Construction of a cDNA library encoding pea seed proteins

Laurene Neil Gatehouse

CONSTRUCTION OF A cDNA LIBRARY ENCODING PEA SEED PROTEINS

A Thesis submitted by :-

LAURENCE NEIL GATEHOUSE BSc.(Hons)

University of Leeds

In accordance with the requirements of the University of Durham, for the degree of Master of Science.

Fe ddechreuwn a disgled o dê

A.Plater

Department of Botany.

December, 1985

Construction of a cDNA library encoding pea seed proteins.

LAURENCE NEIL GATEHOUSE

ABSTRACT

A cDNA library was constructed using mRNA isolated from the cotyledons of developing seeds of pea (Pisum sativum L.) 13-17 days after flowering, the mid-stage of development. The library was systematically screened for clones encoding most of the more abundant seed proteins. Representative clones of each class were isolated and characterised by restriction mapping, hybridisation experiments and in some cases DNA sequencing. The abundances of the various classes of clones were correlated with the known abundance classes of mRNA sequences at this stage of development and shown to fit well.

Two types of cDNA clone corresponding to legumin were found. One type was found to have similar restriction maps to legumin genes A, B and C which encode "major" legumin polypeptides. A representative of the other type of legumin clone was shown to correspond to legumin gene J which encodes a "minor" legumin polypeptide. Restriction mapping was used to show that genes in this subfamily are more diverse than in the Leg A,B,C subfamily.

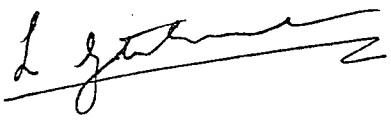
Clones for the three main vicilin types were found. From their abundances and cross hybridisations, it was shown that the three vicilins are roughly equally homologous to each other and that their mRNAs are roughly equally abundant at this stage of development.

A clone was found which encodes one of the major albumins and was completely DNA sequenced. The predicted protein sequence thus obtained was shown to match well with the partial protein sequences previously obtained from total major albumins.

Clones for the soluble seed lectin were found and the DNA sequence of one was shown to be identical with the published sequence of pea lectin mRNA. However its poly A tail started 24 bases further downstream, after a series of overlapping sequences closely homologous to the consensus polyadenylation signal. The size of the lectin mRNA was found to be ~1000 bases, confirming the near completeness of the published sequence.

DECLARATION

No part of this work has been submitted by me for any degree at this or any other University. All the work presented was done by me except where otherwise stated in the text.



L.N. Gatehouse.

CONTENTS

Contents Contd/...

Contents Contd/...

Contents Contd/...

Contents Contd/...

Contents Contd/...

LIST OF FIGURES:

List of Figures Contd/..

LIST OF TABLES

Tables Contd/... Page:

Tables Contd/... Page:

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Professor Donald Boulter, for his guidance, advice and encouragement throughout this project. I also wish to thank Dr. R.R.D. Croy and Dr. J.A. Gatehouse for their inspiration, guidance and many helpful suggestions.

Special thanks are also due to; Dr. M. Evans for critically reading this thesis and her cDNA synthesis protocol; Mr. R. Swinhoe for the kind gift of the poly(A)$^+$ RNA; Mr. D. Bown, Mrs. M. Richards, Mrs. P. Brown and Mr. T. Pickard for all the techniques they taught me; Dr. G. Lycett for his end-labelling dideoxy sequencing protocol; Dr. A.J. Delauney for his excellent advice and inspiration; Mrs. J. Scott for typing this thesis; Mr. I. Burrows for his unstinting help with the main frame computer and many other things; Mr. R.A. Ragab for his albumins probe; Mr. J. Gilroy for his protein sequences; Mr. P. Sidney for the photographs; Dr. R. Waterhouse for many interesting discussions; Dr. A.M.R. Gatehouse for moral support and Miss. J.S. Barker for encouraging me to start this work.

Finally I wish to thank my family, my friends, and the other members of the department for their help and support.

ABBREVIATIONS.

The following abbreviations were used:

Alecs gel buffer	= 40mM Tris-acetic acid pH7.7,2mMEDTA
AMV	= Avian Myeloblastosis Virus (reverse transcriptase).
ATP	= Adenosine Triphosphate.
β-NAD	= β-Nicotinamide Adenine Dinucleotide.
bp	= base pairs.
BSA	= Bovine Serum Albumin.
cDNA	= complementary/copy DNA
CTP	= Cytidine Triphosphate.
d	= deoxy-
dd	= dideoxy-
DMF	= Dimethylformamide.
DMSO	= Dimethyl Sulphoxide.
DNA	= Deoxy-ribonucleic Acid.
ds	= double-stranded.
DTT	= Dithiothreitol.
E.coli	= *Escherichia coli*
EDTA	= Ethylene Diamine Tetra-acetic Acid.
GTP	= Guanosine Triphosphate.
HCl	= Hydrochloric acid.
H.p.l.c	= High performance liquid chromatography.
HPLC grade	= Double distilled and deionised.
IPTG	= Isopropyl-β-D-thiogalactopyranoside.
kb	= kilo base pairs.
lac	= β-galactosidase gene system.
L Agar	= L Broth without glucose and plus 15gm/ litre Bacto-Agar.
L Broth	= Bacto-tryptone 10g, Yeast Extract 5g, Sodium Chloride 5g, Glucose 1g to 1 litre with water.
Miniprep	= Miniature preparation.
Minipreped	= Prepared by the miniprep method.
Mr	= Molecular Weight.
mRNA	= messenger RNA.
O.D.	= Optical Density.
PEG6000	= Polyethylene glycol 6000
pLGA,B	= plasmid L.Gatehouse Filter A, Colony Number B.
PMA-L	= Pea Major Albumin Large

contd/...

Abbreviations contd/...

PMA-S	= Pea Major Albumin Small.
PNK Buffer	= 50mM Tris-HCl pH7.5, 10mM magnesium chloride, 10mM DTT.
Poly A	= Polyadenylic acid.
Poly(A)$^+$	= Retained by an oligo (dT) cellulose column.
RF	= Replicative Form.
RNA	= Ribonucleic Acid.
RNAsein	= RNAase inhibitor.
S	= Svedburgs.
ss	= single-stranded.
SDS	= Sodium Dodecyl-sulphate.
SSC	= 0.15M sodium chloride, 15mM sodium citrate to pH7.0 with sodium hydroxide.
SSPE	= 0.18M sodium chloride, 10mM sodium dihydrogen phosphate pH7.4, 1mM EDTA.
TBE buffer	= 10.8gm Tris, 5.5gm Boric acid, 0.93gm EDTA ($Na_2.2H_2O$) to 1 litre pH$\sim$8.3.
TCA	= Trichloroacetic Acid.
TdT	= Terminal deoxynucleotidyl Transferase.
T.E. Buffer	= 10mM Tris-HCl pH7.5, 1mM EDTA.
Tris	= Tris (hydroxymethyl)-aminomethane.
tRNA	= transfer RNA.
TTP	= Thymidine Triphosphate.
u	= units (of enzyme)
µ	= micro
UV	= Ultra Violet.
v/v	= volume per volume.
w/v	= weight per volume.
X-Gal	= 5-Bromo-4-chloro-3indolyl-β-D-galactopyranoside.
YT Agar	= YT broth plus 15gm/l Bacto-Agar.
YT Broth	= Bactotryptone 8g, Yeast Extract 5g, Sodium Chloride 5g, to 1 litre with water.

CHAPTER 1

INTRODUCTION

1.1. General Introduction

Peas are members of the family of Leguminosae, a group of plants with properties of enormous importance to man. Firstly they, along with the cereals, directly supply about 70% of mankinds dietary protein (Oram and Brock,1972), most of the rest coming from animals which are often fed with seed meals. Although cereals (8-15% dry weight protein content) account for more overall protein consumption, legumes (20-50% dry weight protein content (Danielson, 1949)), are a concentrated protein source and are very important in providing an adequate diet in developing countries where protein malnutrition is common (Shewry et.al., 1981).

Secondly, legumes by means of a symbiotic relationship between their root nodules and various strains of Rhizobium bacteria, are able to fix atmospheric nitrogen. This eliminates the need for the addition of relatively expensive and sometimes harmful nitrogen fertilizers (Nitrates in the water supply and algal "bloom" in rivers.). In addition when grown in rotation or in mixed crops, such as the maize, squash, Phaseolus bean system of the red indians of the North American southwest, legumes maintain or even improve soil fertility over many growing seasons. Nutritionally animals require in their diets sufficient protein to supply the required amounts of essential amino acids as the "raw materials" for protein synthesis. Nine of the twenty protein amino acids are essential for mammalian, and hence human, nutrition. These are leucine, isoleucine, lysine, methionine, phenylalanine, threonine, tryptophan, tyrosine and valine. Legume seed storage proteins are in

general limiting in the sulphur amino acids methionine and cysteine (the latter has a sparing effect on methionine), where-as cereals are usually limiting in lysine, threonine and tryptophan (Boulter et.al., 1975; Shewry et.al., 1981; Croy et.al., 1980; Casey and Short 1981). Alterations and improvements to the seed proteins, or even the introduction of novel and/or alien proteins, to increase their overall nutritional value could improve general health and well being in some parts of the world.

Since the advent of agriculture some 18,000 years ago (Wendorf et.al., 1979), man has been selecting and breeding his crop plants for desirable properties with astounding success. Yields of seed from the ancient hybrid grasses do not begin to compare with those from the modern cereals which have been bred from them. In particular the "green revolution" of recent years in much of Asia, and the high energy and capital-intensive farming practices of the western world, coupled with the carefully bred strains grown, have been producing vast yields, sufficient to feed the world and to spare.

The traditional method of improving strains by cross fertilisation and selecting for desirable characteristics is an empirical trial and error process, and because the selection is at the phenotypic rather than the genotypic level many characteristics can also be inadvertantly bred out, such as the loss of many insect resistance factors in the cultivated cereals compared to their more "primitive" relatives. While this may be annoying for a farming system using herbicide and insecticide sprays, it has been disastrous for some developing-world farmers.

Essentially, conventional plant breeding being limited to crosses of sexually compatible cultivars and often involving much back crossing to isolate the desired phenotype from various undesirable ones, causes an overall decrease in the genetic diversity of the plant and hence a greater chance of vulnerability to stresses, pests and pathogens against which resistance was not specifically selected.

The elucidation of the molecular basis of genetics and the progress in genetic engineering, promise the ability to make controlled changes to the genetic complement of a plant such as switching on or off specific genes or gene families, making specific changes in genes or even introducing entirely alien genes into a plant, potentially increasing the genetic diversity of the cultivars almost without limits.

Various reviews of the potential methods to achieve these goals (Croy and Gatehouse, 1985; Barton and Hill 1983; Larkins 1983; Shewry et.al., 1981) show that the technical problems are being slowly resolved. Using the Ti (tumour-inducing) plasmids of some Agrobacterium strains, vectors have been produced which can introduce foreign DNA into a plant cell and stably integrate it with the nuclear DNA. Within the past two years the expression of alien genes introduced by this method has been achieved : a phaseolin gene, coding for a storage protein of Phaseolus vulgaris, has been transformed into and expressed in sunflower cells (Murai et.al.,1983) as well as the light-regulated pea ribulose 1,5-bisphosphate carboxylase gene in petunia cells (Broglie et.al.,1984).

Regenerating viable plants from the callus tissue resulting from Ti plasmid infection has been achieved for tobacco (Horsch et.al., 1984; De Block et.al., 1984) and so in principal it should be possible to apply these results to any plant, although the technical problems involved are formidable (Flavell and Mathias, 1984).

The Ti plasmid has a limited host range, infecting only the dicotyledonous plants, but this problem should be circumvented by some of the techniques being proposed such as using; the hybrid Ti plasmid - gemini virus vectors of Buck and Coults (1983), vectors containing plant transposable elements (Federoff, 1983) or cauliflower mosaic virus (Covey and Hull, 1985).

The choice of what genes to introduce into a given plant, with a few exceptions, is a far more subtle problem, as the molecular basis of many traits of interest to the plant breeder are poorly understood. An exception is the seed storage proteins. Each protein is the product of a single or small family of genes, they are tissue specific and under strict developmental control, so that these genes are only transcribed in the developing seed and only at certain stages of seed development. Providing this degree of control can be maintained the removal, mutation, introduction or substitution of seed storage protein genes should be possible without major modifications of or upset to the plants normal metabolism. Backing this up is an extensive body of research into seed storage proteins, providing much of the information on their synthesis, transport, processing and deposition that would be needed to introduce controlled modifications

and changes. (Derbyshire et.al., 1976; Boulter 1981; Larkins 1981; Brown et.al., 1982; Gatehouse et.al., 1984; Higgins 1984; Croy and Gatehouse 1985; Spencer 1984).

Much of the early work on seed storage proteins was done because of their great economic value as the major source of protein in the human diet. However, since the elucidation of the molecular basis of genetic inheritance, and the advent of the techniques for the manipulation of DNA, the seed proteins are now seen as an ideal system in which to investigate the control of gene expression in eukaryotes, a detailed understanding of which is probably the key problem in biology today. Other studies along these lines have used the synthesis of haemoglobin in erythroid cells and ovalbumin in chicken oviducts as model systems. (O'Malley et.al., 1977). Progress with the seed protein system has produced isolated and characterised seed storage protein genes from several legumes and cereals, including legumin from pea (Shirsat, 1984).

Although the detailed understanding of the control systems is many years away, some progress towards this goal has been made (Goldberg et.al., 1981; Gatehouse et.al., 1982; Evans et.al., 1984)

One of the ways to screen a genomic library for a specific gene is by means of a complementary DNA (cDNA) probe. To this aim a cDNA library has been constructed from total poly(A)$^+$ RNA from mid-developmental pea

cotyledons. A cDNA library consists of DNA copies of messenger RNAs, which encode for proteins, cloned usually into bacterial plasmids which are used to transform susceptable bacteria and hence can be maintained or grown up as desired. The cDNA library will reflect the relative abundances of the mRNAs present, mid-developmental pea cotyledons contain three abundance classes of mRNA, very abundant $\sim 7 \times 10^5$ copies per cell, abundant $\sim 1.5 \times 10^4$ copies per cell and rare $\sim 7 \times 10^2$ copies per cell (Morton et.al., 1982).

The mRNA is copied into cDNA using the reverse transcriptase enzyme of avian myeloblastosis virus, the synthesis being primed by oligo $(dT)_{12-18}$ which anneals to the polyA tails of the mRNA thus providing a start point for the enzyme (Evans et.al., 1980). The cDNA is made double stranded and inserted into a vector by a variety of methods (Land et.al., 1981; Okayama and Berg 1982; Maniatis et.al., 1982). By maintaining all the cDNA clones produced, the library will statistically represent the population of expressed mRNAs which should be the same as the population in the cells used for isolation. This library of cDNAs should reflect the proteins synthesised in the cell, assuming the amount of a mRNA is correlated with the amount of the relevant protein being synthesised, at the time the mRNA was isolated.

Just a few seed storage proteins are produced, often in large amounts compared with other proteins synthesised in the seed and this synthesis is developmentally regulated and tissue specific. Thus mRNAs extracted from developing seeds at a given stage of development should

contain large amounts of the mRNAs coding for the storage proteins being synthesised at that stage of development. Hence a cDNA library made from mRNA extracted from developing seeds will contain many cDNA clones of seed storage protein mRNAs.

Having obtained a cDNA library the next problem is to identify what individual cDNA clones encode, the simplest initial method being colony hybridization (Grunstein and Hogness,1975; Maniatis 1982). In the absence of probes specific for particular messages, such as cDNA or genomic clones previously identified, labelled mRNA may be used for selection. Probing with the total mRNA used as a template for the cDNA library will give strong signals for cDNAs of common messages and weak ones or no signals for cDNAs of rare messages, or cDNAs which are short in length.

By selecting cDNA clones which give strong signals and by various cross hybridization experiments to group the cDNAs into classes and then positively identifying a member of each class by techniques such as : restriction mapping, DNA sequencing and hybrid release translation with the use of antibodies specific to given proteins, cDNA clones coding for the major seed storage proteins can be found fairly easily. Finding cDNA clones of less common mRNAs by this method is a very hit-or-miss operation, as a large number of clones must be screened to be reasonably certain of finding a clone of the required message and is normally considered impractical in the absence of specific hybridization probes of some kind.

Hybridization probes for the rarer messages can be

obtained in a number of ways though often these probes are not very highly specific. Immunoprecipitating polysomes with antibodies specific to the protein product of the required message will selectively precipitate those polysomes synthesising that protein product and the mRNA extracted from these polysomes will be enriched in the required message. The drawbacks of this method are the small amounts of mRNA obtained and the need for specific antibodies which necessitates the isolation of significant amounts of the relatively pure protein product, a difficult task for the product of a rare message. However in principal this method can be highly specific. A less specific method is to hybridize with mRNA extracted from different tissues or from the same tissue at different stages of development. Combining this with some knowledge of where and when the required message is transcribed will eliminate many false positives by showing they occur at the wrong stages of development etc. In favourable systems this method can be quite specific. Similarly mutants or varieties which express/do not express the gene of interest may be used.

1.2. The Seed Proteins of Pea (Pisum sativum L.)

A seed storage protein may be defined as a protein which occurs exclusively in the seed, usually in large amounts and is hydrolysed upon the germination of the seed to be used as a source of amino acids for the developing seedling. Seed storage proteins are deposited within the cell in the membrane bound organelles called protein bodies, by means of specific transport processes.

Storage proteins tend to be inert and often insoluble in aqueous media which allows them to be deposited. This constraint on solubility combined with the need for deposition and packing in the protein bodies means that most seed storage proteins are large molecules which associate into multimers. Their importance as a nitrogen source is indicated by their high levels of nitrogen rich amino acids i.e. amides and arginine.

Developing and germinating seeds need to continue with normal metabolism, so seed specific proteins could be classified as storage proteins if they are inert and take no part in the metabolism of the cell. Another proposed method of classification is by solubility, metabolic proteins being in general soluble in water (the albumin fraction) and storage proteins in aqueous salt solution (the globulin fraction). However the pea seed major albumins which are often thought of as storage proteins are water soluble. These classifications are often not clear cut, as enzymatic and other functions have been assigned to some storage proteins and some proteins with no known metabolic role which accumulate in seeds are not utilised upon germination.

The Osborn solubility method, a system of classification of proteins by their solubilities in different solvents, showed that the major group of proteins in legume seeds were globulins, soluble in 5% aqueous salt solution. Osborn and Campbell (1898) in their pioneering work on legume storage proteins, separated pea seed globulins into three fractions on the basis of their solubility in dilute salt and reaction to heating: The legumin fraction,

insoluble in dilute salt solution and unaffected by heating to 100°C; the vicilin fraction, soluble in dilute salt solutions and coagulated by heating to 90°C; the legumelin fraction which had properties of both albumins and globulins and precipitated partially upon a reduction in salt concentration.

Danielson (1949) separated this globulin fraction into 2 major fractions at sedimentation coefficients ∿11-13S corresponding to the pea legumin fraction and ∿7-8S corresponding to the pea vicilin fraction; no legumelin fraction could be found. The ratios of these two fractions was a function of the type of legume seed used; Phaseolus vulgaris contains little if any of the ∿11S fraction; Vicia Faba , however, contains the 11S fraction as its major storage protein. (Gatehouse et.al., 1984).

Pea, Pisum sativum L., is a typical legume with roughly equal amounts of each globulin fraction protein. The ratio of the two proteins is under genetic controls and varies from line to line in a Mendelian fashion. Mature pea seeds contain 18-32% dry weight as protein (average 26%) (Schroeder 1982), of which in various estimates 15-38% are albumin proteins and 50-80% globulin proteins, with only small amounts of other proteins (Davies 1976). In pea, one major globulin fraction sediments at 12.6S, apparent molecular weight 381,000 the legumin fraction and the other at 8.1S apparent molecular weight 186,000 the vicilin fraction. (Danielson 1949). Other workers have quoted a range of molecular weights; 380 - 410,000 for the legumin fraction and 150 - 190,000 for the vicilin fraction. (Gatehouse et.al., 1984). The ratio of the legumin fraction

to the vicilin fraction varies from 0.5 to 4.0 (Gatehouse et.al., 1984). These two protein fractions are not pure proteins, which caused some confusion in earlier work, as some authors have not distinguished between fractions and pure proteins. The legumin fraction mainly consists of a single protein species called legumin, the vicilin fraction mainly consists of two distinct, though related, protein species called vicilin and convicilin, both of which can be purified free of the other.

The albumin fraction is a complex mix of proteins including significant amounts of various enzymes such as carbohydrate hydrolases (Gatehouse et.al., 1984). The major fractions within the albumin fraction are the seed lectin fraction representing ∿4% of the total seed protein. (Gatehouse and Bown personal communication), the "major albumin" fraction containing subunits of ∿25,000 and representing ∿8-10% of the total seed protein (Croy et. al., 1984) and two other significant albumin fractions Mrs ∿100,000 and ∿8000. (Gatehouse et.al., 1984). Some of these albumin fractions do not seem to be storage proteins as they are not deposited in the protein bodies and are not degraded upon germination. The albumin fraction is proportionately richer in sulphur amino acids than the globulin fraction and this is nutritionally important although it accounts for only a fraction of the total seed protein.

The soluble seed lectin fraction however, is localised in the protein bodies (Van Dreissche et.al., 1981) and is rapidly hydrolysed upon germination according to Rouge and Labroue (1977), though significant amounts of lectin persist long after the disappearance of the legumin and vicilin fractions according to

Guldager (1978) and Tyler (1981). Its function is not thought to be that of a storage protein.

Lectins are glycoproteins which specifically bind to given sugars, with pea seed lectin being specific for D-mannose and D-glucose. (Lis and Sharron 1981). They were first recognised in plant extracts by their ability to agglutinate erythrocytes, Sumner and Howell (1937) showed that the effect was due to lectin binding to glycoproteins or sugar residues of polysaccarides on the surface of the red cells. Lectins can be detected in various parts of the pea plant such as the roots where sugar specificity has been linked to the binding of the symbiotic rhizobium bacteria (Lis and Sharron 1981). However the lectin concentration in the seed is far higher than anywhere else in the plant and it is thought these non seed lectins are clearly different to the soluble seed lectin (Van Dreissche et.al., 1981). Lectins have been implicated in toxicity and impared growth rates in experimental animals fed unheated plant proteins, presumably the lectins bind to the animals gut walls or digestive enzymes (Jaffe 1969). However not all lectins are similarly active, the removal of soy bean seed lectin from the raw meal make no difference to growth rates or protein absorption efficiency in experimental animals. Their role in the seed is probably to defend against fungal and bacterial pathogens by binding to their cell walls and in controlling animal, particularly insect, pests by their growth inhibitory or toxic properties. The nutritional significance of the seed lectin fraction is limited due to its relatively low contribution to total seed protein. However if eaten uncooked lectins can interfere with digestion. The sugar binding activity of

lectins is destroyed by cooking.

1.2.1. The legumin fraction

Pea legumins of molecular weight within the range 380 - 410,000 in non-reducing denaturing solvents such as SDS dissociate into molecules of Mr∿60,000, as estimated by SDS polyacrylamide gel electrophoresis. In the presence of a reducing agent such as 2-mercapto-ethanol a further dissociation occurs to subunits of molecular weights ∿38,000 and ∿21,000 (Casey 1979).

Analogous experiments on the 11S proteins of Vicia faba had led Wright and Boulter (1974) to propose a model for the structure of the legume 11S proteins. The structure proposed that the small subunits were held together in pairs, one Mr 38,000 to one Mr 21,000, by disulphide bonds, one or more per subunit pair. These subunit pairs then aggregate into hexamers which are not covalently bonded to each other. This model was confirmed for pea legumin by Croy et.al., (1979).

In pea legumin the larger of the subunits (Mr∿38,000) is called the α-subunit or acidic subunit having pI values in the range 4.8-6.1, and the smaller (Mr∿21,000) is called the β-subunit or basic subunit having pI values in the range 6.2 - 8.0. Both types of subunits show considerable variation in size and charge under suitably chosen conditions of gel electrophoresis. Casey (1979) who classified the α-subunits by their relative abundances and later Matta et.al., (1981), showed that each α-subunit type was associated with a specific β-subunit type, and each subunit pair could be classified as either "main

legumin" the most abundant type with α subunits of Mr 38,000, "big legumin" with α subunits of Mr 39-42,000 or "small legumin" with α subunits of Mr∿25,000.

Each type of subunit pair showed further heterogeneity with a total of 22 acidic and 11 basic subunits being resolved by SDS polyacrylamide gel electrophoresis and isoelectric focusing. This diversity is reflected in the isolation of legumin hexamers of entirely main legumin, of roughly equal amounts of "main" and "big" legumins, and of all three types. However the association of subunit pairs is under strict control, with only defined pairs occuring. (Gatehouse et. al., 1981)

This classification scheme is supported by the segregation of the SDS polyacrylamide gel band patterns for legumin in crosses of various pea lines, (Gatehouse et. al., 1981) as the "main" "big" and "small" legumin types segregate independantly but the individual subunit pairs always segregate together.

The amino acid content of the legumin fraction is typical of a legume storage protein with low levels of sulphur and high levels of amides and arginine, although genetic variation in the level of methionine has been reported (Casey,1981). Legumin is not glycosylated (Casey,1979; Gatehouse et.al., 1980), and reports of covalently attached sugars (Basha and Beevers 1976, Davey and Dudman 1979), were based on analysis of impure legumin fractions.

Antibodies raised against pea legumin specifically react with all the legumin subunit pairs (Matta et.al., 1981; Domoney and Casey 1983), and with 11S proteins from closely related species eg. Vicia faba, (Croy et. al., 1979), lentil (Dudman and Millerd 1975) but not with other pea seed protein subunits.

1.2.2. The Vicilin Fraction.

1.2.2.1. Vicilin

Pea vicilin of Mr 150-180,000 (Thomson et.al., 1980; Gatehouse et.al., 1981) in a denaturing solvent dissociates into many subunits, the major ones being of ∿Mrs 50,000, 33,000, 19,000, 16,000, 13,500 and 12,500 as estimated by SDS polyacrylamide gel electrophoresis. Reducing agents do not produce any further subunits, indicating no disulphide bonded subunits, (vicilin contains no sulphur amino acids). The major subunits are present in the approximate molecular ratios 1(50,000) : 1 (33,000) : 1 (19,000): 2 (16,000 + 13,500 + 12,500) (Gatehouse et.al., 1984), with small amounts of minor subunits of Mrs ∿31,000 and ∿35,000. Often subunits of convicilin (see 1.2.2.2) Mr ∿71,000 are present particularly in preparations from solubility fractionation (Grant and Lawrence 1964; Thomson et.al., 1980), this is probably due to molecular rearrangements caused by partial dissociation into subunits in the extraction procedure, or to the low efficiency of fractionation by solubility. Vicilin and convicilin can be readily separated (Croy et.al., 1980).

Estimates of the molecular weights of the various vicilin subunits vary considerably from one literature source to another. Also vicilin, unlike legumin, contains small amounts of covalently bound carbodydrate which further confuses the molecular weight estimations. Vicilin under non-dissociating conditions can be separated into various molecular species of slightly differing subunit composition, one of which contains no Mr 33,000 subunit (Gatehouse et.al., 1981; Thomson et.al., 1978; Grant and Lawrence,1964),

It can also be separated by carbohydrate affinity column chromatography on immobilised concanavilin A into fractions with and without significant amounts of carbohydrate. The carbohydrate was shown to be associated with the Mr 16,000 subunit (Gatehouse et.al.,1980; Davey and Dudman 1976) showing that some vicilin molecules contain the Mr 16,000 subunit and others do not. High resolution 2-dimensional isoelectric focusing reveals diversity in the various subunits and as many as 30 vicilin subunits have been observed (Hirano and Gatehouse, personal communication; Thomson et.al., 1978).

Conversely the close relationships between the various vicilin polypeptides have been shown by tryptic peptide mapping studies (Gatehouse et.al., 1982) and by serological studies on the Mr 16,000, and Mr 12,500 subunits, the Mr 16,000 subunit being probably a glycosylated form of the Mr 12,500 or another very closely related peptide.

Inheritance studies on the vicilin subunits show Mendelian segregation suggesting a separate gene for each subunit, though no rigorous analysis has been performed (Sharon and Lis 1979). This large array of vicilin subunits combined with the inconsistencies in some molecular weight estimates made the molecular structure of native undissociated vicilin obscure. Early models which assembled the molecule from six or seven separate subunits, assuming each to be a separate entity (Thomson et. al., 1980) were too complicated to seem likely and it was thought that the diversity of subunits disguised an underlying simplicity of structure. The structural model of vicilin is derived from its biosynthesis and I shall discuss it in detail later.

Antibodies raised against pea vicilin react with pea convicilin (Croy et.al., 1980; Domoney and Casey,1983), and with the 7S proteins from closely related legume species, eg: Vicia faba, lentil etc., (Dudman and Millerd, 1975), but they do not react with pea legumin.

1.2.2.2.Convicilin.

Pea convicilin of Mr 210-280,000 (Croy et.al., 1980) in a denaturing solvent dissociates into subunits of Mr 71,000 as estimated by SDS-polyacrylamide gel electrophoresis; reducing agents do not produce any further dissociation. Convicilin contains no significant amounts of carbohydrate and on the basis of these molecular weight values is thought to be a simple tetramer. It contains sulphur amino acids, one cysteine and one methionine residues per subunit molecule, in contrast to the vicilin. However, antibodies raised against pure convicilin react with vicilin (Domoney and Casey 1983), their serological simularities having already been commented upon.

Convicilin contains subunits of more than one pI value and also shows variations between pea lines in subunit size as estimated by SDS-polyacrylamide gel electrophoresis. This diversity is less than that shown for legumin or vicilin. Using SDS-PAGE, convicilin was shown to map to a single Mendelian genetic locus on chromosome two of the pea genome (Matta and Gatehouse 1982), designated Cvc, previously referred to incorrectly as VA.(Thomson and Schroeder 1978). Cvc and Lg-1 (a legumin locus) were the first biochemical markers accurately mapped on the pea genome.

1.2.3. The Albumin Fraction.

1.2.3.1. Major Albumins.

The pea major albumins fraction consists of two closely related protein species of Mr∿53,000 and∿48,000, known respectively as PMA-L (Pea major albumin-Large) and PMA-S (-Small), which can be purified free of the other. In denaturing solvents they dissociate to subunits of Mr∿25,000 for PMA-L and ∿24,000 for PMA-S as estimated by SDS-polyacrylamide gel electrophoresis (Croy et.al., 1984), indicating a dimer structure. No mixed dimers, i.e. one of each subunit, were isolated despite the two proteins being structurally very similar. This structural similarity is illustrated by their cyanogen bromide cleavage patterns on high resolution SDS-urea polyacrylamide gel electrophoresis, where both major albumins give an identical fragment of Mr 11,500 together with fragments of Mr 8,400 and 7,200 for PMA-L and 7,600 and 6,600 for PMA-S. Cyanogen bromide cleaves protein at the methionine residues. The major albumins give similar tryptic peptide maps as determined by H.p.l.c. and neither contains significant amounts of carbohydrate. (Croy et. al., 1984)

Some charge heterogeneity was detected in both subunits and this seemed to be unconnected with the known changes (thought to be deamination of amide amino acid residues, Croy et.al., 1984) that occur during their storage or during prolonged purification. Both major albumins show a reaction of complete identity in immunodiffusion experiments against anti total major albumin antibodies, and neither give cross reactions with other seed proteins or with antibodies to lectin (which is an albumin fraction protein), vicilin or legumin (Croy et. al. 1984). Pea major albumin antibodies cross

react with total seed albumin extracts from related species such as lentil, or to a lesser extent sweet pea and chick pea.

1.2.3.2. Lectin.

The soluble pea seed lectin, which is isolated from the seed albumin fraction by affinity chromotography, consists of two major protein species both of Mr 49,000 but of differing pI values (Trowbridge,1974): The acidic or A type lectin, pI∿4.1 as estimated by isoelectric focusing in polyacrylamide gels, and the basic or B type lectin, pI∿6.5, which was roughly twice as abundant. In denaturing solvents such as SDS, lectin dissociates into equal molecular proportions of two subunits of Mr 17,000, the β subunit and 7,000 the α subunit, neither showing any size heterogeneity. Isoelectric focusing further resolves the α subunit into two differing charge forms, roughly in the molecular ratio of 2:1. The more acidic α subunit was exclusively present in the acidic type A lectin and the less acidic in the type B lectin. No charge heterogeneity was observed in the β subunit. Lectin does not contain detectable amounts of carbohydrate. (Trowbridge, 1974)

The structure of pea seed lectin is thought to be a tetramer of two α and two β subunits, which is supported by cross-linking experiments, and by analogy with lectins of closely related species such as lentil, Canavalia ensiformis (whose lectin is called concanavilin A) and Vicia faba (whose lectin is called favin) (Higgins et.al., 1982; Sharon and Lis 1979).

Antibodies raised against the total soluble seed lectin

do not cross-react with the pea major albumins or the pea globulins (Croy et.al., 1984) but do react with lectins from closely related species (Lis and Sharon 1979).

1.3. Biosynthesis of the major seed proteins.

The *in vivo* synthesis of pea major storage proteins, the legumin fraction and the vicilin (and convicilin) fraction, occurs in cells on the rough endoplasmic reticulum of the cotyledons. These are called the storage parenchyma cells. The proteins are then transported and deposited in the protein bodies, the whole process taking about 1 hour (Gatehouse et.al., 1984).

Both legumin and vicilin are found in the same protein bodies which are thought to originate from the 1-2 large vacuoles originally present in the cell or from the dictyosome vesicles, which may also be involved in the transport process, or both, as occurs in *Vicia faba* (Nieden and Neumann 1982; Gatehouse et.al., 1984).

Molecular investigations of storage protein synthesis, transport and deposition have been carried out *in vivo* by radioactive pulse chase techniques, and *in vitro* by cell-free translation experiments on mRNAs isolated from developing cotyledons of varying ages.

1.3.1. The Legumin Fraction

Both main legumin subunits, α and β are synthesised in the developing cotyledon as shown by the uptake and fate of (^{35}S)-labelled methionine in developing pea pods. However cell-free translation systems using mRNA isolated from cotyledons at the same stage of development produce neither

of the subunits, but do produce proteins of Mr∿60,000 by SDS-polyacrylamide gel electrophoresis (Gatehouse et. al., 1984). These proteins react specifically with antilegumin antibodies but are not cleaved into subunits by reducing agents as is native main legumin. They also bind specifically to immobilised legumin on an affinity column (Gatehouse et al 1984).

Croy et.al., (1980) proposed that each main legumin is synthesised as a single protein which is folded, disulphide cross-linked and post-translationally cleaved into the disulphide linked α and β subunits found in vivo. Further work has confirmed this model and shown that main legumin is synthesised as a single molecule with a leader sequence which is a signal peptide for secretion and is removed co-translationally (Gatehouse et.al., 1984). Pulse chase experiments suggest that the α,β cleavage occurs several hours later i.e. in the protein bodies. Legumin molecules assemble into trimers of Mr∿180,000 in the endoplasmic reticulum before completing assembly to the hexamers of Mr 360,000, as isolated under non dissociating conditions from mature pea seeds (Gatehouse et. al., 1984).

1.3.2. The Vicilin Fraction.

Vicilin and convicilin are synthesised in the cotyledons in an essentially identical manner to legumin; cell-free translation systems give only four size classes of protein which react with antivicilin antibodies, one of Mr∿71,000 by SDS-polyacrylamide gel electrophoresis corresponding to convicilin and three at Mr∿50,000 (Gatehouse et.al., 1984). On high resolution gels these can be separated into one at Mr 47,000 and two at Mr 50,000,

collectively these Mr∿50,000 proteins are known as vicilin (or main vicilin). None of the smaller vicilin subunits was detected. Short duration pulse-chase experiments show that initially only Mr∿50,000 proteins become labelled, and over longer periods all the smaller subunits of Mr 33,000 - 12,500 become labelled whereas the Mr 47,000, and one of the Mr 50,000 protein species disappear (Gatehouse et.al., 1984). This suggests that the smaller subunits are derived from the Mr 47,000 and one of the Mr 50,000 main vicilins by a post-translational cleavage system.

Gatehouse et.al., (1981) thus proposed a model for vicilin in which each molecule is a trimer of Mr∿50,000 subunits, some of which are post-translationally cleaved (Figure 1). Different main vicilins have different susceptibilities to post-translational processing which may occur on a different time scale. Thus Mr 47,000 vicilin has a half life of 6 hours in developing seeds, showing that the molecules are assembled and deposited before being processed. No product corresponding to cleavage at the α-β site only (Mr 26,000 or 29,500 if glycosylated) is seen which fits with the three Mr∿ 50,000 main vicilins found. This model has been confirmed in detail by Gatehouse et.al., (1984).

Evidence for a second rapid processing step for both vicilin and convicilin has been found (Gatehouse et.al., 1981; Higgins and Spencer 1980) which has been interpreted as the removal of a short leader or signal peptide co-translationally in accordance with the signal hypothesis of Blobel and Dobberstein (1975). The presence of a short leader sequence has been confirmed in a main

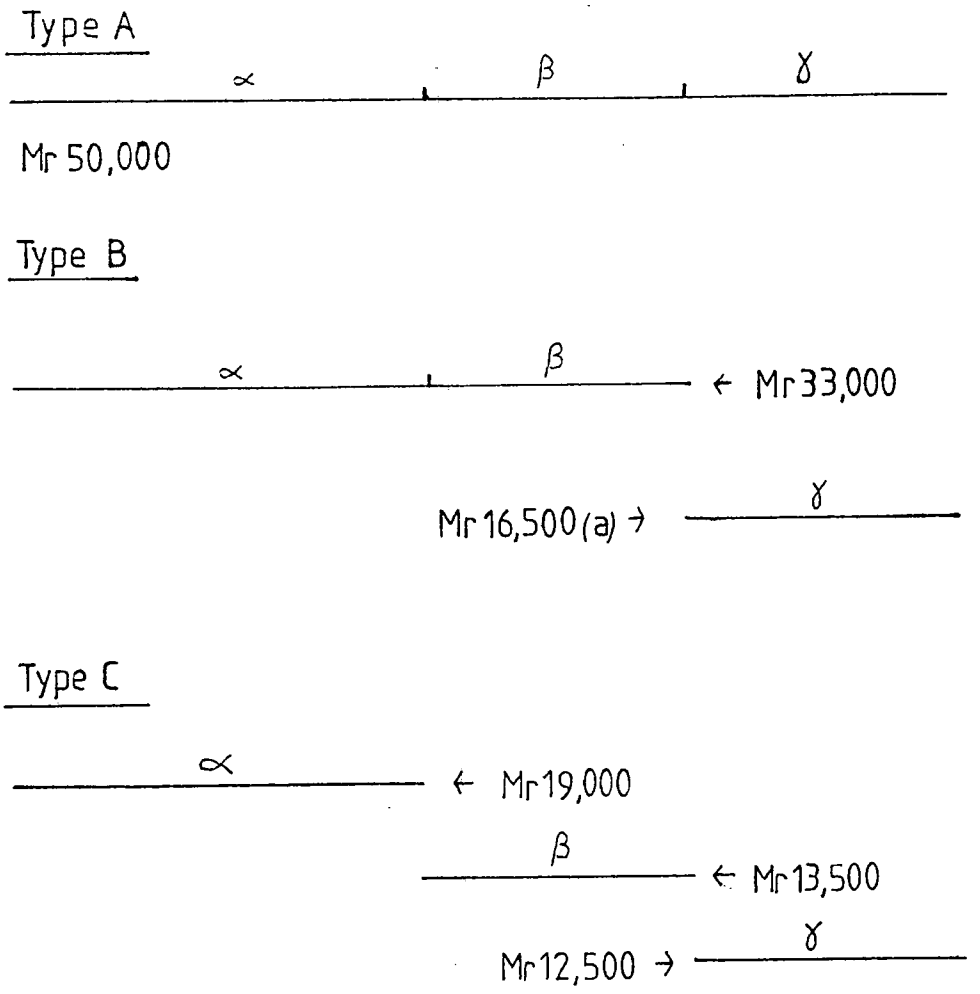


Fig.1. Cleavage Patterns of Vicilins

(a) this subunit is glycosylated.

vicilin by Lycett et.al.,(1983) who showed that it was rich in hydrophobic amino acids as would be expected of a signal for secretion through a membrane. Apart from the leader peptide, convicilin does not undergo any post-translational processing.

Finally, some vicilin is glycosylated sufficiently to bind to concanavilin A, and it has been shown (Higgins and Spencer1980; Badenoch-Jones et.al., 1981; Chrispeels et.al., 1982) that labelling of main Mr∿50,000 vicilin in vivo by radio-labelled sugar is subsequently chased into the Mr 16,000 subunit by the proteolytic cleavages outlined above. Glycosylation was found to be a two-step process, firstly on the rough endoplasmic reticulum the site of synthesis of vicilin, and later in a second step, probably in the Golgi apparatus. Convicilin is not glycosylated although a minor component of some convicilin preparations of Mr∿65,000 is; its relation to convicilin is unknown.

1.3.3. The Albumin Fraction.

1.3.3.1. The Major Albumins.

The small and large major albumins (PMA-S and PMA-L) synthesised in vitro are similar to those isolated in vivo by immunoprecipitation (R.Croy and J.Gatehouse personal communication, R. A. Ragab personal communication), showing only the expected size subunits by SDS polyacrylamide gel electrophoresis. Major albumins synthesised in vitro contain only the Mr~25,000 subunits and no Mr~50,000 species under dissociating conditions. this poses a problem as to why no mixed

major albumins dimers are found in vivo, perhaps indicating subunit interaction limitations.

1.3.3.2. The Lectin.

Higgins et.al., (1982;1983) have shown that, similar to legumin and vicilin, lectin is synthesised as a single chain. This is secreted with the loss of a leader signal peptide, then assembled into dimers and finally subject to post-translational proteolytic cleavage into the α and β subunits in the protein bodies. This probably involves the removal of a small 6 base linker sequence from between the subunits and also a short 4 base tail from the end of the α subunit. The α subunit corresponds to the 3' end of the message and the β to the 5'.

1.3.4. Post-translational Modifications.

Whereas the removal of a leader signal sequence upon crossing the endoplasmic reticulum membrane is a known and accepted model, the function of the glycosylation of some vicilins and the significance of the post-translational cleavage patterns of legumin and some vicilins are however a matter for speculation. The differing glycosylation of different vicilins may indicate different transport mechanisms, perhaps tying in with the two origins of protein bodies hypothesis (Nieden and Neumann 1982; Gatehouse et.al.,1984). It seems that the cleavages of the lectin precursor are necessary for its sugar binding functions (Lis and Sharon, 1981) and those of legumin seem to be necessary for deposition as the legumin cleavage occurs before the legumin trimers associate into hexamers (Spencer

and Higgins 1980; Chrispeels et.al., 1982). The cleavage patterns of the vicilins however seem to have no function unless they speed mobilisation upon germination.

1.4. The Seed-protein Genes.

Cloned complementary DNA (cDNA) and genomic DNA coding for various pea storage proteins have been produced characterized and some of their DNA sequences determined (Review by Sorenson 1984). Before these techniques were available sequence information was severely limited by the practicalities of the chemical protein sequencing methods; for example the legumin sequence was known for approximately 30 amino acids from the N terminus of each subunit (Casey et.al., 1981a; Casey et.al. 1981b). A unique protein sequence can in general be produced from a large number of related but not identical DNA sequences due to; the redundancy of the DNA code, the presence of introns in eukaryote genes and co and post-translational cleavages of precursor proteins to produce the mature protein. Thus although the known microheterogeneity of the seed proteins can be explained in terms of minor differences in protein sequence and cleavage patterns, this is only a reflection of the true differences which exist on the DNA level and can only be studied in detail at that level. These available cloned genes allow comparisons between themselves and with homologous genes from other species to be made, as well as determining the complete protein sequences of individual proteins.

Pea storage protein genes seem to be arrayed into small families of very closely-related genes which have a more distant relationship with other groups of genes.

Using cDNA or genomic hybridisation probes on restricted nuclear DNA the number of copies of a gene can be estimated (eg Domoney and Casey 1983) and by probing chromosomes in situ using the very latest techniques the genes may be located onto a chromosome. These estimates of gene copy number must be treated with some caution as quite closely related genes may differ sufficiently in sequence so as not to hybridise. For example, vicilin and convicilin show a clear relationship in protein sequence as judged by their cross reaction with specific antibodies raised against each other, however their DNA sequence is sufficiently diverse to show no cross hybridisation between cDNA and mRNA under fairly low stringency conditions, as used by Domoney and Casey (1983) in their hybrid release translation experiments.

Protein sequence is under evolutionary control, as changes which disrupt the proteins function can be disadvantageous to the organism. Thus the exon sequence of a gene which gives rise to the protein sequence is under more stringent evolutionary control than the intron sequence which will drift randomly with time (the molecular clock hypothesis), and comparison of intron sequences from related genes or the same gene from different species gives an estimate of the time since the two genes were identical. Similar comparisons of the exon sequences and the flanking sequences potentially can yield information on the molecular basis for differences in protein structure and function, and on the control of gene expression.

1.4.1. Legumin genes.

Various main legumin cDNAs have been sequenced (Croy et.al., 1982; Delauney 1984) and using one of these as a probe three different main legumin genes have been isolated from a pea genomic bank (Shirsat 1984; Croy et. al., 1985) all of which have now been sequenced (Lycett et.al., 1984 ; J.Gatehouse personal communication). The three genes, named legumin A, B and C, are all very similar showing better than 95% homology of DNA sequence.

A fourth gene related to the main legumin genes has also been found, however it contains deletions, frame shift errors, two stop codons near the 5' end of its coding sequence and is not transcribed (Bown et. al., 1985). It has been named legumin D and is a pseudogene. By comparing the sequence from the

cDNAs and the genomic clones main legumin genes have been found to have three introns, two each 88bp long in the coding sequence of the Mr∿40,000 α-subunit and one 99bp long in the Mr∿20,000 β-subunit coding sequence. The boundary sequence of the introns was typical of those of higher plant genes. The flanking sequences of the genes contained the expected transcription control sequences; in the 5' flanking sequence; a 'TATA' box, a 'CAAT' box, an 'AGGA' box and a transcription start codon: in the 3' flanking sequence a number of polyadenylation signals (Lycett et.al., 1984).

Croy et.al., (1982) estimated the number of main legumin genes at 4 by probing nuclear DNA with one of their main legumin cDNAs at a stringency corresponding to ∿95% homology. Shirsat (1984) using as a probe a genomic main legumin clone, found with one of the main legumin cDNAs, estimated there to be at

least 7 legumin genes, Domoney and Casey (1985) in similar experiments using cDNA probes estimated there to be 4 to 6 genes for main legumin. They also estimated there to be 1 to 3 genes for big legumin, which has a molecular weight of 63-65,000 (main legumin has a Mr of 60,000) and 1 to 2 genes for an Mr 80,000 legumin related polypeptide. How these genes relate to the "big" and "small" legumin molecules isolated *in vivo* is not certain. Cross hybridisation experiments show that the Mr∿63-65,000 big legumin cDNA and the Mr∿80,000 legumin cDNA are closely related and cross-react in hybrid selection experiments (Domoney and Casey 1984) but no evidence of cross reaction in the genomic probes was detected. No cross hybridisation of these with the Mr∿60,000 main legumin cDNA was observed under any conditions. The protein products of all three legumin types react with anti legumin antibodies.

Casey et.al., (Personal communication) using their big legumin cDNA have isolated from a pea genomic bank a clone bearing 2 big legumin genes, called legumin J and K. Initial sequencing of legumin J has shown similarities in the coding sequence to the main legumin genes but no significant homologies were found between the 5' flanking sequences except the transcription signals (J.Gatehouse personal communication).

A third legumin related protein, a small legumin, is known *in vivo* but no sequence data has been published.

1.4.2. Vicilin Genes

cDNA clones coding for vicilin have been produced (Croy et.al., 1982; Delauney 1984) and on the basis of their sequences can be divided into three classes. One of them corresponds to the Mr47,000 vicilin and the others to

* Gatehouse et. al. (1983) similarly estimated 3-6 genes homologous to pRC2,2,1 (pDUB2) a cDNA for the Mr 50,000 type A vicilin and 2-3 genes homologous to pRC2,2,10 (pDUB4) a cDNA for the Mr47,000 type B vicilin.

the Mr50,000 vicilins, the DNA sequence homology between any two classes being about 85% (Delauney 1984). The amino acid sequences determined from the DNA sequences confirm the model of the synthesis of the vicilin subunits found in vivo and suggests why some vicilins are proteolytically cleaved and others are not.

