up to an agar plate containing chloramphenicol (10 µg/ml), and incubated for a further 10 h at 37°C for amplification (Clewell 1972). The nitrocellulose filter containing the amplified bacterial colonies was processed by placing it colony side up on blotting paper saturated with the following solutions:

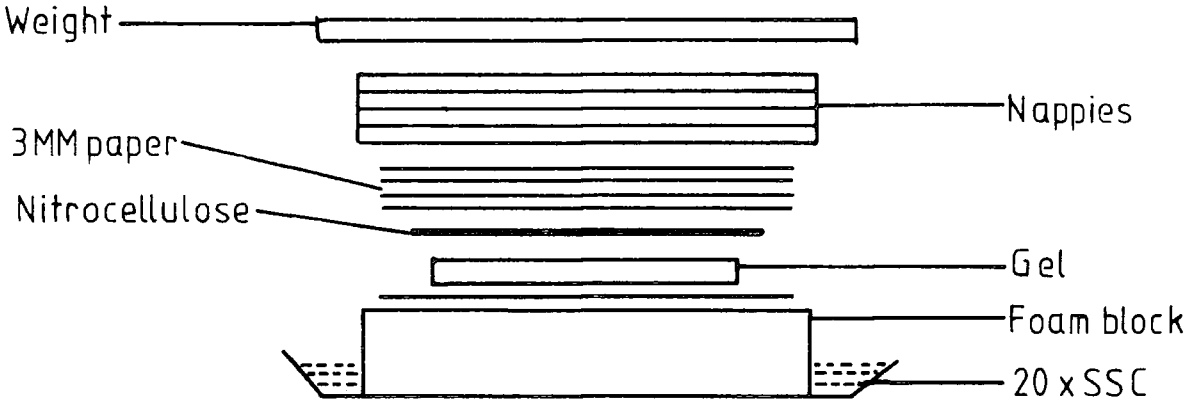
Solutions	Time
0.5 M NaOH	7 min
1.0 M Tris - HCl pH 7.4	2 min
1.0 M Tris - HCl pH 7.4	2 min
1.5 M NaCl, 0.5 M Tris - HCl pH 7.4	4 min

The filter was then dried by blotting with filter paper and then baked under vacuum at 80°C for 2 h. Hybridisation conditions were exactly as for screening gene libraries. Volumes of the solutions were altered, however, so that 50 ml of solution were used in all washing and hybridisation steps. The time required for autoradiography was determined by the number of counts present on the filter after the 0.1 x SSC wash. Once the positive colonies had been identified on the filter, they were picked off, and stored in glycerol at - 80°C. Mini preparations of the plasmid were then performed.

TECHNIQUE OF SOUTHERN TRANSFER AND SUBSEQUENT HYBRIDISATION

The technique used was essentially that described by Southern (1975). After gel electrophoresis was completed, the DNA was stained with ethidium bromide and photographed under u.v. light. The gel was transferred to a glass dish and denatured by soaking in a frequently changed solution of 1.5 M NaCl, 0.5 M NaOH, 1 mm EDTA for 1 h. It was

then neutralised by soaking for another hour in a solution of 3 M NaCl, 0.5 M Tris - HCl pH 7.5, 1mm EDTA. After equilibration for a third hour in 20 x SSC, it was transferred to a blotting apparatus as detailed below:



The gel was carefully placed on the porous foam block, and all air bubbles expelled by gently squeezing the gel. A 20 x 20 cm sheet of nitrocellulose paper was floated onto the surface of some distilled water, and then submerged for 10 min. The distilled water was replaced with 20 x SSC solution and the filter further equilibrated for 0.5 h. It was then placed on top of the gel, expelling all air bubbles, so that one edge extended 1cm over the line of slots at the top of the gel. A piece of Whatman 3MM paper cut to the same size as the gel was wetted in 20 x SSC solution and placed over the filter. This was followed by three more layers of dry 3MM paper. 4 layers of dry disposable nappies were then placed over the 3MM paper, and finally, a glass weight of 500 g placed on top of the nappies. The

The whole apparatus was then transferred to a cold room at + 4°C and transfer of DNA allowed to proceed for 12 - 24 h. After transfer had been completed, the nappies and 3MM paper layers were removed. The position of the gel slots was carefully marked on the filter and the filter gently peeled away from the gel. It was then trimmed to size, and baked at 80°C in a vacuum oven for 2 h between two layers of Whatman 3MM paper. After hybridisation had been performed, the filter was autoradiographed.

AUTORADIOGRAPHY

The following procedure was applied to any DNA sample bound to a solid support - e.g. megafilters used in screening libraries, or sheets of nitrocellulose used in Southern blots. The material to be autoradiographed was dried in a vacuum oven for 10 min, and mounted on a piece of blotting paper using autoclave tape. Radioactive marker ink was spotted in a random orientation on the blotting paper which was then taped to a glass plate. The whole was encased in a plastic bag - this prevented contamination of intensifying screens and stopped the sample from sticking to the film. If the number of counts on the sample was low (2 - 3 cpm), then the X ray film was preflashed, and an intensifying screen used. Exposure times using flashed film and intensifying screens varied depending on the counts present in the sample ; 2 - 3 cpm meant an exposure time of at least 3 days, while 100 - 200 cpm could be exposed for 2 - 3 h. In the darkroom, a sheet of X ray film was placed over the sample encased in it's plastic bag. Flashing of the film was carried out using a Sunpak flashgun at a distance of 3 feet. One DuPont Cronex intensifying screen was placed over the flashed film, followed by a glass plate. The 2 plates were

then held together with rubber bands or autoclave tape, wrapped in several layers of black plastic bags to exclude light, and exposed at -70°C for the required time. At the end of the exposure time, the film was removed from the sample in a dark room, and developed by immersing in Kodak X O MAT developer for 5 min. It was washed for 1 min, and then fixed for 4 min in Kodak fixer. The developed film was then dried at room temperature.

LABELLING DNA FRAGMENTS WITH ^{32}P IN - VITRO

DNA was labelled *in vitro* to a specific activity of 1 x 10^8 cpm/µg, using the nick translation kit supplied by Amersham International. The explanatory booklet supplied with the kit detailed protocols for labelling the DNA to higher and lower specific activities. The nick translation reaction was terminated and the products separated from unincorporated label by separation on a 5 cm column of Sephadex G50 superfine. The Sephadex was hydrated before use by boiling for 1 h in an elution buffer containing 150 mM NaCl, 10 mM EDTA, 50 mM Tris - HCl pH 7.5, and 0.1 % SDS. A sterile, disposable plastic pipette was plugged with sterile, siliconised glass wool, the column poured, and equilibrated with the elution buffer. The nick translated DNA sample was carefully layered on top of the column, and its progress through the column monitored with a Geiger counter. 0.4 ml fractions were collected, 1 µl amounts of the various fractions dispersed in toluene scintillation fluid, and the radioactivity counted in a Packard Tricarb Scintillation Counter. The appropriate fractions were pooled, and used directly in hybridisation reactions.

AGAROSE GEL ELECTROPHORESIS

Horizontal, submerged, agarose slab gels were used to separate, identify, and isolate DNA fragments (Aaij and Borst 1972). Gels were cast when molten in a perspex mould obtained from Shandon Instruments, and allowed to cool for at least 30 min before use. the following agarose gel concentrations were found to be useful:

Amt.of agarose (%)	Separation range of linear DNA molecules(Kb)
0.3	60 - 5
0,6	20 - 1
0.7	10 - 0.8
0.9	7 - 0.5
1.2	6 - 0.4
1.5	4 - 0.2
2.0	3 - 0.1

Gels with an agarose concentration lower than 0.4 % were made at 4°C with a base of 1 % agarose (Fangman 1978), in order to decrease the time required for the gel to set and to provide rigidity to the gel. Samples to be electrophoresed were mixed with agarose loading beads (15 % v/v), and loaded into the gel slots using a Gilson micropipette.

Fast gels were usually electrophoresed at 150 v (15 v/cm), until the bromophenol blue marker dye in the samples had migrated 2/3 of the way along the gel. Slow gels were run at 30 v (3 v/cm) for 12 - 16 h. Ethidium bromide at a concentration of 0.5 µg/ml was included in both the gel and the electrophoresis buffer in order to detect the DNA (Sharp et al 1973). After electrophoresis was complete, the gels were photographed using transmitted ultraviolet light at 254 nm and Polaroid film ASA 3000. An exposure time of 9s at f 5.6 was usually

found to be adequate.

Agarose gel buffer

Tris base	--	96.8 g
0.25 M EDTA	--	80 ml
Acetic acid	--	32 ml
H_2O	--	to 2 l

POLYACRYLAMIDE GEL ELECTROPHORESIS

4 % polyacrylamide gels were usually used. Glass plates were throughly cleaned until a high polish was obtained, and clamped together with perspex spacers and silicone grease on 3 sides. 6 ml of 40 % acrylamide (38 g acrylamide, 2 g bis - acrylamide) and 6 ml of 10x TBE buffer were made up with distilled water to 45 ml, and degassed. 15 ml of glycerol were then added, mixed, followed by 400 µl of 10 % ammonium persulphate and 20 µl of TEMED. The solution was taken up in a syringe and squirted into the space between the two glass plates, taking care to avoid the formation of air bubbles. A comb was then inserted, and the gel allowed to polymerise for at least 1 h. The comb was then removed, the plates clamped in a vertical electrophoresis apparatus and the buffer reservoirs filled with 1 x TBE buffer. The DNA samples were made 1 % with respect to glycerol dyes, and loaded into the slots. Electropharesis was continued until the bromophenol blue was 3/4 of the way down the gel. The gel plates were prised apart, the gel soaked in a solution of 200 µg/ml of ethidium bromide for 30 min, destained in distilled water for 30 min, and photographed under u.v. light.

10X TBE buffer

Tris base	108 g
Boric acid	55 g
EDTA	10.5 g
H_2O	to 1 l

RECOVERY OF DNA FRAGMENTS FROM AGAROSE GELS

The method used was essentially that of Dretzen et al (1981). Strips of Whatman DE81 DEAE cellulose paper were soaked for several hours in 2.5 M NaCl, washed several times in distilled water, and stored dry at room temperature. After electrophoresis of DNA through agarose gels, slits were cut both above and below the fragment to be isolated, and strips of prepared DEAE cellulose paper slipped into each. The insertion of paper above the desired fragment prevented contamination of this fragment with higher mol. wt. DNA. Electrophoresis was continued until the DNA bands had completely entered the paper. The paper was then removed, and extensively washed with distilled water to remove agarose. After drying, the paper was placed in siliconised Corex tubes and 300 - 700 µl of a solution of 20 mM Tris - HCl pH 7.5, 1 mM EDTA, 1.5 M NaCl added per 50 mm^2 of paper. After shredding the paper on a Whirlimixer until a thick slurry was obtained, the tube was incubated at 37°C for 2 h with occasional agitation. The mixture was then transferred to a 1 ml plastic pipette tip which had been plugged with siliconised glass wool. The tip was placed inside a Corex tube, and centrifuged at 3000 rpm for 2 min in a Sorvall HB4 rotor at 4°C . The eluate was then extracted with n - butanol to remove ethidium bromide, and alcohol precipitated as usual. This technique yielded DNA of sufficent purity for labelling,

restriction enzyme cleavage, and ligation.

TRANSFORMATION

The method used was based on the procedure of Dagert and Ehrlich (1973) who incubated *E. coli* cells in calcium chloride to increase the level of transformation. A single colony from a L agar plate was inoculated into 50 ml of L broth and incubated at 37°C with agitation until the A reached 0.2. The culture was chilled on ice for 10 min, and the cells harvested by centrifugation for 10 min at full speed in a MSE bench top centrifuge. The pellet was carefully suspended in 20 ml of cold freshly made up 0.1 M $CaCl_2$ and incubated at 0°C for 20 - 25 min. The cells were harvested by centrifugation as before and resuspended in 0.1 M $CaCl_2$. They were then kept on ice for 24 h, after which they were used for transformation. This was carried out by adding 10 µl of the DNA solution (10 - 100 ng) to 100 µl of the cell suspension. If the DNA used had been subject to ligation, it was recovered from the ligation mix by alcohol precipitation and resuspended in 0.1 M calcium chloride. The mixture was incubated on ice for 10 min, and then at 37°for 5 min. It was then diluted into 2 ml of L broth and incubated at 37°C for 1 h with shaking. 10 µl of this was then spread on a cool dry L plate containing, if desired, 100 µg/ml of antibiotics.

TRANSFECTION

Dilutions of phage from 10^{-2} to 10^{-15}were made in phage buffer (10 mM Tris - HCl pH 7.5, 10 mM $MgSO_4$). Host cells competent for transfection were made by harvesting a 10 ml exponentially growing

culture of cells (in L broth + 0.4 % maltose), by centrifugation in a MSE bench top centrifuge at full speed for 10 min. They were then resuspended in 2 ml of ice cold 10 mM $MgSO_4$. 100 µl of these competent bacteria were mixed with 100 µl of the various phage dilutions, and incubated at 37°C in a water bath for 10 min. 2.5 ml of warm BBL agar was then added to the transfection mixture, the contents mixed, and poured evenly over the surface of a cool BBL agar petri dish. The dish was inverted and incubated at 37°C until phage plaques formed. For the gene libraries, 6.8 x 10^4 pfu were adsorbed to 5 ml of competent cells.

LIGATION CONDITIONS

Conditions for the ligation of the constituents of the gene library are described in the methods. For subcloning cloning fragments with cohesive termini from λ clones into plasmids, 200 ng of linearised pUC 8 (Viera and Messing 1982) were mixed with a three fold molar excess of the fragment to be cloned. If the volume of the DNA solution exceeded 8 µl, the DNA mixture was alcohol precipitated and redissolved in 8 µl of T.E. buffer. 1 µl of 10X ligation buffer was then added, followed by ligase to a concentration of 1 Weiss unit/µg of DNA. Solutions were mixed by vortexing briefly, and incubated at 15°C for 14 h. After this, a small aliquot was checked on an agarose gel to monitor the extent of ligation. The remainder was alcohol precipitated, redissolved in 0.1 M calcium chloride, and used to transform competent bacteria.

10X ligation buffer

0.66 M Tris - HCl pH 7.5

50 mM $MgCl_2$

50 mM DTT

10 mM ATP

DNA SEQUENCING

This sequencing method was based on that of Maxam and Gilbert (1977), as modified by Maat and Smith (1978) and Sief et al (1980).

(A) First restriction

(B) Phosphatase treatment

(C) Kinase labelling for protruding 5′ termini

(D) Second restriction step

(E) Preparative gel

(F) Purification of labelled fragments from the preparative gel

(G) Dideoxy nucleotide reactions

(H) Preparation and running of sequencing gels

(I) Autoradiography

(J) Buffers and solutions for DNA sequencing

(A) First restriction step

10 - 12 µg of DNA were digested thoroughly with a restriction enzyme which gave 5′ protruding termini, in a total volume of 100 µl. The DNA was alcohol precipitated and dried.

(B) Phosphatase treatment

The restricted and ethanol precipitated DNA was resuspended in

193 µl of d H_2O, 5 µl of 2 M Tris - HCl pH 8.0 added, followed by 7 µl of 0.6 U/µl alkaline phosphatase, and the reaction incubated at 37°C for 45 min. The reaction was terminated by the addition of 2 µl of 100 mM EDTA pH 8.0, and the enzyme extracted by the addition of 200 µl buffer saturated phenol, vortexing, removing the lower phenol phase, and repeating the phenol extraction. The phenol was removed and the DNA alcohol precipitated as before. After resuspending the resultant pellet in 250 µl of 0.3 M sodium acetate by vortexing, a further two alcohol precipitations were carried out and the pellet vacuum dried.

(C) Kinase labelling for protruding 5′ termini

The phosphatased and ethanol precipitated DNA was resuspended in 3 µl of H_2O, 2.5 µl of kinase buffer followed by 1 µl of 17.5 mM spermidine, 3.5 µl of 4µM ATP, 13 µl ^{32}P ATP (130 µCi), and 2 µl of polynucleotide kinase (20U) were added and incubated at 37°C for 45 min. The reaction was stopped by the addition of 200 µl 2.5 M ammonium acetate. 1 µl of a 1 mg/ml solution of tRNA followed by 750 µl of ethanol were added, and the DNA recovered by performing 3 consecutive alcohol precipitations.

(D) Second restriction step

The purpose of this was to generate fragments with one labelled end for sequencing. The labelled, precipitated and dried DNA was resuspended in 82 µl of H_2O by vortexing. 4 µl of this was taken and kept as a control sample for the preparative gel. 20 µl of 5X assay buffer and 40U of restriction enzyme were added , and the DNA incubated and precipitated as before.

(E) Preparative gel

20 ml of 40 % v/v acrylamide stock solution (19:1 acrylamide / bisacrylamide) and 16 ml of 10X Tris borate EDTA were made up to 160 ml with dH O and degassed. 1.1 ml of ammonium persulphate followed by 100 µl of TEMED were added to the gel solution, mixed, and poured into a 18 x 40 cm polyacrylamide gel apparatus (1.5 mm thick), using a syringe. After the gel had set, it was pre electrophoresed for 30 min at 500v. The precipitated, dried , DNA was resuspended in 7µl T.E. buffer and 5 µl glycerol dyes added. 3 µl of dyes were added to the sample from step (C), mixed, and both samples loaded in separate wells. Electrophoresis at 500 V was continued until the bromophenol blue was 3 cm from the bottom of the gel. The gel plates were then separated, and the gel covered with cling film. Pieces of filter paper with distinctive patterns of radioactive ink were taped to the cling film, and the gel autoradiographed at room temperature for 30 - 60 min using an intensifying screen and flashed film.

(F) Preparation of labelled fragments from the gel

The autoradiograph from step (E) was examined and the required fragments identified. Gel fragments corresponding to the autoradiographic images were excised from the gel and trimmed so that the minimum amount of gel was present. Blue 1 ml disposable pipette tips were carefully sealed over a bunsen flame and plugged with a piece of siliconised glass wool. The gel slice was placed in the sealed tip and mashed to a paste with a small rod. 600 µl of gel elution buffer were then added to each tip, followed by 6 µl tRNA (1 mg/ml), the tip sealed with two layers of parafilm , and heated at 37°C

for at least 2 h. The parafilm was removed, the sealed ends cut off, and the tips spun in Corex tubes for 2 min. at 3,000 rpm using a Sorvall HB4 rotor. The eluate was collected, 200 µl of gel elution buffer added to the tip, and centrifuged as before. The sample eluates were pooled, and alcohol precipitated The pellet was resuspended in 300 µl of 0.3 M sodium acetate and alcohol precipitated twice. The pellet was then washed with 1 ml of 90 % ethanol and vacuum dried.

(G) Dideoxynucleotide reactions

The vacuum dried and labelled fragment was resuspended in 20 µl H_2O, and 10 µl of 5 x Sief buffer, 8 µl of 5U/µl endonuclease free Dnase 1 added, and kept on ice. Narrow plastic tubes were marked BG, BA, BT, BC, FG, FA, FT, FC, corresponding to the forwards and backwards reactions, and 1.1µl of the appropriate nucleotide triphosphate dispensed into each tube. 4.5µl of the DNA / enzyme mixture was then added to each tube, centrifuged briefly to mix the solution, and incubated at 37°C for 30 min.

The reactions were frozen, 1 µl of 0.1 M EDTA pH 8.0 added to each of the backwards (B) mixtures, centrifuged briefly, and the whole mixture transferred into the appropriate forwards tube ; e.g. EDTA was added to BG and the contents transferred to the FG tube. Small holes were made in the caps, and the tubes frozen at -80°C for 30 min, after which the samples were lyophilised for 1 h. 10µl of formamide dyes were added to each tube, which was then heated at 90°C for 5 min, placed in scintillation vials and counted, and then frozen at - 20°C until the

sequencing gels were ready to be loaded.

(H) Preparation and electrophoresis of sequencing gels

6 %, thin polyacrylamide gels were used for sequencing. Glass plates were cleaned extensively with decon, water, and ethanol until a high polish was obtained. 2 glass plates separated by 0.5 mm perspex spacers were taped together, the gel solution prepared, and the gel poured free of bubbles by holding the plates almost horizontal and pouring the solution down one edge. A 14 tooth comb was placed in the top and the gel allowed to set. The top of the gel was then flooded with buffer and the comb carefully removed. The tape sealing the bottom of the gel plates was cut away and the plates clamped in a vertical electrophoresis apparatus. The wells were cleared of debris by flooding with a syringe.

Gels were pre electrophoresed at 1500 v for 1 h before sequencing samples were loaded. The gels were switched off for as short a time as possible during loading of samples.Samples to be loaded were heated at 90°C for 1 min, cooled rapidly in ice, and 2.5 µl of the separate nucleotide reactions loaded in 4 separate tracks using a drawn out capillary attached to a mouth tube. The first loading was run until the xylene cyanol in the sample had migrated to the bottom of the gel. The second loading was run until the bromophenol blue was at the bottom, and the third loading until the bromophenol blue was 3 cm from the bottom.

(I) Autoradiography

The gel plates were removed from the apparatus and placed on the bench with the siliconised plate uppermost. The tape holding the plates together was cut, and the perspex spacers removed. The plates were then levered apart, the gel adhering to the larger, non siliconised plate. The gel was covered with cling film and autoradiographed at -80°C using a preflashed film and an intensifying screen. Exposure times were calculated using the formula

$$x = 5 \times 10^5 / y$$

where x = exposure time in h.

y = cpm per gel track as determined earlier.

Buffers for DNA sequencing

10 x Kinase buffer

500 mM Tris - HCl pH 7.6, 100 mM $MgCl_2$, 50 mM DTT, 1 mM EDTA.

10 x Tris Borate buffer / litre, pH 8.3

Tris base 108 g, Boric acid 55 g, EDTA $2H_2O$ 10.3 g.

Formamide dyes

80 % v/v formamide, 10 mM NaOH, 1 mM EDTA, 0.1 % xylene cyanol, 0.1 % bromophenol blue

5 x Sief buffer

33 mM Tris - HCl pH 7.5, 33 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl.

6 % polyacrylamide gel

Urea 30 g, 40 % acrylamide / bisacrylamide 9 ml, 10 x Tris borate EDTA 6 ml, water to final volume of 60 ml. Filter through a millipore

filtration apparatus, degas, add 400 µl 10 % ammonium persulphate, 20 µl TEMED, and pour.

Nucleotide mixes

BG ----- 1 mM dd GTP

BA ----- 1 mM dd ATP

BT ----- 1 mM dd TTP

BC ----- 1 mM dd CTP

FG ----- 1 mM dd GTP, 200 µM d ATP, 200 µM d TTP, 200 µM d CTP.

FA ----- 1 mM dd ATP, 200 µM d GTP, 200 µM d TTP, 200 µM d CTP.

FT ----- 1 mM dd TTP, 200 µM d GTP, 200 µM d ATP, 200 µM d CTP.

FC ----- 1 mM dd CTP, 200 µM d GTP, 200 µM d ATP, 200 µM d TTP.

BIOCHEMICAL TECHNIQUES

Glassware and plasticware

All glassware was throughly washed in detergent, rinsed with several changes of distilled water, and autoclaved before use. Items of glassware and plasticware e.g. eppendorf tubes, pipette tips, Corex centrifuge tubes etc, were siliconised by first degreasing in chloroform, and then immersing in Repelcote , a solution of dichlorodimethylsilane. They were then baked at 100°C for 2 h and autoclaved before use.

Restriction endonuclease digestion

The following 4 buffers were used for restriction enzyme cleavage. A 10X solution was usually made, and stored at - 20°C in 1 ml aliquots.

Low salt	Medium salt	High salt
1 mM Tris HCl pH7.5	10 mM Tris HCl pH7.5	50 mM Tris HCl pH7.5
10 mM $MgCl_2$ pH7.5	10 mM $MgCl_2$ pH7.5	10 mM $MgCl_2$ pH7.5
1 mM DTT	1 mM DTT	1 mM DTT
	50 mM NaCl	100 mM NaCl

Sma1 buffer

10 mM Tris HCl pH 8.0

10 mM $MgCl_2$

1 mM DTT

20 mM KCl

1 µg of DNA was usually digested in a final volume of 30 µl, with 5 U of restriction enzyme for 2 h. Where inactivation of the enzyme was desired, the restriction digest was made 10 mM with respect to EDTA. If removal of the enzyme was desired, e.g. prior to ligation, the enzyme was extracted into phenol and the phenol extracted into ether. Where dissociation of cohesive ends was desired, the digest was heated to 70°C for 5 min, and rapidly cooled on ice. Prior to agarose gel electrophoresis, samples were made 1/3 v/v with respect to agarose beads.

Phenol extraction

A mixture of 25 : 24 : 1 of phenol : chloroform : isoamylalcohol was usually used to deprotenise nucleic acids and to inactivate enzymes involved in nucleic acid manipulations. An equal volume of the phenol mixture was added to the solution, vortexed, and the lower phenol phase removed. A second phenol extraction was then performed. After removal of the phenol phase, residual phenol was

removed by extracting twice with water saturated ether (the phenol used in the mix was pre equilibrated with 10 mM Tris - HCl pH 7.5, 10 mM EDTA). After the ether extraction, the DNA was usually precipitated with ethanol and vacuum dried. Vortexing was avoided when the DNA size exceeded 20 Kb.

Alcohol precipitation of nucleic acids

The volume of the nucleic acid solution was estimated, and if necessary, made 0.25 M with respect to sodium acetate. After mixing, exactly 2 volumes of ethanol at - 20°C were added, mixed, and stored at - 80°C for 15 min. If the DNA was of a small size (1Kb), or present in low concentrations (0.1 µg/ml), the storage at - 80°C was extended up to 5 h. The precipitated DNA was recovered by centrifugation for 10 min at full speed in a microcentrifuge for small samples, or for 20 min at 12,000 rpm in a Sorvall HB4 rotor for larger samples. The alcohol was carefully decanted, pellet washed with 80 % ethanol and dried under vacuum. The recovered DNA was usually redissolved in a small amount of T.E.buffer.

