The control of expression of storage protein genes in pisum sativum L

Waterhouse, Rosemary Noelle

The Control of Expression of Storage Protein Genes in *Pisum sativum* L.

by

Rosemary Noelle Waterhouse, B.Sc. (Nottingham)

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

Department of Botany

September, 1985

The Control of Expression of Storage Protein Genes in *Pisum sativum* L.

Rosemary Noelle Waterhouse

ABSTRACT

Pea cotyledon and leaf genomic DNA were found to be methylated in a series of defined methylation states. ^{m}CG and ^{m}C-X-G methylations were detected and the latter was more prevalent in leaf DNA. Pea rDNA was also found to be highly methylated but was relatively undermethylated in the developing cotyledon. The significance of the relative hypo-methylation of cotyledon genomic DNA (and rDNA) is discussed with respect to the endoreduplication phase of seed development.

Two post-expression demethylation events associated with the legumin gene family were detected using a cDNA probe. The methylation of specific CCGG sequences in and around two legumin genes was also investigated. The extent of the methylation of the genes was found to increase in a 5' to 3' direction and one gene was found to have an unmethylated site about 500bp upstream from the transcription start site. Minor changes in the extent of methylation of two sites in the protein coding regions of the two genes were detected and these are thought to represent 'fine-tuning' of gene expression, rather than major gene switching events.

One or two post-expression demethylation events associated with the vicilin gene family, were detected using cDNA probes. In addition, there was evidence that some cytosines associated with the vicilin genes became hypermethylated during cotyledon development. A normal pattern of 50,000-M_r vicilin gene demethylation and hypermethylation was detected in the cotyledon DNA of a mutant pea line, which produces reduced levels of 50,000-M_r vicilin polypeptide and message.

Analysis of the sequence data of two legumin genes indicated that in general the CG dinucleotide was suppressed although one exon was found to have a cluster of CG dinucleotides and an increased usage of the CG-containing arginine codons. The mutability of 5-methylcytosine is discussed in relation to possible legumin protein coding requirements.

CONTENTS

Contents (continued)

Contents (continued)

Contents (continued)

FIGURES

TABLES

ACKNOWLEDGEMENTS

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

I am also grateful to Dr John Gatehouse for the original concept of this project and for his assistance during my time at Durham. I should especially like to thank Dr Marta Evans for her interest, advice and encouragement during the last four years

I would also like to thank Dr R.R.D. Croy for supplying me with the λLeg clones and the cloned legumin genes, and Mrs Phillipa Brown and Mr Dave Bown for introducing me to a wide range of techniques.

I am indebted to the assistance of Dr G. Turner at the University of Newcastle for the use of his laser densitometer, to Dr C.H. Shaw and Mr M. Gibson for allowing me to use their computers, and to Mr M. Levasseur for his Leg D DNA sequence.

Special thanks are due to those who have helped with the preparation of this thesis, namely: Mrs D. Hughes for the diagrams, Mr P. Sydney for the photographs and to Mrs C. Cumming and Mrs K. Davison for the typing of the manuscript.

Finally, my thanks go to my parents, for their continued support of my academic career and without whose assistance this thesis would not have been possible.

This work was supported by funds from the Science and Engineering Research Council.

ABBREVIATIONS

The abbreviations used are as recommended by the Biochemical Society (1983). Instruction to authors. Biochem. J., 209, 1-27. Additions to this list are given below. The one-letter notation for amino acids (used in table 13) is given in: Biochem. J. (1969), 113, 1-4.

bp	= base pairs
BSA	= bovine serum albumin
d.a.f.	= days ofter flowering
kb	= kilobase pairs
LMP agarose	= low melting point agarose
M_n	= number average molecular weight
M_w	= mass average molecular weight
NTS	= ribosomal gene non-transcribed spacer
PEG	= polyethylene glycol
SDS	= sodium dodecyl sulphate
SSC	= saline sodium citrate
dCTP	= deoxycytidine 5'-triphosphate
mRNA	= messenger RNA
rDNA	= ribosomal DNA
rRNA	= ribosomal RNA
tRNA	= transfer RNA
5mC	= 5-methylcytosine
6mA	= 6-methlyadenine

1: INTRODUCTION

1:1 General Introduction: Biotechnology in Perspective

In 1798, in 'An Essay on the Principle of Population, as it affects the Future Improvement of Society', an English clergyman, Thomas Malthus, observed that

> the increase of population will take place if unchecked in a geometrical progression, while the means of subsistance will increase in only an arithmetical progression.
>
> (Encyclopaedia Britannica, 1974, Hill, 1975)

He asserted that the consequence of this was that populations would always tend to increase until prevented by lack of food or catastrophies (such as plagues or wars). However, this view fails to take into account the influence of scientific progress on food supplies or the productive potential of newly discovered lands, eg the prairies (Hill, 1975).

The productivity of an area can be increased in three ways. First, through the influence of government implemented policies. For example, peasants in China have doubled the national wheat production, through increasing their productivity by an annual average of 12% for the last seven years (The Economist, 1985a) whilst, in the 20 years preceeding 1977 they only achieved a 2.1% annual increase in grain production (*idem* , 1985b). Between 1977-79, the Chinese government had increased the price it guaranteed to farmers for produce and allowed them to sell their produce in the open market. The result of these governmental policies is that China has overtaken the Soviet Union as the world's largest wheat producer (*ibid*). There is an added

incentive for governments to initiate such policies, in that success in agriculture complements industry. A World Bank study of irrigation schemes in Malaysia, found that each dollar of extra rice production, generated 75 cents of extra demand elsewhere in the economy (*ibid*). When one considers that 98% of all the world's farmers are peasants but grow only 75% of the world's food and only 25% of its exports, it becomes apparent that they have achieved only a fraction of their potential (The Economist, 1985a).

The second way by which productivity can be increased is by reducing what can be as much as a 70% decrease of the potential yield for an arable crop, due to the combined effect of: preharvest losses (ie loss due to weeds, insects or disease); harvest losses (ie loss due to the selection and recovery of only a fraction of the potential harvest); post-harvest losses (ie loss due to vermin or fungal attack on stored material). Improved management can help reduce these losses.

The third way to improve productivity is by the better management of the crop (eg sowing date, plant density, multicropping etc.) and by the utilisation of the new varieties produced by the plant breeders. In developed countries, better management and plant breeding have each resulted in a 50% improvement in productivity. The use of high yielding varieties of rice, maize and wheat, which had been bred in the mid-1960s, has greatly increased the productivity of areas such as the Punjab, where cereal production rose by 7.3% a year between 1967 and 1982 (The Economist, 1985b). However, such successes are scattered and sometimes have unforeseen consequences. An increase in yield can often result in a decrease in protein content per unit of production

(Cuthbertson, 1970). Part of the success of the Punjabi farmers was as a result of them switching from the low value pulses to high value grain (The Economist, *op cit*) but this has nutritional implications. Cereals in general are deficient in the amino acids lysine and threonine, but this deficiency of the amino acids in the human diet can be partly offset by legume and leaf proteins (Cuthbertson, *op cit*).

It has been optimistically suggested that within the next ten years, agriculture could be transformed by biotechnology (The Economist, 1984a and b) and it is probable that these techniques have the potential to hasten the advances being made through conventional methods to breed new higher yielding varieties, for example with disease and pest resistance. A recent success in this field has been the introduction into tobacco, via the *Agrobacterium tumefaciens* 'Ti plasmid', of a gene encoding a toxin poisonous to some insect caterpillars from a bacterium (*Bacillus thuringiensis* 'Bt') (*idem* , 1984a and 1985c). Similarly, another bacterial gene endowing resistance to the herbicide 'Roundup', has also been successfully introduced into plants (*idem* , 1985c).

Paradoxically, genetic engineering offers the best chance to solve the problem of overproduction in the developed countries and help solve underproduction in the developing countries. If for instance in the long term, wheat could be genetically engineered to fix nitrogen, to become photosynthetically more active or to be pest and disease resistant (which is perhaps the most realistic of these options, as already indicated by the 'Ti plasmid' experiments), then in theory

costs could be substantially reduced and so allow farmers the same profit from a lower yield (for wheat, the variable costs due to chemicals and fertilisers account for 28% of the total costs of the crop (The Economist, 1984a).

Whilst the decisive step of putting into a crop plant a desirable plant gene, perhaps from a totally unrelated and at present incompatible species, is still a fairly distant ambition, there is still much basic research (such as the investigation of factors controlling normal gene expression) to be done, before the achievements of this field can be commercialised.

Adequate nutrition for all human beings must be one of the major goals for the next twenty years and the final objectives should be not only to ensure enough food to meet the effective demand (ie food that someone will pay for) but also to see that human needs for adequate nutrition are met (Scrimshaw and Taylor, 1980).

Whilst economic and political solutions are the most urgent, since food handouts are not the answer, it can also be argued that it is the duty of the 'rich' countries' governments to finance more research to transfer existing scientific information, and the duty of the 'poor' countries' governments to promote the uptake of such advances (The Economist, 1985a) and to realize the importance of agricultural self sufficiency. This has already happened in Bangladesh, where the population has an annual growth rate of 2.6% and yet her increasing productivity has enabled her to cut her cereal imports by roughly 2% a year, since 1974 (*ibid*).

1:2 The Synthesis of Storage Proteins in *Pisum sativum*

1:2:1 The importance of seed proteins

In developing countries, cereals provide about 60% and pulses, oilseed and nuts about 17% of the available proteins (Cuthbertson, 1970). Thus, seed proteins are an extremely important component of nutrition for both humans and animals (Millerd, 1975).

The proteins of seeds can be subdivided into two categories. The storage protein fraction contains large amounts of only a few different species of proteins and accounts for the major part of the total seed protein, whilst the minor 'housekeeping' protein fraction is composed of relatively small amounts of numerous protein species which are essential for the maintenance of normal cell metabolism, for example structural proteins and enzymes (Boulter, 1981). A seed storage protein may be defined as

> any protein accumulated in significant quantities in the developing seed, which on germination is rapidly hydrolyzed to provide a source of reduced nitrogen for the early stages of seedling growth.
>
> (Higgins, 1984)

The proteins must be able to withstand the osmotic changes during seed desiccation and be resistant to hydrolysis during this period, yet they must also tolerate the renewed osmotic changes during imbibition and become susceptible to hydrolysis during seed germination (*ibid*).

The seed storage proteins were first identified according to their solubility (globulins, albumins, prolamins and glutelins) by Osborne in 1924. The major storage proteins of the legumes are the

salt-soluble globulins. Danielsson (1949) showed that the globulin fraction could be characterized by an ultracentrifuge, by the presence of two peaks with sedimentation coefficients of about 7-8S (the vicilin fraction) and 11-12S (the legumin fraction).

In *Pisum sativum* , between 20-50% of the dry weight of the seed is protein and together, legumin and vicilin, constitute about 80% of the total protein in the mature pea seed (Evans *et al* , 1984). Legumin is the major storage protein, M_r 360,000-400,000 (Croy *et al* , 1980a) and vicilin, the secondary storage protein, M_r 180,000-200,000 (Derbyshire *et al* , 1976), is present in the mature seed at a level of between 25-75% that of legumin (Croy *et al* , 1980b).

In *P. sativum* , the legumin fraction is an oligomeric structure of six subunit dimers, 'αβ' (M_r about 60,000), each of which is composed of one acidic 'α' subunit (M_r about 40,000) plus one basic, 'β', subunit (M_r about 20,000), joined together by disulphide bonds (Derbyshire *et al, loc cit* ; Boulter, 1981; Higgins, 1984). The subunits of pea legumin have been shown to vary with respect to both size and charge (Matta *et al* , 1981b).

In pea, the 7S globulin fraction contains a number of subunits ranging in M_r from 12,000 to 71,000. The vicilin fraction comprises the subunits of M_r of 50,000 and less (Croy *et al* , 1980b; Gatehouse *et al* , 1981). In contrast to legumin, however, there is no disulphide bonding between the vicilin subunits. In addition to the vicilin subunits, the 7S globulin fraction contains a third protein, convicilin (M_r 290,000), the subunits of which each have a M_r of 71,000 (Croy *et al* , 1980c; Casey and Sanger, 1980). The amino acid

composition for convicilin and vicilin is similar (*ibid*).

There is much evidence indicating that both legumin and vicilin are initially synthesised as larger precursor molecules which are post-translationally modified (for vicilin see: Higgins and Spencer, 1981, Lycett *et al* , 1983b; Gatehouse *et al* , 1981; Gatehouse *et al* , 1983; for legumin see Croy *et al* , 1980a; Spencer and Higgins, 1980; Domoney and Casey, 1984 and 1985).

Investigation of the storage protein precursors has indicated that: (i) there are distinct variations in the coding sequences for both vicilin and legumin; (ii) these variations can define how a precursor molecule will be modified after translation; (iii) specific cDNA probes exist in several laboratories, which are capable of detecting the different major vicilin or legumin genes.

Recently (1984) Domoney and Casey have measured the number of genes coding for legumin, vicilin and convicilin. They were able to show that there were no major differences in the gene numbers in a number of *Pisum* genotypes of variant protein composition. However, they did detect differences in the number of genes encoding the different proteins. The number of gene copies detected per haploid genome were: for the 7S fraction proteins, 5-7 genes for the 47,000-M_r and 4-6 genes for the 50,000-M_r vicilin subunits and 1 gene for convicilin; for the 11-12S legumin fraction protein, 4-6 genes for the 60,000-M_r polypeptide, 1-3 gene copies for the 63,000-M_r/65,000-M_r polypeptide and 1-2 gene copies for the 80,000-M_r polypeptide. These figures are in reasonable agreement with the previously reported values. Croy *et al* (1982), using a cDNA probe, detected 4 gene copies

per haploid genome for the 60,000-M_r legumin precursor. However, when a cloned legumin gene was used as a probe (Shirsat, 1984), a figure of at least 7 legumin genes per haploid genome was obtained. For the vicilin fraction proteins Gatehouse *et al* ,(1983) reported 5 gene copies per haploid genome for the 50,000-M_r polypeptide and 2-3 gene copies per haploid genome for the 47,000-M_r polypeptide.

1:2:1 Legume seed development

Seed development, or embryogenesis, is one of the most critical stages in the life of the plant; it is a preparation for a successful germination, which is probably the most precarious time in the life of a plant (Dure III, 1975). The storage tissue is of particular interest to molecular biologists because it exhibits a general reversal of metabolism during the onset of germination. The same cells which had previously been synthesising enormous amounts of reserve materials (proteins and carbohydrates), now commence a rapid hydrolysis of these reserves during germination and also exhibit a reversal in the direction of flow of nutrients through the tissue. As this reversal has occurred in an unchanging cell population, considerable changes of gene activation and inactivation must also be occurring (*ibid*). Thus, the synthesis of storage proteins represents a one-time expression of a few genes, in a specific tissue, at a specific point of the plant's life-cycle.

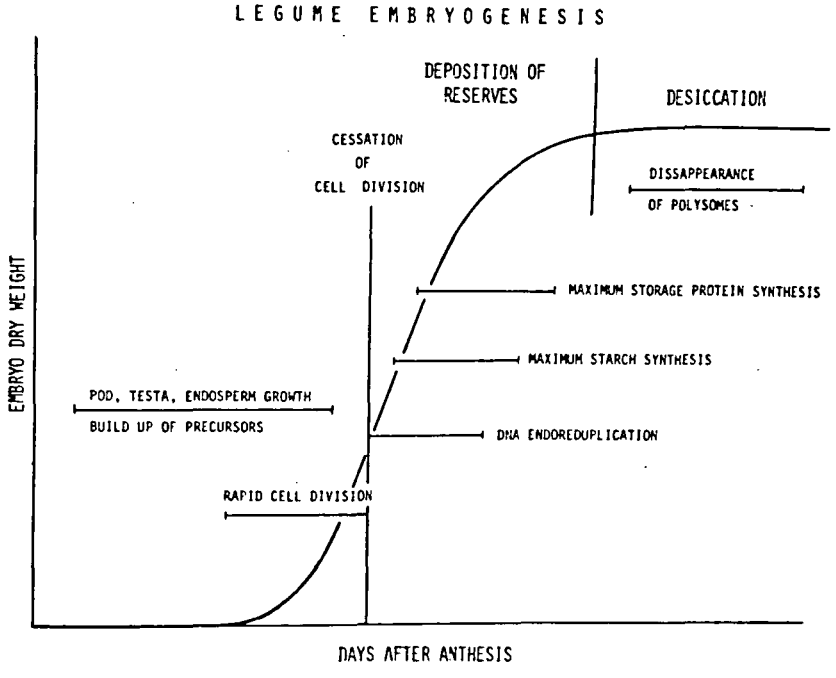
In 1955, McKee *et al* found that in *Pisum sativum* the final number of cells in the pea embryo (1.4×10^6), was reached less than half of the way through seed development. This feature was subsequently found

to be true for other legumes and dicotyledonous plants. This led Bain and Mercer (1966) to identify four morphological developmental phases in seed formation. These were: phase 1- 'Cell Formation' (the embryo appeared meristematic until day 10 and then differentiated); phase 2 - 'Cell Enlargement' (parenchyma cells in the cotyledons expanded fourfold, vacuolation of the cytoplosm was observed and storage proteins began to appear, in small amounts, towards the end of this phase); phase 3 - 'Synthesis of Storage Reserves' (this phase was marked by the onset of rapid synthesis of starch and reserve proteins and by a further increase in cotyledon size and seed volume); phase 4 - 'Maturation' (this phase was characterised by a marked fall off in both the rate of increase in fresh weight and in the rate of seed and embryo enlargement; the cells were observed to dry out as the seeds passed into dormancy and there was a loss of fine structure in cotyledon mitochondria during this final phase of embryogenesis). In the outside growing conditions utilised in this study, these phases corresponded to phase 1, 0-10 days after flowering (d.a.f.); phase 2, 10-18 d.a.f.; phase 3, 18-28 d.a.f. and phase 4, 28-54 d.a.f.

In 1965 Opik suggested that the nuclei in the cotyledons of *Phaseolus vulgaris* became polyploid and in 1971 this was confirmed by Smith, who observed that the cotyledon cells of *Pisum arvense* contained up to the 16C level of the DNA and that during germination this decreased to between 2C and 4C, which suggested that the polyploidy was acting as a form of nucleotide storage for the incipient seedling. For *Pisum sativum* , the level of polyploidy has

been reported as being between 32C and 64C. This high degree of ploidy was achieved shortly after the final number of cells and was concomitant with the beginning of maximum storage protein synthesis in the cells. This, together with the observation that the maximum rate of RNA synthesis occurred prior to the maximum storage protein synthesis phase, provoked the suggestion that its purpose was gene duplication to increase the rate of storage protein synthesis. However, the maximum RNA synthesis occurred before the maximum DNA synthesis and Millerd and Whitfeld (1973) finally clarified the matter by showing that in *Vicia faba* , the endoreduplication involved an increase in total nuclear DNA and therefore was not specific gene amplification.

The seed development in *Pisum sativum* is summarised in Figure 1 (taken from Dure III, 1975).



Over 90% of the protein in *Pisum* seed cotyledons is synthesised during the cell expansion phase (Millerd and Spencer, 1974). Using immunoelectrophoretic techniques the authors were able to show that in rapidly grown peas (25°C/16h light days), vicilin was first detected about 9d.a.f., when 60% of the final cell complement was present; legumin was detected one day later, when about 80% of the cells were present. However, it has since been demonstrated that a very low level of legumin synthesis can be detected in pea embryos during the very early stages of embryogenesis (Domoney *et al* , 1980; Gatehouse *et al* , 1982). Thus, the results of Domoney (*loc cit*) indicate that either some storage protein synthesis preceeds endoreduplication or that the low level of expression, which they were able to detect in immature embryos, reflects synthesis in a small proportion of cells which have already ceased to divide.

Figure 2 summarises the changes in dry weight, legumin and vicilin accumulation during the development of rapidly growing peas (taken from Gatehouse *et al* , 1982).

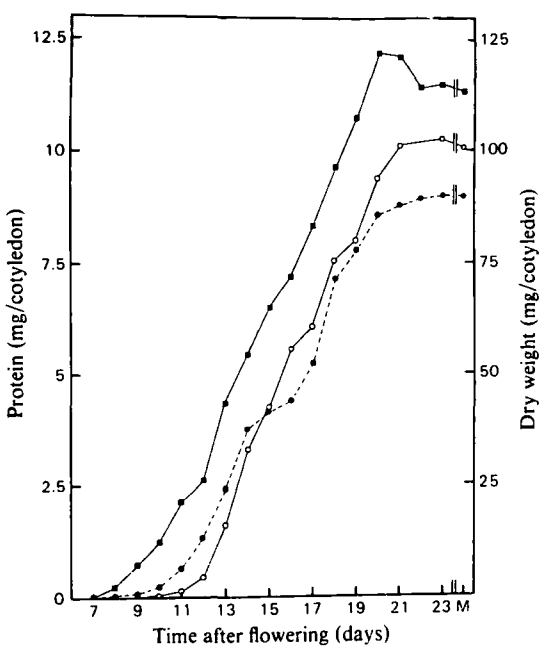


Fig. 2. Cotyledon dry weight accumulation (■) and accumulation of the cotyledon storage proteins legumin (o) and the vicilin fraction (●) in developing pea seeds.

The total amount of protein accumulated by the seed is proportional to either the rate of transcription or the equilibrium between RNA synthesis and degradation. Madison *et al* (1981), were able to demonstrate that the half lives of both legumin and vicilin mRNAs were longer than the maturation time of the seeds. This implies that the accumulation of the storage proteins is not influenced to any great extent by degradation of the message. During normal seed development the major limiting factor in storage protein accumulation is the level (ie synthesis) of mRNA.

Morton *et al* (1983), using hybridization kinetic studies, observed an increase (from 1-6), in the number of very abundant poly$(A)^+$ -RNA sequences (ie those occurring at greater than $5x10^5$ copies per cell), from early to mid-developmental stages. This increased level was maintained through to the late developmental stage. These changes were in agreement with the increased synthesis of storage protein polypeptides during the early- to mid-developmental stages and their continued synthesis through to late cotyledon development. Thus, it was suggested that these six very abundant mRNA species, present at 14d.a.f., be equated with the vicilin 50,000-Mr and 47,000-M_r, the convicilin 71,000-M_r and the legumin 60,000-M_r precursor polypeptides. The other two mRNA species could perhaps represent major seed albumins or lectins.

Northern blotting techniques and cell-free RNA translation systems have aided the investigation of changes in specific mRNAs. Gatehouse *et al* (1982), were able to demonstrate that the changes in the relative levels of the 47,000-M_r vicilin and 60,000-M_r legumin

messages were in agreement with the known accumulation patterns for the storage proteins. However, the appearance of the 50,000-M_r vicilin message at 10d.a.f. appeared to contradict the early onset of vicilin synthesis (9d.a.f.) although it is possible that the cDNA clone used to probe the mRNA was only selecting a mRNA for a 50,000-M_r polypeptide synthesised during later stages of cotyledon development and that mRNAs for 50,000-M_r polypeptides synthesised during early stages of cotyledon development were less homologous and they were therefore not detected by the probe used.

In order to investigate possible regulatory controls on transcription, it is necessary to measure actual RNA synthesis in terms of 'run-off' transcripts, produced by nuclei isolated from the different developmental stages. Evans *et al* (1984) hybridized radioactively labelled transcripts from a series of developmental stages, to Southern blots of specific cDNA probes for the storage proteins. The authors were able to show that storage protein gene transcripts formed an increasing proportion of the total transcription as cotyledon development proceeded, but were undetectable in the leaf transcripts. A strong correlation was observed between the level of storage protein transcripts in the developing cotyledon detected by these experiments (*ibid*), with the amounts of corresponding cytoplasmic mRNAs (Gatehouse *et al* , 1982). The results obtained by Evans (*loc cit*), clearly indicated that there was apparent control of these tissue-specific genes at the transcriptional level.

1:3 The Control of Eukaryotic Gene Expression

The first level of transcriptional control in eukaryotic cells is provided by a series of discrete sequences, usually located in the DNA flanking region, 5' to the gene. A second level of transcriptional control may be a limited accessibility (for example of enzymes such as RNA polymerase II) to regions of chromatin, or to specific sequences of DNA. Eukaryotic chromatin has a dynamic, complex hierarchical structure and active gene transcription occurs on only a small proportion of it at any one time (Reeves, 1984). These two levels of control will be discussed separately.

1:3:1 Regulatory sequences involved in animal gene expression

Potential regulatory regions have been deduced from the comparison of sequence data in normal and mutant organisms. This enables firstly, the identification of common sequences, found in similar locations (with respect to gene position, ie a determinate number of nucleotides 'upstream' or 'downstream' from the site of initiation of transcription (cap-site)). Such 'common' sequences are said to have been conserved through evolution and might therefore be expected to have a functional role.

Secondly, the study of sequence differences occurring in mutants, known to have an altered expression of a particular gene, has frequently shown that the mutant DNA differs at only one nucleotide (ie a point mutation) and yet this mutation may be sufficient to drastically alter gene expression. For example, in the Greek type of 'hereditary persistence of foetal haemoglobin' (HPFH), there is a 'G'

to 'A' substitution in what is known as the distal 'CCAAT-box' (ie 5' - CTTGACCAATAG-3' becomes 5' - CTTAACCAATAG-3') (Gelinas *et al*, 1985). This transition is sufficient to cause the persistance of the expression of one of the foetal globin genes in adults. Similarly, *in vitro* produced deletion mutants can be shown to have abnormal transcription, ie certain DNA sequences have been shown to have functional role in normal gene expression.

The regulatory signals (figure 3) serve two functions. First, they specify the position where RNA synthesis is to begin (this has been identified by biochemical characterization of primary RNA transcripts). Second, they govern the efficiency of transcription initiation (this has been clearly indicated by visual analysis of active genes and by the measurement of the rate of RNA synthesis) (McKnight and Kingsbury, 1982).

The first major regulatory element to be identified was the Goldberg/Hogness or 'TATA'-box, the percentage composition of which is $T_{82}A_{97}A_{85}{}^{A_{63}}T_{37}\ A_{83}{}^{A_{50}}T_{33}$ (Lewin, 1983). This sequence is situated 25-30 nucleotides upstream from the site of initiation of transcription ('cap site') (Breathnach and Chambon, 1981).

The TATA-box falls in an AT-rich region of DNA and is similar in sequence to the prokaryotic 'Pribnow' or TATAAT-box, which is situated only 10 nucleotides upstream from the cap site. It is of interest, that the Pribnow box is approximately one turn and the Goldberg/Hogness box three turns of DNA helix, from the cap site. The exact location of the TATA-box varies slightly from gene to gene, with the T in position '1' falling between positions -34 to -26 nucleotides

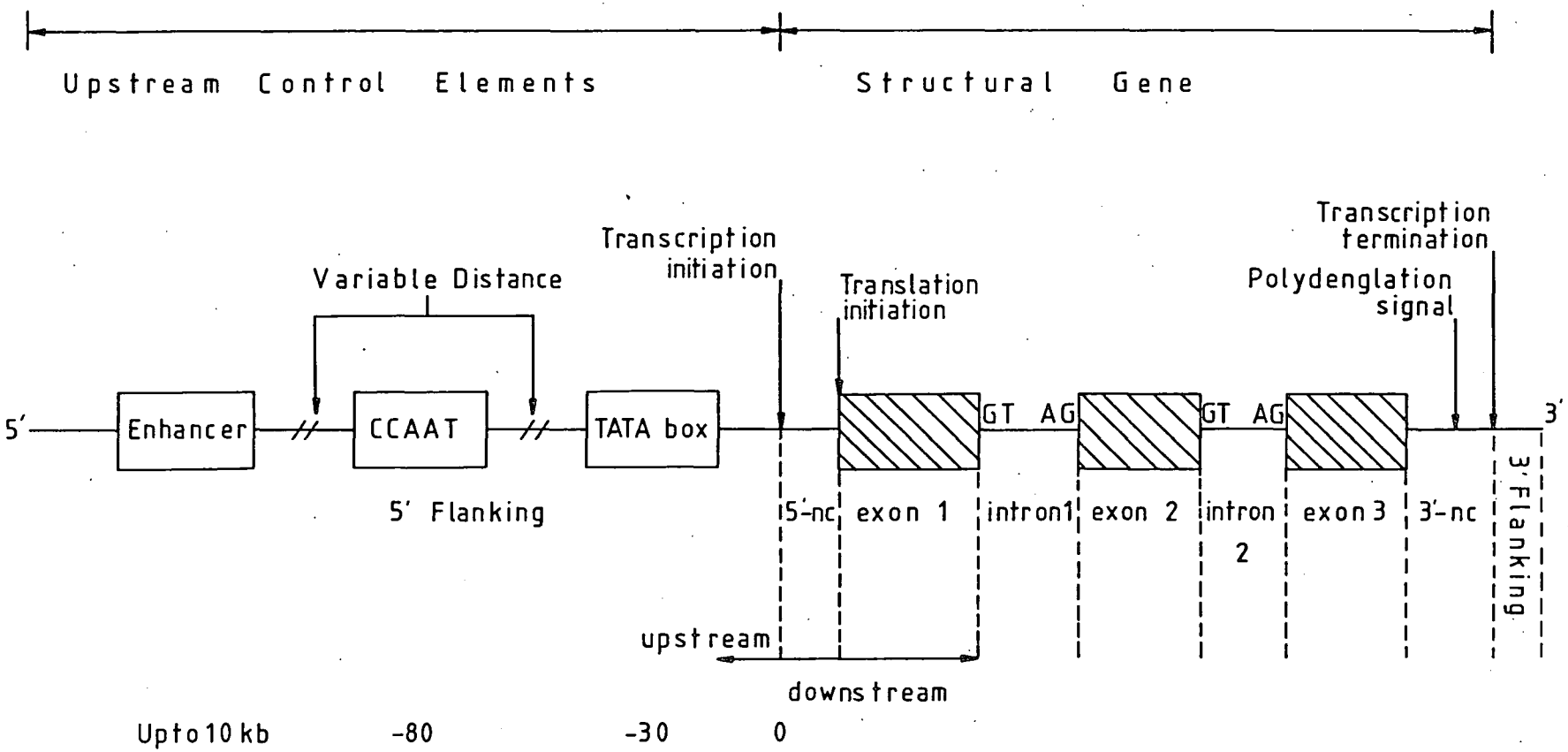


Fig 3. Diagrammatic representation of a typical animal gene transcribed by RNApolymerase II.

upstream from the cap site. However, 54 out of 60 genes examined had this T falling within 2 nucleotides of position -34 (Breathnach and Chambon, *op cit*).

Absence of a TATA-box has been shown to result in micro-heterogeneity of mRNA molecules, thus indicating its importance in defining the cap site (eg Baker *et al* , 1979). In some cases, anomolies in the TATA-box will result in a reduced efficiency of transcription. For example, in the chicken conalbumin gene, a point mutation whereby TATA becomes TAGA, is sufficient of an abnormality to reduce transcription by 95% (Minty and Newmark, 1980).

The second regulatory element is the 'upstream element' or 'CAAT-box', which is found about 80 nucleotides (between 60 to 100) upstream from the cap site (Benoist *et al* , 1980). Deletion of this element has been shown to markedly reduce the transcription of globin genes *in vivo* (but not *in vitro*) (Minty and Newmark, *loc cit*).

The immunoglobulin genes show an interesting variation in this regulatory element. In every heavy chain gene examined, at approximately 70 nucleotides upstream from the cap site, is the sequence ATGCAAAT. At the corresponding location of every light chain gene examined, is the precise inverse of this sequence ATTTGCAT (Parslow *et al* , 1984). Excluding the TATA-box at -30 nucleotides, the remainder of the flanking sequences varies wildly among the different immunoglobulin genes, ie the flanking sequences have diverged extensively in evolution but TATA and these two sets of octanucleotides have been conserved. This may imply they are involved with the tissue specific factors required for immunoglobulin gene

transcription. In addition a pseudogene was shown to have no detectable homology with the octanucleotide blocks.

Also in this second category of regulatory elements, probably come the GC-rich elements. This is a broad class of sequences that come at variable distances from the cap site (Reudelhuber, 1984). For example, in SV40, there are three 21 base pair repeats containing six GC-rich hexanucleotide sequences (CCGCCC) (Khoury and Gruss, 1983).

The third group of regulatory elements are the 'enhancers'. These are short cis-acting regulatory sequences, that strongly stimulate transcription from the promoters of nearly genes. They increase transcriptional efficiency, in a manner which is independent of their orientation, position or distances (which in some cases can be up to 10kb) from a gene (Khoury and Gruss, *op cit* ; Banerji *et al* , 1983). They were first identified in viruses (eg the 72 base pair repeat in SV40). They may act by providing a bidirectional entry site for either RNA polymerase II or one of its subunits (Khoury and Gruss, *op cit* ; Velcich and Ziff, 1984). It has also been proposed that enhancers (or their flanking sequences) may alter either chromatin structure or the superhelicity, to create regions of transcriptional efficiency.

To complicate matters further, there is increasing evidence that some eukaryotic genes harbour regulatory elements within the structural gene itself. For example, the globin gene family, the members of which are differentially expressed during development in the following sequence: ε = embryonic; α = foetal; β = adult. When the promoter from the foetal α-globin gene, was fused to the adult β-globin structural gene and the hybrid introduced into a mouse

erythroleukemia (MEC) cell line, the expression of this hybrid gene was found to be β-globin like in its relative rate of transcription (Wright *et al* , 1984; see also Charnay *et al* , 1984). Similarly, the mouse immunoglobulin genes contain an enhancer in an intron (Picard and Schaffner, 1984) and the chicken thymidine kinase gene has also been shown to have an intragenic control element (Merill *et al* , 1984).

Other regions of the gene with possible roles in gene regulation are the polyadenylation signal and the termination of message signal and the introns. In eukaryotes, termination of the message can occur in one of three ways (Proudfoot, 1982). Firstly, as in the histone genes, the message is terminated at a specific sequence by the RNA polymerase II. Secondly, in yeast, the enzyme both terminates the transcript and in a coupled reaction, polyadenylates it. Lastly, in all other higher eukaryotic RNA polymerase II genes examined the enzyme initially terminates the precursor message, possibly at a similar site to that recognized in the histone genes, the newly formed 3'-end of the precursor message is then cleaved by an endonuclease to 15 bases 3' of an AAUAAA sequence. This implies that probably both the RNA endonuclease and poly(A) polymerase recognizes the AAUAAA as a signal. The consensus DNA sequence AATAAA, is now definitely known to be a polyadenylation signal, and is not also a termination site. Hofer and Darnell (1981) demonstrated that the initial transcripts from the mouse β-globin gene, extended 1,400 nucleotides beyond the poly-adenylation addition site. However, these extended messages did have definite ends, therefore there must be a precise termination sequence.

Gil and Proudfoot (1984) showed that additional sequences are required for efficient termination and polyadenylation of the message in rabbit ß-globin. These are a preceeding pentanucleodide, CATTG, which is complementary to regions of the RNA from the U4 small nuclear ribonucleoprotein and a proceeding GT-rich octnucleotide, which it is thought might be required for correct 3' and formation.

The introns are transcribed regions of a gene which are spliced out during processing of the message. Some introns have been shown to have multiple splice sites. The intron always has a GT-dinucleotide at its 5' end and an AG-dinucleotide at its 3' end (Breathnach and Chambon, 1981) splicing of a message nearly always occurs after polyadenylation. Naora *et al* (1982) found that the overall length of the introns was a function of the total size of the gene.

Some transcription units are complex, their transcripts can give rise to two or more in RNAs (ie they encode two or more different proteins) (Darnell, 1982). These complex messages usually have either two or more polyadenylation sites (eg immunoglobulin heavy chain and calcitonin genes) or two or more splicing sites, which enables the primary transcript to be spliced in different ways (eg viral genes) (*ibid*).

Integration of these different regulatory elements can allow for a differential or tissue-specific expression of genes. Carl Wu (1984) was able to show the presence of two protein binding sites, which covered the TATA-box sequence and an upstream control element in the 5' flanking region of the *Drosophilia* heat shock gene. He proposed that a heat shock activity protein (HAP) was thus able to positively

regulate the gene. In a similar manner, Groner *et al* (1984) have detected common sequences in the 5' flanking regions of unrelated genes which respond to glucocorticoid hormones. These regions were found to bind the glucocorticoid hormone/receptor complex.

Dean *et al* (1983), showed that the 5' flanking sequence of the ovalbumin gene was essential for progesterone regulation of this gene and Ott *et al* (1984) demonstrated that the 5' flanking sequence of the rat albumin gene confers tissue specific expression. Similarly, a hybrid gene formed with the 5' flanking sequence of the small subunit of 'Rubisco' plus the structural portion of the bacterial gene for chloramphenicol, resulted in the bacterial gene becoming light regulated (Shields, 1984). Several authors Reudelhuber, 1984; Picard and Schaffner, 1984; Ott *et al, loc cit*), have proposed that enhancers act as tissue-specific modulators, ie that the TATA-box region may act as a 'selector' whilst the upstream region acts as a 'modulator' (Breathnach and Chambon, *op cit*).

Gene regions, other than the 5' flanking sequences, have also been shown to be necessary for tissue-specific gene regulation. Banerji *et al* (1983) claimed that a downstream enhancer was necessary for lymphocyte-specific expression of immunoglobulin heavy chain genes and Gillies *et al* (1983) detected a tissue-specific transcription enhancer in the major intron of a rearranged immunoglobulin heavy chain gene. Similarly, Milner *et al* (1984) have shown that an 82 nucleotide brain-specific 'identifier sequence' is present in introns of genes expressed in the brain. These are thought to have been inserted as mobile elements.

1:3:2 Putative regulatory sequences involved in plant gene expression

The flanking DNA regions around known plant genes, have been shown to have several sequences in common with those previously discussed as being involved in the control of animal gene expression. Thus it has been inferred that they serve a similar function, with respect to plant gene expression. However, few experiments have been published which demonstrate a functional relationship between the sequences and plant gene expression. Link (1984) using isolated mustard plastids and either the cloned 32,000 -M_r photosystem II protein gene or truncated linear templates was able to identify three DNA regions, upstream from the cap site that were necessary for efficient transcription. These regions were first at 0 to -13bp, 5'-TATACT-3', which resembles the prokaryotic 'Pribnow' box usually found at -10bp. On its own this was found not to be sufficient for transcription of the plastid gene. Second at -13 to -17 bp, there was a sequence 5'-TATAAA-3', which matches the consensus sequence for the eukaryotic TATA-box. Third at -28 to -35 bp upstream from the start site, was 5'-TTGACA-3'. This matches the consensus sequence for the prokaryotic '-35' promoter element and this region was found to be needed for efficient and selective initiation of transcription. In the absence of the '-35' element the TATA-like box was found to give a basic level of transcription. It is interesting to note that the plastids have a combination of prokaryotic and eukaryotic-type regulators.

The only other published work, which conclusively assigns a particular function to a specific DNA sequence, involves the *Agrobacterium tumefaciens* , 'Ti-plasmid T-region' (Shaw *et al* , 1984).

These workers, using a series of deletion mutants, were able to demonstrate that the right copy (but not the left copy) of the 25bp repeat is required for T-DNA transfer and/or integration. Removal of the right copy of the repeat abolished the oncogenicity of the mutant plasmid.

However, plant DNA 5'-flanking regions have been shown to have sequences in common with known DNA regulatory sequences and although it has not yet been demonstrated, by functional assays, that they serve a regulatory purpose, the similarities in DNA sequence, size and location, make it possible to infer that they may have a similar role.

Plant genes have been shown to have a TATA-like box (consensus sequence T $\begin{matrix}C\\G\end{matrix}$ TATA $\begin{matrix}T\\A\end{matrix}$A -- $\begin{matrix}C\\T\end{matrix}$A), at 29-33 nucleotides upstream from the cap site (Messing *et al* , 1983). This similarity in TATA-box is expected because both animal and plant protein genes are transcribed by RNA polymerase II. Plant genes have also been shown to have a translation start site (consensus sequence 5'-$\begin{matrix}C\\G\end{matrix}$ AANNATGG-3'), 5'-GT/AG-3' limits to the introns and an AATAAA polyadenylation signal (*ibid*). There are indications that multiple polyadenylation signals occur more often in plant genes than in animal genes (Lycett *et al* , 1983a). It also appears that there is more variation in the polyadenylation signals of plant genes, for example, leghaemoglobin GATAAA and legumin AATAAG.

Most plant genes have only a limited homology to the CAAT-box and instead show a greater homology to the AGGA-box, the consensus sequence for which is $\begin{matrix}C\\T\end{matrix}$A---- $\begin{matrix}G\\T\end{matrix}$NGA --- $\begin{matrix}CC\\TT\end{matrix}$ (Messing *et al, op cit*). This AGGA-box is situated 70-90 nucleotides upstream from the cap site

but it is not yet known whether this is a novel functional element, involved in the regulation of plant gene expression. The legumin gene, however, does exhibit a good match for the CAAT-box, 126 nucleotides upstream from the cap site, but it has only a partial homology with the AGGA-box (Lycett *et al* , 1984).

1:3:3 Possible diffusible substances involved in the regulation of plant gene transcripts

In the animal system, proteins eg heat shock activator protein (Wu, 1984) and hormones eg glucocorticoid hormones (Groner *et al* , 1984) and progesterone (Dean *et al* , 1983) have been shown to bind to certain DNA sequences in the 5'-flanking region. It can be postulated that a similar method of eliciting gene expression may be present in plant systems.

The expression of the pea storage protein genes has been shown to be apparently co-ordinated, tissue-specific and developmentally regulated at the transcriptional level. Many factors could affect transcription and the developing seed is under the influence of many potential stimuli. It is known to be a rich source of several plant growth substances ('hormones'), which are known modulators of other gene systems. There is also a marked change in osmotic stress during the latter half of seed development and the developing cotyledons act as a sink for many metabolic precursor molecules. It is probable that plant gene expression is dependent upon some or all of these factors and possibly some as yet unknown factors, all of which are inter-related and may act together in a cumulative manner.

To take plant growth substances first, gibberellins, cytokinins and abscisic acid have all been detected in cotyledons and some reports state that growth regulators change their level during development. For example, Skene (1970) isolated gibberellin-like substances from several stages of *Phaseolus vulgaris* seed development. It was demonstrated that two gibberellin-like substances had been isolated, gibberellin A_1-like (GA_1) and gibberellin A_5-like (GA_5) activity. GA_5-like activity, was highest in young seeds and disappeared after cell division in the cotyledons had ceased. However, GA_1-like activity rose to its highest level during the period of rapid cell expansion (ie protein accumulation phase) in the cotyledons. GA_1 is now known to be the active form of this class of hormones.

The evidence for a regulatory role by growth substances is frequently contradictory. For example, abscisic acid (ABA) has been shown to be essential for the synthesis and accumulation of the 12S globulin (napin), in cultured embryos of *Brassica napus* (Crouch and Sussex, 1981); yet it has no observed effect on storage protein accumulation by cottonseed embryos (Dure III and Galau, 1981). The addition of ABA to cultured excised embryos of *Phaseolus vulgaris* was shown to stimulate a 3-5 fold increase in the accumulation of vicilin (Spencer and Higgins, 1982).

The presence of sucrose (0.35M) has been shown to be essential for the synthesis and accumulation of napin (Crouch and Sussex, *op cit*). Domoney *et al* (1980), also showed the importance of sucrose (18%) in culture media, for the initiation of legumin synthesis by immature embryos of *Pisum sativum* (and in doing so contradicted Millerd's

results of 1975). It is not known how sucrose is acting to enhance gene expression, whether it is as a molecule binding directly to a regulatory gene sequence or indirectly to a protein already bound to a gene sequence. Its action could also result from an indirect physiological effect, eg osmotic pressure. Certainly, Millerd's failure to detect the initiation of legumin synthesis by immature embryos is probably due to a relative starvation experienced by the cotyledon cells, thus causing them to degrade available metabolites (possibly including the small amounts of storage proteins), to provide a source of carbon skeletons.