The Mr47,000 vicilin (vicilin B) contains a glycosylation site in its γ subunit (see Figure 1) and a proteolytic cleavage signal at its β-γ junction, and is processed to the Mr 16,000 and 33,000 subunits. One of the Mr50,000 vicilins (vicilin A) has no processing or glycosylation signals and hence is not processed. The other Mr50,000 vicilin (vicilin C) has no glycosylation site but has two proteolytic cleavage signals at the α-β and β-γ subunit junctions and hence gives rise to the Mr 12,500, 13,500 and 19,000 subunits. (Lycett et. al., 1983; Delauney 1984; Croy et.al., 1982). This fits tolerably well with the estimates of the molecular ratios of the subunits. No vicilin D cleaved at the α-β junction but not at the β-γ was detected. Thus, the known vicilin gene family consists of three types of vicilin gene, each of which is probably present in more than one copy: Delauney (1984) found minor DNA sequences differences in two otherwise identical vicilin C cDNAs. Domoney and Casey (1985) estimated, by probing genomic DNA with cDNA probes, 5-7 vicilin B genes using their pCD4 Mr47,000 vicilin cDNA clone and 4-6 genes for a Mr50,000 vicilin using their pCD48 Mr50,000 vicilin cDNA clone, which vicilin this codes for is as yet unknown.* Significant cross hybridisation of the two probes was detected.

1.4.3. Convicilin genes

A cDNA coding for part of a Mr 71,000 convicilin subunit has been produced (Domoney and Casey 1983). Its sequence has considerable homology with that of the vicilin cDNAs but it does not select vicilin mRNA in hybrid release translation experiments (Casey et.al., 1984). An estimate of the number of convicilin genes by Domoney and Casey (1985) of 0.5 to 1 seems at variance with the genetic mapping experiments of Casey and Sanger (1980) and Matta and Gatehouse (1982) which indicate more than one.

1.4.4. Major Albumin Genes

Until the present work no cDNAs or genomic clones of the major albumins have been isolated, however, the microheterogeneity of both small and large major albumins suggest small gene families for each protein.

1.4.5. Lectin genes

cDNAs coding for a seed lectin have been produced (Higgins et.al., 1983), and the complete coding sequence, and the 3'-end untranslated sequence up to the polyA tail, have been determined. It was not specified whether it was the acidic type A lectin or the more abundant basic type B lectin. The biosynthetic scheme for production of pre-prolectin followed by the loss of a leader signal peptide to give prolectin, and the cleavage of prolectin with the loss of a small linker peptide to give the α and β subunits, was confirmed. No estimates of lectin gene number were made although the presence of two types of seed lectin and the suggestion of heterogeneity in the

5' untranslated sequence of the cDNAs indicate at least 3 seed lectin genes. There are also a number of non-seed lectins which may or may not be related to the seed lectin.

1.5. Objectives and Content of the Research.

For reasons outlined above cDNA libraries are the most direct method to obtain protein and DNA sequence data of individual gene products. They allow the determination of the organisation of sequence within a gene, i.e., the location of its introns and the termini of its mRNA, by comparison of cDNA and genomic DNA nucleotide sequences. DNA sequencing is much faster and easier than protein sequencing and the sequence of a cDNA represents the product of a single gene whereas protein sequencing methods cannot be so selective. Genomic clones also carry this information but the introns cannot be unambiguously located without the cDNA or protein sequence. Genomic clones are usually selected from a gene bank by their ability to hybridize to a known cDNA.

This thesis describes the construction of a cDNA library from mRNA isolated from 13-17 day old developing pea cotyledons and the isolation of some cDNAs coding for various legumins and vicilins as well as cDNAs coding for the seed lectin and one of the major albumins. The methods used to isolate and identify these clones were : hybridisation to previously characterised DNAs, insert sizing by agarose gels, restriction enzyme mapping and DNA sequencing (Maniatis et.al., 1982; Maxam and Gilbert 1980; Messing 1983; Davies 1982). The aim of this research

was to produce characterised pea seed cDNAs which could be used to select genomic clones carrying the corresponding genes from a pea gene bank.

To correlate the abundances of the different classes of cDNAs with the messenger RNA sequence complexity data of Morton et. al. (1983), and the abundance of their protein products in mature seeds.

To cross correlate the colony hybridisation screen results with the three different vicilin probes in order to assign colonies as specifically positive for one vicilin type over the other types and hence investigate the relative abundances and inter-relationships of the three vicilin types at the mRNA level.

CHAPTER 2

MATERIALS AND METHODS.

2.1. Materials.

2.1.1. Chemical and Biological reagents.

Most chemical reagents were supplied by B.D.H.Chemicals Ltd., Poole, Dorset, U.K., and were of the purest grade available.

The Sigma Chemical Co., Poole, Dorset, U.K. supplied; ethidium bromide, adenosine triphosphate (ATP), spermidine, bovine serum albumin (BSA), dithiothreitol (DTT), HEPES (N-2-hydroxyethylpiperazine-N-2-ethane sulphonic acid), Tris (tris(hydroxymethyl) aminomethane), ampicillin, chloramphenicol, tetracycline, RNAseA, E.coli ^{t}RNA, polyadenylic acid (poly A), herring sperm DNA, all deoxynucleotide triphosphates, all dideoxynucleotide triphosphates, lysozyme.

Koch-Light Ltd., Haverhill, Suffolk, U.K., supplied; sodium chloride, caesium chloride, acetic anhydride, isoamyl alcohol, chloroform.

Pharmacia Fine Chemicals, Uppsala, Sweden, supplied; Sephadex G-50, Sepharose 6B-CL, Ficoll 400.

Schleicher and Schull, Anderman and Co., Ltd., Kingston-upon-Thames, Surrey, U.K., supplied; nitrocellulose filters (BA85, 0.45µm)

Whatmann Ltd., Maidstone, Kent, U.K., supplied; 3MM paper, DEAE-cellulose (DE 81) paper, glass fibre (GF and GC) filters.

Difco Laboratories, Detroit, Michegan, U.S.A. supplied; Bactotryptone, Bacto-Agar, Bacto-Yeast Extract.

Becton Dickinson and Co., Cockeysville, M.D., U.S.A., supplied; BBL Tripticase peptone.

Collaborative Research Inc., Waltham, M.A., U.S.A. supplied; Oligo dT_{12-18}.

Restriction endonucleases were supplied by : Bethesda Research Laboratories (UK) Ltd., Cambridge, U.K., ; The Boehringer Corporation (London) Ltd., Lewes, East Sussex, UK.; New England Biolabs, C.P. Laboratories Ltd., Bishops Stortford, Herts, U.K.

The Boehringer Corporation (London) Ltd., also supplied; Calf intestine alkaline phosphatase, endonuclease-free E.coli DNA polymerase I, T4 polynucleotide kinase, T4 DNA ligase, S1 nuclease.

Bethesda Research Laboratories(UK) Ltd., also supplied; E.coli DNA polymerase I large fragment (Klenow polymerase), EcoRI synthetic linkers, agarose (electrophoresis grade).

Pharmacia P.L. Biochemicals Inc., Pharmacia (Great Britain) Ltd., Milton Keynes, Bucks, U.K. supplied ;

Calf thymus terminal deoxynucleotidyl transferase, mung bean nuclease.

Worthing Biochemicals Millipore (U.K.) Ltd., London, U.K., supplied; DNAse I (DPFF).

The Division of Cancer Cause and Prevention, National Cancer Institute, N.I.H., Bethesda, MD, U.S.A., supplied; Avian myeloblastosis virus (AMV) reverse transcriptase.

Amersham International plc., Amersham, Bucks, U.K., supplied; Radiochemicals, nick translation kits.

Mr. R. Swinhoe made a generous gift of; Poly A^+ RNA prepared from 13-17 day old cotyledons of Pisum sativum L. variety Feltham first (Sutton Seeds Ltd., Reading, Berks, U.K.).

2.1.2. Bacterial Strains Plasmids and Phages.

All bacterial strains used were derivatives of E. coli K-12 and are listed in Table 1. Plasmids and phages used are also listed in Table 1.

2.2. Methods.

2.2.1. Standard Biochemical Techniques.

2.2.1.1. Plasticware, Glassware, General Apparatus and Reagents.

All apparatus coming directly or indirectly in contact with nucleic acids, bacteria, enzymes, or any sensitive or sterile reagents etc., was sterilised by autoclaving.

TABLE 1

Properties of E.Coli Strains, plasmids and phages used.

Item	Genetic Characteristics	Reference/ Source
Bacterial Strains		
910	$Ap^S Tc^S$ (803 SupE SupF RecBC)	W.J.Brammar, Dept. of Biochemistry, University of Leicester, U.K.
TB1	Ap^S Δlac pro ara thi strA Ø80d lacZ ΔM15 hsdR $r^- m^-$	T. Baldwin Texas A+M University
JM101	Δlacpro supE thi F' traD36 proAB $lacI^q$ lacZ ΔM15	
Plasmids		
pBR322	Ap^R, Tc^R	Bolivar et.al., 1977
pUC8	Ap^R, lac^C	Vieira and Messing,1982.
pBR322-SV40 (map units 0.71-0.86)		-- Okayama and Berg, 1981.
pBR322-SV40 (map units 0.19-0.32)		
Phages		
λNM258		N. Murray Mol. Biol. Dept. Edinburgh University.
M13 mp9	lac^C	M13 Cloning/ Sequencing Manual, BRL (UK) Ltd. / Vieira and Messing, 1982.

Key: Ap^S, ampicillin sensitive. Ap^R, ampicillin resistant.

Tc^S, tetracycline sensitive. Tc^R, tetracycline resistant.

lac^C complements the lac system.

Those items which cannot be autoclaved were sterilized by prolonged immersion in 80% ethanol and if appropriate,flaming. All glassware and plastic eppendorfs were siliconized before autoclaving using "Repelcote" (Hopkins and Williams, Romford, U.K.).

All solutions for DNA manipulations except gels, were autoclaved where possible or made up using sterile water and sterile stock solutions in sterilised containers, minimising the use of non sterile reagents. All water was distilled or where noted as HPLC grade, double distilled and deionised, all ethanol used in DNA manipulations was redistilled.

2.2.1.2. Alcohol Precipitation of DNA

DNA solutions were made to 0.3M sodium acetate, usually by the addition of 1/10th volume of 3M sodium acetate pH5.2, then 2-3 volumes of ethanol were added, mixed and the solution kept at -20°C overnight or -80°C for 15-20 minutes.

The DNA was then pelleted by centrifugation; for small (up to 2ml) samples at ∿12,000g 4°C for 10-15 minutes (M.S.E. Micro Centaur microcentrifuge or equivalent), or for larger samples at ∿25,000g 4°C for 20-30 minutes (Sorvall RC-5B centrifuge). The pellet was then washed in 70% ethanol 2 or 3 times and gently vacuum dried before being resuspended usually in either water or T.E.buffer (10mM Tris-HCl pH7.5, 1mM EDTA). When the concentration of DNA was low or the size small (<200bp) transfer RNA (tRNA) was added prior to the addition of the ethanol. When the sodium concentration in the DNA

solution was high potassium acetate pH5.2 to 0.5M or ammonium acetate and magnesium acetate pH5.2 to 0.5M and 0.1M respectively were used instead of sodium acetate. When it was necessary to minimise the total volume 0.6-1.0 volumes of iso-propanol were used in place of ethanol and the mixture kept at -20°C for at least 20 minutes.

2.2.1.3. Phenol Extraction of DNA Solutions

Proteins were removed from DNA solutions by extracting twice with ∿1 volume of phenol saturated with T.E buffer (10mM Tris-HCl pH7.5 1mM EDTA), then twice with chloroform: iso-amylalcohol (24:1). The layers were mixed by brief vortexing and separated by brief centrifugation if needed. The DNA remains in the top aqueous phase in each case. Occasionally the chloroform iso-amylalcohol extraction steps were replaced with 2 or 3 extractions with diethylether, in this case the DNA containing aqueous layer is on the bottom. This whole process is referred to as phenol extraction.

2.2.1.4. Trichloroacetic Acid Precipitation of DNA

DNA solutions were trichloroacetic acid precipitated essentially according to the protocol in the Amersham Nick Translation kit handbook (Amersham International plc., Amersham, U.K.) using ∿30µg of phage λ DNA as carrier and Whatman GFC glass fibre filters.

2.2.1.5. Scintillation Counting of Radiolabelled DNA

Labelled DNA in solution or precipitated onto glass fibre discs was counted by liquid scintillation in a

Packard PL Tri-Carb liquid scintillation counter.

For ^{32}P and ^{3}H samples precipitated on glass fibre discs 3.0 gm/litre PPO (2,5-diphenoxazole) and 300mg/litre POPOP (1,4-di-2(5-phenoxazoyl)benzene) in toluene was used as scintillation fluid and for ^{32}P samples in solution 100:1 toluene triton containing 5.0gm/litre PPO.

2.2.1.6. Dialysis of Solutions of Macromolecules

Dialysis tubing of various sizes (Medicell International Ltd., London, UK.) was prepared by boiling for 20 minutes in 10mM EDTA followed by rinsing with distilled water. The tubing was knotted at one end at least twice and then filled using two pasteur pipettes one as a pipette and the other as a funnel. A space was left at the top of the tubing to allow for changes in volume of the solution and the tubing sealed with again at least two knots. The dialysis tubing was then placed in a large volume of the relevant buffer which was stirred at 4°C. Dialysis proceeded for at least 24 hours and two changes of buffer.

2.2.1.7. Estimation of DNA Concentration and Purity by Spectrophotometry.

Using a Pye Unicam SP8-150 ultra violet/visible spectrophotometer and 1 cm path length quartz cells both the concentration and purity of DNA solutions were estimated.

The optical density (O.D.) of a DNA solution in water at 260 nm is ∿0.02 times its concentration in microgrammes per ml (µg/ml). A spectrum from 230 to 370 nm was used to

assess purity as follows, for pure DNA:

$OD260/OD280 \simeq 1.8$

$OD260/OD235 > OD260/OD280$

$OD320 \simeq 0$

Should these three conditions not hold then the concentration estimate will be in error.

2.2.1.8. The Storage of Bacterial Strains

In the short term (up to 2 months) bacterial colonies were stored on agar plates kept inverted at 4°C and sealed with Nescofilm (Nippon Shoji Kaisha Ltd., Osaka, Japan). Permanent storage was effected by dispersing bacterial lawns, grown from single colonies on agar plates, in 2ml sterile portions of equal amounts of L-broth and 80% glycerol and keeping at -80°C.

2.2.2. Preparation of DNA

2.2.2.1. Plasmid Minipreps.

The following method of rapidly preparing plasmid DNA was used for all plasmids derived from pBR322 including all pUC8 derived plasmids. Colonies were picked into 10ml of broth (L-broth for pBR322 type plasmids, YT broth for pUC8 plasmids) in a McCartney bottle containing the appropriate antibiotic (50µg/ml ampicillin for pUC8 derived plasmids) and were grown overnight at 37°C in a rotary incubator. The cells were spun down, in a bench centrifuge in the culture bottles, ∿6000g 15 minutes, then resuspended in 200µl of freshly prepared lysozyme solution (50mM glucose, 10mM EDTA, 25mM Tris-HCI pH 8.0, 2mg/ml lysozyme), transferred to a 1.5ml eppendorf tube

and stood at 0°C for 30 minutes. Next 600µl of freshly prepared alkaline SDS was added (0.2M sodium hydroxide, 1% SDS) and gently mixed until the solution became sticky and viscous, the mix was then stood in ice for at least 5 minutes. Sodium acetate solution (450µl of 3M pH4.8) was then added and mixed by inverting and standing on ice for 1 minute. The cell debris were then pelleted by centrifugation at 12,000g for 5 minutes (M.S.E. Micro Centaur bench centrifuge) and 1,100µl of clear supernatant was transferred into a fresh eppendorf. Isopropanol (500µl) was added, mixed and the DNA precipitated at -20°C for 30 minutes followed by centrifugation at 12,000g for 2 minutes. The pellet was resuspended in 200µl of 100 mM sodium acetate, 50mM Tris-HCl pH6.0 and precipitated with 500µl of ethanol in the usual way, resuspended and re-precipitated, washed in 80% ethanol twice, vacuum dried and resuspended in 50µl of HPLC grade water.

This method leaves large amounts of RNA in the DNA, which could be destroyed by treatment with RNAse A if desired. 5µl in a restriction digest allows down to 200 bp to be detected on agarose gels by ethidium bromide staining.

2.2.2.2. Large Scale Preparation of Plasmid DNA

10 ml overnight cultures of the plasmid bearing strains were grown up in L-broth containing the relevant antibiotic(s) on a rotary incubator. 250ml portions of L-broth also containing the relevant antibiotic(s) in baffled 1 litre flasks were innoculated with the overnight cultures (1:25 dilution) and grown with shaking at 37°C to an optical density at 650nm (OD650) of 0.9. Chloramphenicol was then added to 150-200µg/ml and the

incubation continued overnight.

The cells were pelleted by centrifugation at 15,000g for 10 minutes at 4°C, and resuspended in 5ml of 25% w/v sucrose in 50mM Tris-HCl pH8.0 at 0°C. Freshly prepared lysozyme solution (1ml of 10mg/ml lysozyme in the same buffer) was added, mixed gently, and then incubated with shaking at 37°C for 2 minutes and then at 0°C for 10 minutes. Next 5ml of 0.2M EDTA pH8.0 was added and shaking at 0°C continued for a further 10 minutes, followed by the addition of 1ml of 20% w/v SDS and gentle mixing at room temperature until the suspension clarified. Sodium chloride (3ml of 5M in 50mM Tris-HCl pH8.0) was then added and the whole was mixed throughly and kept on ice for at least 4 hours or usually overnight. Cell debris was removed by centrifugation at 38,000g for 90 minutes at 4°C followed by careful removal of the supernatant with a pipette.

The plasmid DNA was precipitated from the supernatant by the addition of 50% w/v polyethylene glycol 6000 (PEG 6000) in 50mM Tris-HCl pH8.0 and 50mM Tris-HCl pH8.0 to give final concentrations of 10% w/v PEG, 0.5M sodium chloride and 50 mM Tris-HCl pH8.0. The mix was then stood at 0°C for at least 4 hours or usually overnight then gently centrifuged (∿2000g) for 5-10 minutes at 4°C. The pellet was drained throughly and resuspended in 4-5 ml of T.E. buffer (10mM Tris-HCl pH7.5, 1mM EDTA) by gentle shaking at 0°C or if necessary by using a wide bore pipette. If this suspension was not clear it was briefly spun, it was kept at 0°C and could be stored.

The plasmid was purified by density gradient

centrifugation; freshly prepared ethidium bromide (10mg/ml in T.E.buffer) was added to a final concentration of 300 µg/ml and caesium chloride to 48.4% w/v or a refractive index of 1.3890, the mixture was then stood in ice for 30-60 minutes prior to centrifugation (Sorvall OTD65B Ultracentrifuge using a Beckmann VTi50 rotor) at 13,000 rpm for 30-60 minutes at 0°C. The clear supernatant was then centrifuged in the same apparatus at 40,000 rpm for 20 hours and the lower plasmid band removed with a wide bore needle and syringe. This final step was repeated and the plasmid isolated by extracting with caesium chloride saturated iso-propanol three times, dialysing against T.E.buffer and ethanol precipitation.

2.2.2.3. Large Scale Preparation of Phage λ DNA

Bacteriophage λ strain NM258 is temperature inducable and only enters the lytic cycle when warmed above ~40°C. A bacterial strain, containing phage λ NM258 (which had been stored as a glycerol) was streaked for single colonies. A colony was selected which grew at 30°C but not at 42°C and overnights in 2YMT broth (bactotryptone 16gm, yeast extract 10gm, sodium chloride 5gm, magnesium chloride 2.5gm to 1 litre) were grown at 30°C in a rotary incubator.

250ml portions of 2YMT broth in 1 litre baffled flasks were innoculated 1:25 with the overnights and grown with shaking at 30°C to an OD_{550} of 0.25. They were then induced by warming to 45°C and incubated at 44°C for 30 minutes with gentle shaking, and then shaken at 37°C for 4 hours. After centrifugation, 2,500g at 4°C for 10 minutes, the pellet was resuspended in 200ml

of 10mM Tris;HCl pH7.9., 100mM potassium chloride, 0.1 mM EDTA and frozen overnight at -70°C. If when thawed the mixture was not viscous 2-3 ml of chloroform was added and the mixture shaken. DNAseI (1mg in 2ml of the same buffer) was added and the mix stirred at 37°C until its viscosity was reduced. Cell debris were removed by twice centrifuging at 12000g for 10 minutes at 4°C isolating the supernatant each time. The phage was polyethylene glycol (PEG) precipitated by adding 5M sodium chloride, 50mM Tris-HCl pH8.0 and 50% w/v PEG 6000, 50mM Tris-HCl pH8.0 to final concentrations of 0.5M sodium chloride and 10% PEG, storing overnight at 0°C and gently centrifuging at ∿2000g for 10 minutes at 4°C. The pellet was well drained and resuspended in 20ml of chloride phage buffer (10mM Tris-HCl pH8.0 10mM magnesium chloride, 100mM sodium chloride), a few drops of chloroform were added if the suspension was to be stored. The phage was purified on a caesium chloride step gradient; caesium chloride in chloride phage buffer was made to 1.35, 1.55 and 1.7 gm/ml densities and a step gradient created by carefully layering these solutions into centrifuge tubes. To the top of the gradient was carefully added the phage suspension and the tubes were balanced by careful addition of chloride phage buffer. The gradients were then centrifuged at 25,000rpm in a MSE Prepspin 65 centrifuge with a 3x25ml rotor at 8°C for 90 minutes and the phage which collects between the 1.55gm/ml and the 1.7 gm/ml layers removed with a wide bore needle and syringe. The phage solution so obtained was saturated with caesium chloride and layered onto the bottom of a

step gradient made as before and recentrifuged and isolated as before.

The phage solution was then dialysed against phage buffer (potassium dihydrogen phosphate 3.0gm, disodium hydrogen phosphate 7.0gm, sodium chloride 5.0gm, magnesium sulphate to 1mM, calcium chloride to 0.1mM in 1 litre). The dialysed phage was phenol extracted gently to avoid shearing the DNA, ethanol precipitated with potassium acetate, and resuspended in T.E. buffer by gentle shaking at 0°C for 3-4 days.

In a typical preparation 4 litres of 2YMT broth gave $\sim 1 \times 10^{14}$ phage particles, by serial dilution and counting plaques on a lawn of susceptible bacteria, before the step gradient and $\sim$ 4.5mg of spectroscopically pure DNA after the ethanol precipitation.

2.2.2.4. M13 Minipreps.

The single stranded form of phage M13 and derivatives of it were prepared by the following method:

Overnight cultures of E.coli strain JM 101 were grown in YT broth made at double the normal concentrations (2 x YT broth). 2ml aliquots of 2xYTbroth in McCartney bottles were innoculated with the overnights 1:500 and infected with the M13 plaque by stabbing the plaque with a sterile toothpick and dropping the pick into the McCartney bottle. The cultures were grown overnight at 37°C on a rotary incubator.

The next day 1.5ml eppendorfs were filled with the cultures and spun in a bench centrifuge for 4 minutes, the supernatant was transferred to a fresh tube and the pellet was kept at 4°C to preserve the clone. The

supernatant was again spun and isolated, giving ∿1ml to which was added 200µl of 20% PEG 6000 and 225µl of 2.5M sodium chloride. After mixing and standing at room temperature for ∿20 minutes the phage was collected by spinning in a bench centrifuge for 3 minutes, removing the supernatant, spinning again and carefully removing the last dregs of supernatant with a capillary. The pellet was resuspended in 100µl of T.E. buffer (10mM Tris-HCl pH8.0, 1mM EDTA) phenol extracted, ethanol precipitated and resuspended in 20µl of T.E. buffer. 1µl should be easily visible by ethidium bromide staining on an agarose gel.

2.2.3. Standard Enzymatic Methods

2.2.3.1. Restriction Endonuclease Digestions.

Restrictions were carried out normally in one of the four buffers set out in the Cold Spring Harbour Molecular Cloning Manual (Maniatis et.al., 1982). (Table 2.), the choice of buffer being determined by the manual and the 1983 B.R.L. data poster (Bethesda Research Laboratories, Cambridge , U.K.) which was especially useful for deciding the appropriate buffer for mixed enzyme digestions. Normally enzymes were used at a concentration of 1-3u/µg of DNA except where higher ratios were recommended by the BRL data sheet and were incubated at the temperature recommended by the supplier for 1-4 hours. Nuclease free BSA was added to 100µg/ml to some digestions particularly those on a large scale or being incubated for a long time. RNAse A (10µg per reaction) which had been boiled for 15 minutes to destroy

Table 2 Restriction endonuclease buffers

Buffer	Components (mM)				
	Tris-HCl pH 7.5	Magnesium Chloride	DTT	Sodium Chloride	Potassium Chloride
Low salt (low)	10	10	1	0	0
Medium salt (med)	10	10	1	50	0
High salt (high)	50	10	1	100	0
SmaI	10 (pH 8.0)	10	1	0	20

any DNAse activity, was added to restrictions of plasmid DNA prepared by the miniprep method.

2.2.3.2. Ligation

Ligations, the joining of dsDNA molecules in the conventional manner by covalent bonds, were carried out in KLP Buffer (50mM Tris-HCl, pH7.5, 10mM magnesium chloride, 10mM DTT) containing 1mM ATP with 1u/µg λT4 DNA ligase. For joining two molecules the volume of the reaction was kept to a minimum, for circularising a molecule greater dilution was used. Blunt end ligations were incubated at 12°C for up to 2 days, often with the addition of more ATP and enzyme after one day. Sticky end ligations were incubated at 15°C overnight.

The enzyme E.coli DNA ligase 1u/µg was also used for ligations where it was important to ligate DNA but not RNA. The buffer used was 20mM Tris-HCl pH7.5, 4mM magnesium chloride, 10mM ammonium sulphate, 100mM potassium chloride, 10mM sodium chloride, 50µg/ml BSA, 0.1mM β-NAD, and the incubation was at 12°C overnight.

2.2.3.3. Making Double Stranded DNA Blunt Ended

2.2.3.3.1. Using S1 Nuclease.

This method is suitable for use on dsDNA molecules with protruding 3' or 5' ends or hairpinloops. It is known to give dsDNA molecules with ends blunt or ∿1 base in length, prolonged incubations with S1 nuclease can significantly shorten dsDNA molecules. S1 nuclease reactions were carried out in 200mM sodium chloride, 50mM sodium acetate pH4.4, 1mM zinc sulphate buffer at

37°C for 30 minutes and then 25°C for 30 minutes followed by the addition of ∿1/10th volume of 100mM EDTA to stop the reaction. A typical reaction containing ∿3µg of dsDNA average length ∿1000bp used 1000 u of S1 nuclease in 34µl with ∿10µg of tRNA added.

2.2.3.3.2. Using Mung Bean Nuclease.

This method is suitable for use on dsDNA with protruding 3' or 5' ends or hairpinloops. Its action is similar to that of S1 nuclease but it is less active and degrades dsDNA to a lesser extent. Mung bean nuclease digestions were carried out in 50mM sodium acetate pH5.2, 50mM sodium chloride, 2mM zinc chloride, 1mM DTT at 22°C for 20 minutes using ∿5u/µg DNA (Kroeker et al., 1976).

2.2.3.3.3. Using DNA Polymerase I

This method is suitable for use on dsDNA with protruding 3' or 5' ends, but not on dsDNA with hairpinloops. It was usually used after S1 nuclease digestion and is often referred to as "polishing" the ends of the dsDNA. The enzyme gives accurately blunt ends by its polymerase and exonuclease activities. DNA polymerase I polishing reactions were carried out in P.N.K. buffer (50mM Tris-HCl pH7.5, 10mM magnesium Chloride, 10mM DTT) at 12.5°C for 30 minutes in the presence of 0.2mM of each deoxynucleotide triphosphate and ∿½u of endonuclease free DNA polymerase I per µg of dsDNA.

2.2.3.4. Phosphorylation of DNA at the 5' ends using Poly-Nucleotide Kinase

Normally dsDNA has phosphate groups attached to its 5' ends, and these are necessary for successful ligation reactions. Should these phosphate groups have been removed or are missing e.g. commercial synthetic linkers are supplied without 5' phosphate groups, poly-nucleotide kinase may be used to replace them, and if supplied with γ^{32}p ATP will specifically "label" the 5' end of the dsDNA molecule.

Phosphorylation reactions were carried out in PNK buffer (50mM Tris-HCl pH7.5, 10mM magnesium chloride, 10mM DTT) in the presence of ATP. In a typical reaction 1µl (0.88µg) of EcoRI linkers were phosphorylated using 3 *u* of enzyme and 40µCi (8 x 10^{-12} mol) of γ^{32}P ATP.

2.2.3.5. Dephosphorylation of DNA at the 5' Ends using Alkaline phosphatase

Calf intestine akaline phosphatase specifically removes the phosphate groups from the 5' ends of DNA.

Phosphatase reactions were carried out in 50mM Tris-HCl pH9.0, 1mM magnesium chloride, 0.1mM zinc chloride and 1mM spermidine (Maniatis et.al., 1982). For dsDNA molecules with protruding 5' ends incubation was for 1 hour at 37°C using 0.2 *u*/µg DNA of enzyme. For blunt ended molecules or those with recessed 5' ends i.e. protruding 3' ends, incubation was for 15 minutes at 37°C then 15 minutes at 56°C then a second aliquot of enzyme was added and both incubations repeated.

2.2.3.6. Labelling dsDNA by Nick Translation

dsDNA molecules may be radiolabelled evenly throughout the molecule by the action of DNA PolymeraseI and DNAseI in the presence of deoxynucleotide triphosphates some of which carry a radio label, the process is known as "Nick Translation". Nick Translation reactions were carried out according to the protocol from Amersham International plc (Amersham UK) using the kits supplied by them. Protocol B was used with α^{32}P dCTP as the radiolabel and with 2½ hour incubations. The claimed specific activities were invariably achieved unless the DNA used was impure. The DNA was isolated from unincorporated nucleotides by column chromotography.

2.2.3.7. Labelling the 5' end of RNA

RNA moleucles were labelled using the Bedbrook method (Bedbrook and Bogorad 1976, Grunstein and Wallace 1978) which uses T4 polynucleotide kinase to specifically label the 5' terminal phosphate groups. More random labeling was achieved by partially hydrolysing the RNA before labelling.

Partial hydrolysis was carried out in hydrolysis buffer (5mM Tris-HCl pH9.5, 10mM EDTA, 0.1mM spermidine) by heating to 95°C for 5 minutes then cooling the reaction in ice.

The labelling reaction was carried out in kinase buffer (50mM Tris-HCl pH9.5, 12½ mM magnesium chloride, 10mM DTT, 5% glycerol) at 37°C for 30 minutes using 10μCi γ^{32}P ATP (2×10^{-12}mol) and 1μ of T4 polynucleotide kinase per μg of RNA. The reaction was stopped by the

addition of excess EDTA and the RNA isolated by the addition of 10µg of ^{t}RNA per µg of RNA, phenol extraction and electrophoresing the mixture through an agarose plug into a piece of dialysis tube using TAES buffer (400mM Tris- acetic acid pH7.8, 20mM sodium acetate, 2mM EDTA, 0.2% SDS) as electrolyte. After electrophoresis the contents of the dialysis tubing were ethanol precipitated and the labelled RNA recovered.

2.2.3.8. Adding Polynucleotide Tails to the 3' end of DNA with Terminal Deoxynucleotidyl Transferase

Calf thymus terminal deoxynucleotidyl transferase (TdT) was used to add "tails" of deoxynucleotides to the 3' ends of DNA molecules. It was found to be extremely sensitive to inhibition by any impurities in the reaction mix and the presence of any nucleases, so all buffers etc., used in TdT reactions were either commercially supplied or prepared with great care using HPLC grade water. All DNA used was extensively purified by phenol extraction, ethanol precipitation and resuspension in HPLC grade water. All radio label was vacuum dried and resuspended in HPLC grade water. No Tris or EDTA was allowed in any TdT reaction mix. The method used was based on those of Roychoudhury et.al., (1976), Deng and Wu (1981) and on the data sheets supplied by Bethesda Research Laboratories Ltd. (BRL) and P.L.Biochemicals inc.

TdT tailing reactions were carried out as follows: DNA (22.5µg of PstI cut pBR322, equivalent to 1.65 x 10^{-11} moles of 3' ends) was suspended in 39µl of TdT

buffer (100mM potassium cacodylate pH7.2, 2mM cobalt chloride, 0.2mM DTT. Supplied as 5xTdT tailing buffer by BRL) containing 60mg/ml of nuclease free BSA and ∿0.3 mM dGTP and 5μCi ^{3}H dGTP. After gentle mixing it was preincubated at 37°C for 5 minutes and two samples taken; one was trichloroacetic acid (TCA) precipitated as zero counts, the other spotted onto a filter and allowed to dry as reaction counts.

TdT enzyme 1μl 25u was added and gently mixed, incubation continued at 37°C for up to 30 minutes, samples being taken and TCA precipitated, and the reaction was stopped by the addition of 5μl of 0.1M EDTA. After scintillation counting the samples, the following formula was used to calculate the average length of the tails:-

$$\text{Bases added} = \frac{(\text{Sample counts} - \text{zero counts})}{\text{Reaction counts}} \times \frac{\text{moles dGTP}}{\text{moles 3' ends}}$$

Similar conditions were used for tailing DNA RNA hybrid duplexes, except for the addition of polyA to mop up the oligo dT and when tailing with dCTP and dTTP. which would otherwise provide 3'ends as primers for the tailing reaction,

2.2.4. Gel Electrophoresis

2.2.4.1. Agarose Gel Electrophoresis

DNA and RNA fragments of various sizes from 100bp up to 50kb were analysed for size on agarose gels of concentrations ranging from 0.5-1.5% (Table 3). Horizontal slab gels submerged in buffer were used throughout.

Table 3 Approximate size range of double stranded DNA fragments analysed on agarose gels of given concentrations

Concentration of agarose gel (%)	Size range of dsNDA analysed (kb)
0.5	20 - 1
0.7	10 - 0.5
1.0	7 - 0.35
1.2	7 - 0.2
1.5	5 - 0.1

2.2.4.1.1. Full Size Agarose Gels

Full size agarose gels of approximately 20x15 x 0.6cm were made using perspex moulds on clean glass plates and using a thin layer of vacuum grease to obtain a seal.

Perspex combs were suspended across the mould to produce the slots. All gels were normally made containing 1µg/ml of ethidium bromide and using Alec's gel buffer (40mM Tris-acetic acid pH7.7, 2mM EDTA), or where close bands needed to be resolved or very small amounts of DNA detected TBE buffer (10.8gm Tris, 5.5gm boric acid, 0.93gm EDTA ($Na_2 \cdot 2H_2O$) per litre pH∿8.3) was used as it gives sharper bands. In all cases the gel was run in the same buffer and ethidium bromide concentration as was used to make it. The gels were normally run at ∿1.6V/cm overnight. Normally the slot size was chosen to be the smallest which would contain the sample plus the beads (see below) so as to allow the maximum number of slots and the easier detection of faint bands. All DNA samples were loaded with the addition of ∿1/3rd volume of beads (10mM Tris-HCl pH8.0, 10mM EDTA, 30% w/v glycerol, 0.1% w/v agarose and either 0.1% w/v xylene cyanol and 0.1% w/v bromophenol blue, or 0.1% fast orange for when fragments <200bp are expected. The mixture was prepared, autoclaved, allowed to set, then extruded through a fine needle with a syringe twice). The DNA bands were detected by the orange fluorescence of the ethidium bromide DNA complex under UV light illumination at 254nm wave length, and were recorded by photography through a red-orange filter (Kodak 23A Wratten) with Polaroid Film type 667 3000 ASA. Gels with high levels of background were destained at

room temperature for 45 minutes in 1mM magnesium sulphate. Exposure times varied with the efficiency of the filter on the UV light source but normally an exposure of 7 seconds at f4 was used and as little as 5ng of dsDNA could be detected.

2.2.4.1.2 Agarose Minigels.

Minigel apparatuses were used essentially as suggested by the suppliers (Uniscience of Cambridge, Cambridge, U.K.) except normally the gels were poured to greater thicknesses than suggested, were run at lower currents to reduce heating, and ethidium bromide to 1µg/ml was added to the gels and the buffer to eliminate the staining step.

2.2.4.1.3. Alkaline Agarose Gels.

Two methods were used for alkaline agarose gels. The first method was that of Maniatis (1982) where the entire gel and running buffer was made alkaline. The second method used an ordinary agarose gel with no ethidium bromide, the DNA being made single stranded by treatment with alkali immediately before being loaded (McDonell, 1977). To 10µl of DNA mix was added 6µl of 1M sodium hydroxide at 0°C, after 5 minutes 6µl of 1M HCl was added also at 0°C and the sample immediately loaded and run as normal.

These gels were stained with acridine orange 30µg/ml in the running buffer for 30 minutes at room temperature, then destained in running buffer for ∿1hour also at room temperature. The DNA was detected by flourescence as for

ethidium bromide staining, dsDNA fluoresces green and ssDNA red when stained with acridine orange. N.B. alkaline gels destroy RNA.

2.2.4.1.4. Glyoxal Gels.

Glyoxalation was used to denature RNA (or DNA) prior to gel electrophoresis so that the RNA would be fully linear and its rate of migration would accurately reflect its size. The procedure used was essentially that of Maniatis (1982). Deionised glyoxal 6M was prepared as described, and was used to glyoxalate up to 20µg of RNA in the following mixture: 6M glyoxal 2.7µl, DMSO 8.0µl, 0.1M sodium dihydrogen phosphate pH7.0 1.6µl, water to 20µl, with incubation at 50°C for 1 hour.

1% Agarose gels were made up using 10mM sodium dihydrogen phosphate pH7.0 buffer and no ethidium bromide. The samples were loaded with the addition of 1/5th volume of loading buffer (50% glycerol, 10mM sodium dihydrogen phosphate pH7.0, 0.4% w/v bromophenol blue). The gel was run conventionally in the same phosphate buffer which was recirculated to maintain the pH. The RNA could be detected by staining with ethidium bromide.

2.2.4.1.5. Analysis of Band Patterns on Gels to Determine Fragment Sizes.

The sizes of dsDNA bands on agarose gels were estimated using various standard size markers; phage λ NM258 cut with restriction endonuclease HindIII, plasmid pBR322 cut with RsaI and plasmid pBR322 cut with AluI. The inverse of the logarithm (to the base 10) of the size of the fragment in base pairs was plotted against the distance

of migration of the band in millimeters. This gave a slightly S-shaped curve which approximates very closely to three straight line segments, and allows accurate estimates of dsDNA fragment sizes from 100-10,000 bp.

Other gels were analysed by similar methods.

2.2.4.1.6. The Recovery of DNA from Agarose Gels.

DNA was recovered from agarose gels by the DEAE paper method of Dretzen et.al., (1981). Whatmans DE81 chromatography paper was prepared by soaking in 2.5M sodium chloride for 3 hours, washing 5 times in water for 10 minutes each, washing in 1 mM EDTA, vacuum drying at 80°C and storing dry and protected. The DNA was isolated by cutting a slot next to the DNA band wanted, inserting a piece of prepared DEAE paper and electrophoresing the DNA band onto the paper. The DNA sticks efficiently to the DEAE paper unless very large amounts of DNA are used when the paper will saturate, or is subject to prolonged electrophoresis. The DNA was recovered from the paper by washing it in cold water blotting dry and then vortexing it into a slurry in elution buffer (20mM Tris-HCl pH7.5, 1mM EDTA, 1.5M sodium chloride) followed by incubation at 37°C for 2 hours. The mix was centrifuged through a pipette tip packed with siliconised glass wool, and the clear solution extracted with isoamyl alcohol saturated with elution buffer. The DNA was recovered by ethanol precipitation.

2.2.4.2. Acrylamide Gel Electrophoresis.

Polyacrylamide slab gels were used to obtain the end

labelled DNA fragments for sequencing and other labelled DNAs. Sequencing gels were used for actual DNA sequencing.

2.2.4.2.1. Acrylamide Slab Gels.

Polyacrylamide slab gels were made in two sizes:- 15 cm long x 18cm wide x 0.15cm thick and 36cm long x 18cm wide x 0.15cm thick, between glass plates and were run vertically top to bottom. The acrylamide mix used was 19:1 acrylamide bisacrylamide and TBE buffer (10.8gm Tris, 5.5gm boric acid, 0.93gm EDTA ($Na_2.2H_2O$) to 1 litre pH ∿8.3) was used for both the gel and the two reservoirs of running buffer. The procedures used were essentially those given by Maniatis (1982) with the DNA, invariably radiolabelled, loaded mixed with 1/3rd volume of glycerol dyes (10mM Tris-HCl pH8.0, 10mM EDTA, 80% v/v glycerol, 0.1% bromophenol blue, 0.1% xylene cyanol) and the gel run at approximately 14V/cm. The DNA bands were located by autoradiography normally after the gel had been dried, or frozen, however, if the DNA was to be recovered the gel was autoradiographed through a sheet of saranwrap immediately.

2.2.4.2.2. The Recovery of DNA from Acrylamide Slab Gels.

When the DNA band of interest had been located on the gel by autoradiography, the band was cut out of the gel using a sharp razor and the autoradiograph as a template. The excised band was placed in an eppendorf, mashed with a pipette tip and the DNA eluted with 1 volume of elution buffer (0.5M ammonium acetate pH8.0, 1mM EDTA) at 37°C overnight. After centrifugation at

10,000g, room temperature for 10 minutes the supernatant is isolated and a further ½ volume of elution buffer added to the pellet, vortexed and centrifuged as before. The combined supernatants were centrifuged through a pipette tip packed with siliconized glass wool and ethanol precipitated.

2.2.4.2.3. Acrylamide Gels for DNA Sequencing.

DNA sequencing gels were 38 x 18 x 0.035 cm in size and were made to 6% acrylamide mix (19 parts acrylamide to 1 bisacrylamide) and 8M urea, using TBE buffer for the gel and the reservoirs, essentially as given by Davis (1982). The gels were pre-electrophoresed, prior to loading, at ∿25mA (∿1,500V) for at least 30 minutes to equlibrate and to warm to ∿70°C. Improved regularity of temperature across the gel and hence reduced "smiling" was attained by clamping a 3mm thick metal plate to one side of the gel to act as a heat spreader.

Up to three loadings of the same sample could be achieved for sufficiently long DNAs using the migration of the marker dyes to allow an approximately 20 base overlap in sequence between loadings. Bromophenol blue migrated with fragments of ∿24 bases and xylene cyanol with fragments of ∿100 bases.

2.2.4.3. Gel Drying.

Agarose gels were dried wrapped in cellophane and acrylamide gels were dried sandwiched between a sheet of 3MM paper and a sheet of polythene. A commercial vacuum drying apparatus was used (Bio Rad laboratories) and

normally heat ∿60°C was applied.

2.2.4.4. Southern Blotting the Transfer of DNA from Agarose Gels to Nitrocellulose Paper.

The method used was basically that of Southern (1975). The DNA in an agarose gel was made single stranded by shaking gently in denaturing buffer (1.5M sodium chloride, 0.5M sodium hydroxide) twice for 15 minutes. The gel was then neutralised by gentle shaking in neutralising buffer (1.5M sodium chloride, 0.5M Tris-HCl pH7.4) twice for 15 minutes. A nitrocellulose filter was prepared by floating on distilled water until throughly wet, then soaking in 20 x SSC (20xSSC is 3M sodium chloride, 0.3M sodium citrate adjusted to pH7.0 with sodium hydroxide.) At no time were the filters handled except with rubber gloves and at the edges.

The gel was blotted in the apparatus shown in Fig. 2. The nitro cellulose filter was laid on the gel, the position and orientation of the slots marked and a sheet of 3MM paper was wetted and laid on top, taking care to avoid "short circuit" liquid paths around the gel. Capilliary action draws the 20 x SSC buffer up through the gel and filter into the paper nappies, transferring the DNA to the nitrocellulose where it sticks. The apparatus was always levelled and great care taken to exclude any air bubbles to avoid uneven transfer of DNA.

The process was complete after 12-24 hours. The DNA was fixed to the filter by washing in 6xSSC for 5 minutes, air drying on a sheet of 3MM paper and baking, sandwiched between 2 sheets of 3MM paper, at 80°C in a vacuum oven, for 2 hours. The filter was stored dry until used,

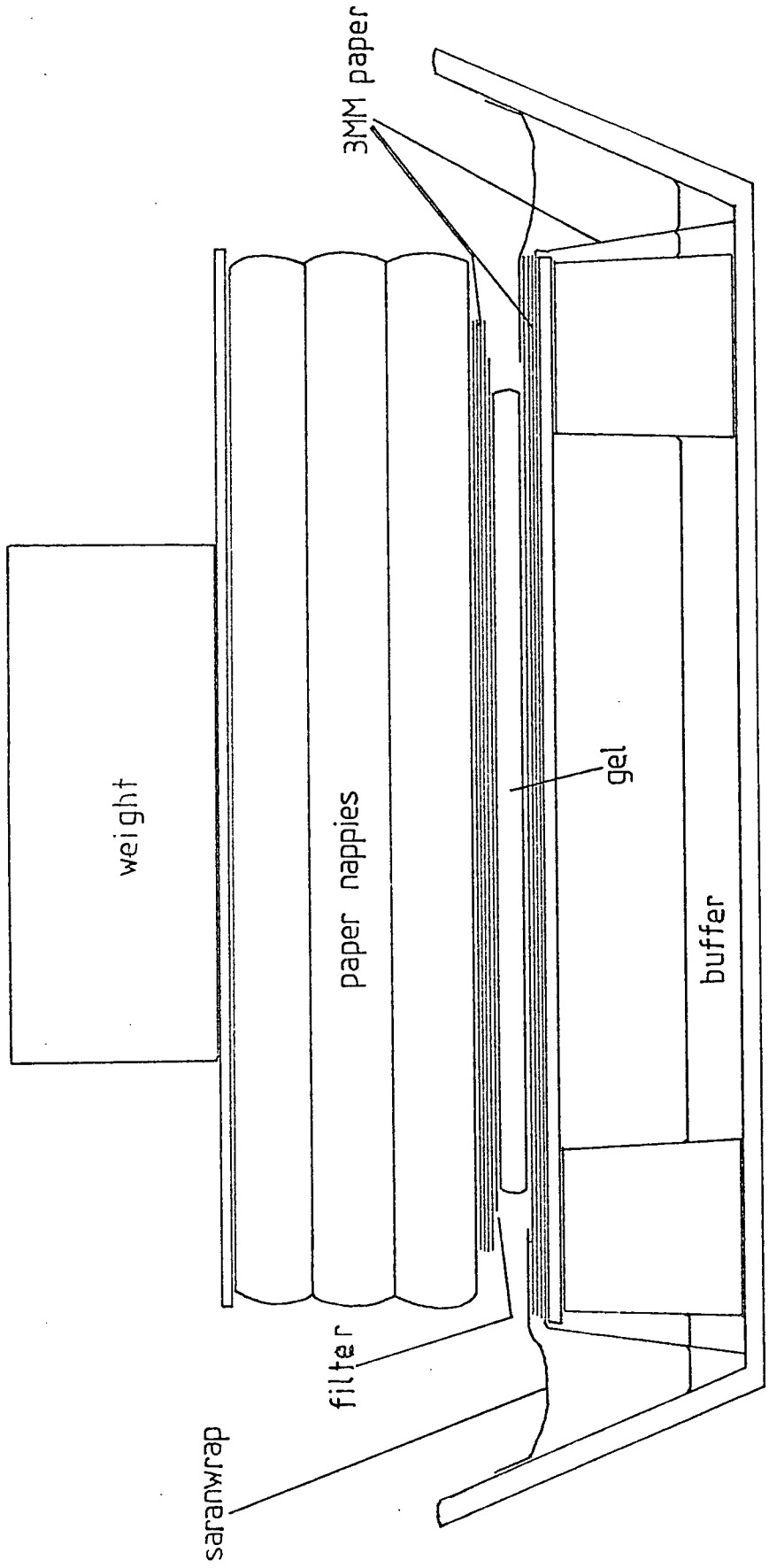


Fig.2. Apparatus for Southern Blotting

protected between sheets of 3MM paper.