Concentration of nucleic acids by extraction with butanol

As recovery of DNA from very dilute solutions by alcohol precipitation is poor, this technique was used to concentrate nucleic acid solutions to a level where they were easily precipitated. The solution was extracted with 2 - butanol which partitions water into the organic phase, excluding DNA or solutes. An equal volumeof 2 - butanol was added to the DNA solution and well mixed. After a brief centrifugation to separate the phases, the upper butanol phase was removed and the extraction repeated with fresh 2 - butanol until the

the volume of the solution had been reduced to a point where the DNA could be easily recovered by alcohol precipitation.

Spectrophotometric determination of nucleic acids

To determine the amount of DNA or RNA, readings were taken of a 1 ml sample in a Pye Unicam SP8 - 150 spectrophotometer at 260 nm and 280 nm. The reading at 260 nm allowed calculation of the amount of nucleic acid in the sample _ an O.D. of 1 corresponding to 50 µg/ml of double stranded DNA, 40 µg/ml of single stranded DNA and RNA, and 20 µg/ml for oligonucleotides. Readings at 260 nm and 280 nm were taken, and the sample purified to give an A_{260} / A_{280} ratio for double stranded DNA > 1.8.

Preparation of dialysis tubing

Dialysis tubing (Medicell) was cut into pieces of convenient length and boiled for 10 min in a large volume of 1 mM EDTA. It was then rinsed throughly in distilled water and stored under water at + 4°C . Before use, the tubing was again throughly washed inside and out with distilled water.

RESULTS

Preparation of the annealed 'arms' of λ L47 DNA

The preparation of the arms is described in the methods. Fig 1 shows the fragments produced upon digestion of the vector with Bam H1. When λ L47 DNA was annealed via its cos sites, cleaved with Bam H1, and centrifuged through a glycerol gradient, the annealed arms were separated from the 6.6 kb replaceable fragment. The structure of the resultant arms is shown in Fig 2. Fig 3 shows an agarose gel analysis of the annealing of λ L47 DNA. The annealed and ligated DNA (tracks 4 & 5) was of significantly higher mol. wt. than either the native λ L47 DNA (track 3), or the λNM 258 marker (track 2), and gave the 34 kb fragment consistent with the formation of annealed DNA upon cleavage with Bam H1 (track 6). Appreciable quantities of both the left and right arms were also present. After centrifugation of the annealed, ligated and cleaved DNA through a glycerol gradient, fractions were collected (Fig 4) and analysed on an agarose gel (Fig 5). Clear separation between the 6.6 kb replaceable fragment, the 10.4 kb right arm, 24.0 kb left arm and the 34.0 kb annealed vector arms was clearly seen. Tracks 7 to 10 (Fig 5), contained 34 kb annealed arms which were only minimally contaminated with the 23 kb left arm. No visible trace of the 6.6 kb replaceable fragment was seen.

Structure of molecules formed upon ligation of the annealed arms with and without foreign DNA

Fig 6 shows the structure of one of the forms produced when the annealed arms of λ L47 were ligated at their Bam H1 t_{λ}^{e}rmini. Cleavage at the cos sites of these molecules by the *in vitro* packaging system led to the formation of two sorts of structures 1) Cleavage

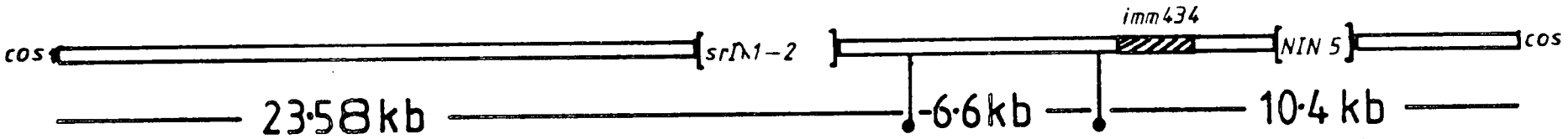

cos
srIλ1–2
imm434
NIN 5
cos
23·58kb
6·6kb
10·4 kb

Fig 1

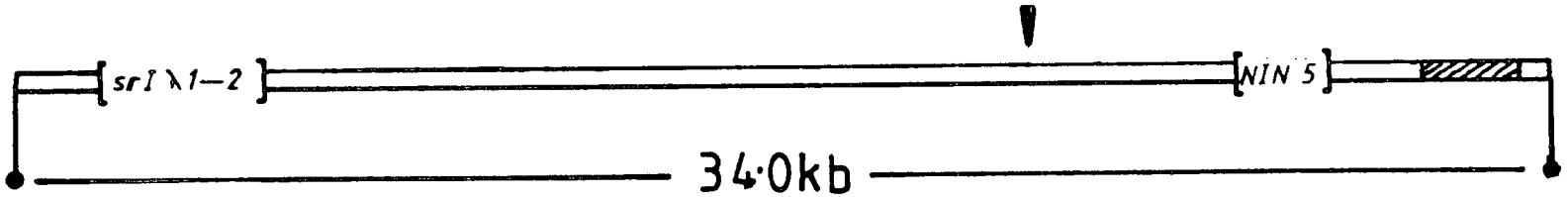

srI λ1–2
NIN 5
34·0kb

Fig 2

Fig 3

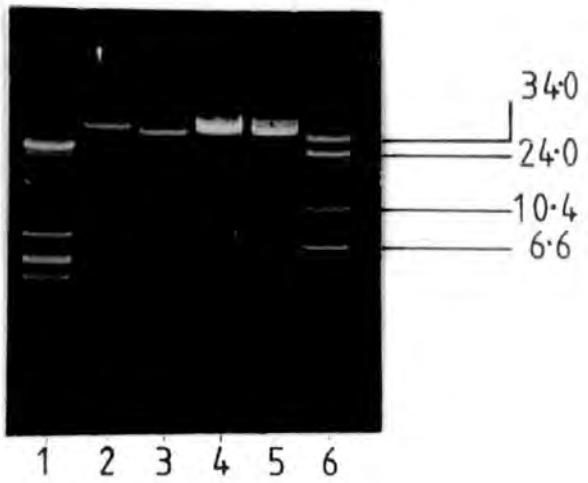


Fig 4

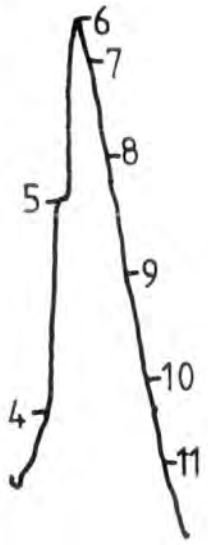
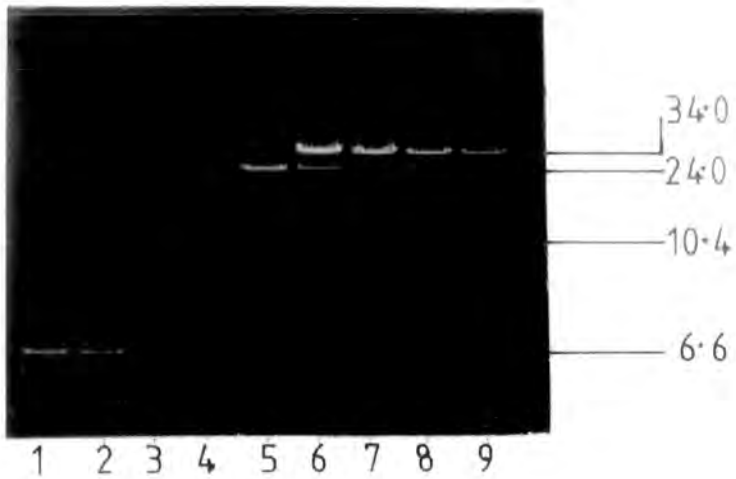


Fig 5



occuring at cos sites 1 and 3 leading to a 54.8 kb molecule too large to be packaged; and 2) cleavage occuring at cos sites 2 and 3 leading to a 20.8 kb molecule too small to be packaged into phage particles. The ligation of foreign DNA at the Bam H1 sites of the arms led to the structure shown in Fig 7. Cleavage at cos sites 1 and 2 Fig 7) led to the formation of the packageable, recombinant structure shown in Fig 8, with an overall size of 53 kb, the upper size limit for packaging into phage λ.

Preparation and characterisation of an in vitro packaging extract.

In vitro packaging extracts were prepared from induced cultures of the λ lysogens BHB 2688 and BHB 2690 as detailed in the methods section. Fig 9 shows the growth curves of the two lysogens, before, during, and after induction. Successful induction of the lysogens was seen by the rapid cessation of bacterial growth after the application of the induction stimulus. The efficency with which the packaging extracts formed mature, viable phage particles using various DNA substrates is shown in Table 1. Packaging due to endogenous phage DNA present in the extracts was nil (line 1), and DNA molecules which did not have the cos sites of bacteriophage λ were not packaged (lines 2 and 3). λ DNA molecules of the incorrect size and orientation were packaged at very low levels (line 4), while native L47 DNA was packaged at high efficencies (line 5), this efficency being reduced if the λ L47 DNA was cleaved with Bam H1 before ligation and packaging.

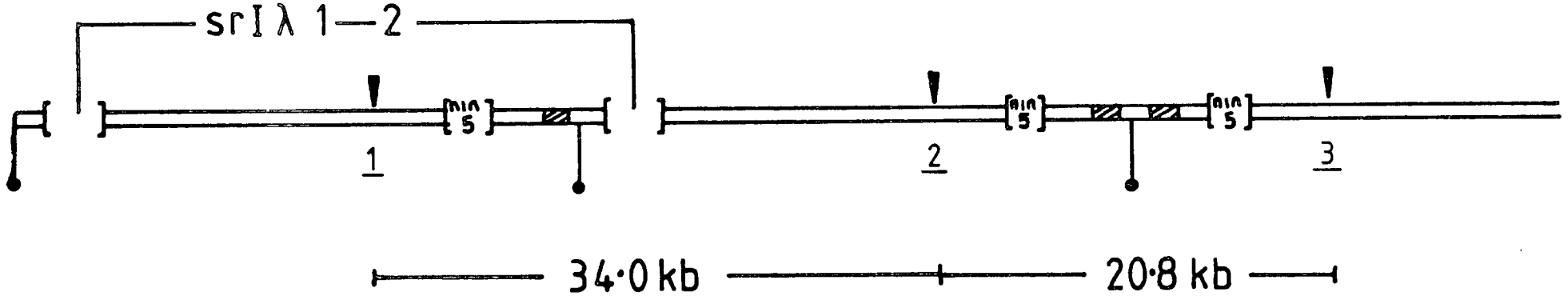

srI λ 1—2
NIN 5
1
2
3
34·0 kb
20·8 kb

Fig 6

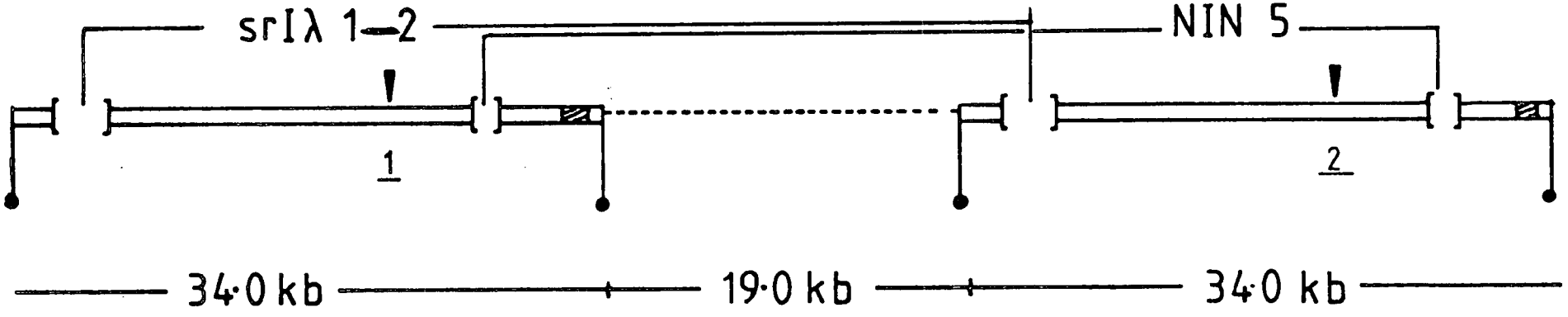

srI λ 1—2
NIN 5
1
2
34·0 kb
19·0 kb
34·0 kb

Fig 7

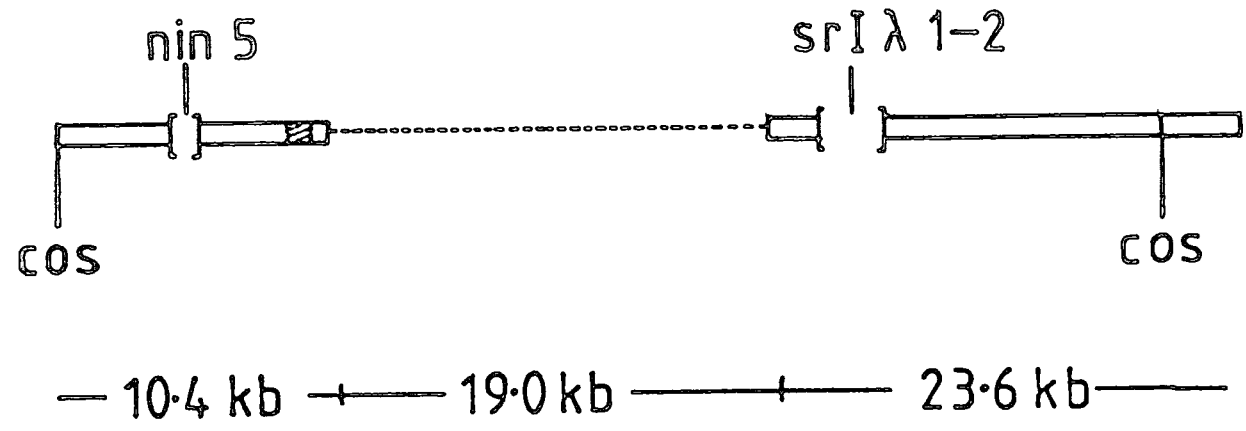


Fig 8

Fig 8

Structure of a recombinant phage molecule, with an insert of 19 Kb, after packaging into a phage particle.

If cos sites 1 and 2 of the structure shown in Fig 7 are cleaved, this molecule is produced.

The order of the fragments is right arm --- foreign DNA --- left arm. The single stranded cos termini produced are shown.

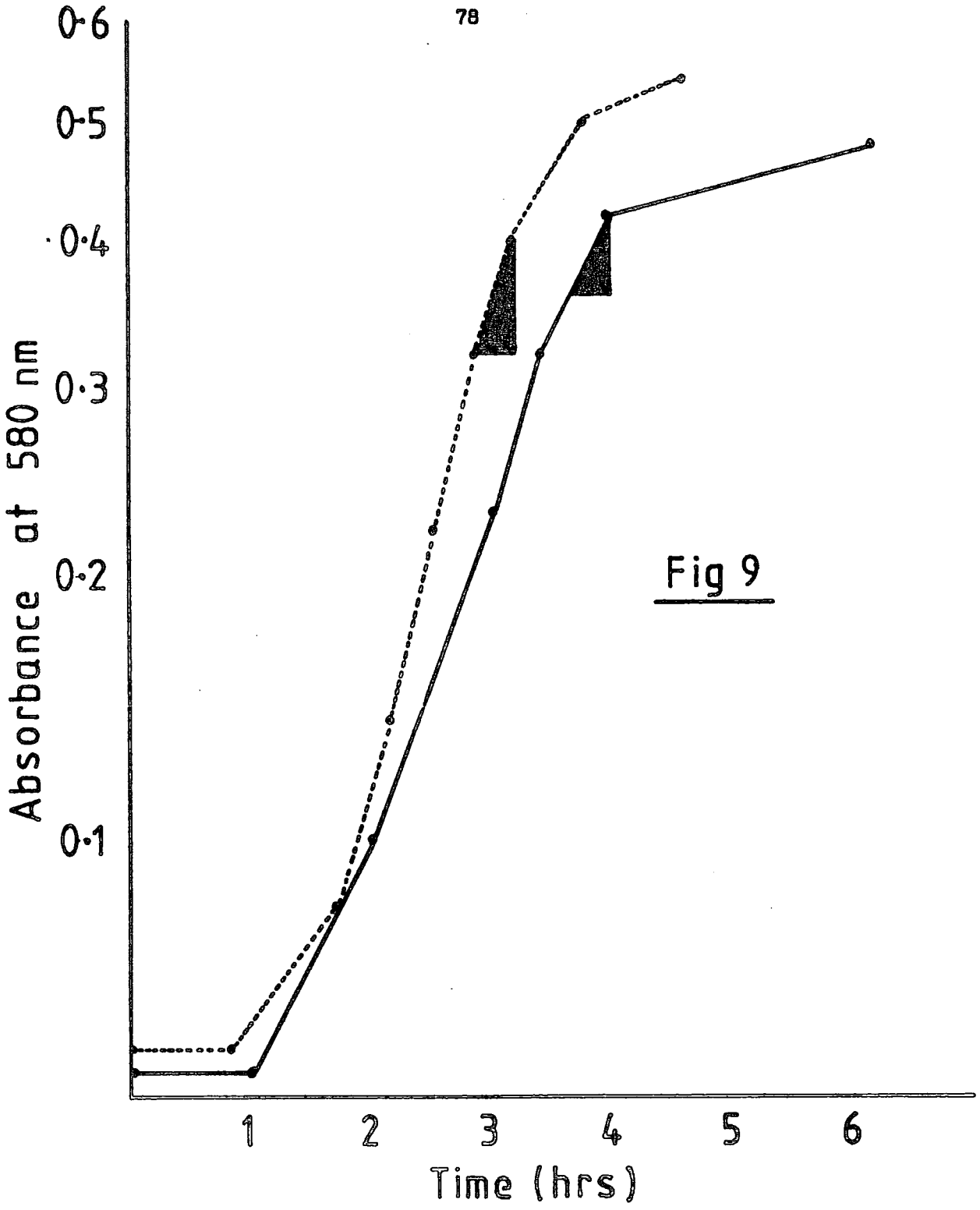


Fig 9

Fig 9

Growth curve of the λ lysogens used to prepare the *in vitro* packaging extracts.

------= BHB 2688

——— = BHB 2690

▬ = duration of the heat stimulus required for induction.

Fig 10

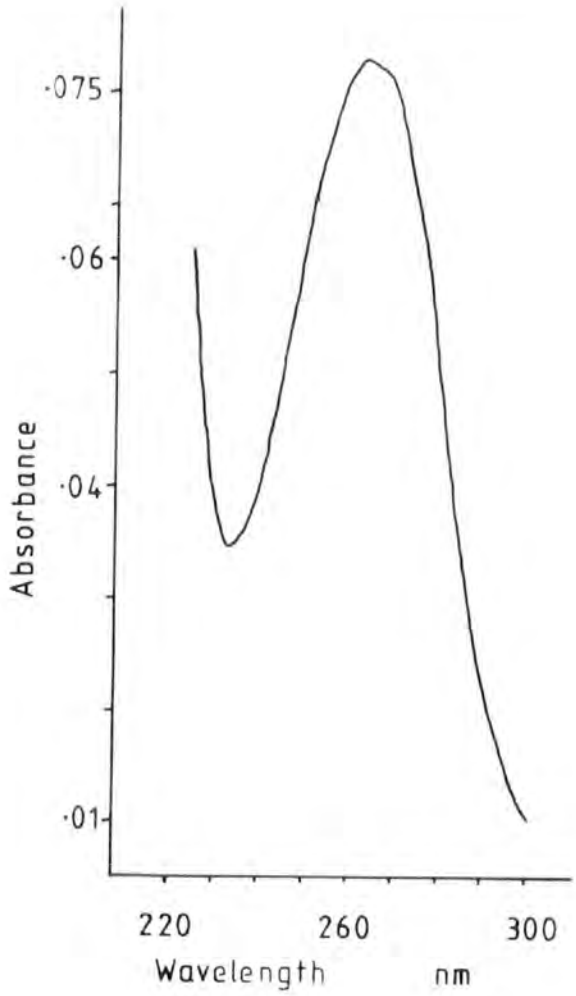
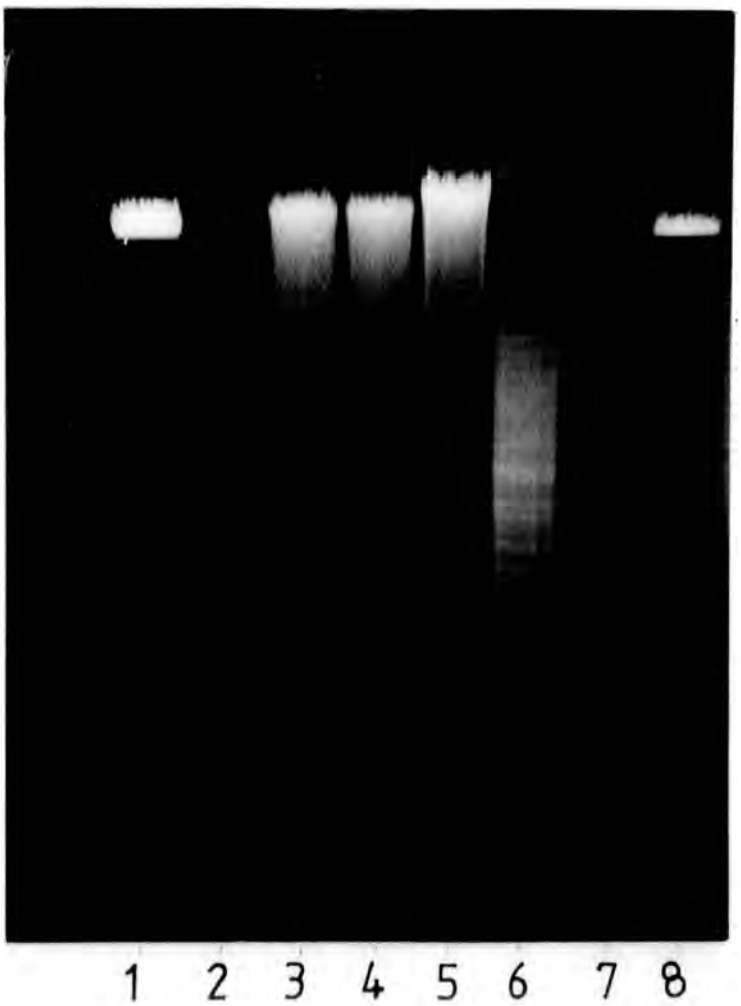


Fig 11



Characterisation of pea DNA.

Pea DNA was prepared as described in the methods section. Fig 10 shows the u.v. absorbance of a sample of this DNA. A classical nucleic acid spectrum, with a peak at 260 nm and a trough at 230 nm was obtained. The O.D. 260 / 280 ratio was 1.97. Fig 11 shows the electrophoresis of DNA through a 0.3 % agarose gel, using intact λ DNA as a high mol. wt. marker. Electrophoresis was at 30 v for 15 h . The majority of pea DNA was greater than 50 kb in size (tracks 3 and 4), and did not show the presence of nuclease activity (track 5). DNA was easily cleaved by the enzymes Bam HI (track 6), and Sau 3A (track 7) indicating its suitability for cloning purposes.

Isolation of 15 - 20 kb partially cleaved pea DNA fractions.