Other simple metabolites have also been shown to have dramatic effects on protein accumulation. Altering the mineral nutrients can cause significant and special changes in the spectrum of seed proteins laid down. An altered proportion of legumin results in *Pisum sativum* grown in a deficient supply of phosphorus or potassium (3-fold increase in the relative amount of legumin observed) or sulphur (greatly reduced to a barely detectable level). The distribution of vicilin subunits was also shown to alter with these deficiencies. However, plants grown with an inadequate supply of magnesium were shown to have a normal legumin content (Randall *et al* , 1979).

Chandler *et al* (1983) demonstrated that in sulphur deficient peas, the legumin mRNA was reduced by 90% and that when an adequate supply of sulphur was resumed, both legumin mRNA and legumin levels were restored to normal. Recently (1985), Beach *et al* have shown that during the first 48h of recovery from sulphur deficiency, transcription of legumin genes is increased by a factor of about 2,

whilst the level of legumin mRNA was shown to increase 20-fold. However, vicilin gene transcription was shown to decrease to normal levels, during the recovery period and this was consistant with the gradual decrease in vicilin mRNA and vicilin synthesis. This indicated that sulphur was influencing legumin synthesis at a post-transcriptional level, whereas its influence on vicilin synthesis was at the transcriptional level.

Obviously, much more research needs to be done, to elicit the mode by which diffusible substances influence plant gene expression.

1:3:4 The influence of chromatin structure on gene expression

Only a small proportion of chromatin is being actively transcribed at any one time (approximately 10-20% of the total). Domains of chromatin which are being transcribed, are apparently packaged in an altered nucleosome structure, which is less condensed and more open than inactive domains (Reeves, 1984). Modification in histones and non-histone proteins, associated with active chromatin, have also been reported.

DNA is normally packed into chromosomes, as a series of structures of increasing complexity (ie a multiple hierarchy). The nucleosome is the lowest level of DNA coiling. The 'core' particle consists of 146bp of double stranded DNA, wrapped in two left-handed superhelical turns around the outside of an octamer of histone proteins ($H4_2$ $H3_2$ $H2A_2$ $H2B_2$) (McGhee and Felsenfeld, 1980; Kornberg, 1977). The nucleosome is a repeating structure (similar to beads-on-a-string) which are joined together by spacer or linker DNA sections, of between 0 to 89bp in

length. This structure is called the extended 10nm chromatin fibre.

The bulk of the inactive (condensed) chromatin consists of 25-30nm thick chromatin fibres, produced as a result of the winding, or coiling, of the 10nm fibre, into a shallow supercoiled 'solenoid' (6 nucleosomes per solenoid turn) (eg McGhee *et al, op cit*). Histone 'H1' is important for the formation and/or stabilization of the 30nm fibre, but its actual location is unknown.

The third level of structure is found in interphase nuclei and metaphase chromosomes and is the folding of the 30nm solenoid into loops or domains of chromatin. These loops are anchored by specific nonhistone proteins located at the base of the chromatin loops to a supported nuclear structure, which has been called the scaffold, cage or envelope (Reeves, 1984). The loops of supercoiled chromatin may be from 35 to in excess of 100kb of DNA.

The extent of chromatin condensation can be monitored by the susceptibility of the chromatin to digestion by pancreatic DNAse I or micrococcal nuclease. It has been suggested that in actively transcribing chromatin, the nucleosome itself may unfold in some way to give easier access of the DNA, to RNA polymerase molecules. It has been frequently suggested that chromatin domains of active genes have an increased susceptibility (relative to inactive or bulk chromatin), to digestion by various nucleases (eg Weisbrod, 1982, McGhee and Felsenfeld, *op cit*). The DNAse I seems to preferentially detect the structure of chromatin that predisposes it to transcriptionl activity, if other cellular conditions are also permissive (Reeves, *op cit*).

Various other features of chromatin structure and composition,

have been correlated with an increased susceptibility to DNAase I. These include undermethylation of the DNA (which, together with Z-DNA will be discussed as separate issue, see section 1:4), modification of histones and the association with active nucleosomes of non-histone proteins, for example the high-mobility group (HMGs) (Reeves, *op cit*).

Histone H1, is generally greatly depleted in active chromatin domains (Igo-Kemens *et al* , 1982) and it is thought that in such regions, the H1 has been replaced by HMGs. HMGs are present in increased amounts in transcriptionally active or potentially active areas of chromatin HMG proteins have been shown to have varied effects on chromatin. HMGs 1 and 2 can either induce superhelicity or unwind and destabilize DNA in closed circular plasmids (Javeherian *et al*). HMG 14 and 17 can (*in vitro*) partially inhibit histone deacetylase in mammalian tissue culture cells. In other studies HMGs have been shown to stabilize the nucleosome core particles at the points of entry and exit of DNA. Weisbrod (1982) showed (*in vitro*) that chicken erythrocytic globin genes lost their selective 'active' DNAase I sensitivity when HMG proteins were removed from proteins but regained the sensitivity when HMGs 14 and 17 were reconstituted onto the chromatin.

Histones, H2B, H3 and H4 (but not H1), are known to be modified by acetylation. This is a post synthetic modification of the amino termini of core histones, which are enriched with basic amino acid residues eg lysine (Isenberg, 1971). Acetylation, neutralises the positive lysine charge of the histone 'tail' and thus loosens the

histone-histone or histone-DNA interactions, ie the chromatin may become opened up or destabilized due to weakened ionic forces (Reeves, *op cit*).

Other histone modifications include phosphorylation, methylation, ubiquitination and poly (ADP) - ribosylation. Ubiquitin-H2A (UH2A) has been shown to be relatively enriched in the transcribed *Drosophila* copia and heat shock genes, whilst being virtually absent from non-transcribed satelite chromatin (Levinger and Varshavsky, 1982). It has been suggested that histone ubiquitination in transcribing nucleosomes may modify some nucleosome-nucleosome interaction, thus preventing the formation of higher-orders of chromosome structure (Levinger and Varshavsky, 1982).

Histones are also modified by poly(ADPribosyl)ation (poly-ADPR) but it is not clear whether poly-ADPR modified histones are associated with transcriptionally active chromatin. Again it has been suggested that poly-ADPR of histones may relax chromatin and prevent the formation of higher-order configurations (Poirier *et al* , 1982; Aubin *et al* , 1983).

It is evident that the structure of chromatin is dynamic and it has been postulated that several known modifications of associated histones or non-histone proteins could 'open-up' regions of DNA, giving transcription enzymes controlled access to certain domains of the chromatin. Two fundamentally different types of mechanism must be involved in gene expression. First, gene activation or commitment, which makes available for transcription a selected subset of cellular genes (eg by changes in the chromatin) and second, a mechanism which

induces and regulates the actual expression of the potentially active genes (Reeves, 1984).

1:4 Cytosine Methylation and Gene Expression

It has been noted in Section 1:3 that the expression of genes requires a loosening of the DNA helix, followed by active transcription of the gene. There is much evidence that cytosine methylation may be involved in these two mechanisms.

There is a widespread occurrence of methylated bases (in particular 5-methylcytosine) amongst eukaryotes and this together with its intragenomic distribution, strongly suggests that it may have some common biological function. The usual site of methylation is the dinucleotide CG, which perhaps significantly is under-represented in the overall gene region, whilst being over-represented ('clustered') in specific regions. The relationship between cytosine methylation and the formation of a different structural form of DNA (the left-handed Z-DNA) is well documented, as is the inverse relationship between the extent of methylation and gene expression. Thus, there is a possible link between cytosine demethylation, a loosening of the DNA helix and subsequent gene expression.

Alternatively, there is a possibility that the methyl moiety by its physical presence, is acting as a signal to protein factors or enzymes involved in the transcriptional process.

The role of cytosine methylation in gene expression will now be discussed more fully.

1:4:1 Occurrence of modified bases in genomic DNA

The only modified bases found in eukaryotic and bacterial DNAs are 5-methylcytosine and 6-methyladenine (Figs.4a/b).

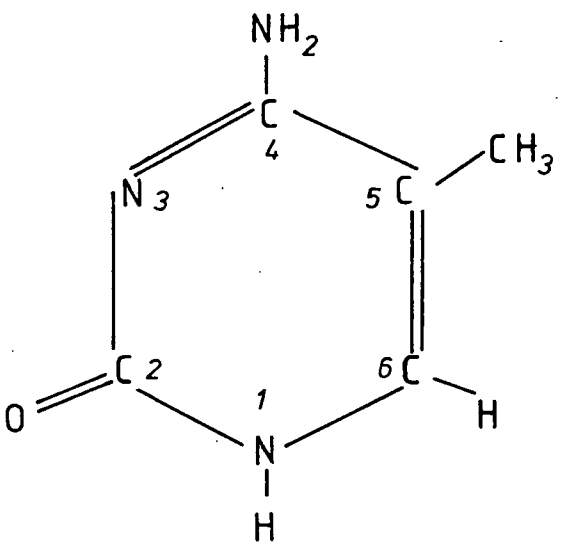


Fig.4a: 5-methylcytosine

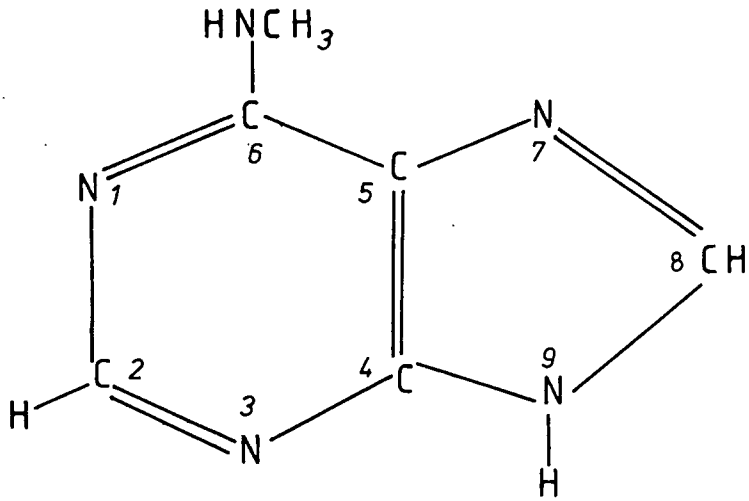


Fig.4b: 6-methyladenine

6-methyladenine (6mA) is a minor methylated base and has been found in bacteria (Vanyushin, 1968) and protozoa (Cummings *et al* , 1974). In higher eukaryotes, Vanyushin *et al* (1971) claim to have detected it in plant DNA but that it is not present to any appreciable degree in vertebrate DNA (Vanyushin *et al* , 1970).

5-methylcytosine (5mC), however, is the major methylated base in eukaryotic DNA. Base composition analyses have shown that levels of 5mC vary in different animal groups from 0.5 to 2.8 moles% of bases (Wyatt, 1951) and that the amount of 5mC in higher plants, is substantially higher than is found in animals (3.6 - 7.1 moles % of bases) (Ehrlich and Wang, 1981) (Table 1).

Table 1:- Examples of percentage 5mC of total cytosine

(Taken from Wagner et al 1981; Wyatt 1955)

Organism	% 5mC of total cytosine
Locust	1.0
Gerbil	3.1
Calf (thymus)	6.2
White Mustard	12.2
Wheat	22.4
Pea	23.2
Mistletoe	23.2
Field Bean	30.5
Tobacco	32.6

In higher eukaryotes, over 90% of methyl moieties are found in the dinucleotide sequence CG. This dinucleotide is more frequent in plant DNA (3-4%) than in animal DNA (0.5-1.0%) and this is one reason for the higher levels of total methylation observed in plant DNA (Gruenbaum et al, 1981b). In higher plants, methylation occurs at 25% of all cytosines but the methylation of the CG dinucleotide cannot account for all the 5mC present in the DNA and additional methyl groups have neen located in the symmetical trinucleotide C-X-G, where X can be C, A or T (*ibid*). It was determined that in wheat germ DNA, over 80% of all CG dinucleotides and C($\overset{A}{T}$)G trinucleotides and 50% of CCG trinucleotides, were methylated.

In vertebrate DNA, there are several different sequences that share in common the CG dinucleotide, in which 5mC is found. CCGG, GCGC and ACGT have all been demonstrated as being partially methylated (Razin *et al* , 1981). These sequences are all simple palindromes, with 180° rotational symmetry and can exist in one of three possible configurations, with respect to methylation level (Figure 5).

Figure 5:- Methylation of CG dinucleotides

5'-NCGN-3'	5'-NC*GN-3'	or 5'-NCGN-3'	5'-NC*GN-3'
3'-NGCN-5'	3'-NGCN-5'	3'-NGC*N-5'	3'-NGC*N-5'
Type 1: Unmethylated	Type 2: Half methylated (Hemi-methylated)		Type 3: Fully Methylated

(*represents a methyl modification)

The methylation status of CG dinucleotides has been investigated by many workers using the isoschizomers Msp I and Hpa II which have a complementary refractoriness for methylation. Both of these restriction enzymes have the recognition sequence 5'-CCGG-3' (McClelland, 1981 ; *idem* , 1983) and will cleave the unmethylated sequence but whenever the internal cytosine is methylated ($C^{m}CGG$), Hpa II is unable to cleave but Msp I can cleave such sites. (Waalwijk and Flavell 1978; Cedar *et al* , 1979). However, methylation of the external cytosine ($^{m}CCGG$) will not be cleaved by Msp I but can be cleaved by Hpa II (Singer *et al* , 1979; Sneider 1980). Neither enzyme will cleave at this recognition sequence when both cytosines are methylated ($^{m}C^{m}CGG$) (McClelland, 1981; *idem* ,1983). It should also be

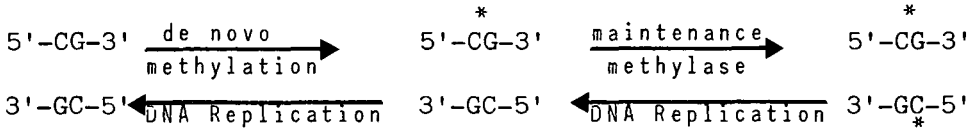
noted that certain flanking sequences to the restriction site, have been reported as influencing the specificity of Msp I cleavage (Busslinger *et al* , 1983b; Keshet and Cedar, 1983).

In a similar manner, to detect methylation at the C-X-G site, isoschizomers with the recognition sequence CC $\frac{A}{T}$GG, can be used. Eco RII will be resistant to cleavage if the internal cytosine is methylated, whereas the specification of Bst NI is unaffected by methylation. (McClelland 1981; Gruenbaum, *et al* 1981b)

1:4:2 Establishment of the Methylation Pattern

The methylation of cytosine residues requires the presence of a methylase (methyltransferase) enzyme (EC2.1.1.37). Two possible forms of methylase activity were proposed by Riggs (1975) and Holliday and Pugh (1975). One was a semi-conservative, maintenance type, methylating symmetrically hemi-methylated sequences and the other responible for methylating *'de novo'* specific, unmethylated sequences as illustrated in Figure 6.

Figure 6: Summary of Eukaryotic Methylation



* represents a methyl modification)

It is now known that these enzymes are located in the nucleus and employ the activated methyl donor, S-adenosyl-L-methionine (Borek and

Srinivasan, 1966), which is itself derived from L-methionine and ATP. These enzymes transfer the chemically active methyl group to the carbon number '5' of the cytosine residue (see figure 4).

When cultured cells are grown under conditions of methionine deprivation, hemimethylated DNA is formed ie one strand is deficient in methyl residues (Turnball and Adams, 1976). Methylases catalyze the addition *in vitro* of very few methyl groups to homologous DNA (Taylor, 1979). The homologous DNA is presumably nearly fully methylated at the sites for which the methylase is specific. However, DNA methylase was shown to methylate heterologous (hemimethylated) DNA to a greater extent than homologous DNA (Razin and Friedman, 1981; Turnball and Adams, *op cit* ; Adams *et al*, 1979; Gruenbaum *et al* , 1982). One bacterial enzyme has been shown to act more than one hundred times faster on a half methylated site than it does on an unmethylated site (Vovis *et al* ,1974).

It is thought that the rat methylase binds to the DNA helix and scans for potential modification sites (Drahovsky and Morris 1971a and b). The production of thymine dimers or double stranded breaks was found to be less inhibitory to mouse spleen methylase action, than the induction of single stranded breaks, apurinic sites, X-ray induced damage or alkylation (Wilson *et al* ,1983). It is not known whether these inhibitions are due to a limitation on the scanning function of the enzyme. It was originally thought that the enzyme having bound to the DNA, 'walked' along it looking for potentially methylatable sites. Drahovsky and Morris (1971a) reported that their enzyme methylated less than 2% of potential CG dinucleotides. Other authors have

indicated that the process of addition of methyl groups is slow and that the enzyme may leave the DNA molecule after each addition, or it may dissociate from the DNAs on encountering a site which is already methylated (Adams *et al* ,1984). Thus it may take several cycles for the enzyme to completely methylate all potential sites. However, such a mechanism would also have the advantage *in vivo* of preventing the enzyme from making fruitless searches for potential sites in fully methylated DNA (*ibid*).

Qureshi *et al*, (1982) and Adams *et al* (*op cit*) whilst attempting to isolate DNA methylase from the nuclei of Krebs II ascites cells, found that after all the soluble enzyme activity had been removed, there remained a residual activity which sedimented with the nuclear matrix. It was suggested that this bound form of activity may be associated with DNA replication and may thus perform the actions of a maintenance type enzyme, whilst the soluble DNA methylase could be a *de novo* type of enzyme, acting on completely unmethylated regions of DNA. There is no evidence for multiple species of DNA methylase to exist (Adams *et al*, *op cit*).

1:4:3 Inheritance of the Methylation Pattern

If the methylation cytosine residues is to play a role in the control of gene expression or in the differentation of cells, then an essential requirement must be that the information encoded in the methylation pattern is stably maintained through each replication cycle.

Replicating DNA has been shown to undergo methylation at or near

the replication fork (Bird, 1978). It is therefore possible for the parental DNA strand, which carries the methylation pattern, to serve as a "template" for a semi-conservative type of methylation, which will copy the methylation pattern from the parental strand onto the newly synthesized progeny strand. The key element in this model, is the symmetry of the methylated sites. Without some form of strand symmetry, it would be impossible to transfer faithfully the methylation pattern from generation to generation. Thus, it is expected that the hemimethylated sites (figure 5-Type 2) should not normally occur and this has been confirmed by the experiments of Bird and Southern (1978) and Bird (*op cit*).

Bird (*op cit*), also demonstrated that following DNA replication in cultured *Xenopus laevis* cells, new methyl groups were added only to the progeny strand and that the parental strand did not become labelled with (methyl-^{3}H) methionine. These experiments clearly indicated that any pattern of methylated and unmethylated paired CG dinucleotides, in the genome of a cell, would be inherited by the descendants of that cell.

Thus, for CG dinucleotides and in plants the additional C($^{A}_{T}$)G trinucleotide, a modification on the parental strand can signal to the methylase enzyme, that the cytosine residue sited symmetrically opposite requires methylation. However, the inheritance of possible modifications in the trinucleotide sequence C-C-G must be slightly different, to that discussed for CG of C($^{A}_{T}$)G. In the C-C-G sequence, either or both of the two cytosines may be modified (Gruenbaum *et al* ,1982). These cytosines have inherited their modification pattern

from the parental complementary strand G-G-C, which contains only one cytosine residue. Therefore, this single parental cytosine must have served as a template for the possible methylation of both cytosines, in the progeny C-C-G DNA sequence. The fact that this single parental cytosine, is part of a CG dinucleotide $\begin{matrix}5'\text{-}C\text{-}C\text{-}G\text{-}3' \\ 3'\text{-}G\text{-}G\text{-}\underline{C}\text{-}5'\end{matrix}$, suggests that if the parental cytosine ($\underline{C}$) is methylated, then the internal cytosine of the progeny strand will always be modified ie $\begin{matrix}5'\text{-}C\text{-}\overset{*}{C}\text{-}G\text{-}3' \\ 3'\text{-}G\text{-}G\text{-}\underset{*}{C}\text{-}5'\end{matrix}$, whilst the external cytosine may or may not be methylated

ie $\begin{matrix}5'\text{-}C\text{-}\overset{*}{C}\text{-}G\text{-}3' \\ 3'\text{-}G\text{-}G\text{-}\underset{*}{C}\text{-}5'\end{matrix}$ or $\begin{matrix}5'\text{-}\overset{*}{C}\text{-}\overset{*}{C}\text{-}G\text{-}3' \\ 3'\text{-}G\text{-}G\text{-}\underset{*}{C}\text{-}5'\end{matrix}$.

The decision by the methylase enzyme, whether or not to methylate the external cytosine may be random or it may be determined by adjacent DNA sequences. Gruenbaum *et al* , (*op cit*) found that only 50% of C-C-G sites are methylated at the external cytosine residue, whilst 80% of all CG or $C(\frac{A}{T})G$ sites are methylated.

Further evidence for the maintenance of an existing methylation pattern is provided by the experiments involving the injection of oócytes with foreign DNA sequences. Harland (1982) demonstrated that a pBR322 derivative plasmid methylated *in vitro* at 5'-CCGG-3' sequences and then injected into *Xenopus laevis* oocytes, had its methylation pattern maintained through replication ie in the progeny molecules, the methylated sites remained methylated and the unmethylated sites remained unmethylated. Stein *et al* (1982) were able to demonstrate the stable inheritance of a methylation pattern for one hundred generations, following the injection of ΦX 174 RF-DNA, into mouse cells (see also Pollack *et al* , 1980; Wigler *etal* , 1981; Wigler, 1981; Stein *et al* , 1982).

Methylations of the plant trinucleotide sequence C-X-G, were found

not to be inheritable in mouse cells (Stein *et al, op cit*) and it is also thought that the animal cell methylase is unable to transmit methylations at the external cytosine of the Msp I site (CCGG).

Harland (1982) also demonstrated that some methylation could occur in the absence of replication but it occurs less efficiently than in the presence of replication. This type of methylation might be associated with either repair replication or *de novo* methylation. Similarly, in lily DNA, methylation was shown to occur at each of 3 intervals of DNA synthesis associated with the meiotic cycle (Hotta and Hecht, 1971). These phases were: a) premeiotic S-phase interval, when bulk DNA is replicated; b) zygotene, when a high-GC satellite component undergoes replication delayed from S-phase; c) pachytene, which is characteristic of repair replication.

De novo methylation is best illustrated by the experiments involving the nuclear integration of viral DNA into cultured mammalian cells. Non-integrated or 'free' viral particles (eg Herpes simplex, Polyoma and Adenovirus) remain unmethylated (Kaye *et al* , 1967; von Acken *et al* , 1979). However, when a hamster tumour cell line is infected with adenovirus, the integrated viral DNA becomes methylated in a highly specific pattern (eg Sutter and Doerfler, 1980). This *de novo* methylation, may be as a result of the process of viral DNA insertion into the host genome and it may imply that chromosomal location is important for methylase action (Doerfler, 1983).

In general, *de novo* methylation is a rare event in eukaryotic

somatic cells, whilst maintenance methylation is very efficient (Razin and Szyf, 1984). However, the higher methylation levels, reported for human and bovine somatic tissue as compared to germline DNA (eg for sperm DNA, see Sturm and Taylor, 1981; Chapman *et al* , 1984), does imply that *de novo* methylation must occur at some stage during embryonic development, prior to the separation of somatic from germ line cells. However, this cannot be the rule because in other organisms eg rabbit, the total 5mC content of sperm is approximately 40% higher than that observed in the liver (Ehrlich and Wang, 1981).

1.4.4 Methylation in relation to DNA sequence type

Reanneling studies have indicated that highly repetitive DNA sequences (eg satellite DNA) contain more methylated cytosine residues than the corresponding bulk DNA, (Russell *et al* , 1976). In several instances, the satellite DNA has been found to contain a high content of the dinucleotide CG (Gruenbaum, 1981). In cultured mouse cells, the DNA of the inverted repeat ('foldback') class is methylated about 50% more than repetitive DNA which itself is three times more methylated than single copy or intermediate repetitive DNA (Drahovsky and Morris, 1971a). The rat satellite DNA, contains ten CGs in a repeating sequence of 370 base pairs and all are methylated. However, calf satellite DNA generally contains unmethylated or partially methylated CG sequences (Pech *et al* ,1979).

Experiments performed by Bird and co-workers (1979) using

Hpa II/Msp I restriction enzyme analysis of *Echinus* (sea urchin) DNA, have indicated that the DNA of this organism can be separated into a highly methylated fraction (about 40% of the genome) and an unmethylated fraction (about 60% of the genome). These domains or "compartments" were present in all developmental stages and tissues examined. The methylated CG sequences of the sea urchin DNA, were found to occur in specific long tracts of the genome (>15 kb) and the 5mC sequences were clustered, so that the methylated domains were separated from one another by stretches of unmethylated DNA.

It has since been demonstrated that the degree of methylation of Hpa II sites, in a number of organisms, does not vary continuously but can be classified into one of three distingushable groups (Bird and Taggart, 1980). In insects, the DNA methylation was barely detectable; an intermediate level of methylation was detected in non-arthropod inverebrates and a high level of methylation was detected in vertebrates. Organisms which are closely related in evolution exhibited similar methylation patterns. Bird and Taggart were able to conclude from these results that in all cases the methylated and unmethylated regions of the DNA were in separate domains and that in non-anthropod invertebrates, the unmethylated domain (m-) is predominant over the methylated domian (m+), whilst the reverse situation is found in the vertebrates. In insects, it must be assumed that the methylated domain is so small as to be undetectable.

The *Echinus* distribution pattern has also been observed in fungi and some plants and it has therefore been suggested that this pattern may be similar to the ancestral pattern of methylation, from which the

other two categories have been derived (Cooper, 1983).

Naveh-Many and Cedar (1982) demonstrated a clustering of methylated Hpa II sites in mouse DNA and of methylated Eco RII sites in wheat germ, tobacco and cauliflower DNA. One reason for this apparently non-random distribution of methyl moieties, may be that the under methylated regions represent the actively transcribed portion of the genome, whereas the inactive or silent DNA is highly methylated. This hypothesis is supported by data obtained from the analysis of the sea urchin, where histone and ribosonal gene sequences were found to be present in the unmethylated (m-) compartment (Bird *et al* , 1979).

The presence of some unmethylated stretches of DNA has been demonstrated in wheat germ and cauliflower (Naveh-Many and Cedar, 1982). Cooper *et al* (1983), using an end-labelling technique, demonstrated the presence of a small unmethylated domain (1-2% of the genome), thus indicating that in fact the echinoderms and vertebrates differ only quantitatively rather than qualitatively, with respect to the presence of these domains (Cooper, 1983).

1.4.5 Methylation in relation to gene expression

In vertebrates, several genes have been found to be methylated at restriction sites in germ line DNA but relatively unmethylated at specific sites in somatic cells which express the gene, but not in those where the gene is silent. For example: chicken ovalbumin gene (Mandel and Chambon, 1979); chicken β-globin (McGhee and Ginder, 1979); rabbit β-globin (Waalwijk and Flavell 1978) and foetal globin genes (van der Ploeg and Flavell, 1980). Other evidence shows that all expressed genes in a particular cell are about 30% as methylated as the average cellular DNA (Kunnath and Locker, 1982a). It has been

suggested that the *de novo* methylation observed during human or bovine embryo development, turns off genes that were previously expressed (eg during oogenesis) but whose continued expression is either not required or may be deleterious to the developing embryo ie *de novo* methylation may 'reset' the genetic programme (Jaenisch and Jahner, 1984).

There are over 36 reported examples of a negative correlation between the methylation of a gene and its expression (for reviews see Cooper, 1983; Riggs and Jones, 1983) but the mode of action, by which the loss of a methyl groups(s), can elicit gene transcription is not known. The conversion of a cytosine residue to 5mC introduces a methyl group into an exposed position in the major groove of the double helix. The binding of histones and hormone receptor proteins to the DNA helix is known to be affected by changes in the major groove. Several bacterial restriction enzymes have been shown to have a strong affinity for unmethylated sites but a reduced affinity for methylated sites.

It is now becoming apparent that adjacent CG dinucleotides within a given gene region, do not always display similar patterns of tissue-specific methylation. It has also become apparent, that whereas it was initially thought that perhaps a hypomethylation in the 5' flanking region (promoter region) of a gene was necessary for gene transcription to occur, that this hypomethylation alone, is not a sufficient condition for transcription. There are now several documented cases of various sites existing in the unmodified form in tissues where the gene is inactive. For example, the α2(type 1)

collagen gene (McKeon *et al* , 1982) had unmethylated DNA around the start site of transcription, whether or not the cells from which the gene was isolated, normally synthesised collagen or not. Similarly, Ott *et al* (1982) demonstrated that hypomethylation of the 5' end of the albumin gene was a necessary but not a sufficient condition for albumin gene expression. However, in general the correlation between hypomethylation and gene expression is strongest for the 5' flanking region (Riggs and Jones, 1983; Cooper 1983). It has been suggested (Riggs and Jones, *op cit*) that methylation 'locks' a gene in an inactive state. Therefore, removal of the methyl groups is essential for efficient gene transcription but in some systems additional factors are also involved.

Analysis of the pattern of the methylation status of individual sites along eukaryotic genes, has led Razin and Szyf (1984) to classify the genes into five paradigmatic groups. At one extreme are genes which are hypomethylated in expressing tissues whilst being fully methylated in non-expressing tissues eg, rat insulin I gene (Cate *et al* ,1983) and chicken β-globin gene cluster (McGee and Ginder (1979). The groups range to the fourth category, which contains the genes that remain fully methylated in all tissues, for example *Xenopus* vitellogenin (Gerber-Huber *et al* , 1983). The final group contains those genes which tissue specific hypomethylation occurs but cannot be correlated with the observed transcription of the gene, for example, rat α-fetoprotein (Kunnath and Locker, 1983) and *Xenopus* albumin (Gerber-Huber *et al, loc cit*). Razin and Szyf (*op cit*) point out from their analysis, that site-specific

hypomethylations in some cases, may be related to the expression of genes, eg 'housekeeping genes', which would be constituitively expressed in all tissues, whereas tissue-specific hypomethylations may be associated with the differential expression of genes.

Any change in the methylation of a specific site from being modified to being hypomethylated requires either the inhibition of the post-replicative maintenance methylase (ie a passive mechanism) or a specific demethylase (ie an active mechanism). There is only one published report of an extract from nuclei, which has a demethylating activity (Gjerset and Martin, 1982). Should a demethylase definitely be shown to occur, the situation would then exist where the DNA can code for two different proteins which in effect could either activate or silence other segments of DNA. At the moment it is difficult to understand how a demethylase could specifically identify and activate one particular gene but it is likely that the folding and packaging of the DNA would be important.

It is easier to speculate about possible mechanism for a passive hypomethylation system. The first would be a physical blocking of the methylase enzyme (Kunnath and Locker, *op cit*). This could be by a site-specific determinator protein (Razin and Szyf, *op cit*) but is so, it is difficult to comprehend why some sites, which have no apparent role in gene expression, are also hypomethylated.

A more plausible mechanism would be if the failure to methylate a specific site occurred as a result of the activation of certain genes (ie their transcription). Bird (1984) has suggested that it is the activation of a gene which leads to a demethylation, which in turn

relaxed control on the gene (ie the demethylation will facilitate the continued expression of a gene, once activation has occurred). Such a mechanism is perhaps indicated by the oestrogen induced expression of vitellogenin in the liver which preceeds the observed hypomethylation of its 5' flanking sequence (Wilks *et al* ,1982; Meijilink *et al* , 1983). Similarily, in chicken, precursor cells, hypomethylation of the globin genes was not detected before the onset of globin synthesis (Groudine *et al* , 1981). The drawback to this theory is that although oestrogen does induce the expression of vitellogenin in the liver and a subsequent 5' hypomethylation is observed, this same hypomethylation is also observed in the oviduct which does not synthesise vitellogenin (although it does respond to oestrogen). Therefore, it is not necessarily reasonable to imply that it is the transcription of the vitellogenin gene, which is causing its hypomethylation.

Szyf *et al* (1984) have proposed a different model for a mechanism by which site-specific hypomethylation can occur. This model suggests that a change in the methylation capacity causes overall changes in the extent of methylation of cellular DNA. The methylation capacity (C) of a cell, is a function of the intracellular level of methylase activity (ie C = M.T where M = number of methylase molecules and T = turnover number of methylase). A further factor to be taken into consideration is the methylation quotient (Q) (where Q = C/N and N is the rate of emergence of newly replicated methylatable sites) (*ibid*). Thus, under conditions where Q is less than 1.0 (eg a high level of replicating sites), hypomethylation will result.

This theory is supported by experiments utilising the cytosine

analogue, 5-azacytidine. This analogue contains a nitrogen atom replacing the carbon atom at position 5 of the pyrimidine ring (see figure 4a), thus preventing the acceptance of a methyl group. Treatment of mouse cell lines with 5-azacytidine has resulted in the differentiation of the cells (Taylor and Jones, 1979). It was demonstrated that the continued presence of 5-azacytidine was not required and that the change in gene activity was clonally heritable. It was later found that 5-azacytidine caused undermethylation of the DNA (Jones *et al* , 1980). Even low levels of incorporation of 5-azacytidine into DNA, reduces DNA methylase activity in cell extracts for 1 or 2 days after treatment (Taylor and Jones, 1982).

Careful interpretation is needed when dealing with experiments involving 5-azacytidine. The amount of demethylation produced by the drug is far in excess of the amount of DNA substitution by 5-azacytidine, which may suggest that replacement of 5mC is not the major element in demethylation. (Razin and Cedar, 1984) or it may reflect the ability of the drug to 'trap' the methylases and thus block their action (Doerfler, 1983).

The action of 5-azacytidine has also been shown to be apparently specific to certain areas of the genome (Razin and Cedar, *op cit*). Groudine *et al* (1981a), were able to demonstrate whilst investigating the 5-azacytidine induced demethylation of chicken endogenous viral genes (AEV), that the globin genes in the same cells remained normally methylated. In addition, whilst 5-azacytidine gives increased expression of HbF(α) genes in adult erythrocytes, other cytotoxic agents (eg hydroxyurea) which interfere with cell division but do not

inhibit methylation give the same result.

Another methylase inhibitor, ethionine (a methionine analogue), is an effective inducer of globin genes in Friend erythroleukemia cells (Christman *et al* , 1977). The results obtained from experiments using the various methylase inhibitors thus imply that the intracellular level of methylase is not in a large excess (Razin and Szyf, 1984). Therefore, under these conditions the methylation capacity of a cell will be reduced and the methylation quotient will be less than 1.0 and hypomethylation will result. Hypomethylation will initially occur at the sites with the lowest affinity for the methylase and then at the sites with an intermediate affinity for the enzyme and finally at the sites with the highest affinity for the enzyme (Szyf *et al* , 1984).

Razin and Szyf (*op cit*) suggest that a change in the affinity of a site (perhaps as a result of a chromatin structural change), may explain the oestrogen induced hypomethylation event, observed in the flanking region of the chicken oviduct vitellogenin gene, previously discussed (Wilks *et al* , 1982).

1:4:6 Methylation and possible DNA structural changes

In 1980, Bird found that on average CG dinucleotides are present in vertebrate DNA at a frequency of only one quarter of that predicted by random distribution ie in effect there was a 'CG-suppression'. One explanation for this effect may be the 3-fold increased mutability of a methylated cytosine residue compared with an unmodified residue (Ehrlich and Wang, 1981). 5mC can be deaminated to thymine and a repair mechanism may replace the now mismatched opposite guanine, with

an adenine residue. Thus, a methylated $\overset{}{\underset{}{}}$ dinucleotide can mutate to a $\overset{}{\underset{}{}}$. Such a mechanism could explain the lower occurrence of the CG dinucleotide.

McClelland and Ivarie (1982) were able to show that the CG dinucleotides were non-randomly distributed along mammalian genes. It was found that the gene regions were deficient in CGs and had an excessive amount of TG + CA dinucleotides. Tykocinski and Max (1984) were able to demonstrate that certain MHC genes had regions where CG suppression was apparently absent (5' to intron 3), ie certain regions existed which had a cluster of CG dinucleotides. Max (1984) and Tykocinski and Max (*loc cit*) have suggested that some CG-rich regions are associated with DNA segments that remain unmethylated in germ-line DNA and other tissues. However, Adams and Eason (1984b) argue that rather than a failure to methylate being the cause for these CG-rich clusters, it is a failure to deaminate 5mC which allows such regions to exist. They point out that regions with a G + C content of over 60% have a greater stability of double helix, even when the cytosines are methylated. Deaminations (*in vitro*) normally only occur on single stranded DNA. Therefore, if a G + C-rich region is held in a tight helix, deamination is unlikely to occur and hence 'CG suppression' will not be observed. The high G + C content may be preserved either by chance or selection.

The *in vitro* methylation of cytosine residues in synthetic polymers has indicated that these helicies will form the left handed, Z-form, at physiological salt condition. Polymers with unmethylated cytosine will normally only form a Z-helix in the presence of a high

salt concentration (Behe and Felsenfeld, 1981). However, it is thought that a DNA sequence, CCGG, may favour an A-form helix (Conner *et al* , 1982). It may even sterically inhibit the B-Z conversion (Doerfler, 1983; Reeves, 1984). Thus, it has been suggested that certain modifications may stabilise certain transitions. Nordheim *et al* (1981), using anti-Z DNA antibodies which reacted specifically with interband regions of *Drosophila* polytene chromosomes, suggested that the Z-form of DNA may be transcriptionally inactive. Nickol *et al* (1982) were able to show that *in vitro* the Z-form of DNA bound histones but did not form nucleosomes. They therefore suggested that if Z-DNA were present in eukaryotic nuclei then it would disrupt normal chromatin structure. Santoro *et al* (1984) have demonstrated that Z-DNA in the flanking regions of genes can have a strong inhibitory effect on eukaryotic tRNA transcription.

Rich and coworkers have shown that Z-DNA-binding proteins (ie proteins which both bind specifically to Z-DNA but can also flip the B-form into the Z-configuration and hold it there), are present in a very high level in wheat germ cells and at low levels in *Drosophila* (Rich, 1983). These proteins are too large to be histones.

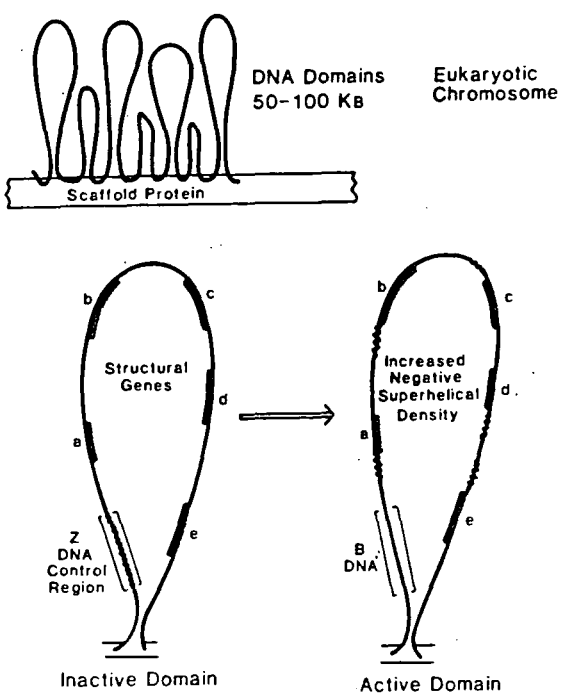
The degree of DNA supercoiling is considered to be an important mechanism for the regulation of gene expression in prokaryotes. All closed circular DNA appears to be negatively supercoiled (ie the two helical strands are relatively slightly underwound with respect to the fully relaxed state). Thermodynamically, negatively supercoiled DNA is in a higher free-energy state than the relaxed form (therefore, any

process which is going to decrease the degree of supercoiling will be energetically favoured) (Reeves, 1984).

Rich (*loc cit*) has envisaged that a system, similar to that proposed for prokaryotes, functions in eukaryotes. Linear, supercoiled 'loops' or 'domains' of DNA (perhaps containing control regions and structural genes and being between 50-100kb in length) could be isolated from one another by being anchored at either end of the loop, by the scaffold protein or nuclear matrix (fig.7a). These domains may then be activated independently of each other (fig.7b), by the induction of a change in the negative supercoiling, which induces a structural change from the inactive Z-DNA to the active B-DNA, in the control regions. This structural change in effect, may slightly uncoil the structure of a portion of DNA, thus making it accessible for RNA polymerase II. According to this theory, a change in the structural form of DNA in one stretch, could be 'felt' by a distant segment of DNA, as long as it is present in the same domain.

Nordheim and Rich have found in SV40 DNA, that Z-DNA induced by negative supercoiling, is only formed in a transcriptional enhancer segment, where alternating purine-pyrimidine sequences are clustered (Nordheim and Rich, 1983). They found that in general, potential viral Z-DNA forming sequences were all located in enhancer regions, which suggests that they might have a biological role (Kolata, 1983). In addition, it has been found in supercoiled plasmids the junctions between Z-form and B-form DNA are particularly susceptible to cleavage by single-strand-specific nuclease (eg Bal 31) (Kilpatrick *et al* , 1983) and this may also relate to the formation of a DNAse I hypersensitive site in chromatin.

Fig.7. Rich's model for DNA-domain activation of chromatin: a) Representation of DNA domains present in the eukaryotic chromosome (as they are visualised in the election microscope once histones have been removed); b) Activation of an individual domain containing five structural genes (taken from Reeves (1984))



The enhancer is a region which increases access for RNA polymerase to the promoter. Therefore, methylation of cytosine residues in these regions, may induce the DNA to change into a Z-helix and alter the access of the polymerase to the promoters. When one considers that

gene regions generally suffer from a CG suppression effect, then any clustering of CG's could have a significant effect on the form adopted by the DNA structure and the fact that such clusters of CG dinucleotides have been reported in a variety of systems, may indicate that they serve a biological function.

1:5 Summary and Objectives of Project

1:5:1 Summary

It is probable that methylation is only one of a complex and interacting set of potential regulatory mechanisms of eukaryotes. In the gene systems examined where no apparent correlation is observed between DNA methylation and gene regulation, it is possible that DNA methylation could be affecting essential biological functions directly or causing or stabilizing structural changes in the DNA or chromatin (ie where the methylation seems to be a necessary but not a sufficient requirement for gene expression).

The developing pea cotyledon provides us with a well documented system with several interesting features: the storage protein genes are tissue-specific in their expression; during the latter phase of seed development, cell division has ceased but endoreduplication occurs; a mutant line of *Pisum sativum* exists, which has on abnormal vicilin subunit synthesis.

1:5:2 Objectives of the project

The overall aim of this project is to investigate the role of

cytosine methylation in gene expression in *Pisum sativum* . Much is documented about cytosine methylation in vertebrates, and relatively little in relation to higher plants. The following areas are to be investigated:-

1) to look for evidence of methylated and unmethylated domains in genomic DNA of *Pisum sativum*
2) to observe any developmentally related changes in the methylation status of the ribosonal genes, using the isoschizomeric pairs Msp I/Hpa II and Bst NI /Eco RII
3) to observe any developmentally related changes in the methylation status of the legumin gene family, using the isoschizomers Msp I/Hpa II, cDNA probes and specific legumin gene probes
4) by analysis of established sequence data of a legumin gene and a legumin pseudogene, investigate the frequency of occurrence of CG dinucleotides and look for evidence of 'CG-suppression'
5) by analysis of established sequence data of a legumin gene and a legumin pseudogene, investigate the microenvironment of the CG dinucleotide
6) to observe any developmentally related changes in the methylation status of the vicilin gene family, using the isoschizomers Msp I/Hpa II and various cDNA probes
7) to look at a *Pisum sativum* vicilin deficient mutant and to investigate a possible altered methylation pattern.