2.2.5. Transformation of Strains of E.coli.

2.2.5.1. With Plasmids.

E.coli cells were made competent for transformation with plasmids essentially by the procedure of Dagert and Ehrlich (1979). Overnights of the relevant strain were grown at 37°C from single colonies and were used to innoculate, at a ratio of 1:50, 100ml aliquots of the relevant broth, normally L broth. These were then grown to an OD_{650} of 0.2 at 37°C on a fast shaker followed by chilling on ice. The cells were harvested by centrifugation at 3000g, 4°C for 15 minutes, the supernatant carefully removed and the cells resuspended in 50ml of ice cold 100mM calcium chloride, great care was taken to keep everything at 0°C. After 20 minutes the cells were then isolated as before and resuspended in 1ml of 100mM calcium chloride at 0°C and stored on ice until use. Competent cells were kept like this for up to 30 hours although transformation efficiencies fell appreciably by that time.

Transformations were effected by adding to 10µl of a suitable dilution of DNA in 100mM calcium chloride , 100µl of competent cell suspension. The mixture was incubated at 0°C for 10 minutes and at 37°C for 5 minutes, followed by the addition of 1ml of L broth and incubation with shaking at 37°C for 1 hour. Samples of this culture, normally 100µl, were then spread on nutrient agar plates containing the relevant antibiotics and marker systems, the rest of the culture was stored at 4°C for up

to 30 hours before being either discarded or spread onto other plates.

2.2.5.2. Transformation of E.coli Strains with Replicative Form M13.

Replicative form (ds)M13mp9 and its derivatives, may be treated as a plasmid and is transformed into suitable strains of E.coli in the same way as for other plasmids with a few changes in method. Strain JM101 was invariably used for M13, competent cells were prepared in the usual manner except that they were grown to an OD_{660} of 0.3-0.4 and some exponentially growing cells were transferred to a McCartney bottle containing nutrient broth and grown at 37°C during the experiment.

The transformation was effected by mixing 10µl of a suitable concentration of DNA with 100µl of competent cells at 0°C and leaving for at least 40 minutes. The mix was then heat shocked at 42°C for 2 minutes and the following were added in rapid succession:

10µl 0.1M IPTG (Isopropyl-β-D-thiogalactopyranoside)

50µl 2%X-Gal (5-Bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside) in dimethyl formamide

200µl of growing cells and

3ml of liquid soft agar (made with 6gm/l instead of 15gm/l of agar) at 42°C, after mixing it was poured onto an agar plate containing 50µg/ml ampicillin, which had been allowed to warm to room temperature, and incubated at 37°C. Blue plaques were M13mp9, clear plaques M13mp9 with an insert in the cloning site.

2.2.6. The Construction of a Pea Seed cDNA library.

The following protocol was used to produce the cDNA library and is based on the personal communications I received from I.M.Evans, R.R.D. Croy, R.Swinhoe, D.Bown, J.A.Gatehouse and others and Molecular cloning a laboratory manual (Maniatis et.al., 1982).

2.2.6.1. Preparation of poly(A)$^+$RNA.

Poly(A)$^+$RNA was kindly prepared by Mr. R. Swinhoe, by the method of Hall et.al., (1978), from pea cotyledons isolated 13-17 days after flowering, and was selected for the presence of polyA tails by passage down an oligo dT-cellulose column twice (Evans et.al., 1980).

2.2.6.2. Synthesis of Single Strand cDNA.

The protocol was based on that of Wickens et.al., (1978). 3.5µg of poly(A)$^+$mRNA was incubated at 37°C for 30 minutes in 100µl containing 40µl of mix 1 (Table 4). 30µ of human placental RNAse inhibitor, 0.377µg of oligo dT_{12-18} and 180u of Avian myeloblastosis virus reverse transcriptase (AMV).

2.2.6.3. Synthesis of Double Strand cDNA.

After the 1st strand incubation the mixture was heated to 100°C for 3 minutes then cooled rapidly in ice to denature the DNA:RNA duplexes. The second strand was then synthesised by incubation at 37°C for 1 hour by adding to the sscDNA mix; 90µl mix 2 (see table 4.), 50µCi of α^{32}PdCTP previously vacuum dried into the reaction vessel, 20u of DNA polymerase I large fragment (Klenow

Table 4 Standard Mixes for cDNA Synthesis

Solution	Volume (µl)	Final concentration in reaction mix
mix 1		
1M Tris-HCl pH 8.3	25	50mM
1M DTT	4	8mM
1M Magnesium chloride	4	8mM
1M Potassium chloride	50	100mM
20mM dATP	20	0.8mM
20mM dCTP	20	0.8mM
20mM dGTP	20	0.8mM
20mM dTTP	20	0.8mM
Water	37	
	200	
mix 2		
1M HEPES pH 6.9	50	100mM
1M Potassium chloride	100	200mM
20mM dATP	8	0.3mM
20mM dCTP	8	0.3mM
20mM dGTP	8	0.3mM
20mM dTTP	8	0.3mM
Water	318	
	500	

polymerase), water to 200µl. The reaction mixture was phenol extracted and the dscDNA separated from the unreacted nucleotides by passage down a 10ml column of G50 sepharose using 300mM sodium chloride, 50mM Tris-HCl pH7.5 buffer. Fractions containing cDNA were detected by scintillation counting and were pooled and ethanol precipitated with the addition of 20µg of tRNA using ammonium acetate.

2.2.6.4. Making the Double Strand cDNA Blunt Ended.

The dscDNA was resuspended in 26µl of resuspension buffer (1mM Tris-HCl pH7.6, 10mM sodium chloride, 0.05 mM EDTA) and treated with 1000µ of S1 nuclease in 34µl volume at 37°C for 30 minutes and 25°C for 30 minutes with 10µg of ^{t}RNA and in S1 nuclease buffer (200mM sodium chloride, 50mM sodium acetate pH4.4, 1mM zinc sulphate). The reaction was stopped by the addition of 1.5µl of 100 mM EDTA. The mix was phenol extracted ethanol precipitated with ammonium acetate and the DNA resuspended in 20µl of PNK buffer (50mM Tris-HCl pH7.5, 10mM magnesium chloride, 10mM DTT). The ends of the dscDNA were then "polished" by the addition of each deoxynucleotide triphosphate to 0.25mM and treatment with 1 u of endonuclease-free DNA polymerase I at 12.5°C for 30 minutes. The reaction was stopped by cooling to 0°C.

2.2.6.5. The Addition of Linkers and Ligation into a Vector of the dsDNA.

Commercially supplied EcoRI linkers ∿1µg were phosphorylated in 10µl and added to the blunt-ended dscDNA together with 5µl of 2 x PNK buffer ∿3.5u of T_4 DNA ligase

and ATP to 1mM. The mix was incubated at 15°C overnight in a total volume of ∿40µl, a further 1µl of 100mM ATP was added and incubation was continued a further 24 hours, before heating to 80°C for 5 minutes. The mix was then adjusted to 10mM Tris-HCl pH7.5, 10mM magnesium chloride 10mM sodium chloride and 5mM DTT in 200µl, 10u of restriction endonuclease EcoRI added and the whole incubated for 1 hour at 37° when a further 10u of EcoRI was added. After a further hour the mix was heated to 80°C for 5 minutes then cooled in ice with the addition of 20µg of tRNA and 5µl of 10% SDS.

The linkered dscDNA was separated from unligated linker fragments by passage down a 10ml column of 6B-CL sepharose in 10mM Tris-HCl pH7.6, 300mM sodium chloride, 1mm EDTA buffer. Fractions containing dscDNA were detected by scintillation counting, pooled and ethanol precipitated using ammonium acetate.

A sample of the vector pUC8 was prepared for use by restriction with excess EcoRI and the enzyme destroyed by heating the mixture to 75°C for 5 minutes. The efficiency of restriction was tested for by agarose gel electrophoresis and by transforming competent cells with the restricted vector. The quality of the "sticky ends" of the prepared vector were tested by ligating it to its self followed by agarose gel electrophoresis and by transforming competent cells.

The amount of dscDNA made could be calculated from the uptake of radiolabelled nucleotide in the second strand synthesis and sufficient vector to give a molecular ratio of cDNA : vector of 1:4 was added to the dscDNA. This mixture was incubated overnight at 12°C

in 20µl of ligase buffer (66mM Tris-HCl pH7.5, 1mM EDTA, 10mM magnesium chloride, 10mM β-mercaptoethanol) containing 1mM ATP and 1µ of T_4 DNA ligase.

2.2.6.6. Transforming and Screening.

The ligated dscDNA vector mix was used to transform competent cells of E.coli strain TB1. Positive colonies were detected by their resistance to 50µg/ml of ampicillin in the agar and by their inability to cleave X-Gal, which when cleaved produces a blue dye.

2.2.6.7. Preserving the cDNA Library.

The cDNA library was preserved in two forms : Individual white colonies were picked with sterile wooden toothpicks and streaked onto identical positions on each of three sterile gridded nitrocellulose filters, each laid on a nutrient agar plate (YT agar plus 50µg/ml ampicillin), and into a uniquely numbered well, in a sterile microtitre plate, containing 50µl of YT broth plus 50µg/ml ampicillin. The first method is essentially that of Hanahan and Meselson (1980), the second a variation on normal glycerols. By means of a master numbering system for the grids and microtitre plates, individual colonys could be uniquely located.

The streaks were grown at 37°C until well developed but not spreading over the filters, and overnight for the plates, and were then preserved: The filters were transferred to YT agar plates containing 25% v/v glycerol and 50µg/ml ampicillin and after incubation at 37°C for 2 hours were sealed with Nesco film and stored inverted at

-20°C. The microtitre plates had 50µl of 80% glycerol added to each well with mixing and after sealing with Nesco film were stored at -30°C.

Individual colonies of interest were normally picked from the filters or microtitre plates with sterile toothpicks, streaked out on YT agar plates containing 50µg/ml ampicillin and individually preserved as glycerols.

2.2.6.8. Producing Replica Filters.

Copies of the master filter were produced essentially as described by Maniatis et.al., (1982). A set of master filters were allowed to warm to room temperature and a set of nitrocellulose filters were sterilized with 6 pieces of 3MM paper per filter each slightly larger than the filters. The nitrocellulose filters were liberally wetted before sterilisation and kept sealed and flat to prevent drying out and buckling. Three pieces of the sterile 3MM paper were laid on an alcohol sterilised glass plate in a sterile air cabinet and were wetted lightly with sterile water. A master filter was removed from its plate and placed colonies up on the 3MM paper, a sterile nitrocellulose filter was then placed accurately over the master filter and marked uniquely to allow individual colonies to be located, a biro sterilised in alcohol was used. A further 3 sheets of sterile 3MM paper were placed on top and the whole stack firmly squeezed under a second alcohol sterilized glass plate. It was important to apply pressure evenly to obtain an even transfer of the colonies. The two filters were then peeled apart, the master filter was transferred to a fresh

glycerol agar plate (YT agar, 25% glycerol, 50µg/ml ampicillin) and grown at 37°C for 1 hour before being re-frozen. The replica filter was transferred to an agar plate (YT agar, 50µg/ml ampicillin) and grown at 37°C until the colonies were visible, then kept at 4°C.

2.2.6.9. Processing Replica Filters for Colony Hybridization.

Replica filters were processed by the procedure of Maniatis et.al., (1982); soaking with 10% SDS for 3 minutes followed by denaturing solution (0.5M sodium hydroxide, 1.5m sodium chloride) for 5 minutes, neutralizing buffer (1.5M sodium chloride, 0.5M Tris-HCl pH7.4) for 5 minutes and 2 x SSPE (0.36M sodium chloride, 20mM sodium dihydrogen phosphate pH7.4, 2mM EDTA) for 5 minutes. The released plasmid DNA was fixed to the filters by air drying on 3MM paper followed by baking sandwiched between sheets of 3MM paper at 80°C for 2 hours in a vacuum oven. The filters were then stored in the dry until used.

2.2.7. Hybridisation of Labelled Nucleic Acid Probes to Filter Bound Nucleic Acids.

This procedure was used to detect nucleic acids, both DNA and RNA, bound to nitrocellulose filters, such as lysed colony filters, Southern blots or Northern blots. The hybridization and all subsequent washing steps were carried out in sealed plastic bags in a shaking water bath at the relevant temperatures. Normally the hybridization step was carried out at 42°C.

The filters were wetted thoroughly in 6 x SSC,

(20 x SSC is 3M sodium chloride 0.3M sodium citrate to pH7.0 with sodium hydroxide), some colony filters stuck to their 3MM paper protectors during the baking step and a few minutes soaking freed them. The colony filters were transferred to plastic bags and 50ml of prewash solution (50mM Tris-HCl pH8.0, 1M sodium chloride, 1mM EDTA, 0.1% SDS) added, the bags sealed and the whole incubated with shaking at 42°C for 1-2 hours. This step was unnecessary for the blots.

The filters were next incubated at 42°C for at least 4-6 hours with shaking in prehybridising solution (50% deionized formamide, 5 x Denhardts solution, 5 x SSPE, 0.1% SDS, 100µg/ml denatured salmon or herring sperm DNA, 1µg/ml polyA. 50x Denhardts is 1% Ficoll 400, 1% polyvinyl-pyrrolidine, 1% BSA. 20 x SSPE is 3.6M sodium chloride, 200mM sodium dihyrogen phosphate pH7.7, 20mM EDTA) using ∿20ml for an 82mm filter disc and ∿40ml for a blot filter.

The labelled probe was then added in a small volume, DNA probes were first denatured at 100°C for at least 6 minutes, and hybridisation allowed to take place under the same conditions as prehybridisation according to the formula;

$$\frac{YZ}{25X} = Cot\tfrac{1}{2} \qquad \text{(Maniatis et.al., 1982)}$$

$Cot\frac{1}{2}$ is the time in hours for half a dsDNA probe to rehybridise to itself, X is the weight of probe added in µg, Y is the complexity of the probe which is normally its length in kilobase pairs and Z is the volume the hybridization is carried out in. Hybridization was normally

* A G+C ratio of 40% was assumed as it is a typical value for plant genes and when roughly checked fitted a legumin cDNA and a vicilin cDNA which had been sequenced in this department.

carried out for 1-3$\times$Cot$\frac{1}{2}$.

When the hybridisation step was finished the mix was poured off, it could be re-used to probe other filters after denaturing by boiling for 6 minutes but the protein in the mixture tended to coagulate so this was avoided where possible.

During the washing process care was taken to ensure that the filters never became dry. First the filters were washed 3 or 4 times for ∿5 minutes at room temperature in 2 x SSC, 0.1% SDS using ∿50ml per 82mm disc and ∿100ml per blot.

Next the filters were washed to the required stringency in the appropriate dilution of SSC plus 0.1% SDS as determined by the following formulae, assuming (G+C)%=40.

1) $Tm = 69.3 + 0.41\ (G+C)\%$ (Marmur and Doty 1962)

2) The Tm of a mismatched duplex decreases by 1°C for every increase of 1% in the number of mismatched base pairs (Bonner et.al., 1973).

3) $(Tm)_{\mu_2} - (Tm)_{\mu_1} = 18.5 \log_{10} \frac{\mu 2}{\mu 1}$

Where μ_1 and μ_2 are the ionic strengths of two solutions (Dove and Davidson (1962).

4) Each increase of 1% in formamide concentration lowers Tm by 0.7 degrees C

(McConaughty et. al., 1969; Casey and Davidson 1977)

Tm is the melting point of the duplex and washing was carried out at Tm-12°C.

Normally filters were washed at higher ionic strengths rather than lower temperatures as better results were obtained. The filters were washed for 1 hour twice under the calculated conditions, then allowed to air dry, before being mounted on 3MM paper and covered in saranwrap, ready

for autoradiography.

2.2.8. Autoradiography

Autoradiography was used to detect ^{32}P and ^{35}S labelled nucleic acids on filters and gels. All manipulations involving undeveloped film were carried out in a dark-room using safe-light illumination. The gels to be autoradiographed were always either dried wrapped in cellophane or covered with a sheet of saranwrap except ^{35}S sequencing gels which were always dried and exposed with no wrapping, filters were always covered in saranwrap. Radioactive ink was used to mark slots etc., and uniquely orientate the autoradiograph.

X-ray film (Fuji RX) was sensitised by exposure to a standardised low intensity flash of light and laid onto the gel/filter which was mounted on a glass plate. An intensifying screen (Dupont Cronex Lighting plus) was laid on top of the film followed by a second glass plate, and the whole clamped together with bulldog clips, or elastic bands. The assembly was then wrapped in three black plastic dust-bin liners and left to expose at -80° for up to a month.

The film was developed by washing in fresh Kodak X-Omat developer at room temperature for 5 minutes, cold water for 1 minute, Kodak fixer for 5 minutes, at which point the film could be exposed to light, and cold water for at least 30 minutes. The films were dried at room temperature.

2.2.9. Restriction Mapping

Plasmids containing unknown inserts were restriction

mapped initially by cutting out their insert and accurately determining its size, and by finding some restriction endonucleases which cut the plasmid only once preferably close to the cloning site and the insert not at all. The insert was then mapped in a series of mixed digestions using one of these enzymes and an enzyme which does not cut the plasmid. Normally only hexanucleotide sequence recognising enzymes were used and pentanucleotide and tetranucleotide enzymes were only used for exhaustive mapping or if sites for them were suspected from other data. The sizes of all DNA fragments were determined by agarose gel electrophoresis.

2.2.10. DNA Sequencing

Two methods of DNA sequencing were employed.

2.2.10.1. The M13 Method

DNA subcloned into the multipurpose cloning site of M13 mp9 was sequenced by the dideoxy method using DNA polymerase I and a commercially supplied 15 base primer essentially as given in the manual supplied by J.Messing (University of Minnesota) and B.R.L. Ltd. HPLC grade water was used throughout. Standard M13 single strand minipreps (2.5µl) were annealed to 1.5µl aliquots of 2ng/µl M13 primer in 10µl of Amersham buffer (10mM Tris-HCl pH8.5, 10mM magnesium chloride) by heating to 85°C for 5 minutes in a beaker of water and then allowing the whole apparatus to cool. After 15 minutes any evaporated liquid was spun down in a centrifuge, this mix could be stored frozen.

The following operations were carried out at 4°C: four tubes were labelled G, A, T and C and 1.5µl of the appropriate nucleotide mix added (Table 5). To the annealed mix was added 1µl (10µCi) of ^{35}S dATP and 1µl (4.5u) of DNA polymerase I large fragment (Klenow polymerase), and after gentle mixing 3µl aliquots were dispensed to each labelled tube. After brief centrifugation incubation was for 35 minutes at 30°C with 1µl of 0.5mM dATP "chase" being added after 20 minutes. The reactions were stopped by the addition of 1µl of 0.25M EDTA and freeze drying.

The DNA was resuspended in 6µl of formamide dyes (80% formamide, 10mM sodium hydroxide, 1mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue), at 90°C for 3 minutes, then loaded 2µl/track onto a sequencing gel.

2.2.10.2. The End Labelling Dideoxy Method

This is based on the methods of Maxam and Gilbert (1980), Sief et.al., (1980) and Smith (1980) and was compiled by G.Lycett. A dsDNA fragment is end labelled, cleaved, the two fragments isolated and in two reactions per nucleotide treated with DNA polymerase I and DNAseI and the appropriate dideoxynucleotide.

The DNA to be sequenced was restricted at the site to be end labelled, enzymes giving 5' protruding ends were preferred. Then after phenol extraction and ethanol precipitation ∿10µg was treated with phosphatase to remove the 5' terminal phosphate groups, kinase and 250µCi of γ^{32}PATP to specifically end label the 5' termini of the DNA, restricted a second time to produce two

Table 5 Nucleotide Mixes for M13 ^{35}S Sequencing Reactions

Solution	Mix T^o	C^o	G^o	A^o
0.5mM dTTP	1µl	20µl	20µl	20µl
0.5mM dCTP	20µl	1µl	20µl	20µl
0.5mM dGTP	20µl	20µl	1µl	20µl
HPLC water	20µl	20µl	20µl	1µl
	61µl	61µl	61µl	61µl
Dideoxynucleotide	1.0mM	0.5mM	0.5mM	0.125mM
	61µl	61µl	61µl	61µl

Each mix consists of equal volumes of the appropriate deoxynucleotide mix and the appropriate dideoxynucleotide mix.

Table 6 Nucleotide Mixes for End Labelled Dideoxy Sequencing Reactions

Mix

BG 1mM ddGTP

BA 1mM ddATP

BT 1mM ddTTP

BC 1mM ddCTP

FG 1mM ddGTP, 0.2mM dATP, 0.2mM dTTP, 0.2mM dCTP

FA 1mM ddATP, 0.2mM dGTP, 0.2mM dTTP, 0.2mM dCTP

FT 1mM ddTTP, 0.2mM dGTP, 0.2mM dATP, 0.2mM dCTP

FC 1mM ddCTP, 0.2mM dGTP, 0.2mM dATP, 0.2mM dTTP

fragments of unequal size and these fragments separated and isolated from a 4% acrylamide slab gel.

The dideoxy sequencing reactions were set up at 4°C as follows:

The DNA was resuspended in 40µl of Sief buffer (6.6 mM Tris-HCl pH7.5, 6.6mM magnesium chloride, 2mM DTT, 2mM sodium chloride) containing 40u of DNA polymerase I and 0.1µg of DNAseI. Eight tubes were marked as follows; BG, BA, BT, BC, FG, FA, FT, FC. The appropriate nucleotide mix (1.1µl) (Table 6) was dispensed into the tubes together with 4.5µl of the DNA enzyme mixture. After gentle mixing the tubes were incubated at 37°C for 30 minutes then frozen. To each of the backward (B) mixtures was added 1µl of 0.1M EDTA pH8.0 and the whole mixture mixed with the appropriate forward (F) mixture, the mixtures were then freeze dried. The DNA was resuspended in 10µl of formamide dyes per tube (80% formamide, 10mM sodium hydroxide, 1mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue) by heating at 90°C for 5 minutes. The entire tube was then placed in a scintillation vial and counted for Cherenkov counts, from this approximate exposure times for the autoradiographs were calculated according to the formula;

$$\text{Exposure time} = \frac{5 \times 10^5}{\text{counts per minute}}.$$

The DNA 1-2.5µl was loaded onto a sequencing gel after being heated to 90°C for 1 minute and quenched in ice. Up to 350 bases could be read on each fragment.

2.2.11. Preparation of Nuclease Free BSA

Nuclease free BSA was prepared according to the

method of Gonzales et.al., (1977) in which nuclease activity was destroyed by the action of acetic anhydride and the nuclease free BSA purified by dialysis against double distilled and deionised water and isolated by freeze drying.

2.2.12. Subcloning into and Determining the Orientation of Inserts in M13

Normally inserts from pUC8 type plasmids were subcloned into M13 mp9 by taking a mixture of the plasmid and the double stranded form of M13mp9 restricting with the enzyme whose site the insert was cloned into, phenol extracting, ethanol precipitating and ligating the mixture. After transformation the clear plaques could be roughly selected as large inserts give small plaques. M13mp9 also selectively clones smaller DNA fragments.

The orientation of the insert in an M13 subclone could be determined absolutely by preparing the double stranded form and restricting it with the appropriate enzyme, or relativly by the "C test" (Maniatis et. al. 1982). The "C test" is based on the effect whereby recombinant phages with their inserts in opposite orientations will hybridize via their inserts and the figure of eight structure resulting runs more slowly on an agarose gel than the unhybridized forms.

Good single strand M13 minipreps are selected by running them on an agarose gel containing no ethidium bromide, staining the gel for approximatly 1 hour in normal running buffer containing 1µg/ml of ethidium bromide and visualising under UV light. Good minipreps give only one band which should run slower than single strand M13. The rate of migration is relativly insensitive to insert size but size differences of <200bp can be detected.

"C tested" by being

Good minipreparations were annealed together by mixing equal amounts of DNA and incubating at 65°C for 1 hour prior to running on a similar gel to that used above against single preparations as a control.

2.2.13. Oligo Deoxynucleotide Cellulose Columns

The protocol used was based on that kindly supplied by Mr. R. Swinhoe (Dept. of Botany, University of Durham). All glassware was siliconized and heat sterilized at 170°C overnight, the plastic tubing was flushed out with sterile buffer and the buffers degassed. For DNA samples of up to 200µg of linearized and tailed pBR322 a column of 50mg of the appropriate oligo deoxynucleotide cellulose in a 1 ml syringe was used. The apparatus was set up as shown in Fig.3 using short runs of small bore tubing. The column was packed and equilibrated in salt buffer (400mM sodium chloride, 10mM Tris-HCl pH7.4., 1mM EDTA, 0.1% SDS) at ~7.5ml/hour. When using a new batch of cellulose the column was next equlibrated in no salt buffer (10mM Tris-HCl pH7.4, 1mM EDTA, 0.1% SDS) then re-equlibrated in salt buffer. The column was then regenerated by pumping for 20 minutes successively with sterile water, no salt buffer and salt buffer. When the output from the detector had stabilised the DNA was loaded in salt buffer and the column run until the output from the detector had returned to its original baseline. The buffer was then changed to no salt buffer and as soon as the detector output fell the collection vessel was changed. When the detector output had returned to its baseline collection was stopped and the eluted dN^+

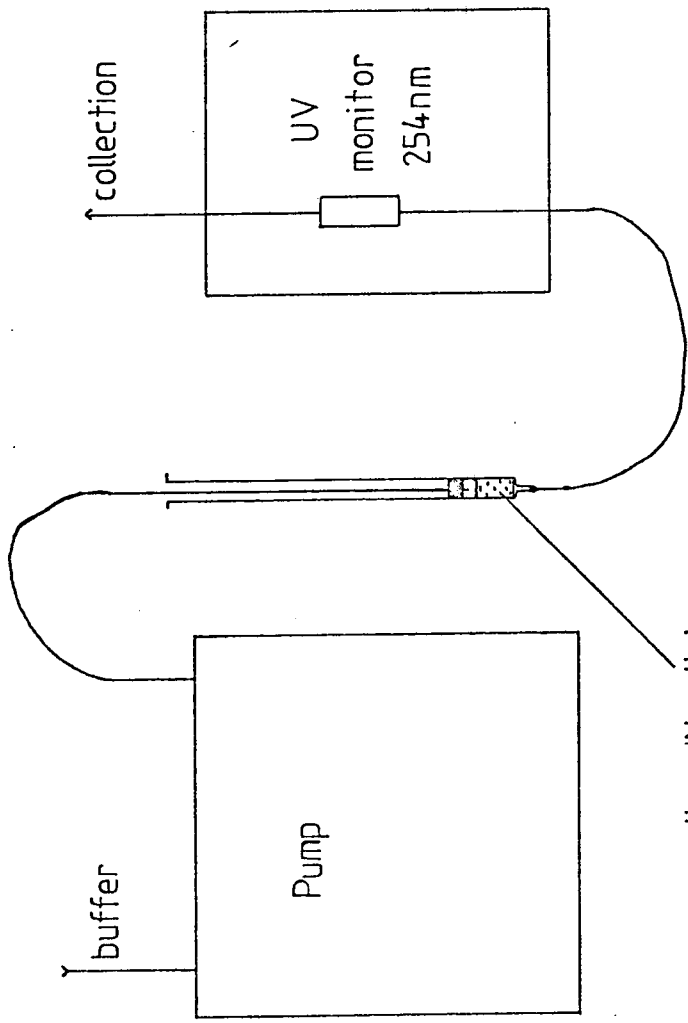


Fig.3. Oligo-deoxynucleotide Cellulose Column

DNA ethanol precipitated. The column was then regenerated as before then pumped with azide buffer (20mM sodium azide, 10mM Tris-HCl pH7.4, 1mM EDTA, 0.1% SDS) and stored sealed at 4°C.

2.2.14. Sephadex G50 and Similar Column Chromatography

Column chromotography with Sephadex G50 and similar substances such as Sepharose CL-4B, fractionates DNA roughly by size, such that the smaller the molecules the slower they pass through the column. Normally the columns were run in sterile 10ml plastic pipettes with a plug of siliconized glass wool at the bottom using ∿6ml of Sephadex. The columns were prepared by slurrying Sephadex in excess elution buffer (150mM sodium chloride, 50mM Tris-HCl pH7.5, 10mM EDTA, 0.1% SDS), and heating this mix to 65°C for 2 hours, this was then stored at 4° C until used. The column was settled and run using elution buffer. The DNA was loaded on by allowing the column to run nearly dry then adding the DNA with a small volume of elution buffer. The progress of the DNA through the column could be followed by a Geiger counter for ^{32}P labelled DNA and fractions of ∿½ml were collected. Labelled DNA was detected by scintillation counting and is eluted in the first peak. The second peak contains unreacted nucleotides or linker fragments etc.

CHAPTER 3

RESULTS

3.1. Preparation of the cDNA Library.

3.1.1. Production of Single-stranded Complementary DNA

Single-stranded complementary DNA (sscDNA) was made according to established protocols: The RNA template was prepared from developing pea cotyledons at the mid-development stage (13-17 days after fertilization). Total RNA was prepared from the cotyledons by the method of Hall et.al., (1978) and was then fractionated by chromotography on an oligo(dT)- cellulose column; the retained fraction was re-chromatographed to give the final poly(A)$^+$RNA preparation. Examination of this fraction by glyoxalation and electrophoresis on agarose gels showed that it was highly enriched in mRNA; less than 10% ribosomal RNA contamination was apparent (R.Swinhoe personal communication).

The poly(A)$^+$RNA was incubated with avian myeloblastosis virus (AMV) reverse transcriptase in a reaction including oligo(dT)$_{12-18}$ as a primer, and human placental ribonuclease inhibitor (RNAsein) to minimise degradation of the template during the reaction. The resulting sscDNA was analysed by agarose gel electrophoresis sometimes after glyoxalation; the cDNA was visualised either by ethidium bromide staining, or, if the sscDNA was made radioactive by incorporation of labelled nucleotide, by autoradiography of a dried gel (Figure 4). The size distribution of the sscDNA reflected that of the mRNA template, and showed the presence of several bands, indicating cDNA species of approximately 3100, 1800, 1500, 1000

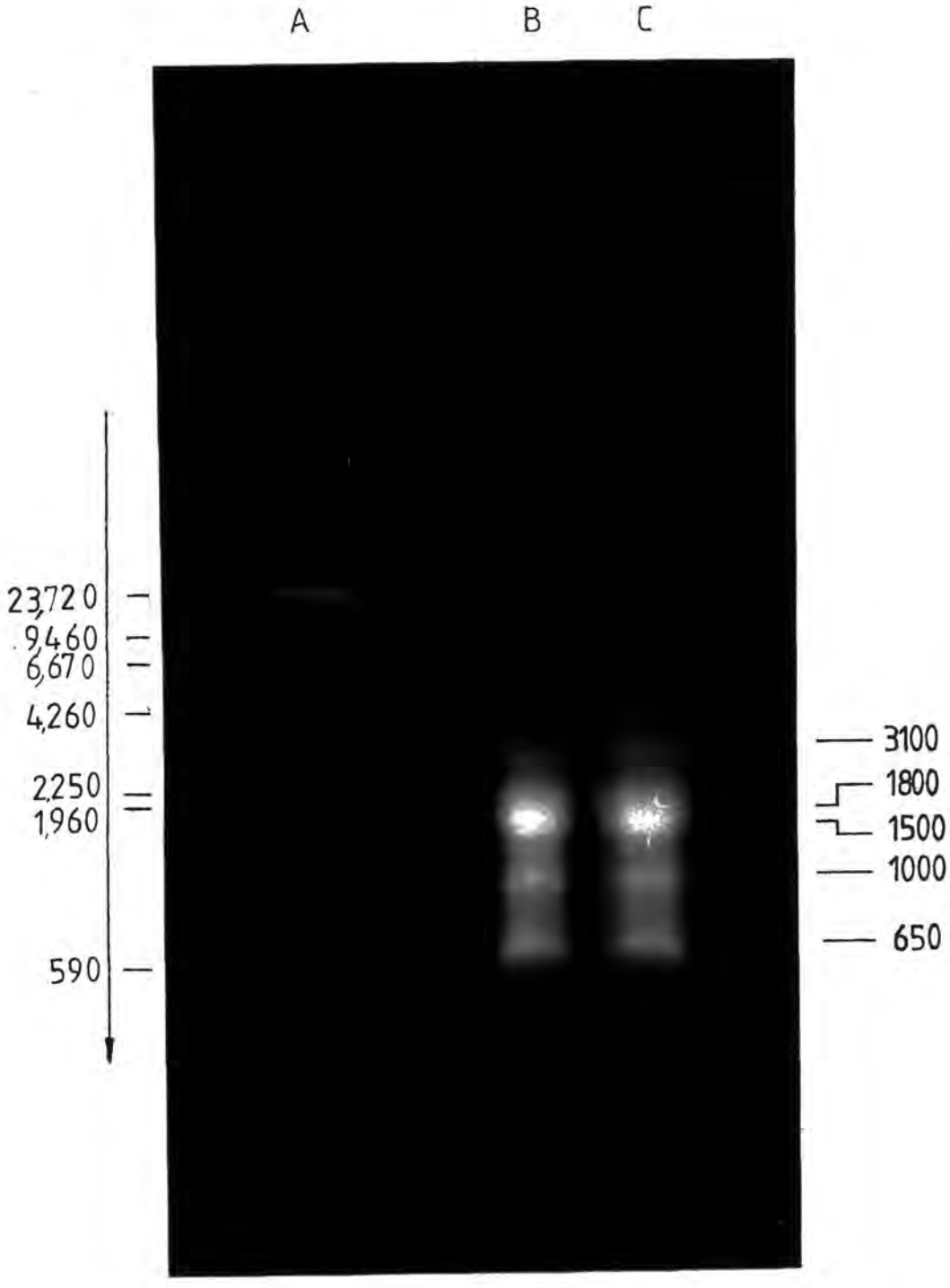


Fig.4. Single-stranded cDNA synthesised with and without Pyrophosphate.

A) Size marker: λNM258 cut with HindIII, this track was much clearer on the original photograph.
B) sscDNA synthesised with 4mM Sodium Pyrophosphate.
c) sscDNA without Pyrophosphate.

Band sizes in bases.

and 650 bases in length. These bands were considered to be full length transcripts of the major mRNA species present in the original poly(A)$^+$RNA.

This pattern was not qualitatively changed by the addition of sodium pyrophosphate to the sscDNA reaction mixture (Murray et.al.,1983). This inhibits the production of overlength sscDNA, which is thought to arise from the formation of a "hairpin loop" at the 3'end of the sscDNA followed by the AMV copying back down the sscDNA to produce an inverted repeat.

Attempts to produce sscDNA primed by fragments of plasmids carrying single stranded "tails" of dT residues on their 3'-end(s) (Okayama and Berg 1981) were essentially unsuccessful as it was not found possible to successfully or consistently produce these dT-tailed fragments. (See Section 3.7, Terminal deoxynucleotidyl transferase).

In a typical preparation of sscDNA made with α^{32}PdCTP as a label the yield estimated from the trichloroacetic acid precipitable radioactivity is 60-70% of the poly(A)$^+$RNA used.

3.1.2. Production of Double-stranded Complementary DNA

Double-stranded complementary DNA (dscDNA) was made by a number of different protocols (Figure 5). The established hairpin loop method uses the Klenow fragment of DNA polymerase I (Klenow polymerase) on sscDNA which has been heated to denature the DNA-RNA duplex. This method relies on the formation of a hairpin loop at the 3'-end of the sscDNA which acts as a primer for the synthesis of the second strand. This loop was later removed by treatment with S1 nuclease. This was the method chosen for the

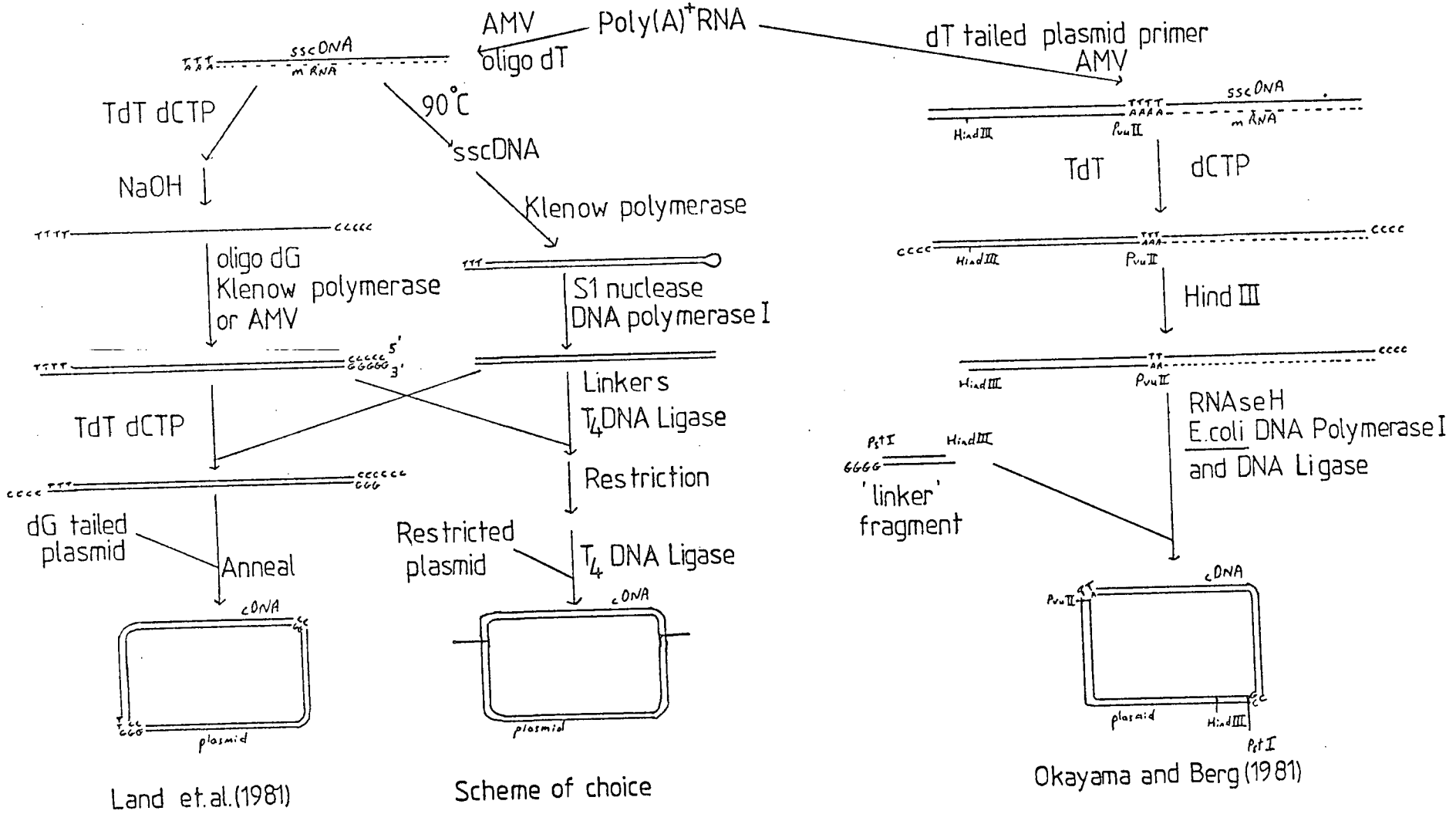


Fig. 5. cDNA Cloning Schemes

production of the cDNA library.

Using a ^{32}P labelled deoxynucleotide in the second strand synthesis with cold sscDNA, the yield estimated from the trichlorocetic acid precipitable radioactivity was ∿80% from sscDNA and ∿50% from poly(A)$^+$RNA. The band pattern observable with dscDNA on an agarose gel was much more ill defined' than for sscDNA (Fig.6). The size distribution ranged from ∿3000 base pairs downwards.

The other dscDNA methods tried (Murray et.al., 1983; Land et.al., 1981; Okayama and Berg, 1981) use a primer to initiate synthesis of the second strand. This is annealed to a tail of deoxynucleotides on the sscDNA, which is added using terminal deoxynucleotidyl transferase (See section 3.7). The RNA is cleaved off with sodium hydroxide before the annealing step or using the Okayama and Berg method, during the synthesis using RNAse H. AMV reverse transcriptase is used to synthesise the second strand except in the Okayama and Berg method where a mixture of endonuclease-free DNA polymeraseI, E.coli ligase and RNAse H is used.

Good results were obtained using either dG or dC tailed sscDNA and an oligo-deoxynucleotide primer12-18 bases long (Fig.7). However, tailed plasmid fragments were very inefficient as primers with very little incorporation of radiolabelled deoxynucleotide when using a dC-tailed plasmid fragment and dG-tailed sscDNA, as shown by autoradiography (Fig.8) and by virtually no trichloroacetic acid-precipitable radioactivity, using a method based on that of Okayama and Berg (1981).

As a consequence of this and of the problems with

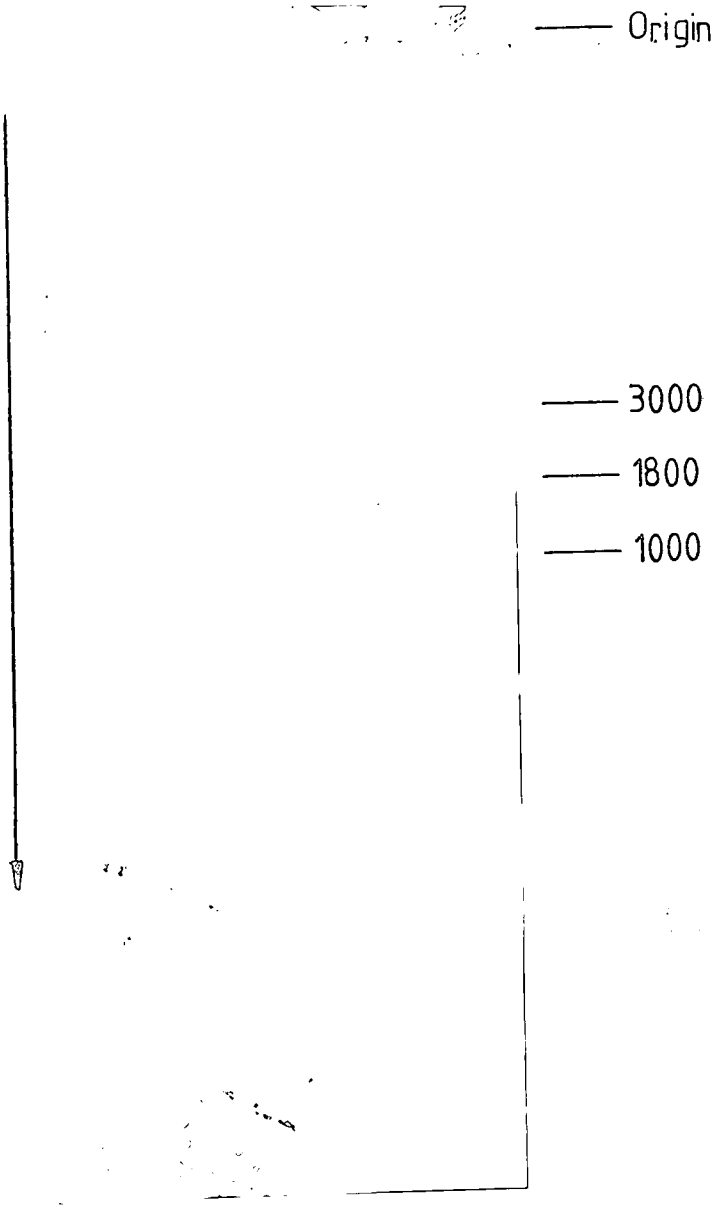


Fig.6. Double-stranded cDNA made by the "hairpin loop" method.

Approximate sizes of bands in base pairs.

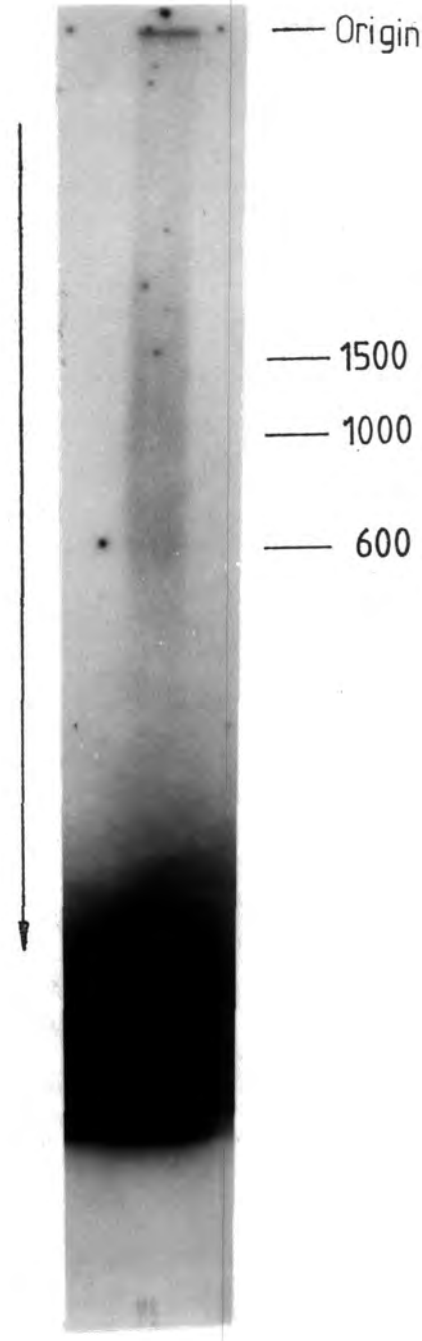


Fig.7. Double-stranded cDNA made from dC-tailed sscDNA primed with oligo-dG.

Band sizes in base pairs.

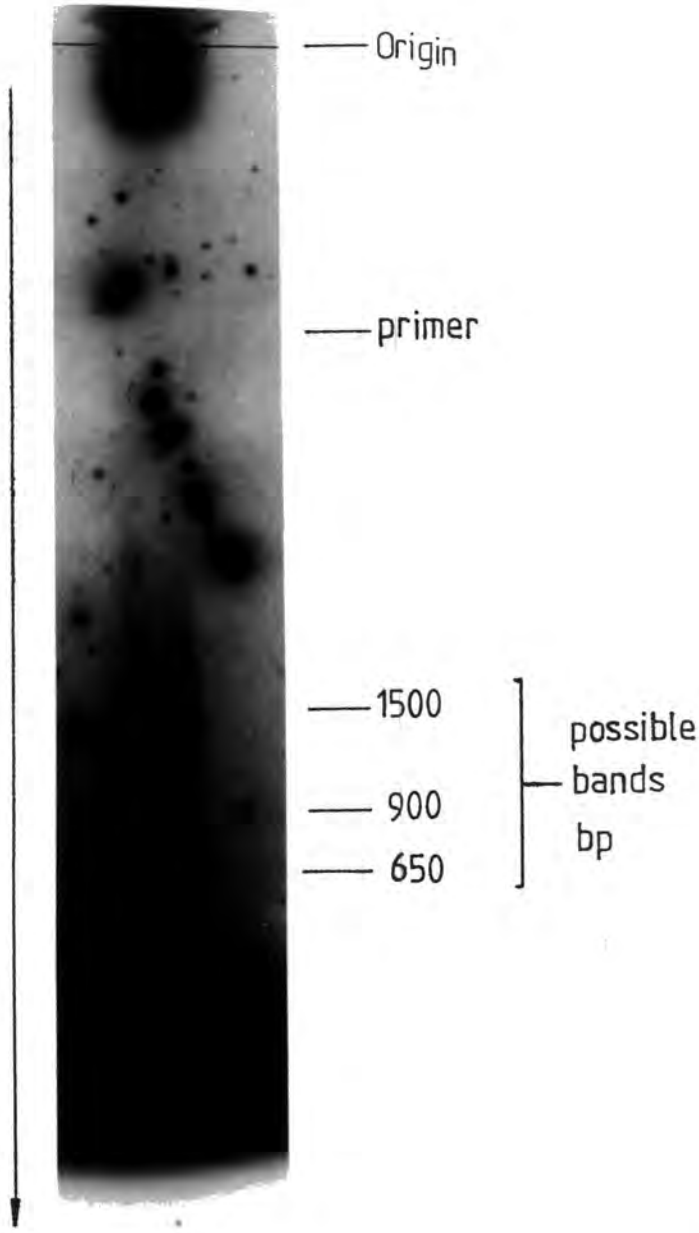


Fig.8. Double-stranded cDNA made from dC-tailed sscDNA primed with a dG-tailed poly(G)^{T} plasmid fragment.