Conditions for the partial cleavage of pea DNA were established as described in the methods. 0.5 µg aliquots of the various restrictions were analysed on an agarose gel (Fig 12) to determine the extent of digestion. A range of digestions were achieved, ranging from incomplete (track 9) to complete (track 2). The maximum amount of DNA in the correct size range appeared to be in track 7, however, the maximum number of molecules were in the digestion (track 8) which used half the enzyme giving the maximum fluorescence in the 15 - 20 kb size range (Seed et al 1982). Therefore, the digestion shown in track 8, which used 0.0156 U of Sau 3A per µg of pea DNA was chosen as the optimal digestion, and 4 large scale digestions of pea DNA were performed, using 100 µg of DNA per digestion and 0.5, 0, 1 and 2 times the amount of enzyme / DNA ratio as in track 8. An analysis of these digestions is shown in Fig 13. The digestions were pooled, and centrifuged through a glycerol

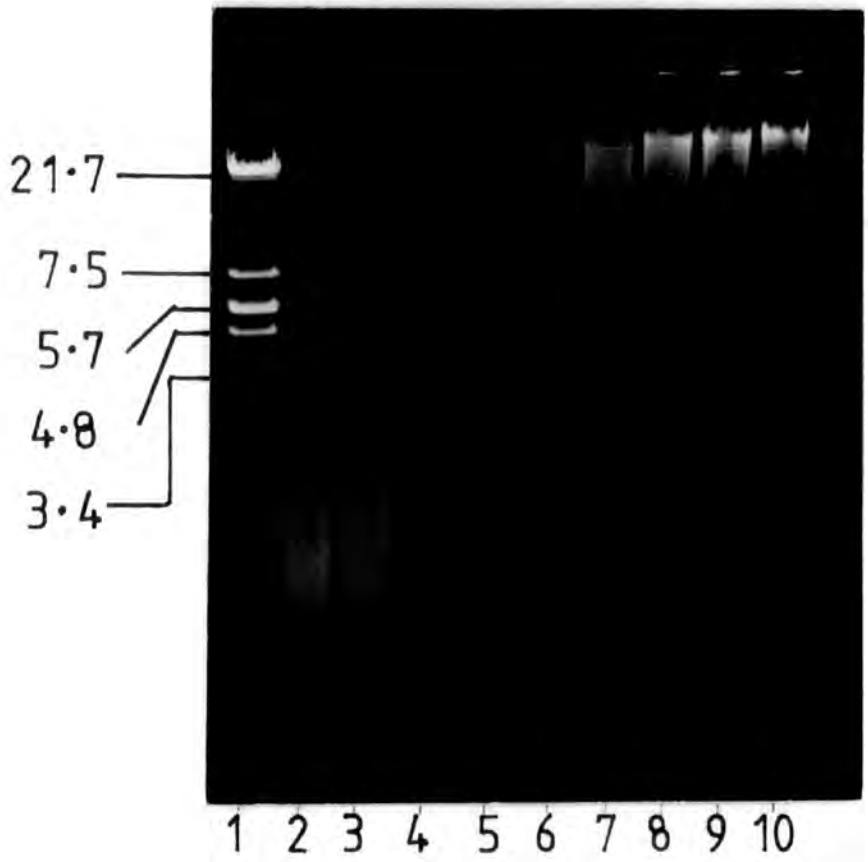
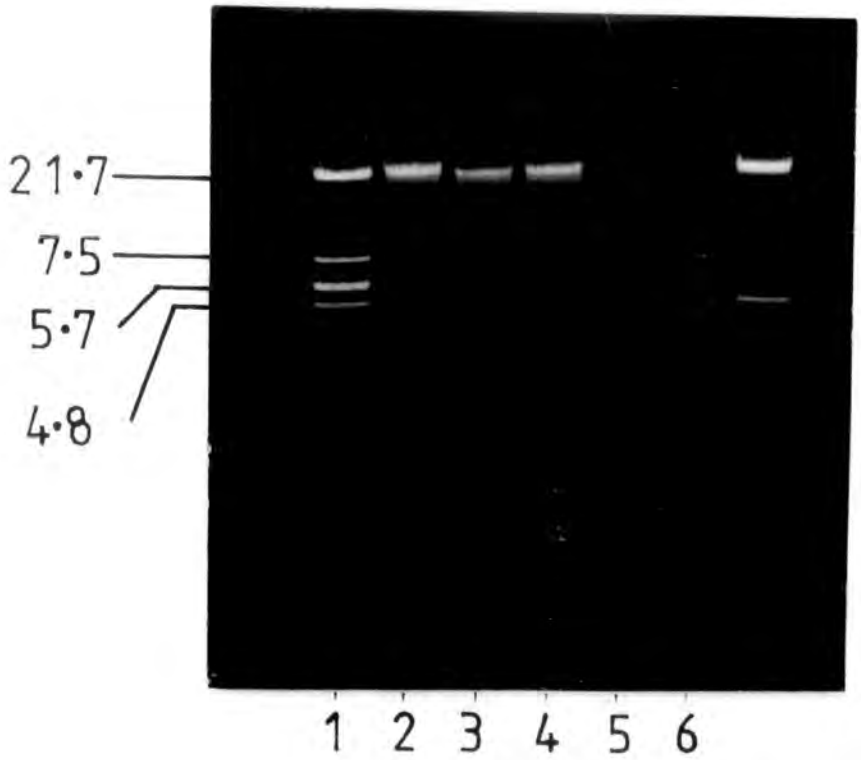


Fig 12

Fig 13



gradient (Fig 14). The analysis of the fractions obtained from the gradient (Fig 15) showed good separation of the various size classes, and indicated that fractions in tracks 3, 4, 8, and 9 were in the desired 15 - 20 kb size range. Fig (16) shows a further agarose gel analysis of fractions from Fig 15 tracks 3, 4, 8 and 9, confirming that they were all in the desired size range.

Construction of the pea gene library : optimisation of ligations and in vitro packagings.

Aliquots of Sau 3A cleaved DNA and λ L47 phage arms were checked as to their ability to ligate in order to see if the cohesive termini of the vector and pea DNA molecules were still present, and if all inhibitors of ligation had been removed. An agarose gel analysis (Fig 17) showed that the 1:1 mixture of L47 arms and Sau 3A cleaved pea DNA (track 3), L47 arms alone (track 4), and Sau 3A cleaved pea DNA alone (track 5) all showed evidence of ligation, as judged by the formation, in all three tracks, of DNA molecules larger than either the unligated L47 arms (track 1), or unligated Sau 3A cleaved pea DNA (track 2). The ratio of L47 arms to Sau 3A cleaved pea DNA which would produce the maximum number of recombinants upon ligation and *in vitro* packaging as judged by the efficency of plaque formation on the selective P2 lysogen *E. coli* was investigated - Table 2. A 1 : 1 ratio of phage arms to cleaved pea DNA was found to be optimal (line 2), while a ratio of 3 : 1 produced the least number of recombinants (line 4). Unligated vector arms produced no plaques at all (line 6), but arms ligated in the absence of Sau 3A cleaved pea DNA produced a small number of plaques on the non selective host strain *E. coli* 5K (line 5). The composition of a ligation reaction for a

Fig 14

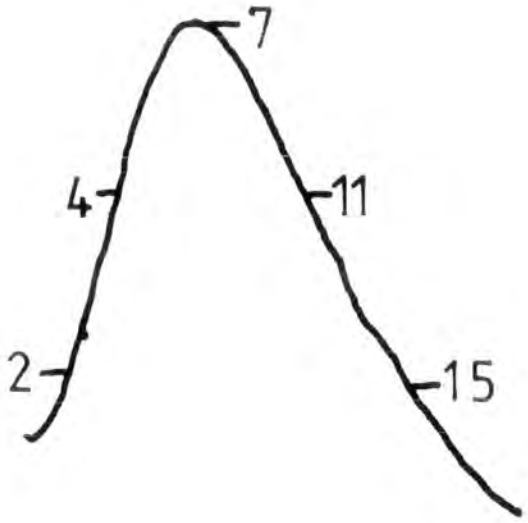


Fig 15

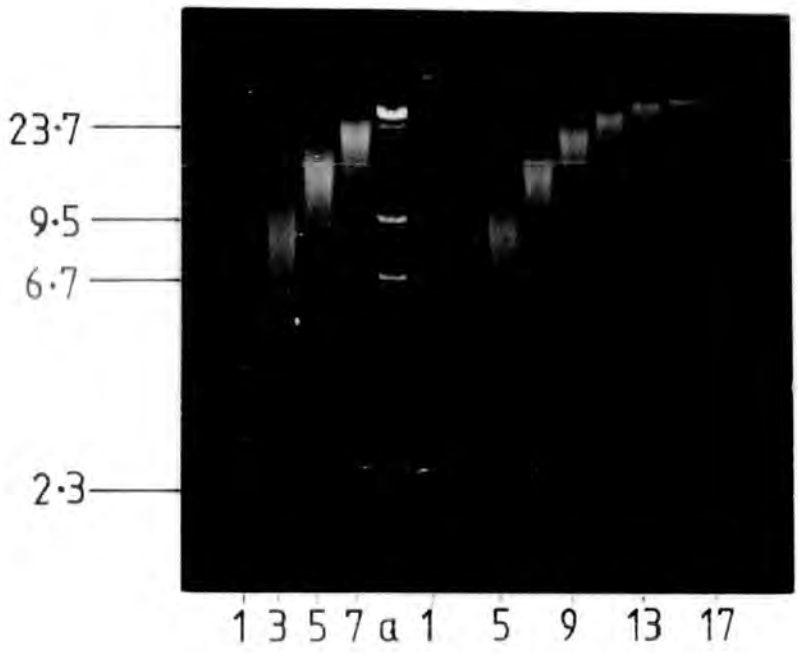
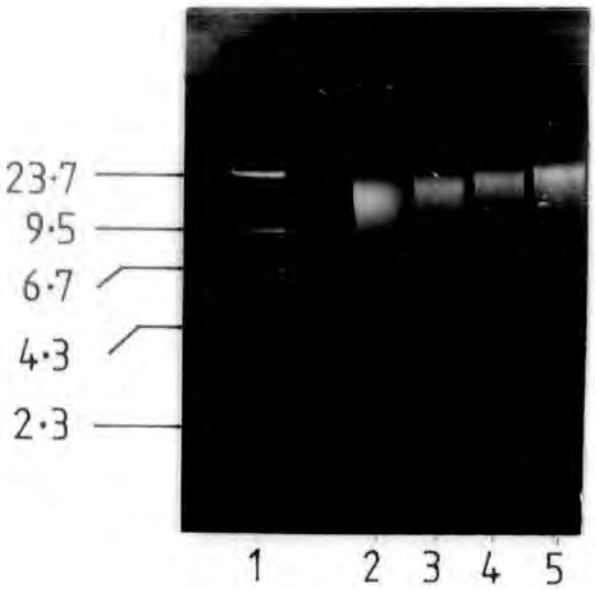


Fig 16



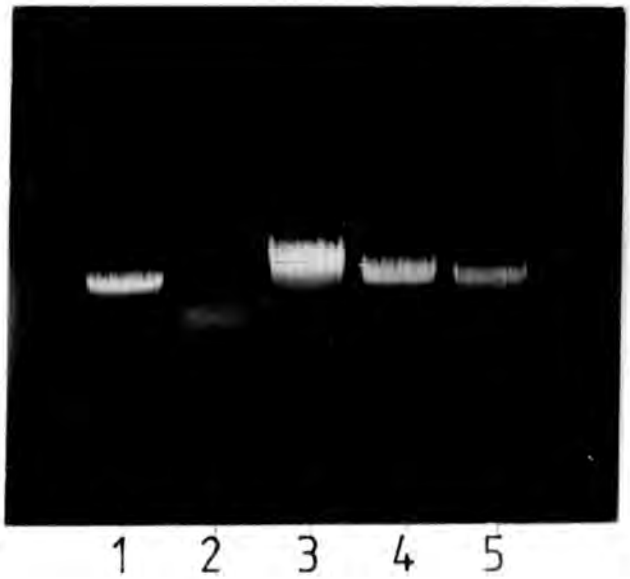


Fig 17

Fig 18

Fig 17

Analysis of the test ligation of λ L47 arms and Sau 3A cleaved, size fractionated pea DNA.

1 µg of DNA was run in all tracks.

1) λ L47 arms DNA (41 kb)
2) 15 - 20 kb Sau 3A cleaved pea DNA
3) 1 : 1 mixture of λ L47 arms and Sau 3A cleaved pea DNA ligated at 25 °C for 2 h.
4) λ L47 arms ligated .
5) Sau 3A cleaved pea DNA ligated.

Fig 18

Analysis of an aliquot of the ligated mixture used to construct the gene library.

Track 1 = λ L47 DNA (41 kb)

Track 2 = 1 µg of the ligated mixture.

Constituents of the ligation mixture are shown in Table 3.

'mini - library' is shown in Table 3. Ligation was at 12°C for 48 h, after which a small aliquot was analysed to monitor the extent of ligation (Fig 18). The gene library ligation mix (track 2) was of appreciably higher mol. wt. than the L47 arms in track 1, showing that successful ligation had been acheived.

Separation of a nick translation reaction on a Sephadex column

The separation of ^{32}P labelled DNA from unincorporated nucleotides is shown in Fig 19. 48 % of the total radioactivity was found in fractions 2 and 3 (peak A).

Hybridisation of labelled pBR 322 DNA and legumin insert cDNA to various putative genomic legumin clones.

Phage clones isolated using the whole plasmid pDUB 6 as a probe were hybridised against pBR 322 DNA to check for contamination of the clones with pBR 322 vector sequences as follows. Aliquots of four putative genomic clones were diluted and spotted onto lawns of *E. coli* strain 5K as rescribed in the methods. After plaques were produced, nitrocellulose filter replicas were made and hybridised separately with ^{32}P labelled pBR 322 DNA (Fig 20), and the legumin cDNA insert from pDUB 6 (Fig 21). Clones 1, 2, and 3 hybridised to pBR 322 DNA , while clone number 4 did not (Fig 20), whereas Fig 21 showed that clone number 4 hybridised to the legumin cDNA insert while clones 1, 2, and 3 did not.

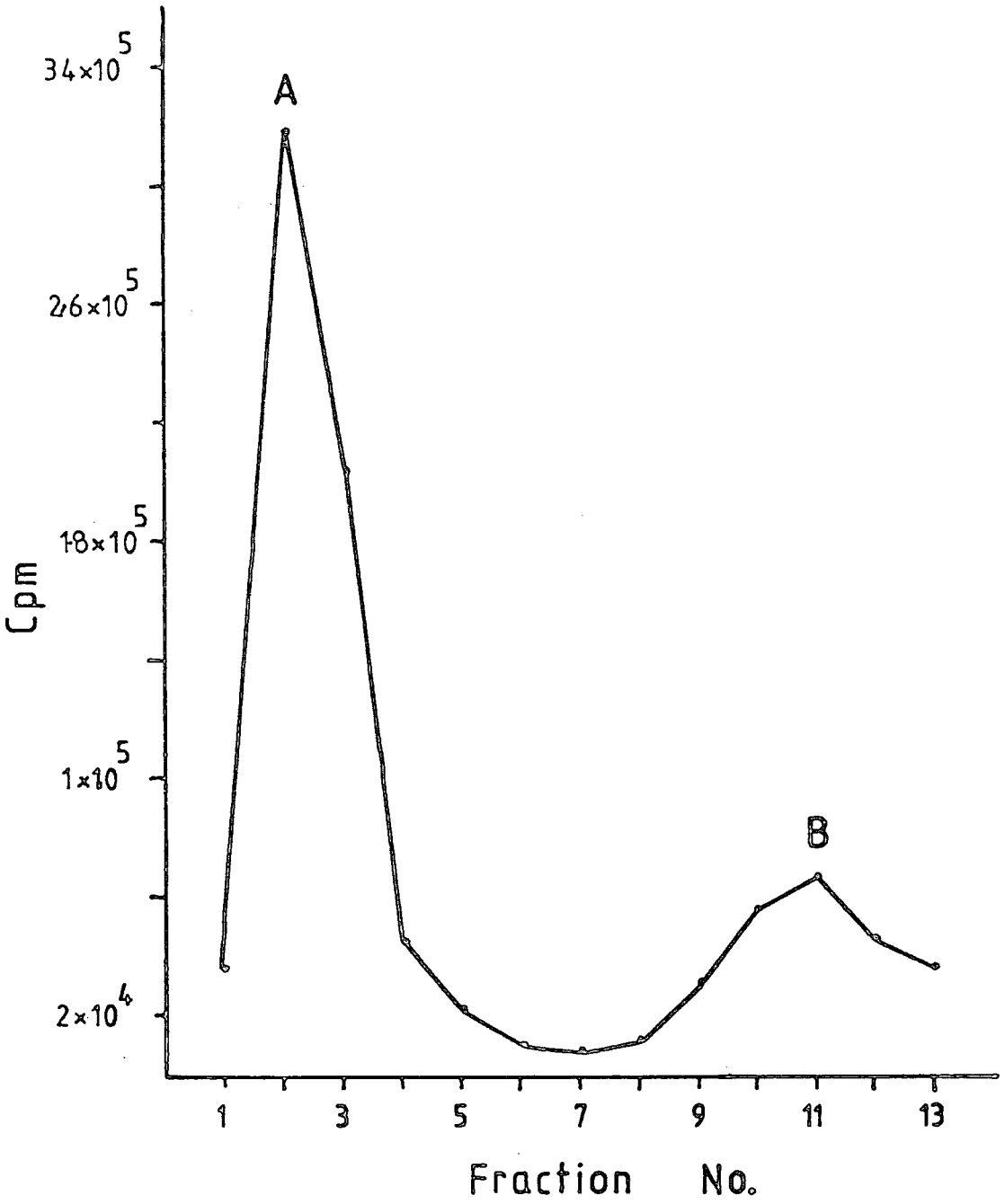


Fig 19

Fig 19

Separation of a nick translation reacton on a Sephadex G - 50 Superfine column.

Peak A = Labelled DNA

Peak B = Unincorporated labelled nucleotides

1 µl amounts of the column fractions were analysed in a scintillation counter to determine the radioactivity.

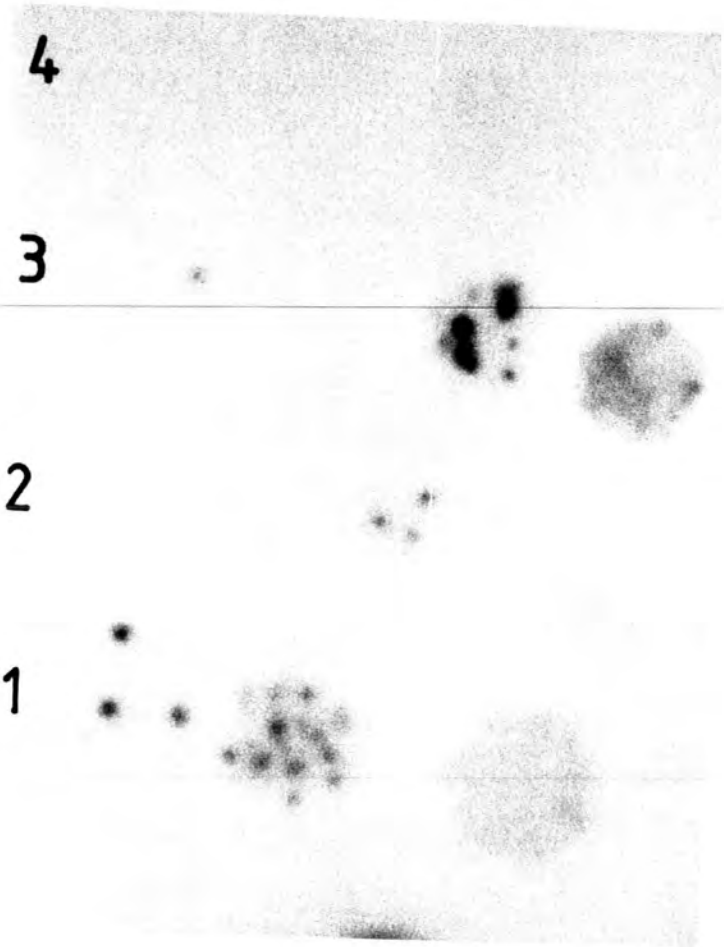


Fig 20

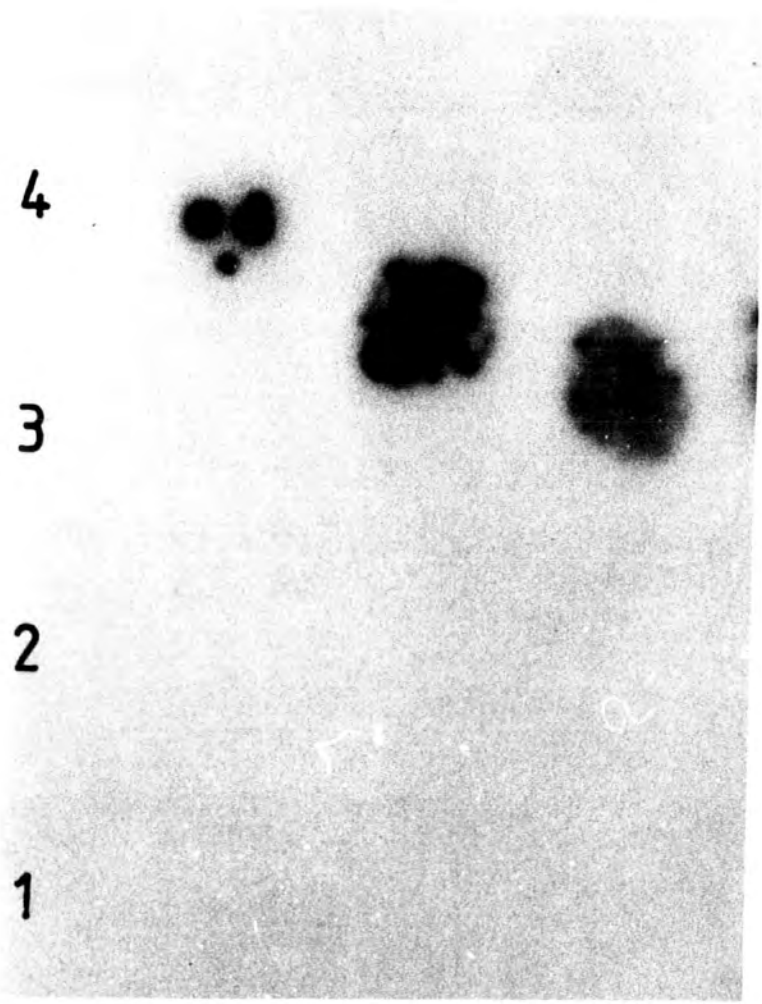


Fig 21

Purification of clone number 4 (λ LEG 3)

λ LEG 3 was purified by repeated cycles of plating out and hybridisation until greater than 95 % of the plaques on a plate were seen to hybridise to pDUB 6. Two types of signal were seen, one reflecting the normal plaque morphology and the other showing a head attached to a comet like tail (Fig 22). This presumably arose due to spreading of DNA from a phage plaque during the plaque lift procedures.

Partial characterisation of genomic ribosomal clones isolated from the library

In order to check whether the library contained sequences present at the same frequency as their occurence in the pea genome, the gene library was screened for genomic ribosomal DNA clones using the pea ribosomal clone pHA 1 (Cuellar 1982) as a probe. Screening of 10,000 phage resulted in the isolation of 85 hybridising clones. Phage minipreps were performed on two isolated clones as described in the methods and the DNA analysed by restriction enzyme cleavage and Southern hybridisation. Fig 23 shows the cleavage pattern of the clones with various restriction enzymes, and Fig 24 shows an autoradiograph of the same gel after hybridisation with ^{32}P labelled pHA 1. Table 4 shows a comparison of the sizes of the restriction enzyme fragments between ribosomal clones 1, 2 and pHA 1. It can be seen that they are almost identical.

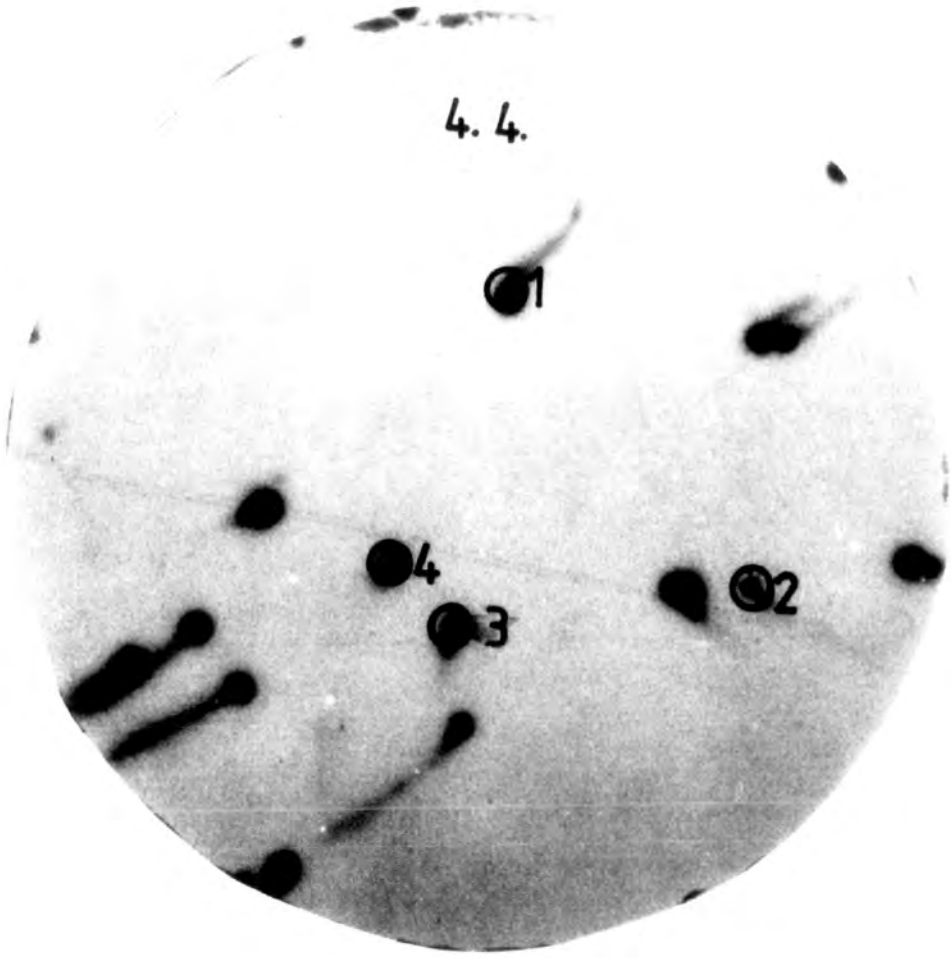


Fig 22

Fig 22

Purification of λ LEG3.

This shows the final stage in the plaque purification of LEG3, when > 95% of plaques on a plate hybridised to pDUB 6 under conditions of high stringency (0.1 x SSC, 65°C, 1 h). Two types of plaque morphology are evident.