2: MATERIALS AND METHODS

2:1 Materials

2:1:1 Biological Materials

Seeds of *Pisum sativum* L. variety 'Feltham First' were obtained from Sutton Seeds Ltd, Reading Berks, UK and those from pea lines 200 and 5478 were obtained from Dr S Blixt, Weibullsholm, Landskrona, Sweden.

2:1:2 Chemicals and non-biological materials

All chemicals (exept those specifically indicated below) were obtained from either British Drug Houses Ltd, Poole, Dorset, UK or Koch-Light Laboratories Ltd, Colnbrook, Berks UK and were of the purest grade available.

The various antibiotics used for plasmid selection, acridine orange, bovine serum albumin (BSA), dextran sulphate (sodium salt), dithiothreitol (DDT), egg white lysozyme, ethylene glycol bis (β-amino-ethyl ether)-N,N,N',N'-tetraacetic acid (EGTA), ethidium bromide, glyoxal, heparin (sodium salt grade II from porcine intestinal mucosa), herring sperm DNA, pronase 'K' and pronase 'P', sodium N-lauryl sarcosine, sucrose and 'Trizma' base [Tris (hydroxy-methyl) amino methane] were all obtained from Sigma Chemical Company, Poole Dorset, UK. 3,5-diaminobenzoic acid was obtained from Aldrich Chemical Company, New Road, Gillingham, Dorset UK.

Bactotryptone, bactoagar and yeast extract came from Difco Laboratories, Detroit, Michigan, USA.

Restriction enzymes were obtained from either Bethesda Research

Laboratories (BRL), Cambridge, (SV40 DNA, low melting point agarose and nuclease-free bovine serum albumin also came from this source) or Boehringer Mannheim Corporation Ltd. (BCL, Bell Lane, Lewes, East Sussex, UK, or DuPont (UK) Ltd, (New England Nuclear products), Stevenage, Herts, UK, or New England Biolabs, CP Laboratories Ltd, Bishops Stortford, Herts, UK.

'Repelcot' siliconising fluid was obtained from Hopkins and Williams, Romford, UK.

Ficoll 400 and Sephadex G-50 superfine were obtained from Phamacia Fine Chemicals, Uppsala, Sweden.

Agarose, high gelling temperature agasose and 'Gellbond' were obtained from Miles Laboratories Ltd, Stoke Poges, Slough, Berks UK.

Guanidinium thiocyanate and guanidinium hydrochloride were obtained from 'Fluka' - Fluorochem Ltd, Peakdale Road, Glossop, Derbyshire, UK.

Nick translation kits and ^{32}P-dCTP were obtained from Amersham International p.l.c. White Lion Road, Amersham, Bucks UK.

Schleicher and Schull nitrocellulose filters (BA 85,0.45µm were obtained from Anderman and Company Ltd, Kingston-Upon-Thames, Surrey, UK.

Diethylaminoethylcellulose (DEAE-cellulose paper (DE 81), glass fibre filters (GF-C) and 3MM paper were all obtained from Whatman Ltd, Maidstone, Kent, UK.

Dialysis tubing (size 1 - 8/32") came from Medicall International Ltd, 239 Liverpool Road, London UK.

Fuji RX-100 X-ray film was obtained from Fuji Ltd, Swindon, Wilts., UK.

Dupont 'Lightning Plus Cronex' intensifying screens were obtained from E.I. DuPont de Nemours and Co (Inc), Photoproducts Department, Wedgward Way, Stevenage, Herts UK.

pDUB6, pDUB7, pDUB9, pDUB24, pDUB25, pDUB27 and λleg1, and λleg2 were provided by Dr R.R.D. Croy and *E.coli* RNA, cauliflower mosaic virus RNA and pea Poly(A)$^-$ RNA were kindly donated by Dr I.M.Evans, Department of Botany, University of Durham, UK.

Sequence data for legumin gene A was taken from Lycett *et al* , 1984 whilst data for the legumin pseudogene D, was kindly provided by Mr M.D. Levasseur, Department of Botany, University of Durham UK.

2:2 Methods

2:2:1 Reagents and equipment

Reagents were either autoclaved after preparation or prepared using sterile water and containers. Plastic microfuge tubes and glassware used in association with nucleic acids, were siliconised whth 'Repelcote' and autoclaved before use. 0.1% 8-hydroxyquinoline was included in all phenol used, to enable it to be stored at 4°C.

2:2:2 Ethanol precipitation and resuspension of nucleic acids

DNA was precipitated with 0.1 vol. acetate solution (5m-ammonium acetate; 100mM-magnesium acetate; pH5.6) and 2 vol. cold absolute ethanol (-20°C). Total RNA was precipitated with 0.1 vol 3M-sodium

acetate and 3 vol. cold absolue ethanol (-20°C).

The precipitated nucleic acid was pelleted by centrifugation (for genomic DNA and total RNA Sorvall RC-5B Superspeed centrifuge; HB4 rotor; 10,000r.p.m.; 20min; 0°C. For gene fragments or plasmid inserts: MSE Microcentaur; 15min; 4°C). The pelleted DNA was washed twice with 80%(v/v) ethanol (total RNA was washed twice with 70% (v/v) ethanol) and the nucleic acids were repelleted by centrufugation as before.

The washed pellet was dried *in vacuo* and resuspended in sterile water. Small fragments of DNA and the total RNA were resuspended on ice (30min. to 2h.). Genomic DNA was resuspended using an ice bucket on a shaker. (Resuspension under these conditions took 2-3d but the resultant DNA was less sheared than when resuspended overnight, on a rotary mixer, at 4°C).

2:2:3 Spectrophotometric analysis of nucleic acids

After resuspension the concentration of the nucleic acids was measured using a Pye-Unicam SP8-150 u.v./vis scanning spectrophotometer. An $O.D._{256}$ of 0.020 (in a 1cm light path, in a quartz cuvette) corresponds to a DNA concentration of about 1µg/ml. Under similar conditions, an RNA concentration of 1µg/ml corresponds to an $O.D._{256}$ of 0.025. The $O.D._{256}$ of 1µl aliquots of the samples, dispersed in 1ml of sterile water, were routinely assessed. In addition, the purity of the nucleic acids was checked by scanning the sample from 320nm to 200nm. A ratio of $O.D._{260}$: $O.D._{280}$ of greater than 1.8 was used to confirm that the nucleic acid preparation was

free from contamination (eg protein or phenol).

It was not found possible to obtain accurate concentrations for genomic DNA samples by this method and consequently their concentration was also determined by a fluorometric method (see Section 2:2:6).

2:2:4 Growth of biological materials

Pea seeds were germinated for 4 days in a 25°C temperature controlled spray room, transferred to water culture bottles in a 'Warren Sherer' growth cabinet, model CEL 511-38 and grown under controlled environmental conditions of light (17h daylength) and temperature (25°C day; 18°C night), similar to that described by Evans *et al* (1979). The pea plants were grown with a liquid nutrient source ('Phostrogen').

The pea plants were grown either for 9 days, when leaf material was harvested, or to maturity when pea seeds were harvested at various states of development. The testas and embryonic axes were discarded and the cotyledons (or 9 day old leaves) were frozen under liquid nitrogen and stored at -80°C until required for DNA extraction.

2:2:5 Extraction and purification of genomic DNA

Two centrifuges were used during the extraction and purification of genomic DNA. Where the term 'Sorvall' is used, unless stated otherwise this refers to a Sorvall RC-5B superspeed centrifuge and an HB4 rotor; where the term 'Prepspin' is used, this refers to an MSE Prepspin 65 centrifuge and a 10 x 10ml rotor.

DNA extraction was by a method similar to that of Graham (1978). Frozen tissue was finely ground under liquid nitrogen, with a pestle and mortar. Homogenising buffer (0.1M-NaCl; 0.025M-EDTA (ethylene diaminotetra-acetic acid) pH 8.0; 2% (w/v) SDS (sodium dodecyl sulphate); 0.1% (v/v) diethyl pyrocarbonate) was added until the extract became a thick suspension. The volume was measured, 5M-sodium perchlorate added to give a final concentration of 1M, quickly followed by 0.5vol. of predistilled phenol and 0.5vol. of chloroform/ octanol (99:1(v/v)).

This suspension was mixed on a rotary shaker (50min; 4°C), centrifuged in corex glass tubes (Sorvall; 8,700 r.p.m.; 5min; 10°C) and the supernatant re-extracted with an equal volume of chloroform/ octanol (99:1 (v/v)) and recentrifuged (Sorvall; 8,700r.p.m.; 1min; 10°C).

2vol. cold ethanol (-20°C) were added to the supernatant, the precipitated DNA collected on a glass hook and surplus ethanol removed *in vacuo* and the DNA dissolved in resuspension buffer 'RB' (50mM-Tris/HCl [Tris (hydroxymethyl) aminomethane] ; 10mM-EDTA; pH 8.0), overnight on a rotary shaker, at 4°C.

Self digested Pronase 'P' (RB buffer; 2H; 37°C) was added to the resuspended DNA, to a concentration of 500µg/ml and incubated for a further 3h at 37°C.

The DNA was isolated by caesium chloride density gradient centrifugation (wt. of CsCl added = 0.94 x wt. of supernatant), in the presence of ethidium bromide (250µg/ml) (Prepspin; 60,000r.p.m.; 36h; 15°C; or a Sorvall ultracentrifuge OTD 65B; vertical rotor; 44,000

r.p.m.; 20h; 15°C). The semi-purified DNA band, removed from the gradient, using a syringe and wide bore hypodermic needle, was subjected to a second density gradient centrifugation, under similar conditions.

The purified DNA band was again removed from the gradient and repeatedly partitioned with RB-saturated amyl alcohol, until all the ethidium bromide had been removed. The DNA was dialysed (36h; 4°C) in freshly prepared dialysis tubing (size 1-8/32", boiled for 20min, 10mM-EDTA, then rinsed in distilled water), with many changes of buffer.

1 vol. phenol was mixed with the contents of the dialysis tubing (Sorvall; 8,000r.p.m.; 5 min; 4°C) and the aqueous phase re-extracted with 1 vol. ether to remove residual phenol. This was allowed to stand (5 min) and most of the upper organic phase was removed by pipette and the residual ether was 'blown off' with nitrogen. The DNA was ethanol precipitated, resuspended (see Section 2:2:2) and its concentration was determined (see Sections 2:2:3 and 2:2:6).

2:2:6 Fluorometric estimation of genomic DNA concentration

Fluorometric analysis of DNA concentration was performed by the method of Thomas and Faraquar (1978), ie ethanol precipitation, in the presence of nuclease-free bovine serum albumin (BSA), of the sample genomic or standard calf thymus DNA, followed by incubation in the presence of 3,5-diaminobenzoic acid (DABA) at 60° for 30 mins. The method was exactly as published except that 20µl of DABA were added to each assay and the incubation was stopped with 2ml of 1M-HCl. The

fluorescence was measured using a Baird Fluoripoint spectro-fluorimeter, set at 400nm excitation wavelength and 505nm emission wavelength.

2:2:7 Restriction endonuclease digestion

DNA samples were digested to completion, in the buffer recommended by the manufactures, using a 5- to 10-fold excess of restriction enzyme. Nuclease-free BSA was included in restrictions (0.1µg BSA/10ml reaction volume), as recommended by BRL. Enzymatic reactions were terminated by heat inactivation of the enzyme (70°C; 10 mins) (Bst NI Normally restricts at 60°C and therefore these samples were transferred to an ice bucket at the end of their restrictions).

Double restriction experiments were performed by heat in activation of the first enzyme and digestion of aliquots of the first reaction mixture by a second restriction enzyme, under appropriate buffer conditions. This second enzyme was in activated as previously described.

2:2:8 Use of an SV40 internal control

For some experiments, 0.1µg SV40 DNA was included in the initial reaction mixture, to monitor the activity of the various isoschizomers used and to ensure that any differential digestion observed, was due neither to a faulty batch of enzyme nor to the presence of inhibitors in the genomic DNA preparation.

2:2:9 Agarose gel electrophoresis

Restricted DNA fragments were separated by electrophoresis in horizontal, submerged agarose gels, containing 0.5-2%(w/v) agrose; 1μg/ml ethidium bromide and a Tris/EDTA buffer (40mM-Tris/acetic acid, pH 7.7;1mM-EDTA). Before loading onto the gel, restricted DNA fragments were mixed with 0.5 vol of agarose loading beads (10mM-Tris/HCL, pH 8.0; 10mM-EDTA; 31.25% (v/v) glycerol; 0.2% (w/v) agarose; 0.1w/v bromophenol blue; 0.1% (w/v) xylene cyanol; 0.2% (w/v) fast orange 'G', autoclaved and when set extruded through a syringe and fine needle). For experiments which included a densitometric scanning of the gel negative loading beads were prepared which contained fast orange 'G' as the only dye. This runs in front of the smallest DNA fragments and thus does not interfere with quantification of the distribution of DNA fragments.

Gels separating restricted genomic DNA fragments were run overnight (16-18h; 30V), whilst gels separating restricted plasmid DNA fragments were run much faster (3-4h; 100-120V).

Separated DNA fragments were visualised using ultraviolet (u.v)(254nm) induced fluorescence of the DNA-ethidium bromide complex and the fluorescence was photographed through a red filter, using a "Polaroid" land camera and 3,000 ASA positive film (type 665). The negatives were fixed in sodium metabisulphite (10% (w/v)) and allowed to dry at room temperature.

2:2:10 Transfer of DNA from agarose gels to nitrocellulose filters.

The method used for the transfer of DNA to nitrocellulose filters

was essentially that of Southern (1975). The gels were treated in flat bottomed glass dishes. Gels which contained small DNA fragments (< 500bp) were initially depurinated (2 x 10min; 0.25M-HCl). All gels were denatured in 2 x 20min washes, with agitation, in denaturing buffer (1.5M-NaCl; 0.5M-NaOH; 1mM-EDTA), then neutralised in 3 x 20min washed (with agitation) in neutralising buffer (3M-NaCl; 1mM-EDTA; 0.5M-Tris/HCl; pH 7.0), or until the fast orange 'G' band had resumed its bright orange colour.

Whilst the gel was being neutralised, a sheet of nitrocelulose paper was prewetted (by floating on top of distilled water, to exclude all the air, and then submerging in the water for 10 min), prior to being equilibrated in 20 x SSC (3M-NaCl; 0.3M-sodium citrate; pH 7.0) for 15min.

The blotting apparatus was essentially a glass plate held above a resevoir containing 20 x SSC. A piece of Whatman 3MM paper (25cm x 45cm) stretched from the resevoir over the glass plate and back into the resevoir. This 'wick' was presoaked with 20 x SSC and all air bubbles, between the wick and plate were gently excluded. The prepared gel was placed on the prewetted wick, again ensuring that no air bubbles were trapped between the wick and gel and the gel surrounded by 'clingfilm' before being overlaid with the equilibrated nitro-cellulose filter. This ensured that all transfer buffer reaching the filter, had passed through the gel and had not been transferred directly from the wicks. The filter was checked to see that no air bubbles were trapped beneath it and was then overlaid with two pieces of Whatman 3MM paper (cut slightly larger than the filter and

presoaked in 20 x SSC) and four pieces of dry Whatman 3MM paper. Three layers of disposable nappies were then placed on top of the Whatman paper. A glass plate and a 5l conical flask containing about 1200ml of water were used as a weight, to ensure the capillary transfer of DNA fragments from the gel to the nitrocellulose filter.

Blotting was allowed to proceed for 20h, when the volume of water in the conical flask was increased to 2l and blotting allowed to continue for a futher 1h. The blotting apparatus was dismantled and the nitrocellulose filter was removed, with the gel attached. The limits of the gel and the position of the wells were marked on the filter in 'biro' ink. The gel was removed from the filter and placed on the u.v. transilluminator to check that transfer of the DNA had occurred. The nitrocellulose filter was blotted dry between two sheets of dry Whatman 3MM paper then baked between four sheets of dry Whatman 3MM paper and two glass plates, in a vacuum oven (80°C; 2h). After baking, the filter was stored at room temperature until used.

2:2:11 Densitometric scanning of the gel negative

Individual tracks, cut from the negatives of Bst N1/Eco RII digests of cotyledon and leaf genomic DNA, were scanned on a Pye Unican SP500 spectrometer. The negatives obtained from Msp 1/Hpa II digested genomic DNA were scanned on an LKB (Bromma) 2202 Ultroscan laser densitameter with an LKB (Bromma) 2220 recording integrator.

Analysis of the densitometer scans, was essentially that of Tanford (1961) as modified by Kunnath and Locker (1982a). The scans were divided into molecular weight intervals, calculated from the

migration of standard DNA molecules. The percentage composition of the digestion was calculated from the areas of these intervals as a percentage of the total area of the scan. Cumulative molecular weight distributions, mass average weights (M_w) and number average molecular weight (M_n) were calculated (see below). M_w analysis is weighted towards larger molecules, whilst M_n analysis is weighted towards smaller molecules. However, because M_n is related to the number of molecules it is also related to the number of sites for a particular restriction enzyme. Therefore, the values obtained can be used to obtain an estimation of the degree of methylation of the genomic DNA.

When determining M_w and M_n the molecular weight midpoint (M_i) was used as the average molecular weight for the interval (i). The mass fraction (W_i) and number fraction (X_i) were calculated for each interval. The mass average molecular weight (M_n) and the number average molecular weight (Mn) were calculated ($M_w = \Sigma W_i M_i$ and $M_n = \Sigma X_i M_i$) (Tanford, *op cit*). The percentage of methylation was calculated (methylation = $1-((M_{nMsp\ I}) / (M_{nHpa\ II})) .100$) (Kunnath and Locker, *op cit*). To observe changes in the distribution of the digested DNA, the ratio 'r' was calculated ($r=M_w/M_n$) (*ibid*).

M_w and M_n were calculated using a 'Supercalc' programme on an Apricot PC1 and also using an 'Abacus' programme on a Sinclair QL.

2:2:12 Extraction of total RNA from pea leaves and cotyledons

The method of Langridge *et al* (1982) was the preferred method used.

2:2:12:1 Isolation of total RNA in the presence of guanidinium thiocyanate

Total RNA was isolated essentially by the method of Chirgwin *et al* (1979) as adapted by the EMBO course on Ti plasmids, at Gent in August 1982. Reagents and basic procedures are as for Chirgwin *et al* but the lengths and conditions of centrifugation have been altered.

RNA was extracted in the presence of guanidinium thiocyanate. The crude extract was cleared (Sorvall superspeed RC5B; HB4 rotor; 9,000rpm; 30min; 4°C). The cleared supernatant was layered onto 6ml CsCl cushions in 25ml polycarbonate tubes and the RNA was pelleted through the cushion (Prepspin 65 centifuge; 3x35ml rotor; 25,000 rpm; 24h; 40°C). The supernatant was removed by water aspiration, the pellet resuspended (2min; 68°C) in guanidinium hydrocholoride solution and insoluble material removed (Sorval RC5B; HB4 rotor; 6,000 rpm; 15 min; 40°C)

RNA was precipitated as described by Chirgwin, washed twice in 70% ethanol (Sorvall RC5B; HB4 rotor; 9,000 rpm; 15 min; 0°C), resuspended in sterile water (see section 2:2:2) and the concentration determined (see section 2:2:3)

2:2:12:2: Isolation of total RNA using hot SDS

The method used was essentially the same as Hall *et al* (1978), with extraction of RNA in the presence of a hot (100°C) SDS-borate buffer with dithiothreitol (DTT). The reagents and initial methods used were the same as those published but the latter procedures were modified from those published.

Centrifugations used a Sorvall RC-5B superspeed and HB4 rotor (10,000 rpm; 10 min; 0°C). The LiCl precipitated and washed pellet was dissoved in 2% potassium acetate, heat shocked (2min; 68°C) and insoluble material was removed by centifugation.

The RNA was precipitated from the supernatant with 3 vol. ethanol (overnight; -20°C), pelleted and dissolved (on ice with agitation), in TE buffer (10mM-Tris/HCl, pH 7.5; 1mm-EDTA). The RNA was extracted twice with phenol/chloroform/isoamylalcohol (25:24:1), ethanol reprecipitated (see section 2:2:2) and the concentration determined (see section 2:2:3).

2:2:12:3 Rapid method for the extraction of total RNA

This method was the same as that used by Langridge *et al* (1982) to isolate zein precursor mRNAs from maize endosperms. It involves extraction of the RNA in a sucrose-Tris buffer containing iodoacetamide. After a brief centrifugation, to remove cell debris, the supernatant is rapidly mixed with 2%(w/v) SDS in phenol/chloroform (1:1). Langridge *et al* (*op cit*) emphasise that the time from the addition of buffer to the tissue, to the extraction of the supernatant with SDS/phenol/chloroform, should be between 3 and 5min. During the extraction of total RNA from pea lines 200 and 5478, this time varied from 4½ to 6 min. Total RNA was recovered from the aqueous phase as described by Langridge *et al* , ethanol precipitated (see section 2:2:2) and concentration determined (see section 2:2:3).

2:2:13 Electrophoreis of RNA on glyoxal gels and subsequent transfer to nitrocellulose filters

2ml of 30% (w/v) glyoxal were deionised by stirring (under nitrogen) with 1g of Amberlite resin. RNA samples were glyoxalated in a manner similar to that of McMaster and Carmichael (1977) and Thomas (1980), by adding in order: 25µl redistilled dimethyl sulphoxide (DMSO); 2.5µl 0.2M sodium phosphate buffer, pH7.0; 7.1µl deionized glyoxal; sample RNA in 15.3 µl H_2O. This mixture was incubated to denature the RNA (60 min; 50°C).

An agarose gel was prepared (1.5%(w/v) high gelling temperature agarose in running buffer (10mM-sodium phosphate buffer; pH6.8)). The gel was poured onto the hydrophillic side of a sheet of 'Gellbond', allowed to set and was transferred to a submarine electropheresis tank.

The long edges of the Gellbond were weighted down with two glass rods and the gel submerged under running buffer. The glyoxalated RNA samples were mixed with 0.5vol of glyoxal agarose loading beads (10mM-sodium phosphate buffer, pH 6.8; 31.25%(v/v) glycerol; 0.2%(w/v) agarose; 0.1%(w/v) bromophenol blue) and loaded onto the gel. *E coli* RNA, cauliflower mosaic virus RNA and pea Poly(A) RNA were used as standards. Electrophoresis was carried out for 4h at 120V, with recirculation of the running buffer.

The gel was stained with freshly prepared acridine orange in 10mM-sodium phosphate buffer, pH6.8 (30mg/l; 15min; 4°C; in the dark) and destained (overnight; in the dark;4°C) in phosphate buffer. The u.v. induced fluorescence of the acridine orange-RNA complex was photographed in a manner similar to that employed for ethidium

bromide-DNA complexes (see section 2:2:9).

RNA gels which were required for 'Northern Blots' were not stained in acridine orange. The marker RNA tracks were removed, stained and photographed as a record, whilst the rest of the gel was immediately blotted, in a similar manner to that described for 'Southern Blots' (see section 2:2:10)(see also Thomas, 1980).

2:2:14 Plasmid Preparation Method

Two methods were used. The first was a miniprep method which isolated RNA-contaminated plasmid DNA in a quantity sufficient to enable identification of appropriate clones. The second method was a maxiprep of purified plasmid DNA.

2:2:14:1 Plasmid miniprep method

The method used was essentially that of Birnboim and Doly (1979). The initial subcloning of legumin gene fragments into a pUC8 vector, the sebsequent traansormation of *E. coli* and identification of transformed clones and the initial lysis of these cells were all done by Mr. D. Bown, Department of Botany, University of Durham.

The initial growth and lysis of the cells was as follows. The cells were grown overnight (37°C) in 2ml YT medium (YT medium is per litre: 8g bactotryptone; 5g yeast extract; 1g NaCL; 50mg ampicillin) Cells were pelleted (MSE Micro-Centaur; 30s), the supernatant discarded and 200 l of freshly prepared lysis solution added (25mM-Tris/HC1, pH 8.0; 10mM-EDTA; 50mM-glucose; 2mg/ml lysozyme). The tubes were mixed and left on ice (30min). 600 l of 2M-NaOH/1%(w/v) SDS was added, the tubes mixed and left on ice for a futher 5 min. The

chromosonal DNA was precipitated by adding 450µl of 3M-sodium acetate (pH 4.8) and leaving the tubes on ice for a further 1h. The cleared supernatant (MSE Micro-Centaur; 5 min) was ethanol precipitated and the pelleted plasmid DNA (MSE Micro-Centaur; 2 min) was redissolved in 200µl 50mM-Tris/HCL, pH 6.0; 100mM-sodium acetate and purified by a further ethanol re-precipitation before being finally dried *in vacuo* and resuspended in 100 µl sterile water and stored at -20°C (See section 2:2:2).

2:2:14:2 Maxi preparation of purified plasmid DNA

The method used was essentially that of Clewell (1972) and Katz *et al* (1977) and involved amplification of the plasmid in the presence of chloramphenicol followed by SDS-lysis of the cells.

Overnight (37°C) 'L broth' cultures were grown from plasmid bearing strains, stored in 20% glycerol at -80°C (L broth per litre is: 10g tryptone; 5g yeast extract; 5g NaCl; 1g glucose; pH7.0; 50mg ampicillin; 50mg tetracyclin). The overnight cultures were diluted (1:25) with fresh L broth (250ml), cultured for a further 6h when chloramphenicol was added to a concentration of 150µg/ml, to amplify the plasmid DNA overnight.

The harvested cells (MSE High Speed 18; 10,000rpm; 10min; 4°C) were resuspended in 5ml of 50mM-Tris/HCl, pH8.0 containing 25% (w/v) sucrose. Spheroplasts were formed by adding 1ml of freshly prepared lysozyme in sucrose/Tris (10mg/ml) and incubating with agitation (2min; 37°C; then 10min; 0°C). 5ml 0.2M-EDTA, pH8.0 was added (10min; 0°C) and the cells were finally lysed with 20% (w/v) SDS and the suspension was mixed gently at 20°C until it clarified. 3ml 5M-NaCl

was added and after mixing, the suspension was allowed to stand on ice for 1h. The extract was cleared, to remove cell debris and the bulk of the chromosomal DNA (M5E High Speed 18; 8,000rpm; 90min; 0°C), and 50% (w/v) polyethylene glycol (PEG) in 50mM-Tris/HCl, pH8.0 was added to a final concentration of 10% PEG and 0.5M NaCl. This was well mixed and allowed to stand on ice overnight.

The precipitated DNA was pelleted (Sorvall RC-5B; HB4 rotor; 3,000rpm, 10min, 4°C), redissolved in 5ml TE buffer (10mM-Tris/HCl, pH 7.5; 1mM-EDTA), purified by caesium chloride density gradient centrifugation, as previously described for genomic DNA (see section 2:2:5), ethanol precipitated (see section 2:2:2) and its concentration was measured (see section 2:2:3).

2:2:15 Recovery of DNA fragments from gels

Recovery of DNA fragments from gels, for the early part of this work, used a modification of the method of Dretzen *et al* (1981) using DEAE-cellulose (diethylaminoethylcellulose). This was not found to be a satisfactory method and recovery of DNA was low. Two other methods were then used. To isolate DNA fragments required for subsequent restriction and not needed for nick-translation, a method utilising low melting point (LMP) agarose was used. DNA fragments which were required as probes, were recovered by a method similar to that of Chen and Thomas (1980) which used perchlorate isolation on Whatman GF-C disks.

2.2.15:1 Recovery of DNA fragments using DEAE-cellulose (Whatman DE81)

The method used was identical to that described by Dretzen *et*

al (*loc cit*), except that the prepared pieces of DEAE-cellulose were dried in a vacuum oven (80°C; 1h) and stored at room temperature until required. In addition, because agarose gels were electrophoresed in the presence of ethidium bromide, it was not necessary to presoak the DEAE-cellulose in ethidium bromide. Recovery of the DNA was otherwise as described.

2:2:15:2 Recovery of DNA fragments using low melting point (LMP) agarose

The method used was adapted from one recommended by BRL. The agarose gel was prepared using LMP agarose and Tris/EDTA gel buffer (see section 2:2:9). The gel solution was cooled to below 37°C before pouring. The gel surround and comb were not removed until the gel was in the electrophoresis tank and submerged under Tris/EDTA buffer (this was found to be necessary because LMP gels were very fragile). The gel was pre-run, increasing the voltage gradually during a 10min period. Apart from this, electrophoresis and sample preparation was as normal (see sections 2:2:7 and 2:2:9).

After electrophoresis the gel was viewed on a u.v. transilluminator, the desired band was cut out with a sterile scalpel and placed in a 1.5ml eppendorf tube and the agarose was melted (10min; 65°C). 2vol. of 50mM-Tris/HCl, pH 8.0; 0.5mM-EDTA were added and the mixture incubated (15min; 37°C). The agarose was precipitated with an equal vol. of buffer saturated phenol (MSE Micro-Centaur; 5min), the upper aqueous phase re-extracted with an equal vol. of phenol (MSE Micro-Centaur; 5min) and residual phenol in the aqueous layer was removed by partitioning with chloroform/isoamyl alcohol

(24:1). The DNA was ethanol precipitated (see section 2:2:2).

2:2:15:3 Recovery of DNA fragments using Whatman GF-C discs

This was adapted from the method of Chen and Thomas (1980). Fragments were separated on a normal agarose gel (0.5%). The gel was viewed on a u.v. transilluminator and the gel slice containing the required DNA fragment was cut out, with a sterile scalpel, transferred to a 1.5ml eppendorf tube and dissolved in 2-3vol 8M-sodium perchlorate.

Two superimposed 6mm Whatman GF-C disks were placed on two pieces of Whatman 3MM paper on a disposable nappy. All solutions were applied to the GF-C disks in 15-20µl aliquots and each aliquot allowed to soak through the GF-C disks, into the lower absorbant material, before the next aliquot was applied. The GF-C disks were prepared with 300µl of 6M-sodium perchlorate before the 8M-sodium perchlorate/DNA/agarose was applied. 2ml of 6M-sodium perchlorate in TE buffer and 1.5ml absolute ethanol (-20°C) were applied to the disks, to remove residual agarose and perchlorate.

The GF-C disks, with their bound DNA, were dried at room temperature (20min) and the DNA eluted by incubating them twice, in a 0.5ml eppendorf tube, with 20µl sterile water (37°C; 1h). The eluted DNA was collected by piercing the bottom of the tube and collecting the elutant in a 1.5ml eppendorf tube (MSE Micro-Centaur; 5min). The DNA was stored at -80°C until required for nicks translation.

2:2:16 ^{32}P-dCTP labelling of DNA probes (nick-translation)

Radioactive probes were ^{32}P-dCTP labelled by nick translation, by

the method stated in the Amersham nick translation kit pamphlet p1/86/81/4. Typically 0.1μg of DNA was nick translated. A modified 'procedure D' was used (ie 0.1μg of DNA in a reaction volume of 20μl) for cDNA and legumin gene fragment probes, to obtain on average a specific activity of 2-4 x 10^8 cpm/μgDNA. For ribosomal (pHA1) and SV40 probes and legumin gene probes required for restriction mapping, a modified 'procedure A' was used (0.1 μg of DNA in a reaction volume of 20μl), to obtain on average a specific activity of 1-2 x 10^7 cpm/μgDNA.

Nick translation was carried out at 15°C for 2h. (A time course study of standard λDNA had indicated that maximum incorporation of ^{32}P-dCTP was achieved after 2h). The reaction was terminated with 50μl 0.2M-EDTA, pH 8.0 and the labelled DNA was separated from the unincorporated nucleotide, on a Sephadex G-50 superfine column, prepared in a flat-bottomed 10ml syringe, with a siliconised glass wool plug at the bottom. The Sephadex was freshly swollen (1h; 90°C) in Amersham's recommended column buffer (50mM-Tris/HCl, pH 7.5; 10mM-EDTA; 150mM-NaCl; 0.1% (w/v) SDS). The column was equilibrated with column buffer before the reaction mixture was loaded. 0.4ml fractions were collected and those containing the labelled probe were initially detected with a Geiger-Muller counter. The hottest fractions were identified by counting 1μl aliquots in 5ml scintillation fluid (toluene/triton (2:1) with 5g/l 2,5-diphenyloxazole (PPO) on a Packard Tricarb scintillation counter, which has a 90% efficiency for counting ^{32}P. This enabled the specific activity of the probe to be calculated. The hottest fractions were pooled (typically about 1.5ml) and used directly for the hybridization of nitrocellulose blots.

Periodically, 1µl aliquots of the pooled fractions were trichloracetic acid (TCA) precipitated. The precipitates were collected on cellulose acetate filters (Millipore), rinsed with ethanol, dried in a vacuum oven (10min; 80°C) and counted using triton-free scintillation fluid (300mg/l 1,4-di-2(5-phenyloxazolyl)benzene (POPOP); 3g/l PPO; in toluene). This procedure was used to confirm the specific activity of the probes used.

2:2:17 Hybridization of the ^{32}P-labelled probe to DNA or RNA blots

Basically, two hybridization procedures were used for the hybridization of ^{32}P-labelled probes to 'Southern' blotted pea genomic DNA. The first was the Schleicher and Schull version employing SSC and Denhardt's solution. A completely different and novel hybridization system was also successfully used. This was Singh and Jones' (1984) heparin/dextran sulphate system. It has a simple protocol and a greatly reduced hybridization volume and therefore the probe is relatively more concentrated.

Hybridization of ^{32}P-labelled probes to RNA (Northern) blots was by a method similar to Thomas (1980) which, in addition to SSC and Denhardt's solution, also uses formamide and dextran sulphate.

Nitrocellulose filters were prehybridized and hybridized in heat-sealed plastic bags. The ^{32}P-labelled DNA probe was denatured by boiling for 10 minutes and cooling rapidly on ice, before being added to the hybridization solution. Air was excluded from the bag before it was resealed. After hybridization the filters were removed from the bags and were washed in sealable plastic boxes. The hybridized filters were then blotted dry between Whatman 3MM paper and allowed to dry at

room temperature between clean dry sheets of Whatman 3MM paper. Any individual modification to the hybridization system is mentioned, where appropriate, in the 'Results' section. Unless otherwise stated, DNA Southern blots were hybridized in the SSC system.

2:2:17:1 Schleicher and Schull's SSC/Denhardt's hybridization system for DNA-blots

The filter was prehybridized for 2-4h at 62°C in 100ml: 5 x SSC; 100µg/ml denatured herring sperm DNA (Maniatis *et al* , 1982); 5 x Denhardt's solution (1 x Denhardt's is 0.02% (w/v) BSA; 0.02% (w/v) Ficoll 400; 0.02% (w/v) polyvinylpyrrolidone). Denhardt's solution prevents non-specific hybridization of denatured or single stranded DNA to the nitrocellulose filter (Denhardt, 1966) but at concentrations of less than 5 x Denhardt's, does not reduce the specific annealing of denatured DNA to complementary DNA (Barinaga *et al* , 1981).

The prehybridization solution was replaced by 50ml of preheated hybridization solution (5 x SSC; 1.5 x Denhardt's, 100µg/ml denatured herring sperm DNA), the denatured ^{32}P-labelled probe added and the filter hybridized with agitation, for 30-40h at 62°C.

After hybridization the filter was washed (50-60°C in 250ml of solution) as follows: 2 x 15min, 2 x SSC, 0.1% (w/v) SDS; then 2 x 15min, 0.1 x SSC, 0.1% (w/v) SDS.

2:2:17:1 Singh and Jones' heparin/dextran sulphate hybridization system

The method used was the same as that published by Singh and Jones

except that filters were prehybridized for 3h and hybridized for 20h, both at 65°C. A note is made in the figure legend, where this method has been used.

2:2:17:3 Hybridization of RNA-blots in SSC and formamide

This method was adapted from that of Thomas (1980) and is similar to that described by Barinaga *et al* (1981). The filters were prehybridized (50ml; 7h; 42°C) in 5 x SSC; 5 x Denhardt's; 50% (w/v) formamide; 100μg/ml denatured herring sperm DNA; 0.1% (w/v) SDS. This solution was then exchanged for 50ml of 5 x SSC; 2 x Denhardt's; 50% (w/v) formamide; 200μg/ml denatured herring sperm DNA; 0.1% (w/v) SDS. The denatured probe was added and the hybridization solution thoroughly mixed. The filters were hybridized at 42°C for 48h. After hybridization, the filters were washed at 42°C for 2 x 10min in 2 x SSC; 0.1% (w/v) SDS and 2 x 10min in 0.1SSC; 0.1% (w/v) SDS.

2:2:18 Autoradiography of hybridized DNA and RNA filters

All filters were air-dried before autoradiography to minimise fracturing of the filter during freezing and thawing. The filters were mounted on a piece of Whatman 3MM paper, on a glass plate. This glass plate was then sealed in a plastic bag, to prevent the filter sticking to the film during autoradiography. The plastic bag also prevented movement of the filter during exposure.

Two sticky labels were stuck to the plastic bag alongside the well-position and the 'origin' was marked with radioactive ink. A sheet of Fuji X-ray film (NIF-RX) was flash sensitised and together with a Dupont intensifying screen was placed over the filter. These

were held in place by a second glass plate and several elastic bands. this exposure assembly was then placed in several black plastic bags and finally into an empty Kodak photographic-paper box.

The genomic filters, probed with storage protein gene or cDNA probes were exposed for 3d at -80°C. The film was developed in Ilford 'Phenisol' developer (8min; room temperature), rinsed in cold water and fixed in Kodak fixer (2min; room temperature). The developed film was washed in running cold water (30min) and air-dried at room temperature. After the filters had completely thawed out, they were re-exposed for appropriate lengths of time (typically 2 weeks).

Southern blots which had been hybridized with the ribosomal probe were usually much hotter and were initially exposed overnight, with the second exposure being adjusted accordingly.

2:2:19 Assessment of the autoradiographs

The extent of $C^{m}CGG$ modification in the legumin gene family was determined by a method similar to that of van der Ploeg *et al* (1980) and Kunnath and Locker (1982b). The relative intensities of hybridization signal, of all Hpa II bands relating to a specific Msp I/Hpa II site, were scored on a 1-16* scale. The percentage methylation at a particular site was then calculated as follows:-

$$\% \text{ methylation} = \frac{\text{Total score of partial digest where the site was uncut}}{\text{Total score for all bands containing the site}} \times 100$$

In some cases it was possible to scan the autoradiographs with an LKB (Bromma) 2202 Ultroscan laser densitometer, with an LKB (Bromma) 2220 recording integrator. From the areas under the individual peaks relating to the hybridization bands in question, it was possible to

get a more accurate measure of the relative signal intensities.

All other autoradiographs were assessed visually.

2:2:20 Analysis of legumin gene sequence data

The sequence data of legumin gene A (Lycett *et al* , 1984) and a pseudogene, legumin gene D (Mr M. Levasseur, Department of Botany, University of Durham) were analysed.

2:2:20:1 Distribution of CG dinucleotides in sequence data

This analysis is based on that of McClelland and Ivarie (1982), McClelland (1983b) and Adams and Eason (1984a). The sequence data was divided into 100 nucleotide segments (working upstream and downstream from the site of initiation of translation). The total number of each of the individual bases was counted and the observed number of the various di- and tri-nucleotides was compared to the number expected if the bases occurred at random. The results were examined for evidence of CG-suppression in the different gene regions.

2:2:20:2 Assessment of the microenvironment of the CG dinucleotides

A method recently published by Adams and Eason (1984a) was used. The method compares the number of dinudeotides observed with the overall percentage composition of the DNA. The environment of both 100 nucleotide stretches and the microenvironment surrounding individual dinucleotides was examined.

Where analysis has been performed on 100 nucleotide segments, the reference line for a random distribution was obtained by plotting the expected frequency of occurrence for all the dinucleotides, against

the percentage composition of the 100 nucleotides. This reference line was then used to compare the observed frequencies for 12 different dinucleotides.

Where analysis of the microenvironment has been performed, the four bases on either side of the dinucleotide were assessed and the number of times that a CG dinucleotide occurred in a decanucleotide containing 2,3,4...9,10 C + G nucleotides, was plotted as a cumulative index against the C + G content of the decanucleotide.

3: RESULTS

3:1 Analysis of the Methylation Status of Pea Genomic DNA

The extent of methylation at the 5'-CCGG-3' sequences was monitored by resistance to digestion by Hpa II (figures 8a and b). Both Hpa II and Msp I cut the sequence 5'-CCGG-3', but only Msp I cuts the methylated version, 5'-C^{m}CGG-3'. Thus, comparison of the molecular weight distributions of genomic DNA cut with each of these enzymes, enabled the quantitation of methylation. Figure 9 (a and b) shows comparative laser densitometer scans of Hpa II and Msp I digested cotyledon and leaf DNA.

Figure 10 (a and b) shows cumulative number average molecular weight distributions calculated from the scans illustrated in figure 9. The cotyledon number average distributions appeared to be homogeneous curves. However, the distributions obtained for Msp I and Hpa II digested leaf DNA were more complex and appeared to consist of two main regions, with a break at about 20kb. The number average molecular weights (M_n) calculated for each genomic DNA are shown in table 2. For cotyledon DNA, digested with Msp I, the M_n is related to the number of molecules and by implication to the number of sites for a particular restriction enzyme and it was therefore used to monitor the extent of methylation at Msp I/Hpa II sites. The percentage methylation of cotyledon genomic DNA was calculated to be 27.3% at 9 d.a.f. and this value decreased rapidly during the next 48 hours reaching a minimum value (11.1%) at 15 d.a.f. before starting to slowly increase again towards the later stages of cotyledon development. The similarity in Msp I and Hpa II M_n values obtained for leaf genomic DNA precluded an estimation of its methylation status.

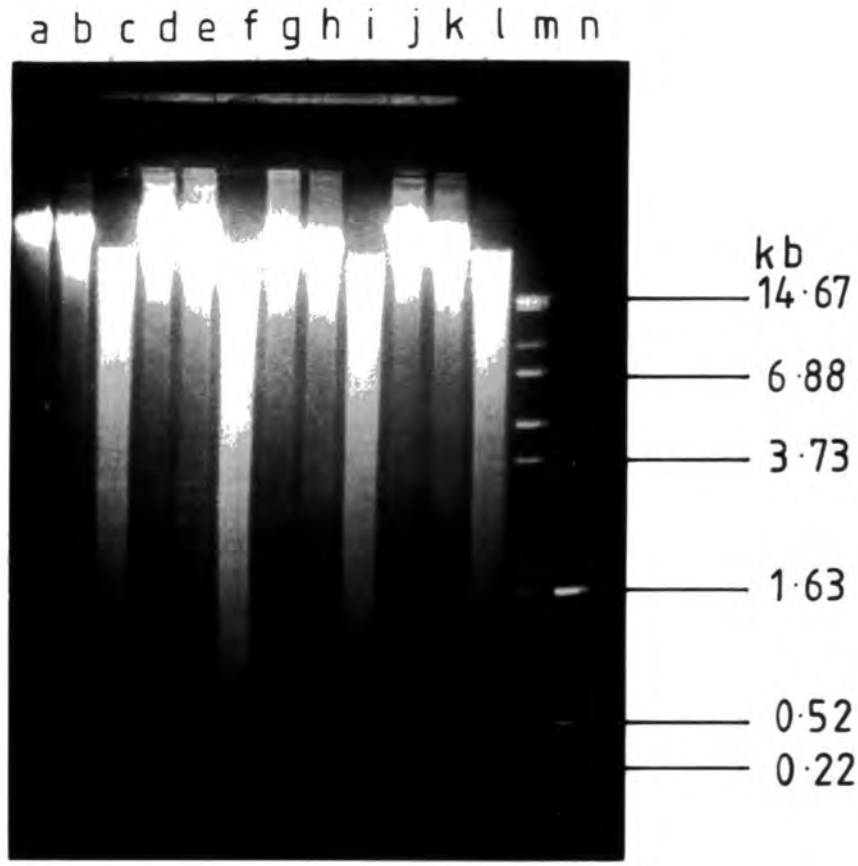
Figure 8a and b:- Distribution of Msp I and Hpa II digested pea genomic DNA.