Primer marks the estimated point to which the dG-tailed plasmid fragment ran.
Any dscDNA primed by this would have run slower than this, and would have left a trace between this point and the origin.

terminal deoxynucleotidyl transferase (Section 3.7) the "hairpin loop" method was adopted as method of choice.

3.1.3. cDNA Cloning

cDNA cloning was attempted by various methods, namely those of Land et.al., (1981), Okayama and Berg (1981), Murray et.al., (1983) and the use of synthetic linkers (M.Evans, personal communication).

The method of choice was removal of the dscDNA's hairpin loop with S1 nuclease and treatment with DNA polymerase I to make the ends blunt. Synthetic EcoRI linkers were ligated onto the dscDNA and after cleaving with EcoRI restriction endonuclease, it was ligated into the EcoRI site of plasmid pUC8. Upon transformation of *E.coli* strain TB1 (Which is a restriction/modification minus derivative of JM83 made especially for cDNA cloning with pUC8.) with half the resulting mixture, a total of 686 positive transformants were picked out starting from 3.48µg of poly(A)$^+$RNA. This is an overall efficiency of 3.6×10^2 transformants/µg of poly(A)$^+$RNA. The efficiencies of ligation and transformation are summarised in Table 8.

Another method which was investigated was that of annealing homopolymer tailed dscDNA into a complementarily tailed linear plasmid; a model system was tried using plasmid pRB322. pBR322 was restricted at the PstI site, tailed with dG and passed down an oligo(dC)-cellulose column to select for those molecules with a viable dG tail at, at least, one end. A second portion of pBR322 was dC tailed and passed down an oligo(dG)-cellulose column to again select tailed molecules. The dC-tailed linear plasmids will anneal to the dG-tailed linear plasmids to form

Table 8 Transformation efficiencies of the DNAs in the cloning of the cDNA library

DNA	Transformation efficiency Transformants per μg of pUC8
pUC8	5.7×10^5
pUC8 cut with EcoRI	3.9×10^2
pUC8 cut then ligated	2.2×10^4
pUC8 cDNA ligation	7.0×10^4 of which $\sim 3\frac{1}{2}\%$ were white colonies

Table 9 Transformation efficiencies of DNAs in the homopolymer tail annealing model system of cDNA cloning

DNA	Transformation efficiency Transformants per μg of pBR322
pBR322	4×10^6
poly(G)$^+$ PstI cut pBR322	$<5 \times 10^3$
poly(C)$^+$ PstI cut pBR322	$\sim 5.5 \times 10^3$
Annealed mix	$\sim 2 \times 10^4$ ‡
Annealed mix ampicillin sensitive	$\sim 7 \times 10^3$ ‡

‡ These efficiencies were calculated on the basis of the amount of poly(C)$^+$ PstI cut - pBR322 so as to compare directly with other figures.

circular structures containing two complete plasmids linked by the annealed homopolymer tails i.e. dimer plasmids. These were used to transform E.coli strain 910 and transformants selected for resistance to tetracycline. A second screen was carred out to find those colonies sensitive to ampicillin. The results are summarised in Table 9.

A random selection of six ampicillin sensitive colonies were minipreped. Restriction analysis of these plasmids showed them to be idential with similarly restricted pBR322 but the unrestricted plasmids were clearly different to unrestricted pBR322 by agarose gel electrophoresis. As predicted the Pst I sites, where tested for, were present.

A significant amount of pBR322-sized plasmid could be seen in each dimer miniprep: presumably these arise when a recombination event occurs between the two homopolymer stretches in the dimer plasmid to produce two plasmids which will be pBR322 with short homopolymer inserts in their Pst I site. As the inserts are small they will appear to be the same size as pBR322 but as the colonies are ampicillin sensitive the homopolymer insert must have caused a frame shift error in the ampicillin resistance gene. Although recombinations of this type may occur less often in a cDNA clone, it rules this method out. Interestingly no dimers when restricted with EcoRI gave fragments of two differing sizes suggesting, only pBR322 dimers with their origins of replication in the same orientation are viable.

The third method tried was that of Okayama and Berg (1981) and variations upon it. This involved priming the synthesis of one or both of the cDNA strands with fragments of plasmids bearing the appropriate homopolymer tails and circularising the hybrid plasmid by a ligation procedure. The unmodified Okayama and Berg method proved unsuccessful

due to inability to tail the first strand-primer plasmid fragment with dT. However, two variations were tried.

In one, the second strand synthesis was primed with Pst I-cut, dC-tailed, oligo(dG)-cellulose retained pBR322, the sscDNA having been dG-tailed and the mRNA hydrolysed away with sodium hydroxide. The resulting plasmids were treated with mung bean nuclease to produce blunt ends, then T4 ligase was used to circularise them. Transformation of E.coli strain 910 with the resulting mix gave no positive colonies. (Table 10). In the other modification, second strand synthesis was primed with an oligo(dC)-cellulose retained dG-tailed fragment of pBR322. The Okayama and Berg conditions and mix of DNA polymerase I, E.coli ligase and RNAse H were used for second strand synthesis. The resulting DNA had EcoRI linkers ligated onto it, the linkers were cleaved and the plasmid circularised. Transformation of E.coli strain 910 with this mix gave no positive colonies (Table 11).

3.1.4. Transformations

Transformations were carried out using the established calcium chloride method. Plasmid pUC8 has had E.coli strains JM83 and TB1 tailored for it, TB1 was chosen as it is restriction/modification minus. Phage M13 mp9 has had E.coli strains JM101 and JM103 tailored for it, JM101 was chosen as it gave consistent results when transformed with the replicative form (double stranded) of M13mp9 (J.Gatehouse, personal communication). For pBR322-like plasmids which do not use the lac system, a range of E.coli host strains are available, from which six were chosen and tested for transforming efficiency and

Table 10 Transformation efficiencies of DNAs in the cloning by tailed plasmid-primed second synthesis scheme

DNA	Transformation efficiency Transformants per µg of vector
pBR 322	$\sim 4.4 \times 10^5$
Pst I cut pBR 322 treated with mung bean nuclease and ligated	$\sim 3.8 \times 10^3$ of which $\sim 1.3 \times 10^3$ were ampicillin sensitive
cDNA ligation mix	$\sim 1.2 \times 10^2$ of which none were ampicillin sensitive

Table 11 Transformation efficiencies of DNAs in the cloning by Okayama and Berg-like conditions scheme

DNA	Transformation efficiency Transformants per µg of vector
pBR 322	$\sim 6.5 \times 10^5$
cDNA mix	$< 1 \times 10^1$

consistency. (Table 12). The two strains often recommended in the literature HB101 and DH1 gave good transformation efficiencies and consistency, but were slightly inferior to strain 910 which was the strain of choice. Strain JM83 gave low transformation efficiencies, strain GL80 gave poor reproducibility and was difficult to resuspend after centrifugation; strain RR28 gave erratic results.

3.1.5. Selecting, Preserving and Replicating the cDNA Library

The cDNA library was produced by ligating dscDNA into plasmid pUC8 and transforming E.coli strain TB1. E.coli strain TB1 has a disabled β-galactosidase gene which is complemented by the lac gene in pUC8 to produce a functional enzyme, which is able to cleave X-gal to produce a blue dye. The insertion of a fragment of DNA into the cloning site of pUC8 disrupts the translation of the lac gene, which contains the cloning site, and hence transformed colonies with an insert will produce no functional β-galactosidase and will be white. Plasmid pUC8 also contains an ampicillin resistance gene and so transformed bacteria are selected for by including ampicillin in the growth medium.

Positive transformants i.e. white or pale blue colonies were picked with sterile toothpicks and streaked in a grid pattern onto agar plates containing both X-gal and ampicillin. 768 colonies were picked, of which all grew, and 146 were obviously blue; some of the them were mixed i.e. partly white, partly blue or pale blue.

The library was picked for a second time onto sterile gridded nitrocellulose filters laid on top of agar containing ampicillin and X-gal, and into microtitre plates, each

Table 12 Transformation efficiencies of various *E. coli* strains with plasmid pBR 322

E. coli strain	Transformation efficiency colonies per µg pBR 322			
	Best	Worst	Average	Consistency
JM83	4.0×10^3	1.1×10^3	2.5×10^3	Good
HB101	4.1×10^5	1.1×10^5	2.7×10^5	Good
GL80	1.9×10^5	2.8×10^4	6×10^4	Poor
DH1	1.1×10^6	1×10^5	6.9×10^5	Fair
910	5×10^6	3×10^5	2×10^6	Fair
RR28	2×10^6	5.5×10^4	7.1×10^5	Poor

well containing 50µg/ml of ampicillin and carefully numbered, so as to be able to find any colony on a plate from its position on a filter. The white, pale blue and white parts of mixed colonies were selected, this left 686 colonies.

The library was preserved by growing up the plates and adding 50µl of 80% glycerol to each well of the microtitre plates, mixing and storing sealed at -30°C, and by growing up the filters and then transferring them onto agar plates containing ampicillin and 25% glycerol, and storing them sealed at -20°C.

The library was replicated by picking individual colonies with sterile toothpicks from the microtitre plate wells or the entire plate with a specially made tool allowing 48 samples to be taken simultaneously.

The library was also replicated by making copies of the master nitrocellulose filter, as detailed in the methods, on a second nitrocellulose filter. This method produced replicas of the master filters which could be used for colony hybridization. After regenerating the master filters further sets of copies could be made up to approximately 5 copies, before the transfer process became patchy and significant numbers of colonies were not present on the replica filters.

3.2. Colony Hybridisation

Sets of replica filters when lysed with SDS and alkali, neutralised and "fixed" can be hybridisation-probed with single-stranded or denatured double-stranded DNA or RNA (Grunstein and Wallis, 1979). The conditions of lysis convert the plasmid DNA in the cells to the single stranded

form and when the filter is dried the DNA binds effectively irreversibly to it. Complementary ssDNA in solution can then bind to the fixed DNA by conventional base pairing, the higher the homology i.e. the more correct base pairs the stronger the binding. Using radiolabelled probes under conditions where binding is favoured and then washing the filters under conditions of temperature and ionic strength chosen to reflect a certain degree of homology, those colonies with cDNAs sufficiently homologous to the probe DNA will bind it and will show up as dark spots when the filter is autoradiographed. The results of all the colony-hybridisation screens are collected in table 38.

3.2.1. Probing with mRNA

Poly(A)$^+$RNA isolated from 14 day old cotyledons of Pisum sativum L., the approximately same stage of development as the RNA used to produce the bank, was radiolabelled and used to probe a set of replica colony filters. The filters were washed in 1 x SSC, 0.1% SDS at 68°C which represents approximately 94% homology. The results listed in Table 13 show that 52% of the library contains cDNA inserts of sufficient length and are copies of messages sufficiently common to be detected under these conditions. The remainder represents cDNAs of rare messages, short cDNAs, delections or non-cDNA inserts.

Table 13 Results of probing the cDNA library with mRNA at ∿94% homology

Positive colonies:-

Intensity	Number	Percentage of the library
Strong	55	8.0%
Medium	45	6.6%
Weak	144	21.0%
Very Weak	112	16.3%
Total	356	52%

Table 14 Results of probing the cDNA library with pAD 4.4 (pDUB6), a main legumin cDNA at ∿94% homology

Positive colonies:-

			mRNA probe				
Intensity	Number	Percentage of library	S	M	W	VW	O
Strong	17	2.5%	3	6	5	1	2
Medium	14	2.0%	2	2	9	1	0
Weak	28	4.1%	1	3	14	9	1
Very Weak	0	0%	0	0	0	0	0
Total	59	8.6%	6	11	28	11	3

These tables should be read as follows:

Intensity: Intensity of the signal on the autoradiograph for a given colony when probed with the probe named in the title at the stringency given.

Number: The total number of colonies showing a signal of this strength.

Percentage of library: The total number of colonies showing a signal of this strength given as a percentage of the total number of colonies in the library.

mRNA probe: Cross index to the results of the colony screen with mRNA. The numbers represent the total number of colonies showing the given strengths of signal with the given probe and with mRNA. S= strong. M= medium. W= weak. VW= very weak. O= no detectable signal.
(Dubious = doubtful signals, these are not included into the totals.)

3.2.2. Probing with DNAs Coding for Main Legumin

The library was intially probed with the insert of plasmid pAD4.4 (pDUB 6) (Delauney,1984) which is a cDNA coding for main legumin of 1105 base pairs in length and covering the 3'-end of the message. The filters were washed in 1 x SSC, 0.1%SDS at 68°C, which represents approximately 94% homology. The results are listed in Table 14.

The cDNA library was later probed with two genomic clones of main legumin genes. These two experiments were carried out successively on the same set of replica filters so as to be able to directly compare the two sets of results. The filters used were slightly patchy and some colonies were missing. The library was probed at high stringency with the legumin C genomic clone as it was suspected at the time that this gene was not transcribed at this stage of development and that it was significantly different to legumin gene A which was known to be transcribed at this stage of development. Thus if this were true then few or no colonies should hybridise to the legumin C probe at high stringency. The low stringency probe with legumin gene A was to provide a control to compare the legumin C results to and to find any colonies bearing cDNAs for the 5'end of the legumin mRNA or showing weak homology to legumin mRNA.

The filters were first probed with the radiolabelled HindIII fragment of pAS2 (pDUB26) (Lycett et.al., 1985) which comprises the legumin C gene from just after the start of transcription to well beyond the end of it. The filters were washed in 0.1 x SSC, 0.1% SDS 65°C which represents approximately 99% homology. After autoradiography the filters were probed with the HindIII to BamHI fragment of pDUB21 (Lycett et.al., 1985) which comprises the legumin A gene from just after the start of transcription to well beyond the end of it. Approximately ten times the amount of radioactivity was used compared to the legumin C probe. The filters were washed in 1 x SSC, 0.1% SDS at 55°C which represents approximately 81% homology. The results are summarised in Table 15.

One legumin C positive was not a legumin A positive, this was pLG2, 147 which was a legumin C very weak,

Table 15 Results of probing a set of filters sequentially with legumin C at high stringency and legumin A at low stringency

Intensity	Probe	
	Legumin C	Legumin A
Strong	2	10
Medium	7	11
Weak	18	21
Very Weak	14	3
(Dubious)	(2)	(4)
Total	41	45

For information on individual colonies see the data table in the appendix.

Table 16 Results of probing the cDNA library with a DNA coding for the (big) legumin J gene at ∿91% homology [pJC 5-2]

Positive colonies:-			mRNA probe				
Intensity	Number	Percentage of library	S	M	W	VW	O
Strong	15	2.2%	7	3	3	2	1
Medium	4	0.6%	1	0	1	1	1
Weak	7	1.0%	1	1	3	2	0
Very weak	2	0.3%	0	1	1	0	0
(Dubious)	(6)	(0.9%)	(2)	(1)	(2)	(0)	(1)
Total	28	4.1%	11	6	10	6	3

See legend on page 101.

also pLG3, 114 which was a legumin C dubious is not a legumin A positive. These probably represent errors in reading the legumin C autoradiograph.

Seven legumin A positives were not legumin C positives they were pLG 1,67, pLG1,68, pLG 1,149, pLG3,155 and pLG 3,177 which were legumin A weak and pLG 4,7 and pLG 4,70 which were legumin A very weak. A further six legumin C positives showed large increases in intensity when probed with legumin A. They were pLG 2,133, pLG 3,20, pLG 3,186 and pLG 4,11 which were legumin C weak and legumin A strong and pLG 3,163 and pLG 3,165 which were legumin C very weak and legumin A medium.

3.2.3. Probing with DNA Coding for Big Legumin

The insert of plasmid pJC5-2 is a subclone of the genomic clone λJC5 (*N. Ellis*, personal communication). It consists of the 1.9kb EcoRI fragment of λJC5 and represents part of the legumin J gene, one of the big legumin genes.

A set of replica filters were probed with this insert and washed in 1 x SSC, 0.1% SDS at 65°C which represents approximately 91% homology. The results are summarised in Table 16. Some degree of overlap with the main legumin positives is seen but only pLG 3,161 gives convincing results for both. Colony pLG 2,75 which is a strong positive with this probe was also a strong positive for the insert of pAD 3,4 (pDUB7) which is a cDNA of 47k vicilin messenger RNA.

3.2.4. Probing with DNAs Coding for Vicilin

Three types of vicilin polypeptides of initial molecular weight around 50,000 are known to occur in large amounts in pea, differing by minor differences in size and in susceptibility to post-translation cleavages (Lycett et.al., 1983; Delauney 1984). There is also a related protein of higher molecular weight ∿71,000, called convicilin.

3.2.4.1. Probing with DNA Coding for the Mr∿47k type B Vicilin

The insert of plasmid pAD3.4 (pDUB7) (Delauney,1984) is a cDNA of a message coding for the 47k type B vicilin. It is ∿1100bp in length and lacks only the last 300-400 bp from the 3'-end. A set of filters was probed with this insert and washed in 1 x SSC, 0.1% SDS at 68°C which represents approximately 94% homology. The results are summarised in Table 17. The same set of filters was then re-probed with the same probe and washed in 1 x SSC, 0.1% SDS at 55°C, representing approximately 81% homology. The results are summarised in Table 18. The same set of filters was then re-probed with the same probe and washed in 1 x SSC, 0.1% SDS at 48°C representing approximately 74% homology. The results are summarised in Table 19. However, under these washing conditions the background on the autoradiograph was sufficiently high to obscure some of the weaker signals and little further information was obtained.

3.2.4.2. Probing with DNA Coding for the Mr∿50k type C Vicilin

The insert of plasmid pAD2,1 (pDUB9) (Delauney,1984)

Table 17 Results of probing the cDNA library with a cDNA coding for type B vicilin at ∿94% homology [47k vicilin; pAD3.4 (pDUB7)]

Positive colonies:-			mRNA probe				
Intensity	Number	Percentage of library	S	M	W	VW	O
Strong	12	1.8%	9	1	2	0	0
Medium	13	1.9%	6	2	4	1	0
Weak	13	1.9%	1	3	7	1	2
Very Weak	1	0.1%	1	0	0	0	0
Total	39	5.7%	17	6	13	2	2

Table 18 Results of probing the cDNA library with a cDNA coding for type B vicilin at ∿81% homology [47k vicilin; pAD3.4 (pDUB7)]

Positive colonies:-

Intensity	Also present at 94% homology	Not present at 94% homology	Total	Percentage of library	mRNA probe S	M	W	VW	O
Strong	27	0	27	3.9%	16	3	8	0	0
Medium	12	3	15	2.2%	3	3	4	2	3
Weak	1	21	22	3.2%	9	16	4	3	0
Very Weak	0	9	9	1.3%	0	1	4	0	5
Total	40	37	73	10.6%	28	23	20	5	8

See legend on page 101.

Table 19 Results of probing the cDNA library with a cDNA coding for type B vicilin at ∿74% homology [47k vicilin; pAD3.4 (pDUB7)]

Positive colonies:-

Intensity	Number	Percentage of library
Strong	36	5.2%
Medium	8	1.2%
Weak	17	2.5%
Very Weak	3	0.4%
(Dubious)	(19)	(2.8%)
Total	64 (83)	9.3% (12.1%)

Table 20 Results of probing the cDNA library with a cDNA coding for type C vicilin at ∿94% homology [50k vicilin; pAD2.1 (pDUB9)]

Positive colonies:-

			mRNA probe				
Intensity	Number	Percentage of library	S	M	W	VW	O
Strong	22	3.2%	11	2	5	4	1
Medium	15	2.2%	9	3	3	0	0
Weak	8	1.2%	3	0	1	1	3
Very Weak	12	1.7%	2	1	3	2	4
Total	57	8.3%	25	6	12	7	8

See legend on page 101.

is a cDNA of a message coding for the Mr50k type C vicilin. It is ∿1500bp long and is very nearly a complete copy of the message. A set of filters was probed with this insert and washed in 1 x SSC, 0.1% SDS at 68°C which represents approximately 94% homology. The results are summarised in Table 20.

Eighteen of these positives were not positives in the type B vicilin screens. The same set of filters was re-probed with the same probe and washed in 1 x SSC, 0.1% SDS at 55°C representing approximately 81% homology. The results are summarised in Table 21.

The same set of filters was re-probed with the same probe and washed in 3 x SSC, 0.1% SDS at 55°C which represents approximately 72% homology. The results are summarised in Table 22.

Of the positives detected 3 weak, 4 very weak and 1 dubious were good positives for other probes and hence were probably spurious.

3.2.4.3. Probing with DNA Coding for the Mr∿50k Type A Vicilin.

The insert of plasmid pRC 2,2,1 (pDUB2) (Croy et.al., 1982; Lycett et.al., 1983) is a cDNA of a message coding for the Mr50k type A vicilin It is ∿900bp in length and stretches from the 3'-end to just into the α-subunit, i.e. it lacks ∿600bp from the 5'-end. A set of filters was probed with this insert and washed in 1 x SSC, 0.1% SDS at 68°C, representing approximately 94% homology. The results

Table 21 Results of probing the cDNA library with a cDNA coding for type C vicilin at ~81% homology [50k vicilin; pAD2.1 (pDUB9)]

Positive colonies:

Intensity	Also present at 94% homology	Not present at 94% homology	Total	Percentage of library	S	mRNA probe M	W	VW	
Strong	27	0	27	3.9%	14	4	5	3	
Medium	13	0	13	1.9%	8	1	3	1	
Weak	6	4	10	1.5%	1	1	1	2	
Very Weak	6	20	26	3.8%	6	3	7	5	
Total	52	24	76	11.1%	29	9	16	11	

Table 22 Results of probing the cDNA library with a cDNA coding for type C vicilin at ~72% homology [50k vicilin; pAD2.1 (pDUB9)]

Positive colonies:-

Intensity	Previously detected	Newly detected	Total	Percentage of library	mRNA positive	mRNA zero
Strong	23	0	23	3.4%	22	1
Medium	16	0	16	2.3%	16	0
Weak	11	2	13	1.9%	9	4
Very Weak	18	20	38	5.5%	26	12
(Dubious)	(9)	(24)	(33)	(5.0%)	(20)	(13)
Total	68(77)	22(46)	90(123)	13.1%(18.1%)	73(93)	17(30)

See legend on page 101.

are summarised in Table 23. Of the positives detected thirteen were not detected with the other two vicilin probes.

3.2.5. Probing with the Insert of Plasmid pRC2,2,26

The insert of plasmid pRC2,2,26 is a cDNA of about 240bp and was cloned into the BamHI site of plasmid pBR322. It is known to select, in hybrid release translation, for a message producing a water soluble (albumin) protein of Mr∿25,000 (R.R.D.Croy and J.A.Gatehouse, personal communication). This message was known to be reasonably abundant and to code for neither legumin nor vicilin. It was suspected that it could code for one of the major albumins Mr∿25,000 and ∿24,000 or the seed lectin Mr∿23,000 before post-translational cleavage.

A set of filters was probed with this insert and washed in 1 x SSC, 0.1% SDS at 68°C, representing approximately 94% homology. The results are summarised in Table 24. Nine colonies constituting ∿1.3% of the library were convincingly positive, all of which except two were also mRNA positive.

3.2.6. Probing with DNA Coding for the Seed Lectin of Phaseolus vulgaris.

The insert of plasmid pPVL 134 is a cDNA of the message coding for the seed lectin of Phaseolus vulgaris which is know to have antigenic and sequence similarities to the pea seed lectin (Hoffman et.al., 1982). The insert is ∿960bp long and contains the entire coding sequence from just 5' of the start codon to the poly A tail. A set

Table 23 Results of probing the cDNA library with a cDNA coding for type A vicilin at 94% homology [50k vicilin; pRC 2,2,1 (pDUB2)]

Positive colonies:-

Intensity	Number	Percentage of library	mRNA probe S	M	W	VW	0
Strong	11	1.6%	5	3	2	1	0
Medium	17	2.5%	10	2	3	1	1
Weak	16	2.3%	3	4	6	2	1
Very Weak	2	0.3%	0	1	0	0	1
(Dubious)	(2)	(0.3%)	(0)	(0)	(0)	(0)	(2)
Total	46	6.7%	18	10	11	4	3

Table 24 Results of probing the cDNA library with the insert of the plasmid pRC 2,2,26

Colony pLG number	pRC 2,2,26 probe intensity	mRNA probe intensity
1.139	strong	weak
2.56	weak	zero
2.73	weak	weak
2.107	medium	very weak
3.86	medium	zero
3.129	strong	very weak
4.10	strong	weak
4.14	strong	weak
4.71	strong	weak
(4.81	dubious	very weak)

See legend on page 101.

of filters was probed with this insert and washed in 3 x SSC, 0.1% SDS at 65.5°C, which represents approximately 82% homology. No strong signals were seen on the autoradiograph and the background was fairly high, however two possible positives were detected; pLG 2,47 which is a mRNA weak positive and pLG3,127 which is a mRNA medium positive. The filters were then re-probed with the same probe and washed in 3 x SSC, 0.1%SDS at 55°C which represents approximately 72% homology. No positive signals were detected by autoradiography, each colony showing roughly uniform background labelling.

3.2.7. Probing with cDNA Enriched in Sequences Coding for the Pea Seed Major Albumins

Single-stranded, labelled cDNA made from mRNA enriched in sequences coding for pea seed major albumins was prepared by Mr. R. A. Ragab. The enriched RNA was prepared by immunoprecipitating polysomes from developing pea cotyledons with antibodies raised against pea seed major albumins, extracting the RNA and passing it down an oligo(dT)-cellulose column and, collecting the retained poly(A)$^+$ fraction. Due to the very limited amount and low activity of the probe only one of the four filters in the set (Number 4) was probed. The filter was washed in 1 x SSC, 0.1% SDS at 65°C which represents approximately 91% homology. The results, which were obscured by a significant background on the autoradiograph, are summarised in Table 25. The eight reasonably certain positives, i.e. 2 medium and 6 weak which were not positives for other probes, represent ∿1.2% of the library.

Table 25 Results of probing the cDNA library with cDNA enriched in sequences coding for the pea seed major albumins

Colony pLG number	albumins probe intensity	mRNA probe intensity	Other probes
4.4	weak	zero	-
4.7	weak	very weak	-
4.9	weak	zero	-
(4.11	dubious	weak	legumin)
4.18	weak	zero	-
(4.25	dubious	weak	-)
(4.28	dubious	very weak	legumin)
(4.30	dubious	weak	possible vicilin)
4.44	weak	zero	possible legumin
4.45	weak	zero	possible vicilin
(4.47	dubious	very weak	-)
4.52	weak	weak	-
4.61	weak	zero	legumin
(4.74	dubious	weak	vicilin)
(4.83	dubious	zero	-)
4.91	weak	weak	possible vicilin
4.92	medium	weak	-
(4.96	dubious	strong	vicilin)
(4.104	dubious	very weak	-)
(4.111	dubious	weak	possible big legumin)
4.115	weak	very weak	-
(4.127	dubious	weak	legumin)
(4.131	dubious	strong	legumin)
(4.132	dubious	medium	-)
4.135	medium	medium	-
(4.137	dubious	medium	possible vicilin)

2 medium and 6 weak positive colonies which were not positives for other probes.

3.3. Investigation of Colonies of Interest

3.3.1. Some Randomly Selected Colonies

A random selection of 11 white and 1 blue colonies were minipreped. Their inserts were sized by restriction with EcoRI to cut out the insert and agarose gel electrophoresis and by restriction with BamHI to linearise the molecule and agarose gel electrophoresis. The second gel was blotted as in the Southern Technique (Southern blotted) and the filter probed with radio labelled mRNA and washed in 1 x SSC, 0.1% SDS at 68°C, which represents approximately 94% homology. The results are summarised in Table 26. The average insert size was ∿590bp.

3.3.2. Some mRNA Positive Colonies

A selection of mRNA positive colonies were minipreped together with one blue colony. Their inserts were sized by restriction with EcoRI which cuts out the inserts, and agarose gel electrophoresis. The results are summarised in Table 27.

3.3.3. Some Main Legumin Positive Colonies

Of the colonies minipreped three had been shown to be main legumin positives in colony hybridization screens with the insert of plasmid pAD4,4 (pDUB6) as a probe. Two were restriction-mapped and their maps compared with that of the legumin A gene and the insert of pAD4,4 in Fig.9.

The colony hybridisation screen with the genomic clones of legumin C and legumin A found three colonies which were positive for these genomic clones but not for the insert of pAD4,4 (pDUB6), and were suspected to represent

Table 26 Insert sizes and results of mRNA probing a Southern blot at ∿94% homology of some randomly selected colonies

Colony pLG number	Insert size base pairs	mRNA colony hybridisation probe	mRNA probe of Southern blot
1.2	300	very weak	weak
1.6	540	zero	zero
1.7	450	zero	zero
1.11	1180	medium	strong
1.13	690	very weak	very weak
1.14	450	very weak	weak
1.16	960	strong	strong
1.17	440	zero	zero
1.18	600	weak	-
1.19	580	zero	zero
1.20	330	weak	medium
1.4 ‡	0	zero	zero

‡ A blue i.e. insertless colony

Table 27 Insert sizes of a selection of mRNA positive colonies

Colony pLG number	Insert size base pairs	mRNA colony hybridisation probe
1.9	1050	weak
1.11	1180	medium
1.16	960	strong
1.18	600	weak
1.20	330	weak
1.40	470	medium
1.46	600	medium
1.60	840	medium
1.63	1530	strong
1.72	710	medium
1.87	1090	medium
1.4 ‡	0	zero

‡ A blue i.e. insertless colony

Table 28 Insert sizes and some colony screen results of possible cDNAs for the 5'-end of main legumin

Colony pLG number	Insert size base pairs	Colony screen results pAD4,4(pDUB6)	Legumin C	Legumin A	mRNA
2.116	390	zero	weak	medium	medium
4.11	230	zero	weak	strong	weak

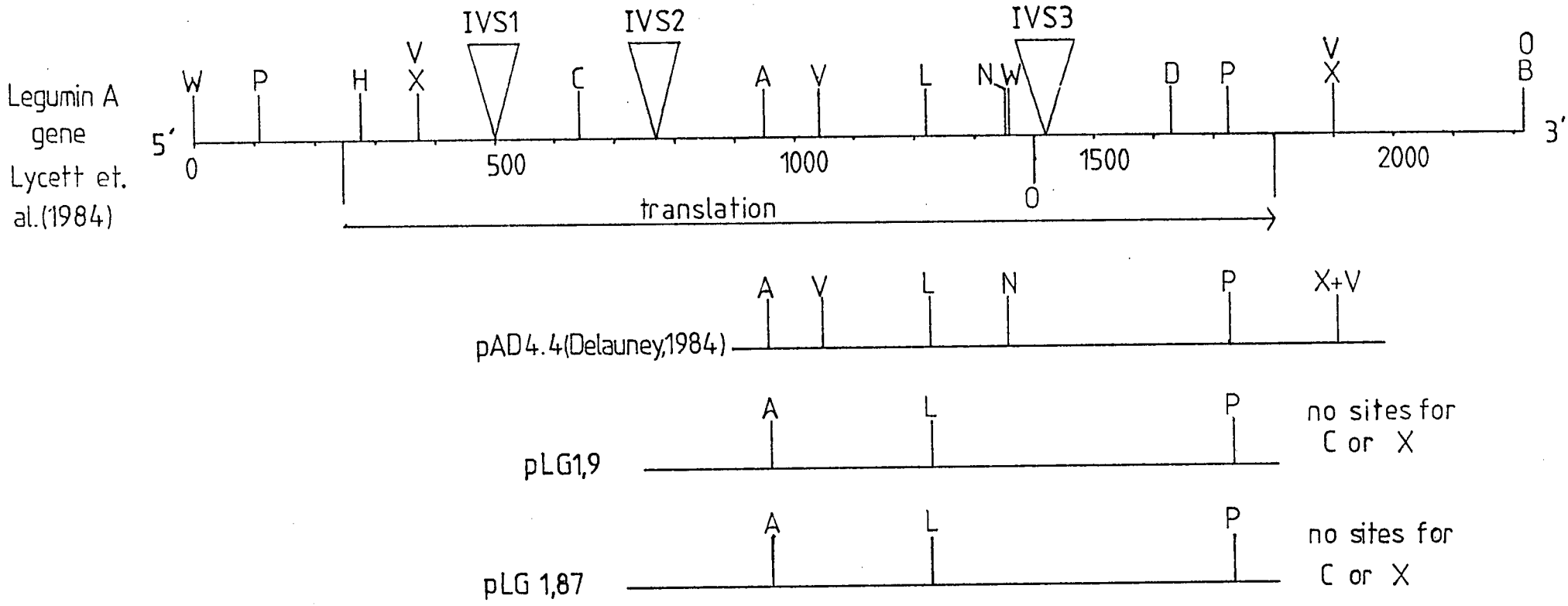


Fig.9. Restriction Maps of the Two pAD4.4(pDUB6) Legumin Positives Compared with those of pAD4.4 and Legumin Gene A

Symbols for restriction enzyme cleavage sites are:

A = AccI, B = BamHI, C = HincII = HindII, D = NdeI, H = Hind III, L = BglI, N = Bst NI, O = XhoII, P = PstI, V = AvaI, W = AvaII, X = XhoI.
Scale in base pairs. Intervening sequences (IVS) 1 and 2 are 87bp long and IVS3 is 98bp long.

parts of the message not present in pAD4,4 (pDUB6), i.e. the 5'-end of the message. Two of them were minipreped and their inserts sized by restriction with EcoRI and agarose gel electrophoresis. The results are summarised in Table 28. The insert sizes were consistent with the known length of pAD4,4(pDUB6) and hence with being cDNAs of the 5' end of the main legumin message.

3.3.4. Some Type B Vicilin Positive Colonies

Of the colonies minipreped, two had been shown to be type B vicilin positive in colony hybridisation screens using the insert of plasmid pAD3,4(pDUB7) as a probe. They were restriction-mapped and their maps compared, in Fig.10, with those of: 1) pAD3,4(pDUB7) and pRC2,2,10 (pDUB4), a combination of which represents all but 100bp of the type B vicilin message; 2) pAD2,1(pDUB9), which represents nearly a complete type C vicilin message and 3) pRC2,2,1(pDUB2) which represents ∿900bp of the Type A vicilin message, with ∿600bp of the 5'-end of the message missing.

3.3.5. Plasmid pRC2,2,26 and some pRC2,2,26 Positive Colonies

3.3.5.1. Plasmid pRC2,2,26

Plasmid pRC2,2,26 (R.R.D. Croy, personal communication) consists of a short cDNA insert, cloned into the BamHI insert of plasmid pBR322, which in hybrid-release translation experiments selects a message coding for a protein of molecular weight ∿25,000, which is reasonably abundant. It was thought to code for one of the major albumins Mr 25,000 and 24,000 or the seed lectin which may have this Mr before post-translational cleavage. A full

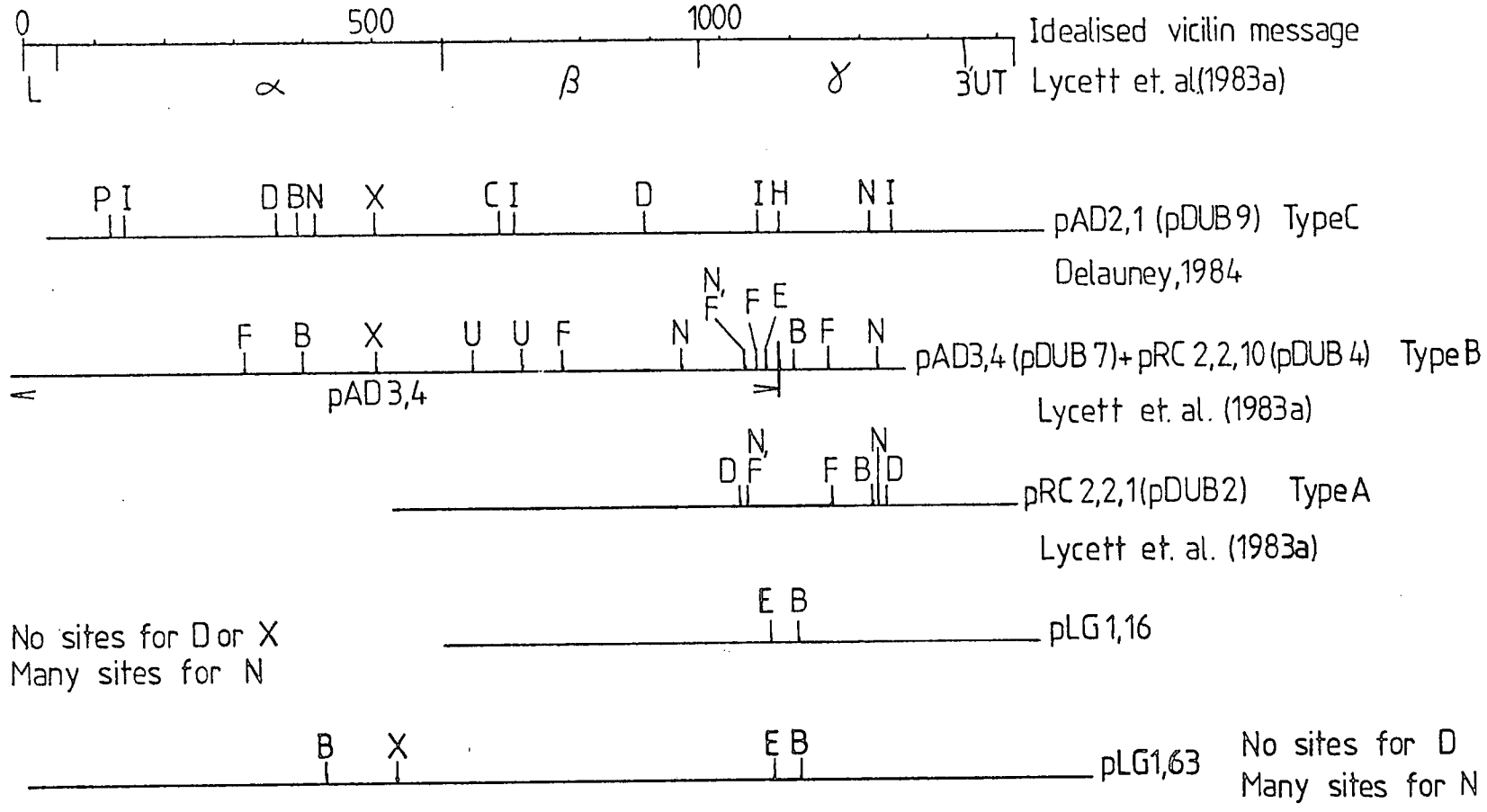


Fig.10. Restriction Maps of Two pAD3.4(pDUB7) Vicilin B Positives Compared with those of the Three Vicilin cDNA Probes

Symbols for restriction enzyme cleavage sites are:

B = BglII, C = AccI, D = HincII = HindII, E = BstEII, F = HphI, H = HindIII, I = HinfI, N = BstNI, P = PstI, U = Sau 96I, X = XbaI.

Scale in base pairs. L = Leader sequence, 3'UT = 3' untranslated region.

scale preparation of pRC2,2,26 was performed and the insert was extensively restriction-mapped (Fig.11), and partially sequenced (Fig.12) by the end-labelling dideoxy method.

3.3.5.2. Some pRC2,2,26 Positive Colonies

Minipreparations were made of a selection of those colonies which hybridized to the insert of plasmid pRC 2,2,26 in a colony hybridisation screen. Their inserts were sized by restriction with EcoRI and agarose gel electrophoresis, (Table 29) and extensive restriction-mapping carried out on two clones with large inserts : pLG2,73 (insert size 800bp, and pRC2,2,26 and mRNA weak) and pLG 4,10 (insert size 860bp, pRC2,2,26 strong and mRNA weak) (Fig.13). A Southern blot of the insert sizing gel was probed with the insert of plasmid pPVL134 a Phaseolus vulgaris lectin cDNA at 37°C (instead of 42°C) and washed in 3 x SSC, 0.1% SDS at 60°C which represents approximately 77% homology. A Southern blot of a gel of various restrictions of the two clones with large inserts (pLG2,73 and pLG4,10) and pRC2,2,26 was probed with the insert of pRC2,2,26 and washed in 1 x SSC, 0.1% SDS at 68°C, which represents approximately 94% homology. These results are summarised in Table 29. Plasmid pLG 2,73 and pLG 4,10 were subcloned into M13mp9 and sequenced by the dideoxy method (Fig.14)

3.3.6. Some Phaseolus vulgaris Seed Lectin cDNA Positive Colonies.

The two colonies pLG 2,47 and pLG 3,127, that may have been positives in the colony hybridisation screen with the insert of plasmid pPVL134, which is a cDNA of the message coding for Phaseolus vulgaris seed lectin,

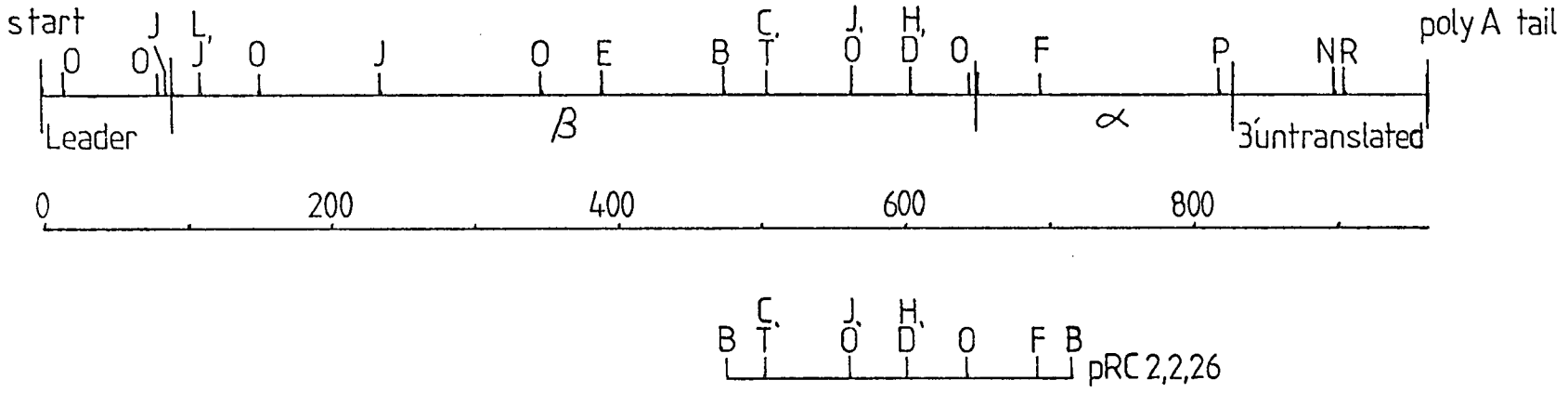


Fig.11. Restriction Map of the Insert of Plasmid pRC2,2,26 Compared with that of Pea Lectin Message

The map of pea lectin message was compiled from the published sequence (Higgins et.al., 1982). Plasmid pRC2,2,26 was a gift from R.R.D. Croy (see Sections 3.2.5 and 3.3.5).

Symbols for restriction enzyme cleavage sites are :

B = BamHI, C = ClaI, D = HincII = HindII, E = EcoRV, F = FokI, H = HpaI, J = HphI, L = BclI, N = NdeI, O = MboII, P = PstI, R = RsaI, T = TaqI.

Scale in base pairs.

Fig.12. The DNA Sequence of the Insert of Plasmid pRC2,2,26 Compared with that of Pea Lectin Message

The pea lectin sequence (Higgins et.al., 1982) runs from the start codon to the poly A tail, and is the upper sequence.

Table 29 Insert sizes and results of various hybridisation screens of some plasmid pRC 2,2,26 positive colonies

Colony pLG number	pRC 2,2,26 colony screen	Insert size base pairs	mRNA colony screen	pRC 2,2,26* Southern blot	pPVL134** Southern blot
1.139	strong	420	weak	-	zero
2.56	weak	210	zero	-	zero
2.73	weak	800	weak	zero	zero
2.107	medium	210	very weak	-	very weak
3.85	medium	290	zero	-	very weak
4.10	strong	850	weak	strong	weak
4.14	strong	210	weak	-	zero
4.71	strong	290	weak	-	very weak
(4.88	dubious	300	very weak	-	zero)
plasmid pRC2,2,26	-	240	-	strong	weak

* Washed in 1 x SSC, 0.1% SDS at 68°C representing ∿94% homology

** Washed in 3 x SSC, 0.1% SDS at 60°C representing ∿77% homology

\- Not done.

In each case the probe used was the cDNA insert of the given plasmid.

Table 30 Insert sizes and the results of various screenings for the pPVL134 positive colonies

	pLG 2.47	pLG 3.127
pPLV134 colony screen	weak	weak
mRNA colony screen	weak	medium
Insert size base pairs	300 and 500	550
pPVL134 Southern blot*	very weak	very weak
pRC2,2,26 Southern blot **	zero	-

* Washed to ∿77% homology. ** washed to ∿94% homology.

blot hybridisation intensities were relative to the size markers.

\- Not done.

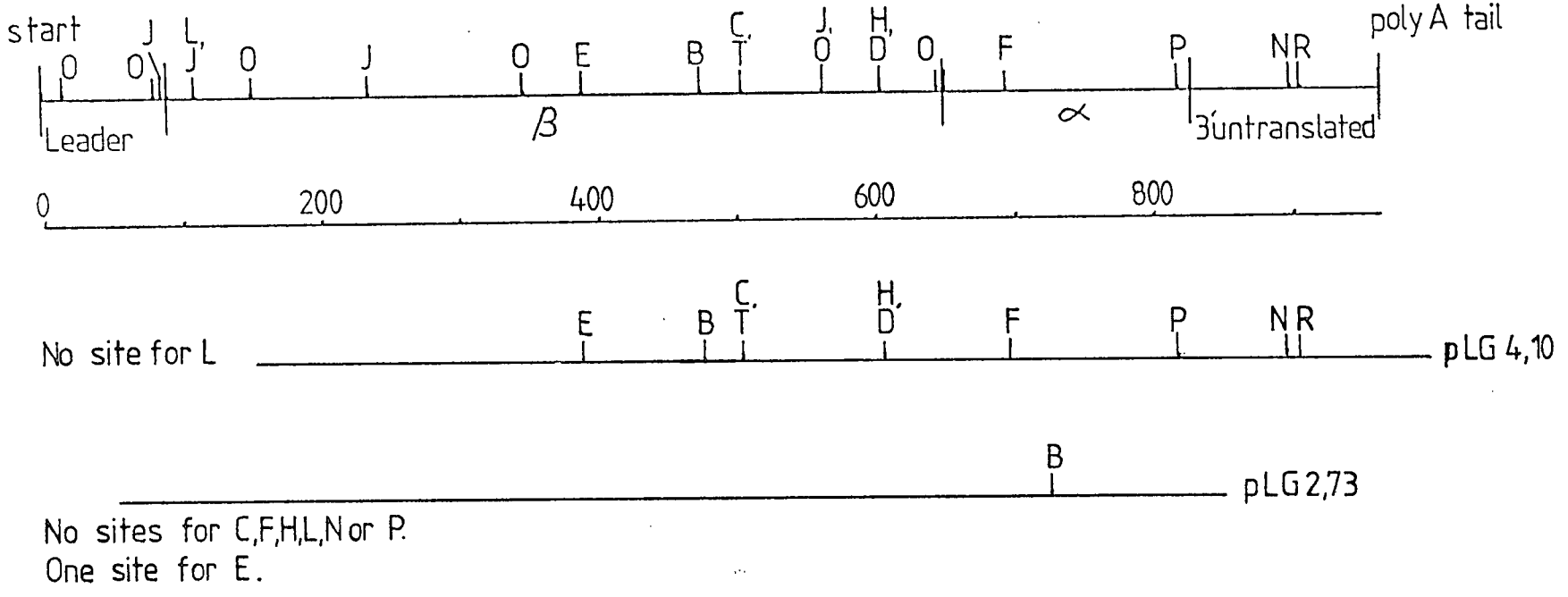


Fig.13. Restriction Maps of Two Plasmid pRC2,2,26 Positives Compared with that of Pea Lectin Message

The map of pea lectin message was compiled from the published sequence (Higgins et.al., 1982).

Symbols for restriction enzyme cleavage sites are as for Fig.11.

Scale in base pairs.