Fig 23

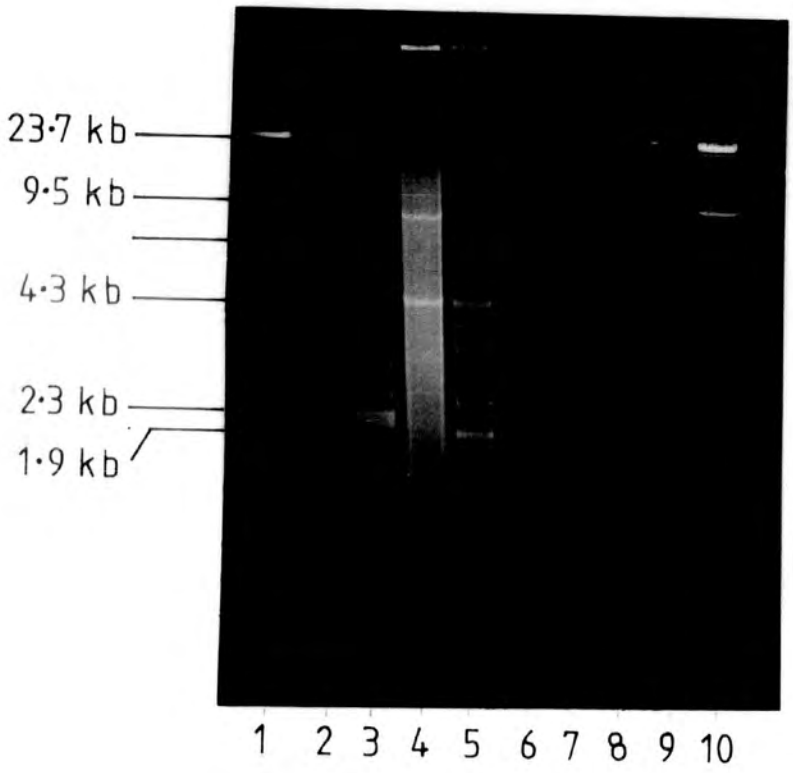
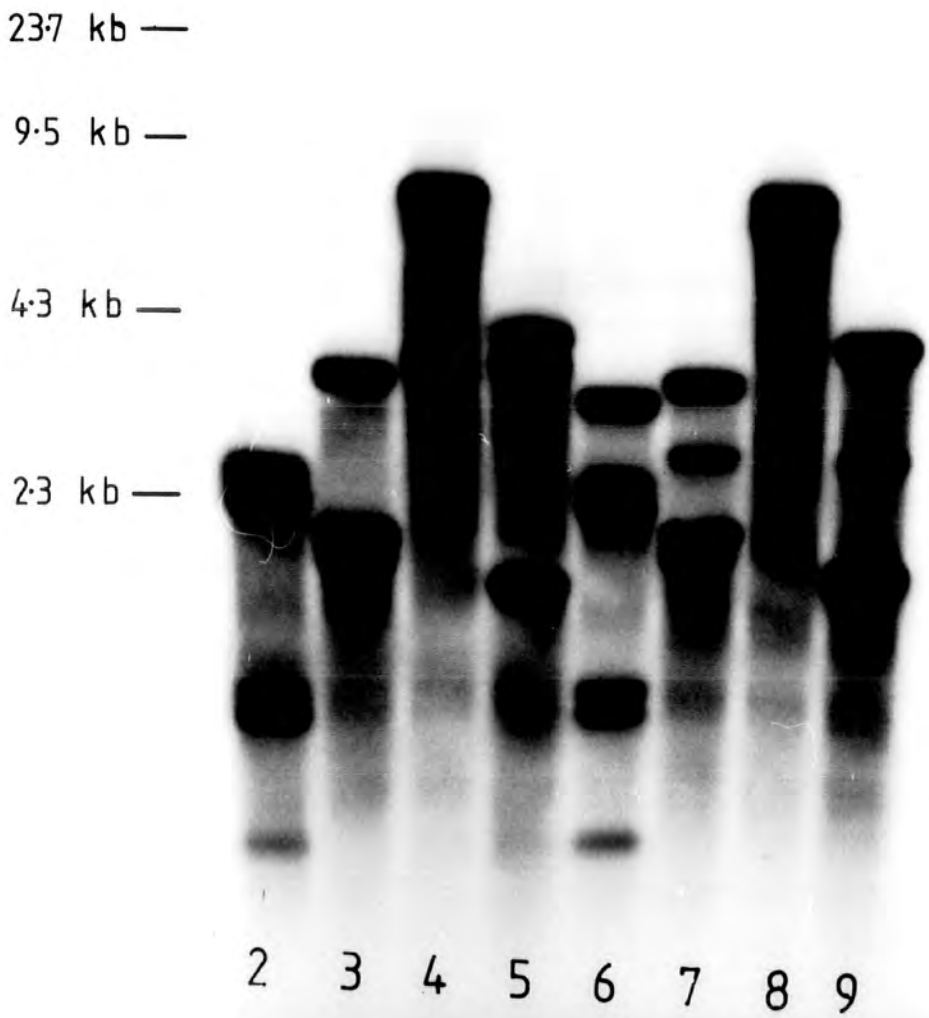

23·7 kb
9·5 kb
4·3 kb
2·3 kb
1·9 kb
1 2 3 4 5 6 7 8 9 10

Fig 24


23·7 kb
9·5 kb
4·3 kb
2·3 kb
2 3 4 5 6 7 8 9

Relation of various cDNA clones to the legumin 60,000 subunit

The two legumin cDNAs used for screening the gene library and characterising the isolated clones were pDUB 6 and pDUB 8 (Lycett *et al* 1984). Fig 25 shows the extent to which the two cDNA clones pDUB 6 & 8 relate to the legumin mRNA and the legumin subunit. pDUB 6 codes for the entire basic and approximately 40 % of the acidic subunit, including the 3′ non coding region and part of the poly A tail, while pDUB 8 codes for almost the whole of the acidic and approximately 20 % of the basic subunit. Taken together, the two clones almost completely code for the entire 60,000 subunit, with the exception of some sequence at the 5′ terminus.

Restriction enzyme and hybridisation analysis of λ LEG 3

λ LEG 3 was mapped with respect to restriction enzyme cleavage sites by a combination of hybridisation analysis and an examination of the sizes of the fragments produced. One of the mapping gels used for this analysis is shown in Fig 26. L47 vector DNA digested with the same enzymes as λ LEG3 was included as an aid to mapping on this gel, as it simplified the task of identifying vector fragments in λ LEG3. Southern blotting of these fragments and subsequent hybridisation against the insert from pDUB 8 gave the autoradiograph produced in Fig 27. The sizes of the hybridising fragments are given below :

Bam H1 - Eco R1 --- 4.8 kb.

Bam H1 - Bgl 2 --- 3.7 kb.

Eco R1 - Bgl 2 --- 6.0 kb.

Hind 3 --- 2.7 kb.

Xho 1 --- 1.8 kb.

Pst 1 --- 3.7 kb, 12.5 kb.

Bgl 1 --- 1.8 kb, 1.5 kb.

Fig 25

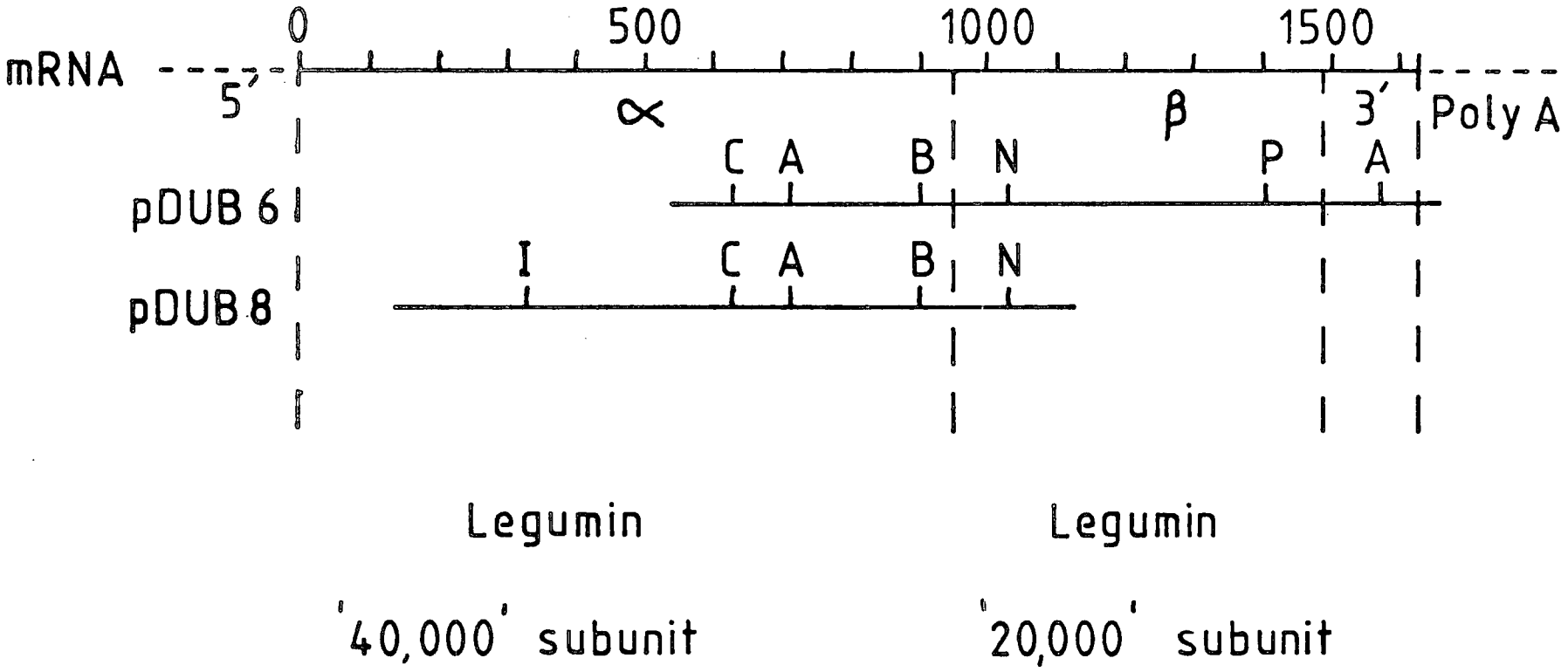

0
500
1000
1500
mRNA
5'
∝
β
3'
Poly A
C A B N P A
pDUB 6
I C A B N
pDUB 8
Legumin
'40,000' subunit
Legumin
'20,000' subunit

Fig 26

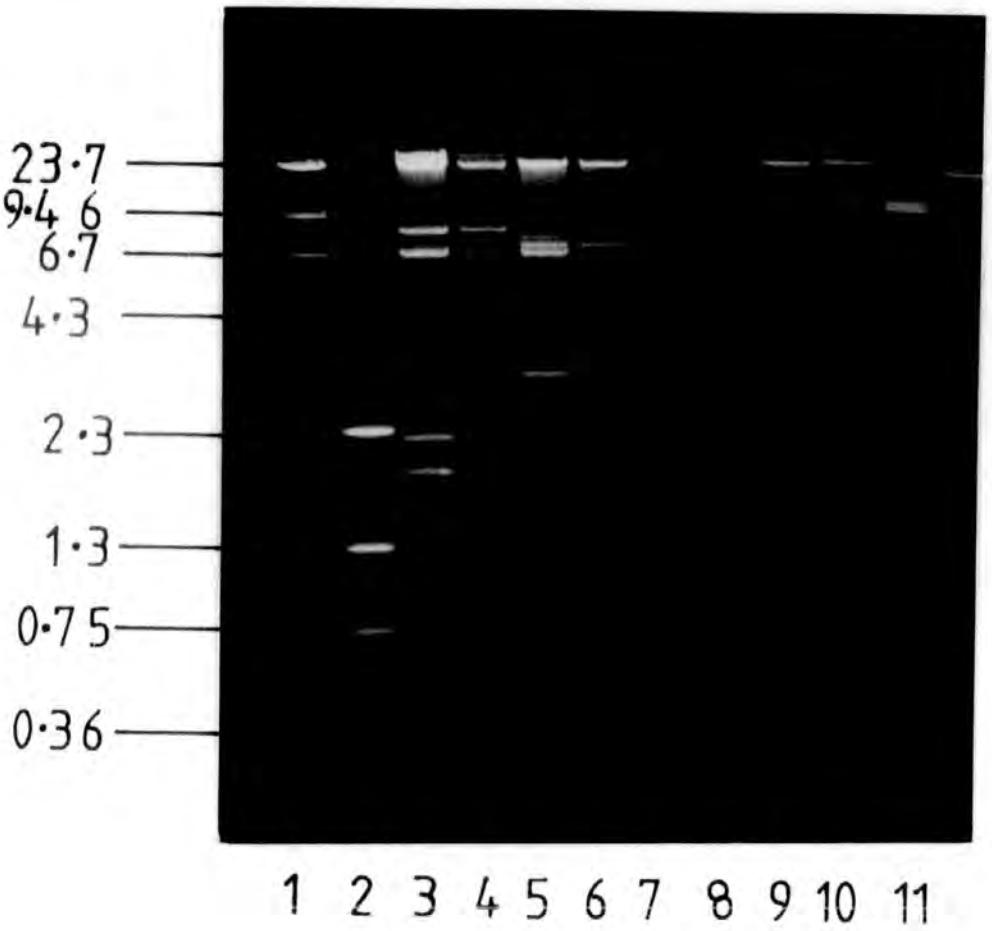
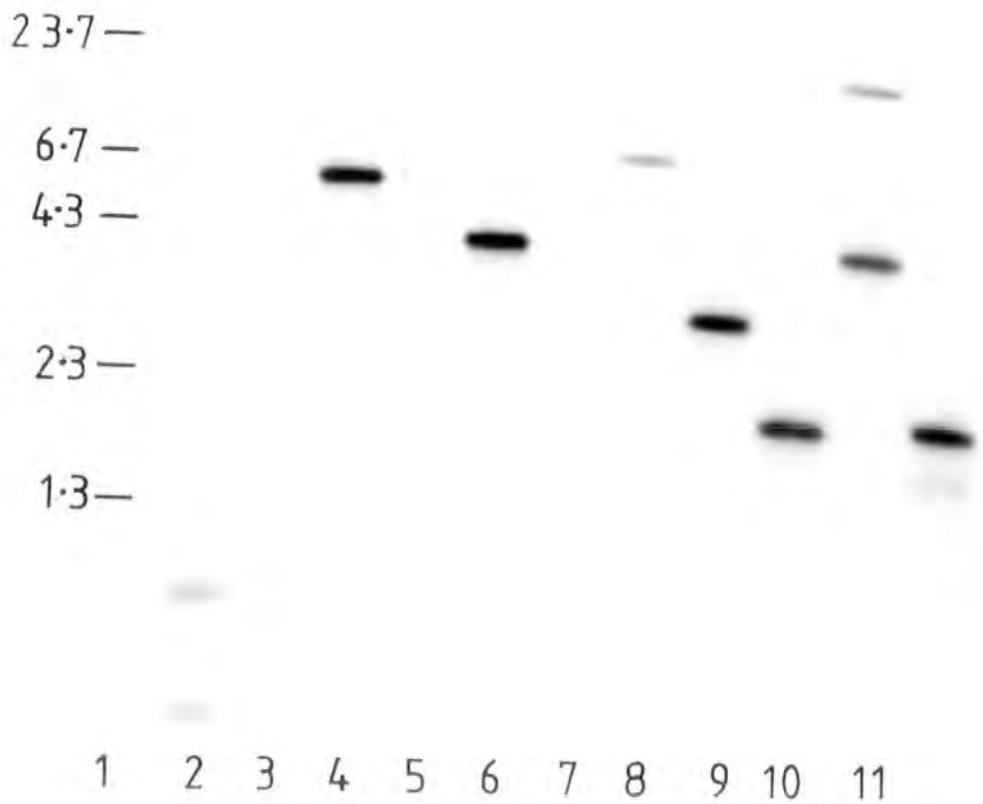

23·7
9·4 6
6·7
4·3
2·3
1·3
0·7 5
0·3 6
1 2 3 4 5 6 7 8 9 10 11

Fig 27


23·7
6·7
4·3
2·3
1·3
1 2 3 4 5 6 7 8 9 10 11

Comparison of the restriction enzyme cleavage maps of λ LEG 1, 2, and 3.

In order to confirm that λ LEG3 was a clone which had not previously been isolated, the restriction enzyme cleavage maps of λ LEG 1 and 2 (Croy et al 1984) were compared with the map of λ LEG3 (Fig 28). The sizes of the Eco R1 and Hind 3 fragments between the three clones were dissimilar. The Eco R1 fragments containing the legumin coding sequence in λ LEG 1, 2, and 3 were, respectively, 12.5, 7.6, and 6.4 Kb, while the Hind 3 fragments were 2.3, 2.3, and 2.7 Kb.

Restriction digests of pea DNA hybridised to λ LEG 3

In order to investigate the sizes of fragments in genomic pea DNA which were complementary to the legumin gene in λ LEG3, pea DNA was cleaved to completion with Eco R1, Hind 3, and Eco R1/Hind 3, separated on 0.5 % agarose gels, transferred to nitrocellulose paper, and hybridised against the ^{32}P labelled Xho 1 fragment of λ LEG3, which contained the legumin coding sequence. Fig 29 shows various restriction digests of pea DNA which were used in this experiment. In order to estimate the approximate gene copy number of the resulting hybridising fragments, various amounts of pAS 2 linearised with Hind 3 were electrophoresed in tracks 7, 8, and 9. The autoradiograph of the hybridised gel is shown in Fig 30. The sizes of the hybridising fragments observed in the various digests are given below :

Eco R1	Hind 3	Bam H1 / Eco R1
14.0 kb	4.2 kb	12.0 kb
12.0 "	2.6 "	9.2 "

Eco R1	Hind 3	Bam H1 / Eco R1
11.0 Kb		6.4 Kb
9.5 "		4.95 "
8.6 "		4.2 "
6.6 "		3.6 "
4.2 "		3.5 "

*

Table 5 shows a comparison of the sizes of the Eco R1 fragments from pea DNA hybridising to (a) the Xho 1 fragment of λ LEG3 and (b), the legumin cDNA 2.2.4 which codes mainly for the legumin basic subunit (Croy et al 1982). It is apparent that certain fragments which are seen in the λ LEG3 hybridisation are not seen in the DNA hybridised with cDNA 2.2 4, and vice versa.

Restriction enzyme and hybridisation analysis of pAS 2.

The 2.7 kb Hind 3 fragment of λ LEG3 which hybridised to the pDUB 6 and pDUB 8 legumin cDNAs was subcloned into the plasmid vector pUC 8 (Viera and Messing 1982). The resultant clone was named pAS 2. One of the gels used in the mapping experiments is shown in Fig 31. An autoradiograph of the same gel after Southern blotting and hybridisation against pDUB 6 is shown in Fig 32. The sizes of the fragments binding to the cDNA are given below :

Ava 1 / Pst 1 --- 0.84 kb

Pst 1 / Bam H1 --- 1.8 kb, 1.56 kb.

Ava 1 / Bam H1 --- 1.0 kb.

Bam H1/ Xho 1 --- 1.95 kb.

Bam H1/ Eco R5 --- 1.8 kb.

Xho 1 / Eco R5 --- 1.5 kb

Pst 1 / Xho 1 --- 1.75 kb.

Pst 1 / Eco R5 --- 1.4 kb, 1.1 kb.

Ava 1 / Xho 1 --- 1.1 kb, 0.96 kb.

Ava 1 / Eco R5 --- 1.05 kb, 0.5 kb.

This data, along with that from other experiments, allowed the construction of a restriction enzyme cleavage map for pAS 2.

Restriction enzyme cleavage map of pAS 2, showing fragments chosen for sequencing

Fig 33 illustrates the restriction enzyme map of pAS 2, showing the sizes of the various fragments in detail. Fragments chosen for sequencing are also shown.

Comparison of the restriction enzyme cleavage maps of pAS 2 with pDUB 6 & 8

The restriction enzyme cleavage maps of pDUB 6 and 8 (Lycett et al 1984) were compared with that of pAS 2 to examine points of similarity. Although most of the restriction enzyme sites appeared to be in the same location in all three clones, certain fragments in pAS 2 seemed to be larger. Fig 34 compares these restriction maps. The Bst N1 - Pst 1 fragment in pDUB 6 was approximately 90 bp smaller than the corresponding fragment in pAS 2, while the Eco R5 - Acc 1 fragment in pDUB 8 was approximately 80 bp smaller than its homologue in pAS 2.

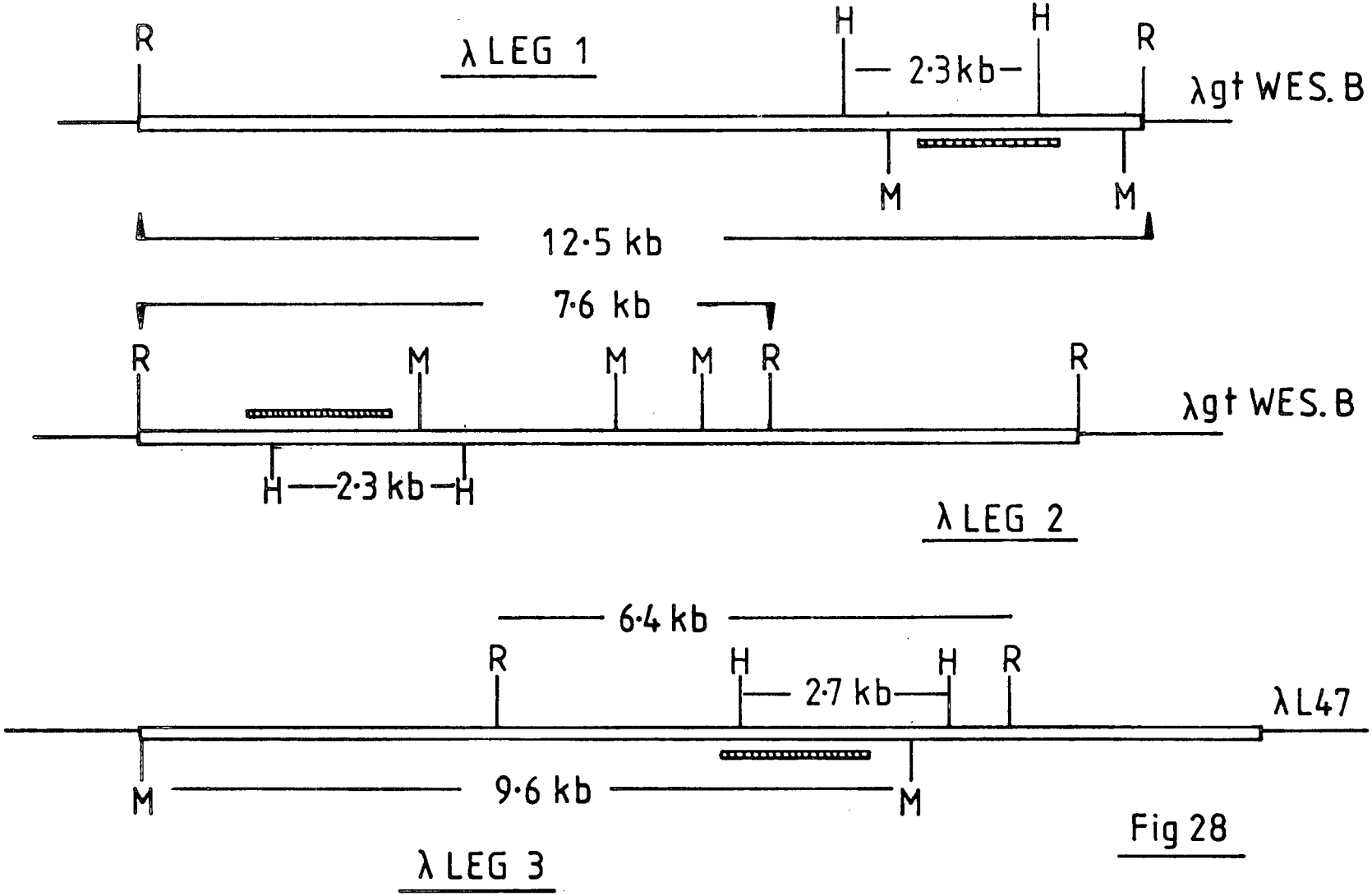

R
λ LEG 1
H
H
R
2·3kb
λgt WES. B
M
M
12·5 kb
7·6 kb
R
M
M
M
R
R
λgt WES. B
H
2·3 kb
H
λ LEG 2
6·4 kb
R
H
H
R
2·7 kb
λL47
M
9·6 kb
M
λ LEG 3

Fig 28

Fig 29

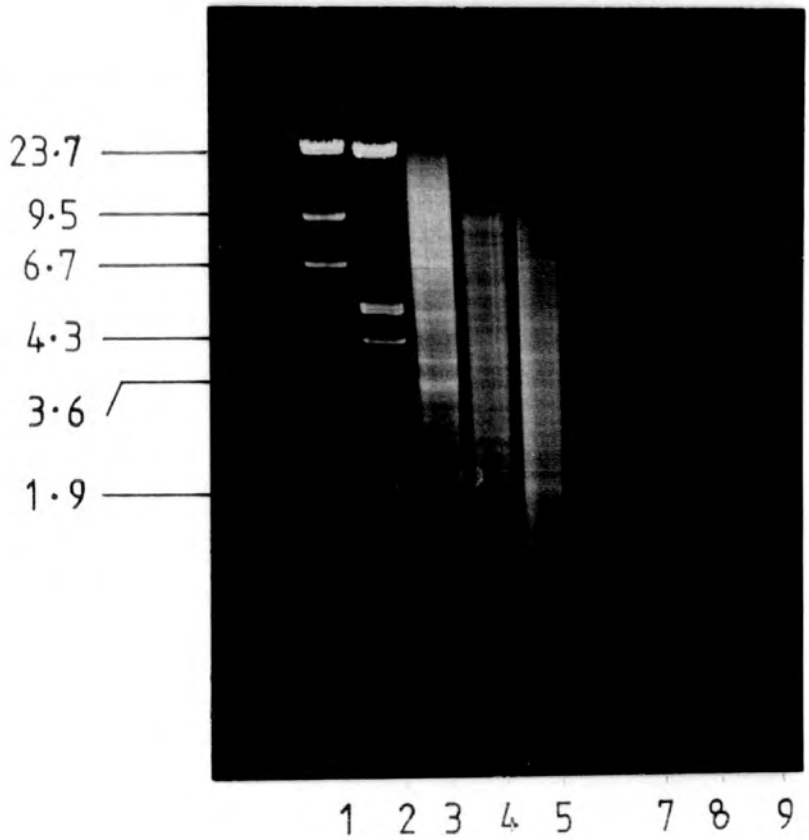
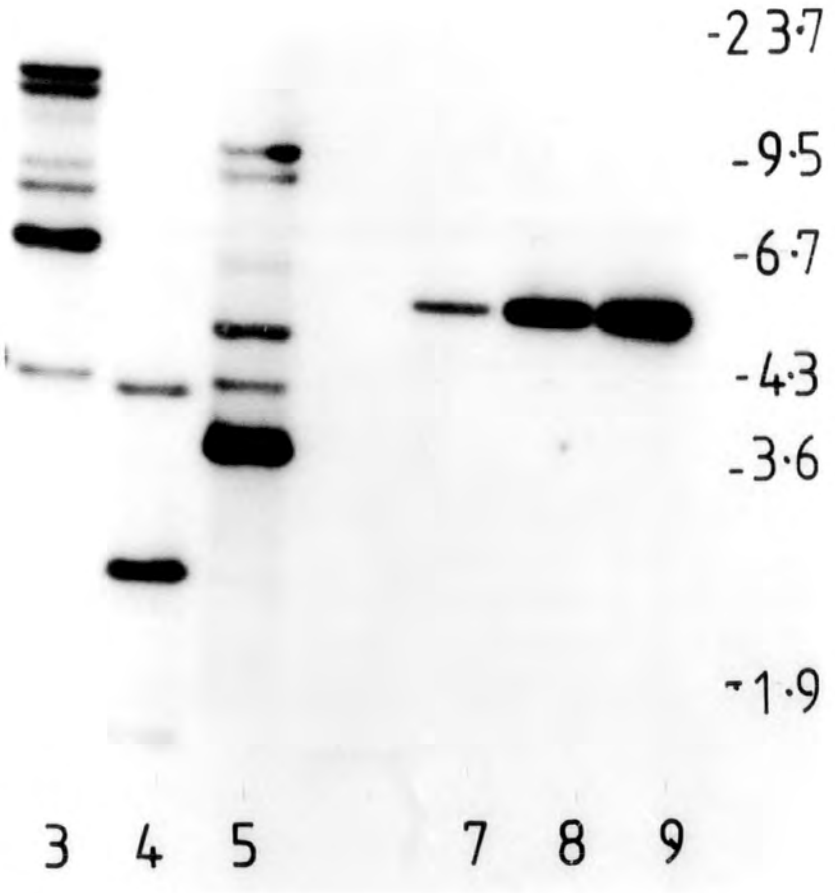