Figure 8a:- a) 9 d.a.f. cotyledon DNA (no enzyme)
b) " " " " + Hpa II
c) " " " " + Msp I
d) 10 d.a.f. cotyledon DNA (no enzyme)
e) " " " + Hpa II
f) " " " + Msp I
g) 11 d.a.f. cotyledon DNA (no enzyme)
h) " " " + Hpa II
i) " " " + Msp I
j) Imbibed cotyledon DNA (no enzyme)
k) " " " + Hpa II
l) " " " + Msp I
m) λNM258 + Ava I
n) pBR322 + Hinf I

Figure 8b:- a) 12 d.a.f. cotyledon DNA (no enzyme)
b) " " " + Hpa II
c) " " " + Msp I
d) 15 d.a.f. cotyledon DNA (no enzyme)
e) " " " + Hpa II
f) " " " + Msp I
g) 21 d.a.f. cotyledon DNA (no enzyme)
h) " " " + Hpa II
i) " " " + Msp I
j) Leaf DNA (no enzyme)
k) " " " + Hpa II
l) " " " + Msp I
m) λNM258 + Ava I
n) pBR322 + Hinf I

All genomic tracks contained 10μg of DNA and digestions were carried out with a 10-fold excess of enzyme. Control genomic DNAs were incubated at 37°C with Msp I buffer but no enzyme.

a.


a b c d e f g h i j k l m n
kb
14·67
6·88
3·73
1·63
0·52
0·22

b.

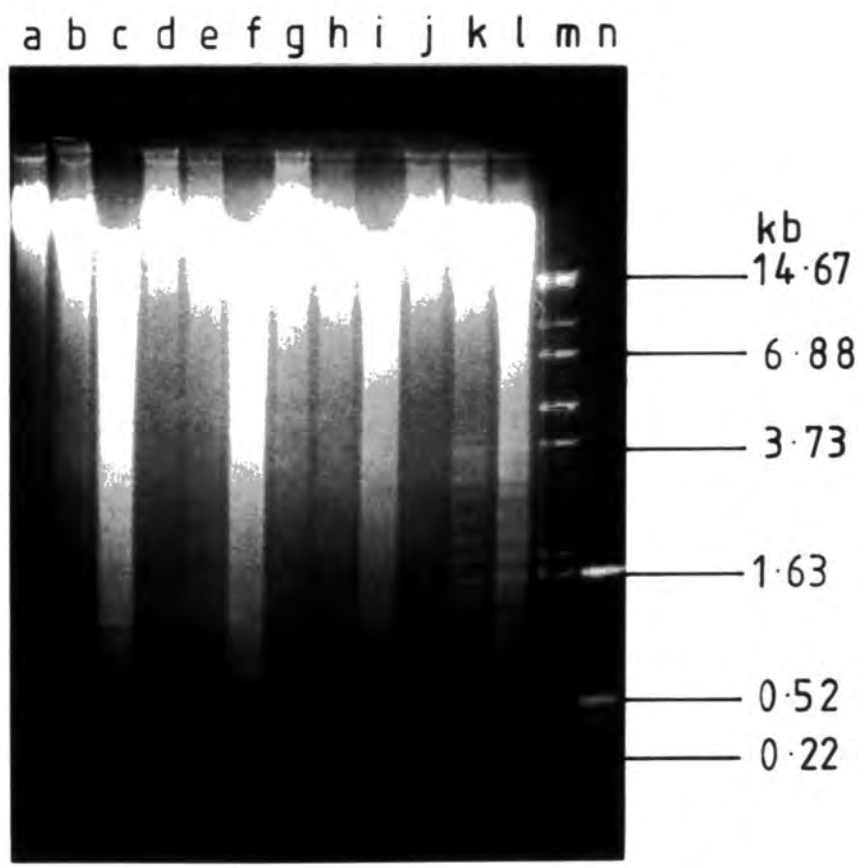

a b c d e f g h i j k l m n
kb
14·67
6·88
3·73
1·63
0·52
0·22

Figure 9a:- Laser densitometer scans obtained from a photographic negative of the gel illustrated in figure 8a.

I 9 d.a.f. cotyledon DNA
II 11 d.a.f. cotyledon DNA
III 15 d.a.f. cotyledon DNA

The solid line represents the distribution of DNA fragments obtained after Hpa II digestion and the dotted line represents the distribution of Msp I derived fragments.

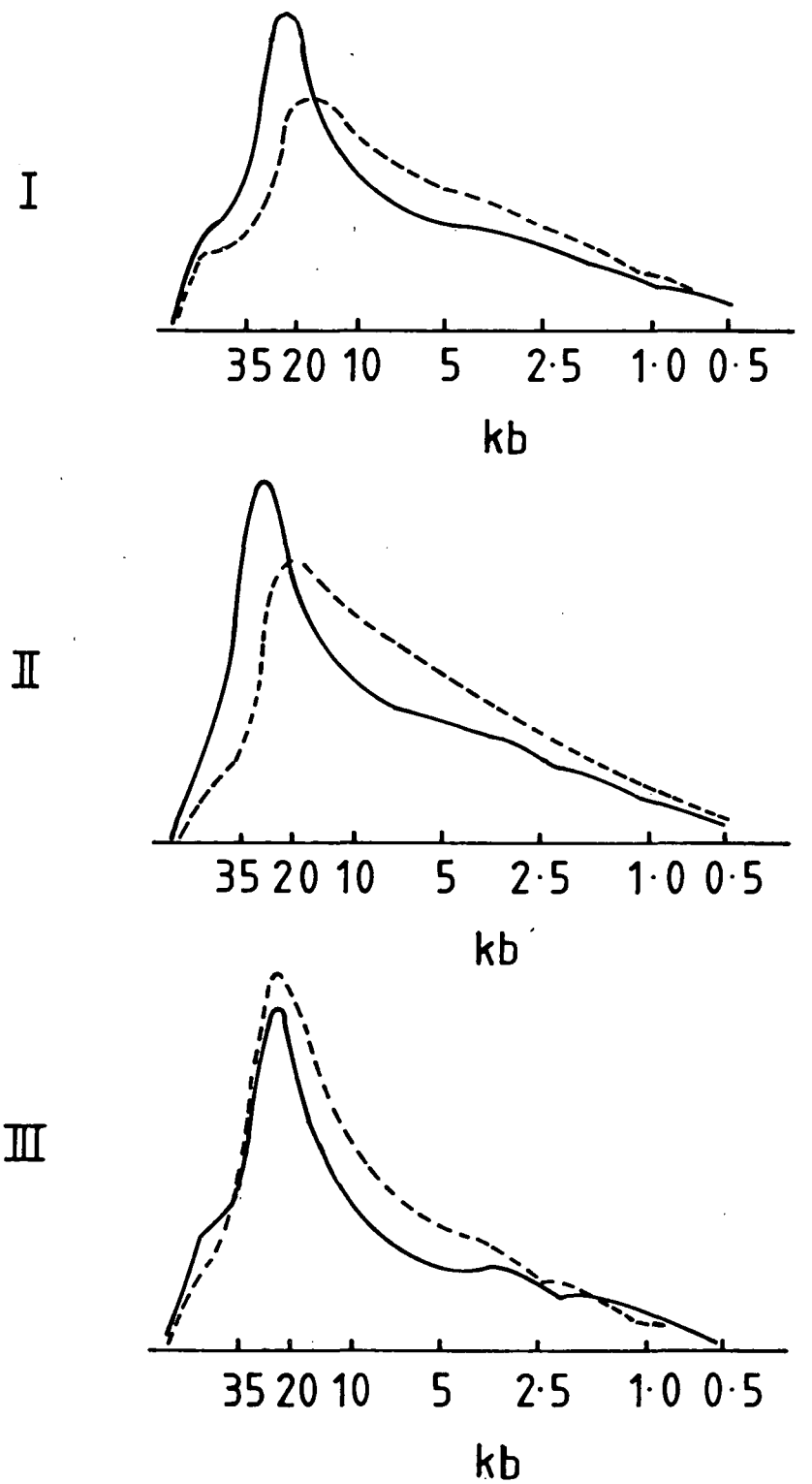
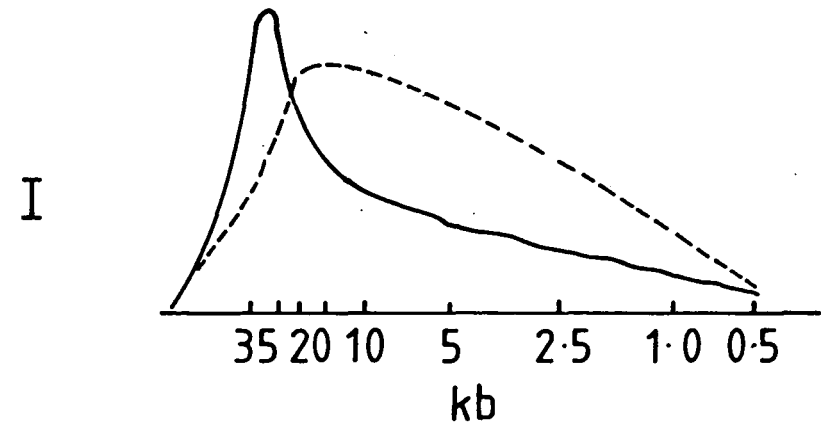
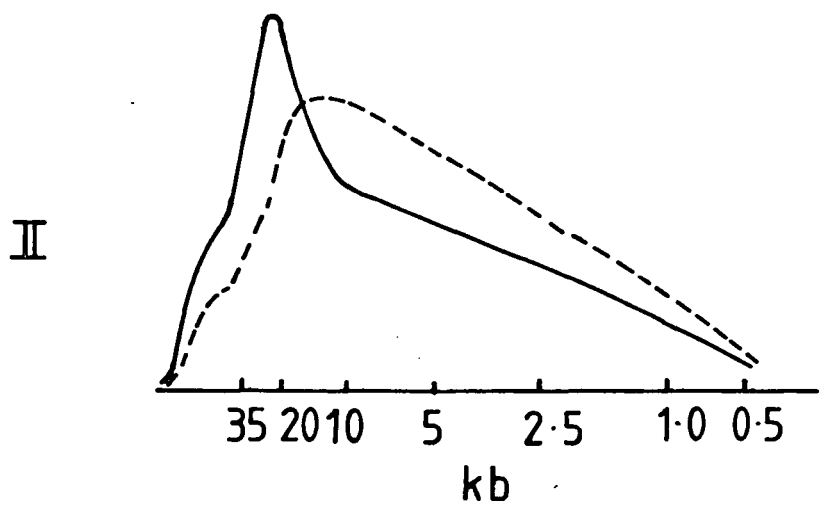

I
35 20 10 5 2·5 1·0 0·5
kb
II
35 20 10 5 2·5 1·0 0·5
kb
III
35 20 10 5 2·5 1·0 0·5
kb

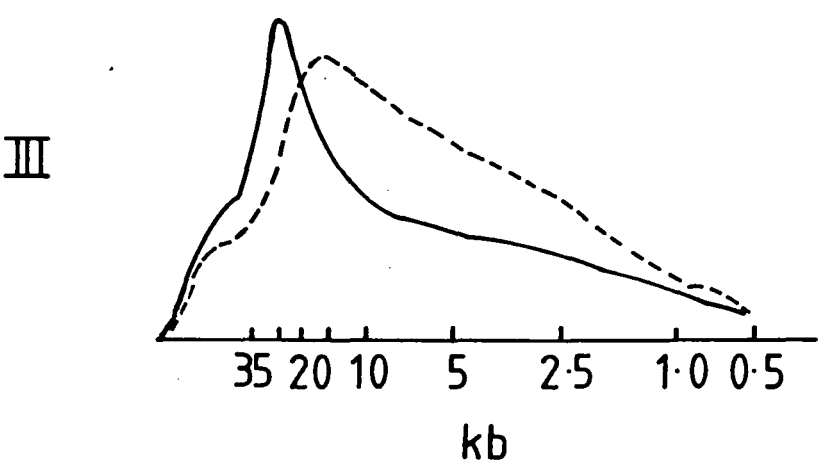
Figure 9b:- Laser densitometer scans obtained from a photographic negative of the gel illustrated in figure 8b.

I 21 d.a.f. cotyledon DNA
II Imbibed cotyledon DNA
III Leaf DNA

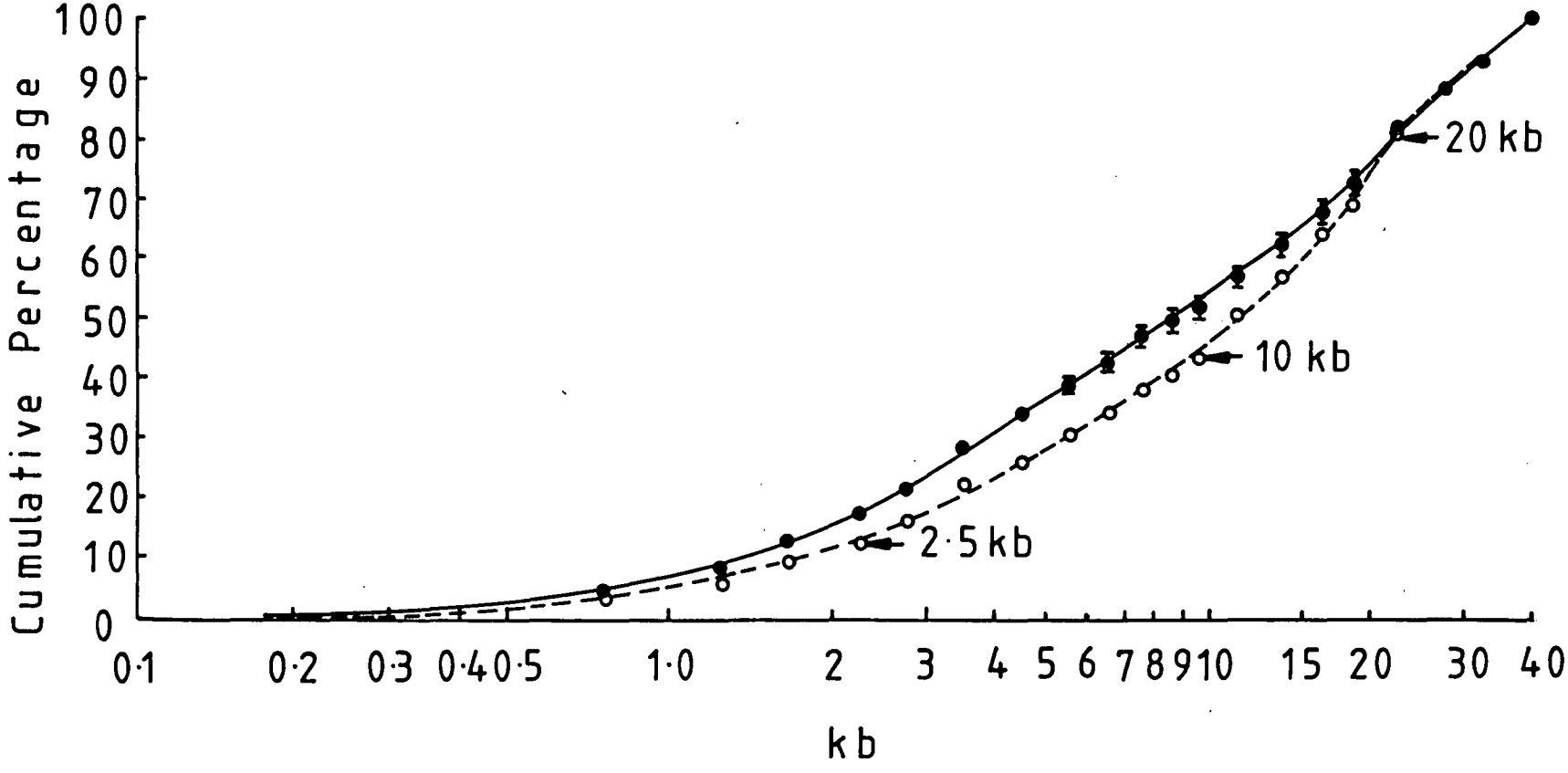
The solid line represents the distribution of DNA fragments obtained after Hpa II digestion and the dotted line represents the distribution of Msp I derived fragments.


I
35 20 10 5 2·5 1·0 0·5
kb


II
35 20 10 5 2·5 1·0 0·5
kb


III
35 20 10 5 2·5 1·0 0·5
kb

b.


Cumulative Percentage
100
90
80
70
60
50
40
30
20
10
0
0·1
0·2
0·3
0·4
0·5
1·0
2
3
4
5
6
7
8
9
10
15
20
30
40
kb
20 kb
10 kb
2·5 kb

Cumulative Percentage Weight
0
10
20
30
40
50
60
70
80
90
100
0·1
0·2
0·3
0·4
0·5
1·0
2
3
4
5
6
7
8
9
10
15
20
30
40
kb
0·1
0·2
0·3
0·4
0·5
1·0
2
3
4
5
6
7
8
9
10
12
15
20
30
40
Kp
20 kb
50 Kb
10 Kb
2·5 kb

B.

Figures 10a and b:- Cumulative number average molecular weight distributions of Msp I and Hpa II digested cotyledon and leaf genomic DNA.

Figure 10a shows the cumulative distributions of Msp I digested cotyledon (◉) and leaf (O) DNA. Figure 10b shows the cumulative distribution for Hpa II digested cotyledon (◉) and leaf (O) DNA. The cotyledon distributions show the mean values (±1 standard error) obtained from the 9, 10, 11, 15, 21, d.a.f. and imbibed densitometer scans (figures 9a and b). The distributions were plotted using the molecular weight midpoint of each interval used for the analysis.

Table 2: Average molecular weights and methylation of Msp I/Hpa II digested pea genomic DNA

Genomic DNA digestion	Mass average molecular weight 'M_w' (kb)	Number average molecular weight 'M_n' (kb)	Percentage methylation	Ratio 'r'
9 d.a.f. coty + Hpa II	24.92	4.19	27.31	5.95
9 d.a.f. coty + Msp I	18.34	3.04		6.03
10 d.a.f. coty + Hpa II	23.97	3.53	21.16	6.80
10 d.a.f. coty + Msp I	18.28	2.78		6.65
11 d.a.f. coty + Hpa II	22.35	3.36	14.67	6.65
11 d.a.f. coty + Msp I	18.42	2.87		6.42
15 d.a.f. coty + Hpa II	24.43	3.99	11.13	6.13
15 d.a.f. coty + Msp I	19.77	3.54		5.58
21 d.a.f. coty + Hpa II	23.02	4.12	14.67	5.59
21 d.a.f. coty + Msp I	20.50	3.53		5.84
Imbibed coty + Hpa II	22.97	4.23	17.84	5.43
Imbibed coty + Msp I	18.13	3.48		5.21
Leaf + Hpa II	23.42	4.88	(-1.61)	4.81
Leaf + Msp I	21.36	4.96		4.31

These values were calculated (see section 2:2:9) from the laser densitometer scans (figure 9) of the negatives of the gels shown in figure 8.

Figure 11 (a and b) shows the cumulative mass average molecular weight distributions for cotyledon and leaf genomic DNA. Again the cotyledon distributions appeared to be simple, smooth curves, whilst those obtained for leaf DNA were complex with apparent breaks in their continuity at about 17kb (Msp I) and 30kb (Hpa II). The mass average molecular weight (M_n) for Msp I digested cotyledon DNA remained fairly constant during the early stages of cotyledon development whilst increasing during later stages of development. However, for Hpa II digested cotyledon DNA, M_n decreased from 24.9kb (9 d.a.f.) to 23.3kb at 11 d.a.f. before increasing again. Leaf genomic DNA had M_w of 21.4kb (Msp I) and 23.4kb (Hpa II).

The apparent differences in M_n and M_w, between cotyledon and leaf DNA, are summarised in table 3.

Figure 12 shows an example of the distribution of DNA fragments obtained, following digestion of leaf and cotyledon genomic DNA with the isoschizomers Bst NI and Eco RII (recognition sequence 5'-CC GG-3'). An example of the comparative densitometer scans is shown in figure 13. Figures 14 and 15 show the cumulative number average and mass average molecular weight distributions. The mass average distributions appeared to be homogeneous curves, however, the cumulative number average distributions had breaks in their continuity at 12.5kb(Eco RII) and 17.5kb (Bst NI).

Table 4 shows the average molecular weights obtained for Bst NI and Eco RII digested cotyledon and leaf DNA. The mass average molecular weights obtained for leaf DNA digestions were greater than those obtained for cotyledon DNA digestions (Bst NI: 18.3kb for leaf

Figures 11a and b:- Cumulative mass average molecular weight distributions of Msp I and Hpa II digested cotyledon and leaf genomic DNA.

Figure 11a shows the cumulative distributions for Msp I digested cotyledon (●) and leaf (O) DNA. Figure 11b shows the cumulative distribution for Hpa II digested cotyledon (●) and leaf (O) DNA. The cotyledon distributions show the mean values (±1 standard error) obtained from the 9, 10, 11, 15, 21, d.a.f. and imbibed densitometer scans (figures 9a and b). The distributions were plotted using the molecular weight midpoint, of each interval used for the analysis.

Cumulative Percentage
100
90
80
70
60
50
40
30
20
10
0
0·1
0·2
0·3
0·4
0·5
1·0
2
3
4
5
6
7
8
9
10
15
20
30
40
kb
17 kb

a b.

Table 3: Comparison of leaf and cotyledon average molecular weight

Average molecular weights	Cotyledon DNA Mean	Cotyledon DNA S.D. n-1 (kb)	Leaf DNA (kb)
Hpa II M_w	23.61	0.99	23.42
Msp I M_w	18.91	0.99	21.36
Hpa II M_n	3.90	0.37	4.88
Msp I M_n	3.20	0.35	4.95

Table 4: Average molecular weights and methylation of Bst NI/Eco RII digested pea genomic DNA

Genomic DNA digestion	Mass average molecular weight 'M_w' (kb)	Number average molecular weight 'M_n' (kb)	Percentage methylation	Ratio 'r'
11 d.a.f. coty + Bst NI	20.57	3.28	67.30	6.27
11 d.a.f. coty + Eco RII	31.60	10.03		3.15
22 d.a.f. coty + Bst NI	21.93	2.40	83.48	9.14
22 d.a.f. coty + Eco RII	31.49	14.52		2.17
Leaf + Bst NI	18.25	2.44	78.73	7.45
Leaf + Eco RII	34.92	11.45		3.05

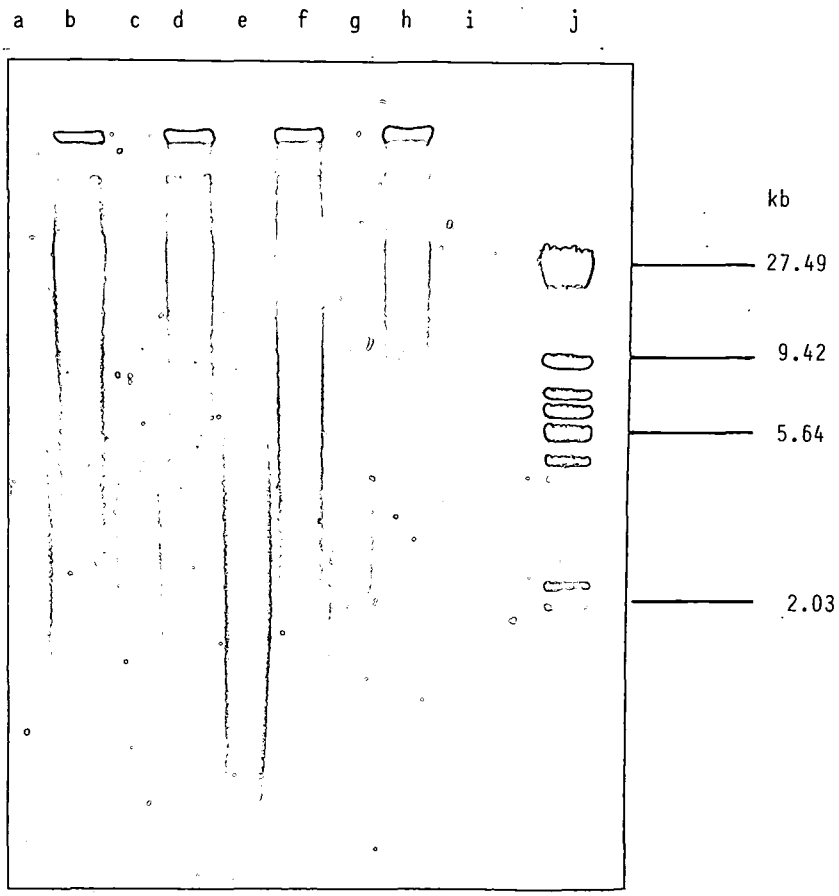


Figure 12:- Gel photograph showing the distribution of Bst NI and Eco RII digested genomic DNA.

a) 11 d.a.f. cotyledon DNA + Bst NI
b) " " " " + Eco RII
c) 15 d.a.f. cotyledon DNA + Bst NI
d) " " " " + Eco RII
e) 22 d.a.f. cotyledon DNA + Bst NI
f) " " " " + Eco RII
g) Leaf DNA + Bst NI
h) " " + Eco RII
i) 250pg pHA 1
j) λNM258 + Hind III; λNM258 + Eco RI

All genomic tracks contained 2µg of DNA and digestion were carried out with a 5-fold excess of enzyme.

Figure 13:- Densitometer scans obtained from a photographic negative of the gel illustrated in figure 12.

I 11 d.a.f. cotyledon DNA
II 22 d.a.f. cotyledon DNA
III Leaf DNA

The solid line represents the distribution of DNA fragments obtained after Eco RII digestion and the dotted line represents the distribution of Bst NI derived fragments.

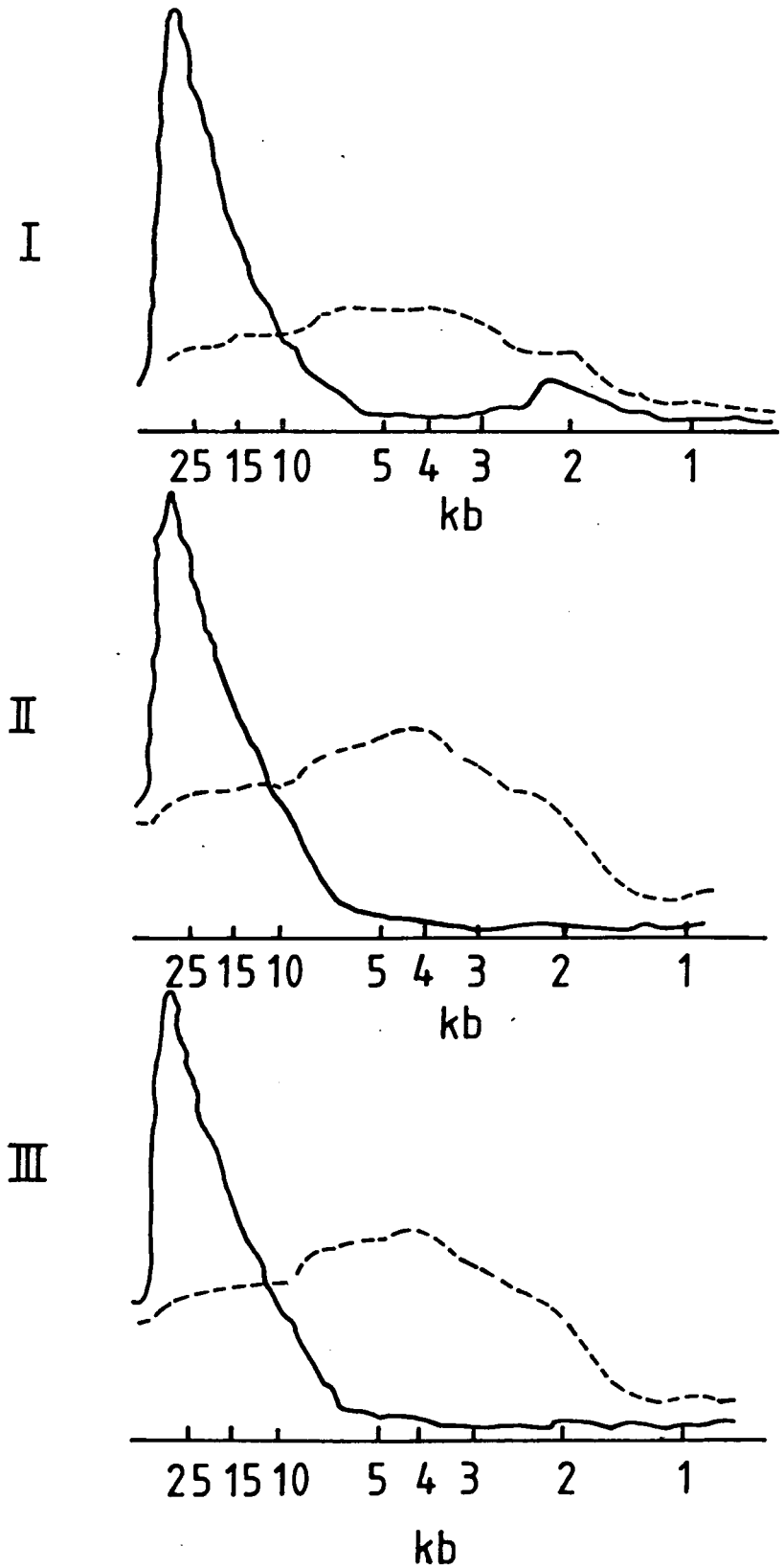

I
25 15 10 5 4 3 2 1
kb
II
25 15 10 5 4 3 2 1
kb
III
25 15 10 5 4 3 2 1
kb

Figure 14:- Cumulative number average molecular weight distributions of Bst NI and Eco RII digested genomic DNA.

The cumulative distributions show the mean values (±1 standard error) of Bst NI (●) and Eco RII (O) digested genomic DNA. The distributions were plotted using the molecular weight midpoint of each interval used in the analysis.

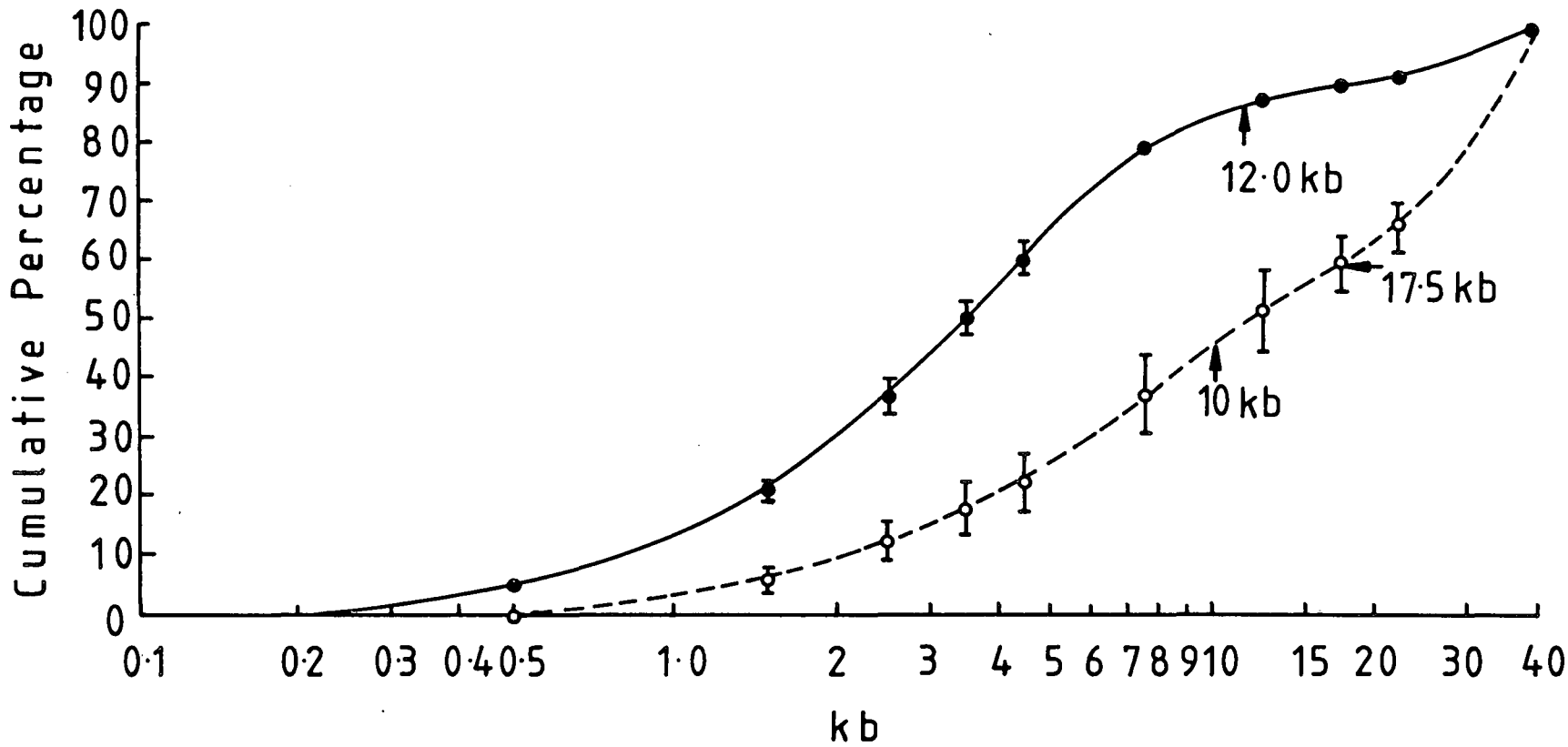
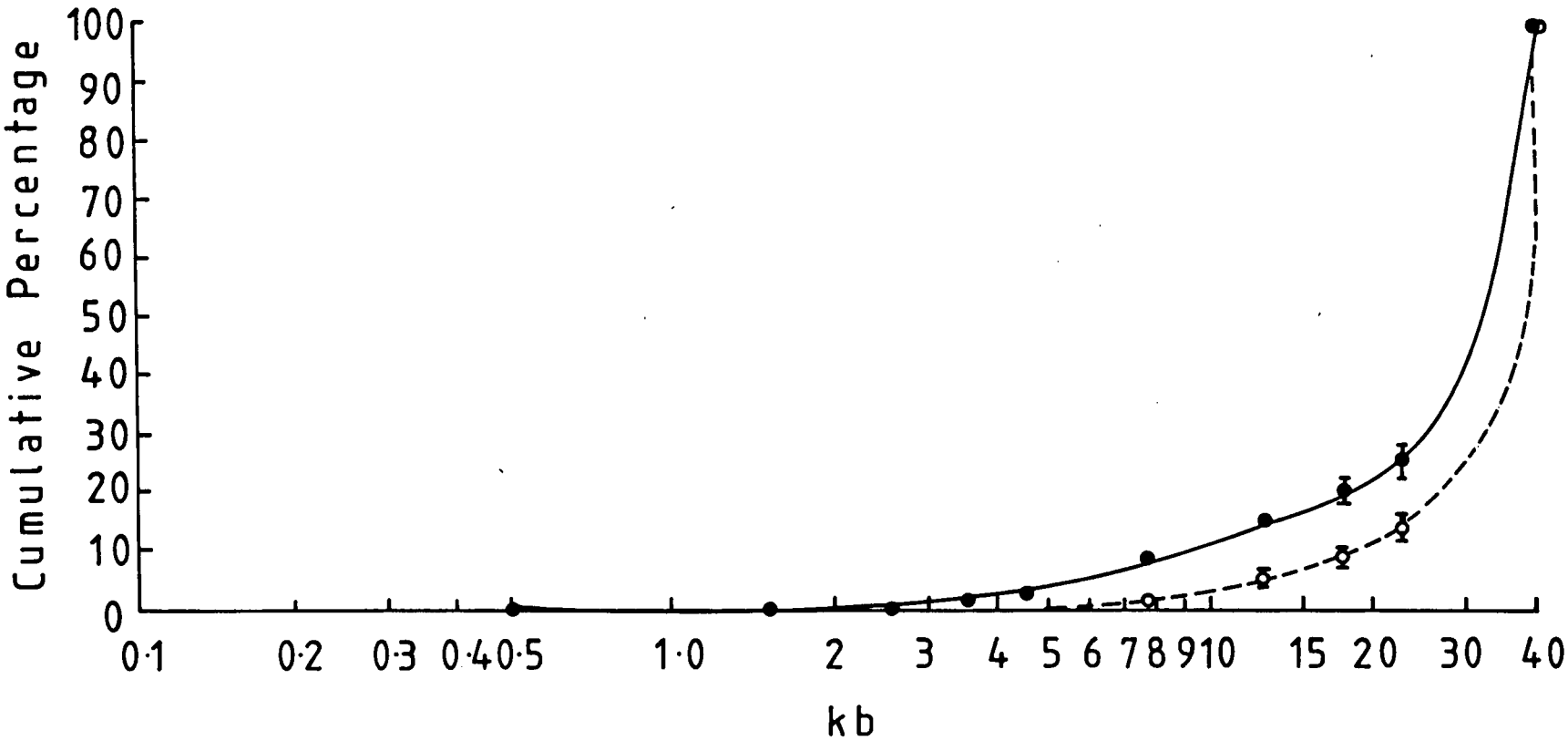

Cumulative Percentage
100
90
80
70
60
50
40
30
20
10
0
0·1
0·2
0·3
0·4
0·5
1·0
2
3
4
5
6
7
8
9
10
15
20
30
40
kb
12·0 kb
17·5 kb
10 kb

Figure 15:- Cumulative mass average molecular weight distributions of Bst NI and Eco RII digested genomic DNA.

The cumulative distributions show the mean values (±1 standard error) of Bst NI (●) and Eco RII (O) digested genomic DNA. The distributions were plotted using the molecular weight midpoint, of each interval used in the analysis.


Cumulative Percentage
100
90
80
70
60
50
40
30
20
10
0
0·1
0·2
0·3
0·4
0·5
1·0
2
3
4
5
6
7
8
9
10
15
20
30
40
kb

compared with 20.6 and 21.9kb for cotyledon DNA; Eco RII 34.9kb for leaf compared with 31.6 and 31.5kb for cotyledon DNA).

The number average molecular weights obtained for Bst NI digestions of genomic DNA ranged from 2.4 to 3.3kb, whilst greater variations was observed following Eco RII digestion, with 11 and 22 d.a.f. cotyledon DNA having M_n values of 10.0 and 14.5kb respectively and leaf DNA, 11.5kb. The percentages methylation calculated from these M_n values were for cotyledon DNA, 67.3 and 83.5% and for leaf DNA, 78.7%.

3:2 Analysis of the Methylation of rRNA Genes in Pea

3:2:1 Methylation of rDNA at Msp I/Hpa II sites

Figure 17 (a and b) shows the results obtained when pea genomic DNA, restricted fragments were blotted and hybridized with a labelled cloned pea ribosomal gene. All four restriction enzymes used had a four nucleotide recognition sequence (Hae III 5'-GGCC-3' ; Hha I 5'-GCGC-3'; Msp I and Hpa II 5'-CCGG-3'). Hae III cleaved the genomic DNA into fragments with an average size of less than 2kb (figure 17a). Following digestion by both Hha I and Hpa II, a considerable amount of DNA remained as fragments of greater than 25kb (figure 17a).

The cloned pea ribosomal gene probe (pHA 1) hybridized to a wide range of fragment sizes (figure 17b). Following Hha I digestion of cotyledon DNA (tracks b, f, j and n), major bands of hybridization were observed at 2.69, 3.39, 3.85, 4.84, 6.73 and 10.00kb. The

Figure 16:- Simplified restriction map of a cloned pea ribosomal RNA gene (pHA 1) (from Cuellar, 1982).

The thickened line represents the cloned gene (8.6kb) and the single line the pACYC 184 vector (3.9kb). The positions of the Bam HI sites (B) and Hind III sites (H) are indicated. All fragment sizes are in kilobases.

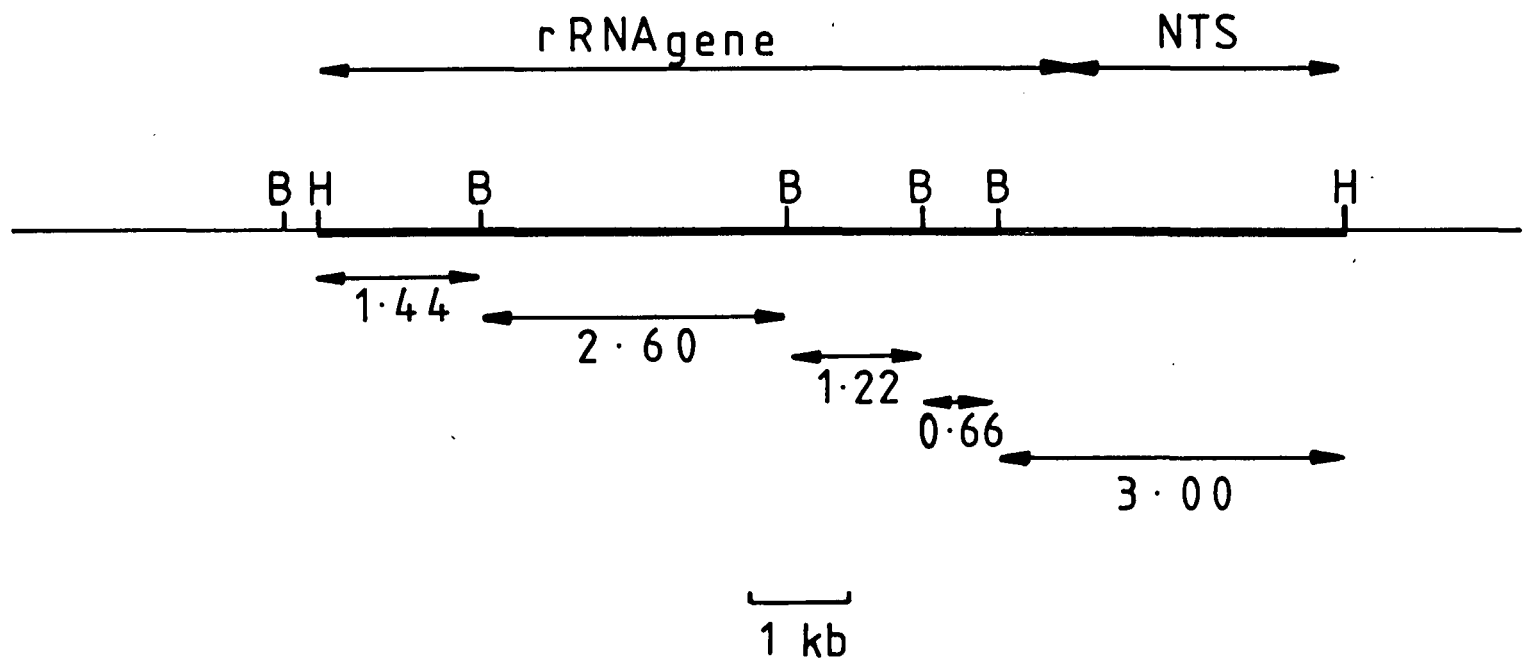

rRNA gene
NTS
B H
B
B
B
B
H
1·44
2·60
1·22
0·66
3·00
1 kb

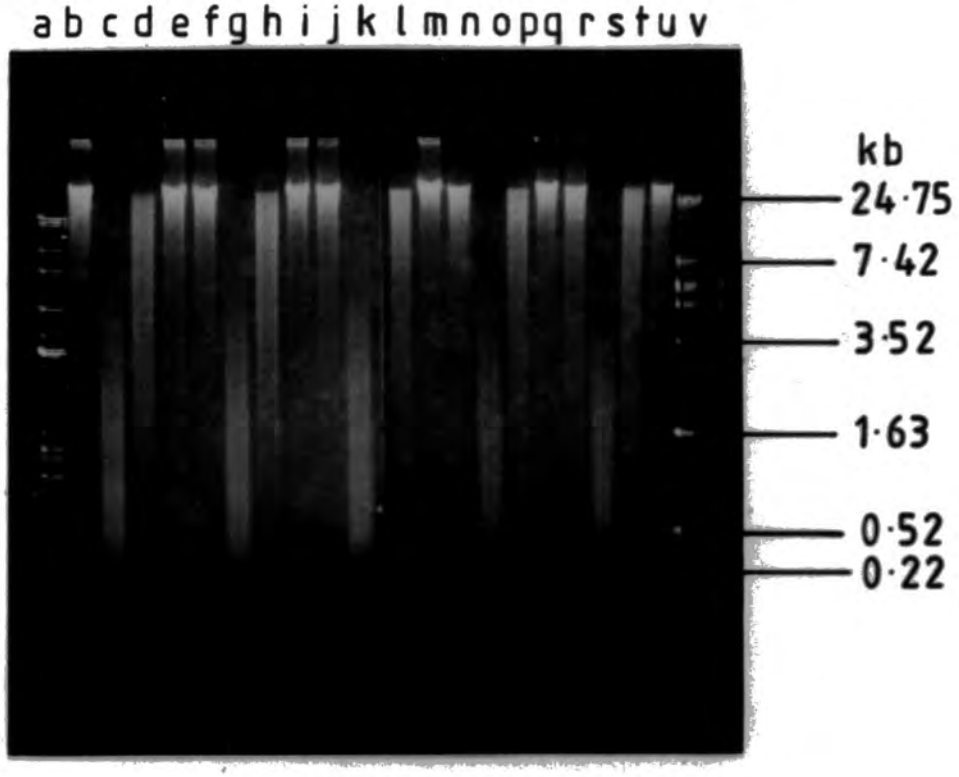
Figure 17a:- Gel photograph of Hha I, Hae III, Msp I and Hpa II digested genomic DNA.

a) λNM258 + Ava I; pBR322 + Hinc II; pBR322 + Taq I
b) 10 d.a.f. cotyledon DNA + Hha I
c) " " " " + Hae III
d) " " " " + Msp I
e) " " " " + Hpa II
f) 12 d.a.f. cotyledon DNA + Hha I
g) " " " " + Hae III
h) " " " " + Msp I
i) " " " " + Hpa II
j) 15 d.a.f. cotyledon DNA + Hha I
k) " " " " + Hae III
l) " " " " + Msp I
m) " " " " + Hpa II
n) 21 d.a.f. cotyledon DNA + Hha I
o) " " " " + Hae III
p) " " " " + Msp I
q) " " " " + Hpa II
r) Leaf DNA + Hha I
s) " " + Hae III
t) " " + Msp I
u) " " + Hpa II
v) λNm258 + Eco RI; pBR322 + Hinf I.