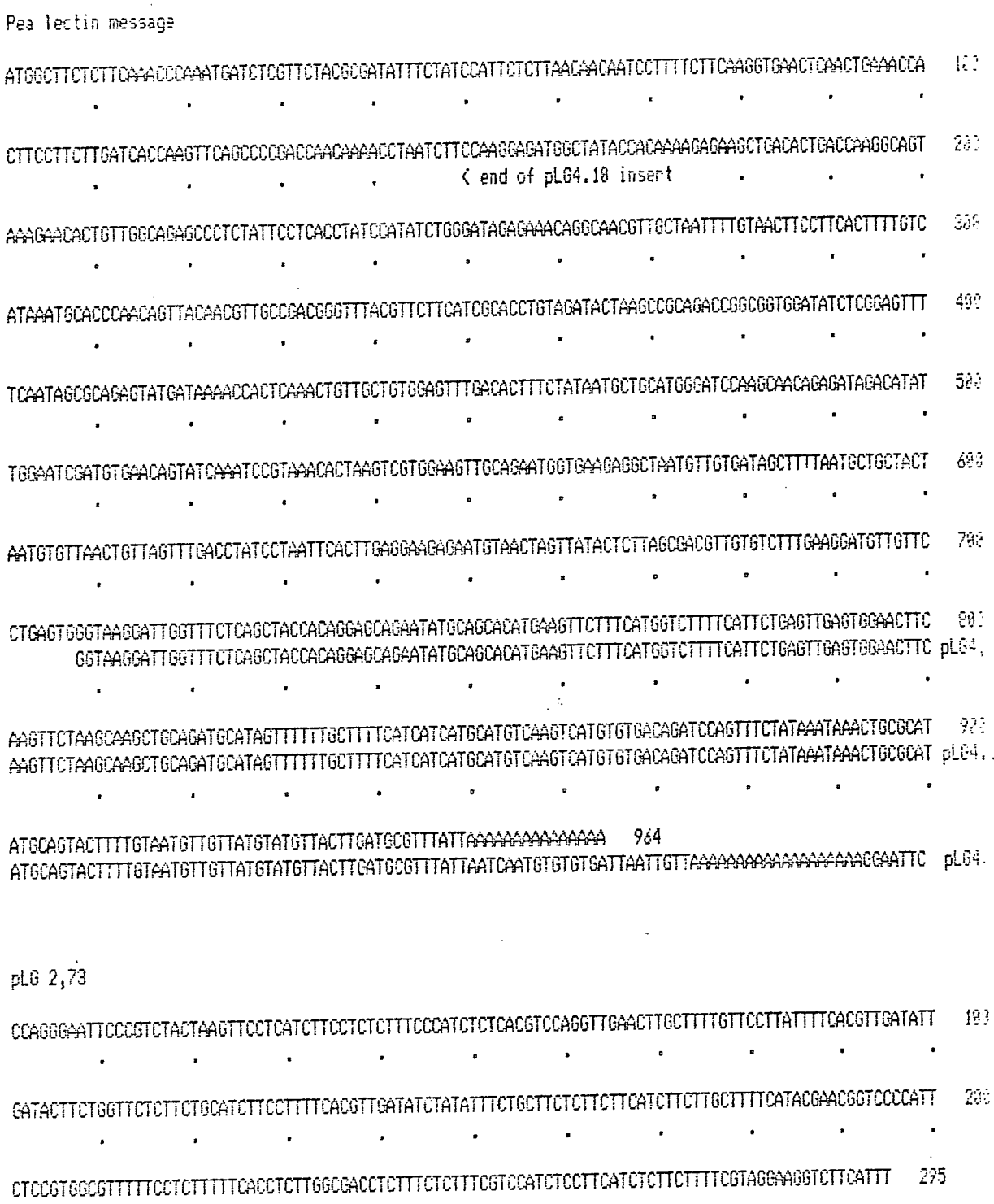


Fig.14. DNA Sequences of Two Plasmid pRC2,2,26 Positives Compared with that of Pea Lectin Message

The sequence of pea lectin message (Higgins et.al., 1982) runs from the start codon to the poly A tail, and is the upper sequence.

The insert of pLG 2,73 runs on to ~800

were minipreped and the inserts sized by restriction and agarose gel electrophoresis. A Southern blot of these plasmids was probed with the insert of plasmid pPVL134 and washed in 3 x SSC, 0.1% SDS at 60°C which represents approximately 77% homology, and one of pLG2,47 was probed with the insert of pRC2,2,26 (See section 3.3.5) and washed in 1 x SSC, 0.1% SDS at 68°C, which represents approximately 94% homology. These results are summarised in Table 30. Some rough restriction-mapping was carried out : Plasmid pLG2,47 has in its inserts: 1 HpaI site, 2 EcoRV sites approximately 270bp apart and 1 PstI site approximately 470bp from the PstI site in the multipurpose cloning site of pUC8. It has no sites for BamHI, ClaI or PvuII in its insert. Restriction with PvuII gave only two fragments and as PvuII cuts pUC8 twice either side of the cloning site this confirms that pLG2,47 has two inserts.

Plasmid pLG3,127 has no HpaI site. Thus neither inserts restriction map fits that of pea lectin.

3.3.7. Some Major Albumins Positive Colonies

Minipreparations were made of a selection of those colonies which hybridised to the cDNA enriched with sequences coding for the major albumins (Section 3.2.7.). Their inserts were sized by restriction with endonucleases and agarose gel electrophoresis (Table 31). No strong correlation was seen between insert size and strength of hybridisation signal with the major albumins probe.

R.A. Ragab (1985) investigated some of these further and showed by hybrid-release translation that the larger of the inserts of pLG 4,135 selects a message for a protein of Mr ~25,000, the same size as the pea seed large major

Table 31 Insert sizes and the results of various screenings for the major albumins positive colonies

Colony pLG number	Major albumins colony screen	Insert size base pairs	mRNA colony screen	Other colony screens
4.7	weak	410 and 230	very weak	weak legumin
4.9	weak	870	zero	-
4.44	weak	580	zero	-
4.45	weak	720	zero	possible vicilin
4.52	weak	580	weak	-
4.61	weak	640	medium	legumin positive
4.92	medium	1070	weak	-
4.115	weak	<200 ±	very weak	-
4.135	medium	840 and 330	medium	-

± i.e. no insert visible on the gel

albumin. This protein was further shown to react with antibodies raised against the pea seed major albumins. The large insert of plasmid pLG 4,135 did not cross hybridise with the other possible albumin positives, but it was used to select two colonies from another cDNA library.

The large insert of pLG 4,135 was restriction-mapped (Fig.15) and DNA sequenced (Fig.16). This restriction-map was found to be similar to but not identical to the maps of the two albumin positive colonies it has selected from the other cDNA library.

3.3.8 Plasmid pJC5-2, pJC5-2 Positive Colonies, and (big) Legumin Genes J and K

3.3.8.1. Plasmid pJC5-2 and (big) Legumin Genes J and K

Plasmid pJC5-2 is a subclone of the 1.9kb EcoRI fragment of genomic clone λJC5 (*N. Ellis*, personal communication) in the EcoRI site of plasmid pUC9. Plasmid pUC9 is essentially identical to plasmid pUC8 but with the multipurpose cloning site in the opposite orientation. This 1.9kb fragment was shown to hybridise to a cDNA (pCD40; Domoney and Casey, 1984) of a message coding for a Mr ∿65,000 legumin-related protein, i.e. big legumin, and the gene, of which this fragment was part of, was called legumin J. A second EcoRI fragment of 3.5 kb, which did not border the 1.9kb fragment also hybridized with the big legumin cDNA probe and was similarly subcloned in pUC9 as plasmid pJC5-11. This gene was named legumin K (Fig.17).

The insert of pJC5-2 (legumin J) was extensively restriction-mapped. The insert of pJC5-11 (legumin K) and part of the legumin J gene not on pJC5-2 were also mapped

Fig.15 Restriction Map of the Large Insert of Plasmid pLG4,135 a Major Albumins Positive.

This map was compiled in conjuction with the DNA sequence. Symbols for restriction enzyme cleavage sites are:

A = AluI, C = BclI, F = FokI, H = HindIII, J = HphI, L = BalI, N = BstNI,
O = MboII, R = RsaI, S = SphI, T = TaqI, V = AvaII.

Scale in base pairs.

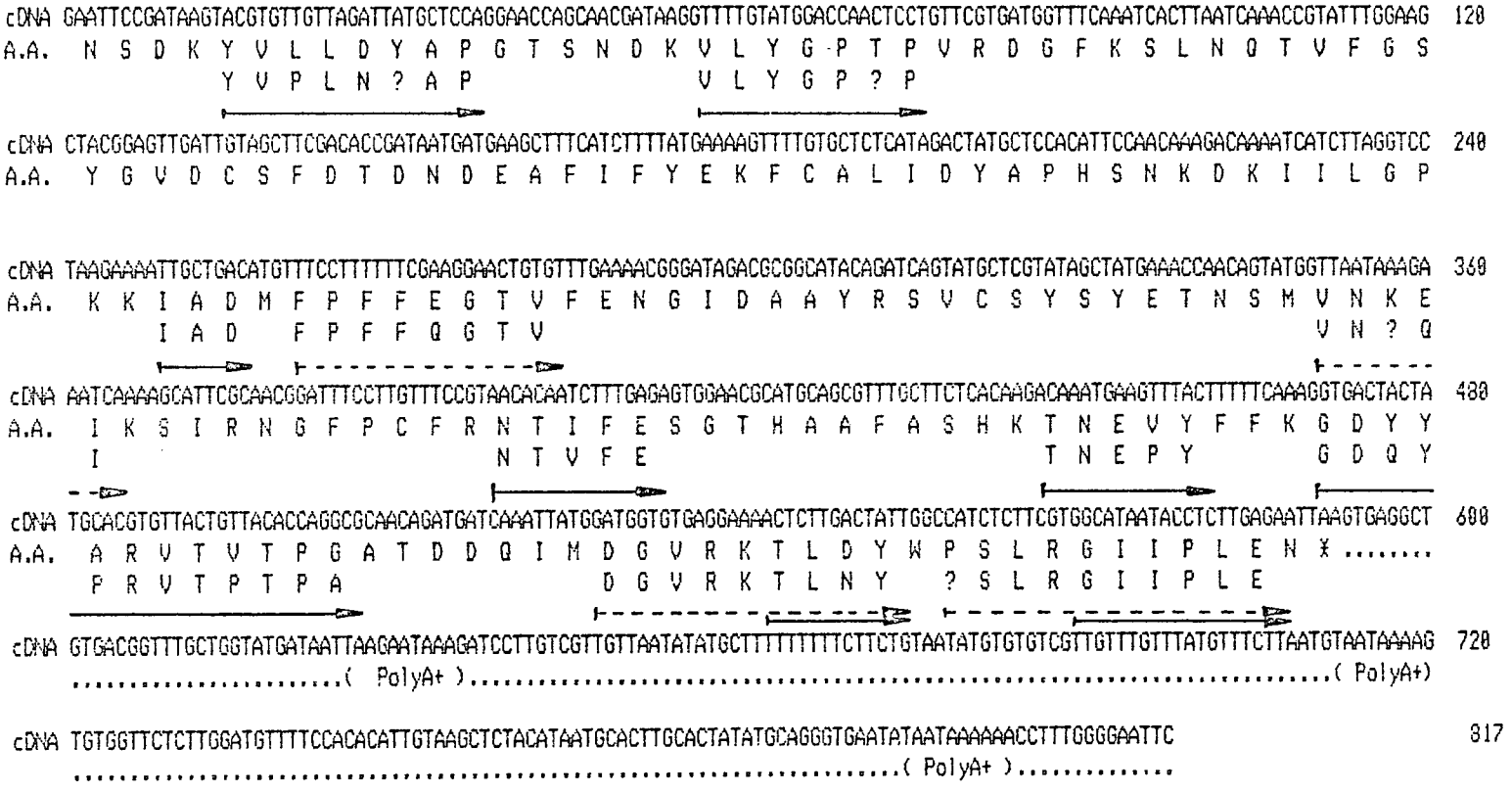


Fig.16. The DNA and Amino Acid Sequences of the Large Insert of pLG4,135 Compared with some Partial Amino Acid Sequence Data from Pea Major Albumins

Solid lines are tryptic peptides, dashed lines are cyanogen bromide peptides. The protein sequence was a personal communication from J. Gatehouse and J. Gilroy.

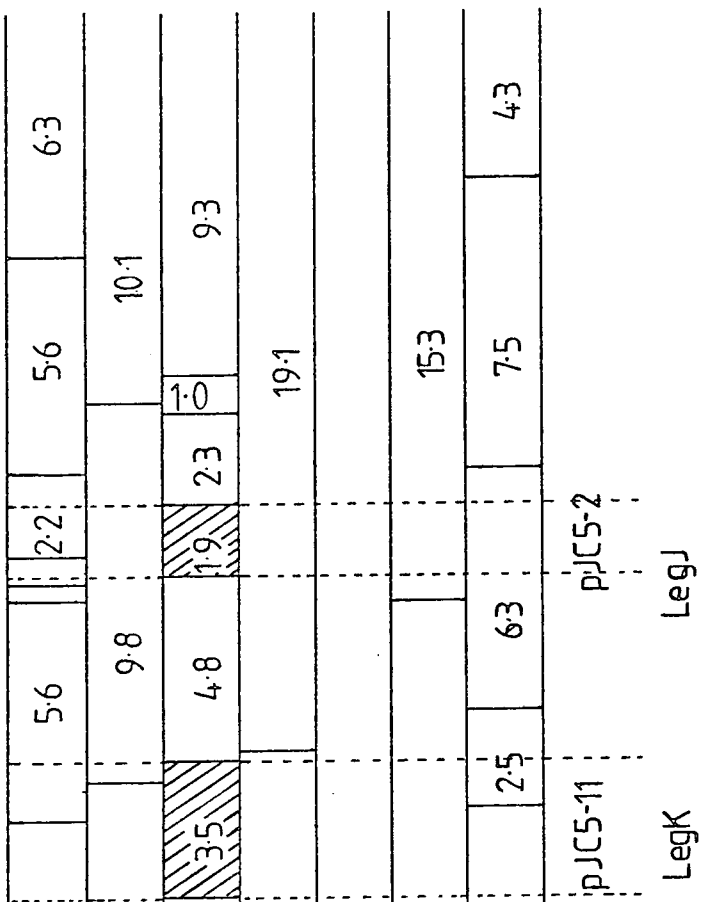


Fig.17. Restriction Map of the Genomic Clone λJC5

N. Ellis (personal communication). The shaded areas represent regions of homology with big legumin cDNA pCD40 (Domoney and Casey 1983) and were subcloned as marked. The approximate position of legumin genes J and K are indicated.

in less detail by myself, D. Bown, I.M. Evans and M. Levasseur (Fig.18). The two maps are similar, but not identical.

3.3.8.2. Plasmid pLG 3,121 a pJC5-2 Big Legumin Positive Colony

A miniprep of plasmid pLG 3,121, which hybridised strongly to the insert of plasmid pJC5-2 and to mRNA in colony hybridisation screens, was prepared. Its insert was sized and restriction-mapped, its map was found to be essentially similar to that of legumin J, taking into account the intervening sequences and different to those of legumin K and pCD40, which were similar to each other (Fig.18).

3.4. Cross Hybridisation of Main and Big Legumin DNAs

A Southern blot of an agarose gel of EcoRI digests of some main legumin positive colonies, i.e. the isolated inserts of these positive colonies, was probed with the legumin J gene (a big legumin gene) probe, the insert of plasmid pJC5-2 (Section 3.2.3). The filter was washed in 1 x SSC, 0.1% at 65°C, which represents approximately 91% homology. No cross hybridisation could be seen. (Fig.19).

3.5. Sizing the Message Corresponding to the cDNA Inserts of Plasmids

Northern blots of agarose gels of the relevant mRNA can be probed with a cDNA and from the hybridisation pattern the size of the mRNA from which the cDNA was made can be estimated from various standard size markers.

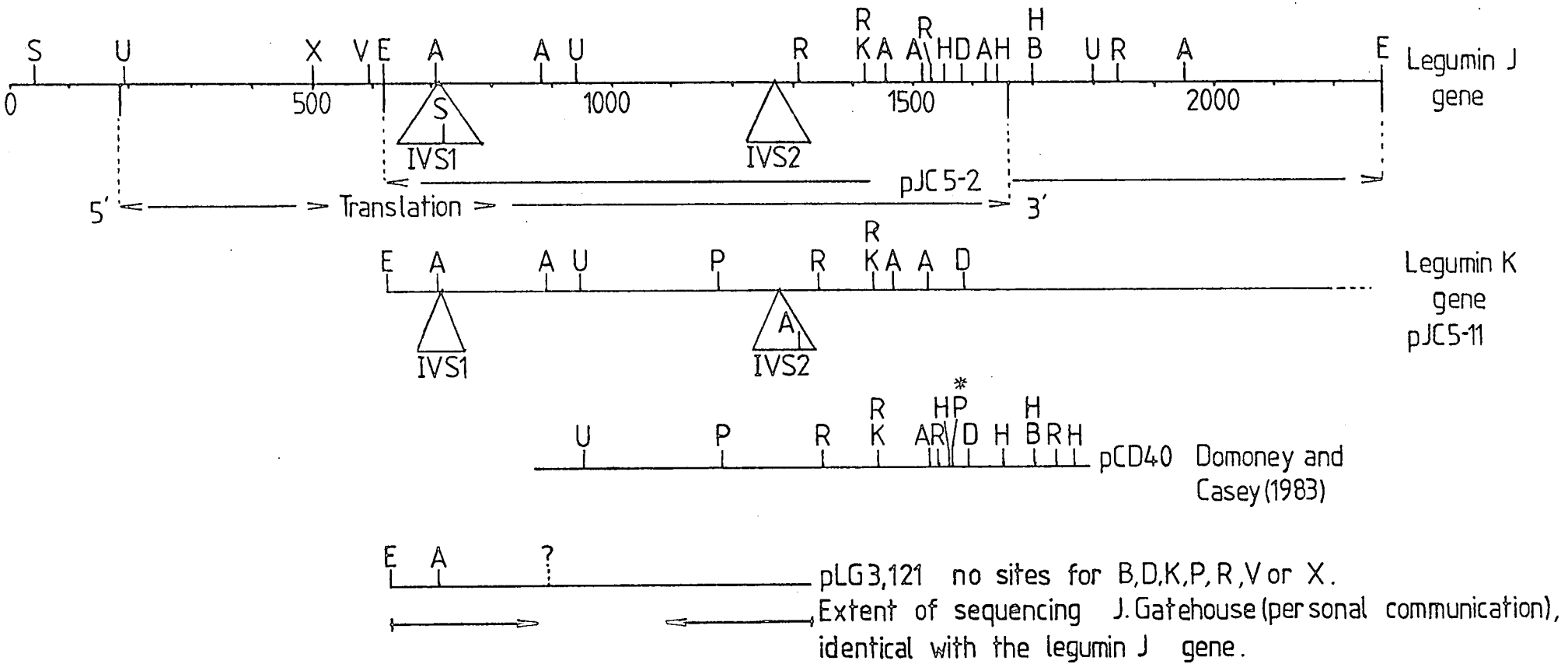


Fig.18. Restriction Map of a Plasmid pJC5-2 Big Legumin Positive Compared with those of Legumin Genes J and K and Plasmid pCD40

Plasmid pCD40 is a big legumin cDNA. Intervening sequence (IVS) is 138 base pairs long in legumin J and 81 in K, IVS2 is 98 base pairs long in legumin J and 105 in K.

Symbols for restriction enzyme cleavage sites are:

A = AluI, B = BalI, D = HincII = HindII, E = EcoRI, H = HaeIII, K = KpnI, P = PstI, R = RsaI, S = SphI, U = Sau3A, V = EcoRV, X = XbaI. * See Page 166

Scale in base pairs.

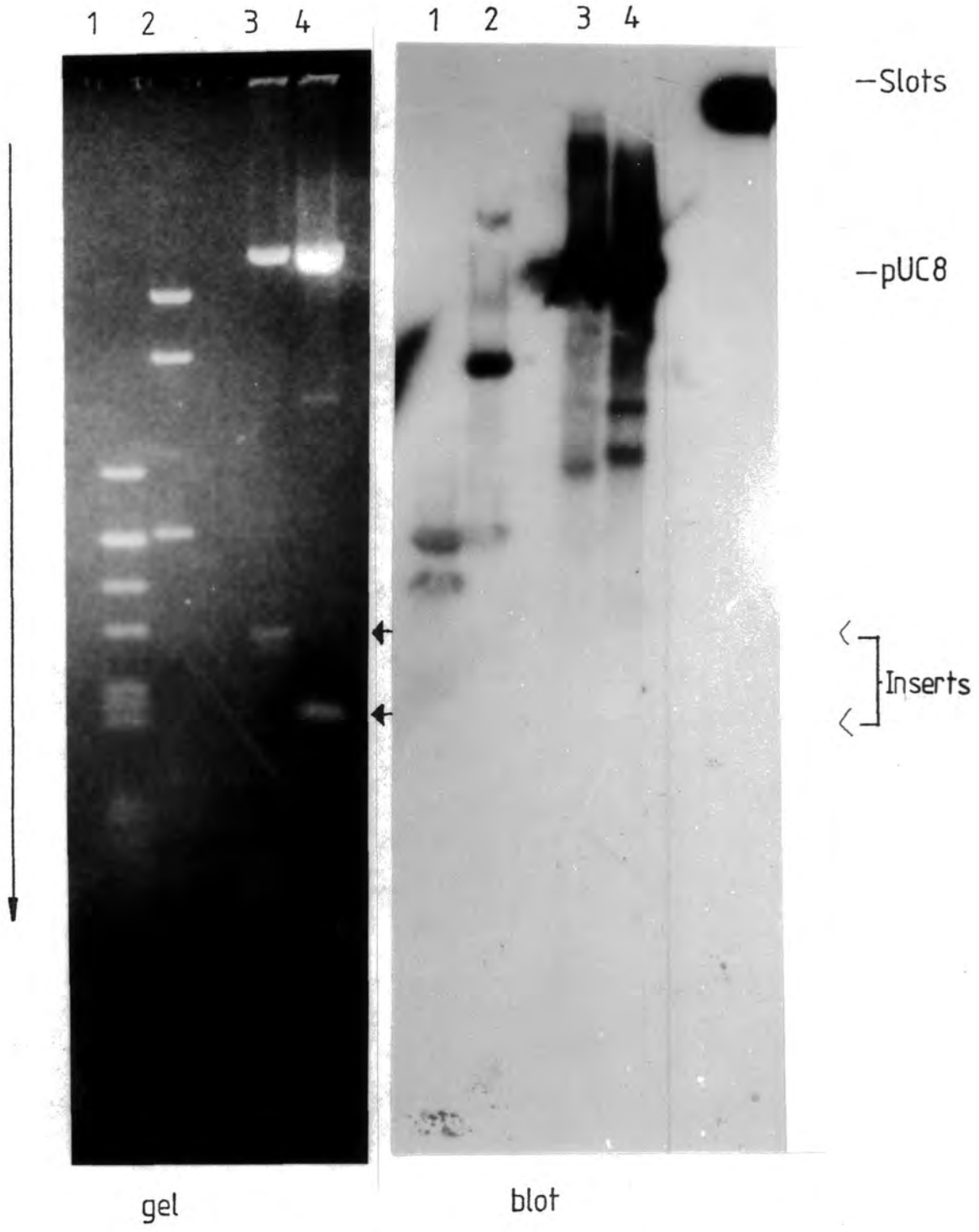


Fig.19. Attempted hybridisation of big to main legumin at ∿91% homology.

It was attempted to hybridise a big legumin probe (the insert of plasmid pJC5-2) to the inserts of two main legumin positives which had been Southern blotted. The washing conditions correspond to approximately 91% homology.

1) Size marker pBR322 AluI
2) Size marker pBR322 RsaI
3) & 4) Main legumin positives

Inserts i.e. main legumin cDNAs.

3.5.1. Estimating the Size of the mRNA Corresponding to the Insert of Plasmid pLG 4,92

Plasmid pLG 4,92 was a plausible major albumin positive (Sections 3.3.7 and 3.2.7), a sample of which was nick-translated and used to probe a Northern blot of an agarose gel of pea cotyledon polysomal RNA from various stages of development (Fig.20). The size markers were run on the same gel, but instead of being blotted were ethidium bromide stained and photographed in the conventional manner. The filter was washed in 0.1 x SSC x 0.1% SDS at 50°C, which represents approximately 84% homology. The results are summarised in Table 32.

3.5.2. Estimating the Size of the mRNA Corresponding to the Insert of Plasmid pRC 2,2,26 (lectin)

A sample of plasmid pRC 2,2,26 (Sections 3.2.5 and 3.3.5) was nick-translated and used to probe a Northern blot of an agarose gel of pea cotyledon RNA from various stages of development, and isolated by different methods. (Fig.21). The size markers were run on the same gel, but instead of being blotted were ethidium bromide stained and photographed in the conventional manner. The filter was washed in 0.1 x SSC, 0.1% SDS at 50°C, which represents approximately 84% homology. The results are summarised in Table 33.

3.6. Ligation of DNA

In a series of experiments test ligations were carried out on plasmid pBR322 restricted with endonuclease PvuII, which produces blunt ends. It was found that the conditions suggested by Maniatis et.al., (1982) gave good

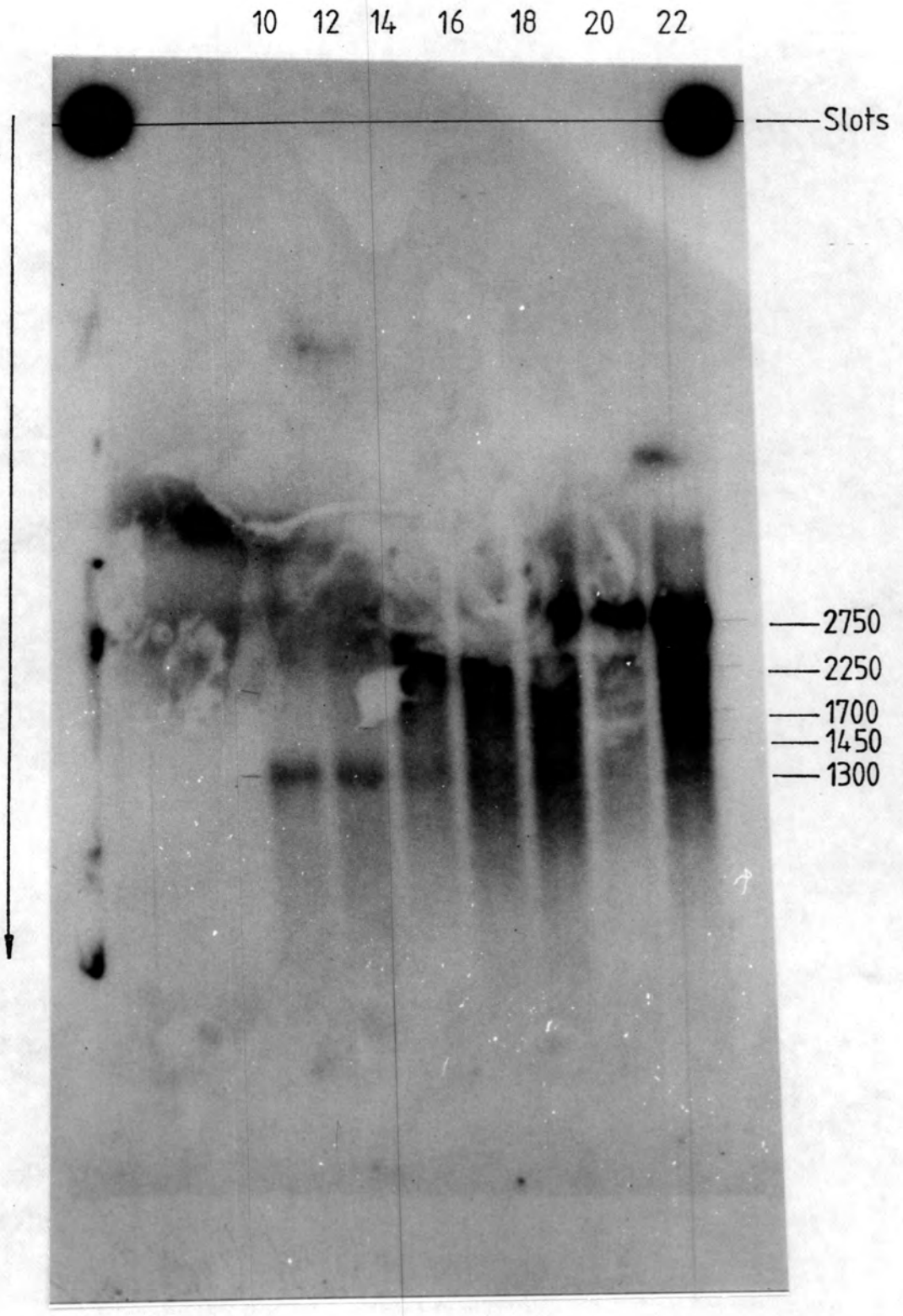


Fig.20. Sizing the mRNA for pLG4,92 at ∿84% homology.

Tracks are pea cotyledon polysomal RNA isolated the given number of days after flowering. Band sizes are in bases.

Table 32 Results of probing a Northern blot of pea cotyledon RNA isolated at various stages of development, with plasmid pLG 4.92 at ∿84% homology

Stage of development when RNA isolated (days after flowering)	Intensity of hybridisation with bands of		
	∿1300 bases	∿2250 bases	∿2750 bases
10	medium	zero	zero
12	medium	zero	zero
14	very weak	medium	zero
16	zero	medium	-
18	zero	medium	strong
20	zero	medium	strong
22‡	dubious	medium	strong

‡ Also ∿1700 weak and ∿1450 very weak

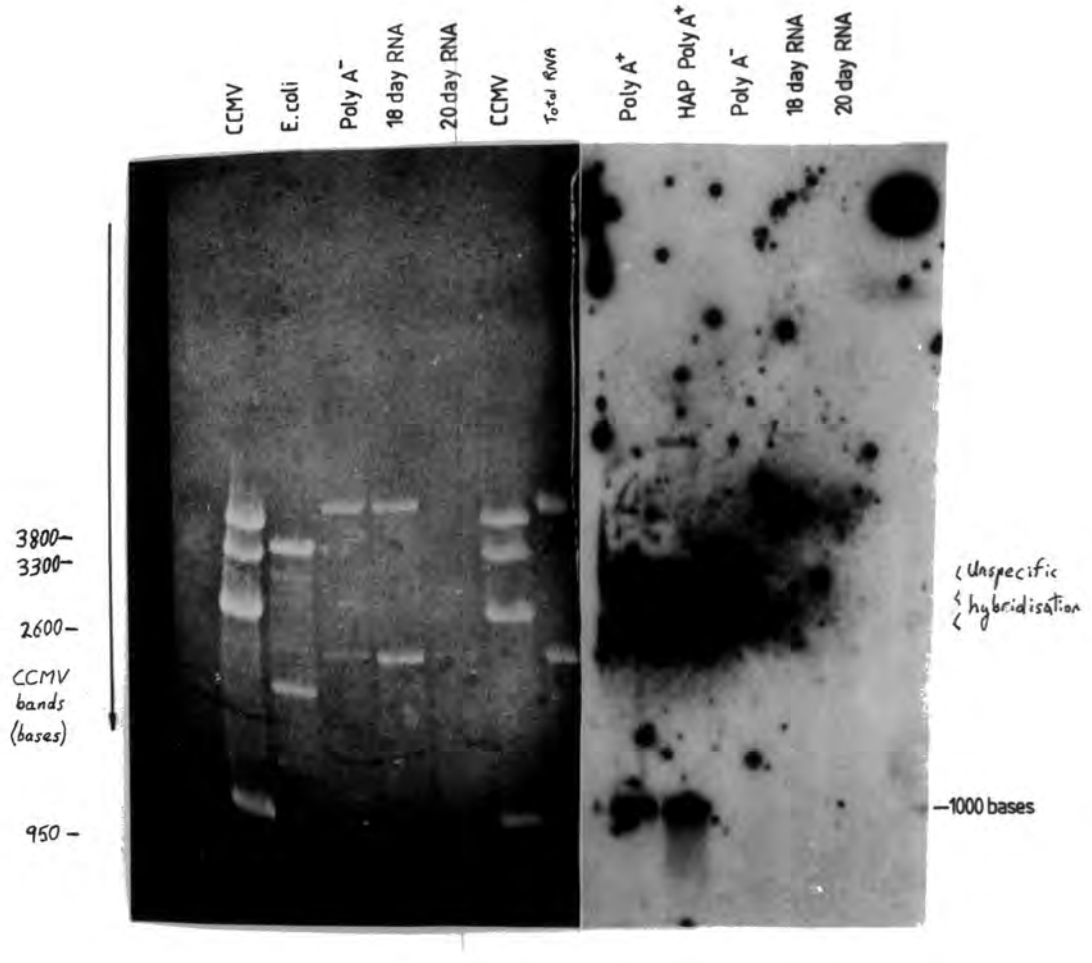


Fig.21. Sizing the mRNA for pea soluble seed Lectin.

CCMV = Cowpea chlorotic mottle virus RNA. E.coli = E.coli RNA. Total RNA = RNA from pea cotyledons isolated 13-17 days after flowering. Poly A^- = Poly A^- RNA from pea cotyledons (Hall et. al., 1978). 18/20 day RNA = pea cotyledon RNA isolated 18/20 days after flowering. Poly A^+/ HAP Poly A^+ = Poly A^+ RNA from pea cotyledons ∿14 days after flowering isolated by differing methods (Hall et.al., 1978).

Table 33 Results of probing a Northern blot of various pea cotyledon RNAs with plasmid pRC 2,2,26 at ∿84% homology

Pea cotyledon RNA	Size of band (bases)	Intensity of Hybridisation
poly(A)$^+$ RNA mid-development	∿ 1000	medium
HAP poly(A)$^+$ RNA mid-development ‡	∿ 1000	medium
poly(A)$^-$ RNA mid-development	-	zero
18 day RNA	-	zero
20 day RNA	-	zero

‡ HAP poly(A)$^+$ RNA is prepared by the alternate method of Hall et.al.,(1978)

Poly(A)- RNA is the RNA which is not retained by an oligo(dT) cellulose column and hence carries no poly(A) tail. It mainly consists of ribosomal RNA, transfer RNA and degraded messenger RNA.

results and were adopted as the method of choice.

3.7. Terminal Deoxynucleotidyl Transferase Tailing Reactions

Calf thymus terminal deoxynucleotidyl transferase was used to add homopolymer tails to the 3'-ends of single- and double-stranded DNA.

3.7.1. Terminal Transferase Commercial Assay.

Terminal transferase was assayed according to the protocol supplied by BRL (Bethesda Research Laboratores) on their data sheet. The substrate was a sample of phage λ NM258 restricted with endonuclease HaeIII and purified by multiple phenol extractions and ethanol precipitation using HPLC-purity grade water throughout.

Following the protocol for tailing with dG residues, results essentially identical with those given in the data sheet could be obtained, but not reproducibly. The best assay gave tails of 197, 217 and 233 base after 30, 60 and 120 minutes compared with 118, 113 and 110 bases claimed in the data sheet for the same conditions. Using an identical protocol with dCTP instead of dGTP very rapid initial uptakes of radiolabel after 2 minutes were followed by apparent loss of the tail, i.e. essentially no uptake of radioactivity after 5-10 minutes and subsequently.

3.7.2. Terminal Transferase a Functional Assay

A functional assay of TdT was devised : Plasmid pBR322 was cut at its PstI site to leave exposed 3'-ends, the best substrate for TdT, and purified carefully by phenol

extractions and ethanol precipitation using HPLC water throughout.

Samples of PstI cut pBR322 were tailed with dC and dG (Section 3.7.4), and were purified by passing down the appropriate oligo(deoxnucteotidyl)-cellulose column (Section 3.8). The DNA retained by the column was identical to the original substrate when examined by agarose gel electrophoresis and in each case the overall yield was approximately 12%.

The results of the assay are contained in section 3.1.3 and in particular Table 9, which shows the results of annealing the tailed plasmids and transforming competent bacteria. The resulting transformation rate of dimers of $7 \times 10^3/\mu g$ of vector compares badly with that of Land et.al., (1981) of $8 \times 10^5/\mu g$ of vector. However the oligo(deoxynucleotidyl)-cellulose column result would suggest only 12% of the DNA had functional tails at both ends. Thus the transformation rate is $4 \times 10^5/\mu g$ vector for molecules with viable tails at both ends.

3.7.3. Tailing Single-stranded cDNA

sscDNA/mRNA hybrid duplexes and sscDNA from which RNA had been cleaved were tailed with both dC and dG residues under the same conditions found best for tailing plasmid DNA (Section 3.7.4). Typical results were the addition of approximately 1 base per 3'-end of DNA. Best results were 1.5 and 3 for dG and dC bases respectively. Despite the disappointingly low tailing efficiency, viable tails were produced as shown by the results in Section 3.1.2.

3.7.4. Tailing Plasmid and other Double Stranded DNA

In a series of experiments it was determined that the maximum incorporation of radiolabel into dsDNA substrates, using either dGTP or dCTP, occurred after approximately 2½ minutes in the presence of 60µg/µl of nuclease free bovine serum albumin (BSA).

3.8. Oligo(deoxynucleotidyl)-Cellulose Columns

In a series of experiments it was found to be possible to separate microgram quantities of tailed and untailed DNA using the appropriate oligo(deoxynucleotidyl)-cellulose. The retained "tail positive" fraction was eluted in a sufficiently small volume to be readily ethanol-precipitated, though 10µg tRNA was added as a matter of course. The progress of the column was monitored by the absorption of UV light, at 254 nm wavelength, of the eluate (Fig.22) and was sensitive enough to detect sub-microgram quantities of DNA.

3.9. A Restriction Map of Plasmid pUC8

Plasmid pUC8 (Viera and Messing, 1982) is a purpose-built cloning plasmid and was the cloning vector of choice, however no detailed restriction map of pUC8 was available. From the details of construction of plasmid pUC8 the DNA sequence was assembled and exhaustively searched for the presence of sites for most common restriction endonucleases as well as some less common ones. Figure 23, Figure 24, Table 34.

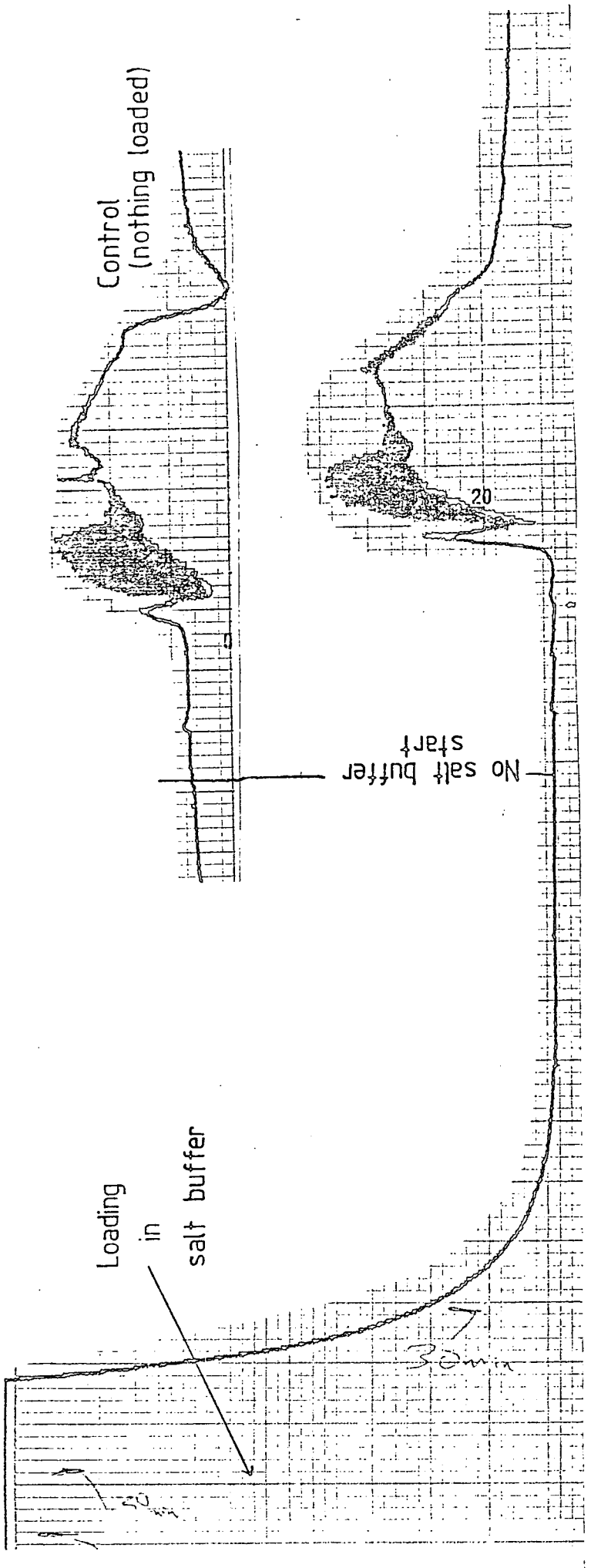


Fig.22. Detector Output of an oligo(deoxynucleotide)- Cellulose Column.

Note the larger peak for the sample compared to the control and the lack of a tail on the control peak.

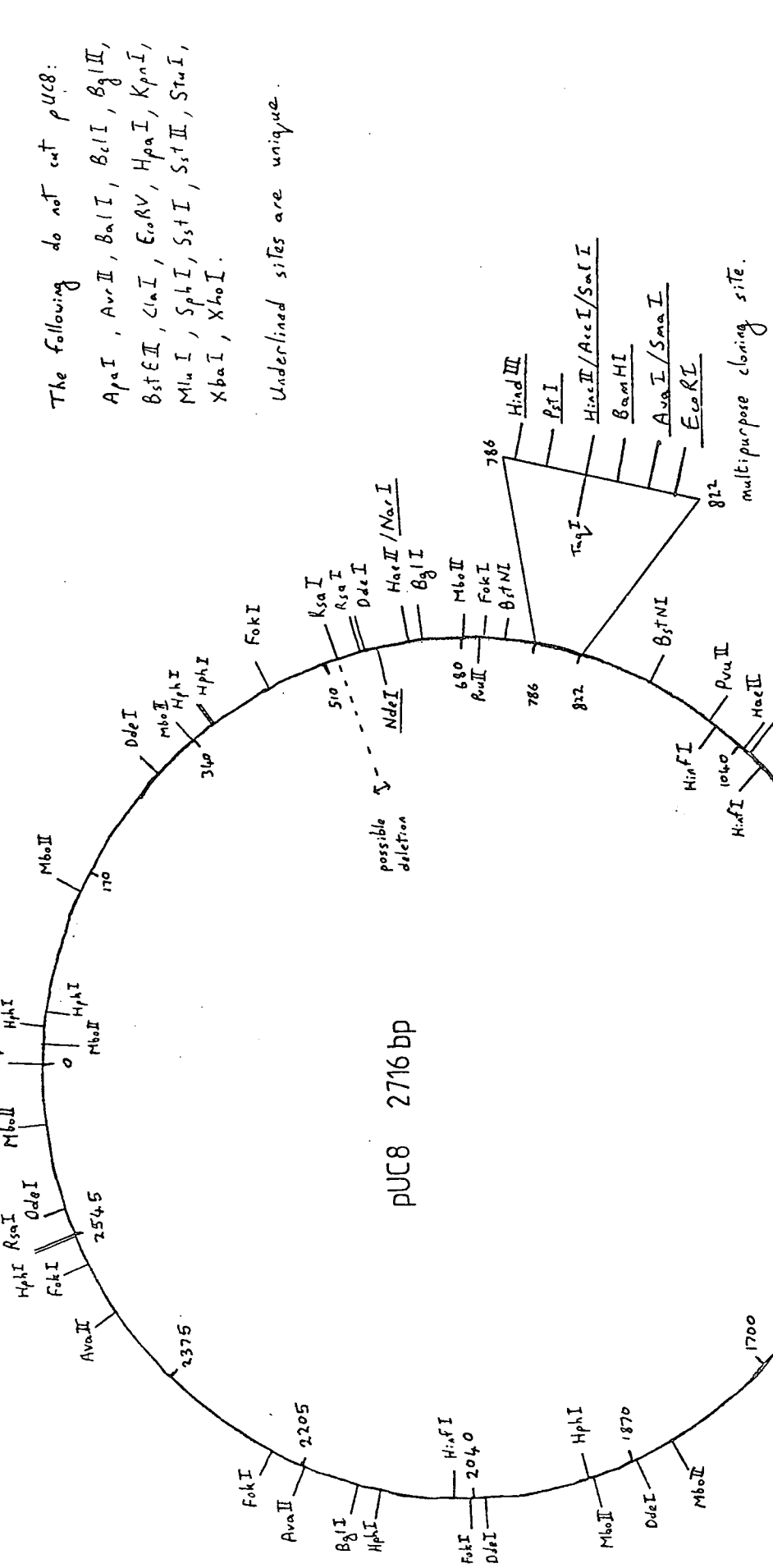


Fig.23. Restriction Map of Plasmid pUC8

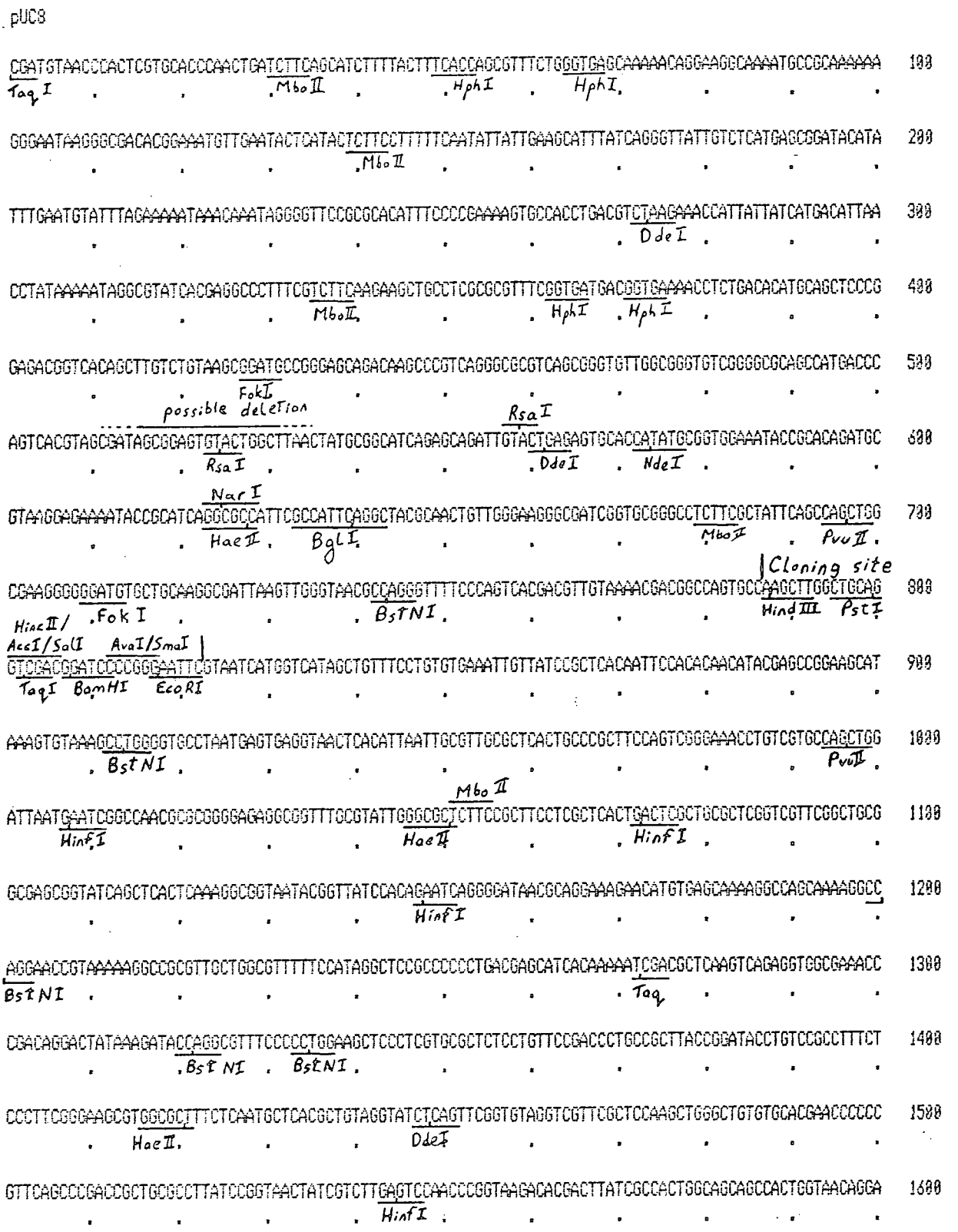


Fig.24. DNA Sequence and Restriction Endonuclease Recognition Sites of Plasmid pUC8

The DNA sequence of pUC8 was compiled from the details of its construction and tested by some restriction mapping. No exaustive restriction mapping or DNA sequencing was carried out.

pUC8 continued

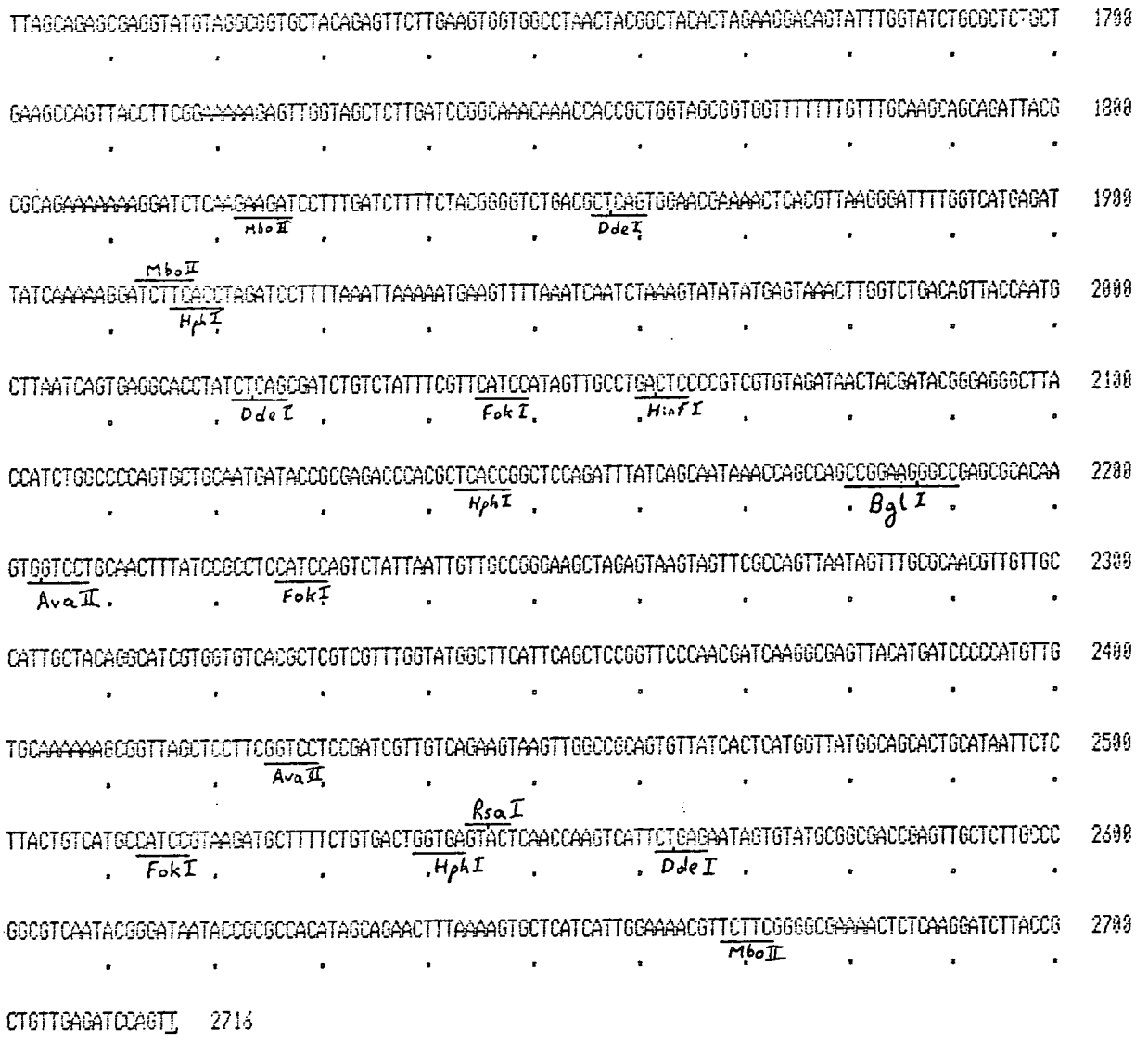


Figure 24 Continued

Table 34 A restriction map of plasmid pUC8

The multipurpose cloning site has unique restriction sites for:-

AccI, AvaI, BamHI, EcoRI, HincII, HindIII, PstI, SalI, SmaI

Restriction endonuclease	Number of sites ‡	Position(s)	Fragment sizes* (base pairs)
AccI	1	∿803	2716*
AluI	14X		
ApaI	0	-	-
ApyI	up to 4‡		
AvaI	1	814	2716*
AvaII	2	2203, 2425	2494*, 222
AvrII	0	-	-
BalI	0	-	-
BamHI	1	807	2716*
BclI	0	-	-
BglI	2	630, 2186	1160, 1556*
BglII	0	-	-
BstEII	0	-	-
BstNI	5	743, 913, 1200, 1321, 1334	2125, 170*, 287, 121 23
ClaI	0	-	-
DdeI	6		
EcoRI	1	817	2716*
EcoRII	5	see BstNI	
EcoRV	0	-	-
FokI	5	430, 710, 2050, 2230, 2515	613, 280, 1340*, 180, 285
HaeII	3	627, 1050, 1420	1923, 423*, 370
HaeIII	17X		
HhaI	17X		
HincII	1	∿803	2716*
HindIII	1	787	2716*
HinfI	5		
HpaI	0	-	-
HpaII	13X		
HphI	7		

Table 34 (Continued)

Restriction endonuclease	Number of sites ‡	Position(s)	Fragment sizes* (base pairs)
KpnI	0	-	-
MboI	15X		
MboII	8		
MluI	0	-	-
NarI	1	624	2716*
NciI	6X		
NdeI	1	573	2716*
PstI	1	799	2716*
PvuII	2	696, 996	2416, 300*
RsaI ‡‡	3 or 2	524?, 558, 2545	695, 34?, 1987*
SalI	1	∿803	2716*
Sau3A	15X		
Sau96I	6X		
SmaI	1	∿814	2716*
SphI	0	-	-
SstI	0	-	-
SstII	0	-	-
StuI	0	-	-
TaqI	3	1, 802, 1272	1445, 801*, 470*
XbaI	0	-	-
XhoI	0	-	-

‡ X indicates these sites are not marked on the wheel map.