Fig 30



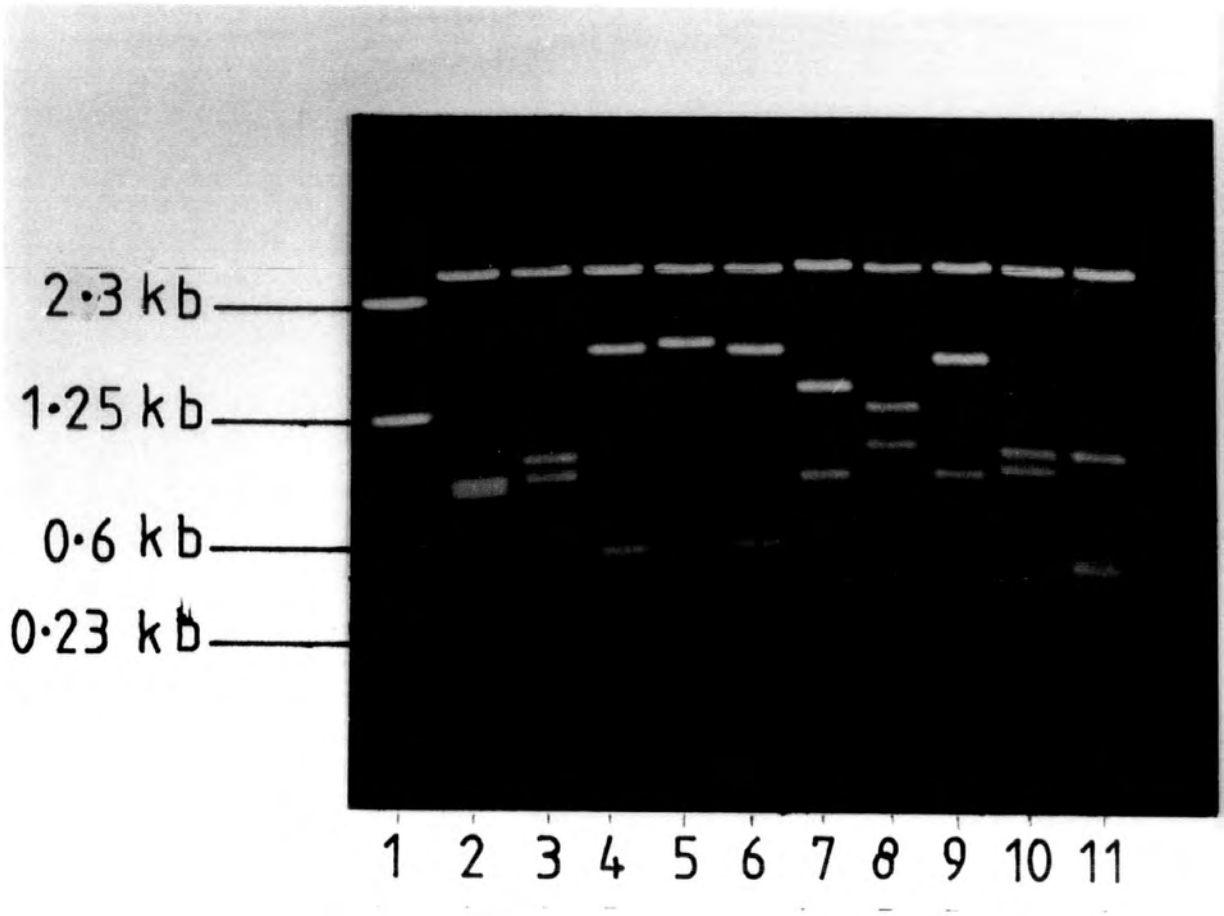
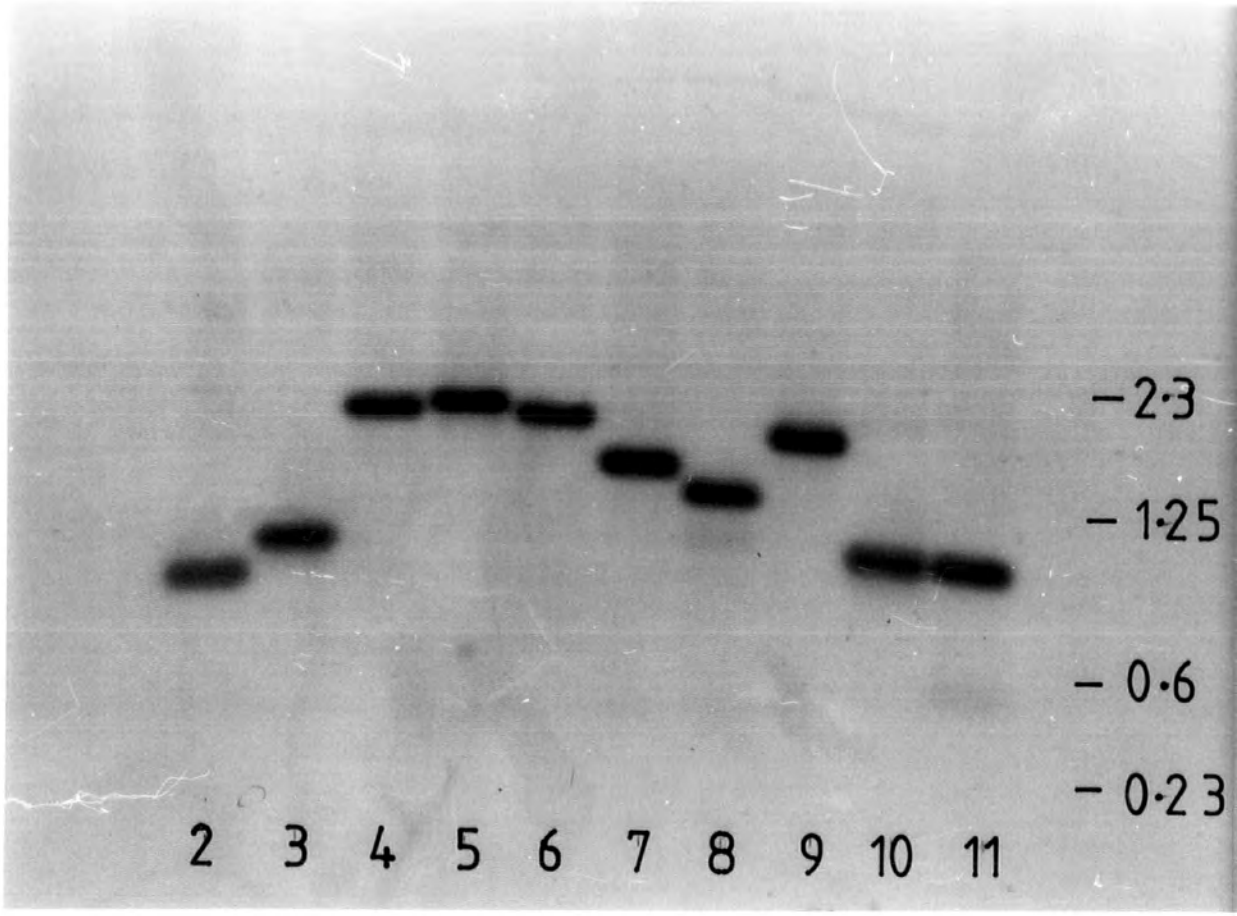
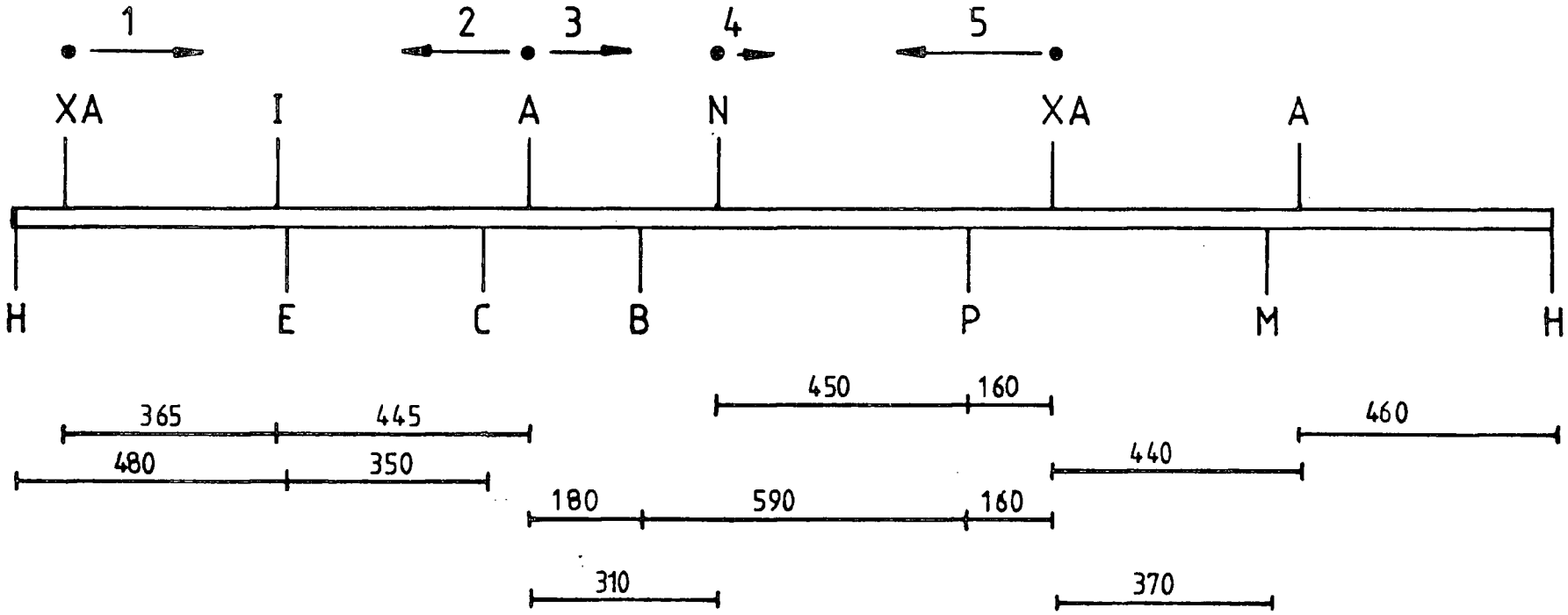

2·3 kb
1·25 kb
0·6 kb
0·23 kb
1 2 3 4 5 6 7 8 9 10 11

Fig 31

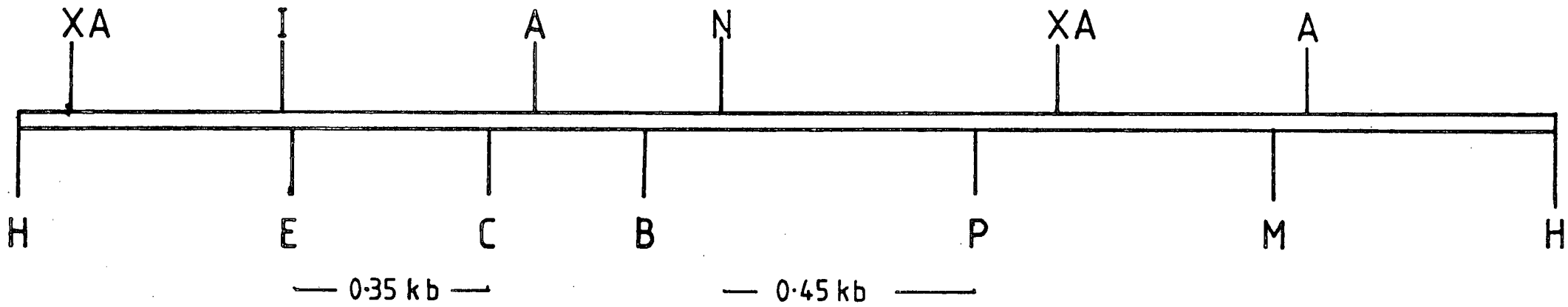
Fig 32

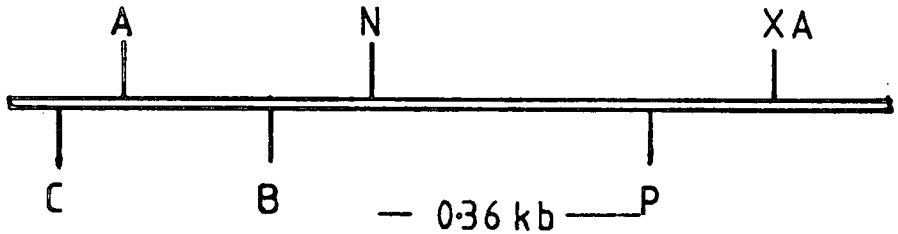

2·3
1·25
0·6
0·23
2 3 4 5 6 7 8 9 10 11



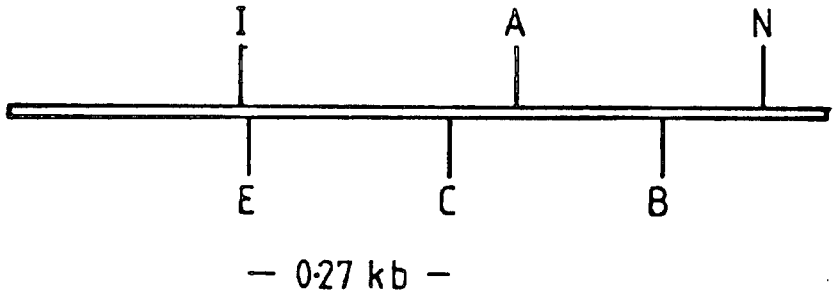
2·7 kb Hind 3 fragment.

Fig 33





pDUB 6

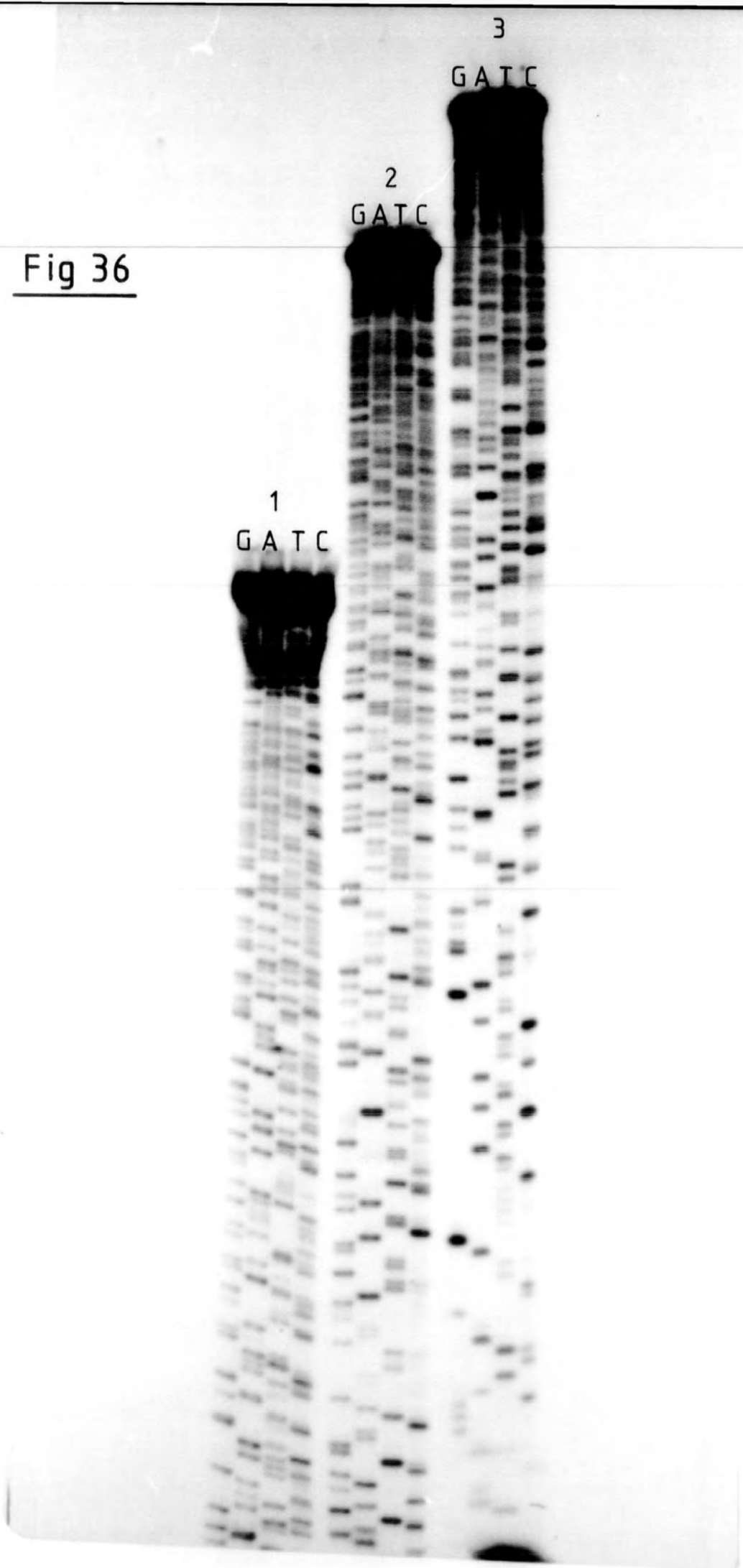


pDUB 8

Fig 34

Fig 35

Fig 36



Comparison of the sizes of homologous restriction enzyme fragments between the cDNAs pDUB 6 & 8 with the legumin genomic fragment of pAS 2 (Fig 35)

Data from Nielsen (pers. comm.), indicated that intervening sequences were present in the soybean glycinin gene (which codes for a protein analagous to pea legumin), in fragments defined by the restriction enzyme sites Bst N1 - Pst 1 and Acc 1 - Eco R5. A comparison of the restriction enzyme maps of pDUB 6 & 8 (Lycett et al 1984a) with the map of LEG3 (Fig 33 and 34), showed that these pea legumin cDNA fragments were smaller than their homologues in pAS 2. A direct demonstration of the occurence of intervening sequences in the pea legumin coding sequence by a comparison of homologous fragments between the cDNAs and pAS 2 was therefore attempted. Tracks 2 & 3 (Fig 35) show that the Bst N1 - Pst 1 fragment of pDUB 6 was approximately 90 bp smaller than its homologue in pAS 2. Similarly, tracks 4 & 5 showed that the Acc 1 - Eco R5 fragment of pDUB 8 was approximately 80 bp smaller than its homologue in pAS 2.

An example of an autoradiograph produced during the sequencing of pAS 2 (Fig 36).

Sequencing was performed by the method of Sief et al (1980).

Sequence of the legumin cDNA clones pDUB 6 & 8 (Fig 37).

This diagram was adapted from Lycett et al 1984. Although complete at the 3′ terminus, the sequence was incomplete at the 5′ end, since pDUB 8 lacked coding sequence for the 47 more amino acids required to reach the N terminus of the acidic legumin subunit. The sequence

differences between pDUB 6 & 8 show that they must have been synthesised from mRNAs coding for nonidentical legumin genes. Regions of repetition of the amino acid sequence are shown.

Partial sequence from 5 regions of the 2.7 kb insert in pAS 2 (Fig 38).

Fragment 1 extended 19 amino acids beyond the 5′ terminus of pDUB 8, approaching the N terminus of the acidic legumin subunit. It also included part of an intervening sequence (IVS 1). The first two nucleotides of the intervening sequence were GT. The intervening sequence was rich in A/T nucleotides. Five differences between the amino acid sequence between LEG 3 and pDUB 8 were found in Fragment 1 and a further five in Fragment 2. Fragment 3 possessed part of the Repeat 1 sequence found in the legumin cDNAs, the whole of Repeat 2, and part of Repeat 3. The remainder of the repeat regions were not sequenced. Fragment 4 showed the beginning of another intervening sequence (IVS 3), the first two nucleotides of the intron again beginning with GT. Fragment 5 contained 174 bp of sequence coding for the legumin basic subunit, an amber translation termination codon, and 156 bp of 3′ non coding sequence. No polyadenylation signals were found in this region.

Fig 35

Comparison of homologous fragments between the cDNAs pDUB 6 & 8 with the legumin genomic fragment from pAS 2.

pDUB 6 & 8 were cleaved with Bam H1 to isolate the cDNA inserts, and pAS2 was cleaved with Hind 3 to isolate the genomic fragment.

1) pBR 322 cleaved with Bam H1 and Bgl 1
2) pDUB 6 " Bam H1, Bst N1, and Pst 1
3) pAS 2 " Hind 3, Bst N1, and Pst 1.
4) pDUB 8 " Bam H1, Eco R5, and Acc1.
5) pAS2 " Hind 3, Eco R5, and Acc1.

The sizes of the homologous fragments between pDUB 6 and pAS 2 (tracks 2 & 3), and pDUB 8 and pAS 2 (tracks 4 & 5) are shown.

Fig 37

The sequence of the cDNA clones pDUB 6 & 8 is shown (adapted from Lycett *et al* 1984a). The uppermost sequence is shown in full and the other sequence is shown only where it differs from that sequence. The positions of the repeats is indicated below the sequences, and the presence of the small internal duplication in the second repeat is shown by brackets. The amino acid sequence predicted from the DNA sequence is also shown (A.A.). Asterisks (*) indicate the region for which comparable sequence data was obtained from the legumin genomic clone (Fig 38). The fragment numbers 1 - 5 refer to the regions sequenced in the genomic legumin clone (Fig 38).

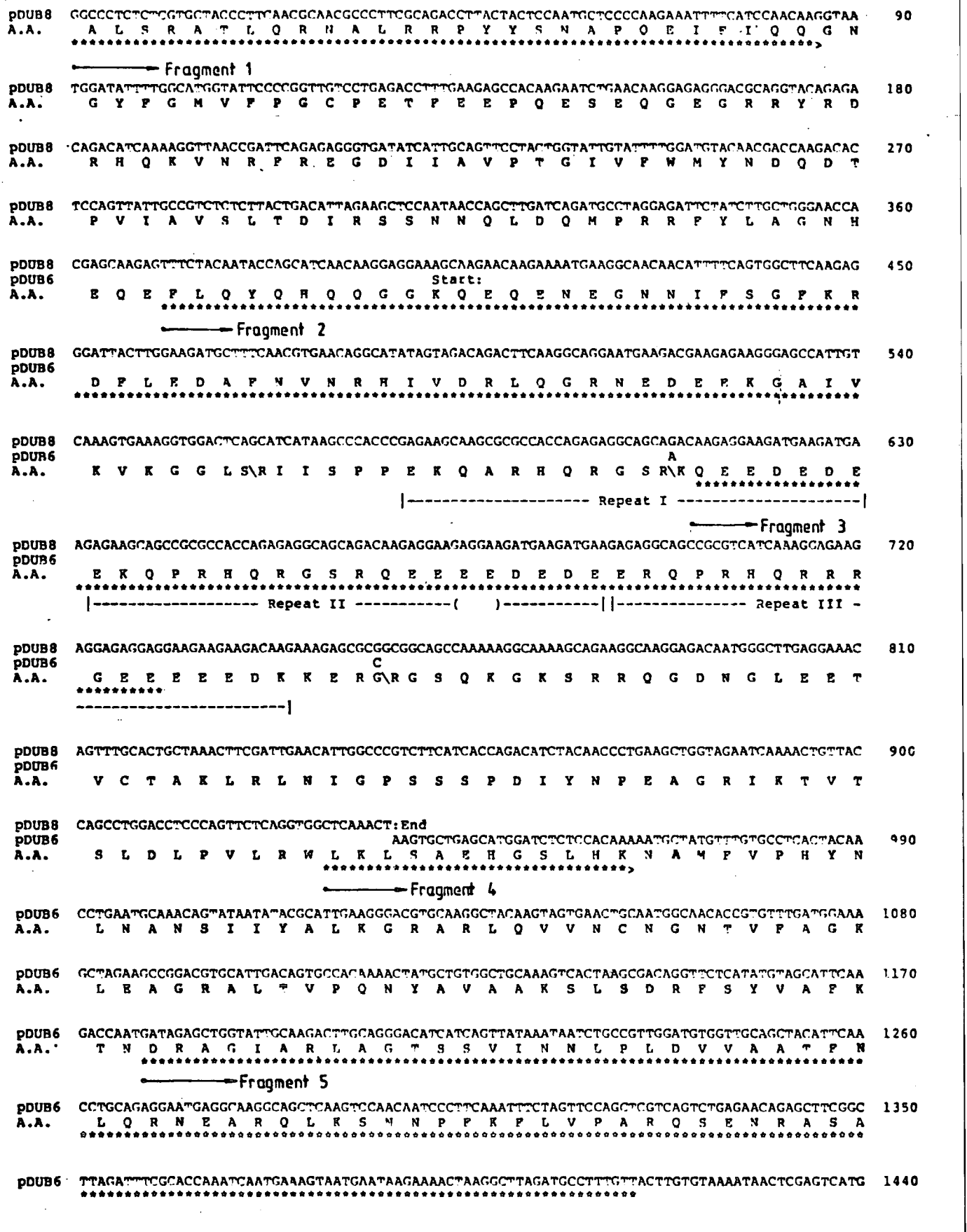

pDUB8
A.A.
Fragment 1
Start:
Fragment 2
Repeat I
Fragment 3
Repeat II
Repeat III
End
Fragment 4
Fragment 5

Fig 37

Fig 38

Partial sequence from five regions of the 2.7 kb insert in pAS 2 is shown. The position of these regions is shown in Fig 33. Homology between the cDNA sequence (Fig 37) and the genomic sequence can be seen by comparing regions 2, 3, 4 and 5 in Fig 38 with the corresponding regions in Fig 37. The cDNA sequence of Fragment 1 (Fig 37) commences 57 bp after the start of the genomic sequence of fragment 1 (Fig 38). Thus, data for 19 amino acids 5′ to the end of the pDUB 8 sequence is shown. Discrepancies between the amino acid sequences between LEG3 and the cDNA sequence are underlined. Silent base changes in the genomic sequence are indicated by (v) . Intervening sequences in fragments 1 and 4 are shown by (-----). The beginning of the intervening sequence is indicated by ()). The 3′ untranslated region in Fragment 5 is shown by (.....).