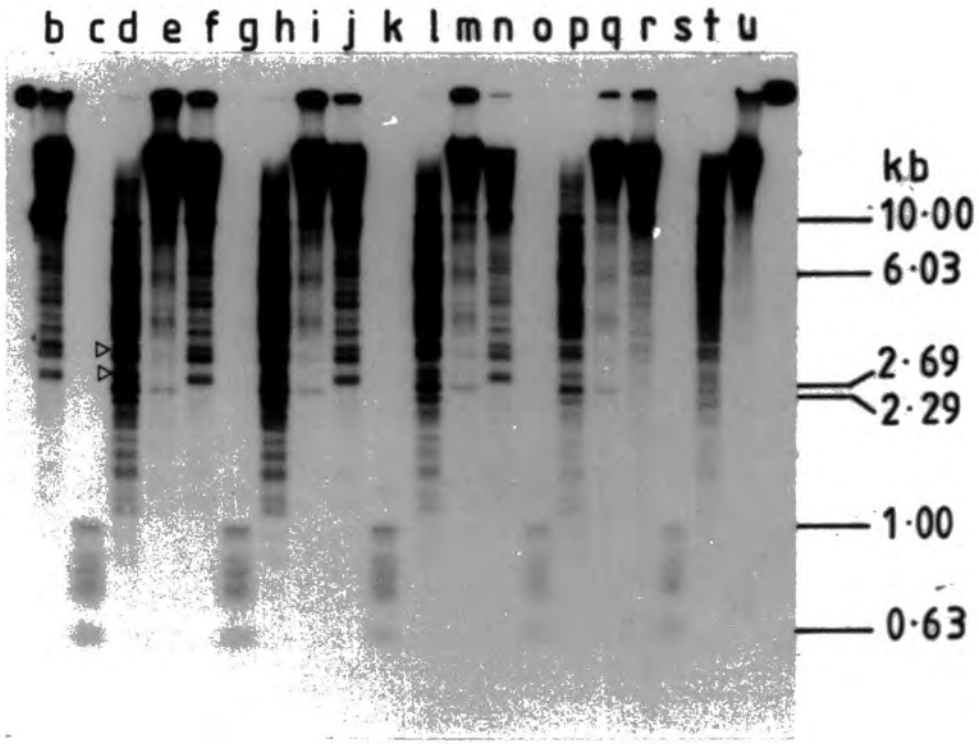
Genomic tracks contained 2 μg DNA digested with a 15-fold excess of enzyme.

Figure 17b:- An autoradiograph showing the hybridization of ^{32}P-pHA 1, to a Southern blot prepared from the gel illustrated in figure 17a. The λNM258 and pBR322 marker tracks were removed from the gel before blotting, The identities of the tracks are indicated above, The specific activity of the probe was 4.9 x 10^7 cpm/μg and the filter was washed at 50°C.

a.


a b c d e f g h i j k l m n o p q r s t u v
kb
24·75
7·42
3·52
1·63
0·52
0·22

b.


b c d e f g h i j k l m n o p q r s t u
kb
10·00
6·03
2·69
2·29
1·00
0·63

smallest Hha I fragments to hybridize to the probe were 1.00 and 1.27kb and these had only a weak signal. However, in the leaf DNA (track r) the basic pattern of hybridization remained the same but the relative intensities of the signal for the 2.69, 3.39, 3.85, 4.84 and 6.73kb bands were all reduced, whilst that of the 10.00kb remained of a similar intensity, to that observed in the cotyledon digests.

Following Hae III digestion of leaf and cotyledon DNA (tracks c, g, k, o and s), pHA 1 hybridized to a series of fragments ranging in size from 1.04kb to 0.21kb.

The 1.04kb band had a much weaker signal strength than the other bands, which were all of a similar intensity, pHA 1 also hybridized to a wide range of cotyledon, Msp I digested DNA fragments (tracks d, h, l, and p). The probe hybridized most strongly with the larger fragments, which ranged in size from 10.00 to 2.29kb. The signal intensity was reduced when pHA 1 hybridized to fragments of between 2.09 and 0.93kb and was further reduced when the probe hybridized to fragments of less than 0.8kb. Following Msp I digestion of leaf DNA (track t), a similar pattern of hybridization was obtained, except that fragments of between 2.29 and 3.78kb (which had hybridized very strongly in the cotyledon tracks) now had a reduced signal level, equivalent to only the intermediate level of hybridization observed in Msp I digested cotyledon DNA.

pHA 1 hybridized to fewer distinct bands in the Hpa II digested cotyledon DNA (tracks e, i, m and q) than in Msp I digested cotyledon DNA. Only a weak signal of hybridization was detected for the intermediate range (2.09 to 0.93kb) and a reduced signal strength was

also observed for hybridization of the probe, to fragments of greater than 2.29kb. Two fragments noted in the higher molecular weight range of Msp I digested cotyledon DNA (3.45 and 2.83kb), were absent in the Hpa II cotyledon tracks. Hybridization of pHA 1 to Hpa II digested leaf DNA (track u), was also far less than had been noted for cotyledon DNA. Even when the filter was over exposed, only a weak signal was observed for hybridization to the 10.00kb fragment.

Figure 18 shows the results obtained when an SV40 internal control, was mixed with the genomic DNA, prior to digestion by Msp I or Hpa II. The 5.23kb SV40 DNA has only one 5'-CCGG-3' sequence. Figure 18a, track k, shows the mobility of the uncleaved, supercoiled SV40 molecule, whilst tracks a, to h, show the position of the linearised molecule, which stands out strongly from the background genomic DNA.

Figure 18b shows the hybridization of pHA 1 to a filter prepared from the gel in 18a. Again the three levels of signal intensity in the Msp I tracks were apparent, as was the marked reduction in the number of leaf + Hpa II hybridization bands (track g). Tracks i and j show the position of DNA fragments produced following cleavage of pHA 1 by Msp I (i) and Hpa II (j). Although these hybridization signals were not strong, they appeared to correspond to the regions of the Msp I tracks (a, c, e and h) which showed only a weak hybridization to pHA 1 (ie less than 1.0kb). Even when this filter was over exposed, no hybridization bands of a corresponding size, were observed in any of the Hpa II tracks.

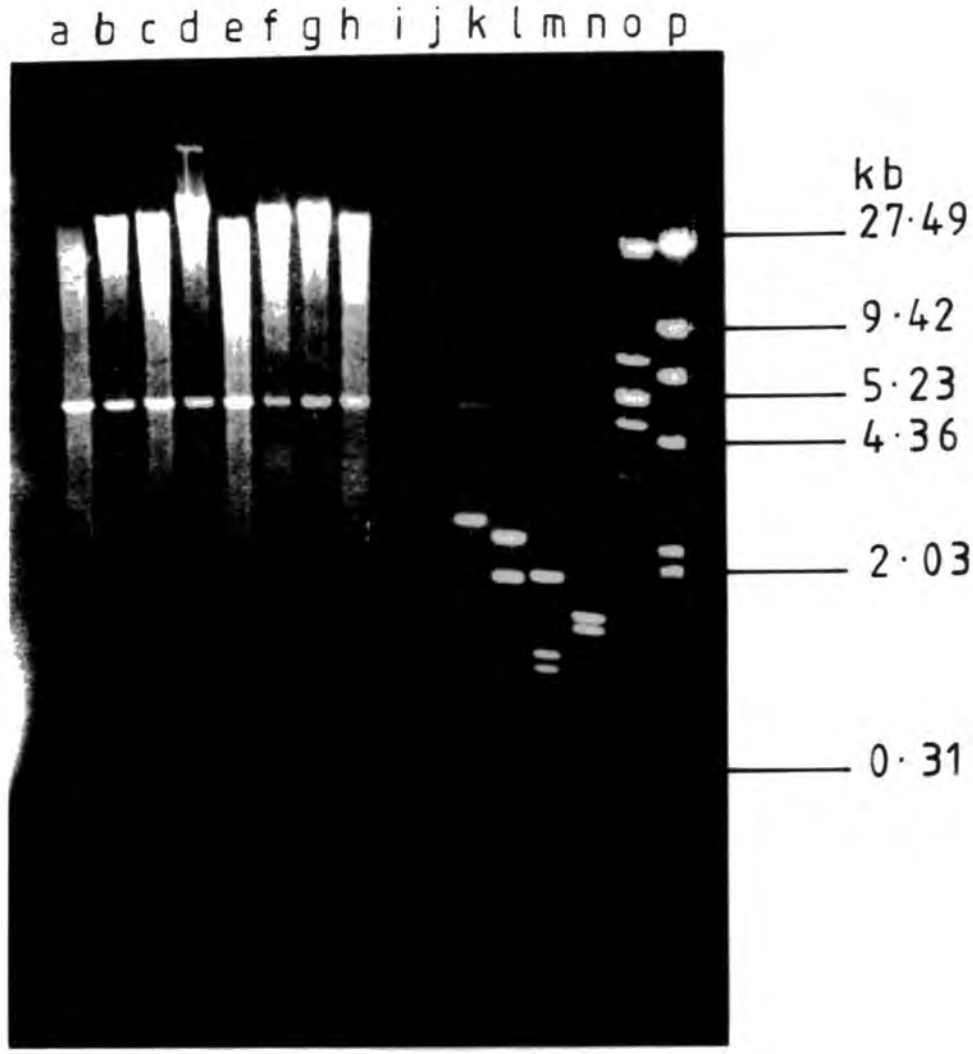
Figure 18a:- Gel photograph of Msp I and Hpa II digested genomic DNA containing an SV40 (5.23kb) internal control.

a) 11 d.a.f. cotyledon DNA + Msp I
b) " " " " + Hpa II
c) 15 d.a.f. cotyledon DNA + Msp I
d) " " " " + Hpa II
e) 22 d.a.f. cotyledon DNA + Msp I
f) " " " " + Hpa II
g) Leaf DNA + Hpa II
h) " " + Msp I
i) pHA 1 (10ng) + Msp I
j) " (10ng) + Hpa II
k) SV40 (0.1µg) Unrestricted
l) pBR322 + Bgl I
m) " + Bst NI
n) " + Taq I
o) λNM258 + Eco RI
p) " + Hind III

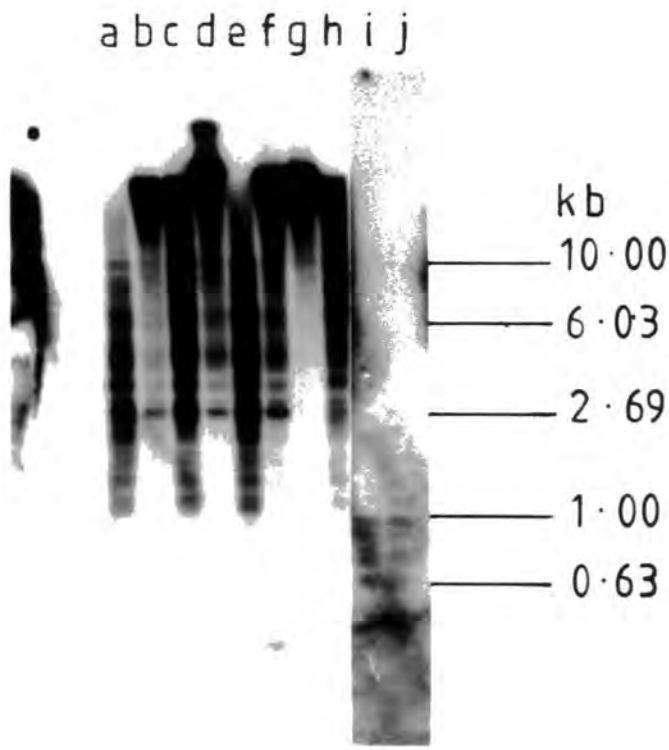
Genomic tracks contianed 2µg of DNA plus 0.1µg of SV40 DNA and were digested with a 15-fold excess of enzyme.

Figure 18b:- Hybridization of ^{32}P-pHA 1 to a Southern blot prepared from the gel illustrated in figure 18a. The specific activity of the probe was 9.7×10^7 cpm/µg and the filter was hybridized and washed at 65°C. The tracks are as indicated above.

a.


a b c d e f g h i j k l m n o p
kb
27·49
9·42
5·23
4·36
2·03
0·31

b.


a b c d e f g h i j
kb
10·00
6·03
2·69
1·00
0·63

3:2:2 Methylation of rDNA at Bst NI/Eco RII sites

Figure 19 shows the distribution of genomic DNA following digestion with Hind III plus either Bst NI of Eco RII. These genomic digestions also contained an SV40 internal control and some of the SV40 fragments can be visualised in the genomic DNA. Subsequent cleavage of Hind III digested DNA, with Bst NI produced a range of DNA fragments with an average size of less than 2kb, whereas, when the second restriction was with Eco RII, no significant further reduction in DNA fragment size was observed.

When a Southern blot prepared from this gel was hybridized to pHA 1, a major Hind III band was observed at 9.33kb, in a position corresponding to that of pHA 1 + Hind III (track 1). pHA 1 was cleaved by Bst NI (track m) and produced a range of fragments (4.07, 1.17, 1.06, 0.97, 0.85, 0.49kb). pHA 1 hybridized to a series of genomic fragments when cotyledon and leaf DNA was digested with Hind III plus Bst NI (tracks b, e, h and j). A fairly major hybridization band in this region, at 4.07kb, corresponded to the band in the pHA 1 + Hind III + Bst NI (track 1). In addition some smaller genomic bands (1.17 and 1.06kb) also corresponded to the expected bands (track 1).

The Hind III digested genomic DNA was subsequently digested with Eco RII and hybridized to pHA 1. The 9.33kb Hind III was only partially cleaved (tracks c, f, i and k) but two new bands were observed at 4.50 and 4.07kb.

This Southern blot was subsequently hybridized to ^{32}P-lablled SV40 (figure 19c). The Hind III SV40 bands (1.77, 1.17, 1.12, 0.53. 0.45 and 0.22kb) were clearly visible (tracks a, d and g). Published SV40

Figure 19a:- Ethidium bromide stained gel photograph of double restricted (Hind III plus Bst NI/Eco RII) genomic DNA containing an SV40 internal control.

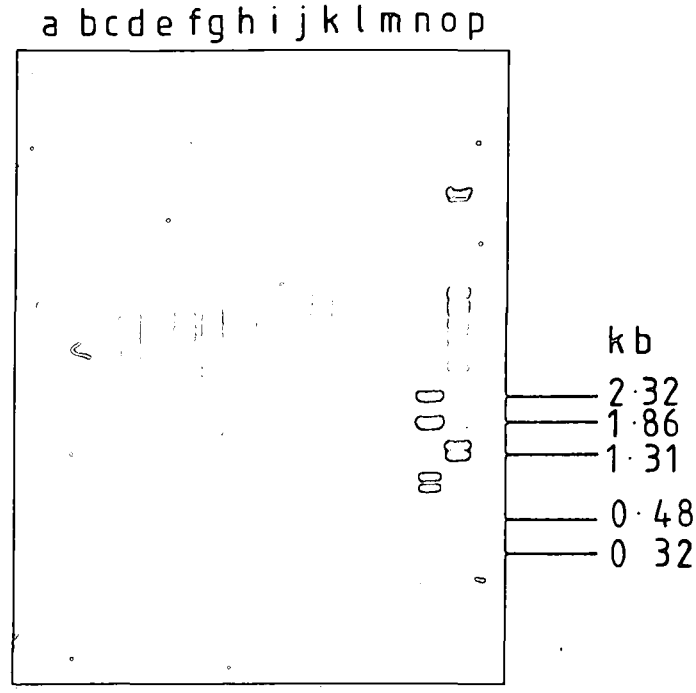
a) 11 d.a.f. cotyledon DNA + Hind III
b) " " " " + Hind III + Bst NI
c) " " " " + Hind III + Eco RII
d) 15 d.a.f. cotyledon DNA + Hind III
e) " " " " + Hind III + Bst NI
f) " " " " + Hind III + Eco RII
g) 22 d.a.f. cotyledon DNA + Hind III
h) " " " " + Hind III + Bst NI
i) " " " " + Hind III + Eco RII
j) Leaf DNA + Hind III + Bst NI
k) " " + Hind III + Eco RII
l) pHA 1 (5ng) + Hind III
m) pHA 1 (5ng) + Hind III + Bst NI
n) pHA 1 (5ng) + Hind III + Eco RII
o) pBR322 + Bgl I; pBR322 + Bst NI
p) λNM258 = Eco RI; λNM258 + Hind III; pBR322 + Taq I

Genomic tracks contained 3μg of DNA digested with an 8-fold excess of Hind III and a 10-fold excess of Bst NI and Eco RII.

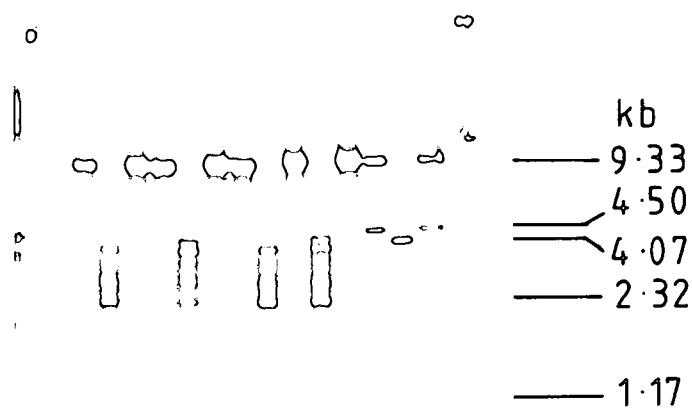
Figure 19b:- Hybridization of ^{32}P-pHA 1 to a Southern blot prepared from the gel shown in figure 19a. The specific activity of the probe was 1.4 x10^7 cpm/μg and the filter was hybridized and washed at 65°. The identities of the tracks are indicated above.

Figure 19c:- Rehybridization of the nitrocellulose filter shown in figure 19b, with ^{32}P-SV40. The specific acivity of the probe was 8.8 x 10^7 cpm/μg. The filter was hybridized and washed at 65°C.

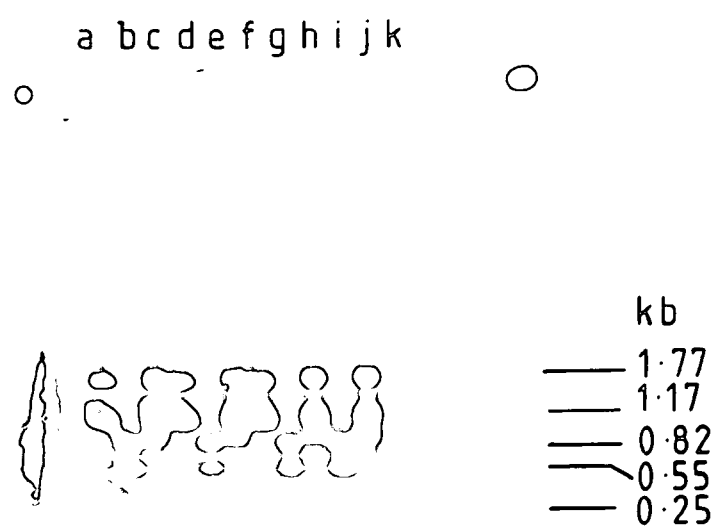
a.


a b c d e f g h i j k l m n o p
kb
2·32
1·86
1·31
0·48
0 32

b.


a bc de f g h i j k l m n
kb
9·33
4·50
4·07
2·32
1·17

c.


a b c d e f g h i j k
kb
1·77
1·17
0·82
0·55
0·25

sequence data indicated that Bst NI should cleave all of these Hind III bands except for the 0.53kb band. New bands should be detected at 0.82, 0.55, 0.36, 0.25kb (plus several fragments of less than 200bp). These bands were observed in tracks b, e, h and j. However, after a second digestion with Eco RII, although new bands were observed at 0.82, 0.55 and 0.36kb most of the SV40 had hybridized to the residual, uncleaved Hind III bands.

Figure 20 shows the results obtained when Bam HI digested genomic DNA (plus an internal SV40 control) and with a second digestion by either Bst NI or Eco RII, was blotted and hybridized to pHA 1. (The Bam HI sites in pHA 1 are shown in figure 16.)

The internal SV40 control has only one recognition sequence for Bam HI and the linearised molecule was clearly visible in the Bam HI digests (figure 20, and c, tracks a, d and g). The hybridization bands of SV40 in the Bam HI plus Bst NI digests (figure 20c, tracks b, e, h and j) corresponded in size to the largest theoretically sized fragments (0.99, 0.82, 0.67, 0.55, 0.44, 0.37. 0.25 and 0.20kb). The remaining smaller fragments were either not resolved or were not efficiently transferred during the blotting process.

In the Bam HI plus Eco RII digests (figure 20c, tracks c, f, i and k) these bands were also visible but most of the SV40 hybridized to a series of SV40 partial digests ranging in size from 1 to 4kb.

When this Southern blot was hybridized to pHA 1, a distinctive pattern of bands was produced (figure 20b). When genomic DNA was digested by Bam HI (tracks a, d and g), a prominent doublet (5.55 and 4.95kb) was observed with a third relatively major band at 2.92kb.

Figure 20a:- Gel photograph of double restricted (Bam H1 plus Bst NI/Eco RII) genomic DNA containing an SV40 internal control.

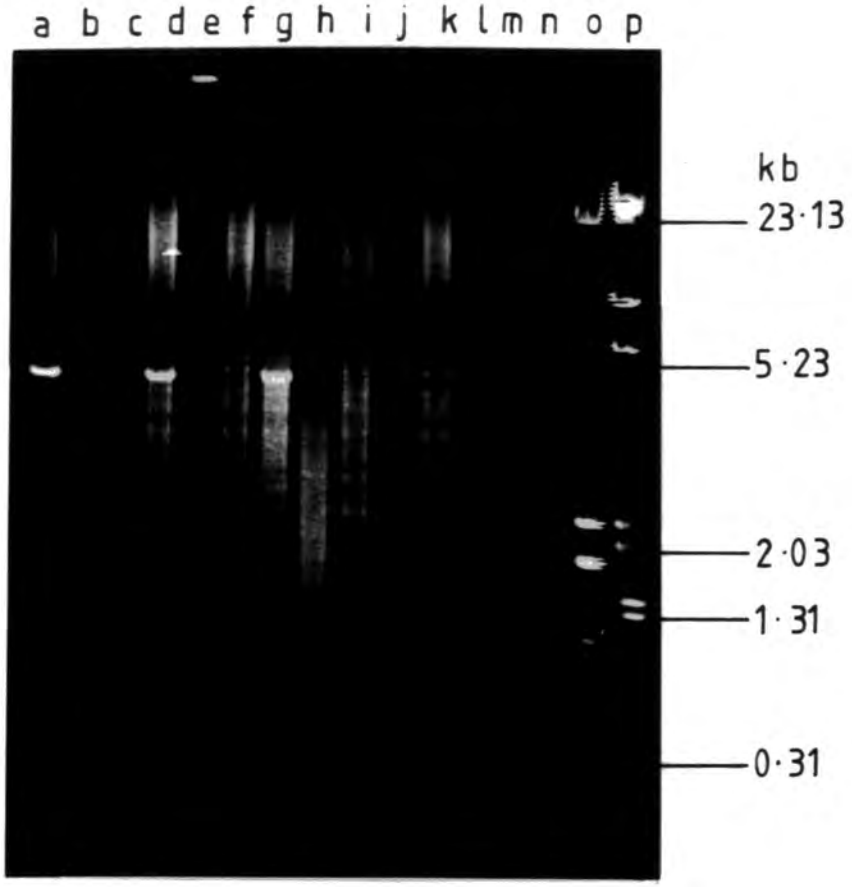
a) 11 d.a.f. cotyledon DNA + Bam HI
b) " " " " + Bam HI + Bst NI
c) " " " " + Bam HI + Eco RII
d) 15 d.a.f. cotyledon DNA + Bam HI
e) " " " " + Bam HI + Bst NI
f) " " " " + Bam HI + Eco RII
g) 22 d.a.f. cotyledon DNA + Bam HI
h) " " " " + Bam HI + Bst NI
i) " " " " + Bam HI + Eco RII
j) Leaf DNA + Bam HI + Bst NI
k) " " + Bam HI + Eco RII
l) pHA1 (5ng) + Bam HI
m) " (5ng) + Bam HI + Bst NI
n) " (5ng) + Bam HI + Eco RII
o) χNM258 + Eco RI; pBR322 + Bst NI; pBR322 + Bgl I
p) χNM258 + Hind III; pBR322 + Taq I

Genomic tracks contained 3 μg DNA plus 0.1 μg SV40 DNA and were restricted with an 8-fold excess of Bam HI and a 10-fold excess of Bst NI or Eco RII.

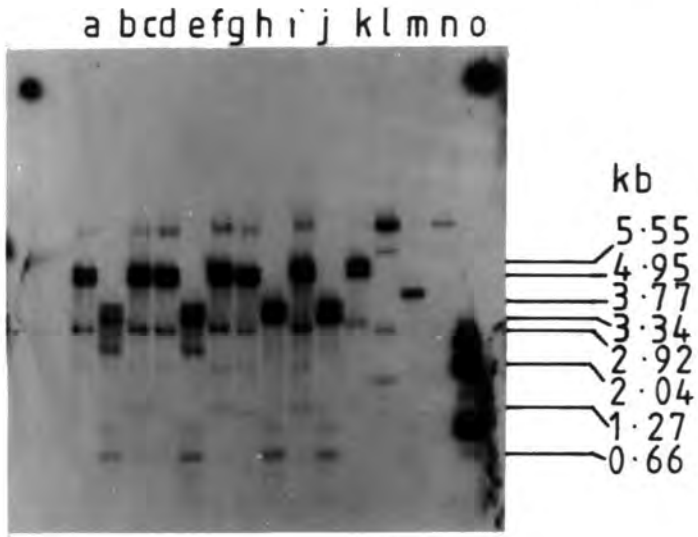
Figure 20b:- Hybridization of ^{32}P-pHA 1 to a Southern blot prepared from the gel illustrated in figure 20a. The specific activity of the probe was 1.4 x 10^7 cpm/μg and the filter was hybridized and washed at 65°C. The identities of the tracks are as indicated above.

Figure 20c:- The pHA 1 probe was washed off and the filter was rehybridized to ^{32}P-SV40. The specific activity of the probe was 8.8 x 10^7 cpm/μg. The filter was hybridized and washed at 65°C.

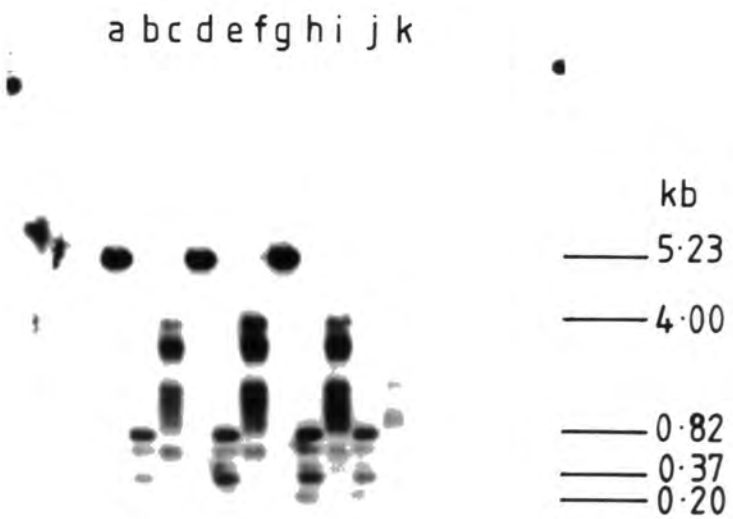
a.


a b c d e f g h i j k l m n o p
kb
23·13
5·23
2·03
1·31
0·31

b.


a b c d e f g h i j k l m n o
kb
5·55
4·95
3·77
3·34
2·92
2·04
1·27
0·66

c.


a b c d e f g h i j k
kb
5·23
4·00
0·82
0·37
0·20

Minor bands were observed at 8.32, 2.04, 1.27 and 0.66kb. The bands seen at positions equivalent in size to 1.27 and 2.92kb, appeared to be slightly larger than the equivalent bands (1.20 and 2.82kb) seen in the marker pHA 1 plus Bam HI track (1). All cotyledon and leaf DNA samples gave the same pattern of hybridization bands. A prolonged exposure of the autoradiograph indicated a band equivalent in size, to the 4.95kb fragment in the marker pHA 1 plus Bam HI track (1). This marker band corresponded to the smaller of the two bands seen in the doublet, in tracks a,d and g.

Bst NI cleaved all the major Bam HI fragments containing rDNA, both in cotyledon and leaf DNA (tracks b, e, h and j). However, Eco RII did not cleave any of the Bam HI generated rDNA fragments (tracks c, f, i and k). Two of the bands produced by Bst NI digestions, at 3.77 and 3.34kb had a greater signal strength than the rest. Other bands were noted at 2.44. 0.91 and 0.66kb. No band larger than 3.77kb was observed following Bst NI digestion. Comparison of the relative intensities of the Bst NI doublet (3.77 and 3.34kb) with the 2.44kb band, indicates a marked difference between the 11 and 15 d.a.f. cotyledon DNAs and the 22 d.a.f cotyledon and leaf DNAs.

Figure 21 shows the results obtained when Eco RII from different manufacturers was used to restrict different plasmid DNAs. Eco RII was shown to produce a large number of SV40 partial digest bands (track b). The addition of BSA did not improve the digestion (track d). Eco RII was also inefficient at cleaving ϕx174 and pBR322 (tracks f, g and i) both of which had been derived from methylase minus strains of bacteria.

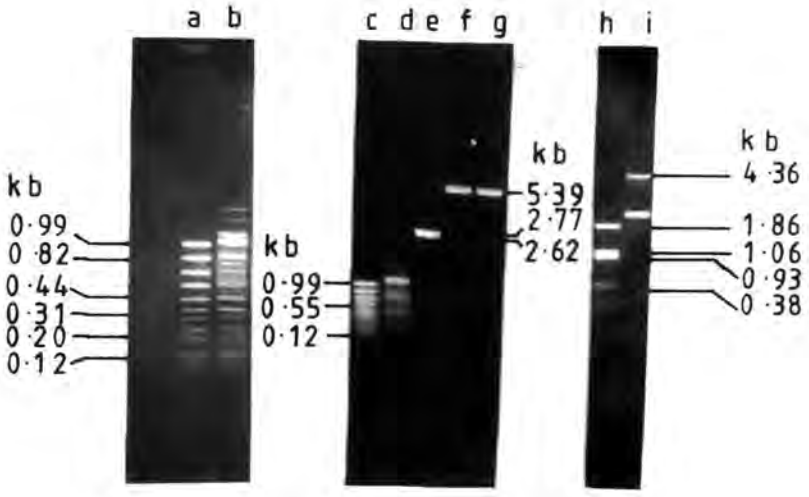


Figure 21:- Investigation of the restriction of SV40, ϕX174 RF and pBR322 by Bst NI and Eco RII (obtained from different suppliers) with and without bovine serum albumin (BSA).

a) SV40 + Bst NI (NEB)
b) SV40 + Eco RII (BRL)
c) SV40 + Bst NI (NEB) + BSA
d) SV40 + Eco RII (Uniscience) + BSA
e) ϕX174 + Bst NI (NEB) + BSA
f) ϕX174 + Eco RII (BRL) + BSA
g) ϕX174 + Eco RII (Uniscience) + BSA
h) pBR322 + Bst NI (NEB)
i) pBR322 + Eco RII (BRL-new batch of enzyme)

3:3 Methylation of the Legumin Genes

3:3:1 Changes in CG-methylation at sites associated with the legumin gene family .

Figure 22a shows the distribution of DNA fragments obtained following digestion by Hae III, Msp I and Hpa II. These distributions were similar to those previously described for figure 17a but the 5-fold higher loading of DNA on this gel, has enabled the visualisation of the bands of repetitive DNA. These were most prominent in the tracks containing leaf genomic DNA (k and l).

When a legumin cDNA was hybridized to the Southern blot prepared from this gel (figure 22b) major Hae III (5'-GGCC-3') bands were observed at 2.96, 2.63 and 0.99kb (tracks a, d, g and j). In addition three bands of a lesser intensity (3.49, 2.25 and 0.74kb)were also noted. Sequence data obtained from legumin gene A (figure 23) indicated that for this gene, the cDNA probe should detect two fragments (1.96kb plus a fragment at the 3' end of the gene, which is $\geq$1.2kb).

The cDNA probe (pDUB 6) hybridized to two major Msp I fragments (3.74 and 1.17kb) and also to two other bands with a lower signal intensity (3.10 and 0.51kb) (tracks b, e, h and k). The 0.51, 1.17 and 3.74kb were also evident in the Hpa II digests (tracks c, f, i and l). Four additional Hpa II bands were detected, of which the 5.37kb was a major band, whilst the 6.50, 4.19 and 1.69kb bands were of a lower signal strength. The 6.50kb band was quite distinct in track m (leaf + Hpa II) and track c (10 d.a.f. cotyledon + Hpa II), whilst being of a

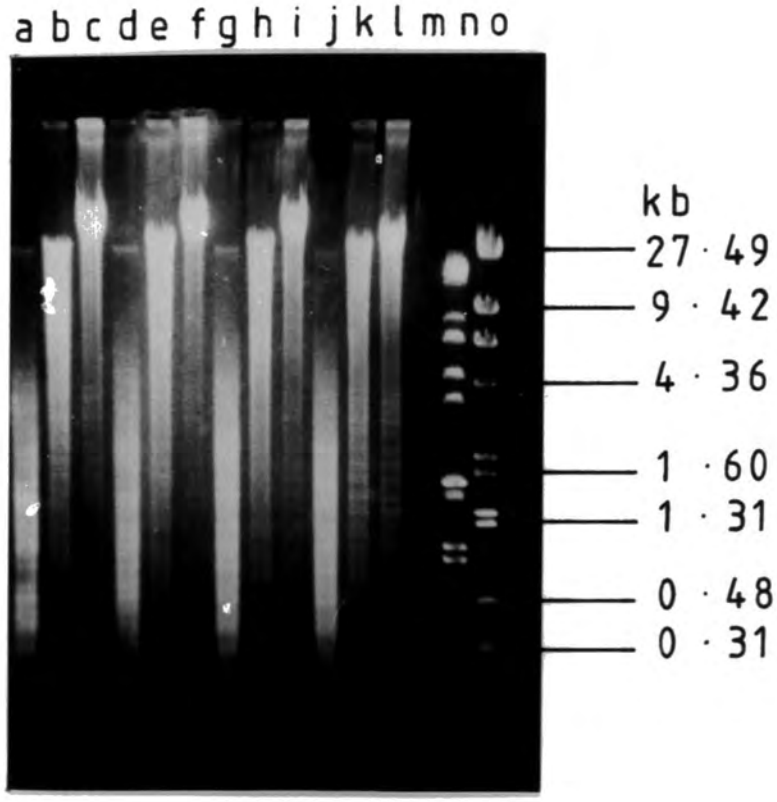
Figure 22a:- Gel photograph of Hae III, Msp I and Hpa II digested genomic DNA.

a) 10 d.a.f. cotyledon DNA + Hae III
b) " " " " + Msp I
c) " " " " + Hpa II
d) 15 d.a.f. cotyledon DNA + Hae III
e) " " " " + Msp I
f) " " " " + Hpa II
g) 21 d.a.f. cotyledon DNA + Hae III
h) " " " " + Msp I
i) " " " " + Hpa II
j) Leaf DNA + Hae III
k) " " + Msp I
l) " " + Hpa II
m) pDUB 1 (20pg)
n) χNM258 + Ava I; pBR322 + Bst NI
o) λNM258 + Hind III; pBR322 + Taq I

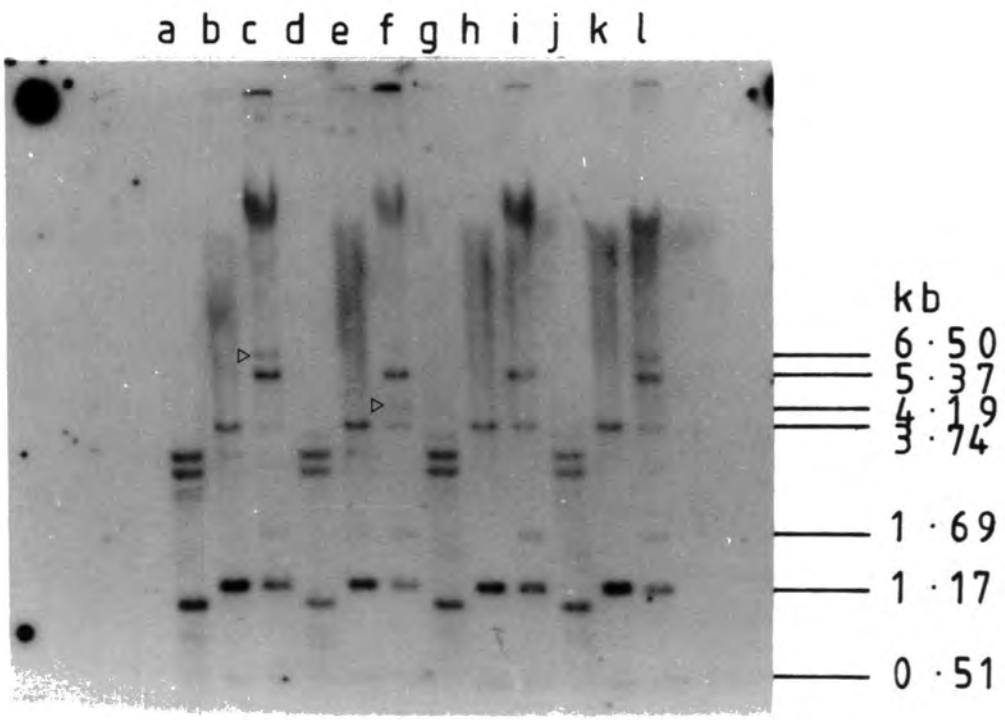
Genomic tracks contained 10 µg DNA and were digested with a 6-fold excess of enzyme.

Figure 22b:- Hybridization of ^{32}P-pDUB 6 to a Southern blot prepared from the gel shown above. The specific activity of the probe was 1×10^8 cpm/µg and the filter was hybridized at 60°C and washed at 50°C. The identities of the tracks are as indicated above.

a.


a b c d e f g h i j k l m n o
kb
27 · 49
9 · 42
4 · 36
1 · 60
1 · 31
0 · 48
0 · 31

b.


a b c d e f g h i j k l
kb
6 · 50
5 · 37
4 · 19
3 · 74
1 · 69
1 · 17
0 · 51

lower relative intensity in the 15 d.a.f. cotyledon + Hpa II (track f) and absent in 21 d.a.f. cotyledon DNA plus Hpa II (track i) (table 5b). The 4.19kb band, however, was absent in leaf and 10 d.a.f. cotyledon DNA (c and l), became visible 15 d.a.f. (track f) and was of a reduced intensity at a later stage on cotyledon development (track i). The relative intensities of the 3.74 and 1.69kb Hpa II bands, appeared to increase in intensity during cotyledon development, (table 5b). The Msp I and Hpa II 0.51kb bands appeared to be of a similar intensity, whilst the 4.19 and 1.17 bands were always of a greater relative intensity in Msp I tracks (Table 5a).

3:3:2 Methylation of three Msp I/Hpa II sites in legumin gene A

The coding and 5'flanking sequence of the cloned legumin gene, Leg A, predicts the presence of three Msp I/Hpa II sites (figure 23). Restriction fragments specific for each of these three sites were therefore selected as probes for genomic DNA. These were (i) Probe A, an Rsa I-Rsa I fragment, 1370 nucleotides, containing the Hpa II/Msp I site, M3, at position 1596 in the final exon of the coding sequence; (ii) Probe B, a Taq I-Taq I fragment, 1090 nucleotides containing the Hpa II/Msp I site, M2, at position 435 in the second exon of the protein coding sequence (near the 5' end); (iii) Probe C, a Bam I-Taq I fragment, 1349 nucleotides, containing the Hpa II/Msp I site, M1 at position -546 in the 5' flanking sequence. The probes were isolated and labelled as described in the Methods section.