* Fragments marked thus contain all or part of the multipurpose cloning site.

‡‡ ApyI only cuts specifically methylated DNA.

‡‡ The RsaI site at position 524 may have been deleted in the construction of pUC8.

CHAPTER 4

DISCUSSION

4.1. Methods and Strategies for Producing a cDNA Library.

An ideal method for the synthesis of cDNA would consist of the following steps:

First strand synthesis would have to initiate from the 3' end of undegraded mRNA; this is efficiently achieved by priming the polymerase with oligo (dT)12-18, and should proceed with no "mistakes" completely to the 5' end of the mRNA where it would stop. The enzyme used for this process, avian myeloblastosis virus (AMV) reverse transcriptase, has a number of properties and requirements which are both advantageous and disadvantageous. The AMV enzyme will only initiate from a primer and will polymerise DNA 5' to 3' in the conventional manner using either ssDNA or ssRNA as a template. The primer may be either DNA or RNA with a 3' hydroxyl group and must form a region of double stranded nucleic acid. The new DNA strand is formed on this 3' hydroxyl group and progresses along the template.

The AMV enzyme also has 5'-3' exoribonuclease and 3'-5' exoribonuclease (RNAse H) activities, which specifically degrade RNA in DNA:RNA duplexes. These properties, combined with the transient formation of hairpin loop structures, from which AMV reverse transcriptase is known to prime if somewhat inefficiently (Land et.al., 1981; Murray et.al., 1983), may give rise to inverse repeat artifacts (Fagan et.al., 1980; Volckaert et.al., 1981; Weaver et.al., 1981; Geraghty et.al., 1982; Rasmussen et. al., 1983; Delauney 1984). Murray et.al.,(1983) exploited this to produce near complete dscDNA in one step. The AMV

enzyme is also slightly prone to making point "mistakes" where the wrong base is added to the chain (Gopinathen et. al., 1979).

The AMV enzyme is thought to be reasonably efficient at producing full length transcripts, provided the mRNA is undegraded. However, secondary structures formed by the mRNA or some specific sequences within it, may cause the AMV enzyme to pause which would cause under representation of 5' ends in cDNA libraries.

Control of the undesirable RNAse activities is achieved by the use of vanadyl ribonucleoside complexes or with human placental or rat liver RNAse inhibitor. Both inhibit RNAse activity without interfering with other enzyme activities, but RNAse inhibitor being a protein of Mr ∿40,000 is easily removed by phenol extraction, whereas the vanadyl ribonucleoside complexes require multiple extractions to remove them efficiently (Berger and Birkenmeir 1979; Maniatis et.al., 1982; Scheel and Blackburn 1979; de Martynoff et.al., 1980). The 3'-end hairpin-loop formation and the production of inverse repeats is inhibited by including 4 mM sodium pyrophosphate in the reaction mix (Murray et.al., 1983); however, these experiments used no RNAse inhibitor or vanadyl ribonucleoside complexes. The present investigations showed no size differences (on agarose gel) for sscDNA made with and without pyrophosphate, using RNAse inhibitor in the reaction mix (Fig.6. Section 3.1.1). The use of pyrophosphate will necessitate its removal before second strand synthesis by the hairpin-loop method.

The scheme of Okayama and Berg (1981) (Fig.5) differs

by priming with a fragment of plasmid carrying at one end a tail of dT residues about 60 bases long, and a 1.5-3 times molecular excess of poly(A)$^+$ RNA. This has the advantage of always uniquely orientating the cDNA in the final plasmid, but is otherwise identical with conventional methods of sscDNA synthesis. The present results (Fig.4.) show that the sscDNA synthesis method used gives sscDNA bands of similar size to the major bands in the poly(A)$^+$ RNA (Evans et.al., 1980).

Second strand synthesis should, as for the first strand synthesis, initiate at the 3'-end of the first strand cDNA and proceed without errors to the 5'-end, where it should stop. To accomplish this step many methods have been used and suggested, as well as avoiding it entirely by ligating the sscDNA:mRNA hybrid duplex into a plasmid and using this to transform a suitable strain of bacteria (Wood and Lee 1976; Zain et.al., 1979). However, this method is at least 10 times less efficient than using dscDNA; e.g., Okayama and Berg (1981) obtained just 50 transformants per microgram of mRNA by essentially this method compared with 100,000 for their full synthesis method

The process of second strand synthesis can be broken down into three processes which do not necessarily occur in a set order : removal of the RNA from the DNA:RNA hybrid duplex, the production of a priming site and the synthesis of the cDNA from this priming site.

Generally, one of three methods for the removal of the RNA is employed: 1) heat-denaturing, which melts the duplex followed by rapid cooling to prevent rehybridisation;

2) alkaline hydrolysis, whereby the RNA is completely hydrolysed by mild alkali (e.g.0.3M NaOH): 3) treatment with RNAse H which specifically degrades RNA in RNA:DNA duplexes (Okayama and Berg 1981). However some schemes do not specifically remove the RNA and presumably rely on the RNAse activities of the polymerases used in first and second strand synthesis.

The Okayama and Berg scheme takes the sscDNA:mRNA duplex, formed on the end of a plasmid fragment, and tails it with ∿20 dC residues, using terminal deoxynucleotidyl transferase. The plasmid fragment is then cleaved with restriction endonuclease HindIII, (HindIII cleaves sites in DNA:RNA duplexes very inefficiently) and in one step primed with a second plasmid fragment carrying a HindIII "sticky" end at one end and a dG tail ∿20 bases long at the other; the RNA is removed; the second strand is synthesised; and the total plasmid is circularised using a mix of DNA polymerase I, E.coli DNA ligase and RNAse H (Fig.5).

The production of a priming site for second strand synthesis is normally achieved by one of two methods, though others have been suggested.

The hairpin-loop or self-priming method relies on the transient formation of base mismatched loop structures at the 3'-end of the sscDNA, which provide a length of duplex DNA with a 3' hydroxyl group from which a polymerase can prime. There has been much speculation but the mechanism of formation of these structures remains largely unknown.

The second method uses TdT to tail the sscDNA at its 3'-end and then anneals it with a primer, either an oligdeoxynucleotide or a tailed plasmid fragment (Land et.al., 1981; Murray et.al., 1983; Okayama and Berg 1981).

Another possible method would be to ligate a piece of dsDNA to the sscDNA:mRNA duplex using T4 DNA and T4 RNA ligases. The dsDNA region would provide the priming site after the removal of the RNA from the sscDNA:mRNA region, the polymerase effectively filling in a region of ssDNA on one end of a piece of dsDNA. Delauney (1984) suggested using synthetic linkers as this dsDNA primer, which would produce dscDNA with linkers ready attached. The problems of using linkers are discussed below.

A number of polymerase enzymes have been used to produce the second strand. DNA polymerase I was originally used in the method of Efstratiadis et.al., (1976), but because of its 5'-3' exonuclease activity is now normally superceded by its large fragment (Klenow polymerase), which lacks this activity. Both enzymes have 3'-5' exonuclease activities but they are blocked by the polymerase activities. AMV reverse transcriptase is also used and has no exonuclease activities, but is normally used only in the specifically primed methods as it primes inefficiently from hairpin loops (Land et.al., 1981; Murray et.al., 1983). Maniatis et.al., (1982) recommend the use of both enzymes sequentially to overcome any problems of enzymes pausing at certain sequences; presumably the two enzymes pause at different sequences. The chosen hairpin-loop scheme gave a significant increase in the number of molecules of sizes differing from those of the major bands in the poly(A)+ RNA as shown by the ill defined band pattern in Figure 6.

However, there seems to be mainly complete or near complete dscDNAs as the band pattern still persists to a certain extent.

Having produced dscDNA the next step was to insert it into a suitable vector. This step should be efficient and should not degrade the cDNA. This vector (in combination with its host) should carry some convenient recognition systems so that vectors containing inserts are easily sorted from those without, and a "selectable gene" such as an antibiotic-resistance gene to select hosts carrying vectors and place the host under selection pressure to maintain the vector. The vector should be relatively small and contain a unique restriction site in the recognition system to clone the cDNA into. It should be stable in the host used so that inserts are not deleted or mutated and carry a convenient range of restriction endonuclease sites to allow the insert to be manipulated and mapped. High copy number to give good yields in preparations, ease of handling and storage, safety and other properties must also be considered.

The chosen vector, plasmid pUC8 (Vieira and Messing, 1982), carries an ampicillin resistance gene and a multi-purpose cloning site in its lac gene which is disrupted by the insertion of DNA into the cloning site. This lac gene in combination with the chosen host E.coli strain TB1 produces an enzyme which cleaves X-gal to produce a blue dye. Thus colonies with plasmid bearing inserts grown on media containing X-gal are white and those without inserts are blue.

Many other plasmids are available for cDNA cloning, carrying a variety of antibiotic resistance genes and other markers. Most of these are essentially very similar to pUC8 and hence pBR322 (Bolivar et.al., 1977; Sutcliffe

1978, 1979)) being plasmids based on a common origin of replication. One other class of vectors forms a strong contender for normal cDNA cloning, namely those derived from the E.coli male-specific single-stranded filamentous DNA phage M13. Messing (BRL Ltd. M13 cloning/ dideoxy sequencing manual) (Vieira and Messing, 1982;) has developed these as vectors using a similar lac gene system containing a multipurpose cloning site as for pUC8.

The life cycle of phage M13 takes it through both single and double stranded circular forms. The existance of the double stranded replicative form (RF) allows M13 to be treated as a plasmid and allows its manipulation with restriction endonucleases, ligase and other enzymes. RF M13 naturally exists only within a host bacterium from which it can be isolated (similarly to plasmids) and competent bacteria can be transformed with RF M13 as with a plasmid. A bacterium containing M13 will be constantly secreting, without lysis, single stranded M13 packaged in a protein coat, the phage form. Only one strand (the + strand) is secreted in this manner and these phage particles are able to infect other susceptible bacteria.

Messing's M13 vectors are relatively small, 7-7.5 kb in length, exist in a high copy number at ∿200 copies per cell of the RF, have a simple system for selecting bacteria bearing phage with an insert (essentially the lac system as in pUC8) and have in their multipurpose cloning sites a variety of unique restriction sites. Selection pressure to maintain M13 in its host is provided by its phage life cycle. Any host cell not containing a phage will be susceptable to infection by the phage particles being constantly secreted into the medium by infected cells. The phage form of M13 is also ideal for DNA sequencing and hybrid-release translation experiments

(Chandler, 1982).

The most significant defect in M13 vectors is that they are appreciably more efficient at cloning small DNA fragments than large ones, which would introduce significant distortions of representation in a cDNA library. It has also been suggested that inserts in M13 vectors are more prone to partial or complete deletion than inserts in plasmids, but this is probably a function of the host strain. All host strains for M13 are rec^+ and this increases the possibility of artificial deletions in the cDNA clones.

The preparation of RF M13 is fairly inefficient as the host bacteria will only grow to fairly low densities, thus making restriction mapping less convenient. The secretion of the phage form into the media makes isolation of the + strand easy, however should both strands of the insert be required, either the RF form or a subclone carrying the insert in the opposite orientation would have to be isolated.

The vector of choice, pUC8, was preferred over other plasmid vectors, because of its lac insert screening system and over M13 vectors, as extensive restriction mapping of promising clones was envisaged.

The method chosen for inserting the cDNA into the vector was to make the cDNA blunt-ended, attach synthetic linkers and insert into the relevant site in the vector's multipurpose cloning site. EcoRI linkers were chosen as they are cut by a very "well behaved" enzyme and because previous libraries made in Durham had used BamHI linkers.

The "blunt-ending" was achieved by sequential use of S1 nuclease and DNA polymerase I. S1 nuclease is a

single-strand specific nuclease which will trim off lengths of single stranded DNA on the ends of duplex DNA to leave blunt or near-blunt ends. The hairpin-loop structure as it is base-mismatched will also be cleaved. Unfortunately S1 nuclease will also cleave duplex DNA at "nicks" or small gaps (Kroeker and Kowalski, 1976) and exposure to large amounts of enzyme will completely digest dsDNA. This is an acknowledged major defect in the self-primed method of producing dsDNA; even should the S1 nuclease perform perfectly removing only the base mismatched hairpin loop, inevitably part of the 5'end of the message is lost. In practice, S1 nuclease shortens the dscDNA, presumably by nibbling off transient single stranded regions of the duplex, particularly at the ends. Certainly this loss of some of the 5' end of the message introduces an over representation of the 3' end of messages in cDNA libraries.

An alternative enzyme is available for cleaving the hairpin loop, namely mung bean nuclease which has essentially the same activities as S1 nuclease. However, it is thought (Kroeker and Kowalski, 1976) that mung bean nuclease is more gentle in its action, being unable to cleave DNA duplexes at nicks until they are enlarged into gaps of several nucleotides.

The second step in the production of accurately blunt ends is "polishing" with DNA polymerase I, using its 5'-3' polymerase and 3'-5' and 5'-3' exonuclease activities to fill in or cleave off any small irregularities. Maniatis et.al., (1982) obtained improved transformation efficiencies using this step. However, again the cDNA maybe slightly

degraded.

The attachment of synthetic linkers to the dscDNA is an efficient process, as a large molecular excess of linkers over dscDNA makes the chance of a dscDNA molecule not having linkers attached to both ends, being circularised or being attached to another dscDNA molecule, remote. However, the use of linkers involves a possible major degradation of the dscDNA. In order to produce sticky ends for the ligation of the dscDNA into the vector, the linkers must be cleaved with the appropriate restriction enzyme and any sites for the enzyme within the dscDNA will also be cleaved. Treatment of the dscDNA with the appropriate methylase (to stop it being cleaved by the cloning enzyme) before attaching the linkers would raise another problem : should a plasmid be isolated and found to have two inserts, it would be impossible to simply tell whether there are genuinely two inserts or one insert containing a site for the cloning enzyme. For a cDNA library made from non methylase-treated dscDNA any clone bearing multiple inserts may be treated as carrying two independant and unconnected inserts. The complete sequences of cDNAs carrying recognition sites for the cloning restriction enzyme can be deduced from comparisons of sequence and cross hybridisation data with a second cDNA library made using a different cloning restriction enzyme, or with a genomic bank.

Finally, the dscDNA is ligated with the chosen vector, prepared by restriction with the cloning restriction enzyme, to produce circular plasmid molecules, suitable for transforming competent cells of the chosen host strain. Normal practice is to use a molecular excess of vector over

dscDNA. Ratios between 1.5:1 and 5:1 being commonly employed. This will reduce the incidence of multiple inserts. Plasmid pUC8's lac insert detecting system makes phosphatase treatment of the vector to prevent recircularisation unnecessary; phosphatase treatment reduces ligation efficiency.

Alternative methods of inserting dscDNA into a vector are: Attachment of different linkers to either end of the cDNA, one before one after S1 nuclease treatment (Maniatis et. al., 1982). Here, S1 nuclease may well destroy the first linkers, and two sets of linkers double the chances of the dscDNA being cleaved at recognition sites for the cloning restriction enzymes, making it more difficult to reconstruct the sequences of complete messages. The orientation of the cDNA in the vector is however uniquely defined. Another method involves annealing homopolymer-tailed dscDNA with complementarily homopolymer-tailed vector; yet another method is the scheme of Okayama and Berg (1981).

These latter two methods rely on terminal deoxynucleotidyl transferase (TdT) to produce the homopolymer tails. TdT is notoriously difficult and erratic and will be discussed later (Section 4.7). These tailing schemes are normally designed to regenerate restriction sites at the ends of the insert. Slight nuclease activity which is often found in even extensively purified TdT preparations, may destroy these sites and make isolation of the insert extremely difficult. The use of tails in any case, makes the exact ends of the message uncertain. Finally, it has been reported that polymerases read through extended regions of G or C bases poorly. This would be a big draw-

back in DNA sequencing and nick translation. In addition it has been reported that annealed plasmids do not transform competent bacteria as efficiently as ligated plasmids, but this may merely reflect a poor choice of host.

Having produced plasmids containing inserts the next step in the procedure is to choose a strain of bacteria and to transform it with the plasmids. There is a vast range of strains of E.coli, mostly derived from E.coli K-12, which have been used for transformation. For those vectors using host specific systems, such as the lac complementation system of pUC8 with TB1 or JM83, the host strain is fixed. Otherwise ease of handling, high transformation efficiencies, safety and the deletion or inclusion of specific traits are the criteria for choice.

Two major methods (with variations) are used to make cells competent for transformation with plasmid DNA. The calcium chloride method (Mandel and Higa, 1970), which was the method used here, can give up to 10^7 transformants per μg of plasmid pBR322 with E.coli strain HB 101. Efficiencies approaching this were achieved in the test transformations. The major variation of this method is the prolonged incubation method of Dagert and Ehrlich (1979) where the optimum transformation efficiency is achieved after incubation for 24 hours at 0°C in 100 mM calcium chloride. The present results show reduced transformation efficiency after this procedure, however this effect is probably a function of the strain used.

The other method is the rubidium chloride method of Kushner (1978), which was designed for E.coli strain SK1590 but does give higher transformation efficiencies than does

the calcium chloride method for some strains.

Maniatis et.al., (1982) recommends a method developed by D.Hanahan (unpublished) for use on E.coli χ1776, which with minor changes gives transformation efficiencies of 10^7-10^8 transformants per μg of pBR322 with some strains. The commonly used strain HB101 does not respond well to this method.

The final steps in the production of a cDNA library are the selection of those bacterial colonies carrying plasmids with inserts and their preservation in some convenient form. Maniatis et.al., (1982) provides a series of methods. The choice of a method is dictated essentially by practicalities and purpose for the library produced.

4.2. General Assessment of the cDNA Library.

The cDNA library produced contained 686 colonies after final picking, of which ∿52% showed as positive in colony hybridisation screens with mRNA as a probe. This agrees well with the sequence complexity data of Morton et.al. (1983) which, for 14 days after flowering pea seed mRNA, show the "very abundant" sequences representing ∿42% and the "abundant" sequences ∿24% of the mRNA. Not all of the "abundant" sequences will be abundant enough to give a signal in the mRNA colony screen; cDNAs which are short relative to the length of the probe will give relatively weak signals for their abundance class; there will be some non-cDNA inserts and some colonies with no inserts. Not all blue colonies were rejected and some colonies may contain pUC8 with a short deletion. Thus the mRNA probe should have found all the "very abundant" cDNAs and about half the "abundant" cDNAs.

The average insert size of ∿590bp is significantly shorter than any estimate of the average mRNA length (See Fig. 4). The isolation of many cDNAs which represented only parts of their mRNAs confirmed that the cDNA had been significantly degraded before being cloned. However, the insert of pLG1, 63, a vicilin B positive, at 1530bp in length could be a very nearly complete cDNA (Fig.10).

The method chosen for screening the library, i.e. colony hybridisation screening, was shown to be fully effective by the results in tables 26, 27 and columns E and F in table 38. Only a weak correlation between insert size and strength of hybridisation to mRNA, and a strong correlation between strength of hybridisation to mRNA between colony hybridisation screens and Southern blots are seen. The differences between the colony hybridisation screen and the Southern blot results may be explained by the higher "signal to noise" ratio of Southern blots. Columns E and F of table 38 show identical mRNA colony hybridisation screenings of different sets of replica filters. The results correlate well, taking into account the greater sensitivity of the autoradiography technique used for the column F results. The results in column E were obtained from a single autoradiograph exposure, and the filters were unfortunately destroyed, so a second longer exposure to find weaker signals was not performed. The differences between the results in columns E and F are mainly weak and very weak signals not being detected in column E, other greater differences, due to uneven growth of the colonies on the filters, are rare.

The mRNA positives were classified (Table 35) according

TABLE 35

Identification of clones that are positive with the mRNA probes.

mRNA positives			Identification	
Strength	Number	Positive	Dubious	Unidentified
Strong	55	41	7	7
Medium	45	29	8	8
Weak	144	63	28	53
Very Weak	115	30	20	65
Total	359	163	63	133

to strength of hybridisation and homology with other probes. The trend observed is that, the strong and medium positives are mainly identified as positives for one of the other probes used, the weak roughly half-identified half-unidentified and the very weak mainly unidentified. These screens are not perfect as some strong positives for legumin and vicilin probes have not shown up as mRNA positives, which is probably caused by uneven growth of the colonies on the filter.

The abundant seed proteins not specifically probed for were : the major albumins on filters 1, 2 and 3. These would not be expected to show up more strongly than mRNA medium,as major albumins constitute only ∿8-10% of the total seed protein (Croy et.al., 1984). The Mr ∿8000 albumin which represents ∿4% of the total seed protein (J.Gatehouse, personal communication) would probably show as mRNA weak, as it would have a short mRNA. Convicilin however, is known to be roughly half to a third as abundant as any one type of vicilin, as estimated from gene copy numbers, and would have a longer mRNA than vicilin. Thus it would be expected to account for some of the mRNA strong unknowns. Finally, the Mr ∿80,000 legumin-related polypeptide (Domoney and Casey 1984), which is thought to give rise to the small legumin subunits isolated *in vivo*, would have a similar abundance, by gene copy number, and an even larger mRNA than convicilin and hence, should account for some more mRNA strong unknowns.

4.3. Legumin cDNAs

4.3.1. Main Legumin cDNAs.

The two main legumin positives which were restriction

mapped (Fig.9) share the same restriction pattern as the main legumin cDNA used as a probe, pAD4.4 (pDUB6), and the legumin A gene taking into account introns. Legumin genes A, B and C have nearly identical restriction maps, and hence these two positives are main legumin cDNAs.

The other two main legumin screens, using as probes the legumin C gene at ∿99% homology and the legumin A gene at ∿81% homology, found a further three main legumin positive colonies over the screen with pAD4.4 (pDUB6). The inserts of two of those were sized (Table 28) and found to be small enough to be cDNAs of the 5' end of main legumin message, which is not covered by the cDNA probe pAD4.4 (pDUB6). Thus they are not cDNAs of some atypical main legumin message.

The strengths of hybridisation of the main legumin cDNA probe (Table 14) on the whole reflect the strengths of hybridisation of the mRNA probe to the same colonies.

The total number of 67 main legumin positives in the library, when probed at ∿94% homology, represents 9.8% of the library or ∿23% of the 14 day "very abundant" sequences of Morton et.al., (1983). This total may well rise if low stringency screens were performed and more legumin cDNAs corresponding to small parts of the probe found. Thus main legumin message represents at least one of the estimated six "very abundant" sequences.

It is known (Domoney and Casey, 1985) that main legumin cDNAs do not cross hybridise with either big legumin DNAs, which was confirmed (Fig.19), or with the Mr∿80,000 legumin-related polypeptide cDNA. Also none of the colonies which showed a large increase in hybridisation intensity between the

the legumin C at ∿99% homology and the legumin A at ∿81% homology screens, or only showed as positives in the legumin A screen, were big legumin positives. Since legumin A and C are very closely homologous (J.Gatehouse, personal communication) the differences between the two screens are due to the general increase in hybridisation intensity at the lower stringency. This occurs particularly for GC poor and relatively short inserts, as they are washed at effectively higher stringency than the average insert. Thus there is no evidence for the presence of main legumin related cDNAs from the colony screens.

4.3.2. Big Legumin.

The big legumin positive which was restriction mapped (Fig.18) had a restriction map similar to that of legumin gene J and differing from that of legumin gene K. The two ends of its insert were sequenced (J.Gatehouse, personal communication) and the sequences found to match perfectly with that of legumin gene J, despite the legumin J gene being isolated from a different variety of pea Dark skinned perfection than the cDNA library (Feltham first). Plasmid pCD40, the big legumin cDNA used to isolate legumin genes J and K has an identical restriction map to that of legumin gene K, taking into account introns and as far as sites have been tested for, except for one extra Pst I site marked with a * . This difference may be due to the cDNA library pCD40 was isolated from being made from mRNA from the Birte variety of pea. However, it is known there is a third big legumin gene (Legumin L) which could be as similar to

legumin K as are the main legumin genes to each other.

The strengths of hybridisation of the big legumin probe to positive colonies is greater than that of the mRNA probe to the same colonies; however, nearly all big legumin positives are mRNA positives (Table 16). When probing the library at ~91% homology, the total number of 28 big legumin positives represents 4.1% of the library or ∿9.8% of the 14 day "very abundant" sequences, (Morton et.al., 1983). This number may increase in a lower stringency probe as legumin genes J and K are much less homologous than the main legumin genes. Big legumin message thus represents one of the six "very abundant" sequences.

The cDNA for the Mr∿80,000 legumin-related polypeptide of Domoney and Casey (1984,1985) is known not to cross hybridise with big legumin genes in genomic blots, which were washed in 0.1 x SSC, 0.1% SDS at 65°C. This corresponds to ∿99% homology. However, in hybrid selection experiments and Northern blot analyses, considerable cross hybridisation was seen. In these experiments the washing was to similar stringencies to that of the colony hybridisation screen. Thus weak and very weak positives for the big legumin probe, particularly those with large inserts may be cDNAs of the Mr∿80,000 legumin related polypeptide. Alternatively, a lower stringency colony screen with the big legumin probe may find some of these Mr∿80,000 legumin cDNAs.

4.3.3. The legumin genes.

Six legumin genes have been isolated and the existence of at least three more can be deduced from the evidence

of genomic blots and indirectly, from the heterogeneity in size and pI values of legumin subunits. From cross hybridisation data and DNA sequence homologies an "evolutionary" diagram of the legumin gene family has been compiled (Fig.25). The isolated main legumin genes A, B and C are so homologous they have not been separated on the diagram, and all are closely related to the pseudo legumin gene D. Legumin genes J and K are much less homologous to each other than are the main legumin genes and are closer to the putative legumin X gene which gives rise to the Mr80,000 legumin-related polypeptide than to the main legumin genes.

4.4. Vicilin cDNAs.

The two vicilin-positive clones restriction-mapped (Fig.10), have the same restriction pattern as the vicilin B (pAD3.4 plus pRC2,2,10 (pDUB7 plus pDUB4)) composite map. The three types of vicilin have differing restriction maps, thus the two vicilin positives mapped are both cDNAs of type B (47k) vicilin. The restriction map of pLG 1,63 suggests that it must be close to being a full length cDNA, being, at 1530 bp long, 100 bp longer than the idealised vicilin mRNA of Lycett et.al., (1983).

Colony 1,63 showed stronger hybridisation to pAD3.4 (pDUB7), the vicilin B probe, than to the other vicilin probes at ∿94% homology. However, colony 1,9 showed strong hybridisation at ∿94% homology to all the vicilin probes. This suggested that colony hybridisation screens were sufficiently sensitive to at least partially discriminate between cDNAs of different types of vicilin.

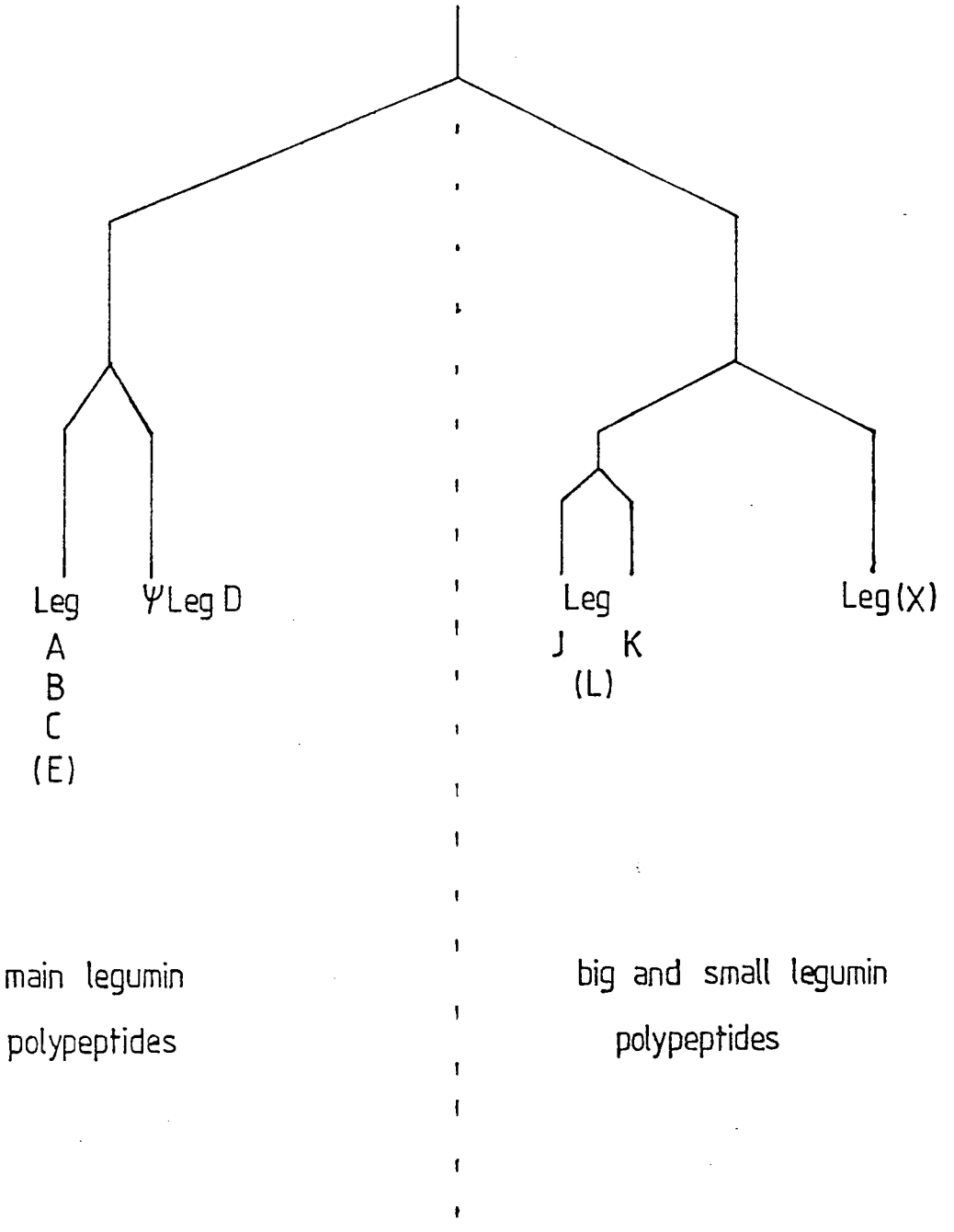


Figure 25: The legumin gene family.

( ) = Genes known to be present from genomic blots but not isolated.

Delauney (1984) has estimated the sequence homology between any two types of vicilin mRNA to be ≈85% and this close homology is demonstrated by the extensive overlapping of positives for the different vicilin probes, especially in those screens washed at lower stringencies.

The related protein, convicilin, does not show sufficient DNA sequence homology to the vicilins to select vicilin message in hybrid-release translation experiments (Domoney and Casey, 1983). Thus convicilin cDNAs would not show as positives in colony hybridisation screens with a vicilin cDNA as probe, except possibly at very low stringency.

Thus, using the results from the colony hybridisation screens at different stringencies for the different vicilin type probes, the vicilin positives were classified according to the intensity of hybridisation into type A, B or C vicilin cDNAs (Table 36). Only those colonies which showed positive at ∿94% homology were considered to avoid a bias against type A vicilin for which there were no lower stringency screens. The lower stringency data was used to distinguish between colonies hybridising equally intensely to two probes at ∿94% homology. This is reflected in the low number of colonies equally positive for types B and C vicilins. The total number of positives for each probe, which was obtained by splitting evenly the numbers of colonies equally positive for more than one probe, are roughly equal. This reflects the roughly equal amounts of the three vicilin type proteins isolated *in vivo* and found in *in vitro* translations of total mRNA.

In order to try to decide if any of the vicilin types

TABLE 36

Classification of vicilin positive clones at ∿94% homology into type A, B or C vicilin cDNAs.

Vicilin type	Specific Positives	Total Positives	Percentage of Library
A	18	∿24	∿3.5%
B	22	∿26	∿3.8%
C	19	∿23	∿3.4%
A+B	5		
A+C	6		
B+C	2		
A+B+C	1		
Total	73	73	10.6%

Specific positives represents the total number of colonies which show most/equally strongly in the high stringency colony screens when probed with the given probe/probes.

Total positives represents the total number of colonies which carry cDNAs for the given vicilin type. The numbers were arrived at by adding to the specific positives for each vicilin type a share of the colonies specifically positive for more than one vicilin type.

is less homologous to the other two than those two to each other, the vicilin positives were classified on the basis of order of hybridisation intensities with the vicilin probes (Table 37). Again, only those colonies showing as vicilin positives at ∿94% homology were considered, but the lower stringency probe data was used to distinguish between these colonies hybridising equally to the vicilin B and C probes at ∿94% homology.

Due to the lack of low stringency screens with the vicilin A probe some slight distortions may be expected, i.e. there may be an excess of colonies to which vicilin A hybridises equally as intensely as one of the other vicilins. Actually only type C positives equally homologous to types B and A vicilin probes show a significantly anomalous value. These effects would be expected to balance out somewhat. It would also be expected that greater homology between vicilins A and C and B and C than between vicilins A and B would be observed due to the extent of overlap between the probes (Fig.10).

Table 37 may be summarised as showing : 24 colonies where vicilins A and B are most similar in hybridisation strengths, 23 colonies where vicilins B and C are most similar, 27 colonies where vicilins A and C are most similar and 3 colonies where vicilins A, B and C are equally similar. This shows no real difference in hybridisation strengths of cDNAs of one vicilin type to probes for the other vicilins. This confirms Delauney's (1984) estimate of ∿85% homology between any two vicilin types, although it should be remembered that colony hybridisation screening is not the most sensitive of

TABLE 37

Cross classification of vicilin positive clones with each other.

Strongest (a)	Next strongest (b)	Number
A	B	3
B	A	1
A	C	5
C	A	5
B	C	9
C	B	3
A+B	C	3
C	A+B	17
A+C	B	6
B	A+C	11
B+C	A	2
A	B+C	9
A+B+C	-	3

a) Type(s) of vicilin the colony hybridises most strongly to

b) Type(s) of vicilin the colony hybridises second most strongly to.

This table shows the total number of colonies which hybridise most strongly to the given vicilin probe/probes and next most strongly to the other given probe/probes. It should be read as showing if any one pair of vicilin probes and hence types are more or less closely related to each other than they are to the third.

techniques.

The total number of vicilin positives in the library when probed at 94% homology is 73, representing 10.6% of the library or ∿25% of the 14 day "very abundant" sequences of Morton et.al., (1983). The individual type vicilins represent ∿8.3% for type A, 9.1% for type B and 8.1% for type C of these "very abundant" sequences. These vicilin cDNAs probably represent three of the estimated six "very abundant" sequences. If one includes vicilin positives found only in the lower stringency screenings, the total number of vicilin positives increases to 122, representing ∿17.8% of the library and ∿42.4% of the 14 day "very abundant" sequences. This may increase still further if low stringency screens with the vicilin A probe were performed. Strength of hybridisation of mRNA to vicilin positives matches reasonably well with the strengths of hybridisation with the relevant vicilin probe.

No vicilin genes have been isolated and fully characterised, but it seems clear that the vicilin gene family consists of three closely related groups of main vicilin genes, corresponding to the three vicilin types. More distantly related would be a smaller group of convicilin genes, and possibly some other distantly related genes, e.g. the glycosylated Mr∿65,000 minor component in convicilin preparations.

4.5. Lectin cDNAs

The probe used to find soluble seed lectin positives was the insert of plasmid pRC2,2,26. This cDNA clone was

isolated from a previously prepared pea seed cDNA library and had been partially characterised by hybrid release translation, where a polypeptide of Mr∿25,000 was produced by mRNA hybridising to the clone. From the size of the polypeptide produced and the abundance of the corresponding mRNA,it was thought to code for one of the major albumins or the soluble seed lectin. The insert of pRC2,2,26 was extensively restriction-mapped and the map compared with that deduced from the published sequence of a pea lectin cDNA (Higgins et.al., 1982) (Fig.11). Corresponding regions of the maps were identical and hence pRC2,2,26 is a pea seed soluble lectin cDNA.

The insert of pRC2,2,26 was sequenced and found to be identical to the published sequence of pea lectin mRNA (Fig.12), thus confirming its identity. The insert of pRC2,2,26 was also used to probe a Northern blot of pea cotyledon mRNA (Fig.21., Table 33). The only band to which it hybridised was 1000 bases in length, the other hybridisation in these tracks can be seen on the autoradiograph to be unspecific, thus confirming that the published lectin sequence is nearly complete, lacking only some 5' untranslated sequence. This also suggests that any other lectin gene if present is either 1) not transcribed in pea seeds at the mid stage of development, 2) not closely homologous to the soluble seed lectin, or 3) produces a message of almost identical length to the one sequenced and published.

The lectin mRNA was only present in detectable amounts in the mRNA isolated from pea cotyledons at the mid-developmental stage (∿14 days after flowering). None was detected in the 18 and 20 day mRNA suggesting that lectin is not synthesised in the later stages of seed development.

In comparison, vicilin mRNA is also not detectable in late-development stage cotyledons, however legumin mRNA is detectable at 22 days after flowering (Gatehouse et.al., 1984).

A second probe used to find lectin positives was the insert of pPVL134, which is a cDNA coding for Phaseolus vulgaris lectin (Hoffman et.al., 1982). It was originally selected from a Southern blot using a pea seed soluble lectin cDNA. It was known however, that the degree of homology between the lectins was not great and may well be too little to be detected in a colony hybridisation screen due to the relatively higher background. The sequence of Phaseolus vulgaris lectin mRNA, which extends from the translation start codon to the poly(A) tail is given in Fig.26., along with one of the regions of homology between pea and Phaseolus vulgaris lectin messages, which corresponds to the 5' end of the plasmid pRC2,2,26 insert.

The regions of homology between the two lectin message sequences were found using a dot matrix comparison program on a computer (Figs.27 and 28). The DNA sequence comparison (Fig.27), was used to look for stretches of 30 bases in one sequence, where at least 18 bases in the other sequence match. Whenever this occurs the computer prints a dot. Regions of homology show up as diagonal lines. Looking for matches over longer stretches of sequence reduces the background of chance homologies, and changing the number of matches for a given sequence length changes the degree of homology being probed for (similar to washing blots at higher or lower stringencies).

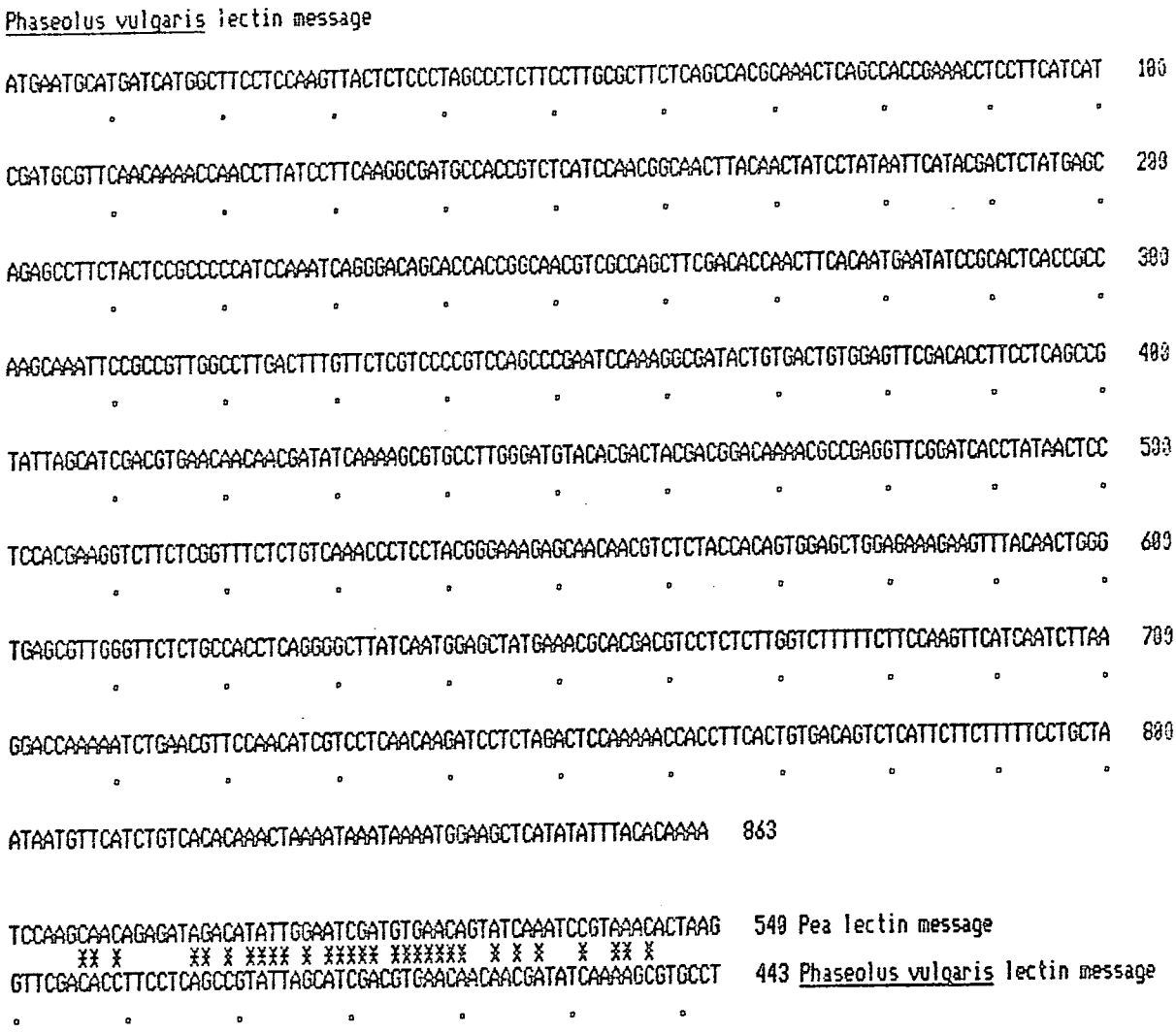


Fig. 26. The sequence of Phaseolus vulgaris lectin message and one of its regions of homology with pea lectin message.