Fragment 1

```
LEG 3 AGGTGGGCTCATTGAGACTTGGAATCCCAACAACAAGCAATTCCGATGTGCTGGTG
A.A.    G  G  L  I  E  T  W  N  P  N  N  K  Q  F  R  C  A  G

LEG 3 GCATCATCTCGCGAGCTACCCTTCAACGCAACGCCCTTCGCAGACCTTGCTACTCCAA
A.A.  G  I  I  S  R  A  T  L  Q  R  N  A  L  R  R  P  C  Y  S  N

LEG 3 TGCTCCCCAAGAAATTTTCGACCAACAACAAG>GTTACTTATTTTGATCTTATACCT
A.A.    A  P  Q  E  I  F  D  Q  Q  Q  ----------------------

LEG 3 TCTTCTTTACGTACATTACATGCATATTAGCATA
A.A.  ---------------------------------
```

Fragment 2

```
LEG 3  TTTCTACGATACCAGCATCAACAAGGAGGAAAGCAAGAACAAGAAAATGAAGGCAAC
 A.A.   F  L  R  Y  Q  H  Q  Q  G  G  K  Q  E  Q  E  N  E  G  N

LEG 3 AACATTTTCAGTGGCTTCAAGAGGGTTTTCTTGGAAGATGCTTTCAACGTGAACGGC
A.A.   N  I  F  S  G  F  K  R  V  F  L  E  D  A  F  N  V  N  G

LEG 3 TATATAGTAGACAGACTTCAAGGCAGGAATGAAGACGAAGAGATTGGAGCCATTGTC
A.A.  Y  I  V  D  R  L  Q  G  R  N  E  D  E  E  I  G  A  I  V
```

Fragment 3

```
LEG 3 CAAGAGGAAGATGAAGATGAAGAGAAGCAGCCGCGCCACCAGAGAGGCAGCAGACAA
 A.A.  Q  E  E  D  E  D  E  E  K  Q  P  R  H  Q  R  G  S  R  Q

LEG 3 GAGGAAGAGGAAGATGAAGATGAAGAGAGGCAGCCGCGTCATCAAAGGAGAAGAGGA
A.A.   E  E  E  E  D  E  D  E  E  R  Q  P  R  H  Q  R  R  R  G

LEG 3 GAGGAG
 A.A.  E  E
```

Fig 38

Fragment 4

```
LEG 3 CTCAAACTAAGTGCTGAGCATGGATCTCTCCACAAA>GTAT
A.A.   L  K  L  S  A  E  H  G  S  L  H  K  ----
```

Fragment 5

```
LEG 3 GATAGAGCTGGTATTGCAAGACTTGCAGGGACATCATCAGTTATAAATAATCTGCCG
 A.A.  D  R  A  G  I  A  R  L  A  G  T  S  S  V  I  N  N  L  P

LEG 3 TTGGATGTGGTTGCAGCTACATTCAACCTGCAGAGGAATGAGGCAAGGCAGCTCAAG
A.A.   L  D  V  V  A  A  T  F  N  L  Q  R  N  E  A  R  Q  L  K

LEG 3 TCCAACAATCCCTTCAAATTTCTAGTTCCAGCTCGTCAGTCTGAGAACAGAGCTTCG
A.A.   S  N  N  P  F  K  F  L  V  P  A  R  Q  S  E  N  R  A  S

LEG 3 GCTTAGATTTCGCACCAAATCAATGAAAGTAATGAATAAGAAAACTAAGGCTTAGAT
A.A.   A Amb.....................................................

LEG 3 GCCTTTGTTACTTGTGTAAAATAACTCGAGTCATGTACCTTTTTGCGGAAACAGAAT
A.A.  ..........................................................

LEG 3 AAATAAAAGGTAAAATTTCAGTGCTCTATGCTTTTCTACTCCAAGTTA
A.A.  ..................................................
```

TABLES

Table 1

In vitro packaging efficencies using different substrates.

	SUBSTRATE	Plaques /ug
1	Phage buffer	0
2	pBR322 DNA	0
3	Pea DNA	0
4	L47 arms DNA	1×10^2
5	L47 native DNA	4×10^9
6	L47 DNA, restricted with BamH1 and ligated	2×10^7

Table 2

Determination of the optimum ratio of phage arms to pea DNA inserts which produced the maximum number of recombinants.

	Arms:Inserts	5K	L95	% Recombinant
1	0.5 : 1	2.4×10^4	9.6×10^3	40
2	1 : 1	4.7×10^4	3.6×10^4	76
3	2 : 1	6.4×10^4	1.7×10^4	26
4	3 : 1	6.3×10^4	1.9×10^4	30
5	Ligated arms	3.0×10^3	0	0
6	Unligated arms	0	0	0
7	L47 vector	6.6×10^8	0	0

Total DNA concentration in ligations was 200 µg / ml. Titres are expressed per µg.

Table 3

Constituents of a ' mini ' gene library ligation

Reactants	Volume /µls.
5 µg L47 arms DNA	14.3
3 µg Sau 3A cleaved pea DNA	6.0
10X Ligase buffer	3.2
10mM ATP	3.2
T4 ligase (4WU)	4.0
H_2O	1.3

Total DNA concentration in the ligation was 200 µg/ml.

Table 4

Comparison of the restriction enzyme fragments between Ribosomal clones 1, 2 and pHA 1 hybridising to pHA 1.

Bam H1 / Hind 3			Bgl 1 / Hind 3			Hind 3		
Ribo 1	Ribo 2	pHA 1	Ribo 1	Ribo 2	pHA 1	Ribo 1	Ribo 2	pHA 1
	4.0	4.0	4.1	4.1	4.2	8.3	8.3	8.8
							8.1	
				3.0				
2.5	2.5	3.0	2.2	2.2	2.4			
2.4	2.4	2.6	1.9	1.9	2.1			
1.3	1.4	1.4						
1.2	1.4	1.3						
0.65	1.4	0.65						

Bgl 2 / Hind 3

Ribo 1	Ribo 2	pHA 1
4.2	4.2	4.6
4.1		
2.0	2.0	2.2

Table 5

A comparison of fragments in genomic pea DNA hybridising to the legumin coding sequences of cDNA 2.2.4. (Croy etal 1982), and pAS 2.

Probe	2.2.4.	pAS 2	2.2.4.	pAS 2	pAS 2
DNA cleaved with	Eco R1	Eco R1	Hind 3	Hind 3	Bam H1/Eco R1
Fragment sizes (kb)	14.0	14.0	4.2	4.15	10.2
	12.0	12.0	2.5	2.6	9.5
		11.0		1.5	6.4
		9.5			4.95
		8.6			4.15
	7.7				3.6
		6.6			
	4.2	4.2			

Table 6

Comparison of the sequence around IVS 1 and IVS 3 with plant and animal intron/exon boundary consensus sequences.

	-3	-2	-1	+1	+2	+3	+4	+5	+6
Plant	X	T	G	G	T	A	A	G	T
Animal	A/C	A	G	G	T	A	A	G	T
LEG IVS 1	A	A	G	G	T	T	A	C	T
LEG IVS 3	T	A	T	G	T	T	T	T	T
	<---------			----------					----->
	EXON			INTRON					

X = any nucleotide

DISCUSSION

Before the advent of recombinant DNA technology, the isolation of single copy structural genes from eucaryotic genomes was considered an impossible task. However, along with the discovery of DNA restriction / modification systems, three important technical advances helped to make this goal a practical reality. First,procedures for the rapid isolation of the desired sequence from many thousands of recombinants were developed (Benton and Davis 1977). Second, a variety of vectors in which large amounts of DNA could be cloned were constructed (Blattner et al 1977, Leder et al 1977, Hohn and Murray 1977). Third, in vitro packaging systems were developed, which allowed a substantial increase in the efficency with which lambda derived vectors could be introduced into bacterial cells (Hohn and Murray 1977, Sternberg et al 1977).

The construction of the first eucaryotic gene libraries in lambda vectors was reported by Maniatis et al (1978), and the principles elucidated in this paper have been used in all subsequent library constructions. Central to this is the choice of cloning vector, which influences the method chosen to generate fragments of DNA from the genome under study. In the Maniatis method, (Maniatis et al 1978), a Charon phage (Blattner et al 1977), was used as an Eco R1 replacement vector, and DNA fragments generated either by mechanical shearing or by non limit digestion with restriction enzymes which produced blunt ends. These fragments were then size fractionated, methylated with Eco R1 methylase, and had Eco R1 linkers ligated to their termini, before being cleaved with Eco R1 and the resultant cohesive termini ligated into the vector. This involved procedure was used in an attempt to construct a random library in which no particular sequence was under represented. Other workers have used

the Charon series of vectors to construct gene libraries (Kemp et al 1979, Robbins et al 1979, Blattner et al 1978, Lewis et al 1981, Fischer and Goldberg 1982, Nagao et al 1981), but in most cases, these libraries have been non random in theory, since digestion of the genomic DNA with Eco R1 was performed before ligation into the vector (see below).

Another vector which has been widely used to clone fragments of eucaryotic DNA is λgtWES (Leder et al 1977). Although ths vector has been used in cloning unique pieces of DNA, it is less useful in the construction of random gene libraries, since it lacks sites for Bam H1, and thus fragments produced by random cleavage using non limit digestion with the enzyme Sau 3A cannot be cloned. As explained in the introduction, partial cleavage of DNA with a frequently cleaving enzyme such as Sau 3A which recognises a 4 base pair sequence, is the easiest method of producing randomly cleaved DNA. In addition, it has the advantage that fragments with Sau 3A cohesive ends can be ligated directly into a Bam H1 cleaved vector, thus obviating the need for linkering, or methylation of genomic DNA, as was necessary in the Maniatis library.

Vectors which allow the easy construction of random gene libraries, by cloning Sau 3A partial digests of genomic DNA into the Bam H1 site of the vector, have been described (Karn and Brenner 1980, Loenen and Brammar 1980, Mizusawa and Ward 1982, Rimm et al 1980). The vectors described by Karn et al (1980) and Loenen and Brammar (1980) rely on the replacement of the central region of phage lambda derivatives, which contains the red and gam genes, by foreign DNA. Recombinants produced using these vectors are therefore red$^-$ gam$^-$, and

consequently exhibit the Spi⁻ phenotype, i.e. they will grow in bacterial hosts lysogenic for the phage P2 (see introduction). Non recombinants are termed Spi⁺ , since the products of the exo, bet, and gam genes of λ interact with the old gene product of phage P2, leading to inhibition of growth (Lindahl et al 1970). Thus, growth of a recombinant on a host lysogenic for phage P2 provides a selection system. Although both these vectors can be used to clone long fragments of DNA, λ 1059 (Karn et al 1980) has a major disadvantage in that it contains the pBR 322 origin of replication. Although this allows the clone to be propagated as a plasmid (phasmid) in λ lysogens, the presence of the pBR322 sequence in the vector makes the screening of libraries constructed in λ 1059 difficult if not impossible, since most cDNA probes are constructed using pBR 322 or its derivative vectors. As cDNA inserts can never be completely purified away from the vector sequence, spurious hybridisation (false positives) due to pBR 322 homology between probe and clone is bound to occur. For this reason, λ L47 (Loenen and Brammar 1980), which does not have this pBR322 sequence, was chosen as a vector to construct the Pisum sativum L. gene library.

It is desirable, in a cloning experiment, to reduce the number of non recombinant transformants, since this will lead to a reduction in the total number of clones which have to be screened for the sequence of interest. This is usually done, in the case of a bacteriophage vector, by isolating the annealed arms of the phage. In this work, the right and left ends of λ L47 (Fig 1) were annealed by their cos ends, and T4 ligase used to covalently seal the single stranded nicks left after annealing. After cleavage with Bam H1 to excise the central replaceable fragment (Fig 1), the arms (Fig 2) were

separated from the replaceable fragment by centrifugation through a glycerol density gradient. The successful annealing and ligation of the λ L47 arms is shown in Fig 3, and an analyses of fractions obtained from the glycerol gradient (Fig 4) indicates that the arms were purified, as judged by agarose gel electrophoresis (Fig 5).

As only DNA molecules of the correct length (between 38 and 53 kb long) will be packaged (Feiss et al 1977), the unligated vector arms should not be packaged, as they are only 34 kb in length (Fig 2). Ligation of these arms at their cos ends will lead to the formation of concatenates of DNA, which are the substrates for in vitro packaging; however, cleavage at the cos sites of these molecules will produce forms which are either too small or too large to be packaged (Fig 6). Hence, only recombinant molecules, in which DNA has been inserted at the Bam H1 sites, will be the right size for packaging into phage particles (Fig 7 and 8).

As stated earlier, another essential factor in the construction of gene libraries is the development of an in vitro packaging system, as it allows efficencies of at least two orders of magnitude above those obtained by $CaCl_2$ transfection (Thomas et al 1974 and introduction). Before it can be used in packaging DNA for a gene library, the system must a) have efficencies of 10^7 pfu/μg or higher when intact λ DNA is used as a substrate and b) must not show any packaging due to the presence of endogenous λ DNA in the extracts. The packaging system used in this project satisfied both the above criteria, since 1) the packaging efficency was 4×10^9 pfu/μg (Table 1), at least 2 orders of magnitude above that reported by Maniatis et al (1978), and 2) no endogenous DNA was packaged (Table 1). The

efficency of packaging, although high, varied between preparations, and it was found that the most critical step in determining this was the sonication of the induced preparation of sonic lysate. Too little sonication resulted in insufficent lysis of cells, and therefore decreased availability of phage particles, while too much sonication led to the disruption of the liberated phage components (results not shown). Both factors led to a decrease in packaging efficency. The other critical step in the preparation of packaging extracts was induction; cells had to be in the exponential phase of growth and the heat stimulus had to be rapidly applied, otherwise the cultures were not efficently induced. Typical growth curves of the two extracts, which led to high packaging efficencies, are shown in Fig 9. The specificity of the system for DNA molecules with cos sites was demonstrated, since molecules lacking these sites, e.g. pBR 322 and pea DNA, were not packaged. The packaging efficency fell to about 2 x 10^7 pfu/µg when cleaved and religated λ L47 DNA was used as a substrate (Table 1); this is because a variety of different combinations of fragments are made in a ligation reaction, only a few of which form viable phage. In any case, this value was 4 orders of magnitude higher than that obtained when a similiar cleaved and religated DNA substrate was used in a $CaCl_2$ transfection (Mandel and Higa 1970).

The preparation of the annealed arms of λ L47 and the production of an efficent *in vitro* packaging system meant that only one additional component - the partially digested pea DNA, was necessary, before the gene library could be constructed. DNA used in library constructions must be of high molecular weight, free of nuclease contamination, and with few or no single strand breaks. It can be prepared from whole tissues such as leaf, embryo, and root, or

from isolated cellular organelles such as nuclei, mitochondria, or chloroplasts. Since the aim of this work was to prepare a total gene library, in which all DNA sequences would be represented, pea leaf tissue was chosen as an abundant and convenient source of material. The ultraviolet absorbance spectrum of a sample of this DNA, together with the A_{260} / A_{280} ratio obtained, showed the high purity of the DNA preparation, and indicated that contamination with protein and phenol was negligible (Fig 10). As judged by gel electrophoretic analysis, the majority of the DNA was of a size greater than 50 kb (Fig 11). It is impossible to accurately estimate DNA sizes greater than 50 kb, as the resolution of agarose gels in this region is very poor. No endogenous nuclease activity was present in the preparation, since incubation of a sample of this DNA in restriction enzyme buffer at 37°C for 1 h showed no appreciable degradation (Fig 11). However, it was not possible to absolutely rule out the presence of nucleases which required conditions different to those present in this experiment. The DNA was easily cleaved by restriction endonucleases, indicating that no inhibiting materials were present in the preparation, indeed, the presence of families of sequences bounded by Bam H1 sites was clearly seen (Fig 11 track 6). The extent of single strand nicking of the pea DNA preparation was also tested by using the DNA in a DNAse free nick translation reaction, to assay the percent incorporation of label, this was found to be 3 % of that obtained when DNAse 1 was used in the experiment (results not shown). Thus, single strand breakage of the DNA was assumed to be negligible.

Conditions for the partial cleavage of the pea DNA were established (Fig 12). The maximum amount of 15 - 20 kb fragments appeared to be in track 7 as judged by the intensity of fluorescence.

However, as Seed et al(1982) have calculated, this fraction represents twice the extent of digestion necessary for optimal representation of sequences in a library. Therefore, track 8, which used half the amount of enzyme as in track 7 actually yielded the greatest number of molecules in the desired size range. On the basis of this, 4 large scale digestions of pea DNA were performed (Fig 13) and pooled, using 0.5, 0, 1 and 2 times the amount of Sau 3A which yielded the maximum amount of 15 - 20 kb fragments. This was done in order to randomise the 15 - 20 kb size fractionated DNA population, by ensuring that molecules which varied in their degree of susceptibility to cleavage by Sau 3A were included in the preparation for cloning e.g. the 15 - 20 kb fraction in track 3 Fig 13 contained molecules which were relatively easily cleaved by Sau 3A, while track 6 contained a fraction of molecules which were relatively resistant to cleavage. After unloading the gradient (Fig 14), the fractions were checked to obtain the correct size range for cloning. Good separation of the various sizes of DNA was obtained (Fig 15), and appreciable quantities of DNA in the correct size range were obtained (Fig 16). Nevertheless, despite care taken in library construction, the randomness of a gene library can only be judged by the end product; i.e. the ability to isolate from that library, sequences present at about the same frequency as their occurence in the original genome. Investigators have sometimes failed to isolate certain sequences from libraries which were previously considered to be complete (Lawn et al 1978), and new libraries, using different restriction enzymes and digestion conditions, have had to be constructed before the desired sequence could be isolated.

When ligating together the vector arms and Sau 3A cleaved DNA,

two factors have to be considered 1) the molar ratio of arms to inserts and 2) the concentration of the two DNA species. As shown in Fig 7, the ideal substrate for *in vitro* packaging is a concatemer of the form left arm -- insert -- right arm. Since each vector has only one terminus compatible with the two termini of the insert, in theory, a 2 : 1 molar ratio of arms to inserts should give an equimolar amount of the two species of ligatable termini. However, this assumes an ideal situation, in which all the molecules in the ligation reaction have cohesive termini. Since some molecules will have lost their cohesive ends in extraction and purification procedures, this ratio will alter, and a series of test ligations with varying arms : insert ratios was set up to monitor this (Table 2). The DNA concentration in the ligations was chosen so as to favour intermolecular ligation leading to the formation of concatemers, over self ligation which leads to the circularisation of the DNA species (Dugaiczak *et al* 1975). In order to show that all possible inhibitors of ligation had been removed, and that the cohesive termini generated on the vector and pea DNA were capable of ligation, small aliquots of both DNA samples were ligated and the products analysed by agarose gel electrophoresis (Fig 17). Both vector and arms DNA were seen to form molecules of λ size or larger, indicating that concatenates had been formed. Although these test ligations were performed at 23°C for 2 h, it is generally better to ligate at lower temperatures (12°C) for up to 48 h, since this minimises breakage due to contaminating nucleases, and stabilises annealed sequences (Dugaiczak *et al* 1975).

The results obtained in Table 2 showed that a 1 : 1 molar ratio of arms to inserts produced the greatest number of recombinants, as judged by the ability of the Spi$^-$ phage to form plaques on the P2

lysogen E. coli strain L95. However, difficulties were experienced when using L95 as a host strain, and it is possible that the figure of 76 % recombinant is an underestimate. When using 100 µl of prepared cells per transfection (see methods), no plaques were seen after an overnight incubation on E. coli L95, and plaques were only seen if the plating bacteria were reduced to 1/100th the original. Even so, plaques were generally small, poorly defined, and difficult to see. Other workers using the λ L47 system have also experienced similar problems, and workers in this laboratory have found a similiar situation with the cloning vector λ gtWES.B and the host strain E. coli LE392 (Croy.pers.comm.). A possible explanation of this situation might be that E. coli L95 host cells infected by recombinant phage do not have the same growth characteristics as uninfected cells. It is known that the introduction of recombinant DNA into bacteria sometimes results in a slow growth rate of the transformed cells. Since bacteriophage only infect and lyse cells in exponential phase, it is probable that by the time lysis of one infected E. coli L95 cell occurs, the surrounding cells have already reached stationary phase, and are thus unable to be infected. If a smaller amount of plating cells are used, however, a larger proportion of host cells will be in the exponential phase as lysis occurs, leading to further infection and ultimately larger plaques.

Another way of ascertaining the number of background (non recombinant) phage is to package a ligated preparation of vector arms. This gave a figure of 3 x 10^3 pfu / µg - only 0.001 % of that obtained when intact vector DNA was used as a substrate (Table 2). No such background was seen when unligated arms were packaged, indicating that the non recombinant phage were not due to the presence, in the arms preparation, of unrestricted intactλ DNA. Thus, the background

must have been due to the presence of small amounts of contaminating replaceable fragment, although no bands corresponding to this fragment were seen in an agarose gel analysis of the arms preparation (Fig 5). Titering of the ligated preparation on *E. coli* strain L 95 resulted in no plaques, showing that foreign DNA had not been ligated into the arms (Table 2). Other workers (Maniatis *et al* 1978, Blattner *et al* 1978), have also reported the presence of a small proportion of non recombinant phage in their libraries. Although it is possible to lower this fraction by,for example, phosphatasing the vector and replaceable fragment after cleavage, or by loading smaller amounts of annealed DNA onto the gradients and thereby enhancing the resolution between the arms and the replaceable fragment, in practice, the proportion of non recombinants is so small that it can safely be ignored.

The number of recombinant phage produced per µg of ligated DNA was in excess of 3×10^4 pfu/ug (Table 2), this was about the same as the titre obtained in the construction of the rabbit and Drosophila libraries by Maniatis *et al*(1978). Using the equation derived by Clarke and Carbon (1976), and assuming that 1) the *Pisum sativum* L. genome has a size of 4.8×10^9 bp (Thompson and Murray 1981) and that 2), the average size of the cloned insert is 16 Kb, 1.8×10^6 recombinant phage will be required to find any given single copy sequence with a probability of 99 %. As 3×10^4 recombinant phage per µg were produced in the most efficent packaging, this meant that 60 µgs of ligated DNA were needed to construct the complete pea library. In order to simplify the problem of screening such a large collection of clones, it was decided to construct the library in a series of 8 mini libraries, each of which would be screened before proceeding to the

construction of the next. The composition of the ligation reaction for one of these mini libraries is shown in Table 3 , and its successful ligation, as judged by the formation of DNA molecules larger than λ size is shown in Fig 18. Sixteen *in vitro* packaging reactions were used, each to package 0.5 µg of ligated DNA. Maniatis *et al* (1978) concentrated and purified the packaged phage on CsCl step gradients, but this was not done in the present work in order to avoid loss of the phage during unloading and subsequent processing of fractions from the gradient.

Before a full scale screening of this library was carried out, a small portion (10^5 phage) was plated out using the non selective *E. coli* strain 5K as a host, and hybridised with a ^{32}P labelled pBR 322 probe. This was done to check for contamination of the library with pBR 322 sequences, an occurrence which appears to be common in most laboratories where extensive work with pBR 322 has taken place (Ish Horowitz pers.comm.). From such a screening, two clones were isolated which contained the whole of the pBR 322 molecule present as part of a 16 Kb insert (data not shown). This contamination occurred despite the fact that stringent precautions were taken to ensure purity of reagents and cleanliness of glassware. It is therefore important to check all putative clones isolated from genomic libraries against common cloning vectors, especially pBR 322, in order to guard against this kind of contamination. The chance of detecting such contamination can be reduced if purified insert DNA is used as a probe rather than whole recombinant plasmids, but even so, since inserts can never be wholly separated away from vector sequences, the problem still remains. Due to the possible presence of pBR 322 contaminants,

the library was not amplified, as this would have led to the amplification of the pBR 322 containing clones and a consequent increase in the number of putative clones which would have to be rescreened. Instead, the library was plated out directly using E coli 5K as a host, to give about 6.8 x 10^4 phage per megaplate. Although a higher plaque density could have been used (Benton and Davis 1977), this would have resulted in small plaques, and plating out at the lower density allowed relatively large plaques to form, with a consequent increase in the strength of the hybridisation signal.

The library was screened using intact ^{32}P labelled legumin cDNA plasmid pDUB 8 (legumin cDNA cloned into pBR 322 - Lycett et al 1984 - Fig 25), as a probe. The separation of a typical nick translation labelling reaction on a Sephadex column is shown in Fig 19. Incorporation rates of 48 % were routinely obtained. After autoradiography of the screened filters, 4 positive signals were detected amongst 6 x 10^4 phage. In order to check whether any of these contained contaminating pBR 322 sequences, aliquots of phage were hybridised separately against labelled pBR 322 and the purified insert from pDUB 8 (Fig 20 and 21). 3 of these clones were found to hybridise to pBR 322, and were not analysed further, while one hybridised to the legumin cDNA insert alone. This clone (λLEG3), was purified by repeated screenings and hybridisations until 95 % of the plaques on a plate showed positive signals (Fig 22).