Table 5: Results from laser densitometer scanning of autoradiograph (22b)

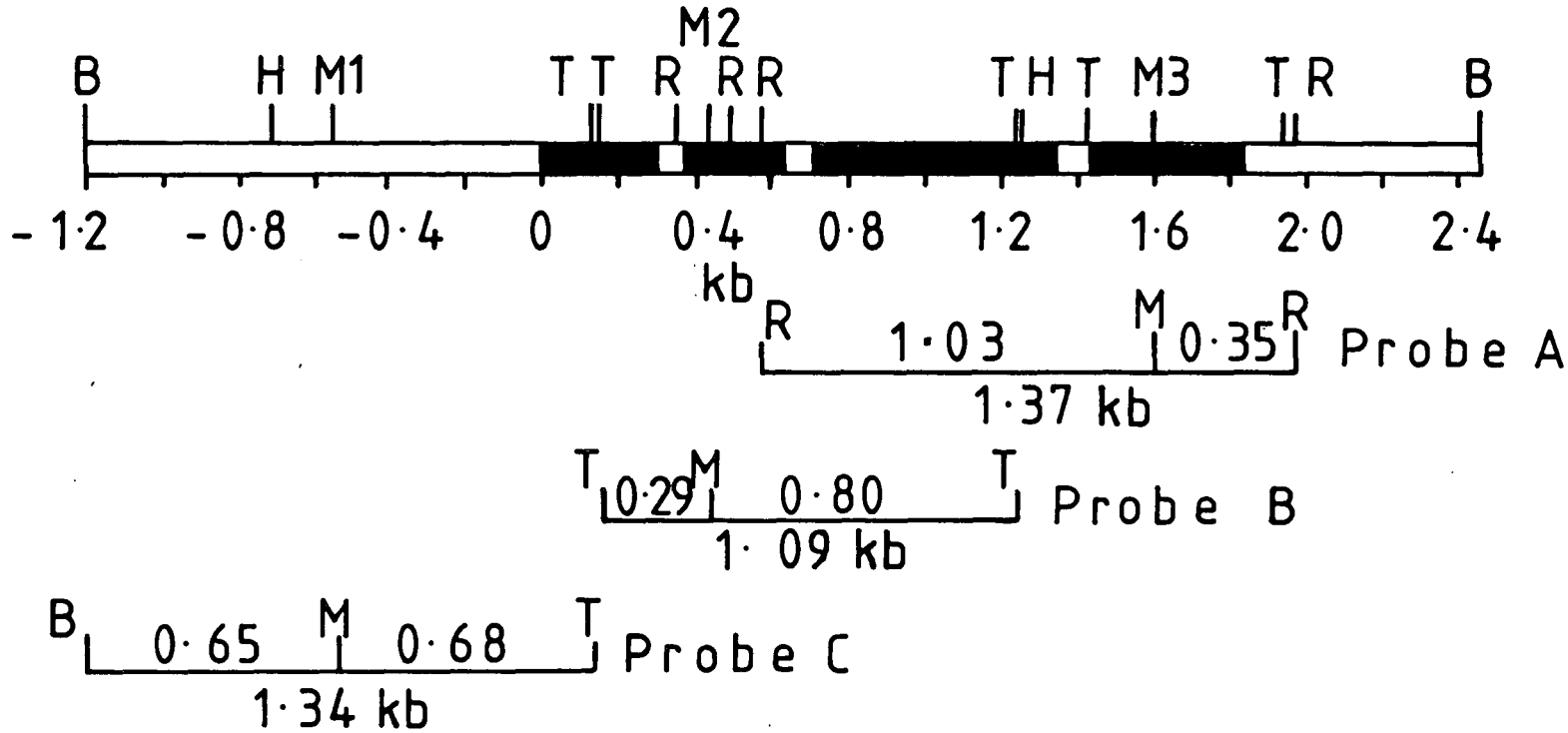
5a

	PEAK AREA ($x10^7$)							
BAND SIZE (kb)	COTYLEDON 10 d.a.f. Msp I	Hpa II	15 d.a.f. Msp I	Hpa II	21 d.a.f Msp I	Hpa II	LEAF Msp I	Hpa II
6.50		0.82		0.42				1.26
5.37		3.12		2.17		1.04		1.78
4.19				0.55				
3.74	1.57	0.60	1.80	0.66	1.01	0.74	1.30	0.33
1.69		0.79		0.55		0.63		0.43
1.17	5.23	3.68	3.92	1.73	4.07	2.67	5.52	2.98
0.51	0.52	0.64	0.37	0.18	0.30	0.31	0.30	0.27

5b

	AREA OF PEAK/AREA OF ALL PEAKS IN TRACK (%)							
BAND SIZE (kb)	COTYLEDON 10 d.a.f. Msp I	Hpa II	15 d.a.f. Msp I	Hpa II	21 d.a.f. Msp I	Hpa II	LEAF Msp I	Hpa II
6.5		8.52		6.78				17.93
5.37		32.30		34.62		19.24		25.24
4.19				8.73				
3.74	20.67	6.22	29.53	10.60	18.74	13.74	18.31	4.72
1.69		8.15		8.72		11.63		6.11
1.17	68.66	38.16	64.38	27.60	75.66	49.63	77.52	42.24
0.51	6.78	6.67	6.09	2.95	5.61	5.76	4.16	3.77

Figure 23:- A simplified restriction map of pDUB 24 (legumin gene A), showing the three fragments isolated as probes to ascertain the methylation status of the three Msp I sites (M1, M2 and M3). The following recognition sites are indicated: B = Bam HI; T = Taq I; R = Rsa I; M= Msp I; H = Hae III. The solid boxes represent the coding regions (exons) and the open boxes the intervening sequences (introns). The unthickened line represents the flanking sequences.


M2
B H M1 T T R R R T H T M3 T R B
-1·2 -0·8 -0·4 0 0·4 0·8 1·2 1·6 2·0 2·4
kb
R 1·03 M 0·35 R
1·37 kb
Probe A
T 0·29 M 0·80 T
1·09 kb
Probe B
B 0·65 M 0·68 T
1·34 kb
Probe C

3:3:2:1 Methylation of Msp I site M3, in the 3' coding region of Leg A

The hybridization of probe A to digests of genomic DNA, isolated from leaves and developing cotyledons, is shown in figure 24b. This probe detected a major band (1.37kb) in the Rsa I digests, (tracks a,d,g and i), by hybridization to its identical sequence in genomic DNA. Minor bands at 1.88, 0.85 and 0.56kb gave relatively very faint hybridization signals.

In the Rsa I + Msp I digest (tracks b, e, h and k), two bands of hybridization (1.03 and 0.35kb) were observed at the positions predicted for the Rsa I + Msp I fragments of this 1.37kb probe (figure 23). In all Rsa I + Hpa II digests (lanes c,f, i and l), the 1.37kb band was only partially cleaved by Hpa II and the three fragments, 1.37, 1.03 and 0.35kb were visible in all Hpa II digests. The 1.88kb minor band observed after Rsa I digestion, was cleaved by both Msp I and Hpa II but the other two minor bands (0.85 and 0.56kb) remained uncleaved after the second restriction.

Overall the strength as well as the pattern of hybridization signal were similar in all genomic DNA digests. However, in the 12 and 21-day digests (track c and i) the 1.03kb hybridization band gave a stronger signal than the 1.37kb hybridization band, whilst in the 15-day cotyledons and leaf digests (tracks f and l), these two signals were of a similar intensity (table 6).

A similar pattern of hybridization and of signal strength of these bands was also observed when Rsa I and Rsa I+Msp I or Rsa I + Hpa II digests of cotyledon and leaf genomic DNA, were probed with a (^{32}P)-1.09kb Taq I fragment of legumin gene A (probe B in figure 23)

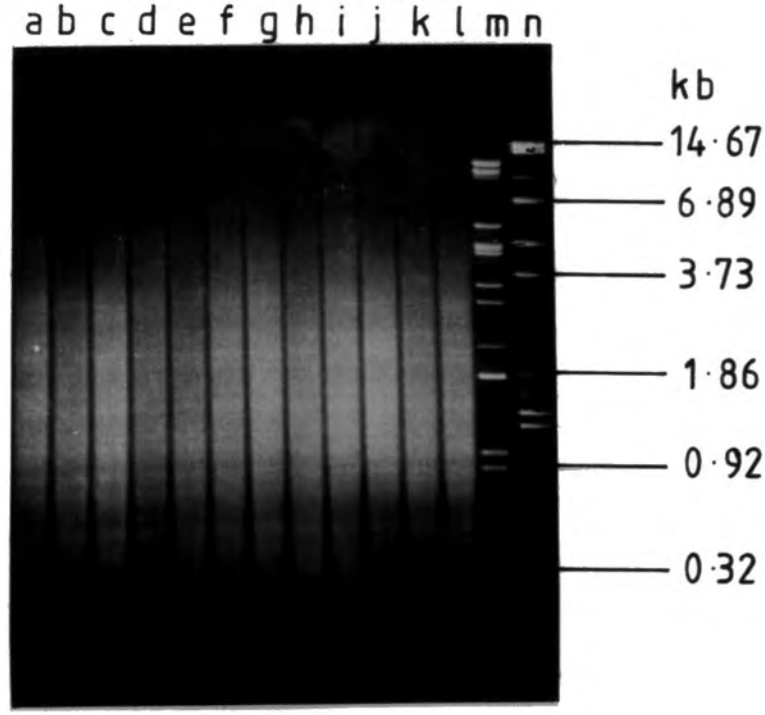
Figure 24a:- Gel photograph of double restricted (Rsa I plus Msp I/ Hpa II) genomic DNA.

a) 12 d.a.f. cotyledon DNA + Rsa I
b) " " " " + Rsa I + Msp I
c) " " " " + Rsa I + Hpa II
d) 15 d.a.f. cotyledon DNA + Rsa I
e) " " " " + Rsa I + Msp I
f) " " " " + Rsa I + Hpa II
g) 21 d.a.f. cotyledon DNA + Rsa I
h) " " " " + Rsa I + Msp I
i) " " " " + Rsa I + Hpa II
j) Leaf DNA + Rsa I
k) " " + Rsa I + Msp I
l) " " + Rsa I + Hpa II
m) λNM258 + Hpa I; pBR322 + Bst NI
n) λNM258 + Ava I; pBR322 + Taq I

Genomic tracks contained 10 µg DNA digested with a 6-fold excess of Rsa I and an 8-fold excess of Msp I or Hpa II.

Figure 24b:- Hybridization of ^{32}P-Probe A (1.37kb RsaI fragment of legumin gene A, figure 23), to a Southern blot prepared from the gel illustrated above. The specific activity of the probe was 2×10^8 cpm/ µg. The filter was hybridized at 60°C and washed at 50°C.

a.


a b c d e f g h i j k l m n
kb
14·67
6·89
3·73
1·86
0·92
0·32

b.

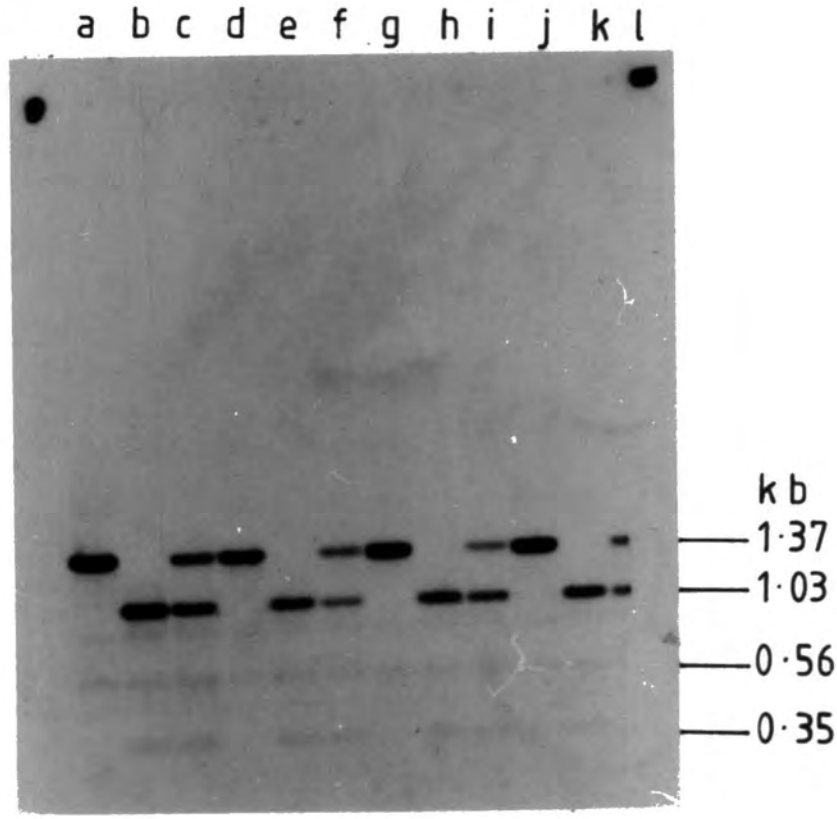

a b c d e f g h i j k l
kb
1·37
1·03
0·56
0·35

Table 6: Percentage methylation of sites M2 and M3 in Legumin Genes A and B

Developmental Stage	PERCENTAGE METHYLATION			
	M2		M3	
	Leg A	Leg B	Leg A	Leg B
12 d.a.f. coty	25.8	15.3	36.9	38.9
15 d.a.f. coty	23.9	16.0	45.6	44.2
21 d.a.f. coty	22.2	14.0	37.4	42.9
Imbibed coty		6.06		32.6
Leaf	26.8	11.25	47.3	47.5

The 12, 15 and 21 d.a.f. cotyledon Leg B results represent the mean values obtained from densitometric scanning of figures 28a and b and figure 35. The Leg A results are the mean values obtained from several experiments, including the results illustrated in figures 24b and 25b.

and when a Taq I and Taq I + Msp I or Taq I + Hpa II genomic digests, were probed with a (^{32}P)-0.52kb Taq I fragment of legumin gene A (which contains site M3) (results not shown).

3:3:2:2 Methylation of Msp I site M2, in the 5' coding region of Leg A

Figure 25b shows an autoradiograph of cotyledon and leaf Taq I and Taq I plus Msp I/Hpa II DNA fragments, which hybridized to a 1.09kb legumin gene A Taq I-Taq I fragment, (^{32}P)-DNA probe (figure 23, probe B). A major band of hybridization at 1.09kb was observed in the Taq I digested genomic DNA (tracks a, d, g and j) representing hybridization of the probe to its identical sequence in the genomic DNA. 2.75 , 2.54 and 1.28kb Taq I fragments also hybridized to this probe.

When Taq I digested genomic DNA was subsequently restricted with Msp I (tracks b, e, h and k) the 1.09 and 2.75kb fragments were cleaved and the signal strength of the 2.54kb band was increased.

Following Taq I plus Hpa II digestion of genomic DNA (lanes c, f, i and l), the 1.09kb Taq I fragment was only partially cleaved by Hpa II and the three fragments at 1.09, 0.80 and 0.29kb, all hybridized to the probe. The 2.75kb fragment was also only partially cleaved.

The Hpa II, 0.80kb fragment always had a greater signal strength than the uncleaved 1.09kb Taq I fragment. However, both the pattern of hybridization signal and the relative strength of the signal in the Hpa II digested cotyledon and leaf DNA were similar (table 6).

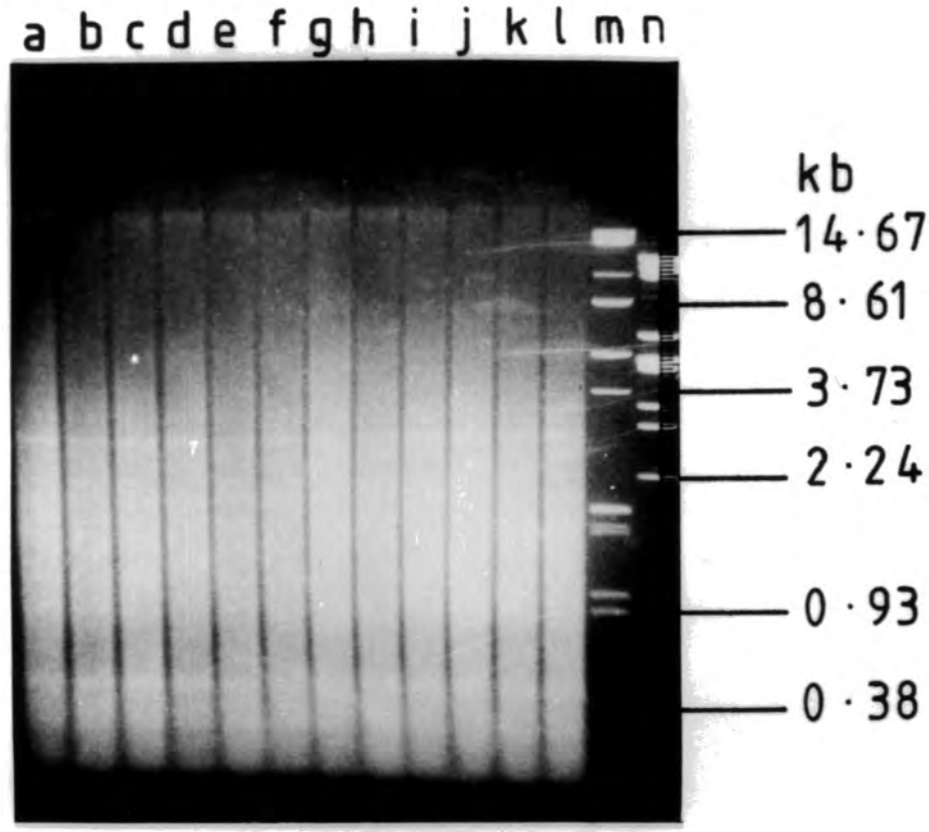
Figure 25a:- Gel photograph of double restricted (Taq I plus Msp I/Hpa II) genomic DNA.

a) 12 d.a.f. cotyledon DNA + Taq I
b) " " " " + Taq I + Msp I
c) " " " " + Taq I + Hpa II
d) 15 d.a.f. cotyledon DNA + Taq I
e) " " " " + Taq I + Msp I
f) " " " " + Taq I + Hpa II
g) 21 d.a.f. cotyledon DNA + Taq I
h) " " " " + Taq I + Msp I
i) " " " " + Taq I + Hpa II
j) Leaf DNA + Taq I
k) " " + Taq I + Msp I
l) " " + Taq I + Hpa II
m) λNM258 + Ava I; pBR322 + Bst NI
n) λNM258 + Hpa I

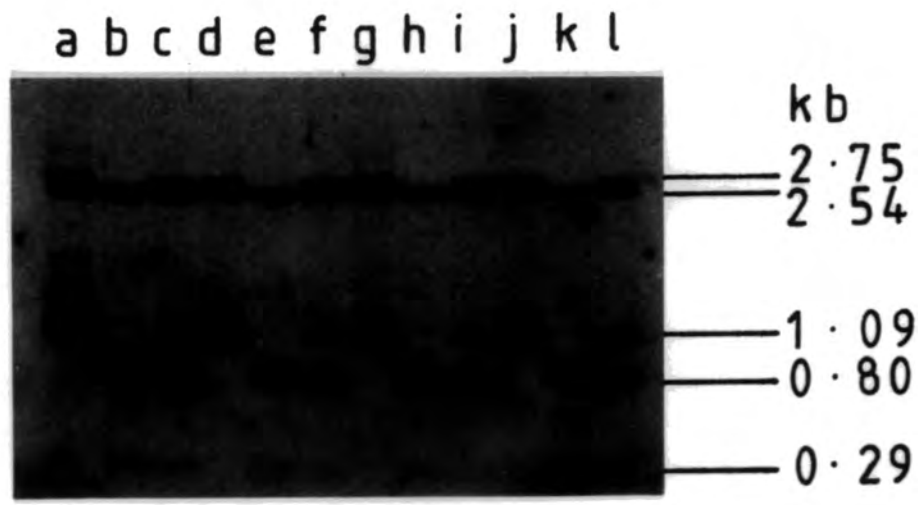
All genomic tracks contained 10 µg DNA and digestions were with a 4-fold excess of Taq I and a 7-fold excess of Msp I and Hpa II.

Figure 25b:- Hybridization of ^{32}P-Probe B (1.09kb Taq I fragment of legumin gene A - figure 23), to a Southern blot prepared from the gel illustrated above. The specific activity of the probe was 1.1 x 10^8 cpm/µg. The filter was hybridized at 62°C and washed at 50°C.

a.


a b c d e f g h i j k l m n
kb
14·67
8·61
3·73
2·24
0·93
0·38

b.


a b c d e f g h i j k l
kb
2·75
2·54
1·09
0·80
0·29

3:3:2:3 Methylation of Msp I site M1, in the 5'flanking region of legumin gene A

Figure 26 shows an autoradiograph of Taq I plus Msp I/Hpa II DNA fragments hybridizing to a 1.33kb Bam HI-Taq I fragment of legumin gene A, (^{32}P)-DNA probe (figure 23: probe C). This probe hybridized strongly to four Taq I fragments of 1.57, 1.49, 1.40 and 1.25kb (tracks a, d, g and i). The signals for the hybridization of this probe to the 1.40 and 1.49kb fragments were greater than those observed for the hybridization of the probe to the 1.57 and 1.25kb fragments. Minor bands of hybridization were detected at 1.06, 2.51 2.66 and 4.22kb.

When Taq I digested genomic DNA was subsequently restricted with Msp I (tracks b,e,h and k) or with Hpa II (tracks c, f, i and l), the major Taq I 1.49kb fragment was cleaved and two fragments of 0.88 and 0.66kb were detected by the probe. This 1.49kb band was taken to represent the identical sequence in genomic DNA to the probe, since it contained a Hpa II/Msp I site and gave a correctly sized fragment on cleavage. Overall, the strength as well as the pattern of hybridization signal, with Msp I and Hpa II digested cotyledon and leaf genomic DNA were similar.

3:3:3 Methylation of legumin gene B

Figure 27 shows the hybridization pattern obtained when a cDNA probe (pDUB 6) was hybridized to Southern blotted, restricted fragments of legumin genes A, B and C (pDUB 24, pDUB 25, pDUB 27). The

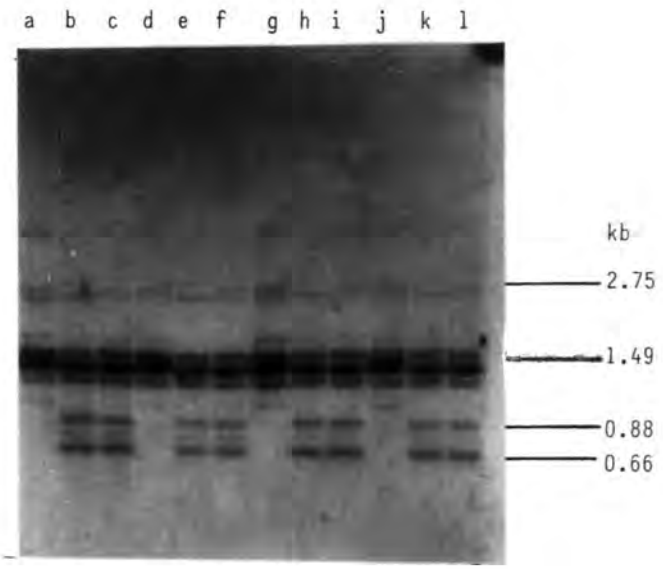


Figure 26:- Hybridization of ^{32}P-Probe C (1.34kb Taq I - Bam HI fragment of legumin gene A, figure 23), to a Southern blot prepared from a gel similar to that illustrated in figure 25a.
The specific activity of the probe was 7 x 10^7 cpm/µg and the filter was hybridized at 62°C and washed at 50°C. The identities of the tracks are as indicated in figure 25a.

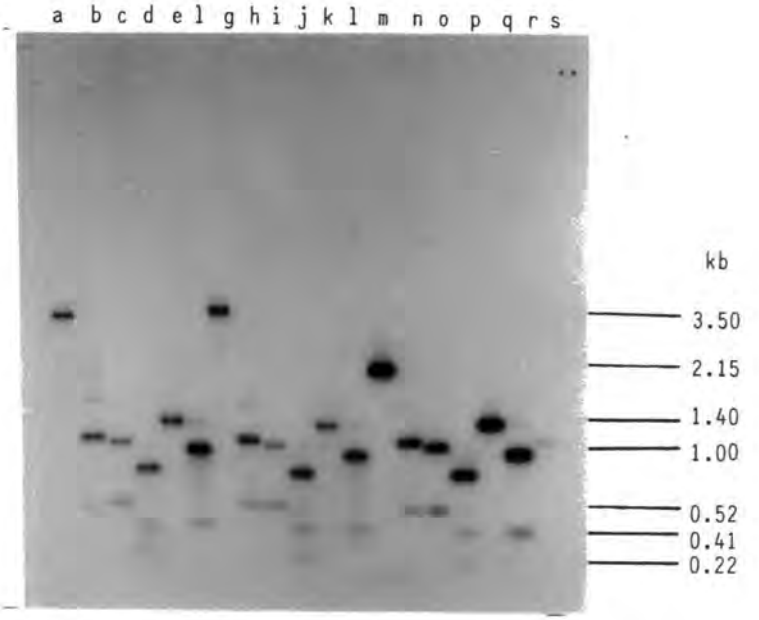


Figure 27:- Hybridization of ^{32}P-pDUB 6 insert (legumin cDNA) to Southern blotted, restricted fragments of the cloned legumin genes A,B. and C (pDUB 24, pDUB 25 and pDUB 27).

a) Legumin gene A (pDUB 24 insert)
b) " " " + Msp I
c) " " " + Taq I
d) " " " + Taq I + Msp I
e) " " " + Rsa I
f) " " " + Rsa I + Msp I
g) Legumin gene B (pDUB 25 insert)
h) " " " + Msp I
i) " " " + Taq I
j) " " " + Taq I + Msp I
k) " " " + Rsa I
l) " " " + Rsa I + Msp I
m) Legumin gene C (pDUB 37 insert)
n) " " " + Msp I
o) " " " + Taq I
p) " " " + Taq I + Msp I
q) " " " + Rsa I
r) " " " + Rsa I + Msp I
s) pDUB 6 insert

The specific activity of the probe was 1.8×10^8 cpm/μg and the filter was hybridized and washed at 65°C.

single digest (Msp I, Taq I and Rsa I) fragments and those of the double digests (Rsa I + Msp I and Taq I + Msp I), were shown to be of a similar size from each of the three legumin gene coding sequences. The lengths of the isolated genes (figure 27, tracks a, g and m; figure 29a) were different and this accounted for the differences observed in the sizes of the larger restriction fragments, derived from the 5' ends of these genes.

When Taq I and Rsa I probes, prepared from Leg B (pDUB 25) were hybridized to restricted genomic DNA (figures 28a and b), similar patterns of hybridization were observed as have been previously described for legumin gene A (figures 24b and 25b). A comparison of the relative intensities of the bands (as determined by a laser densitometer) is shown in table 6.

The extent of methylation of Msp I site M3, was similar in both Leg A and Leg B. However, the M2 site, appeared to be less methylated in Leg B than in Leg A. In both legumin genes, the M3 site was more highly methylated than the M2 site.

3:3:4 Methylation of the flanking sequences of Leg A and Leg B

The location of the flanking regions probes is shown in figure 29b. Legumin gene B does not have an Msp I site equivalent to Leg A's M1 site. It was therefore necessary to select a suitable restriction enzyme (Hind III) which would identify the genomic sequence immediately 5' to that isolated in the λLeg 2 genomic clone.

Figure 28a:- Hybridization of ^{32}P-1.09kb Taq I fragment of legumin gene B (pDUB 25), to a Southern blot of a Taq I plus Msp I/Hpa II gel, similar to that illustrated in figure 25a.

a) Imbibed cotyledon DNA + Taq I
b) " " " + Taq I + Msp I
c) " " " + Taq I + Hpa II
d) 12 d.a.f. cotyledon DNA + Taq I
e) " " " + Taq I + Msp I
f) " " " + Taq I + Hpa II
g) 15 d.f.a. cotyledon DNA + Taq I
h) " " " + Taq I + Msp I
i) " " " + Taq I + Hpa II
j) 21 d.a.f. cotyledon DNA + Taq I
k) " " " + Taq I + Msp I
l) " " " + Taq I + Hpa II

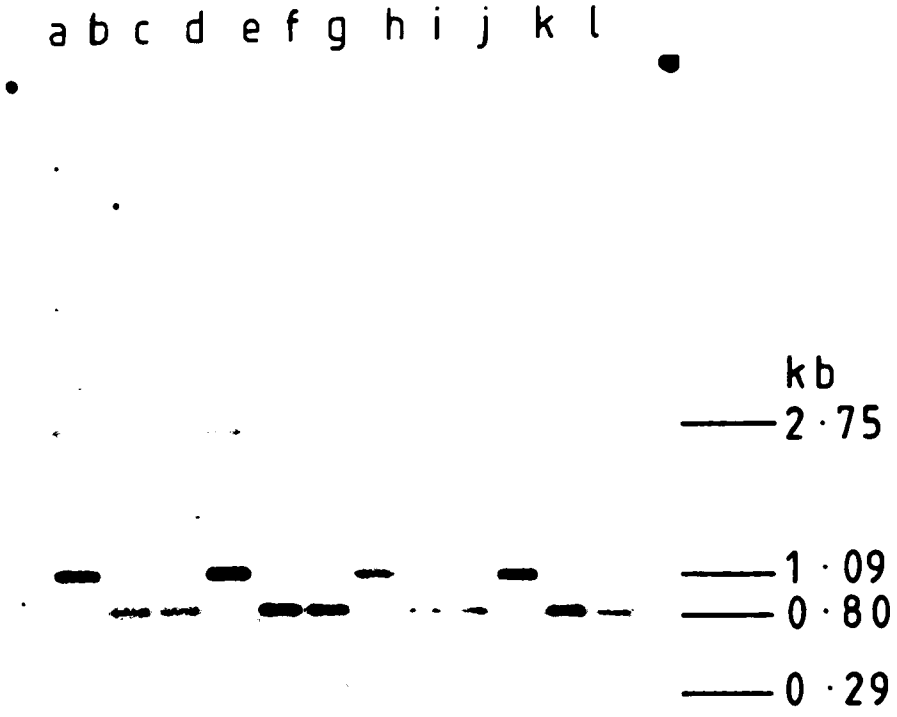
Each genomic track contained 7.5 μg DNA digested with a 12-fold excess of Taq I and a 10-fold excess of Msp I/Hpa II. The specific activity of the probe was 7.2. x 10^7 cpm/μg and the filter was hybridized at 65°C in the heparin system.

Figure 28b:- Hybridization of ^{32}P-1.37kb Rsa I fragment of legumin gene B (pDUB 25), to a Southern blot of a n Rsa I plus Msp I/Hpa II gel, similar to that shown in figure 24a.

a) Imbibed cotyledon DNA + Rsa I
b) " " " + Rsa I + Msp I
c) " " " + Rsa I + Hpa II
d) 12 d.a.f. cotyledon DNA + Rsa I
e) " " " + Rsa I + Msp I
f) " " " + Rsa I + Hpa II
g) 15 d.a.f. cotyledon DNA + Rsa I
h) " " " + Rsa I + MspI
i) " " " + Rsa I + Hpa II
j) 21 d.a.f. cotyledon DNA + Rsa I
k) " " " + Rsa I + Msp I
l) " " " + Rsa I + Hpa II

Each genomic track contained 7.5 μg DNA restricted with a 10-fold excess of Rsa I and a 10-fold excess of Msp I or Hpa II. The specific activity of the probe was 6.5 x 10^7 cpm / μg and the filter was hybridized at 65°C (in the heparin system).

a.



b.

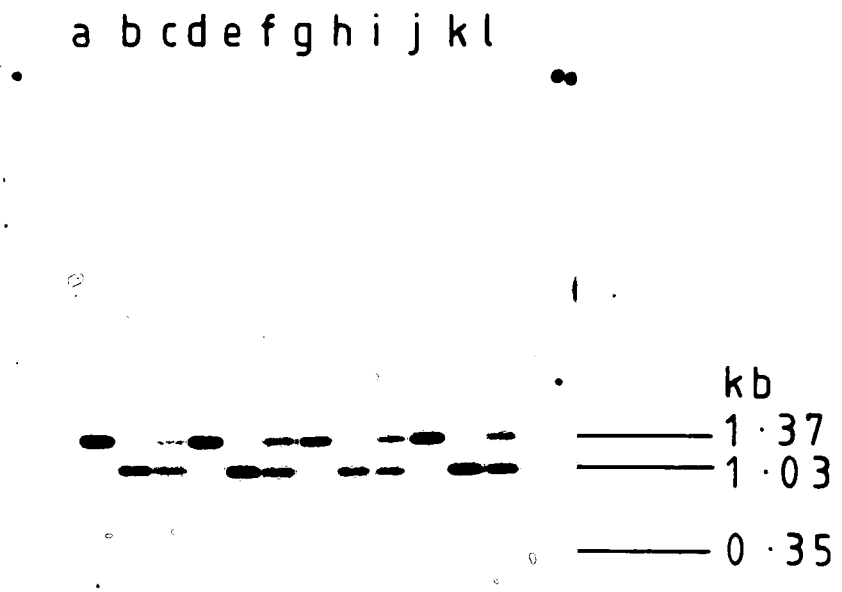
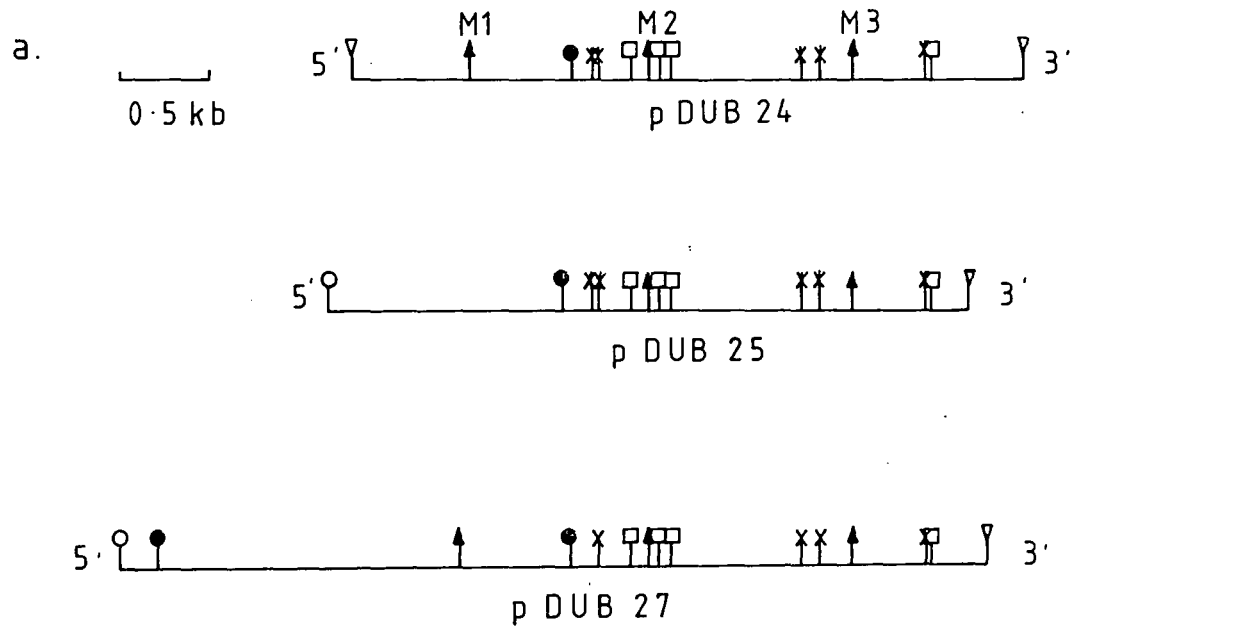


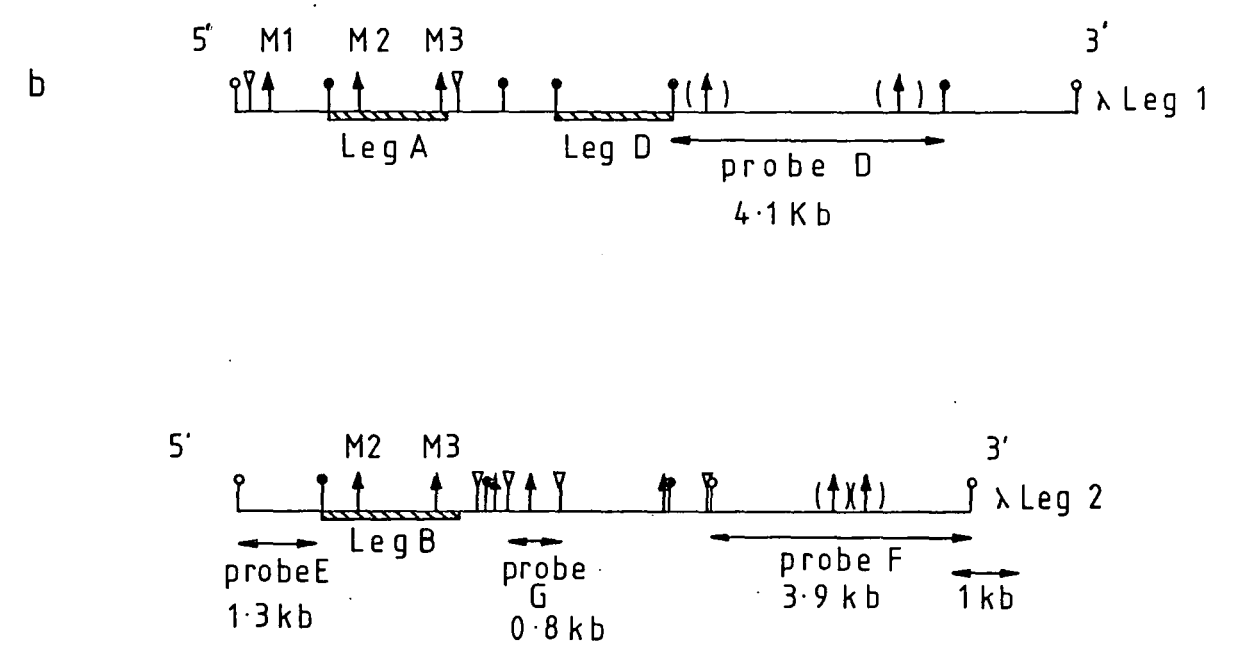
Figure 29a:- Restriction map of legumin genes A,B, and C (pDUB 24, 25 and 27).

The relative positions of Rsa I ( ), Taq I ( ), Msp I ( ), Bam HI ( ), Eco RI ( ) and Hind III ( ) sites in the three genes are indicated. The Msp I sites M1, M2 and M3 in legumin gene A are also indicated and the equivalent sites have been used to align legumin genes B and C.

Figure 29b:- Restriction maps of the genomic clones, λ Leg 1 and λ Leg 2.

The symbols used are explained above. An Msp I symbol in brackets, ( ), indicates that the site is in one of two potential positions. The relative positions of legumin genes A and B and the pseudogene D ae indicated. The origins of the flanking region probes (D,E,F and G), are also indicated.


a.
M1
M2
M3
5'
3'
0·5 kb
p DUB 24
5'
3'
p DUB 25
5'
3'
p DUB 27


b
5'
M1
M2
M3
3'
λ Leg 1
Leg A
Leg D
probe D
4·1 Kb
5'
M2
M3
3'
λ Leg 2
Leg B
probeE
1·3kb
probe
G
0·8 kb
probe F
3·9 kb
1kb

3:3:4:1 Methylation of an Msp I site, 3' to legumin gene A

The next Msp I site 3' to site M3 in λLeg 1, is in a 4.17kb Hind III fragment (probe D). When this fragment was isolated from the genomic clone λLeg 1 and restricted with Msp I, two fragments (3.40 and 0.77kb) were produced (results not shown). When the 4.17 kb fragment was labelled and hybridized to double digested genomic DNA (figure 30), a series of Hind III fragments (10.00, 8.51. 6.79. 4.17. 3.29. 2.90 and 2.54kb) were detected (tracks a, d, g, and j).

Following digestion by Msp I the 10.00, 2.90 and 2.54kb fragments were completely cleaved whilst the 8.51, and 4.17kb bands were only partially cleaved (tracks b, e, h and k). New bands were observed at 3.40, 2.32, 2.06. 1.66. 1.16. 0.60 and 0.41kb. When the relative intensities of the 4.17 and 3.40kb bands were assessed it was calculated that between 40 - 50% of the 4.17kb band remained uncleaved, which is equivalent to a 40-50% methylation of the external cytosine (^{m}CCGG).

Following digestion with Hpa II, the 4.17kb band remained uncleaved, whilst the 2.54kb band was partially cleaved and the 2.90kb band was completely cleaved. New bands were seen at 2.32, 1.16. 0.60 and 0.41kb. No obvious change in the relative intensities of any of the bands in cotyledon or leaf DNA was observed.

3:3:4:2 Methylation of 5' flanking region to legumin gene B

When Hind III digested genomic DNA was hybridized to a probe prepared from the 5' flanking region of legumin gene B (figure 29b,

Figure 30:- Hybridization of ^{32}P-Probe D (4.1kb Hind III fragment of Leg 1) to a Southern blot of double digested (Hind III plus Msp I/Hpa II) genomic DNA.

a) 12 d.a.f. cotyledon DNA + Hind III
b) " " " " + Hind III + Msp I
c) " " " " + Hind III + Hpa II
d) 15 d.a.f. cotyledon DNA + Hind III
e) " " " " + Hind III + Msp I
f) " " " " + Hind III + Hpa II
g) 21 d.a.f. cotyledon DNA + Hind III
h) " " " " + Hind III + Msp I
i) " " " " + Hind III + Hpa II
j) Leaf DNA + Hind III
k) " " + Hind III + Msp I
l) " " + Hind III + Hpa II

Each genomic track contained 10μg DNA restricted with a 10-fold excess of Hind III, Msp I and Hpa II. The specific acivity of the probe was 1.3×10^8 cpm/μg and the filter was hybridized at 65°C and washed at 50°C.

Figure 31:- Hybridization of ^{32}P-Probe E, a 1.3kb Hind III fragment, the 5' flanking region of legumin gene B (pDUB 25), to a Southern blot of double restricted (Hind III plus Msp I/Hpa II) genomic DNA.

a) Leaf DNA + Hind III
b) " " + Hind III + Msp I
c) " " + Hind III + Hpa II
d) 12 d.a.f. cotyledon DNA + Hind III
e) " " " " + Hind III + Msp I
f) " " " " + Hind III + Hpa II
g) 15 d.a.f. cotyledon DNA + Hind III
h) " " " " + Hind III + Msp I
i) " " " " + Hind III + Hpa II
j) 21 d.a.f. cotyledon DNA + Hind III
k) " " " " + Hind III + Msp I
l) " " " " + Hind III + Hpa II

All tracks contained 8μg DNA digested with a 10-fold excess of Hind III, Msp I and Hpa II. The specific activity of the probe was 1.2×10^8 cpm/μg and the filter was hybridized in the heparin system at 62°C and washed at 50°C.