Phaseolus vulgaris lectin message is from Hoffman et.al., (1982), pea lectin message is from Higgins et.al., (1982) and the illustrated region of homology corresponds to the 5'end of the insert of pRC2,2,26, * marks matching bases.

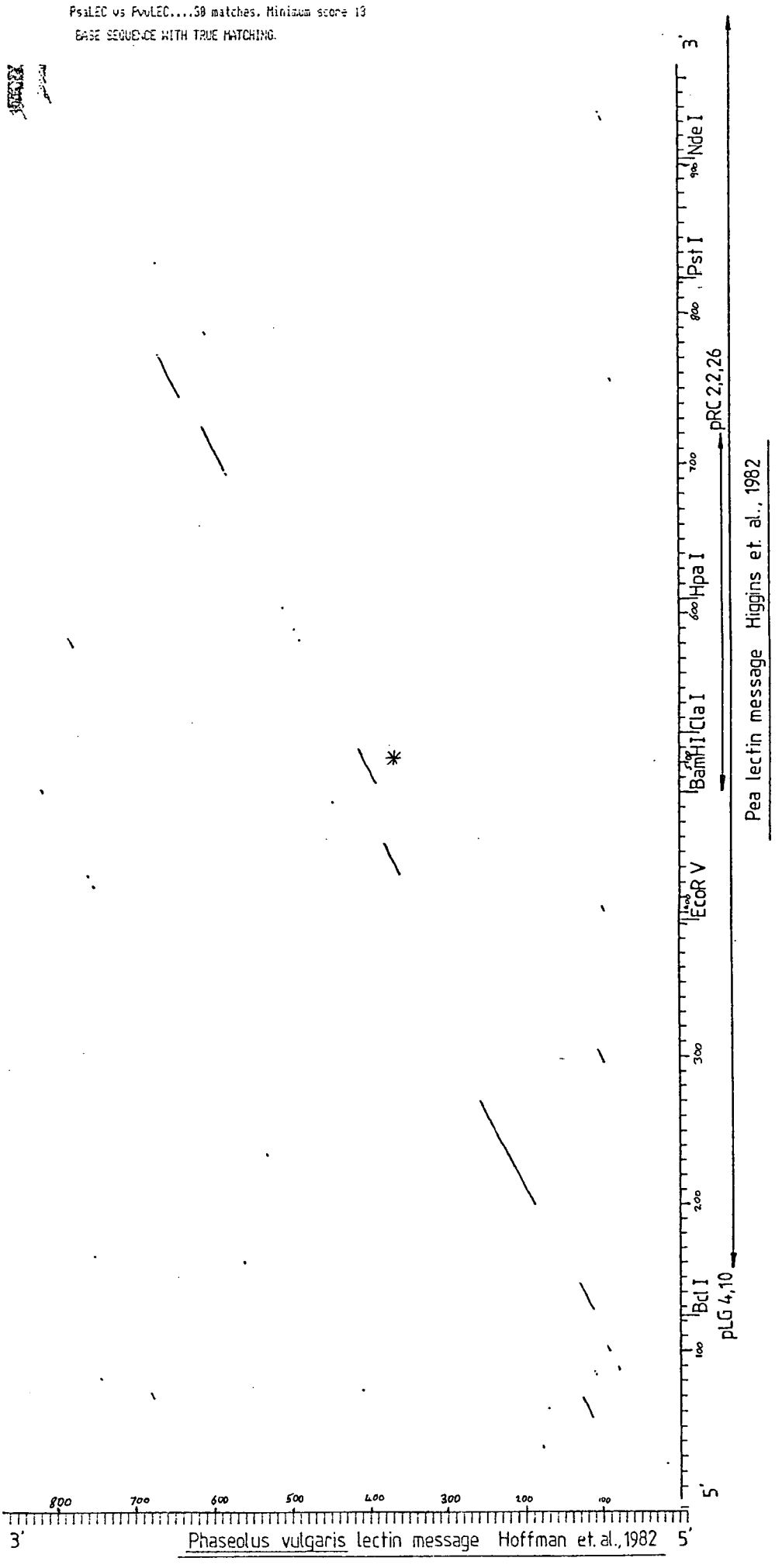


Fig.27. Dot matrix comparison of Pea and Phaseolus vulgaris lectin messages.

* Region of homology shown in Fig. 26. Scale bp

The chosen conditions, a minimum score of 18 matches in 30 base stretches, corresponds to regions of 60% homology extending over at least 30 bases, i.e. sufficient homology over large enough regions to be detected in Southern blot hybridisation probes.

The amino acid sequences of the two lectins were also compared using a dot matrix program and a matrix scoring system for the amino acid matching (Fig.28). The amino acid residues are matched by the scoring system of Staden (1982) which was developed to show homologies between the regions of divergent sequence, but maintained function. This shows somewhat more extensive homology than the DNA sequence does, as would be expected on the basis of conservation of functional domains in the protein. However, the homology is not very close.

Of the two putative lectin positives found with pPVL134 neither were pRC2,2,26 lectin positive and none of the pRC2,2,26 positives show above the non-specific background in the pPVL134 colony hybridisation screens. The rough restriction mapping carried out on these pPVL134 positives showed no similarities with the map derived from the published pea lectin message of Higgins et.al., (1982). Further when the insert of pRC2,2,26 was used as a probe on a Southern blot of the pPVL134 positives (Table 30), no specific hybridisation was detected. It was thus decided that these two pPVL134 positives were not seed lectin positives but spurious signals on the autoradiograph.

The colony hybridisation screen with the insert of pRC2,2,26, at~94% homology showed 9 colonies as positives, representing ~1.3% of the library and ~5.4% of the 14 day

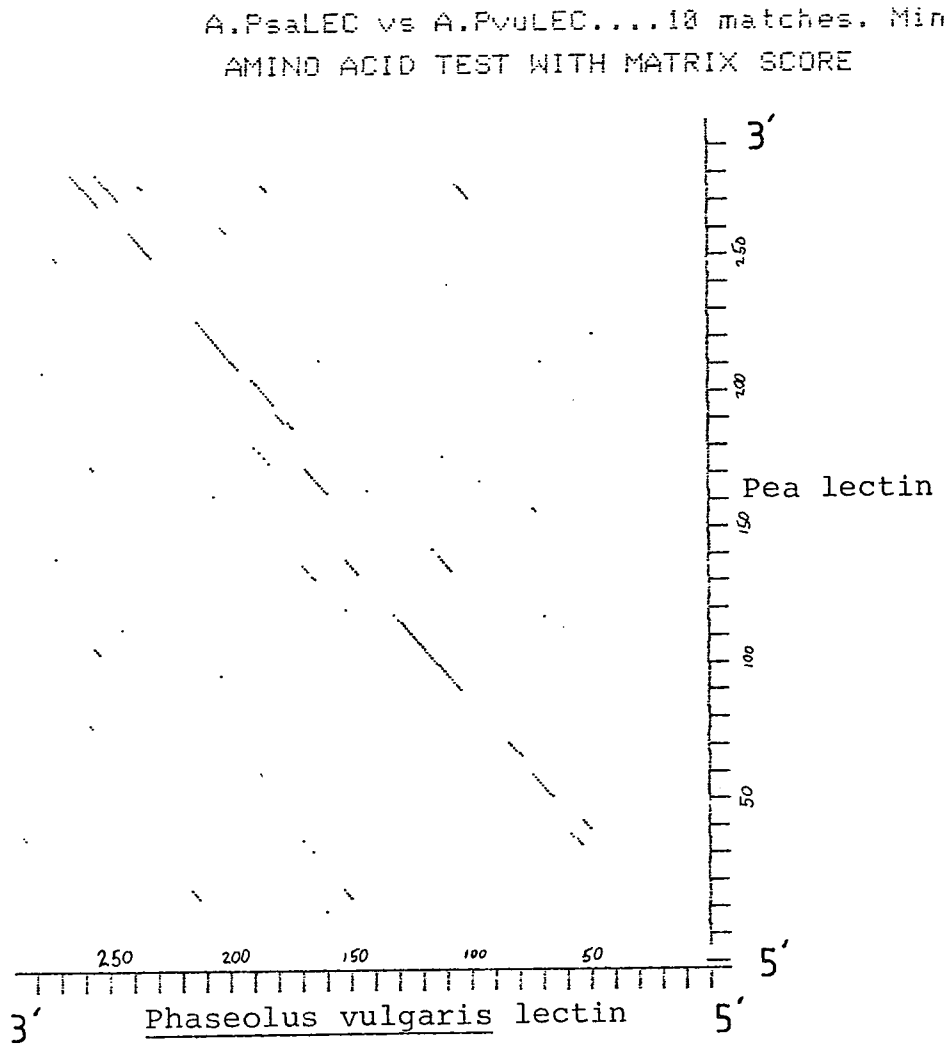


Fig.28. Dot matrix comparison of the amino acid sequences of Pea and Phaseolus vulgaris lectins.

Pea lectin : Higgins et.al., (1982)
Phaseolus vulgaris lectin : Hoffman et.al., (1982)

Scale Amino acid residues.

"abundant" sequences of Morton et.al., (1983). This total would probably rise if a more complete lectin probe was used, as pRC2,2,26 covers only ∿¼ of the lectin mRNA. Whether any further positives would be discovered at lower stringencies would depend on there being other transcribed lectin genes (see above). Most of the lectin positives were mRNA positives thus lectin message is probably one of the more common of the 120 14 day "abundant" sequences of Morton et.al., (1983).

Two lectin positives were chosen to be restriction mapped because they had the largest inserts. The map of pLG4,10 is identical with that derived from the published sequence of pea lectin. However, the map of pLG2,73 is very different and as it hybridised only weakly to pRC2,2,26 it was initially thought to represent some lectin-related message. However, in hybridisation probes of Southern blots the insert of pRC2,2,26 hybridised strongly the pLG4.10, but not significantly to pLG2,73 at~94% homology, the same stringency as the colony screen. Also the insert of pPVL134 (the Phaseolus vulgaris lectin cDNA) hybridised to both pRC2,2,26 and pLG4,10 reasonably well at~77% homology, but not specifically to pLG2,73.

These two lectin positives were sequenced (Fig.14), and pLG4.10 was found to be identical to the published pea lectin mRNA except that its poly(A) tail starts 24 bases further downstream. The sequence of pLG2.73 was found to be very different to that of pea lectin. In dot matrix comparisons of the part of the insert of pLG2.73 sequenced, with pea lectin message at an equivalent to 60% homology over 30 base stretches (Fig.29), the complementary

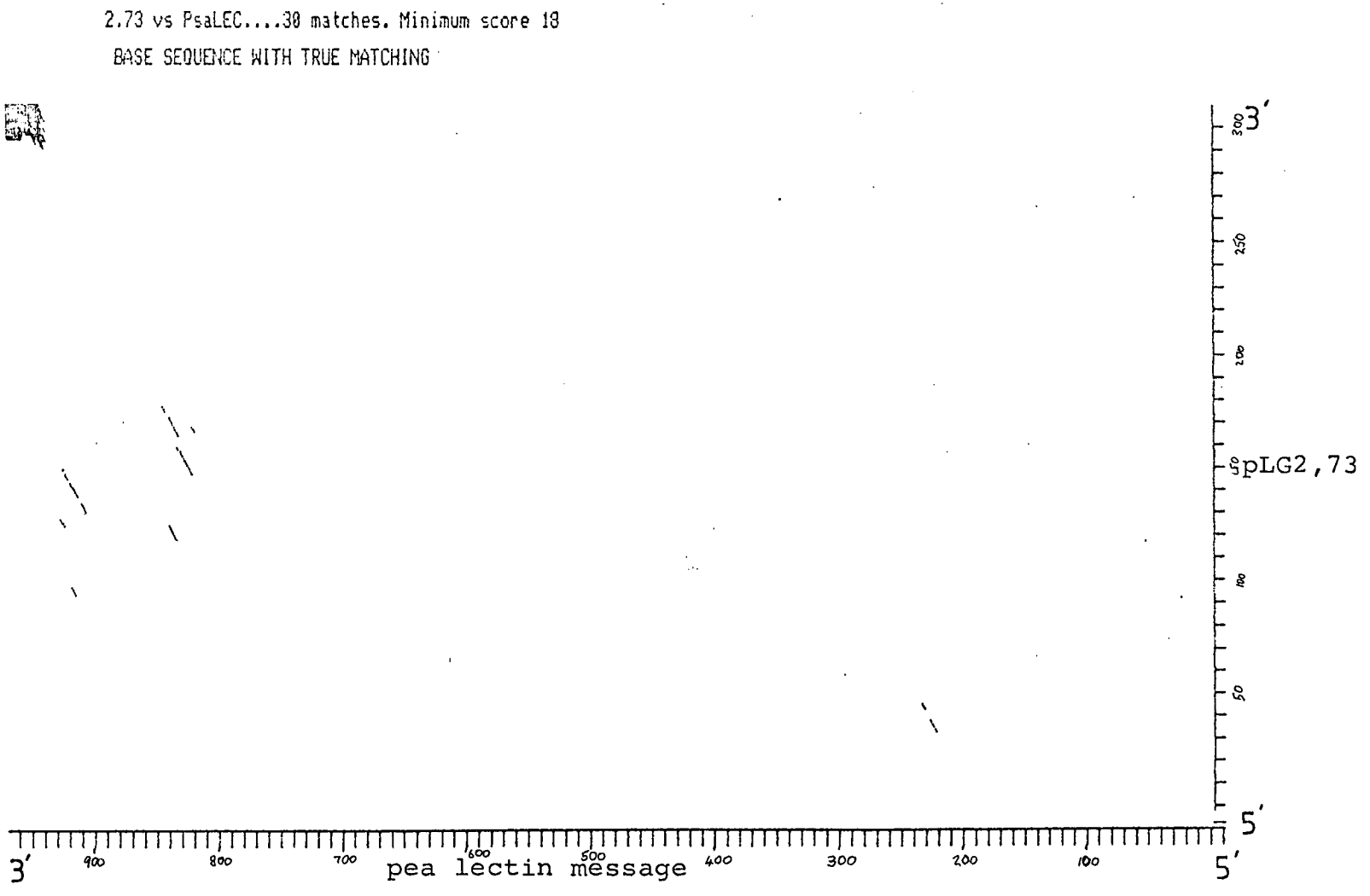


Fig. 29. Dot matrix comparison of Pea lectin message with pLG2,73

Pea lectin message : Higgins et. al., (1982). pLG 2,73 sequenced strand. Scale bp.

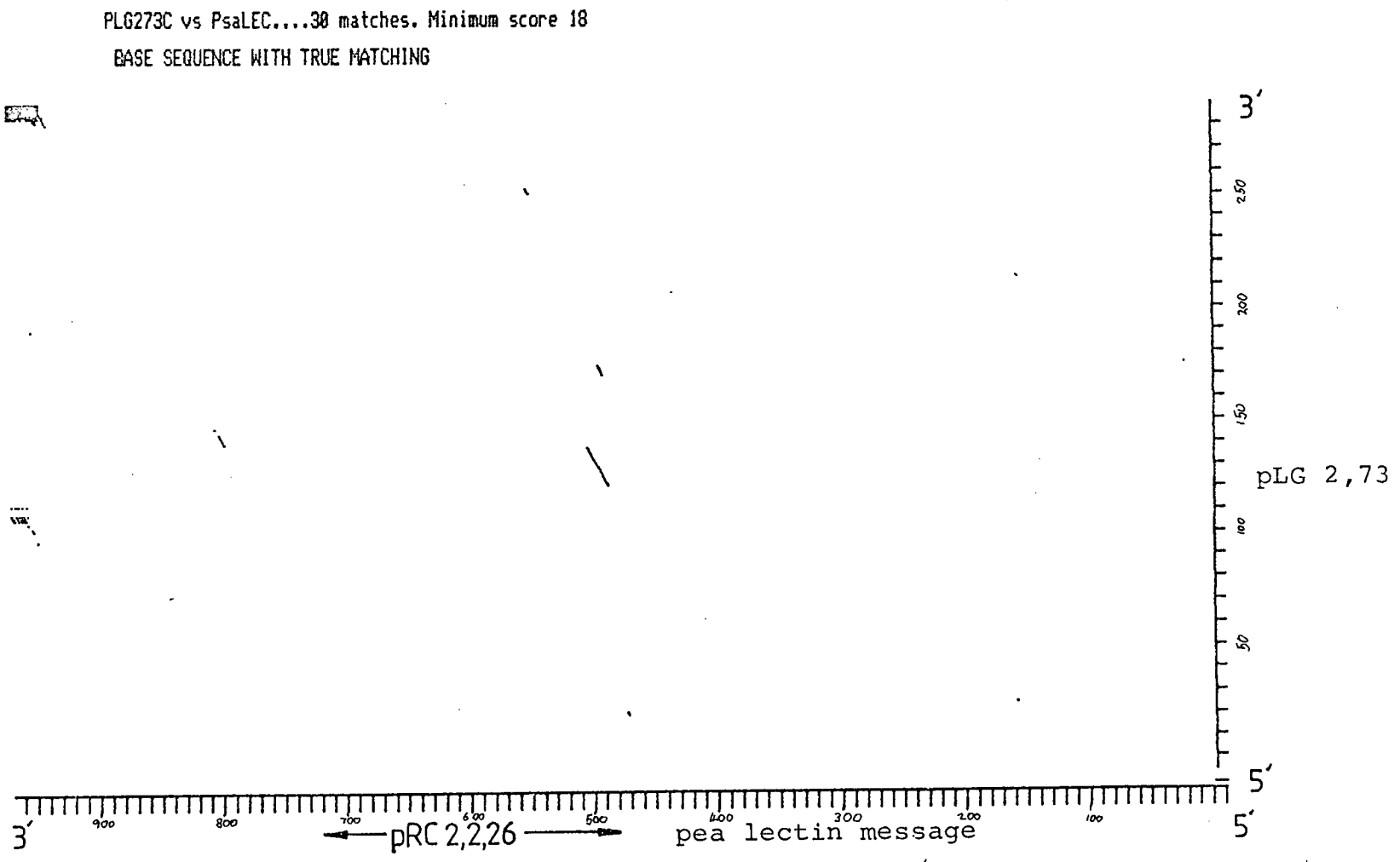


Fig. 29. (Continued) pLG2,73 complementary strand. Scale bp.

sequence of pLG2.73 shows a region of homology with pea lectin message which corresponds to the 5' end of the insert of pRC2,2,26, but few other regions of homology were evident. This area of homology may explain why pLG2,73 was found to be weakly positive in the colony screen with pRC2,2,26. Thus pLG2.73 is not a pea seed soluble lectin cDNA, and pLG4.10 is.

The six possible amino acid sequences of pLG2.73 were derived from the nucleic acid sequence (Fig.30), and some possible polyadenylation signals marked. There is one open reading frame and one frame with a stop codon near the end. No homology was seen between these protein sequences and the protein sequences of the pea major albumins, pea soluble seed lectin or main legumin. The DNA sequence of pLG2.73 is compared with itself and with its complementary sequence (Fig.31), by dot matrix program at an equivalent stringency of 80%, a repeat sequence can be seen. The identity of this cDNA remains unknown.

The sequenced region of pLG4.10 and the derived protein sequence is compared to that of pea lectin message in Fig.32. As previously noted, the sequences are identical except for the start of the poly(A) tail which is 24 bases further downstream in pLG4.10. The consensus polyadenylation signal in plants as in animals is AATAAA except in the B49 sub-family of zein (a maize storage protein) where it is AATAAT (Messing et.al., 1983), but whereas in animals the signal occurs 11-30 bases upstream of the start of the poly(A) tail (Lewin, 1983), the position of the signal in plants varies and there are often multiple polyadenylation signals.

All leghaemoglobin, most zein and alcohol dehydrogenase

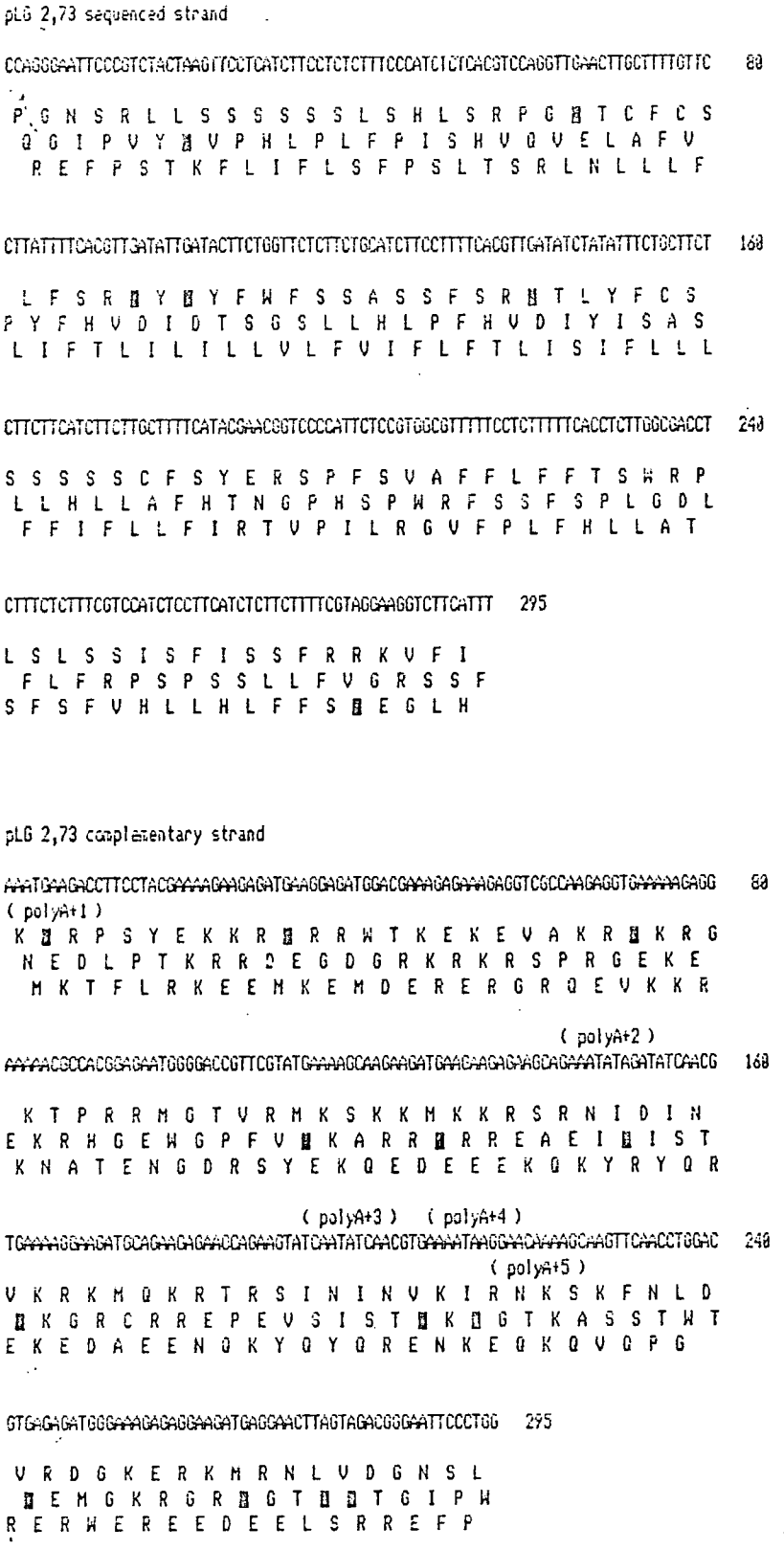


Fig.30. Possible amino acid sequences for pLG 2,73

Sequences close to the consensus polyadenylation signal are marked polyA+, the black symbol stands for a stop codon.

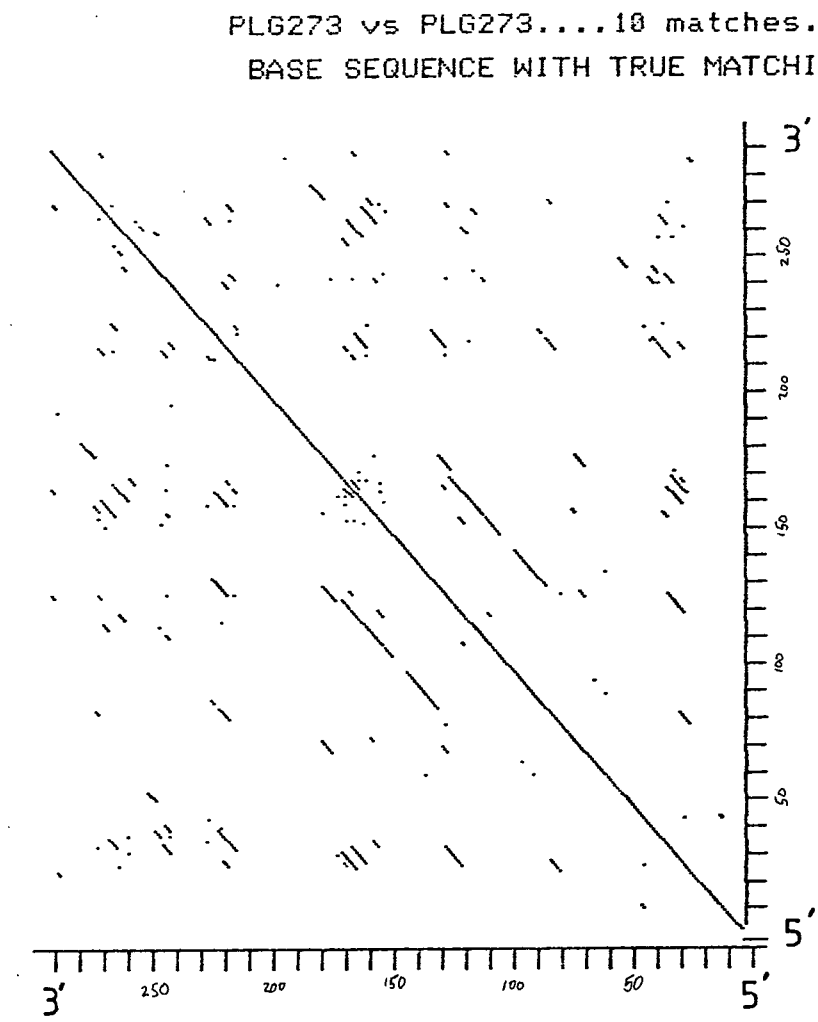


Fig. 31. Dot matrix comparison of pLG 2,73 with itself.

pLG 2,73 sequenced strand v's pLG 2,73 sequenced strand. Scale bp.

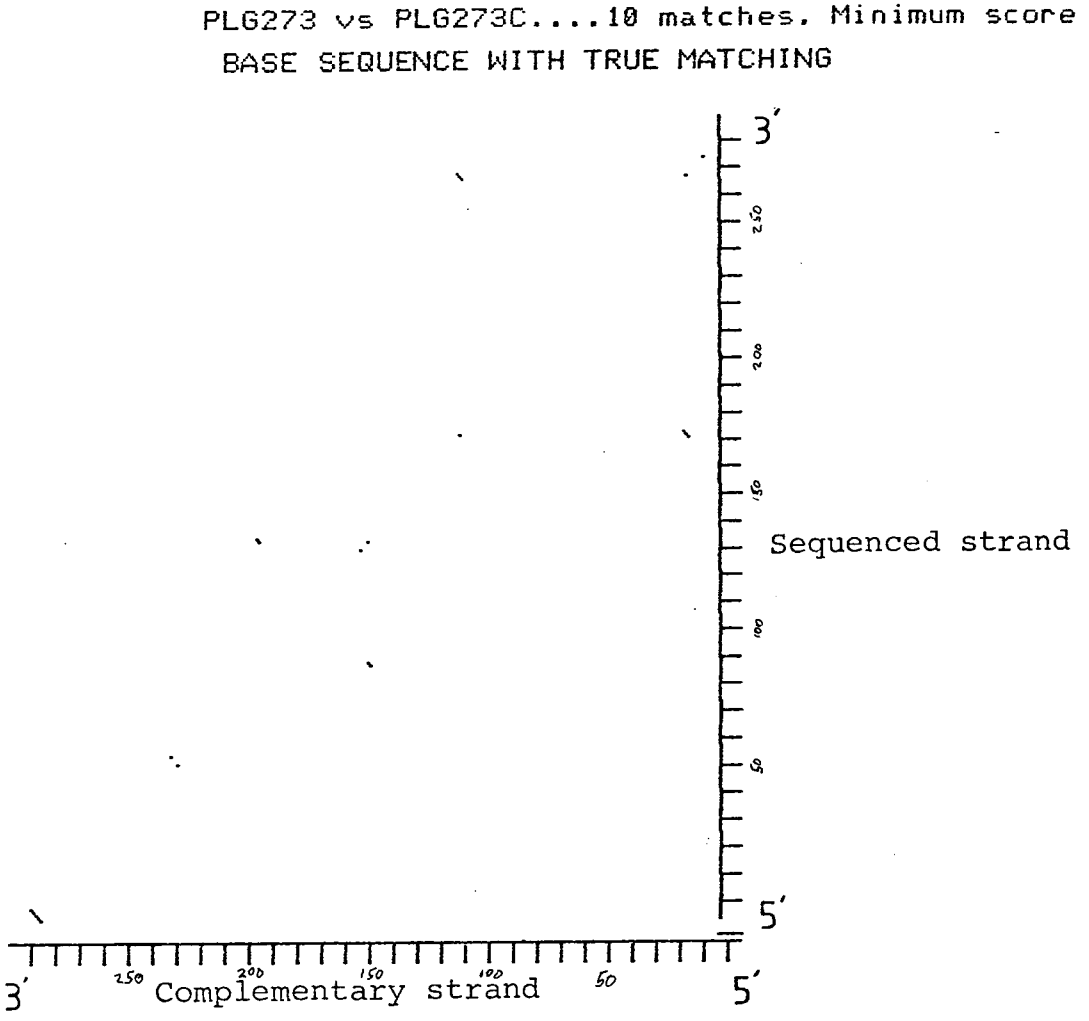


Fig. 31. (Continued)

pLG 2,73 sequenced strand v's pLG 2,73 complementary strand. Scale bp.

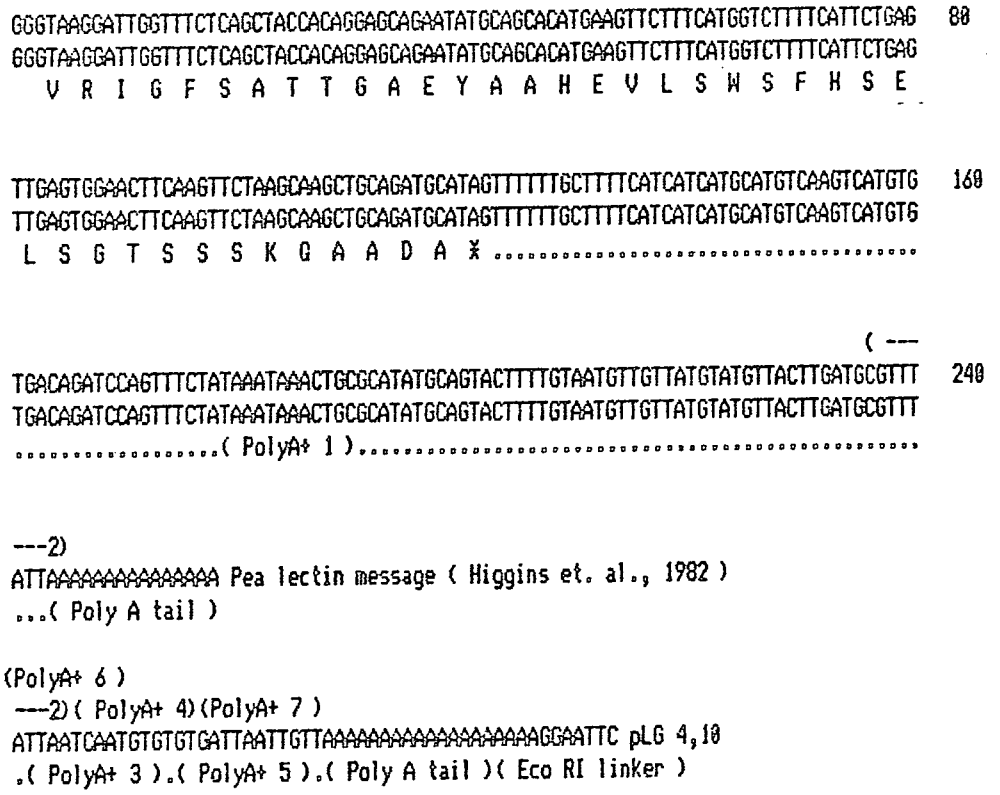


Fig.32. The 3' untranslated region of pLG 4.10 compared with that of Pea lectin message.

Pea lectin message : Higgins et.al., (1982)
PolyA+ = Sequence homologous to the consensus polyadenylation signal. (------2) see text.

genes have an AATAAA polyadenylation signal closer to the stop codon than the poly(A) tail region. Several have two polyadenylation signals; leghaemoglobin genes have a GATAAA signal and alcohol dehydrogenase genes a AATGAG signal close to the polyadenylation region (Messing et.al., 1983). However, legumin has a AATAAG signal near the stop (actually an overlapping double signal AATAAATAAAA) codon and a consensus AATAAA signal near the polyadenylation region. More divergent are the wild type octopine synthase genes of Ti plasmids which have AATAAT and AATAATATA before its polyadenylation sites (Froman et.al., 1984).

The published pea lectin sequence has only one polyadenylation signal, marked (polyA+1), the consensus AATAAA 78 bases downstream of the stop codon and 56 bases upstream of the polyA tail. The pLG4,10 sequence shares this and a further five possible polyadenylation signals, marked (polyA+ 3-7) : 3) AATCAA is 19 bases upstream of the polyA tail, well within the consensus distance, 4) AATGTG is similar to the AATGAG of alcohol dehydrogenase genes, and similarly it is close (14 bases) to the polyA tail, 5) GATTAA is similar to the GATAAA of leghaemoglobin and is very close (5 bases) to the polyA tail, 6) ATTAAT is similar to the wild type octopine synthase genes and at 21 bases upstream from the polyA tail within the animal consensus distance, 7) ATTAAT which is also similar to the wild type octopine synthase genes but is very close (4 bases) to the polyA tail. Of the above possible polyadenylation signals only the GATAAA second signal of leghaemoglobin occurs in the coding sequence of pea lectin message. Interestingly, in both sequences a complementary

polyadenylation signal TTTATT, marked (------2), occurs at the start of the poly(A) tail in the published sequence and 24 bases upstream in pLG4.10.

Whether the polyA+1 signal was the signal used in pLG4.10, where the polyA tail was initiated 102 bases downstream from it, or one of the cluster of putative polyadenylation signals was used, it seems significant that the poly A tail initiates in an AT-rich region, bearing many sequences close to the consensus polyadenylation signal. Multiple overlapping polyadenylation signals are known to occur in plant genes (Croy et.al., 1982; Schuler et.al., 1982; Edens et.al., 1982; Lycett et.al., 1984), in particular in pea legumin, but not in pea vicilin. However, for the legumin cDNAs with polyA tails investigated (Lycett et.al., 1984) all the tails started 19 or 20 bases downstream of the second polyadenylation signal. It is possible that the published lectin message has had the extra 24 bases, which occur in pLG4.10, deleted in some way. This would explain why the polyadenylation signal is so far from the polyadenylation site compared with the consensus distance. This problem should be resolved by sequencing other lectin cDNAs and by isolating and sequencing the lectin gene.

4.6. Albumin cDNAs.

The probe used to find major albumins positives (Sections 3.2.7 and 3.3.7) was only available in a very small amount, so only filter number four was probed in a colony hybridisation screen. The background on the screen was quite high, but a large number of possible

positives were found (Table 25). Nine of these were selected for further investigation, including the two colonies showing as fairly definite positives (Table 31). No strong correlation was seen between the insert size and strength of hybridisation to either the major albumins probe or mRNA. This suggested that some, atleast, of the positives were spurious.

(pLG 4.135 has two inserts)

R.A.Ragab showed, that the large insert of pLG4,135, one of the two fairly definite positives, in hybrid release translation experiments did select a major albumins message and he used it to select two other major albumin cDNAs from another cDNA library. The large insert of pLG4,135 was restriction mapped (Fig.15) and this map was somewhat different to the maps of the other major albumin positives isolated by R.A.Ragab (personal communication). He also found that the large insert of pLG4,135 did not cross hybridise to any of the other eight possible albumins positives investigated in detail. Thus suggesting that the major albumin sequence represents only 1 in 143, ∿0.7% of the library or ∿2.9% of the 14 day "abundant" sequences of Morton et.al., (1983), when probing at~91% homology.

This estimate is undoubtedly low as the major albumins represent 8-10% of the total seed protein (Croy et.al., 1984), double the abundance of the soluble seed lectin. Had the large insert of pLG4,135 been used to probe the library, a more representative number of positives would have doubtless been detected. Also, it is known (Croy et.al.,(1984) that the two major albumins are similar, as shown by their cyanogen bromide cleavage patterns.

However, the degree of DNA sequence homology between them is unknown and hence, lower stringency screening may find both major albumins, whereas at higher stringency only one may be detected.

It is suspected that pLG4,135 large insert codes for the large major albumin (PMA-L) as the polypeptide produced in hybrid release translation is closer in molecular weight to PMA-L (subunit Mr∿25,000) than PMA-S (subunit Mr∿24,000) (R.A.Ragab, personal communication).

The major albumins are produced without signal peptides or leader sequences (R.A.Ragab), although molecular weight estimations from protein gels are usually insufficiently accurate to decide unequivocally between proteins of such similar molecular weights.

The combined DNA and amino acid sequence of pLG4,135 large insert (Fig.16) shows a good match with the partial protein sequences obtained from cyanogen bromide and tryptic cleavage fragments (J.Gilroy and J.Gatehouse, personal communication). The discrepancies are accounted for by errors in the protein sequences and by the protein sequence being derived from total major albumins and the cDNA coding only for one of them.

The cyanogen bromide cleavage patterns observed for the two major albumins; fragments of ∿Mr11,500, 8,400 and 7,200 for PMA-L and∿Mr11,500, 7,600 and 6,600 for PMA-S (Croy et.al., 1984), can be fitted to the predicted amino acid sequence tolerably well. Cyanogen bromide cleaves polypeptides at the methionine (M) residues and sometimes at tryptophan (W) residues. The cyanogen bromide cleavages expected from the predicted amino acid sequence would produce

fragments: 1) greater than 85 amino acid residues corresponding to the Mr∿11,500 fragment; 2) 30 amino acid residues which seems too short, but must correspond to the Mr∿7,200 or Mr∿6,600 fragments; 3) 58 amino acid residues corresponding to the Mr∿8,400 or Mr∿7,600 fragments; 4) 9 amino acid residues and 5) 11 amino acid residues,both of which would not be detected.

The 3' untranslated sequence does not include the polyA tail but does have 3 consensus polyadenylation signals at 36, 120 and 202 bases downstream from the stop codon. The first and third of these can be interpreted as multiple overlapping signals such as occur in the legumin gene (Lycett et.al., 1984; Croy et.al., 1982), and there is a number of other sequences which are close to the consensus polyadenylation signal. The final polyadenylation signal is 17 bases upstream from the end of the cDNA which is within the consensus distance from signal to polyadenylation region. However, without a cDNA having a poly(A) tail, this is a speculation.

Thus the large insert of pLG4,135 is a major albumin cDNA and the other colonies selected in the colony screen do not carry cDNAs coding for the major albumins.

Colony 4,92 was the other reasonably definite major albumins positive. However, as it does not hybridise to pLG4,135 large insert it is not a major albumins cDNA. It's insert was 1070 bp in length and was used to probe a Northern blot of pea cotyledon mRNA from various stages of development. The intention had been to size the major albumins mRNAs . The results (Table 32, Fig.20) show the insert of pLG4,92 hybridises to three different sizes of mRNA which occur at differing stages of development. In addition

two further bands show up in the 22day result at ∿1450 and ∿1700 bases in length but these may be due to unspecific hybridisation.

The strongest hybridisation occurs with the 2750 base band and thus, this is probably the mRNA pLG4,92 codes for. The weak smeared signal for this band for the 16 days after flowering mRNA appears to be due to a defect in the blot. A mRNA of 2750 bases in length should code for a protein of Mr∿69,000. This polypeptide is only produced in the later stages of development. The other two bands, 2250 bases and 1300 bases in length, must represent mRNAs from related genes and should code for proteins of Mr∿56,000 and Mr∿33,000 respectively.

This set of related messages show strong developmental control with the 1300 base message gene being switched off as the 2250 base message gene is switched on at the start of the mid-stage of development and the 2750 base message gene being switched on at the start of late stage development. Alternativly this pattern may be due to changes in transcription of one gene or changes in post transcriptional processing of the mRNA. As the library was made from 13-17 day poly(A)+RNA the smeared 16 day result for the 2750 base band most probably should be positive.

The suggested protein molecular weight of the 2750 base message of about 79,000 raise the speculation that the cDNA may code for convicilin Mr 75,000 or the Mr 80,000 legumin related polypeptide. Both of these possibilities would be expected to show as mRNA positive in colony screens as does pLG4,92. Why pLG4,92 was a fairly definite positive in the colony screen with the major albumins enriched probe is unexplained by these hypothesises.

4.7. Terminal Deoxynucleotidyl Transferase.

Much time and effort was expended in trying to find conditions under which terminal transferase (TdT) added homopolymer tails, to the 3' ends of DNA molecules reliably and consistently and in trying to trace and eliminate all the factors which inhibit the action of TdT. Workable results could be attained for dG and dC tails by taking great care : 1) to purify the DNA, eliminating all organics such as phenol or impurities in the ethanol used for alcohol precipitation, 2) by using deionized and double distilled (HPLC grade) water, 3) by using commercially supplied reaction buffer, 4) by using nuclease free BSA in the reaction mix and 5) by using the other precautions listed in section 2.2.3.8. The best conditions were however never fully reliable and the results of Roychoudhury et.al., (1976) and Deng and Wu (1981) proved unrepeatable despite the results in the assays of the TdT suppliers (BRL and PL) being reproduced reasonably consistently.

In particular the cDNA cloning scheme of Okayama and Berg (1981) requires a plasmid fragment to be tailed with tails of ∿60 bases in length of dT residues (see Fig.5). The total failure to produce samples of this plasmid fragment bearing tails longer than 1-2 bases was the reason for its abandonment as a method of producing

a cDNA library despite its other major advantages.

From personal communications, various mentions in the literature and commercial data sheets, it seems that problems of reproducibility from one laboratory to another are very common with TdT. However, the greater ease of tailing and shorter tails required with dG and dC compared with dT and the recent commercial availability of ready tailed Okayama and Berg plasmid fragments make this scheme for cDNA cloning very attractive.

4.8. Summary.

A cDNA library has been constructed and cDNAs for various seed proteins isolated; those for legumins, vicilins, lectin and major albumins have been characterised. The relative abundancies of these cDNAs has been correlated with the sequence complexity data of Morton et.al., (1983) and shown to fit them well. The total number of main legumin, big legumin and vicilins positive clones represent over 75% of the "very abundant" sequences, and probes with convicilin, with the Mr$\sim$80,000 legumin related polypeptide and with the probes already used at lower stringencies will undoubtedly find the remaining 25%. Thus the library corresponds statistically and is a representation of the population of mRNAs in the pea cotyledons at the time they were harvested.

REFERENCES

Badenoch-Jones, J. Spencer, D. Higgins, T.J.V and Millerd, A. (1981) Planta 153 201-209

Barton, K.A., and Hill, W.J. (1983) Science 219,671-676.

Basha, S.M.M. and Beevers, L. (1976) Plant Physiol. 57 93-97

Bedbrook, J.R., and Bogorad, L. (1976) Proc. Natl. Acad. Sci. USA 73 4309-4313.

Berger, S.L., and Berkenmeier, C.S. (1979) Biochem. 18 5143 - 5149

Blobel, G., and Dobberstein, B. (1975) J. Cell Biol. 67 835- 851

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betloch, M.V., Heynecker, H.L., Boyer, M.W., Crosa, J.H., and Falkow, S. (1977) Gene 2 95-113.

Bonner, T.I., Brenner, D.J., Neufeld, B.R., and Britten, R.J., (1973) J.Mol.Biol. 81 123- 135

Boulter, D. Ellis, R.J., and Yarwood, A. (1972) Biol. Rev. 47 113-175.

Boulter, D. (1981), Adv. Bot. Res. 9 1-31.

Bown,D., Levasseur, M., Croy, R.R.D., Boulter, D., and Gatehouse, J.A. (1985), Nucl. Acids Res. 13 4527-4538.

Broglie, R., Coruzzi, G. Fraley, R.T., Rogers, S.G., Horsch, R.B., Niedermeyer, J.G., Fink, C.L., Flick, J.S., and Chua N.U. (1984) Science 224 838-843.

Brown, J.W.S., Ersland, D.R., and Hall, T.C. (1982) in "The Physiology and Biochemistry of Seed Development, Dormancy and Germination" Elsevier Biomedical Press, p.3-42.

Buck, K.W., and Coults, R.H.A. (1983) Plant Mol. Biol. 2. 57-71.

Casey, J., and Davidson, N. (1977) Nucl. Acids Res. 4 1539- 1545

Casey, R. (1979) Heredity 43 265- 272

Casey, R. (1979) Biochem. J. 177 509- 520

Casey, R., and Sanger, E. (1980) Biochem. Soc. Trans. 8 657-658.

Casey, R., and Short, M.N. (1981) Phytochem. 20 21-23.

References Contd/...

Casey, R., March, J.F., Sharman, J.E. and Short, M.N. (1981a) Biochem. Biophys. Acta 670 428-432.

Casey, R., March, J.F., and Sanger, E. (1981b) Phytochem. 20 161-163.

Casey, R., Domoney, C., and Stanley, J. (1984) Biochem. J. 224 661-666

Chandler, P.M. (1982). Analyt. Biochem. 127 9-16.

Chrispeels, M.J., Higgins, T.J.V., Craig, S. and Spencer,D. (1982) J. Cell Biol. 93 5-14

Covey, S.N. and Hull, R. (1985) Oxford Surveys of Plant Molecular and Cell Biol. 2 339-346.

Croy, R.R.D., Derbyshire, I.E., Krishna, T.G. and Boulter, D., (1979) New Phytol. 83, 29-35.

Croy, R.R.D., Gatehouse, J.A., Evans, I.M. and Boulter, D. (1980a) Planta 148 49-56.

Croy, R.R.D., Gatehouse, J.A., Tyler, M. and Boulter, D., (1980b) Biochem. J. 191 509-516.

Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N. and Boulter, D. (1982) Nature 295 76-79.

Croy, R.R.D., Hoque, M.S., Gatehouse, J.A. and Boulter, D. (1984) Biochem. J. 218 795-803.

Croy, R.R.D., and Gatehouse, J.A. (1985) in Plant Genetic Engineering (Dodds, J.H. ex) Cambridge University Press. Chapter 8 143-168

Dagert, M. and Ehrlich, S.D. (1979) Gene 6 23-28.

Danielsson, C.E. (1949) Biochem. J. 44 387-400.

Davey, R.A. and Dudman, W.F. (1979) Aust. J. Plant Physiol. 6 435-447

Davies, D.R. (1976) Euphytica 25 717-724.

Davies, R.W. (1982) In "Gel Electrophoresis of Nucleic Acids - A Practical Approach" IRL Press Ltd., Oxford, 117-172.

De Block, M., Herrera-Estrella, L., Vanmontague, M., Schell, I. and Zambryski, P. (1984) EMBO J. 3 1681-1689.

Delauney, A.J. (1984) PhD Thesis, Dept. of Botany, University of Durham.

de Martynoff, G., Pays, E. and Vassart, G. (1980) Biochem. Biophys. Res. Commun. 93 645-653

References Contd/...

Deng, G. and Wu, R. (1981) Nucl. Acids Res. 9 4173-4188.

Derbyshire, E., Wright, D.J. and Boulter, D. (1976) Phytochem 15 3-24.

Domoney, C. and Casey, R. (1983) Planta 159, 446-453

Domoney, C. and Casey, R. (1984) Eur. J. Biochem. 139 321-327.