With a pea genome size of 4.8 x 10^9 bp and a cloned insert size of 16 Kb, only one in 3 x 10^5 independent recombinant phage will carry a particular single copy sequence. Southern blot analysis of genomic

DNA has shown that approximately 4 copies of the legumin gene are present in the Pisum sativum L. genome (Croy et al 1982), and so one would expect to isolate one of these in every 7.5 x 10^4 independently generated phage screened, a value close to that obtained in practice. In order to further check the sequence representation within the library, screening for ribosomal DNA sequences using the pea genomic ribosomal clone pHA 1 as a probe (Cuellar 1982) was carried out. Screening of 10,000 phage from the library resulted in the isolation of approximately 85 hybridising clones. Assuming that there are approximately 3000 rRNA genes in the Pisum sativum L. genome, (Ingle and Sinclair 1972), one would expect to isolate one in every 100 independently generated phage screened, a value very similar to that obtained in practice. Analysis of two of these clones using restriction enzyme cleavage and Southern blotting (Figs 23 and 24) confirmed that they were indeed ribosomal genes, due to the close correspondence between the sizes of the restriction fragments between pHA 1 and ribosomal clones 1 and 2 (Table 4). Thus, it was concluded that the pea gene library was near complete.

A restriction cleavage map of λ LEG3 constructed using a combination of single and double digests with various restriction enzymes, showed that the smallest single fragment which hybridised to the legumin cDNA in pDUB 6 was the 1.8 Kb moiety produced upon cleavage of the clone with Xho 1 (Figs 26 and 27). Since the legumin subunit has a Mr of approximately 60,000 (Croy et al 1979), consisting therefore of about 500 amino acids (equivalent to 1500 bp of coding sequence), it was considered likely that this 1.8 Kb Xho 1 fragment contained a major part of the gene, assuming that introns, if present, were small. In confirmation of this, hybridisation of the legumin

cDNA pDUB 8 (Lycett et al 1984 and Figs 25 & 37), which coded for most of the basic and 95 % of the acidic subunit, to a Xho 1 digest of λ LEG 3, failed to show binding to any additional fragment (data not shown). If long introns had been present, a significant portion of the coding sequence would have extended past the 1.8 kb Xho 1 fragment, and hybridisation with pDUB 8 should have detected this. However, pDUB 8 is not a complete cDNA clone; it lacks some of the sequence encoding the N terminus of the acidic legumin subunit. If this additional 5' sequence had been present on a fragment adjacent to the 1.8 kb Xho 1 moeity of λ LEG3, it would be difficult to detect by hybridisation methods alone.

Croy et al (1982), found 4 Eco R1 fragments hybridising to the legumin cDNA 2.2.4, and therefore postulated the existence of at least 4 genes coding for the family of legumin proteins. In order to see whether λ LEG 3 corresponded to any of these fragments, genomic pea DNA digests cleaved with various enzymes were blotted and hybridised against the labelled 1.8 kb Xho 1 fragment of λ LEG 3 (Fig 29). The autoradiograph produced (Fig 30), indicated that a 6.6 kb strongly hybridising fragment was present in the Eco R1 digest which did not correspond to any of the bands obtained by Croy et al (1982). In addition, the 7.7 kb Eco R1 fragment seen by Croy et al (1982) was not observed in this experiment. Three weakly hybridising Eco R1 fragments of mol.wt. 11.0 kb, 9.5 kb, and 8.6 kb which were also not previously observed, were detected in this experiment. In contrast, 3 Eco R1 fragments of mol.wt. 14.0 kb, 12.0 kb, and 4.2 kb were found to be identical in both experiments. These results are summarised in Table 5. It thus seems likely that the legumin gene family consists of more than four genes. The

absence of the previously seen 7.7 kb fragment and the presence of the previously unobserved 6.4 Kb moiety in genomic Eco R1 digests probed with λ LEG 3 (Fig 30), suggests that differences between the nucleotide sequences of 2.2.4 and λ LEG 3 must exist. A similiar situation must arise in the case of the other 3 previously unobserved fragments. Before this can be proved, however, complete sequencing of λ LEG3 is essential.

Comparison of the restriction map of λ LEG 3 with the map of the other 2 genomic legumin clones showed points of similarity and divergence (Fig 28). The clones λ LEG 2 and λ LEG 1 (Croy et al 1984) were isolated from a library obtained by ligating pea DNA partially cleaved with Eco R1 into the bacteriophage vector λ gtWES (Tiemeir et al 1976). The Eco R1 fragments containing the legumin gene within these two clones corresponded to the 12 kb and the 7.6 kb fragments observed by Croy et al (1982) in digests of genomic DNA, thus showing that they were different from the 6.4 kb Eco R1 fragment of λ LEG3, as already found using genomic DNA hybridisation studies (Fig 30). Conversely, λ LEG 1, 2 and 3 all had a small Hind 3 fragment containing most of the legumin coding sequence. However, the 2.7 kb Hind 3 fragment in λ LEG3 was longer by about 0.46 kb at its 3′ end than the corresponding fragments in λ LEG1 and λ LEG2 (Fig 28).

In order to further characterise the legumin coding sequence of λ LEG 3, the 2.7 kb Hind 3 fragment (Figs 27 & 28) was ligated into the plasmid vector pUC 8 (Viera and Messing 1982), and transformants selected as described in the methods. pUC 8 is a useful vector to use in sub cloning experiments, since the variety of restriction sites present on the linker fragment within the plasmid allows the insertion

of foreign DNA fragments with a wide variety of termini. Hybridisation of pDUB 6 to single and double restriction digests of pAS 2 (Figs 31 and 32), enabled the construction of a detailed restriction enzyme cleavage map (Fig 33). Fig 34 compares this map with that of the cDNAs pDUB 6 & 8. The striking similarity between the distribution of restriction sites in both the cDNAs and the genomic clone indicated that the genomic sequence was closely related to the cDNA. The presence of two small intervening sequences (introns) in the genomic sequence was inferred by a close examination of the sizes of homologous fragments between the cDNAs and the genomic clone (Fig 35). The Eco R5 - Acc1 fragment in pAS 2 was approximately 80 bp bigger than the corresponding fragment in pDUB 8, while the Bst N1 - Pst 1 fragment was approximately 90 bp bigger than its homologue in pDUB 6. These particular fragments were chosen to look for the existence of introns because data for the soybean glycinin gene (which codes for a protein closely homologous to pea legumin) indicated that introns were present within these glycinin gene fragments (Nielsen pers. comm.). Another intron, at the 5′ end of the pea legumin gene, whose existence was impossible to deduce due to the lack of the appropriate restriction sites, was later discovered by DNA sequencing (see below). It was recently shown that the soybean glycinin gene was split by three introns of 238, 292, and 624 bp (Nielsen 1984). Although the sizes of the introns in the λ LEG3 gene was not accurately known, it was obvious that they were considerably smaller than the ones in the glycinin gene. Since the position of the introns in the two genes appears to be identical (Nielsen pers.comm.), it is possible that both the soybean glycinin and pea legumin genes arose by

duplication from an ancestral gene. This type of relationship has been shown, amongst others, for the plant actin gene family (Shah et al 1983). The difference in intron length between the closely related glycinin and legumin genes may indicate that the parameters governing intron evolution are less stringent than those governing the evolution of structural coding sequences.

Five small introns have been detected in the gene for the french bean storage protein phaseolin, ranging in size from 72 to 108 bp (Sun et al 1981, Slightom et al 1983), four (between 85 to 132 bp) in the gene for soybean conglycinin (Schuler et al 1982) and three in the soybean leg haemoglobin gene (Hyldig - Nielsen et al 1982). However, the occurrence of introns in plant genes is not a universal phenomenom, since the genes for the maize storage protein zein (Hu et al 1982, Pedersen et al 1982, Wienand et al 1981, Pintor - Toro et al 1982), have been shown to lack introns. The genes for another soybean seed protein - lectin, and those for the soybean Kunitz trypsin inhibitor also contain no introns (Pedersen 1982, - pers. comm. Goldberg). Thus, the situation in plants with respect to intron occurrence within protein coding genes appears to be similiar to that found in animals, where some genes, notably those for histones (Kedes et al 1974) and interferons (Nagata et al 1980) are intron free, whilst others are split to varying extents (see later).

Sequencing was used as a final and definitive method of gene identification. The sequencing strategy used is shown in Fig 33 and a typical sequencing autoradiograph obtained for λ LEG3 is shown in Fig

36. A comparison of the legumin genomic sequence with the published legumin cDNA sequences shows that the amino acid coding regions between the two are almost identical, confirming that a legumin gene was indeed isolated (Figs 37 & 38). Although base differences betwen the cDNA sequence and the genomic sequence can be seen, further sequencing runs are required to confirm these differences. The sequenced region extends for some 57 bp (19 amino acids) beyond the 5′ end of pDUB 8 , approaching the N terminus of the α subunit, which is 84 bp further on (Casey et al 1979). Although it was not possible to obtain confirmatory cDNA sequence in this region due to the fact that pDUB 8 was not a complete cDNA, protein sequence for the first 12 amino acids 5′ to the end of p DUB8 (Casey et al 1979) indicated that the obtained legumin genomic sequence in this region was correct. As the sequence data did not pass the N terminus of the acidic subunit, it was not possible to speculate on the existence of a signal peptide region (Blobel and Dobberstein 1975). However, as phaseolin (Slightom et al 1983), vicilin (Lycett et al 1983), conglycinin (Schuler et al 1982), zein (Pedersen et al 1983), glycinin (Nielsen 1984) and gliadin, the wheat storage protein (Bartels and Thompson 1983), have been shown to possess signal peptides, it is likely that legumin, which is also a secretory protein, follows this trend. Furthermore, protein cell free translation products driven by poly A+ mRNA and microsomes indicate a leader sequence in legumin. At the 3′ end of the clone, sequence data were obtained past the C terminus of the basic subunit, into the 3′ non coding region. Polyadenylation signals were not found in this region, and it was be assumed that they were present further along into the unsequenced 3′ non coding region.

The occurence of two of the three postulated introns was

confirmed by comparison of the cDNA and genomic sequence data (Fig 37 and 38). Sequence data was not available to confirm the existence of the third intron, predicted by the fine mapping of the cDNA and the genomic clone. The 5′ end of both introns began with the dinucleotide GT, conforming to the universal GT / AG dinucleotide sequences observed for the 5′ and 3′ intron boundaries in eucaryotic protein coding genes (Breathnach et al 1978, Lerner et al 1980). The sequence around the 5′ splice junction in IVS 1 appeared to bear greater homology to the animal gene consensus sequence (Breathnach et al 1978), than to the plant gene consensus sequence (Slightom et al 1983) - Table 6. Plant donor 5′ sequences frequently show a strong preference for T rather than A in position - 2 (Slightom et al 1983), whereas both the pea legumin donor sequences have A in this position (Table 6). Also. the sequence of IVS 1 is unusual in having C instead of G in position + 5 - this is not shown by either the french bean phaseollin (Slightom et al 1983), soybean leg haemoglobin (Hyldig - Nielsen et al 1982), or the soybean actin genes (Shah et al 1982). However as the animal and plant consensus sequences are fairly homologous to each other, the difference was probably not significant. The similarity between the two consensus sequences also argues for a common RNA splicing mechanism between animal and plant species, perhaps mediated by the small nuclear RNAs (Lerner et al 1980) e.g. U 1 RNA (Reddy et al 1974). Further insights into the splicing mechanism will be obtained once splicing enzymes have been purified; indeed, splicing in vitro using isolated nuclei and cytoplasmic extracts has already been demonstrated (Hamada et al 1980, Goldenburg and Raskas 1980). It is important to note that the GT / AG intron / exon boundaries and associated consensus sequences are not found in split ribosomal (Bos et al 1980) or tRNA genes (references in

Abelson 1979), and thus it is likely that the processing of these genes requires splicing enzymes different from those which process protein coding genes.

The nucleotide composition of the intron IVS1 showed that it was very rich in A/T nucleotides which comprised 67 % of the sequenced length. In contrast, the A/T composition of the coding regions of the gene averaged 48 % . This is in agreement with the composition of other plant gene introns e.g. phaseolin, where the average A / T composition is 72.4 % (Slightom et al 1983). Soybean actin gene introns (Shah et al 1982) average 69.3 % A/T, while soybean leg haemoglobin gene introns (Hyldig – Nielsen et al 1982) have 76.8 % , and soybean glycinin gene introns are 72 % A/T rich. The A/T richness of plant introns is a feature in variance with animal introns, which generally have a much lower level of A / T content e.g. the human foetal globin genes average 55 % (Smithies et al 1981). The significance, if any, of this difference is not yet known.

Another difference between plant and animal genes lies in the number of intervening sequences. Animal storage protein genes, such as those for ovalbumin, have 7 introns (Benoist et al 1980), while vitellogenin genes have 33 introns (Wahli, Dawid(1980).The pro α 2 (1) collagen gene (Wozney et al 1981) has 50 introns. By contrast, the number of introns in plant genes ranges from 0 to 5 (see above). The reason for this difference is unknown, and it is possible that as more plant genes are cloned and sequenced, some will be found with many introns.

An interesting feature shown by the λ LEG 3 clone was the

occurence of sequences of tandem repeats in the region coding for the C terminal of the acidic legumin subunit (Fig 38). These repeats were imperfect, the first two showing more homology between each other than either showed to the third. Although sequencing of genomic repeat No 1 was incomplete, it is likely that the whole of the 54 bp repeat sequence shown by pDUB 8 was present, as homology between the rest of the repeat genomic sequence and the cDNA sequence was total. It is unlikely that these repeats represent some kind of cloning artefact, since actual amino acid sequence data in this region of the protein confirms the existence of part of the tandem repeat sequence (Lycett et al 1984). Since the previously isolated legumin cDNA pDUB 3 (Croy et al 1982) lacks the repeat region which clone pDUB 8 possesses, Lycett et al (1984a) put forward the suggestion that the existence of these two types of cDNA represented the existence of 2 classes of legumin gene, one class having the repeated region, and one class lacking it. The corn storage protein zein (Hu et al 1982) is another case where a sequence of repeats has been demonstrated in the gene. This repeat is a tandem duplication of a 96 bp sequence, which is interestingly absent from the cDNA probes used to isolate the genomic sequence. This situation is exactly analogous to the occurence of repeats in the legumin gene, and the absence of them from the cDNA. In contrast, repeated sequences have not been found in the soybean glycinin gene (Nielsen 1984).

Lycett et al (1984a) suggest that the existence of these two classes of legumin cDNA might account for the heterogeneity exhibited by the legumin molecule (Matta et al 1981). It is interesting that more heterogeneity is seen amongst the legumin acidic subunits (Matta

et al 1981), where the two proposed classes of legumin cDNA are said to differ, than amongst the basic subunit, where no repeat sequence is seen. It is obvious that confirmation of this theory depends upon the isolation of a legumin gene which is lacking the repeat sequence, and is homologous to p DUB3. Since the positions of the Eco R1 fragments within genomic DNA hybridising to legumin cDNAs are known, (Croy et al 1982, this thesis), a preliminary experiment to show the lack of repeats amongst some of the legumin genes would be to hybridise a labelled copy of the cDNA repeat sequence to a genomic digest of pea DNA. The theory would gain weight if some of the Eco R1 fragments hybridised to this sequence, whilst others did not. The finding that one of the genes coding for soybean glycinin did not contain a repeated sequence region (Nielsen 1984) lends support to the idea that two types of legumin gene exist. It is also possible as pointed out by Lycett et al(1984a), that the absence of repeats from pDUB 3 is a cloning artefact. If this is so, then only one class of gene, all containing the repeat sequence, will be isolated. The first member of this class is λ LEG3. The heterogeneity in the legumin acidic and basic subunits may however, still be due to an alteration in the nucleotide sequence of the various legumin genes, as analysis of the glycinin polypeptides revealed considerable sequence heterogeneity (Nielsen 1984). This means that more than one coding sequence must direct the synthesis of the glycinin subunit polypeptides. Nielsen (1984) speculated that the heterogeneity might reflect the existence of alleles of a small number of genes, since only 4 or 5 glycinin genes were shown to be present (Fischer and Goldberg 1982, Goldberg et al 1981). However, as shown in the present work, the number of genes

picked out in hybridisation studies depends on the type of probe used, and it is possible that more than 5 glycinin genes will eventually be shown to be present. This, combined with possible allelism and post translational modifications, probably accounts for most of the heterogeneity shown by both glycinin and legumin.

The patterns of codon usage in this gene are different from those shown by most animal messages, and follow the usage pattern of legumin cDNA (Lycett et al 1983). Since most of these differences are common to other plant genes, this may be a universal difference between plant and animal systems, but the significance of these differences is unknown.

This work has described the isolation of one of the genes coding for legumin, a major storage protein of Pisum sativum L. As the pea is a nutritionally important seed, the isolation of one of the genes coding for its seed proteins opens the way for direct manipulation of the protein quality , by alteration of the nucleotide sequence of the gene. In doing this, one must not grossly alter the tertiary structure of the new protein, since this might result in inefficent packaging of the altered protein into protein bodies. Before this can be done, however, significant advances in the stable inheritance and expression of transferred genes in the host plant have to be made. As the build up of storage proteins within the seed must be due to the action of highly efficent gene promoters, it is possible that these promoters and their associated sequences, will function as efficent transcribers of other plant genes , thus leading to their possible general use in plant gene expression systems.

REFERENCES

Aaij, C. and Borst, P. (1972), The gel electrophoresis of DNA. Biochem. Biophys. Acta. 269, 192-200.

Abelson,J. (1979), RNA processing and the intervening sequence problem. Ann. Rev. Biochem. 48, 1035-1069.

Adesink, M. and Darnell, J.E. (1972), Biogenesis and characterisation of histone mRNA in HeLa cells. J. Mol. Biol. 67 , 397-406.

Aviv, H. and Leder, P. (1972), Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid - cellulose. Proc. Nat. Acad. Sci. U.S.A. 69, 1408-1412.

Bachmann, B.J. and Low, K.B. (1980), Linkage map of E. coli K12. Microbiol. Rev. 44, 1-56.

Badenoch-Jones, J., Spencer, D., Higgins, T.J.V., Millerd, A. (1981). The role of glycosylation in storage protein synthesis in developing pea seeds. Planta 153, 201-209.

Bartels, D. and Thompson, R.D. (1983), The characterisation of cDNA clones coding for wheat storage protein. Nucl. Acids. Res. 11 2961-2976.

Basha, S.M.M. and Beevers, L. (1975), The development of proteolytic activity and protein degradation during the germination of Pisum sativum L. Planta 124, 77-87.

Baumgartner, B., Tokuyasu, K. T., Chrispeels, M.J. (1980), Immunocytochemical localisation of reserve protein in the endoplasmic reticulum of developing bean Phaseolus vulgaris L. cotyledons. Planta 150, 419-425.

Bedbrook, J.R., Jones, J., O'Dell, M., Thompson, R.D., Flavell, R.B. (1980), A molecular description of telomeric heterochromatin in Secale species. Cell. 19, 545-560.

Beevers, L. and Poulson, R. (1972), Protein synthesis in cotyledons of Pisum sativum L. Plant. Physiol. 49, 476-481.

Bennet, M.D. and Smith, D.B. (1976), Nuclear DNA amounts in angiosperms. Philos. Trans. R. Soc. London. B274, 227-273.

Benoist, C. and Chambon, P. (1981), In vivo sequence requirements of the Simian Virus 40 early promoter region. Nature (London) 290, 304-310.

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

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

Bernard, H.U. and Helinsky, D.R. (1981), Bacterial plasmid cloning vehicles. In " Genetic Engineering " Vol 2 133-168. Eds. Setlow and Hollaender. Plenum Press.

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

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., Smithies, O. (1977), Charon phages - safer derivatives of bacteriophage lambda for DNA cloning. Science 196, 161-169.

Blattner, F.R., Blechl, A.E., Thompson, K.D., Faber, H.E., Richards, J.E., Slightom, J.L., Tucker, P.W., Smithies, O. (1978), Cloning human fetal γ globin DNA: preparation and screening of shotgun collections. Science 202 1279-1284.

Blobel, G. and Dobberstein, B. (1975), Transfer of proteins across membranes 1. 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. and Backmann, K. (1979), Plasmids of E. coli as cloning vehicles. In " Methods in Enzymology "
Vol 68' 220-245 Ed. Grossmann and Moldave. Academic Press. New York.

Bollini, R. and Chrispeels, M.J. (1978), Characterisation and subcellular localisation of vicilin and phytohaemagglutinin, the two major reserve proteins of Phaseolus vulgaris L. Planta 142 291-298.

Bos, J.L., Osinga, K.A., Vanderho, G., Hecht, N.B., Tabak, H.F., Vanommen G.J., Borst, P. (1980), Splice point sequence and transcripts of the intervening sequence in the mitochondrial 21 S ribosomal RNA gene of yeast. Cell 20, 207-214.

Boulter, D., Ellis, R.J., Yarwood, A. (1972), Biochemistry of storage protein in plants. Biol. Rev. 47, 113-175.

Boulter, D. (1981), Biochemistry of storage protein synthesis and deposition in the developing legume seed. In Adv. Bot. Res. 9. 1-28

Brawerman, G. (1974), Eucaryotic messenger RNA. Ann. Rev. Biochem. 43, 621-642.

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

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

Briarty, L.G., Coult, D.A., Boulter, D. (1969), Protein bodies of developing seeds of Vicia faba. J. Exp. Bot. 20, 358-372.

Brinegar, A.C., and Petersen, D.M. (1982), Synthesis of oat globulin precursors. Analogy to legumin 11 S storage protein synthesis. Plant. Physiol. 70 1767-

Brisson, N. and Verma, D.P.S. (1982), Soybean leghaemoglobin gene family : normal, pseudo and truncated genes. Proc. Nat. Acad. Sci. U.S.A. 79, 4055-4059.

Britten, R.J. and Davidson, E.H. (1969), Gene regulation for higher cells - a theory. Science 165, 349-357.

Britten, R.J., and Davidson, E.H. (1976), DNA sequence arrangement and preliminary evidence on its evolution. Fed. Proc. 35, 2151-2157.

Britten, R.J., Graham, D.E., Neufeld, B.R. (1974) In " Methods in Enzymology " Vol 29, 363-418. Eds. Grossman and Moldave. Academic Press, New York.

Britten, R.J. and Kohne, D.E. (1968) Repeated sequences in DNA. Science 161, 529-540.

Caro, L.G. and Palade, G.E. (1964), Protein synthesis, storage, and discharge in the pancreatic exocrine cell. J. Cell. Biol. 20, 473-495.

Casey, R. (1979a) Genetic variability in the structure of the α subunits of legumin from Pisum, a 2 - dimensional gel electrophoretic study. Heredity 43, 265-272.

Casey, R. (1979b), Immunoaffinity chromatography as a means of purifying legumin from Pisum (pea) seeds. Biochem. J. 177, 509-520.

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

Casey, R., March, J.F., Sharman, J.E., 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. Biochem. Biophys. Acta. 670, 428-432.

Chilton, M.D. (1975), Ribosomal DNA in a nuclear satellite of tomato. Genetics 81, 469-483.

Chrispeels, M.J., Bollini, R., Harris, N. (1979), Biosynthesis, accumulation, and catabolism of reserve proteins in legume seeds. Ber. Dt. Bot. Ges. 92, 535-551.

Clarke, L. and Carbon, J. (1976), A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire E. coli genome. Cell 9, 91-99.

Clewell, D.B. (1972), Nature (London) of plasmid replication in E. coli in the presence of chloramphenicol. J. Bact. 110, 667-676.

Cocking, E.C., Davey, M.R., Pental, D., Power, J.B. (1981), Aspects of plant genetic manipulation. Nature (London) 293, 265-270.

Collins, J. and Bruning, H.J. (1978), Plasmids useable as gene cloning vectors in an *in vitro* packaging by coliphage λ : 'cosmids'. Gene *4*, 85-107.

Collins, J. and Hohn, B. (1979), Cosmids: a type of plasmid gene cloning vector that is packageable *in vitro* in bacteriophage λ heads. Proc. Nat. Acad. Sci. U.S.A. *75*, 4242-4246.

Conrad, S.E. and Botchan, M.R. (1982), Isolation and characterisation of human DNA fragments with nucleotide sequence homologies with the Simian Virus 40 regulatory region. Mol. Cell. Biol. *2*, 949-965.

Crain, W.R., Davidson, EH., Britten, R.J. (1976). Contrasting patterns of DNA sequencing arrangement in *Apis mellifera* (honey bee) and *Musca domestica* (house fly). Chromosoma *59*, 1-12.

Crouch, M.L., Tenbarge, K.M., Simon, A.E., Ferl, R. (1983), cDNA clones for *Brassica napus* seed storage proteins - evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. J. Mol. App. Genet. *2*, 273-283.

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

Croy, R.R.D., Gatehouse, J.A., Evans, I.M., Boulter, D. (1980), Characterisation of the storage protein subunits synthesised *in vitro* by polyribosomes and RNA from developing pea *Pisum sativum* L. 1. Legumin. Planta *148*, 49-56.

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

Croy, R.R.D., Shirsat, A., Boulter, D. (1984), Cloning and characterisation of genes for legumin from Pisum sativum L. Nucl. Acids. Res. (in press).

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

Dagert, M. and Ehrlich, S.D. (1979), Prolonged incubation in calcium chloride improves the competence of E. coli cells. Gene 6, 23-28.

Danielsson, C.E. (1949), Seed globulins of the Graminae and Leguminosae. Biochem. J. 44, 387-400.

Darnell, J.E. (1979), Transcription units for mRNA. Prog. Nucleic. Acid. Res. Mol. Biol. 22, 327-353.

Davidson, E.H. and Britten, R.J. (1979), Regulation of gene expression : possible role of repetitive sequences. Science 204, 1052-1059

Davidson, E.H., Galau, G.A., Angerer, R.C., Britten, R.J. (1975), Comparative aspects of DNA organisation in metazoa. Chromosoma 51, 252-259.