Fig. 30

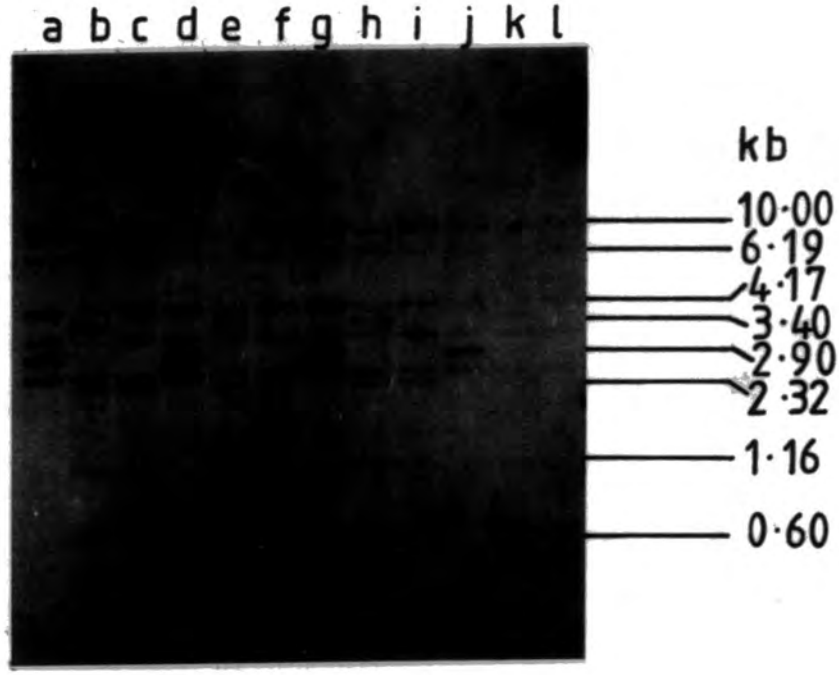
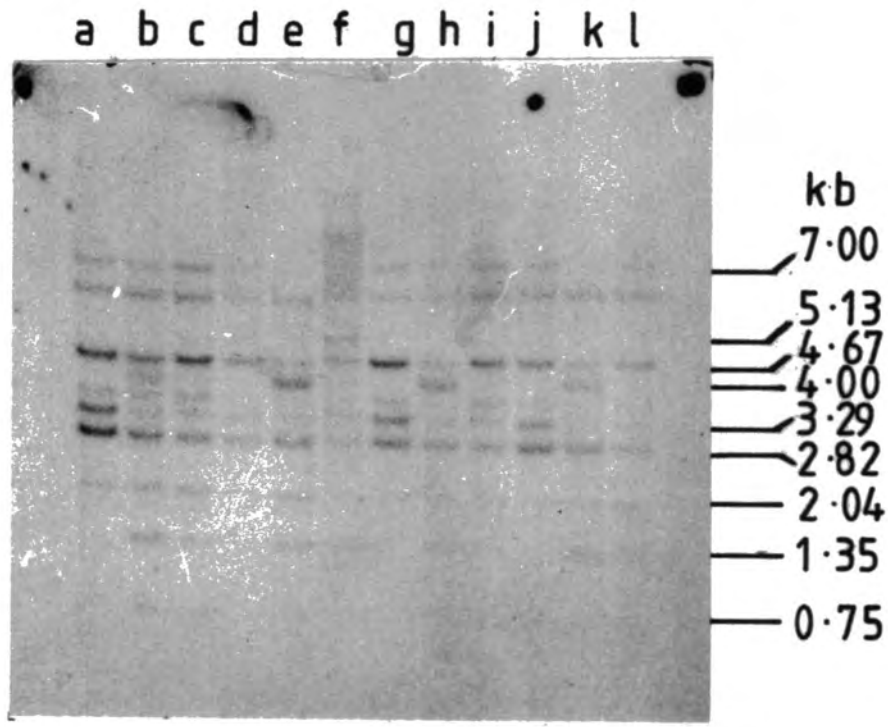

a b c d e f g h i j k l
kb
10·00
6·19
4·17
3·40
2·90
2·32
1·16
0·60

Fig. 31


a b c d e f g h i j k l
kb
7·00
5·13
4·67
4·00
3·29
2·82
2·04
1·35
0·75

probe E) three major bands were identified (4.67, 3.29 and 2.82kb) (figure 31: tracks a, d, g and j). In addition several fainter bands were also detected at 8.51. 7.00, 3.65, 2.04 and 1.86kb.

Following digestion with Msp I (tracks b, e, h and k), all three major Hind III bands were partially cleaved and new bands were apparent at 4.00, 2.38. 1.35 and 0.75kb. Following Hind III plus Hpa II digestion (tracks c, f, i and l), the 4.67kb Hind III fragment was not cleaved but the 3.29 and 2.82kb bands were again partially cleaved and new bands were found at 2.38, 1.35 and 0.75kb. The 3.29kb band was cleaved to a greater extent by Msp I than by Hpa II and also the 1.35kb band was of a greater intensity following Msp I digestion. The 0.75kb band appeared to be of a similar intensity after both Msp I and Hpa II digestion.

The relative intensities of the hybridization bands in the leaf and 15 and 21 d.a.f. cotyledon patterns were similar. However, the pattern obtained in all three tracks of the 12 d.a.f. cotyledon had reduced intensities for the larger fragments as well as an additional Hpa II band (5.13kb, track f). It should perhaps be noted that this was a fresh preparation of 12 d.a.f. cotyledon genomic DNA.

3:3:4:3 Methylation of Msp I sites 3' to legumin gene B

A 3.75kb Eco RI fragment (figure 29b: probe F) was isolated from the genomic clone (λLeg 2). This Eco RI fragment can be restricted with Msp I to give two fragments of 2.00 and 1.75kb (results not shown).

Probe F detected a wide range of Eco RI derived fragments (1.53 to >11kb) (figure 32). Following restrictions with either Msp I or Hpa II, an Eco RI band of 2.95kb was cleaved to give two new bands 2.30 and 0.56kb. However, the 3.75kb was not cut by either Msp I or Hpa II. The multiplicity of genomic fragments cross hybridizing to this probe, precluded further assessment.

A simpler hybridization pattern was obtained (figure 33), when a smaller length of the legumin gene B 3' flanking region was used as a probe (ie the 0.8kb Bam HI-Bam HI fragment, Probe G, figure 29b), which it is known has one Msp I site and if cleaved will give two fragments (0.65 and 0.15kb).

Following digestion by Bam HI a major band at 0.80kb was detected (figure 33: tracks a, d, g and j). Several minor bands were also noted, the most prominent being 6.60 and 5.57kb. Following digestion by either Msp I or Hpa II the major band at 0.80kb remained uncleaved, however, the 6.60kb minor band was cleaved by both of these enzymes. In the Bam HI + Msp I digests (tracks b, e, h and k) the loss of the 6.60kb band correlated with the appearance of new bands at 3.89, 1.25, 0.89 and 0.56kb, whilst following Hpa II digestion, new bands equivalent in size to 4.57 and 2.00kb were observed (tracks c, f and i).

Figure 32:- Hybridization of ^{32}P-Probe F, a 3.8kb Eco RI fragment of λLeg 2 (3' to legumin gene B), to a Southern blot of double restricted (Eco RI plus Msp I/Hpa II) genomic DNA.

a) Leaf DNA + Eco RI
b) " " + Eco RI + Msp I
c) " " + Eco RI + Hpa II
d) 12 d.a.f. cotyledon DNA + Eco RI
e) " " " " + Eco RI + Msp I
f) " " " " + Eco RI + Hpa II
g) 15 d.a.f. cotyledon DNA + Eco RI
h) " " " " + Eco RI + Msp I
i) " " " " + Eco RI + Hpa II
j) 21 d.a.f. cotyledon DNA + Eco RI
k) " " " " + Eco RI + Msp I
l) " " " " + Eco RI + Hpa II

All tracks contained 8μg DNA digested with a 10-fold excess of enzyme. The specific activity of the probe was 4 x 10^7 cpm/μg and the filter was hybridized at 65°C in the SSC system, modified to include 10 x Denhardts. The filter was washed at 65°C.

Figure 33:- Hybridization of ^{32}P-Probe G, a 0.8kb Bam HI fragment of λLeg 2 (3' to legumin gene B), to a Southern blot of double restricted (Bam HI plus Msp I/Hpa II) genomic DNA.

a) Imbibed cotyledon DNA + Bam HI
b) " " " + Bam HI + Msp I
c) " " " + Bam HI + Hpa II
d) 12 d.a.f. cotyledon DNA + Bam HI
e) " " " + Bam HI + Msp I
f) " " " + Bam HI + Hpa II
g) 15 d.a.f. cotyledon DNA + Bam HI
h) " " " + Bam HI + Msp I
i) " " " + Bam HI + Hpa II
J) 21 d.a.f. cotyledon DNA + Bam HI
k) " " " + Bam HI + Msp I

All tracks contained 7.5μg DNA digested with a 12-fold excess of enzyme. The specific activity of the probe was 5 x 10^7 cpm/μg and the filter was hybridized at 65°C in the SSC system, modified to contain 10 x Denhardts. The filter was washed at 65°C.

Fig. 32

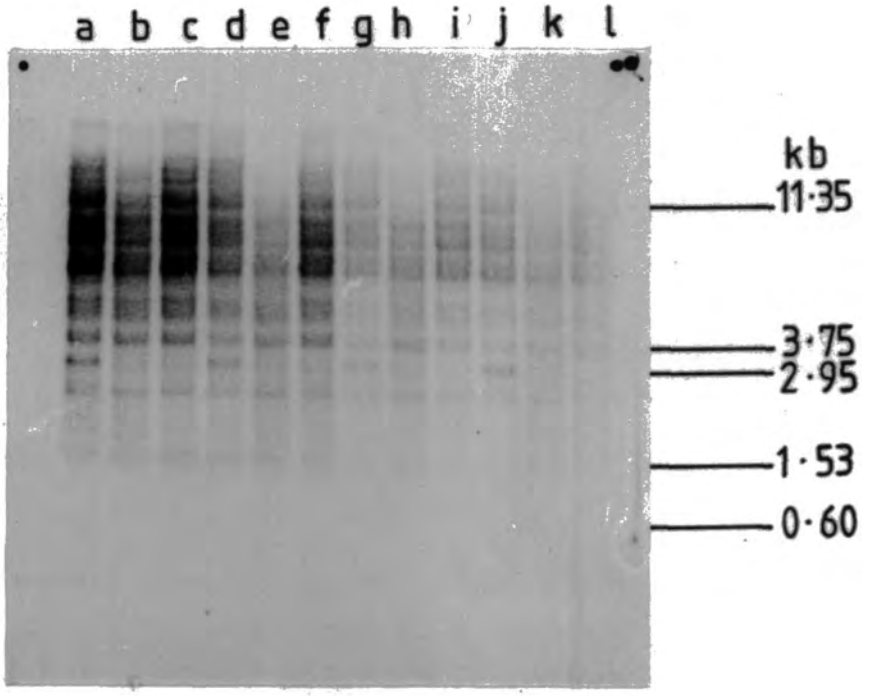
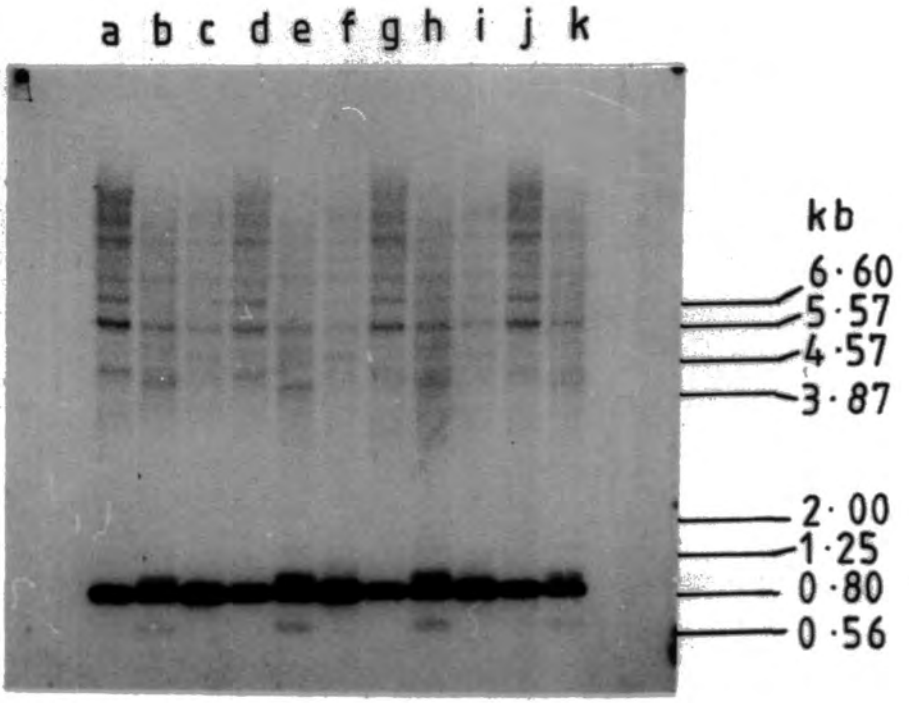

a b c d e f g h i j k l
kb
11·35
3·75
2·95
1·53
0·60

Fig. 33


a b c d e f g h i j k
kb
6·60
5·57
4·57
3·87
2·00
1·25
0·80
0·56

3:3:5 Further work to relate the Hpa II site, whose change in methylation is detected by the cDNA probe, to a specific legumin gene

Figures 34 and 35 show the results obtained when double digested DNA (Taq I plus either Msp I or Hpa II) was probed with two different legumin genes (ie coding regions plus flanking regions), legumin gene A (pDUB 24) and legumin gene B (pDUB 25).

In figure 34 the Taq I fragments containing the 3 sites M1, M2 and M3 were easily identified.

In Leg B, the M1 site is absent and the Taq I genomic fragment equivalent to the 5' end of this gene (1.40kb) remained uncleaved by Msp I and Hpa II. The Taq I fragments containing the other two Msp I sites M2 and M3 were identified. The major difference between the hybridization patterns obtained with these two probes, was in the number of fragments detected in the 1.25 to 1.57kb range. The Leg A probe (figure 34) detected four fragments (1.57, 1.49, 1.40 and 1.25kb) (see also figure 26), whilst the Leg B probe (figure 35) detected only the 1.40kb fragment.

Both legumin gene probes have detected the larger Taq I fragment (2.75kb) which was cleaved by Msp I to yield a 2.54kb fragment. There was an indication of a differential extent of cleavage of this fragment by Hpa II (ie 12 d.a.f. <15 d.a.f. = leaf <21 d.a.f.).

Figure 36 shows the results obtained when genomic DNA was double digested with Eco RI plus either Msp I or Hpa II and the Southern blot hybridized with a legumin cDNA (pDUB 6). Major Eco RI bands were detected at 12.45. 10.84, 7.16 and 4.20kb (tracks a, d, g and j). All

Figure 34:- Rehybridization of the Southern blot shown in figure 25, with ^{32}P-legumin gene A (pDUB 24 insert).

The specific activity of the probe was 2.5 x 10^8 cpm/$_\mu$g and the filter was hybridized and washed at 65°C. The identities of the tracks are indicated in figure 25. The positions of the Leg A, Taq I fragments (▶) and the Msp I/Hpa II cleavage products (▷) are indicated.

Figure 35:- Rehybridization of the Southern blot shown in figure 26, with ^{32}P-legumin gene B (pDUB 25 insert).

The specific activity of the probe was 2 x 10^8 cpm/μg and the filter was hybridized and washed at 65°C. The identities of the tracks are as indicated in figure 25. The positions of the Leg B, Taq I fragments (▶) and the Msp I/Hpa II cleavage products (▷) are indicated.

Fig. 34

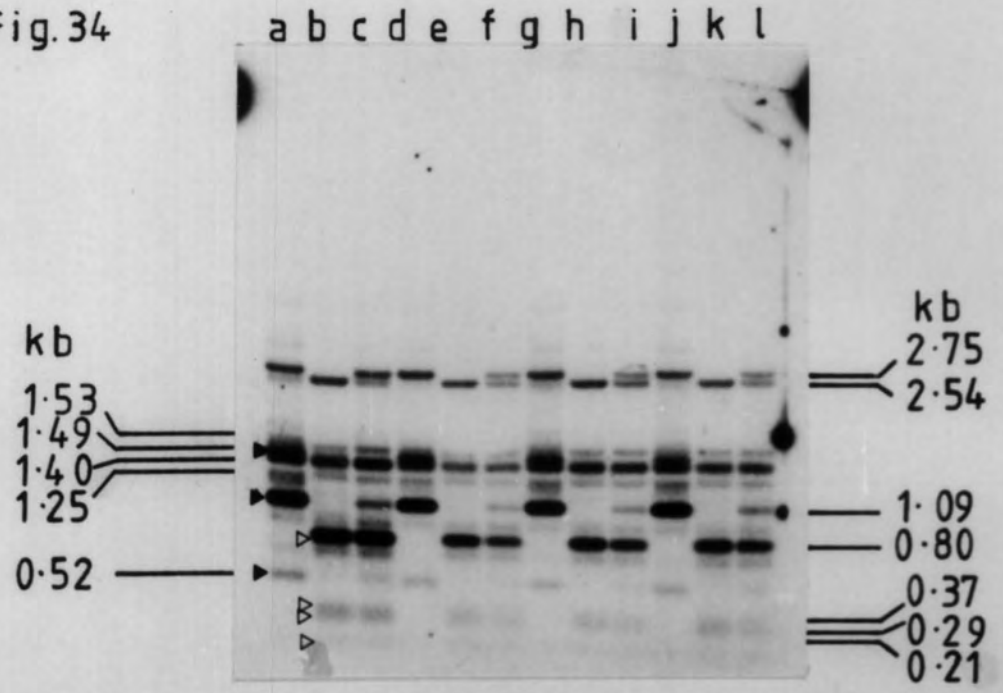
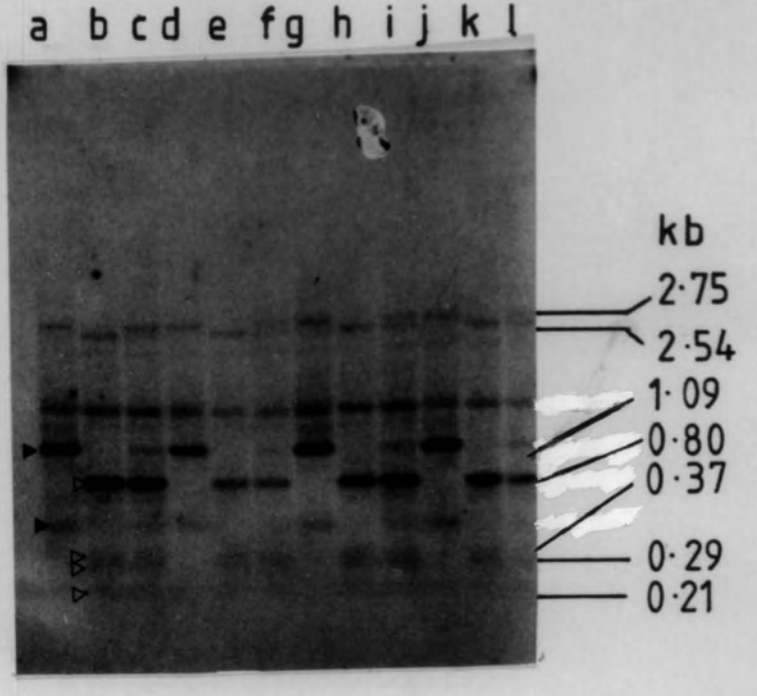

a b c d e f g h i j k l
kb
1·53
1·49
1·40
1·25
0·52
kb
2·75
2·54
1·09
0·80
0·37
0·29
0·21

Fig. 35


a b c d e f g h i j k l
kb
2·75
2·54
1·09
0·80
0·37
0·29
0·21

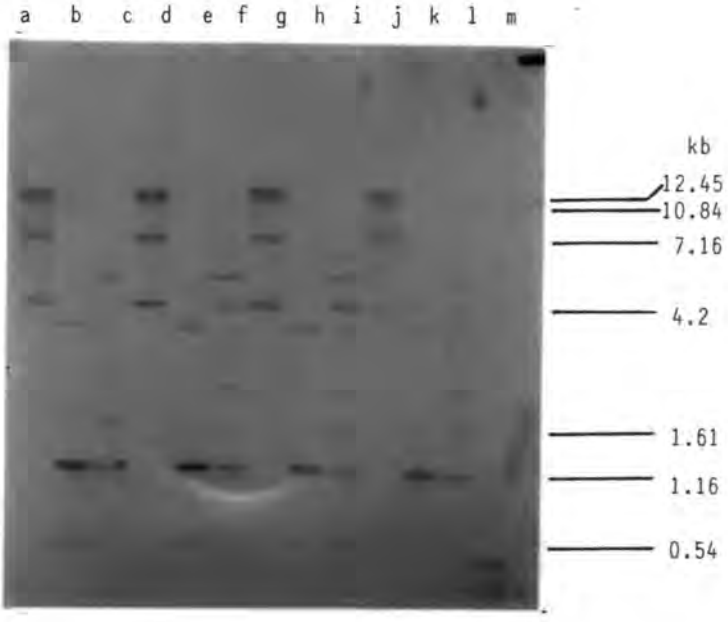


Figure 36:- Hybridization of ^{32}P-legumin cDNA (pDUB 6 insert), to a Southern blot of double restricted (Eco RI plus Msp I/Hpa II) genomic DNA.

a) Leaf DNA + Eco RI
b) " " + Eco RI + Msp I
c) " " + Eco RI + Hpa II
d) 9 d.a.f. cotyledon DNA + Eco RI
e) " " " " + Eco RI + Msp I
f) " " " " + Eco RI + Hpa II
g) 10 d.a.f. cotyledon DNA + Eco RI
h) " " " " + Eco RI + Msp I
i) " " " " + Eco RI + Hpa II
j) 15 d,a,f, cotyledon DNA + Eco RI
k) " " " " + Eco RI + Msp I
l) " " " " + Eco RI + Hpa II
m) pDUB 1 insert (50pg) + Msp I

The tracks contained 5µg DNA restricted with a 20-fold excess of Msp I and Hpa II and a 10-fold excess of Eco RI. The specific activity of the probe was 5 x 10^8 cpm/µg and the filter was hybridized at 65°C and washed at 50°C.

of these bands were cleaved by Msp I (tracks b, e, h and k) and three new bands 0.54, 1.16 and 3.65kb were seen.

Following digestion with Hpa II (tracks c, f, i and l), the 12.45, 10.84 and 7.16kb Eco RI fragments were all cleaved but the 4.20kb fragment was only partially cleaved. Major new bands were seen at positions equivalent to 6.24, 5.07, 3.65, 2.32, 1.70, 1.16, and 0.54kb. However, the 6.24kb band was absent in the 15 d.a.f. cotyledon DNA track (l).

The sizes expected for the Eco RI fragments derived from the genomic clones, λLeg 1, 2 and 3 are 12.5kb (λLeg 1), 7.0kb (λLeg 2) and 9.4kb (λLeg 3) (figure 29b).

3:4 Investigation of the Vicilin Gene Family

3:4:1 Specificity of the vicilin cDNAs used

Two different vicilin cDNA probes were used to investigate methylation of the vicilin gene family. The cDNA, pDUB 7 is an 1100bp cDNA which detects the 47,000-M_r vicilin polypeptide, whilst pDUB 2 is a 900bp cDNA, which detects the 50,000-M_r vicilin polypeptide.

Figure 37b shows that although these two different vicilin cDNA probes have different specificities of hybridization they do detect several common bands of hybridization (eg 7.40 and 5.37kb Eco RI fragments, 6.90kb Bam HI fragment, 7.83 and 5.81kb Kpn I fragments and the 4.96 and 1.93kb Hind III fragments).

Figure 37a:- Photograph of an ethidium bromide stained gel, showing leaf genomic DNA restrictions.

a) Leaf DNA + Eco RI
b) " " + Bam HI
c) " " + Kpn I
d) " " + Hind III
e) pDUB2 (40pg) + pDUB 7 (40pg)
f) Leaf DNA + Eco RI
g) " " + Bam HI
h) " " + Kpn I
i) " " + Hind III
j) λNM258 + Hind III; pBR322 + Hinf I
k) λNM258 + Ava I

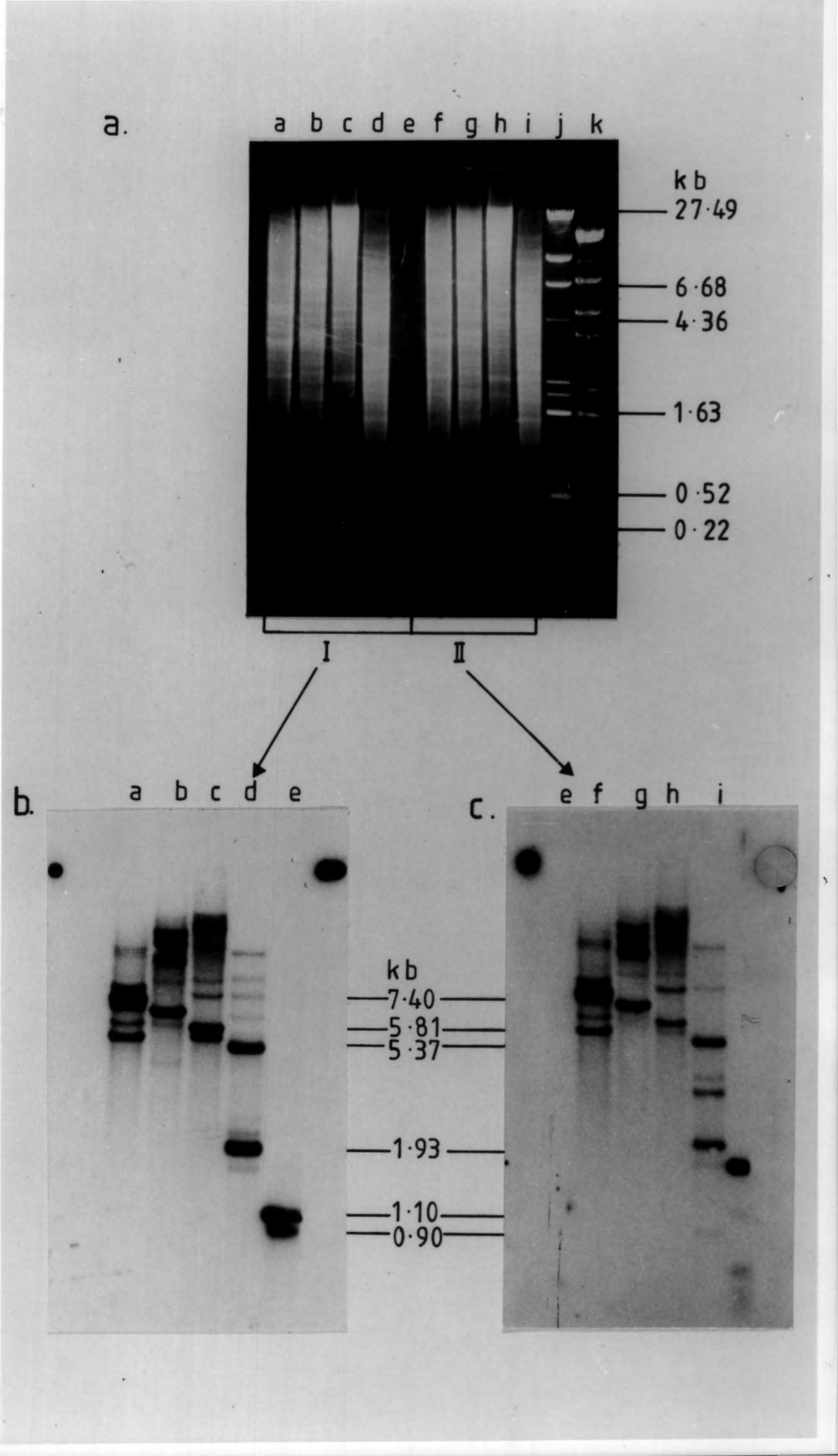
Each genomic track contained 10μg DNA digested with a 9-fold excess of the enzyme.

Figure 37b:- Hybridization of ^{32}P-labelled 50,000-vicilin cDNA (pDUB 2 insert), to a Southern blot of part I of the gel illustrated in figure 37a.

The specific activity of the probe was 6 x 10^7 cpm/μg and the filter was hybridized at 60°C and washed at 50°C.

Figure 37c:- Hybridization of ^{32}P-labelled 47,000-vicilin cDNA (pDUB 7 insert), to a Southern blot of part II of the gel shown in figure 37a.

The specific activity of the probe was 5.3 x 10^7 cpm/μg and the filter was hybridized at 60°C and washed at 50°C.


a.
a b c d e f g h i j k
kb
27·49
6·68
4·36
1·63
0·52
0·22
I
II
b.
a b c d e
kb
7·40
5·81
5·37
1·93
1·10
0·90
c.
e f g h i

3:4:2 Methylation of the vicilin gene family

Figure 38 shows the hybridization pattern obtained when the 47,000-M_r polypeptide vicilin cDNA probe, pDUB 7, was hybridized to Hae III, Msp I and Hpa II digested gemonic DNA. Three major Hae III bands were detected at 8.32, 4.96 and 4.17kb (tracks a, d, g and j). Three minor Hae III fragments were also detected (6.95, 1.35 and 1.11kb).

Three major Msp I bands were observed at 11.40, 8.51 and 4.67kb (tracks b, e and h). In addition, several minor bands were seen at 14,13, 9.59, 7.50, 6.46 and 2.37kb. However, in the leaf plus Msp I track (k) only the 11.40, 9.59 and 4.67kb bands were visualised. Following Hpa II digestion, a single band (11.61kb) was seen in the 15 and 21 d.a.f. cotyledon tracks (f and i) and this band was absent from the leaf and 10 d.a.f. cotyledon DNA tracks (c and l). In all of the Hpa II tracks, a large proportion of the probe had hybridized to fragments of greater than 20kb in length.

The 50,000-M_r polypeptide, vicilin cDNA probe (pDUB 2) detected basically a similar pattern of hybridization to the larger DNA fragments (i.e. >6.4kb) (figure 39: tracks a, b, c, d, e and f). However, this probe did detect slightly different smaller Msp I fragments (5.59, 3.60. 2.90 and 1.82kb). The 11.61kb Hpa II fragment was again detected in cotyledon DNA, 15 and 22 d.a.f. (tracks b and d) but was absent from the leaf plus Hpa II track (f). The 1.82kb Msp I and Hpa II fragment was not detected by the 47,000-vicilin cDNA probe pDUB 7 (figure 38).

Fig. 38

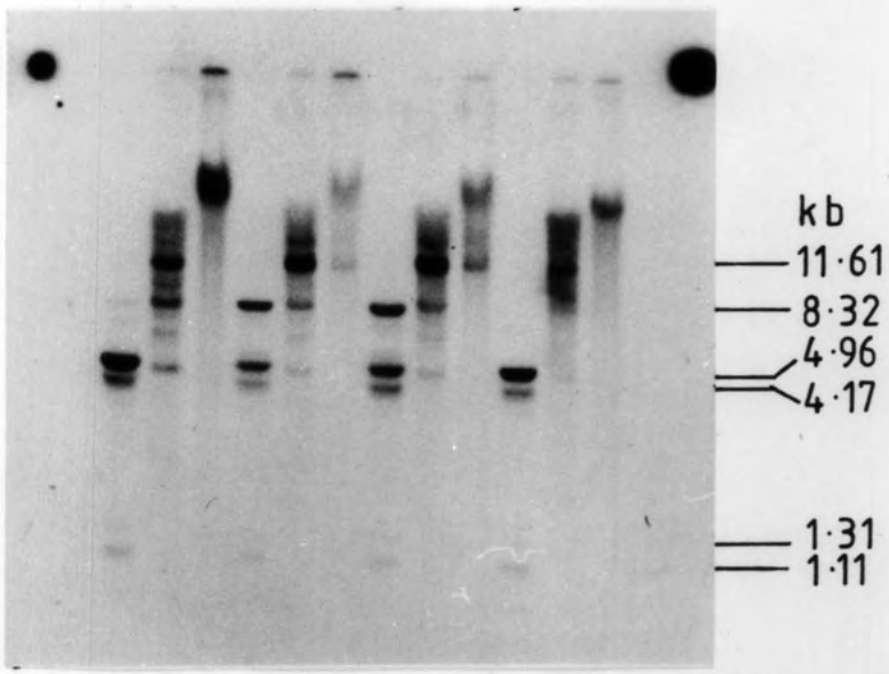

a b c d e f g h i j k l
kb
11·61
8·32
4·96
4·17
1·31
1·11

Fig. 39

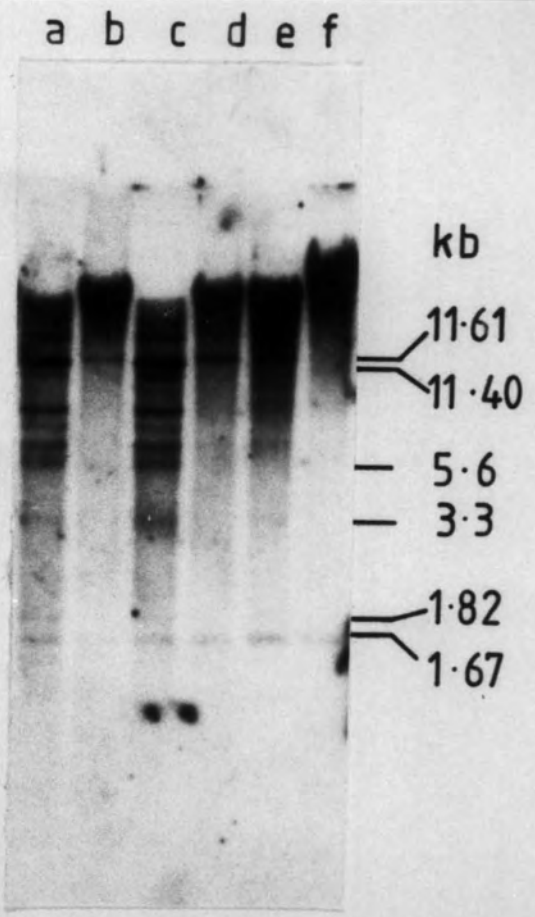

a b c d e f
kb
11·61
11·40
5·6
3·3
1·82
1·67

Figure 38:- Hybridization of ^{32}P-labelled 47,000-vicilin cDNA (pDUB 7 insert), to a Southern blot prepared from a gel similar to that illustrated in figure 22a.

a) 10 d.a.f. cotyledon DNA + Hae III
b) " " " + Msp I
c) " " " + Hpa II
d) 15 d.a.f. cotyledon DNA + Hae III
e) " " " + Msp I
f) " " " + Hpa II
g) 21 d.a.f. cotyledon DNA + Hae III
h) " " " + Msp I
i) " " " + Hpa III
j) Leaf DNA + Hae III
k) " " + Msp I
l) " " + Hpa II

The filter was hybridized and washed at 42°C using the formamide system. The specific activity of the probe was 3 x 10^8 cpm/$_\mu$g.

Figure 39:- Hybridization of ^{32}P-labelled 50,000-vicilin cDNA (pDUB 2 insert) to a Southern blot of Msp I and Hpa II digested genomic DNA.

a) 15 d.a.f. cotyledon DNA + Msp I
b) " " " + Hpa II
c) 22 d.a.f. cotyledon DNA + Msp I
d) " " " + Hpa II
e) Leaf DNA + Msp I
f) " " + Hpa II

Each track contained 10$_\mu$g DNA digested with a 10-fold excess of enzyme. The specific activity of the probe was 3 x 10^8 cpm/$_\mu$g and the filter was hybridized at 65°C and washed at 50°C.

3:5 Work Involving the Mutant Pea Line Witham Wonder 5478

3:5:1 Growth of mutant pea line

Feltham First and Witham Wonder lines 200 (parental) and 5478 (mutant) pea plants were easily distinguished by certain characteristics. Feltham First plants had mottled leaves and the leaflets were not attached by petiolules (figure 40a). Witham Wonder line 200, did not have the mottling and neither did it have petiolules (figure 40b). However, line 5478 had no mottling but the leaflets were attached by petiolules (figure 40c). Line 5478 also showed less apical dominance than the parental line 200, as the side shoots had a tendency to grow out and produce mature stems. It also took considerably longer to reach maturity (ie flowering) and the newly formed pods emerged contorted and twisted but straightened out after a few days. There was a greater tendency for flowers to drop off or to form infertile pods and those pods which did develop mormally, had fewer peas per pod.

3:5:2 Comparison of 50,000-M_r vicilin coding sequences in Feltham First and Witham Wonder pea lines 200 and 5478

Leaf genomic DNA, isolated from Feltham First and Witham Wonder lines 200 and 5478, was digested with a series of restriction enzymes and after blotting was hybridized to the cDNA for the 50,000-M_r vicilin polypeptide (pDUB 2). Hybridization bands ranged in size from 17kb (Kpn I) to less than 0.5kb (Alu I). In all cases, the same bands were observed in all three genomic digests and the relative

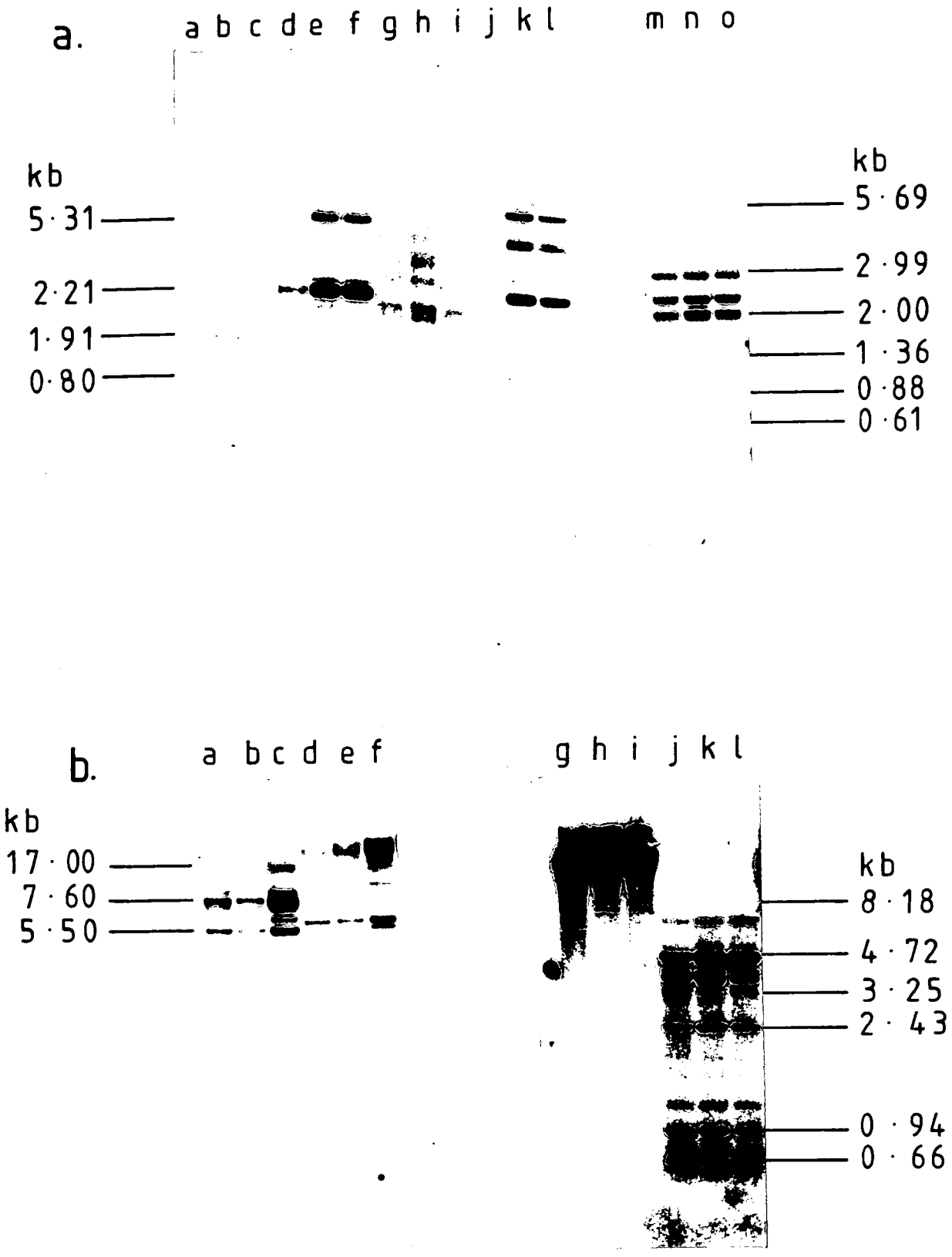
Figure 40:- Photographs of the leaflets and petiolules of a) Feltham First, b) Witham Wonder line 200, c) Witham Wonder line 5478.

Figure 41a and b:- Hybridization of ^{32}P-labelled 50,000-vicilin cDNA (pDUB 2 insert) to Southern blotted, restricted leaf genomic DNA from Feltham First and Witham Wonder lines 200 and 5478.

Figure 41a:- Tracks a, d, g, j and m were Feltham First restricted DNA; tracks b, e, h, k and n were line 200 restricted DNA; trakcs c, b, i, l and o were line 5478 restricted DNA. The restrictions were as follows: a, b and c - Alu I; d, e and f - Hind III; g, h and i - Mbo I; j, k and l - Bcl I; m, n and o - Taq I.

Figure 41b:- Tracks a, d, g and j - Feltham First restricted DNA; b, e, h and k - line 200 restricted DNA; c, f, i and l - line 5478 restricted DNA. Restrictions were as follows: a, b and c - Eco RI; d, e and f - Kpn I; g, h and i -Apa I; j, k and l - Hae III.

The specific activity of the probe was 1.5 x 10^8 cpm/μg and the filters were hybridized at 60°C and washed at room temperature.


a.
a b c d e f g h i j k l
m n o
kb
5·31
2·21
1·91
0·80
kb
5·69
2·99
2·00
1·36
0·88
0·61
b.
a b c d e f
g h i j k l
kb
17·00
7·60
5·50
kb
8·18
4·72
3·25
2·43
0·94
0·66

intensities of the bands, within each track, were similar between the tracks (figure 41)

3:5:3 Analysis of total RNA extracted from Feltham First and Witham Wonder lines 200 and 5478

Feltham First total RNA was isolated using three different methods. The RNA was glyoxalated and after electrophoresis was stained with acridine orange (figure 42). The 25S and 18S ribosomal RNAs were clearly visible in all tracks. In addition the leaf RNA, isolated by Chirgwin's method, was partially degraded (tracks i and j). Where the rapid method of Langridge *et al* was being used (tracks d, e, f and g), tRNA was also isolated and was clearly visible in these tracks.

The method of Langridge was preferred and was therefore used to isolate total RNA from developing cotyledons of the two Witham Wonder pea lines. This RNA was Northern blotted and probed with three different vicilin cDNAs (pDUB 2, 7 and 9) (figure 43). When the cDNA probe for the 47,000-vicilin polypeptide was used (figure 43a), mRNA for this polypeptide was detected in both parental and mutant lines, at all stages of cotyledon development examined. In 9 d.a.f. cotyledons, only a small amount of the message was detected (tracks a and b). There was then a rapid increase in accumulation of this mRNA to 11 d.a.f. (tracks c and d) (over a 10-fold increase), followed by a slower increase to 13 d.a.f. (tracks e and f) (2-fold increase) and a gradual decline between 13 d.a.f. and 17 d.a.f. (tracks g, h, i and j). The message for the 47,000-vicilin was barely detectable in the parental line 19 d.a.f. (track k). The relative amounts of the message

a b c d e f g h i j k

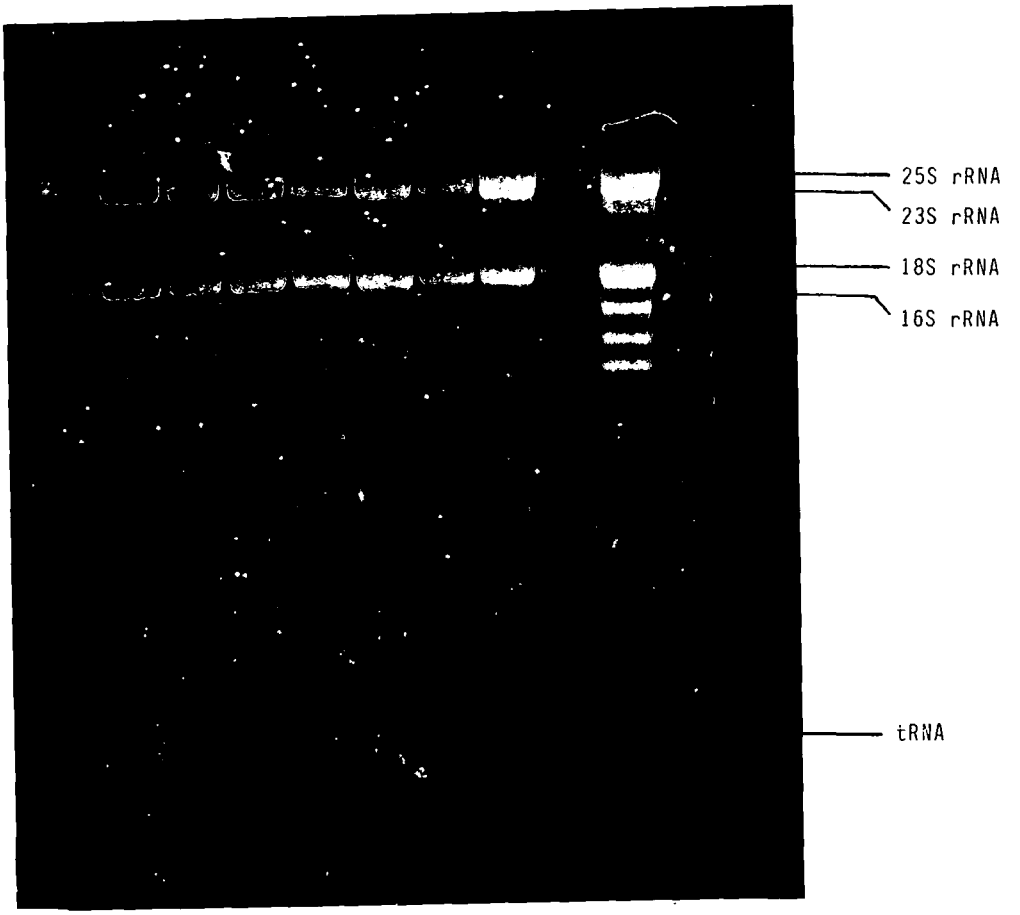


Figure 42:- Acridine orange stained glyoxal gel showing total RNA preparations from Feltham First, using different methods of extraction.

a) 14 d.a.f. cotyledon total RNA
b) 16 d.a.f. " " " } Hall and Buchbinder's
c) 18 d.a.f. " " " hot SDS method
d) 12 d.a.f. " " "
e) 14 d.a.f. " " " Langridge *et al* 's phenol/
f) 14 d.a.f. " " " chloroform/SDS method
g) 16 d.a.f. " " "
h) 11 d.a.f. " " " Chirgwin *et al* 's
i) Leaf total RNA guanidinium thiocyanate method
j) " " "
k) *E. coli* RNA

10μg RNA per track.