Domoney, C. and Casey, R. (1985) Nucl. Acids Res. 13 687-699.

Dove, W.F. and Davidson, N. (1962) J. Mol. Biol. 5 467-478

Dretzen, G., Bellart, M., Sassone-Corsi, P. and Chambon, P., (1981) Anal. Biochem. 112 295-298

Dudman, W.F. and Millerd, A. (1975) Biochem. Syst. Ecol. 3 25-36

Edens, L., Heslinga, L., Klok, R., Ledeboer, A.M., Maat, J., Tooner, M.Y., Visser, C. and Verrips, C.T., (1982) Gene 18 1-12.

Efstratiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T. (1976) Cell 7 279-291

Etzler, M.E., McMillan, S., Scates, S., Gibson, D.M., James, D.W. Jr., Cole, D. and Thayer, S. (1984) Plant Physiol 76 871-878.

Evans, I.M., Croy, R.R.D., Brown, P. and Boulter, D. (1980) Biochem. Biophys. Acta 610 81-95.

Evans, I.M., Gatehouse, J.A., Croy, R.R.D. and Boulter,D. (1984) Planta. 160 559-568.

Fagan, J.B., Pastan, I. and deCrombrugghe, B. (1980) Nucl. Acids Res. 8 3055-3064.

Federoff, N. (1983) Controlling elements in maize. In "Mobile Genetic Elements" (Shapiro, J. ed.) Academic Press, New York, 1-63.

Flavell, R., and Mathias, R. (1984) Nature 307 108-109.

Froman, B.F., Tait, R.C. and Rodriguez, R.L. (1984) Gene 31 257-261.

Gatehouse, J.A., Croy, R.R.D., Morton, H., Tyler, M. and Boulter, D. (1981) Eur. J. Biochem. 118, 627-633.

Gatehouse, J.A., Lycett, G.W., Croy, R.R.D. and Boulter, D. (1982) Biochem. J. 207 629-632.

Gatehouse, J.A., Croy, R.R.D., McIntosh, R., Paul, C. and Boulter, D. (1980) 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., ECSC, EEC, EAEC, Brussels, 173.

References Contd/...

Gatehouse, J.A., Lycett, G.W., Delauney, A.J., Croy, R.R.D. and Boulter, D. (1983) Biochem. J. 212 427-432.

Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) CRC Critical Reviews in Plant Science Vol.1. Issue 4. 287-314.

Geraghty, D.E., Messing, J. and Rubenstein, I. (1982). EMBO J. 1 1329-1335.

Goldberg, R.B., Hoschek, G., Tarn, S.H., Ditta, G.S. and Breidenbach, R. (1981) Dev. Biol. 83 201-218.

Gonzales, N., Wiggs, J., and Chamberlin, M.J. (1977) Arch. Biochem. Biophys. 182 404-408.

Gopinathan, K.P., Weymouth, L.A., Kinkel, T.A. and Loeb, L.A. (1979) Nature 278 857-859.

Grant, D.R. and Lawrence, J.M. (1964) Arch. Biochem. Biophys. 108 552-559

Grunstein, M. and Hogness, D. (1975) Proc. Natl. Acad. Sci. 72 3961-3965

Grunstein, M. and Wallace, J. (1978) Methods in Enzymology 68 379-389.

Guldager, P. (1978) Theor. Appl. Genet. 53 241-247.

Hall, T.C., Ma, Y., Buchbinder, B.U., Pyne, J.W., Sun, S.M. and Bliss, F.A. (1978) Proc. Natl. Acad. Sci. USA 75 3196-3200.

Hanahan, D. and Meselson, M. (1980) Gene 10 63-77

Higgins, T.J.V., and Spencer, D. (1980) Plant Physiol. 67 205-211

Higgins, T.J.V., Chandler, P.M. Zurawski, G., Button, S.C. and Spencer, D. (1982) J. Biol. Chem. 258 9544-9549.

Higgins, T.J.V., Chrispeels, M.J., Chandler, P.M. and Spencer, D. (1982) J. Biol. Chem. 258 9550-9552.

Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C. and Spencer, D. (1983) J. Biol. Chem. 258 9544-9549.

Higgins, T.J.V., (1984) Ann. Rev. Plant Physiol. 35 191-221.

Hoffman, L., Ma, Y. and Barker, R.F. (1982) Nucl.Acids Res. 10 7819-7828.

Horsch, R.B., Fraley, R.T., Rogers, S.G., Sanders, P.R., Lloyd, A. and Hoffman, N. (1984) Science 223, 496-498.

References Contd/...

Jaffe, W.G. (1969) In "Toxic constituents of Plant Foodstuff" (I.E. Liener ed.) 69-101. Academic Press, New York.

Kroeker, W.D., Kowalski, D. and Laskowski, M. (1976) Biochem. 15, 4463-4467.

Kushner, S.R. (1978) In: Genetic Engineering (eds. H.B.Boyer and S.Nicosia) p.17 Elsevier/North Holland, Amsterdam. Chapter 1

Land, H.M., Grez, H., Hansen, H., Lindermaier, W. and Schuetz, G. (1981) Nucl. Acids Res. 9 2251-2266.

Larkins, B.A. (1981) The Biochemistry of Plants Vol.6: Chapter 11. 309 - 341

Larkins, B.A. (1983) in Genetic Engineering of Plants (Hollaender, A., Kosuge, T. and Meredith, C.P. eds.) Plenum Press, New York, Pages 93 - 118

Lewis, B. (1983) Genes, Wiley and Sons Inc.

Lis, H. and Sharon, N. (1981) The Biochemistry of Plants Vol.6 Chapter 10. 263 - 308

Lycett, G.W., Delauney, A.J., Gatehouse, J.A., Gilroy, J. Croy, R.R.D. and Boulter, D. (1983) Nucl. Acids Res. 11 2367-2380.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) Nucl. Acids. Res. 12 4493-4506.

Lycett, G.W., Croy, R.R.D., Shirsat, A.H., Richards, D.M. and Boulter, D. (1985) Nucl. Acids Res. 13 6733-6743.

McConaughty, B.L., Laird, C.D. and McCarthy, B.T. (1969) Biochemistry 8 3289-3295

McDonnell, M.W., Simon, M.N. and Studier, F.W. (1977) J.Mol. Biol. 110 119-146

Mandel, M. and Higa, A (1970) J.Mol. Biol. 53 154-159

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) "Molecular Cloning-A Laboratory Manual" (Cold Spring Harbour Laboratory, New York).

Marmur, J. and Doty, P. (1962) J. Mol. Biol. 5 109-118

Matta, N.K., Gatehouse, J.A. and Boulter, D. (1981) J. Expt. Bot. 32 1295-1307.

Matta, N.K. and Gatehouse, J.A. (1981) Phytochem. 20 2621 - 2623

Matta, N.K. and Gatehouse, J.A. (1982) Heredity 48 383-392

Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology 65 499-513

Messing, J., Geraghty, D., Heidecker, G., Hu, N-T., Kridl, J. and Rubenstein, I. (1983) in "Genetic Engineering in Plants" Plenum Press, New York, chapter 4 211-227.

Morton, H., Evans, I.M., Gatehouse, J.A. and Boulter, D. (1983) Phytochemistry 22 807-812.

Murai, N., Sutton, D.W., Murray, M.G., Slightom, J.L., Merlo, D.J., Reichert, N.A., Singupta-Gopalan, C., Stock, C.A., Barker, R.F., Kemp. T.D. and Hall, T.C. (1983) Science 222, 476-482.

Murray, M.G., Hoffman, L.M. and Jarvis, N.P. (1983) Plant Mol. Biol. 2 74-84.

Neiden, U.Z. and Neumann, D. (1982) Eur. J. Cell. Biol. 26 228-240

Okayama, H. and Berg, P. (1981) Mol. Cell Biol. 2 161-170.

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

Oram, R.N. and Brock R.D. (1972) J. Aust. Inst. Agric. Sci. 38 163-168.

Osborn, T.B. and Campbell, G.R. (1898) J. Am. Chem. Soc. 29 348-356

Rasmussen, S.K. Hopp, H.E., and Brandt, A. (1983) Carlsberg Res.Commun. 48 187-199.

Roychoudhury, R., Jay, E. and Wu, R. (1976) Nucl. Acids Res. 3 101-116.

Scheel, G. and Blackburn, P. (1979) Proc. Natl. Acad. Sci. USA 76 4898-4902

Schroeder, H.E. (1982) J. Sci. Food Agric. 33 623-633.

Schuler, M.A., Schmitt, E.S. and Beachy, R.N. (1982). Nucl. Acids Res. 10 8225-8244.

Schuler, M.A., Ladin, B.F., Polacco, J.C., Freyer, G. and Beachy, R.N. (1982) Nucl. Acids Res. 10 8245-8261.

Shewry, P.R., Miflin, B.J. Forde, B.G. and Bright, S.W.J. (1981) Sci. Prog. Oxf. 67 575-600.

Shirsat, A.H. (1984) PhD Thesis Dept. of Botany, Durham University.

Sief, I., Khoury,G. and Dhar, R. (1980) Nucl. Acids Res. 8 2225-2239.

References Contd/....

Smith, H.D. (1980).,Methods in Enzymology 65 371-392

Sorenson, J.C. (1984) Adv. Genet. 22 109-144.

Southern, E. (1975) J. Mol. Biol. 98 503-517

Spencer, D., Higgins, T.J.V., Button, S.C. and Davey, R.A. (1980) Plant Physiol. 66 510-521

Spencer, D., Chandler, P.M., Higgins, T.V.J., Inglis, A.S. and Rubira, M. (1984) Plant. Mol. Biol. 2 (5) 259-268.

Staden, R (1982) Nucl. Acids Res. 10 2950-2961.

Sumner, J.B. and Howell, S.F. (1937) J. Biol. Chem. 115 583-588.

Sutcliffe, J.G., (1978) Nucl. Acids Res. 5 2721-2727

Sutcliffe, J.G. (1979) Cold Spring Harbour Symp. Quant. Biol. 43 77-90.

Thomson, J.A., Schroeder, H.E. and Dudman, W.F. (1978) Aust. J. Plant Physiol. 5 263-274

Thomson, J.A., Schroeder, H.E. and Tassie, A.M. (1980) Aust. J. Plant Physiol.7 271-282

Trowbridge, I.S. (1974) J. Biol. Chem. 249 6004-6012.

Tyler, M. (1981) PhD Thesis, Dept. of Botany, University of Durham.

Van Driessche, E., Smets, G., Dejaegere, R and Kanarek, L. (1981) Planta 153 287-296.

Vieira, J. and Messing, J. (1982) Gene 19 259-273.

Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) Cell 34 1023-1031.

Volckaert, G., Tavernier, J., Derynck, R., Devos, R. and Fiers, W. (1981) Gene 15 215-233.

Weaver, C.A., Gordon, D.F. and Kemper, B. (1981) Proc. Natl. Acad. Sci. USA 78 4073-4077.

Wendorf, F., Schild, R., El Hadidi, N., Close, A.E., Kobusiewicz, M., Wieckowska, H., Issawi, B. and Haas, H. (1979) Science 205 1341-1347.

Wood, W.B. and Lee, J.C. (1976) Nucl. Acids Res. 3 1961-1968

Wright, D.J. and Boulter, D. (1974) Biochem. J. 141 413-418.

References Contd/...

Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979) Cell 16 851-861

TABLE 38

cDNA Library Colony Hybridisation Screen Data Table.

Column

A	Colony number, pLG number of plasmid.
B	Noticeable growth on original master filter.
C	Blue colour when grown on plates containing X-gal. Positive colonies have no inserts or are mixed.
D	Approximate size of the insert(s) in base pairs.
E) F)	14 day mRNA at ∿94% homology. (Section 3.2.1).
G	Main legumin cDNA pAD4.4 (pDUB6) at ∿94% homology. (Section 3.2.2).
H	Main legumin gene C pAS2 (pDUB26) at ∿99% homology (Section 3.2.2). The results are not complete.
I	Main legumin gene A pRC3.1 (pDUB21) at ∿81% homology (Section 3.2.2). The results are not complete.
J	Big legumin gene J pJC5-2 at ∿91% homology (Section 3.2.3).
K	Vicilin B cDNA pAD3.4 (pDUB7) at ∿94% homology. (Section 3.2.4.1).
L	Vicilin B cDNA pAD3.4 (pDUB7) at ∿81% homology. (Section 3.2.4.1).
M	Vicilin B cDNA pAD3.4 (pDUB7) at ∿74% homology. (Section 3.2.4.1).
N	Vicilin C cDNA pAD2.1 (pDUB9) at ∿94% homology. (Section 3.2.4.2).
O	Vicilin C cDNA pAD2.1 (pDUB9) at ∿81% homology. (Section 3.2.4.2).
P	Vicilin C cDNA pAD2.1 (pDUB9) at ∿72% homology. (Section 3.2.4.2).
Q	Vicilin A cDNA pRC2,2,1 (pDUB2) at ∿94% homology. (Section 3.2.4.3).
R	Phaseolus vulgaris lectin cDNA pPLV134 at ∿82% homology (Section 3.2.6).
S	Pea lectin cDNA pRC2,2,26 at ∿94% homology. (Section 3.2.5).
T	cDNA enriched in sequences coding for the major albumins at ∿91% homology (Section 3.2.7).

Symbols used are : +++ = Strong, ++ = medium, + = weak, o+ = very weak, o = not detectable, ? = dubious i.e. a signal but probably spurious, ∿ = a very weak signal which seems non specific, - = this colony did not appear to have grown or results from this colony should be ignored, blank = this colony was not screened with this probe.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1,1	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,2	+	0	300	0	0+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,3	+	0+		0	0+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,4	+	++	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	
1,5	+	++		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,6	+	0+	540	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,7	+	+	450	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,8	+	++		0	0	0	0	0	0	0	0	0	~	0	0	?	0	0	0	
1,9	+	0+	1050	+	+	++	+	+	0	0	0	?	0	0	0	0	0	0	0	
1,10	+	++		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,11	+	0	1180	+	++	0	0	0	0	0	++	+	0+	+	+	+	0	0		
1,12	+	0+		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,13	+	+	690	0	0+	0	0	0	0	0	0	0	0	~	?	+	0	0	0	
1,14	+	+	450	0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,15	+	0+		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,16	+	0	960	++	+++	0	0	0	0	+++	+++	+++	+++	+++	+++	+++	0	0		
1,17	+	0+	440	0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,18	+	0	600	0	+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,19	+	0	580	0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,20	+	0	330	+	+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,21	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,22	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,23	+	0		+	+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,24	+	0		+	0	0	0	0	0	0	0	0+	0+	0	0	0+	0	0	0	
1,25	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,26	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,27	+	0		0	0+	0	0	0	+++	0	0	~	0	0	0	0	0	0		
1,28	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0		
1,29	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0		

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1,30	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	.
1,31	+	0		0	0	0	0	0	0	0	0	~	0	0	0+	0	0	0	
1,32	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,33	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,34	+	0+		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,35	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,36	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,37	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,38	+	0		0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,39	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,40	+	0+	470	++	++	+++	+++	+++	0	0	0	~	~	0	0	0	0	0	
1,41	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,42	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,43	+	0		0	0+	0	?	0	0	0	0	0	0	0	0	++	0	0	
1,44	+	0		+	0+	0	0	0	+++	0	0	~	0	0	0	0	0	0	
1,45	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,46	+	0	600	0	++	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,47	+	0+		0	+	0	0	0	0	0	0	~	0	0	0+	0	0	0	
1,48	+	0		0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,49	+	0		+	+	0	0	0	+	0	0	0	0	0	0	0	0	0	
1,50	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,51	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,52	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,53	−						0	0											
1,54	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,55	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,56	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,57	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,58	+	0		0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1,59	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,60	+	0	840	++	++	0	0	0	0	0	0	0	~	0	?	0+	0	0	0	-
1,61	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,62	+	0		0	0+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,63	+	0	1530	++	+++	0	0	0	0	+++	+++	+++	++	++	++	++	0	0		
1,64	+	0		0	+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,65	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,66	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,67	+	0		0	0+	+	+	+	0	0	0	~	0	0	0	0	0	0	0	
1,68	+	0		0	0	+	0	+	0	0	0	0	0	0	0	0	0	0	0	
1,69	+	0		0	0+	0	0	+	0	0	0	0	0	0	0	0	0	0	0	
1,70	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,71	+	0		0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	
1,72	+	0	710	0	++	0	0	0	0	0	0	+	?	0	0	0	0	0	0	
1,73	+	0		0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	
1,74	-	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,75	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,76	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,77	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,78	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,79	+	0		0	+	0	0	0	0	0	+	~	0	0	0	0	0	0		
1,80	+	+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,81	+	0		0	0	0	0	0	0	0	0	0+	~	0	0	0	0	0	0	
1,82	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,83	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,84	+	0+		0	0+	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,85	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,86	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,87	+	0	1090	+	++	+	++	+++	0	0	0	~	0	0	0	0	0	0		

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1,88	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,89	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,90	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,91	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,92	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,93	+	0		+	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,94	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,95	+	0		0	+	+	++	+	0	0	0	~	0	0	0	0	0	0	
1,96	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,97	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,98	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,99	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,100	+	0		0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	
1,101	+	0+		0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	
1,102	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,103	+	0		+	0	0	0	0	0	0	0	?	0	?	+	0	0	0	
1,104	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,105	+	0		0	0	0	0	0	0	0	0	~	0	0	0	?	0	0	
1,106	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,107	+	0		0	+	0	0	0	0	0	0	~	0	0	0	+	0	0	
1,108	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,109	+	0		0	+	0	0	0	0	0	0	+	0	0	0	+	0	0	
1,110	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,111	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,112	+	0		0	+	+	+	+	0	0	0	~	0	0	0	0	0	0	
1,113	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,114	+	0		0	+	++	++	++	0	0	0	?	0	0	0	0	0	0	
1,115	+	0		0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	
1,116	+	0		+	+	+	+	+	0	0	0	~	0	0	0	0	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1,117	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,118	+	0		0	0	0	0	0	0	0	0	?	0	0	0	0	~	0	
1,119	+	0		0	0+	0	0	0	0	0	0	?	0	0	0	0	~	0	
1,120	+	0+		0	0	0	0	0	0	0	0	?	0	0	0	0	~	0	
1,121	+	0+		0	0	0	0	0	0	0	0	?	0	0	0	0	~	0	
1,122	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,123	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,124	+	+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,125	+	0		0	0+	0	0	0	0	0	0	+	0	0	?	0	0	0	
1,126	+	0+		0	0+	0	0	0	0	0	0	+	0	0	0	0	0	0	
1,127	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,128	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,129	+	0		0	0	0	0	0	0	0	0	+	0	0	?	0	0	0	
1,130	+	0+		+	+	++	0	0	0	0	0	~	0	0	0	0	0	0	
1,131	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,132	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,133	+	0		0	+	0	0	0	0	0	0	?	0	0	0	0	0	0	
1,134	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,135	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,136	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,137	+	+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,138	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,139	+	0	420	0	+	0	0	0	0	0	0	~	0	0	0	0	0	+++	
1,140	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
1,141	+	0+		0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,142	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,143	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,144	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,145	+	0		0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1,146	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,147	+	0			+	+	0	0	0	0	0	0	~	+++	+++	+++	++	0	0	
1,148	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,149	+	0			0	0+	+	0	+	0	0	0	0	0	0	0	0	0	0	
1,150	+	0			0	0	0	0	0	0	0	0	~	0	?	0+	0	0	0	
1,151	+	0			0	0	0	0	0	0	0	0	~	0	0	~	0	0	0	
1,152	+	0			++	+++	0	0	0	0	++	+++	+++	+++	+++	+++	+++	0	0	
1,153	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,154	+	0			+	++	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,155	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,156	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,157	+	0			+	+	0	0	0	0	0	0	~	+++	+++	+++	+++	0	0	
1,158	+	0			+	++	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,159	+	0			+	+	+	0+	+	0+	0	0	~	0	0	0	0	0	0	
1,160	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,161	+	0+			+	+	+	++	++	0	0	0	~	0	0	0	0	0	0	
1,162	+	0			+	++	0	0	0	0	+	++	++	++	+++	++	+++	0	0	
1,163	+	0			0	+	++	?	+	?	0	0	~	0	0	0	0	0	0	
1,164	+	0+			0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,165	+	0			0	0+	0	0	0	0	0	+	~	0	0	0	0	0	0	
1,166	+	0			0	+++	0	0	0	0	+++	+++	+++	++	+++	++	0	0	0	
1,167	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,168	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,169	+	0			0	+	+	+	++	0	0	0	~	0	0	~	0	0	0	
1,170	+	0			0	+	0	0	0	+++	0	0	~	0	0	0	0	0	0	
1,171	+	0			0	++	+	0	0	0	0	0	~	0	0	0	0	0	0	
1,172	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,173	+	0			0	+	+	0	0	0	0	0	~	0	0	0	0	0	0	
1,174	+	0			0	0+	0	0	0	0	0	+	+	0	0	~	+++	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1,175	+	0+		0	0+	?	0	0	0	0	+	+	0	0	0	0	0	0	
1,176	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
1,177	-			0	0														
1,178	-			0	0														
1,179	-			0	0														
2,1	+	0		0	+	0	0	0	0	0	0	+++	0	0	0	0	0	0	
2,2	+	0+		0	+++	0	0	0	0	++	+++	+++	+	++	++	+	0	0	
2,3	+	0		0	0	0	0	0	0	0	0	+++	0	0	0	0	0	0	
2,4	+	0		+	+++	0	0	0	+++	++	+++	+++	++	++	++	0	0	0	
2,5	+	0		0	?	0	0	0	0	?	?	~	0+	+	?	0	0	0	
2,6	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,7	+	0+		0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	
2,8	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,9	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,10	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,11	+	0+		0	+	0	0	0	0	0	0	0	~	?	?	0	0	0	
2,12	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,13	+	0		0	+	0	0	0	0	0	0	0	~	?	?	0	0	0	
2,14	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,15	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,16	+	+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,17	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,18	+	0		0	0+	0	0	0	0	0	0	0	0	0+	0+	0	0	0	
2,19	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,20	-	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,21	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,22	-	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,23	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,24	+	0		+	+++	0	0	0	0	+	++	++	+++	+++	+++	++	0	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2,25	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,26	+	0			0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	
2,27	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,28	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,29	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,30	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	~	0	
2,31	+	0			0	0	0	0	0	0	0	0	0	0	0+	0+	0	0	0	
2,32	+	0			0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	
2,33	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,34	+	0			+	+	0	0	0	0	0	0	0	0	0	~	0	0	0	
2,35	+	0			0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,36	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,37	+	0			0	0+	0	0	0	0	0	0	~	0	0+	0	0	0	0	
2,38	+	0			0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,39	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,40	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,41	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,42	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,43	+	0			0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,44	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,45	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,46	+	0+			0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,47	+	0		500 300	+	+	0	++	+++	0	0	0	0	0	0	0	0	+	0	
2,48	+	0			0	+	0	0	0	0	0	0	~	+	++	++	0	0	0	
2,49	+	0			0	0+	0	0	0	++	0	0	0	0+	+	+	0	0	0	
2,50	+	0			+	++	0	0	0	0	0	0	~	0	0	0	+++	0	0	
2,51	+	0			+	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,52	+	0			0	+	++	+	+	0	0	0	0	0	0	0	0	0	0	
2,53	+	0			0	0+	0	0	0	0	0	0	0	~	+	?	0	0	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2,54	+	0			0	0	0	0	0	0	0	0	0	+	?	0	0	0	0	
2,55	+	0			0	0	0	0	0	0	0	0	0	0+	?	0	0	0	0	
2,56	+	0		210	0	0	0	0	0	0	0	0	~	0	0	0	0	~	+	
2,57	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,58	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,59	−	0			0	+	0	0	0	0	0	0	0	~	0	?	0	0	0	
2,60	+	0			0	0	0	0	0	0	0	0+	?	0	0	0	0	0	0	
2,61	+	0			0	++	+++	0	0	0	0	0	0	0	0	0	0	0	0	
2,62	+	0			0	0	0	0	0	0	0	0	~	0+	+	+	0	0	0	
2,63	+	0			0	0+	0	0	0	0	0	0	0	+	++	++	0	0	0	
2,64	+	0			0	0+	0	0	0	0	0	0	~	+++	+++	+++	0	0	0	
2,65	+	0			0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,66	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,67	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,68	+	0			+	++	0	0	0	0	0	0+	~	0	0	0	0	0	0	
2,69	+	+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,70	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,71	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,72	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,73	+	0		800	0	+	0	0	0	0	0	0	0	0	0	0	0	0	+	
2,74	+	0			0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,75	+	0			0	+++	0	0	0	+++	+++	+++	+++	+	+	+	0	0	0	
2,76	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,77	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,78	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,79	+	0			0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	
2,80	+	0+			0	+	0	0	0	0	0	0	~	0	?	0	0	0	0	
2,81	+	0			0	+	0	0	0	++	0	0	0	0	0	0	0	0	0	
2,82	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2,83	+	0		0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,84	+	0		+	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,85	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,86	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,87	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,88	+	0		0	0	0	0	0	++	0	0	0	0+	0+	+	0	0	0	
2,89	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,90	-	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,91	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	~	0	
2,92	-	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,93	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,94	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,95	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	+	0	0	
2,96	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,97	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,98	+	0		0	++	0	0	0	0	0	0	0	0	0+	?	0	0	0	
2,99	+	0		++	++	+++	+	+	0	0	0	0	0	0	?	0	0	0	
2,100	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,101	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,102	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,103	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,104	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,105	+	0		0	+	+	0	0	0	0	0	0	0	0	0	0	0	0	
2,106	+	0		0	+	+++	0	0	0	0	0	0	0	0	0	0	0	0	
2,107	+	0+	210	0	0+	0	0	0	+	0	0	~	0	0	0	0	0	++	
2,108	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,109	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,110	+	0+		0	+++	0	0	0	?	0	0	0	0	0+	0+	0+	0	0	
2,111	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2,112	+	O+			O	+++	O	O	O	+	O	+	O	O	O+	O+	O	O	O	
2,113	+	O			O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,114	+	O			+	O	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,115	+	O+			O	+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,116	+	O+		390	O	++	O	+	++	O	O	O	O	O	O	O+	+	O	O	
2,117	+	O			O	+	O	O	O	O	O	O	O	O	O	O+	O+	O	~	O
2,118	+	O			O	+++	O	O	O	O	O	+	O	+++	+++	+++	+++	O	O	
2,119	+	O			O	+	O	O	O	O	O	O	O	O	+	?	O	O	O	
2,120	+	O+			O	+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,121	+	O			O	O+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,122	+	O			O	O	O	O	O	O	O	O	~	O	O	O	O	O	O	
2,123	+	O			O	+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,124	+	O			O	O	O	O	O	O	O	O	O	~	?	?	O	O	O	
2,125	+	O			O	+++	O	O	O	?	O	O	~	~	?	?	O	O	O	
2,126	+	O+			O	+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,127	+	O			O	+	+	O	O	O	O	O	O	O	O	O	O	O	O	
2,128	+	O+			O	++	O	O	O	+	O	+	O	O	O	O	O	O	O	
2,129	+	O			O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,130	+	O+			O	+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,131	+	O			O	++	+	O	O	O	O	O	O	O	O	O	O	O	O	
2,132	+	O			O	++	O	O	O	O	O	+	O	O	O	O	O	O	O	
2,133	+	O			O	+++	+++	+	+++	O	O	O	O	O	O	O	O	O	O	
2,134	+	O			O	+	O	O	O	O	O	O	O	O	O	O	O	~	O	
2,135	+	O+			O	+	+	O+	+	O	O	O	O	O	O	O	O+	O	?	
2,136	+	O			O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,137	+	O			O	O	O	O	O	O	O	O	O	O	O	O	O+	O	O	
2,138	+	O			O	O+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,139	+	O			+	O+	O	O	O	O	O	O	O	O	O	O	O	O	O	
2,140	+	O			++	+++	O	O	O	O	+	+++	+++	+++	+++	+++	++	O	O	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2,141	+		0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,142	+		0+		0	0	0	0	0	0	0	0	0	0	0+	0+	0	0	0	
2,143	+		0		0	0	0	0	0	0	0	0	0	0	0	0+	0	0	0	
2,144	+		0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,145	+		0+		0	++	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,146	+		0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,147	+		0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,148	+		0		+	++	0	0+	0	0	+	++	+++	0	0	0	+	0	0	
2,149	+		0+		0	+	0	0	0	0	+	++	+++	0	0	0	0	0	0	
2,150	+		0		0	+	0	0	0	0	0	0	0	+++	+++	+++	++	0	0	
2,151	+		0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,152	+		0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,153	+		0		+	+	0	0	0	+++	0	0	0	0	0	0	0	0	0	
2,154	+		0+		+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,155	+		0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,156	+		0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,157	+		0+		0	++	0	0	0	0	0	0	0	0	0+	0+	0	0	0	
2,158	+		0		+	+++	0	0	0	0	0	0	0	0	0+	0+	0	0	0	
2,159	+		0+		0	+++	++	+	+	0	0	0	0	0	0	0	0	0	0	
2,160	+		0		0	+++	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,161	−		0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,162	+		0		++	+++	0	0	0	+++	0	0	~	0	0	0+	0	0	0	
2,163	+		0		+	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,164	+		0		+	+	0	0	0	0	0	+	~	++	?	?	+++	0	0	
2,165	+		0+		0	0	0	0	0	+++	0	0	0	0	0	0	0	0	0	
2,166	+		0		0	0	0	0	0	0	0	0	0	0	?	?	0	0	0	
2,167	+		0		0	+	0	0	0	0	0	0	0	~	0	0	0	0	0	
2,168	+		0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,169	+		0+		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
2,170	+	0		0	+	0	0	0	0	0	0	~	0	0+	0+	0	0	0	
2,171	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,172	+	0+		0	+	0	0	0	0	0	0	~	0	0+	0+	0	0	0	
2,173	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,174	+	0+		0	+	++	++	+++	0	0	0	~	0	0	0	0	0	0	
2,175	+	0		+	+	0	0	0	0	+	++	++	+++	+++	+++	0	0	0	
2,176	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
2,177	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,178	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,179	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2,180	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,1	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,2	+	0		0	+	0	0	0	0	0	0	0	0	~	~	0	0	0	
3,3	+	0		0	0+	0	0	0	0	0	0	0	0	?	?	0	0	0	
3,4	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,5	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,6	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,7	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,8	+	0		0	+	0	0	0	0	+	++	+++	0	0	0	+	0	0	
3,9	+	0		+	+++	0	0	0	0	++	+++	+++	++	++	++	++	0	0	
3,10	+	+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,11	+	0		0	+++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,12	+	0+		0	0	0	0	0	0	+	++	++	~	~	0	0	0	0	
3,13	+	0		0	0+	0	0	0	0	0	0	0	~	0	0	0	0	0	
3,14	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,15	+	0		0	0+	0	0	0	0	0	0	0	~	~	~	0	0	0	
3,16	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,17	+	0		0	+++	+	0+	+	0	0	0	0	0	0	0	0	0	0	
3,18	+	0+		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3,19	+	0+		0	0+	0	0	0	0	+	++	+++	~	?	?	0	0	0	
3,20	+	0		0	+	++	+	+++	0	0	0	0	~	0	0	0	0	0	
3,21	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,22	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,23	+	0		0	0	0	0	0	0	+	++	++	~	+	+	0	0	0	
3,24	+	0		0	0	0	0	0	0	0	0	0	+	+	+	0	0	0	
3,25	+	0		0	0+	0	0	0	0	0	0	0	0	+	+	0	0	0	
3,26	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,27	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,28	+	0		0	+++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,29	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,30	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,31	−	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	
3,32	+	0		0	0+	0	0	0	0	0	0	0	0	0+	0+	0	0	0	
3,33	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,34	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,35	+	0		0	+++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,36	+	0		++	+++	0	0	0	0	0+	++	+	+++	+++	+++	+++	0	0	
3,37	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,38	+	0		0	0	0	0	0	0	0	0	~	0	0+	0+	0	0	0	
3,39	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,40	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,41	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,42	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0+	?	0	
3,43	+	0		0	+++	0	0	0	0	0	0	0	0	0	0	~	?	0	
3,44	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	~	0
3,45	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	
3,46	+	0		0	+++	0	0	0	0	0	0	+	~	0	0	0	0	0	
3,47	+	0+		0	+++	0	0	0	0	0	0	0	~	0	?	?	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3,48	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,49	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,50	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,51	+	0		++	+++	0	0	0	0	+++	+++	+++	+	++	++	++	0	0	
3,52	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,53	+	0+		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,54	+	0		0	0+	+++	0	0	0	0	0	0	0	0	0	0	0	0	
3,55	+	0		+	+	++	0+	+	0	0	0	0	0	0	0	0	0	0	
3,56	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	~	0	
3,57	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	~	0	
3,58	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	~	0	
3,59	+	0+		0	+++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,60	-	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,61	+	0		0	0	0	0	0	0	0	0	0	0	?	?	0	0	0	
3,62	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,63	+	0		0	++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,64	+	0		0	+	0	0	0	0	0	0	0	0	~	?	0	0	0	
3,65	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,66	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,67	+	0		0	++	+++	+	++	0	0	0	0	0	0	0	0	0	0	
3,68	+	0+		0	0	?	0	0	0	0	0	~	0	0	0	?	0	0	
3,69	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,70	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,71	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,72	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	~	0	
3,73	+	+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,74	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,75	+	0		0	+++	0	0	0	0	++	+++	+++	++	++	++	++	0	0	
3,76	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3,77	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,78	+	0		0	0	0	0	0	?	0	0	0	0	0	0	0	0	0	
3,79	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,80	+	0		+	++	0	0	0	+++	0	0	0	0	0	0	0	0	0	
3,81	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,82	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,83	+	0		0	0	?	0	0	0	0	0	0	0	0	0	0	0	0	
3,84	+	0		0	+	?	0	0	+	0	0	0	0	0	0	0	0	0	
3,85	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,86	+	0	290	0	0	0	0	0	0	0	0	0	0	0	0	0	0	++	
3,87	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,88	+	0		0	+++	0	0	0	0	0	+	++	0+	0+	+	0	~	0	
3,89	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,90	+	0		0	+++	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,91	+	0		0	+++	0	0	0	+++	0	0	0	0	0	0	0	0	0	
3,92	+	0+		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,93	+	0		0	++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,94	+	0		0	+	0	0	0	0	0	0	0	++	++	++	0	0	0	
3,95	+	0		0	++	0	0	0	0	0	0	0	0	0	0	+	0	0	
3,96	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,97	+	0		+	+++	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,98	+	0		0	0+	?	0	0	+	0	0	~	0	0	0	0	0	0	
3,99	+	0		0	+	+++	0	0	0	0	0	0	0	0	0	0	0	0	
3,100	+	0		0	0	+++	0	0	0	0	0	0	0	0	0	0	0	0	
3,101	+	0		0	0	+++	0	0	0	0	0	0	0	0	0	0	0	0	
3,102	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,103	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	~	0	
3,104	+	0		0	+++	+++	+	++	0	0	0+	~	0	0	0	0	0	0	
3,105	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3,106	+	0			0	+++	0	0	0	0	0	0+	+	+++	+++	+++	0	0	0	
3,107	+	0			0	0+	0	0	0	0	0	0	?	0	0	0	0	0	0	
3,108	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,109	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,110	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,111	+	0			0	+	0	0	0	0	++	+++	+++	0	0	0	0	0	0	
3,112	+	0+			0	0	0	0	0	0	0	0+	~	0	0	0	0	0	0	
3,113	+	0			0	+	0	0	0	0	+	++	+++	0	0	0	0	0	0	
3,114	+	0			0	+	+	?	0	0	0	0	0	0	0	0	0	0	0	
3,115	+	0+			0	++	++	0	0	0	0	+	+	+++	+++	+++	++	0	0	
3,116	+	0			0	0	0	0	0	0	0	0	0	0	0	0	++	0	0	
3,117	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,118	+	0			0	0	0	0	0	0	0	0	0	0	0	~	0	0	0	
3,119	+	0			0	0	0	0	0	0	0	0	~	0	0	0	+	0	0	
3,120	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,121	+	0		790	0	+++	0	0	0	+++	0	0	0	0	~	~	0	0	0	
3,122	+	0+			0	+++	0	0	0	+++	0	0	0	0	0	0	0	0	0	
3,123	+	+			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,124	+	0+			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,125	+	0+			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,126	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,127	+	0		550	0	++	0	0	0	0	0	0	0	0	0	0	0	+	0	
3,128	+	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,129	+	0+			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	+++	
3,130	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,131	+	0			0	0+	+	0+	+	0	0	0	0	0	0	0	0	0	0	
3,132	+	0+			0	++	0	0	0	0	+++	+++	+++	++	+++	+++	++	0	0	
3,133	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,134	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	~	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3,135	+	0+			0	++	0	0	0	0	0	0	0	0	0	0	0+	0	0	
3,136	+	0			0	++	0	0	0	0	0	+	~	0	0	0	0	0	0	
3,137	+	0			0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,138	+	0			0	++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,139	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,140	+	0			0	0	0	0	0	0	0	0	~	0	0	?	0	0	0	
3,141	+	0			0	++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,142	+	0+			0	0	0	0	0	0	0	0	0+	0	0	0	0	~	0	
3,143	+	0+			0	0	0	0	0	0	0	0	0	0	?	~	0	~	0	
3,144	+	0+			+	0	0	0	0	0	0	0+	0+	0	0	0	0	0	0	
3,145	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,146	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,147	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,148	−	0+			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,149	+	0			+	+++	0	0	0	0	0	+	+	0	0	0	0	0	0	
3,150	+	0			0	0	0	0	0	0	0	0+	+	0	0	0	0	0	0	
3,151	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,152	+	0			0	+	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,153	+	0			++	+++	0	0	0	0	+++	+++	+++	+++	+++	+++	++	0	0	
3,154	+	0			++	+++	0	0	0	0	++	+++	+++	+++	+++	+++	++	0	0	
3,155	+	0+			0	+	+	0	+	0	0	0	0	0	0	0	0	0	0	
3,156	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	~	0	
3,157	+	0			0	0	0	0	0	0	0	0	0	0	?	?	0	0	0	
3,158	−	0+			0	0+	0	0	0	0	0	0	0	0	?	?	0	0	0	
3,159	+	0+			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,160	+	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,161	+	0+	1450 230,170		+	+++	++	0	0	+++	0	0	0	~	0+	+	0	0	0	
3,162	+	0			0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	
3,163	+	0			0	0	0	0+	++	0	0	0	0	0	0	0+	0	0	0	

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3,164	+	0			0	+++	0	0	0	0	0	+	+	0	0	0+	0	0	0	
3,165	+	0			0	+++	0	0	0	0	0	0	~	+++	+++	+++	0	0	0	
3,166	+	0			0	++	+++	0+	++	0	0	0	~	0	0	0	0	0	0	
3,167	+	0			0	++	0	0	0	0	++	+++	+++	?	?	0	0	0	0	
3,168	+	0			0	0+	0	0	0	0	0	0	+	0	0	0	0	0	0	
3,169	+	0			0	+	0	0	0	0	0	0	0	~	0	?	0	0	0	
3,170	+	0			0	0	0	0	0	0	0	0	0	~	0	?	0	0	0	
3,171	+	0			0	++	0	0	0	0	+	+	++	++	++	+++	0	0	0	
3,172	+	0			0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,173	+	0			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,174	+	0			0	+++	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,175	−																			
3,176	−																			
3,177	−																			
3,178	+	0+			0	0+	0	0	0	0	0	0	0	0	0	~	0	0	0	
3,179	+	0			0	0	0	0	0	0	0	0+	+	0	0	0	0	0	0	
3,180	+	0+			0	+	0	0	0	0	0	0	0	0	0	0+	0	0	0	
3,181	+	0+			0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,182	+	0			0	0+	0	0	0	0	0	0	~	0	0	0+	0	0	0	
3,183	+	0+			0	++	++	+	+++	0	0	0	~	~	0	0+	0	0	0	
3,184	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,185	+	0			+	+++	0	0	0	0	0	0	+	0	0	0	0	0	0	
3,186	+	0			0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	
3,187	+	0+			0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,188	−	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,189	+	0			0	0+	0	0	0	0	0	0	~	0	0	?	0	~	0	
3,190	+	0			+	+++	0	0	0	0	+++	+++	+++	++	+++	++	+	0	0	
3,191	+	0			0	0	0	0	0	0	0	0	~	0	?	?	0	0	0	
3,192	−	0			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3,193	–	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3,194	–	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4,1	+	0		0	0	0	0	0	0	0	0	0	0			0	0	0	0
4,2	+	0		0	0+	0	0	0	0	0	0	0	0			0	0	0	0
4,3	+	0		0	0	0	0	0	0	0	0	0	0			0	0	0	0
4,4	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	+
4,5	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,6	+	0		+	+	+++	0	0	0	0	0	0	0			0	0	0	0
4,7	+	0	410 230	0	0+	+	0	0+	0	0	0	0	0			0	0	0	+
4,8	+	0		0	0+	+	0	0	0	0	0	~	0	0	0	0	0	0	0
4,9	+	0	870	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	+
4,10	+	0	850	+	+	0	0+	?	0	0	0	?	0	0	0+	0	0	+++	0
4,11	+	0	230	0	+	0	+	+++	0	0	0	0	0	0	0	0	0	0	?
4,12	+	0		+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,13	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,14	+	0	210	0	+	?	0+	?	0	0	0	0	0	0	0	0	0	+++	0
4,15	+	0		+	+	+++	0+	?	0	0	0	0	0	0	0	0	0	0	0
4,16	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,17	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,18	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	+
4,19	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,20	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,21	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0+	0	0	0
4,22	+	+		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,23	+	0		+	+	0	0	0	0	0	+	?	+++	+++	+++	0	0	0	0
4,24	+	0		0	0	0	0	0	0	0	0	0	+	0	?	0	0	0	0
4,25	+	0+		0	+	0	0	0	0	0	0	~	~	0	0	0	0	0	?
4,26	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,27	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	0

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
4,28	+	0		+	0+	++	0	0	0	0	0	~	0	0	0	0	0	0	?
4,29	+	0		0	0+	+	++	+++	0	0	0	0	0	0	0	0	0	0	0
4,30	+	0		0	+	0	0	0	0	0	0	~	0	0	0+	0	0	0	?
4,31	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,32	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,33	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,34	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,35	-	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,36	+	0		+	+	+	0	0	0	0	0	0	0	~	0+	0+	0	0	0
4,37	+	0		0	0	0	0	0	0	0	0	~	~	0	0	0	0	0	0
4,38	+	0+		0	++	0	0	0	+++	0	0	0	0	0	0	0	0	0	0
4,39	+	0		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,40	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,41	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,42	+	0+		0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,43	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	+	0	0	0
4,44	+	0	580	0	0	0	0+	0+	0	0	0	0	0	0	0	0	0	0	+
4,45	+	0	720	0	0	0	0	0	0	0	0	0	0	0	0+	0	0	0	+
4,46	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,47	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	?
4,48	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,49	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,50	+	0+		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,51	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,52	+	0	580	0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	+
4,53	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,54	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,55	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,56	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
4,57	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,58	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0+
4,59	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,60	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,61	+	0+	640	0	0	0	+	++	0	0	0	0	0	0	0	0	0	0	+
4,62	+	0		+	++	+++	0	0	0+	0	0	0	0	0	0	0	0	0	0
4,63	+	0		+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,64	+	0		+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,65	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,66	+	0		0	0+	+	0	0	0	0	0	0	0	0	0	0	0	0	0
4,67	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,68	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,69	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,70	+	0		0	0	0	0	+	0	0	0	~	0	0	0	0	0	0	0
4,71	+	0	290	0	+	0	0	0	0	0	0	0	0	0	0	0	0	+++	0
4,72	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,73	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0+	0	0	0
4,74	+	0		0	+	0	0	0	0	0	++	+++	+++	0+	0+	0+	0	0	?
4,75	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,76	+	0		0	0	0	0	?	0	0	0	0	0	0	0	0	0	0	0
4,77	+	0+		0	0	0	0	0	0	0	0	0	0	0+	0+	0+	0	0	0
4,78	+	0		+	+	0	0	0	0	0	++	+++	+++	++	++	++	0	0	0
4,79	+	0		0	0+	0	0	0	0	0	0	0	0	~	0	0	0	0	0
4,80	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	+	0	0
4,81	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	+	0	0
4,82	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,83	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
4,84	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,85	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0+	0	0	0

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
4,86	+	0		0	0+	0	0	0	0	0	0	0	0	0+	0+	0	0	0	0
4,87	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,88	+	0	300	0	0+	0	0	0	0	0	0	0	0	0	0	0	0	?	0
4,89	+	0		0	+	0	0	0	0	+	+++	+++	0+	0+	0+	0	~	0	0
4,90	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,91	+	0		+	+	0	0	0	0	0	+	0	0	~	~	0	0	0	+
4,92	+	0	1070	0	+	0	0	0	0	0	0	~	0	0	~	0	0	0	++
4,93	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,94	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,95	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,96	+	0		+	+++	0	0	0	?	0	+	+	++	++	++	++	~	0	?
4,97	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,98	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,99	+	0		0	0+	+	0	0	0	0	0	~	0	0	0	0	0	0	0
4,100	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,101	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,102	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,103	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,104	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,105	+	0		0	+	0	0	0	0+	0	0	0	0	0	0	0	0	0	0
4,106	+	0+		0	++	0	0	0	0	++	+++	+++	+++	+++	+++	+++	0	0	?
4,107	+	0		0	0+	0	0	0	0	0	0	0	+++	+++	?	0	0	0	0
4,108	+	0		+	+	0	0	0	0	+	+++	+++	0	0	0	++	0	0	0
4,109	+	0+		0	+	0	0	0	0	+	+++	+++	0	0	0	+	0	0	0
4,110	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,111	+	0		0	+	0	0	0	?	0	0	~	0	0	0	0	0	0	0
4,112	+	0		0	+	0	0	0	0	0	0+	~	0	0	0	0	0	0	0
4,113	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,114	+	0		0	+++	0	0	0	0	+++	+++	+++	+++	+++	+++	++	0	0	0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
4,115	+	0	<200	0	0+	0	0	0	0	0	0	~	0	0	0	0	0	0	+
4,116	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,117	+	0		0	0+	+	0	0	0	0	0	0	0	0	0	0	0	0	0
4,118	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,119	+	0		0	+++	0	0	0	0	0	0	0	~	0	0	0	0	0	0
4,120	+	0		0	+++	0	0	0	++	0	0	~	0	0	0	0	0	0	0
4,121	+	0+		0	0	0	0	0	0	0	++	+++	+++	+++	+++	0	0	0	0
4,122	+	0		0	0+	0	0	0	0	0	0	0	0	+++	+++	+++	0	0	0
4,123	+	0		0	0	0	0	0	0	0	0	0	~	0	0	0	0	0	0
4,124	+	0		0	+	+	0+	0+	0	0	0	0	~	0	0	0	0	0	0
4,125	+	0		0	+	0	0	0	0	0	+++	+++	+++	0+	0	0	0	0	0
4,126	+	0		0	+	+++	+	++	0	0	0	0	0	0	0	0	0	0	0
4,127	+	0		0	+	++	+	++	0	0	0	~	0	0	0+	0	0	0	?
4,128	−	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,129	+	0		0	+	0	0	0	0	++	+++	+++	~	~	~	0	0	0	0
4,130	+	0		0	++	0	0	0	+++	0	0	~	0	0	0	0	0	0	0
4,131	+	0		++	+++	+++	+++	+++	0	0	0	~	0	0+	0+	0	0	0	?
4,132	+	0		+	++	0	0	0	0	0	0	~	0	0	0	0	0	0	?
4,133	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,134	+	0		++	+++	0	0	0	0	+++	+++	+++	++	+++	++	0	0	0	0
4,135	+	0	840 330	0	++	0	0	0	0	0	0	~	0	0	0	0	0	0	++
4,136	+	0		0	0	0	0	0	0	0	0	~	0	0	0	0	0	0	0
4,137	+	0+		+	++	0	0	0	0	0	0	~	0	0	0	+	0	0	?
4,138	+	0+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,139	+	0		++	+++	0	0	0	0	0	++	++	++	++	++	+++	0	0	0
4,140	+	0		0	0+	0	0	0	0	++	++	+++	0+	0+	?	0	0	0	0
4,141	+	0		0	0+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,142	+	0		0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4,143	+	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0