Davidson, E.H., Jacobs, H.T., Britten, R.J. (1983), Very short repeats and coordinate induction of genes. Nature (London) 301, 468-470.

Denhardt, D.T. (1966). A membrane filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. 23, 641-646.

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

Domoney, C., Davies, D.R., Casey, R. (1980), The initiation of legumin synthesis in immature embryos of Pisum sativum L. grown in vivo and in vitro. Planta 149, 454-460 .

Donahue, T.F., Daves, R.S., Lucchini, G., Fink, G.R. (1983), A short nucleotide sequence required for regulation of HIS 4 by the general control system of yeast. Cell 32, 89-98.

Doolittle, W.F. and Sapienza, C. (1980), Selfish genes, the phenotype paradigm and genome evolution. Nature (London) 284, 601-603.

Dretzen, G., Bellard, M., Sassone-Corsi, P., Chambon, P. (1981), A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. 112, 295-296.

Dugaiczyk, A., Boyer, H.W., Goodman, H.M. (1975), Ligation of Eco R1 endonuclease generated DNA fragments into linear and circular structures. J. Mol. Biol. 96, 171-184.

Dure III, L.S. (1975), Seed formation. Ann Rev. Plant. Physiol. 26, 259-278.

Echols, H. and Murialdo, H. (1978). Genetic map of bacteriophage λ. Microbiol. Rev. 42, 577-591.

Edens, L., Heslinga, L., Klok., Ledeboer, A.M., Maat, J., Toonen, M.Y., Visser, C., Verrips, C.T. (1982), Cloning of cDNA encoding the sweet tasting plant protein thaumatin and its expression in E. coli. Gene 18, 1-12.

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

Evans, I.M., Croy, R.R.D., Brown, P., Boulter, D. (1980), Synthesis of complementary DNAs to partially purified mRNAs coding for the storage proteins of Pisum sativum L. Biochim. Biophys. Acta. 610, 81-95.

Fangman, W.L. (1978), Separation of very large DNA molecules by gel electrophoresis. Nucl. Acids. Res. 5, 653-665.

Feiss, M., Fischer, R.A., Crayton, M.A., Egner, C. (1977), Packaging of the bacteriophage λ chromosome : effect of chromosome length. Virology 77, 281-293.

Fischer, R.L. and Goldberg, R.B. (1982), Structure and flanking regions of soybean seed protein genes. Cell, 29, 651-660.

Flavell, R.B. and Smith, D.B. (1976), Nucleotide sequence organisation in the wheat genome. Heredity. 37 231-252.

Flavell, R.B. (1980). The molecular characterisation and organisation of plant chromosomal DNA sequences. Ann. Rev. Plant. Physiol. 31, 569-596.

Food and Agriculture Organisation of the United Nations (1970), Amino acid content of foods and biological data on proteins. Nutritional Studies, 24, Rome.

Galau, G.A. and Dure, L, III. (1981), Developmental biochemistry of cotton seed embryogenesis and germination : changing messenger RNA populations as shown by reciprocal heterologous complementary DNA hybridisation. Biochemistry 20, 4169-4178.

Galau, G.A., Legocki, A.B., Greenway, S.C., Dure, L, III. (1981), Cotton mRNA sequences exist in both polyadenylated and non polyadenylated forms. J. Biol. Chem. 256, 2551-2560.

Gatehouse, J.A., Croy, R.R.D., McIntosh, R., Paul, C., Boulter, D. (1980a), Quantitative and qualitative variation in the storage proteins of material from the EEC joint field bean test. In " Vicia faba feeding value, processing and viruses". Bond, D.A. ed. 173-190. E.C.S.C., E.E.C., E.A.E.C. Brussels, Luxembourg.

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

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

Gatehouse, J.A., Lycett, G.W., Delauney, A.J., Croy, R.R.D., 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.

Gerlach, W.L., Pryor, A.J., Dennis, E.S., Ferl, R.J., Sachs, M.C., Peacock, W.J. (1982), cDNA cloning and induction of the alcohol dehydrogenase gene (Adh 1) of maize. Proc. Nat. Acad. Sci. U.S.A. 79, 2981-2985.

Gillies, S.D., Morrison, S.L., Oi, V.T., Tonegawa, S. (1983), A tissue specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33, 717-728.

Goldberg, R.B., Hoschek, G., Kamalay, J.C. (1978), Sequence complexity of nuclear and polysomal RNA in leaves of the tobacco plant. Cell 14, 113-131.

Goldberg, R.B., Hoschek, G., Tam, S.H., Ditta, G.S., Briedenbach, R.W. (1981), Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis. Dev. Biol. 83, 201-217.

Goldenberg, C.J., and Raskas, H.J.(1980), In vitro processing of intervening sequences in the precursors of messenger ribonucleic acid for adenovirus 2 deoxyribonucleic acid binding protein. Biochemistry 19, 2719-2723.

Gordon, M.E. and Payne, P.I. (1976), In vitro translation of the long lived messenger ribonucleic acid of dry seeds. Planta 130, 269-273.

Graf, H. (1979), Optimisation of conditions for the in vitro formation of hybrid DNA molecules by DNA ligase. Biochim. Biophys. Acta. 564, 225-234.

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

Gray, R.E., and Cashmore, A.R. (1976), RNA synthesis in plant leaf tissue, the characterisation of messenger RNA species lacking and containing polyadenylic acid. J. Mol. Biol. 108, 595-608.

Grosschedl, R., Wasylyk, B., Chambon, P., Birnsteil, M.L. (1981), Point mutation in the TATA box curtails expression of sea urchin H2 A histone genes in vivo. Nature (London) 294, 178-180.

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

Hall, T.C. (1979), Plant mRNA. In " Nucleic Acids In Plants" Vol. 1 217-251. Eds. Hall, T.C. and Davies, J.W. CRC press.

Hall, T.C., Slightom, J.L., Ersland, D.R., Murray, M.G., Hoffman, L.M., Adang, M.J., Brown, J.W.S., Ma, Y., Matthews, J. A., Cramer, J.H., Barker, R.F., Sutton, D.W., Kemp, D.J. (1983), Phaseolin. Nucleotide sequence explains molecular weight and charge heterogeneity of a small multigene family and also assists vector construction for gene expression in alien tissue. In " Structure and function of plant genomes " Eds. Ciferri and Dure. New York. Plenum Press 1984.

Hamada, H., Igarashi, T., Muramatsu, M.(1980), In vitro splicing of Simian Virus 40 mRNA in isolated nuclei from CV-1 cells. Nucl. Acids. Res. 8, 587-599.

Heinze, H., Herzfeld, F., Kiper, M. (1980), Light induced appearance of polysomal poly A rich messenger RNA during greening of barley plants. Eur. J. Biochem. 111, 137-144.

Hentschel, C., Irminger, J.C., Bucher, P., Birnsteil, M.L. (1980), Sea urchin histone mRNA termini are located in gene regions downstream from putative regulatory sequences. Nature (London) 285, 147-151.

Hershfield, V., Boyer, H.W., Yanofsky, C., Lovatt, N., Helinski, D.R. (1974), Plasmid Col E1 as a molecular vehicle for cloning and amplification of DNA. Proc. Nat. Acad. Sci. U.S.A. 71, 3455-3459.

Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C., Spencer, D. (1983a), The biosynthesis and primary structure of pea seed lectin. J. Biol. Chem. 258, 9544-9549.

Higgins, T.J.V., Chrispeels, M.J., Chandler, P.M., Spencer, D. (1983b), Intracellular sites of synthesis and processing of lectin in developing pea cotyledons. J. Biol. Chem. 258, 9550-9552.

Hohn, B. (1975), DNA as substrate for packaging into bacteriophage lambda in vitro. J. Mol. Biol. 98, 93-106.

Hohn, B. (1979), In vitro packaging of λ and cosmid DNA. In " Methods in Enzymology " Vol 68 299-309.Eds. Grossman and Moldave. Acafemic Press. New York.

Hohn, B. and Collins, J. (1980), A small cosmid for efficent cloning of large DNA fragments. Gene 11 291-298.

Hohn, B. and Hinnen, A. (1980), Cosmids in E. coli and yeast. In " Genetic Engineering " Vol 2 169-183. Eds. Setlow and Hollaender, Plenum Press.

Hohn, B. and Hohn, T. (1974), Activity of empty, headlike particles for packaging of DNA into bacteriophage λ in vitro. Proc. Nat. Acad. Sci. U.S.A. 71, 2372-2376.

Hohn, T. and Katsura, I. (1977), Structure and assembly of bacteriophage lambda. Curr. Top. Microbiol. Immunol. 78, 69-110 .

Hohn, B. and Murray, K. (1977), Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc. Nat. Acad. Sci. U.S.A. 74, 3259-3263.

Hu, N., Peifer, M.A., Heidecker, G., Messing, J., Rubenstein, I. (1982), Primary structure of a genomic zein sequence of maize. The EMBO J 1, 1337-1342.

Huez,G., Marbaiy, G., Gallwitz, D., Weineberg, E., Devos, R., Hubert, e., Cleuter, Y. (1978), Functional stabilisation of HeLa cell histone mRNAs injected into Xenopus oocytes by 3′ OH polyadenylation. Nature (London) 271, 572-573.

Hyldig-Nielsen, J.J., Jensen, E.O., Paludan, K., Wiborg, O., Garret, R., Jorgensen, O.P., Marcker, K.A. (1982), The primary structure of two leghaemoglobin genes from soybean. Nucl. Acids. Res. 10, 689-701.

Ingle, J., Pearson, G.G., Sinclair, J. (1973), Species distribution and properties on nuclear satellite DNA in higher plants. Nature (London) New. Biol. 242, 193-197.

Ingle, J. and Sinclair, J. (1972), Ribosomal RNA genes and plant development. Nature (London) 235, 30-32.

Ingle, J., Timmis, J.N., Sinclair, J.(1975), Relationship between satellite DNA, ribosomal RNA gene redundancy and genome size in plants. Plant Physiol. 55, 496-501.

Jelinek, W.R., Toomey, T.P, Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M, Deininge, P.L. (1980), Ubiquitous, interspersed, repeated sequences in mammalian genomes. Proc. Nat. Acad. Sci. U.S.A. 77, 1398-1402.

Jensen, O.E., Paludan, K., Nielsen, H.J.J., Jorgensen, P., Marcker, K.A.(1981), The structure of a chromosomal leghaemoglobin gene from soybean. Nature (London) 291, 677-679.

Kahn, M., Kolter, P., Thomas, C., Figurski, D., Meyer, R., Remaut, E., Helinski, D.R. (1979), Plasmid cloning vehicles derived from plasmids Col E1, F, R 65, and RR 2. In " Methods in Enzymology " Vol. 68 268-281.

Kamalay, J.C. and Goldberg, R.B. (1980), Regulation of structural gene expression in tobacco. Cell 19, 935-946.

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

Katinakis, P.K., Slater, A., Burdon, R.H. (1980), Non polyadenylated mRNAs from eucaryotes. FEBS Lett. 116, 1-7.

Katz, L., Williams, P.H., Sato, H., Leavitt, R.W, Helinsky, D.R. (1977), Purification and characterisation of covalently closed replicative intermediates of Col E1 DNA from E. coli. Biochemistry 16, 1677-1683.

Kedes, L.H. (1976), Histone messengers and histone genes. Cell 8, 321-331.

Kemp. D.J., Cory, S., Adams, J.M. (1979), Cloned pairs of variable region genes for immunoglobulin heavy chains isolated from a clone library of the entire mouse genome. Proc. Nat. Acad. Sci. U.S.A. 76, 4627-4631.

Khoury, G. and Gruss, P. (1983), Enhancer elements. Cell 33, 313-314.

Kiper, M. and Herzfeld, F. (1978), DNA sequence organisation in the genome of Petroselenum sativum. Chromosoma 65, 335-351.

Kiper, M., Bartels, D., Herzfeld, F., Richter, G. (1979), The expression of a plant genome in hnRNA and mRNA. Nucl. Acids. Res. 6, 1961-1978.

Klotz, L.C., and Zimm, B.H. (1972), Size of DNA determined by viscoelastic measurements : results on bacteriophages, Bacillus subtilis, and Escherichia coli. J. Mol. Biol. 72, 779-800.

Korn, L.J. and Gurdon, J.B. (1981), The reactivation of developmentally inert 5S genes in somatic nuclei injected into Xenopus oocytes. Nature (London) 289, 461-465.

Kozak, M. (1980), Evaluation of the scanning model for initiation of protein synthesis in eucaryotes. Cell 22, 7-8.

Krishna, T.G., Croy, R.R.D., Boulter, D. (1979), Heterogeneity in subunit composition of the legumin of Pisum sativum L. Phytochemistry 18, 1879-1880.

Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., Maniatis, T. (1978), The isolation and characterisation of linked and β globin genes from a cloned library of human DNA. Cell, 15, 1157-1174.

Leder, P., Tiemier, D., Enquist, L. (1977), EK 2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms : the λ gtWES system. Science 196, 175-177.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.A., Steitz, J.A. Are snRNAs involved in splicing? Nature (London) 283, 220-224.

Lewin, B. (1975). Units of transcription and translation : the relationship between heterogeneous nuclear RNA and messenger RNA. Cell 4, 11-20.

Lewis, E.D., Hagen, G., Mullins, J.I, Mascia, P.N., Park, W.D., Benton, W.D., Rubenstein, I. (1981), Cloned genomic fragments of Zea mays homologous to zein mRNAs. Gene 14, 205-215.

Lindahl, G., Sironi, G., Bialy, H., Calendar, R. (1970). Bacteriophage lambda; abortive infection of bacteria lysogenic for phage P2. Proc. Nat. Acad. Sci. 66, 592-596.

Loenen, W.A.M. and Brammar, W.J. (1980), A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene 20, 249-259.

Lycett, G.W., Delauney, A.J., Gatehouse, J.A., Gilroy, J., Croy, R.R.D., Boulter, D. (1983a), The vicilin gene family of pea (Pisum sativum L.); a complete cDNA coding sequence for preprovicilin. Nucl. Acids. Res. 11, 2367-2379.

Lycett, G.W., Delauney, A.J., Croy, R.R.D. (1983b). Are plant genes different? FEBS Lett. 153, 43-45.

Lycett, G.W., Delauney, A.J., Zhao, W., Gatehouse, J.A., Croy, R.R.D, Boulter, D. (1984a). Two cDNA clones coding for the legumin protein of Pisum sativum L. contain sequence repeats. Plant. Mol. Biol. (in press).

Lycett, G.W., Croy, R.R.D., Shirsat, A., Boulter, D. (1984b). The complete nucleotide sequence of a legumin gene from pea (Pisum sativum L.). Nucl. Acids. Res. (submitted).

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

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connel, C., Quon, D. (1978), The isolation of structural genes from libraries of eucaryotic DNA. Cell 15, 687-701.

Mans, S.R.J. (1967), Protein synthesis in higher plants. Ann. Rev. Plant. Physiol. 18, 127-146.

Mathis, D.J. and Chambon, P. (1981), The SV 40 early region TATA box is required for accurate in vitro initiation of transcription. Nature (London) 290, 310-315.

Matsubara, K. and Takeshi, M. (1982), Identification of sequences necessary for packaging DNA into lambda phage heads. Gene 20, 267-279.

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

Matta, N.K. and Gatehouse, J.A. (1982), Inheritance and mapping of storage protein genes in Pisum sativum L. Heredity 48, 383-392.

Maxam, A.M. and Gilbert, W. (1977), A new method for sequencing DNA. Proc. Nat. Acad. Sci. 74, 560-564.

Meincke, D.W., Chen, J., Beachy, R.N. (1981), Expression of storage protein genes during soybean development. Planta 153, 130-139.

Messing, J., Geraghty, D., Heidecker, G., Hu, N.T., Kridl, J., Rubenstein, I. (1983), Plant gene structure. In " Genetic engineering of plants" ed. Meredith, C.P., Kosuge, T., Hollaender, A.

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

Minty, A. and Newmark, P. Gene regulation : new, old and remote controls. Nature 288, 210-211.

Mizusawa, S. and Ward, D.F. (1982), A bacteriophage vector for cloning with Bam H1 and Sau 3A. Gene 20, 317-322.

Morton, H., Evans, I.M., Gatehouse, J.A., Boulter, D. (1983), Sequence complexity and abundance classes in messenger RNA from cotyledons of deueloping pea (Pisum sativum L.) seeds. Phytochemistry 22, 807-812.

Murray, M.G., Cuellar, R.E., Thompson, W.F., (1978), DNA sequence organisation in the pea genome. Biochemistry 17, 5781-5790.

Murray, M.G., Peters, D.L., Thompson, W.F. (1981), Ancient repeated sequences in the pea and mung bean genomes and implications for genome evolution. J. Mol. Biol. 17, 31-42.

Murray, M.G. and Thompson, W.F. (1982), Repeat sequence interspersion in coding DNA of peas does not reflect that in total pea DNA. Plant. Mol. Biol. 1, 143-153.

Murray, N.E., Brammar, W.J., Murray, K. Lambdoid phaged that simplify recovery of in vitro recombinants. Mol. Gen. Genet. 150, 53-61

Murray, K. and Murray, N.E. (1975). Phage lambda receptor chromosomes for DNA fragments made with restriction endonuclease 3 of Haemophilus influenzae and restriction endonuclease 1 of Escherichia coli.

Nagao, R.T., Shah, D.M., Eckenrode, V.K., Meagher, R.B. (1981), Multigene family of actin related sequences isolated from a soybean genomic library. DNA 2, 1-9.

Nagata,S., Mantei, M., and Weissmann. C. (1980), The structure of one of eight or more distinct chromosomal genes for human interferon. Nature (London) 287, 401-408.

Nelson, O.E. (1980), Advances in cereal science and technology. 3, 41.

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

Orgel, L.E. and Crick, F.H.C. (1980), Selfish DNA: the ultimate parasite. Nature (London) 284, 604-607.

Osborne, T.B. (1924), "The vegetable proteins" Longmans, Green and Company. London.

Palmiter, R.D. (1972), Regulation of protein synthesis in chick oviduct. J. Biol. Chem. 247, 6450-6459.

Pedersen, K., Devereux, J., Wilson, D.R, Sheldon, E., Larkins, B.A. (1982), Cloning and sequence analysis reveal structural variation among related zein genes in maize. Cell 29, 1015-1026.

Pintor-Toro, J.A., Langridge, P., and Felix, G. (1982), Isolation and characterisation of maize genes coding for zein proteins of the 21,000 dalton size class. Nucl. Acids. Res. 10, 3845-3860.

Ragg, H. and Weissmann, C. (1983), Not more than 117 base pairs of 5′ flanking sequence are required for inducible expression of a human ifn - α gene. Nature (London) 303, 439-442.

Reddy, R., Ro-Choi T.S., Henning, D., Busch, H. (1974), Primary sequence of U 1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. J. Biol. Chem. 249, 6486-6494.

Rimm, D.L., Horness, D., Kucera, J., Blattner, F.R. (1980), Construction of coliphage lambda Charon vectors with Bam HI cloning sites. Gene 12, 301-309.

Roberts, R.J. (1981), Restriction and modification enzymes and their recognition sequences. Nucl. Acids. Res. 9, 75-96.

Robbins, J., Rosteck, P., Haynes, J.R., Freyer, G., Cleary, M.L., Kalter, H.D., Smith, K., Lingrell, J.B. (1979), The isolation and partial characterisation of recombinant DNA containing genomic globin sequences from the goat. J. Biol. Chem. 254, 6187-6195.

Rogers, J.C. and Milliman, C. (1983), Isolation and sequence analysis of a barley α amylase cDNA clone. J. Biol. Chem. 258, 8169-8174.

Ross, J., Gielen, J., Packman, S., Ikawa, Y., Leder, P. (1974), Globin gene expression in cultured erythroleukaemic cells. J. Mol. Biol. 87, 697-714.

Schroeder, H.E. (1982), Quantitative studies on the cotyledonary proteins in the genus Pisum. J. Sci. Food. Agric. 33, 623-633.

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

Seeburg, P.H., Shine, J., Marshall, J.A., Baxter, J.D., Goodman, H.M. (1977), Nucleotide sequence and amplification in bacteria of the structural gene for rat growth hormone. Nature (London) 270, 486-499.

Seed, B., Parker, R., Davidson, N. (1982), Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. Gene 19, 201-209.

Sengupta, C., Deluca, V., Bailey, D.S., Verma, D.P.S. (1981), Post translational processing of 7S and 11S components of soybean storage proteins. Plant. Mol. Biol. 1, 19-34.

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

Shah, D.M., Hightower, R.C., Meagher, R.B. (1983), Genes encoding actin in higher plants : intron positions are highly conserved but the coding sequences are not. J. Mol. Appl. Genet. 2, 111-126.

Sharp, P.A., Sugden, B., Sambrook, J. (1973), Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose. Biochemistry. 12, 3055-3063.

Shatkin, A.J. (1976), Capping of eucaryotic mRNAs. Cell 9, 645-653.

Sherrat, D.J. (1979), Plasmid vectors for genetic manipulation in vitro. Biochem. Soc. Symp. (1979), 44, 29-36.

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

Seif, I., Khoury, G., Dhar, R. (1980), A rapid enzymatic DNA sequencing technique : determination of sequence alterations in early SV 40 temperature sensitive deletion mutants. Nucl. Acids. Res. 8, 2225-2240.

Skinner, D.M. (1977), Satellite DNAs. Bioscience 27, 790-796.

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

Smith, D.F., Searle, P.F., Williams, J.G. (1979), Characterisation of bacterial clones containing DNA sequences derived from Xenopus vitellogenin mRNA. Nucl. Acids. Res. 6, 487-506.

Smithies, O., Engels, W.R., Devereux, J.R., Slightom, J.L, Shen, S. (1981), Base substitutions, length differences, and DNA strand asymmetries in the human γ and γ fetal globin genes. Cell 26, 345-353.

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

Spencer, D. and Higgins, T.J.V. (1979), Molecular aspects of seed protein biosynthesis. Curr. Adv. Pl. Sci. 34, 1-13.

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

Sternberg, N., Tiemeier, D., Enquist, L. (1977), In vitro packaging of a λ Dam vector containing fragments of E.coli and phage P1. Gene 1, 255-280.

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

Suzuki, Y., Gage, P., Brown, D.D.(1972), The genes for silk fibroin in Bombyx mori. J. Mol. Biol 70, 637-649.

Thomas, M., Cameron, J.R., Davis, R.W. (1974), Viable molecular hybrids of bacteriophage lambda and eucaryotic DNA. Proc. Nat. Acad. Sci. U.S.A. (1974), 71, 4579-4583.

Thomas, P.S. (1980), Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Nat. Acad. Sci. U.S.A. 77, 5202-5205.

Thompson, J.A., Millerd, A., Schroeder, H.E. (1979), Seed protein improvement in cereals and grain legumes. Proc. Symp. Neuberberg. 1, 231-240.I.A.E.A. Vienna.

Thompson, J.A., Schroeder, H.E., Dudman, W.F. (1978), Cotyledonary storage proteins in Pisum sativum L. 1) Molecular heterogeneity. Aust. J. PL. Physiol. 5, 263-279.

Thompson, W.F. and Murray, M.G.(1981), The nuclear genome: structure and function. In " The biochemistry of plants " Vol 16, ed. Marcus, A. Academic Press, London.

Viera, J. and Messing, J. (1982), The pUC plasmids, an M13 mp 7 derived system for insertion mutagenesis and sequencing with universal primers. Gene 19, 259-268.

Volloch, V, and Housman, D. (1981), Stability of globin mRNA in terminally differentiating murine erythroleukaemia cells. Cell. 23, 509-514.

deVries, S.C., Springer, J., Wessels, J.G.H. (1983), Sequence diversity of polysomal mRNAs in roots and shoots of etiolated and greened pea seedlings. Planta 158, 42-50.

Wahli, W. and Dawid, I.G. (1980), Comparative analysis of the structural organisation of two closely related vitellogenin genes in Xenopus laevis. Cell 20, 107-117.

Walbot, V. and Dure, L.S. (1976), Developmental biochemistry of cotton seed embryogenesis and germination. J. Mol. Biol. 101, 503-506.

Wall, L.V.M. and Bryant, J.A. (1981), Isolation and preliminary characterisation of cryptic satellite DNA in pea. Phytochem. 20, 1767-1771.

Weber, E. et al (1981), Transfer of in vitro synthesised Vicia faba globulins and barley prolamins across the endoplasmic reticulum membrane of Vicia faba. Carlsberg. Res. Comm. 46, 383

Weiber, H., Konig, M., Gruss, P. (1983), Multiple point mutations affecting the Simian Virus 40 enhancer. Science 219, 626-631.

Wensink, P.C., Finnegan D.J., Donelson, J.E., Hogness, D.S.(1974), A system for mapping DNA sequences in the chromosomes of Drosophila melanogaster. Cell 3, 315-326.

Wienand, U., Langridge, P., Felix, G. (1981), Isolation and characterisation of a genomic sequence of maize coding for a zein gene. Mol. Gen. Genet. 182, 440-444.

Williams, B.G. and Blattner, F.R. (1979), Construction and characterisation of the hybrid bacteriophage lambda charon vectors for DNA cloning. J. Virol. 29, 555-575.

Wiskocil, R., Bensky, P., Dower, W., Goldberger, R.F., Gordon, J.I. (1980), Coordinate regulation of two estrogen dependent genes in avian liver. Proc. Nat. Acad. Sci. U.S.A. 77, 4474-4478.

Wozney, J., Hanahan, D., Tate, U., Boedtker, H., Doty, P. (1981), Structure of the pro α 2 (I) collagen gene. Nature (London) 294, 129-135.

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

Yamagata, H., Sugimoto, T., Tamaka, K., Kasai, Z. (1982), Biosynthesis of storage proteins in developing rice seeds. Plant. Physiol. 70, 1094-1100.

Yamamoto, K.R., Alberts, B.M., Benzinger, L., Lawhorne, L., Treiber, G. (1970), Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large scale virus purification. Virology, 40, 734-744.