Figure 43a:- Hybridization of ^{32}P-labelled 47,000-vicilin cDNA (pDUB 7 insert); to a Northern blot formamide prepared from total RNA isolated at different stages of cotyledon development in Witham Wonder lines 200 and 5478.

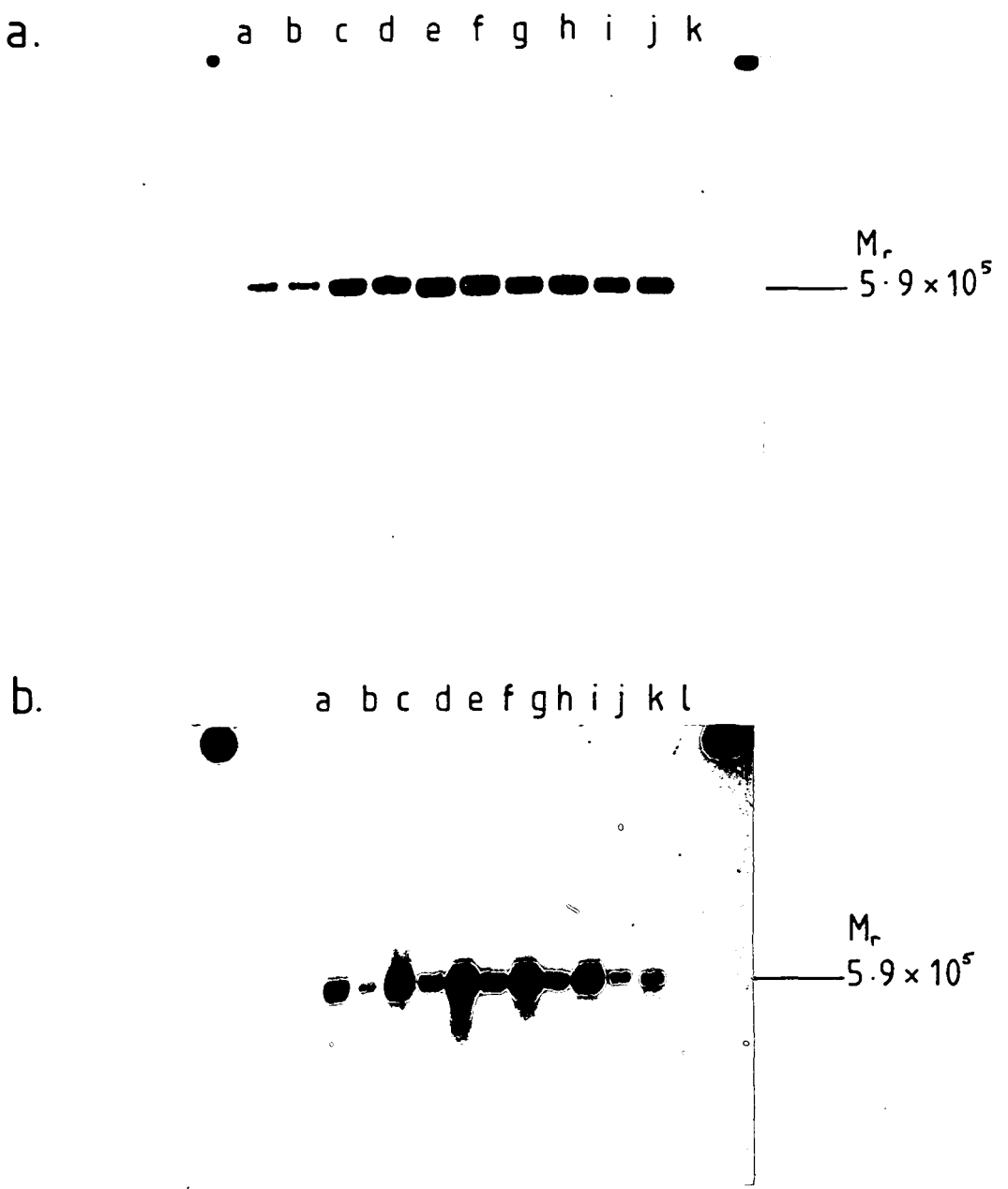
Tracks a, c, e, g, i and k were RNAs isolated from line 200 and tracks b, d, f, h and j were the RNAs isolated from line 5478. The stages of cotyledon development were: 9 d.a.f. (a and b), 11 d.a.f. (c and d), 13 d.a.f. (e and f), 15 d.a.f. (g and h), 17 d.a.f. (i and j) and 19 d.a.f. (k).

10 µg of total RNA were loaded per track and the specific activity of the probe was 3 x 10^8 cpm/µg. The filter was hybridized in the formamide system at 42°C and was also washed at 42°C.

Figure 43b:- Hybridization of ^{32}P-labelled 50,000-vicilin cDNA (pDUB 2 insert), to a Northern blot prepared from total RNA isolated at different stages of Witham Wonder (lines 200 and 5478) cotyledon development.

The tracks are as indicated for figure 43a, with the additional track 'l' being RNA isolated from line 5478 19 d.a.f. The specific activity of the probe was 1.3 x 10^8 cpm/µg and the filter was hybridized as described for figure 43a.

Figure 43c:- Rehybridization of the Northern blot, the autoradiograph of which is shown in figure 43b, with a different ^{32}P-labelled 50,000-vicilin cDNA (pDUB 9 insert).


a.
a b c d e f g h i j k
M_r
$5 \cdot 9 \times 10^5$
b.
a b c d e f g h i j k l
M_r
$5 \cdot 9 \times 10^5$
c.
a b c d e f g h i j k
M_r
$5 \cdot 9 \times 10^5$

detected in parental and mutant lines, at the different stages of cotyledon development, were similar.

When pDUB 2 was used as a probe for the 50,000-M_r vicilin polypeptide message (figure 43b), significant amounts of the message were detected in the parental line 9 d.a.f. (track a). This was followed by a 4 to 5-fold increase in the amount of detectable message at 11 d.a.f. (track c) and a smaller increase (2-fold) to 13 d.a.f. (track e). After this time the level of message gradually declined to 17 d.a.f. (tracks g and i) followed by a rapid decrease thereafter (track k). In the mutant line, the amounts of mRNA detected at all stages of cotyledon development were far less than were detected at the equivalent stage of cotyledon development. The message was barely detectable 9 d.a.f. (track b), increased steadily to 13 d.a.f. (tracks d and f), remained at a fairly constant level 15 d.a.f. (track h) and then declined rapidly and was barely detectable 19 d.a.f. (tracks j and l).

When a cDNA probe for a different 50,000 -M_r vicilin polypeptide message was used (figure 43c), a different pattern of message accumulation was observed. Again, in the parental line, the message was detected in appreciable amounts 9 d.a.f. (track a). However, the amount of message detected increased by only a factor of 3 to 4-fold between 9 d.a.f. and 13 d.a.f. (track e) and this was followed by a gradual decline to 15 d.a.f. (track g) and a rapid decline between 15 and 17 d.a.f. (tracks i and k) to become barely detectable 19 d.a.f. (track k). In the mutant line, relatively higher amounts of this message were detected at all stages of development than had been

detected by pDUB 2 (figure 43b), as judged by the relative intensities of the parental and mutant bands at each stage of cotyledon development. The pattern of accumulation of this message in the mutant line was similar to that described for the parental line except that the amount of message detected after day 15 declined more rapidly (track j) and was below the level of detection 19 d.a.f. (track 1).

3:5:4 Analysis of parental and mutant genomic DNA using MspI and HpaII

When Msp I and Hpa II digested cotyledon and leaf genomic DNA was hybridized to a 50,000-M_r vicilin polypeptide cDNA (pDUB 2), the pattern of hybridization (figure 44) was similar to that described previously (figure 39). In particular, the 11.61kb Hpa II band, absent in both 200 and 5478 leaf DNA (tracks h and j), was barely detectable in 15 d.a.f. 5478 cotyledon DNA (track d) but was clearly visible in the 21 d.a.f. 200 and 5478 cotyledon DNA (tracks d and f). The 1.84kb band was also present in all Msp I and Hpa II tracks.

3:6 Analysis of legumin gene A and pseudo legumin gene D sequence data

Figure 45 shows the distribution of CG dinucleotides along the two genes. In general, the protein coding regions (exons) were richer in CG dinucleotides. This was particularly noticeable in the first exon of legumin gene A (figure 45c). Figure 46 shows the relationship between the observed number of CG and GC dinucleotides and the expected number calculated from the base composition of gene regions (figure 46a: II and III) and 100 nucleotide segments (figure 46b: II

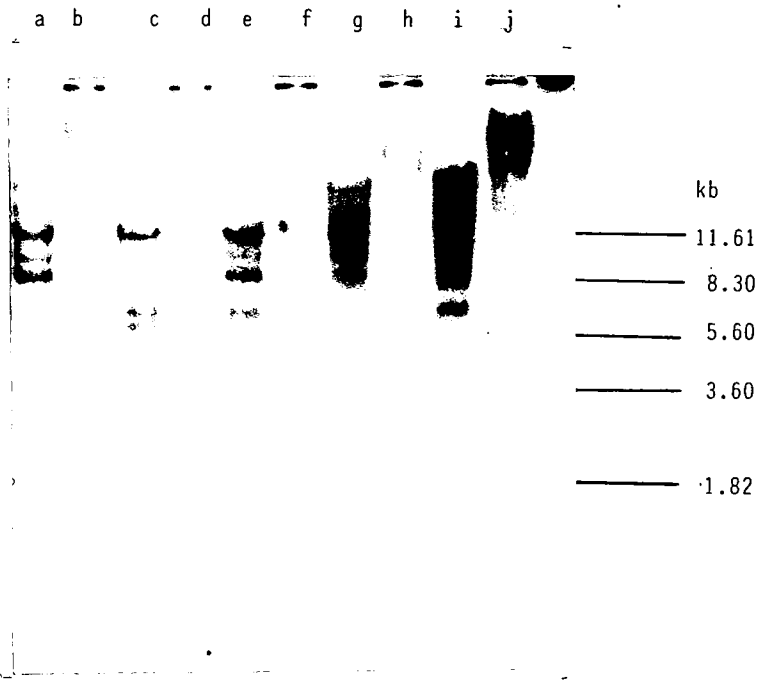


Figure 44:- Hybridization of ^{32}P-labelled 50,000-vicilin cDNA (pDUB 2 insert), to a Southern blot prepared from Msp I and Hpa II digested genomic DNA, from lines 200 and 5478.

a) Line 5478 15 d.a.f. cotyledon DNA + Msp I
b) " " " " " + Hpa II
c) Line 200 22 d.a.f. cotyledon DNA + Msp I
d) " " " " " + Hpa II
e) Line 5478 22 d.a.f. cotyledon DNA + Msp I
f) " " " " " + Hpa II
g) Line 200 leaf DNA + Msp I
h) " " " " " + Hpa II
i) Line 5478 leaf DNA + Msp I
j) " " " + Msp II

Each genomic track contained 7.5μg DNA digested with a 12-fold excess of Msp I and Hpa II. The specific activity of the probe was 1 x 10^8 cpm/μg and the filter was hybridized at 65°C in the heparin system and was also worked at 65°C.

Figure 45:- Distribution of CG dinucleotides in legumin genes A (pDUB 24) and D.

Figure 45a:- Positions of CG dinucleotides in Leg D.

Figure 45b:- Location of protein coding sequences (solid boxes = exons) in Leg D.

Figure 45c:- Positions of CG dinucleotids in Leg A.

Figure 45d:- Location of protein coding sequences (solid boxes = exons) in Leg A.

Figure 46:- Schematic representation of the ratio of observed to the expected number of CG and GC dinucleotides in legumin gene A (pDUB 24).

Figure 46a:-

I Map showing coding and non-coding regions of Leg A.

II Ratio of the observed number of GC dinucleotides to the number calculated for the various gene regions of Leg A.

III Ratio of the observed number of CG dinucleotides to the number calculated for the various gene regions of Leg A.

Figure 46b:-

I Percentage composition (C+G) of 100 nucleotide segments of Leg A.

II Ratio of the observed number of GC dinucleotides to the number calculated for 100 nucleotide segments of Leg A.

III Ratio of the observed number of CG dinucleotides to the number calculated for 100 nucleotide segments of Leg A.

and III) (see also table 7 and 8). When the GC ratio was examined (figure 46b:II), it was noticed that there was considerable variation in this ratio (ranges from 0 to 2.0), and 23 out of the 36 segments had a ratio of over 1.0.

When the observed to expected ratio was examined in terms of gene regions (figure 46a: II), exon 2, and introns 2 and 3 both had ratios of less than 1.0 whilst other gene regions had a ratio of 1.0 or more. In particular, exon 4 had a ratio of 1.50 and it was noticed that all the individual 100 nucleotide segments in this region had a high ratio (figure. 46b:II). In contrast, the 5' flanking region had an area which was relatively deficient in GCs (-500bp to -1200bp from the start of protein coding regions), followed by a region which was relatively rich in GCs (-100 to -500bp). The combination of these two features, gave the 5' flanking region an overall ratio of 1.0.

When the CG ratio was examined (figure 46b:III), it was immediately apparent that the values obtained for 100 nucleotide segments were generally much lower than had been observed for the GC ratio. They ranged from 0 to 1.27 but only 1 out of the 36 segments had a value of over 1.0. These low individual values were reflected in the overall values obtained for the gene regions (figure 46a:III). No single region had a ratio of 1.0 or more. Exon 1 had the highest ratio (0.68) but in general the ratios were between 0.4 to 0.5.

The overall percentage composition (C + G per 100 nucleotides) is shown in figure 46b:I and when the gene map (figure 46a:I) was superimposed on this, it was seen that exon 1 and exon 3 were relatively C + G rich, whilst the 5' and 3' flanking regions were

Table 7: Legumin gene A ratio of observed to expected number of dinucleotides

Length (bp)	Total Sequence 3692	5'F 1231	E1 286	I1 89	E2 251	I2 89	E3 626	I3 99	E4 391	3'F 630
Dinucleotide										
CA	1.27	1.21	1.12	1.36	1.31	1.35	1.34	2.05	1.51	1.26
AG	1.17	1.13	1.09	1.10	1.33	1.46	1.42	0.92	1.16	0.88
GC	1.17	0.98	1.36	1.16	0.64	0.93	1.21	0.51	1.50	0.99
TG	1.13	1.16	1.33	0.82	1.09	1.32	1.14	1.27	1.52	1.30
GA	1.07	1.08	1.13	0.37	1.36	0.77	1.18	1.18	0.82	0.95
TC	1.07	1.16	0.99	0.44	0.97	1.19	1.38	1.41	0.94	1.16
CT	1.02	0.93	1.24	1.02	0.97	0.99	1.38	0.63	1.19	1.04
AT	1.00	1.02	0.79	1.01	0.99	1.16	0.81	1.14	0.95	1.06
AC	0.82	0.73	0.82	1.94	1.11	0.84	0.73	0.91	0.89	0.85
GT	0.80	0.90	0.57	1.37	0.77	1.32	0.59	1.09	1.03	0.92
TA	0.78	0.83	0.47	1.47	0.78	1.06	0.39	0.65	0.73	0.84
CG	0.44	0.36	0.68	0.58	0.40	0	0.41	0.51	0.45	0.33

The abbreviations used are: 5'F = 5' flanking region; E1 = exon 1; I1 = intron1; E2 = exon 2; I2 = intron 2; E3 = exon 3; I3 = intron 3; E4 = exon 4; 3'F = 3' flanking region.

Table 8: Frequency ratios for selected di and trinucleotides in legumin gene D

	Total Sequence	5'F	E1	I1	E2	I2	E3	I3	E4	3'F
Length (bp)	2509	775	289	84	250	113	442	92	289	195
Dinucleotide										
CA	1.88	1.01	0.80	1.53	1.23	0.84	1.39	1.38	1.35	0.72
CT	1.13	1.07	1.28	0.83	1.06	1.18	1.28	1.01	1.40	0.96
GC	1.02	0.95	1.02	0	0.77	1.18	0.99	1.21	1.40	0.61
GT	0.86	1.02	0.78	0.74	0.95	0.88	0.68	0.48	1.01	1.20
TG	1.16	1.30	0.78	0.74	1.25	1.24	1.30	1.44	1.48	1.40
CG	0.71	0.26	0.54	1.14	0.26	0.59	0.22	0	0.51	0.61
CA + TG	1.18	1.14	0.79	1.24	1.24	1.05	1.36	1.40	1.41	1.04
CAG	1.07	0.46	1.38	3.06	1.78	1.48	1.31	5.30	1.06	0
CTG	1.10	0.63	1.55	0	1.24	3.33	1.54	0	1.37	1.95

The abbreviations used are: 5'F = 5' flanking region; E1 = exon 1; I1 = intron1; E2 = exon 2; I2 = intron 2; E3 = exon 3; I3 = intron 3; E4 = exon 4; 3'F = 3' flanking region.

relatively C + G deficient.

When the same analyses were performed with the sequence data for legumin gene D, similar results were obtained (figures 47a and b). 15 out of 25 100 nucleotide segments had a GC observed to expected ratio, of 1.0 or more, whilst only 4 had a CG ratio in excess of 1.0. The values obtained for the gene regions were again similar to those previously detailed for legumin gene A. For the GC ratio, the values ranged from 0.8 to 1.40 (intron 1 and 3' flanking region excluded), whilst for the CG ratio, the values ranged from 0.2 to 0.6 (introns 1 and 3 excluded). When the percentage C + G per 100 nucleotides was examined, the flanking regions were again relatively deficient, whilst exons 1 and 3 were relatively rich in C + G.

Tables 6 and 7 summarize the observed to expected ratios calculated for various dinucleotides in the two genes. Most of these dinucleotides (except for CG) had an observed to expected ratio of about 0.8 to 1.2 for most of the gene regions. Exceptions to this generalisation were usually associated with the introns, which were usually short in length (<100 nucleotides) eg in Leg A, GA and AC in intron 1, AG and GT in intron 2 and CA and GC in intron 3.

Table 9 shows a summary of the observed to calculated frequency ratios of all Leg A, CG-containing trinucleotides. All CG-containing trinucleotides were found to be deficient. No trinucleotide was found to be suppressed to a greater or lesser extent than the rest.

Figure 48 (a to q) shows the results obtained when the sequence data was analysed with respect to percentage composition of either 100 nucleotide segments or gene regions. The solid line in each of these

0·5 kb
I
% C + G
60
40
20
0
II
GC ratio =
obs / calc
2·0
1·6
1·2
0·8
0·4
0
III
CG ratio =
obs / calc
1·2
0·8
0·4
0

b.

Table 9: Frequency ratios (observed/calculated) of all CG-trinucleotides in Leg A.

Length (bp)	Total 3692	5'F 1231	E1 286	I1 89	E2 251	I2 89	E3 626	I3 99	E4 391	3'F 630	TotalNo. observed
Trinucleotide											
ACG	0.32	0.30	0.50	1.92	0.55	0	0.23	0	0.49	0	14
CCG	0.74	0.72	0.44	0	1.17	0	0.53	0	0.70	1.00	17
GCG	0.40	0	0.30	0	0	0	0.69	0	0.21	0.36	10
TCG	0.39	0.32	1.13	0	0	0	0.19	1.20	0.42	0.44	14
CGA	0.32	0.20	0.76	0	0.55	0	0.30	1.73	0.16	0.17	14
CGC	0.61	0.48	0.88	0	0	0	0.80	0	0	1.00	14
CGG	0.44	0.49	0.44	0	0.34	0	0.20	0	0.42	0.72	11
CGT	0.41	0.32	0.45	1.44	0.31	0	0.57	0	1.04	0	15

The abbreviations used are 5'F = 5' flanking region; E1 = exon 1; I1 = intron 1; E2 = exon 2; I2 = intron 2; E3 = exon 3; I3 = intron 3; E4 = exon 4; 3'F = 3' flanking region.

Figure 48:- Relationship between the frequency of occurrence and the percentage composition of legumin gene A (pDUB 24) and the pseudo legumin gene D.

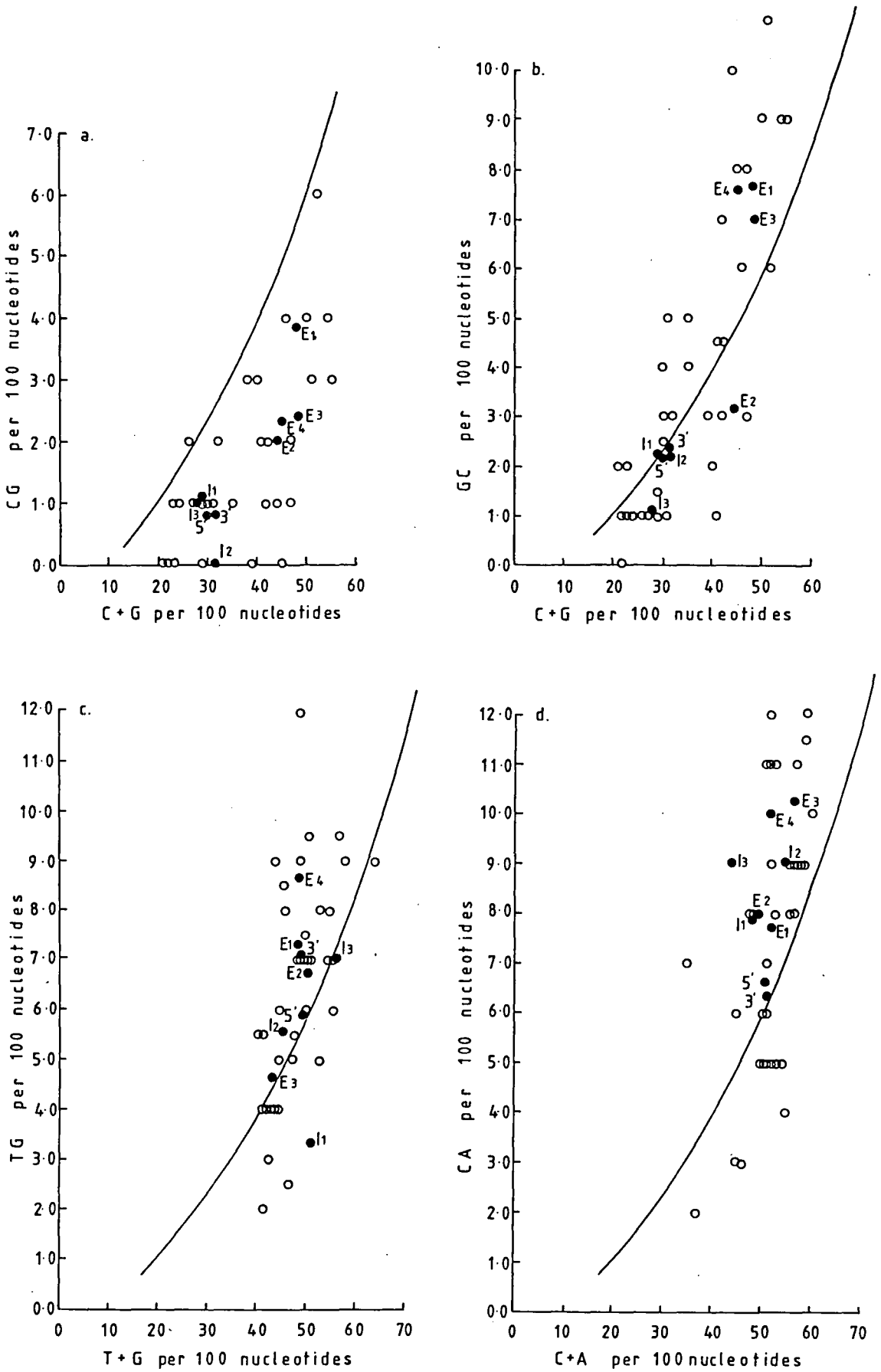
The individual graphs are:-

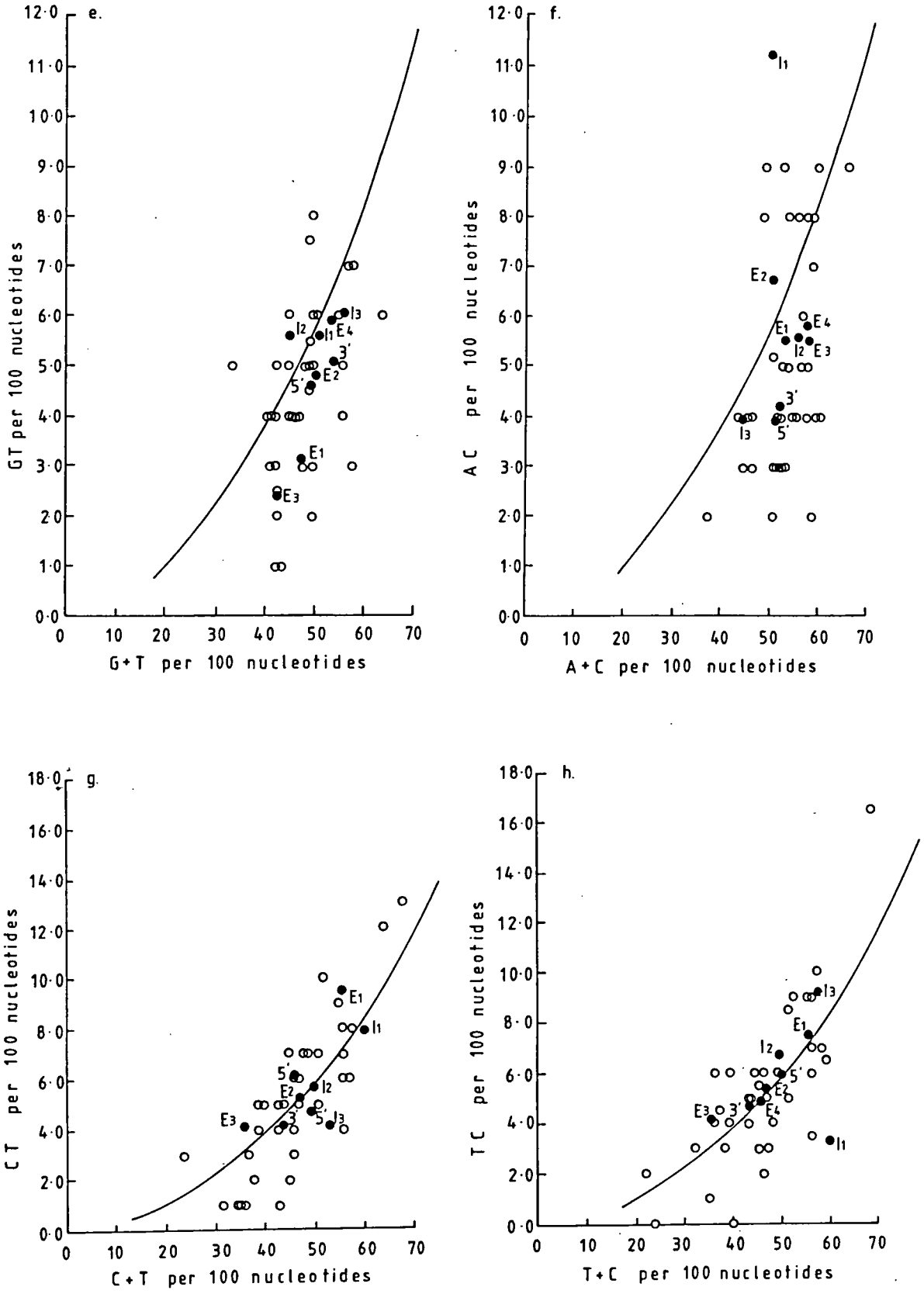
a) Legumin gene A - CG dinucleotides
b) " " " - GC "
c) " " " - TG "
d) " " " - CA "
e) " " " - GT "
f) " " " - AC "
g) " " " - CT "
h) " " " - TC "
i) " " " - GA "
j) " " " - AG "
k) " " " - TA "
l) " " " - AT "
m) Legumin gene D - CG dinucleotides
n) " " " - GC "
p) " " " - TG "
q) " " " - CA "

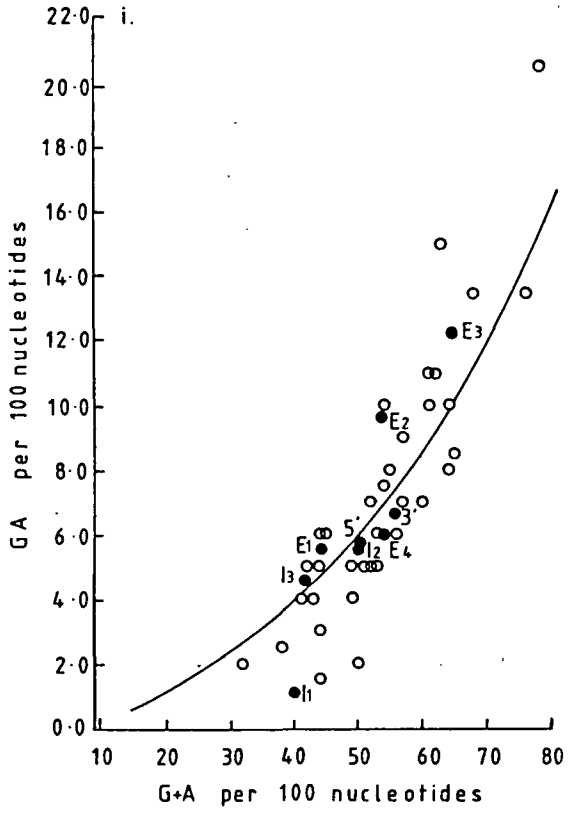
The symbols used are: (o) observed number of dinucleotides in 100 nucleotide segments and (•) observed number of dinucleotide in a gene region (corrected to allow for the varying length of the different regions ie calculated in terms of 100 nucleotide lengths, which are representative of the gene region as a whole).

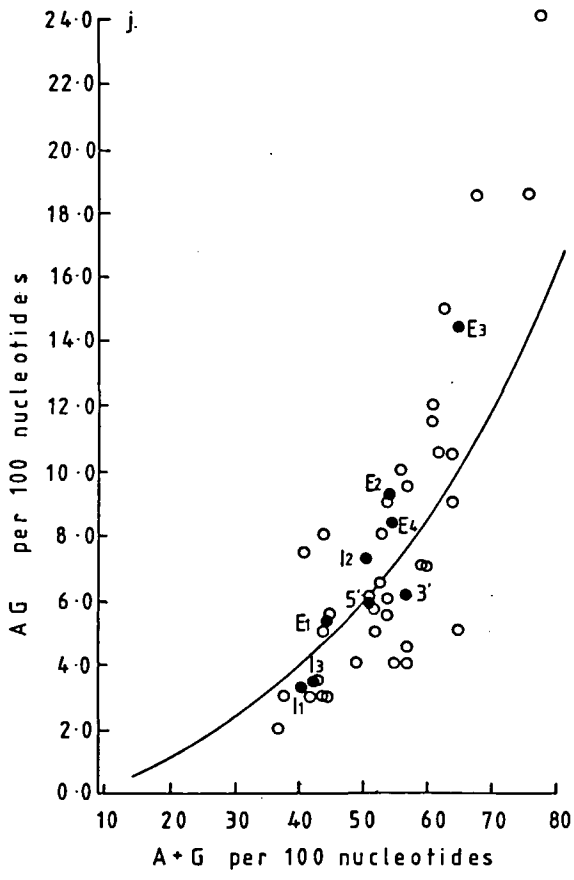
The abbreviations used are: 5' = 5' flanking region; E1 = exon 1; I1 = intron 1; E2 = exon 2; I2 = intron 2; E3 = exon 3; I3 = intron 3; E4 = exon 4; 3' = 3' flanking region.

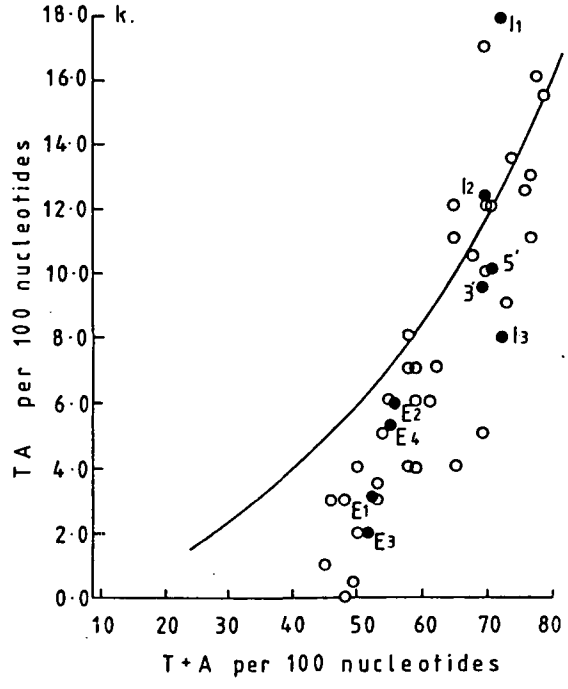
In each graph, the solid line represents the results that would be expected from a random distribution of dinucleotides (see 'Methods').

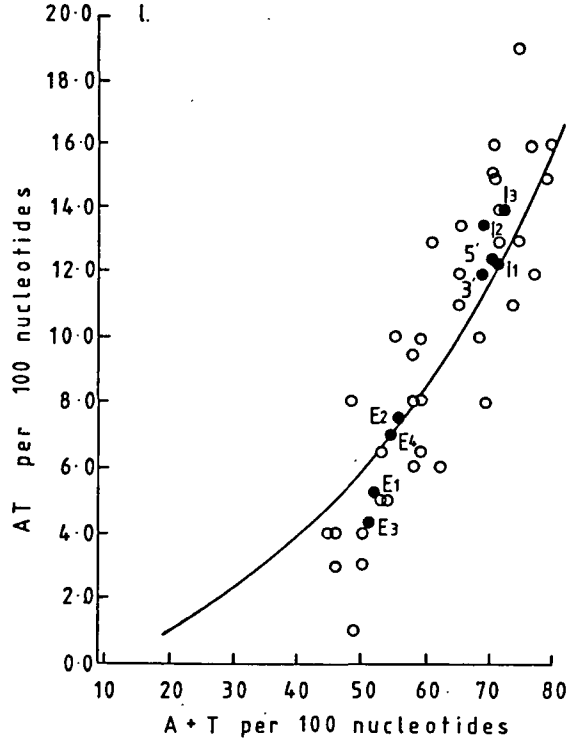

a.
CG per 100 nucleotides
C+G per 100 nucleotides
b.
GC per 100 nucleotides
C+G per 100 nucleotides
c.
TG per 100 nucleotides
T+G per 100 nucleotides
d.
CA per 100 nucleotides
C+A per 100 nucleotides

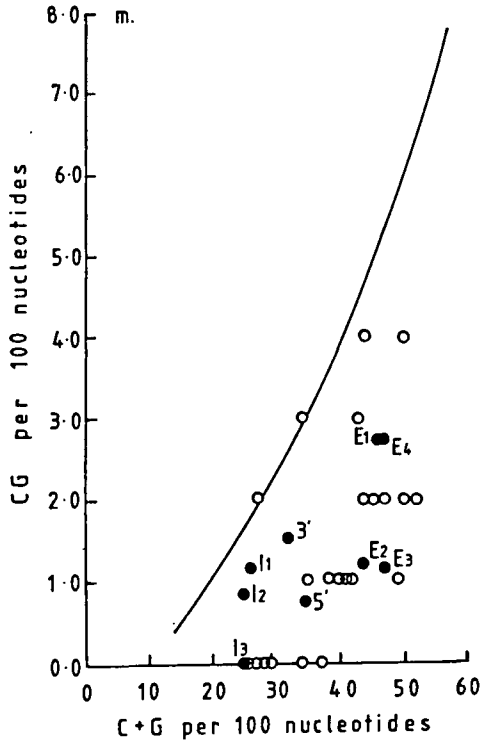

e.
GT per 100 nucleotides
G+T per 100 nucleotides
f.
AC per 100 nucleotides
A+C per 100 nucleotides
g.
CT per 100 nucleotides
C+T per 100 nucleotides
h.
TC per 100 nucleotides
T+C per 100 nucleotides

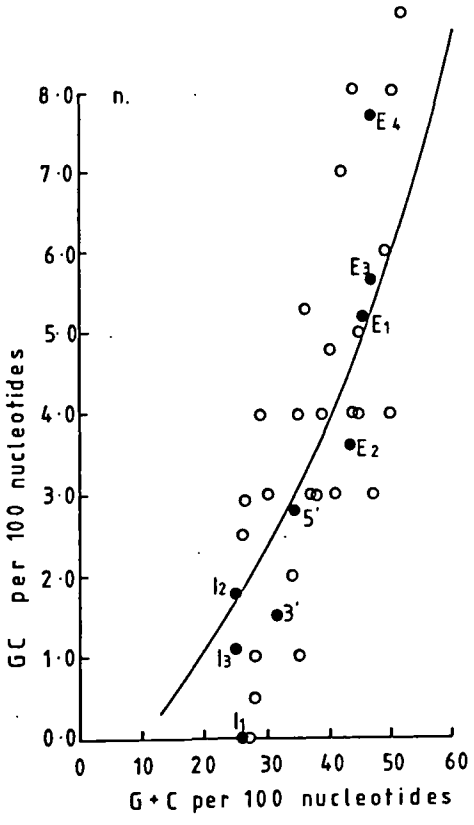

i.
GA per 100 nucleotides
G+A per 100 nucleotides
E1
E2
E3
E4
I1
I2
I3
5'
3'

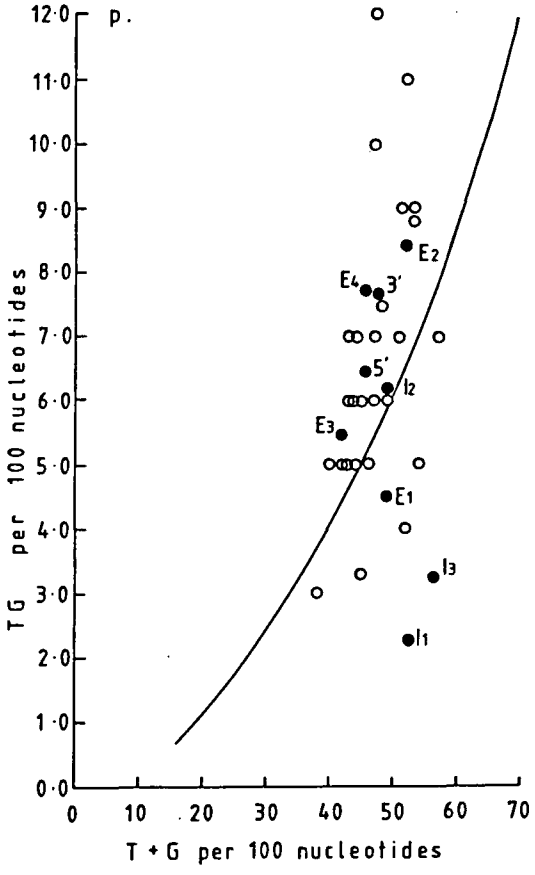

j.
AG per 100 nucleotides
A+G per 100 nucleotides
E1
E2
E3
E4
I1
I2
I3
5'
3'

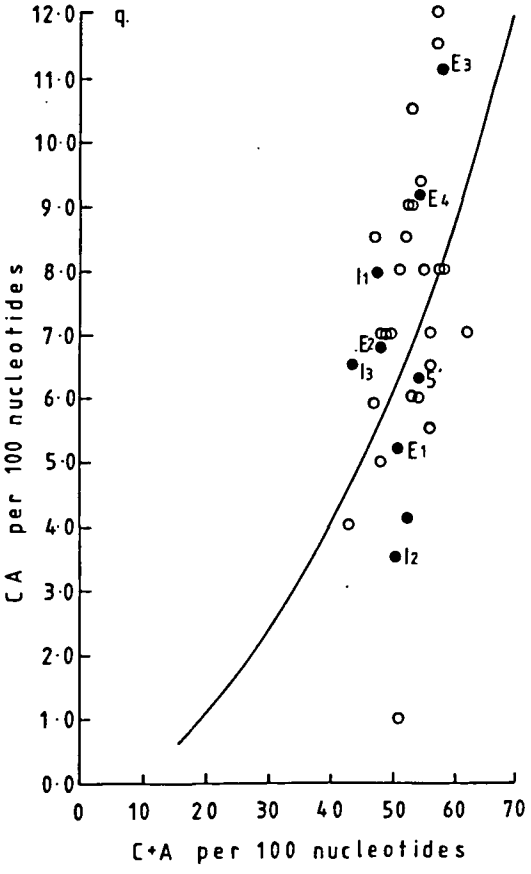

k.
TA per 100 nucleotides
T+A per 100 nucleotides
E1
E2
E3
E4
I1
I2
I3
5'
3'


l.
AT per 100 nucleotides
A+T per 100 nucleotides
E1
E2
E3
E4
I1
I2
I3
5'
3'


m.
CG per 100 nucleotides
C+G per 100 nucleotides
E1
E4
3'
I1
E2
E3
I2
5'
I3


n.
GC per 100 nucleotides
G+C per 100 nucleotides
E4
E3
E1
E2
5'
I2
3'
I3
I1


p.
TG per 100 nucleotides
T+G per 100 nucleotides
E2
E4
3'
5'
I2
E3
E1
I3
I1


q.
CA per 100 nucleotides
C+A per 100 nucleotides
E3
E4
I1
E2
I3
5'
E1
I2

graphs indicates the results that would be expected if the distribution of bases were random. This analysis indicated several interesting features. First, in Leg A, for the dinucleotides CT, TC, GA, AG and AT (g, h, i, j and l), the distribution of the dinucleotides was similar to that predicted.

For the CG dinucleotides (graph a) the observed frequencies (with one exception) were lower than were expected, even at the higher C+G percentage compositions ie the CG suppression was not restricted to regions with a low C+G percentage composition. In addition, the CG suppression was not restricted to either the coding or non-coding regions of the gene. The exception to this generalisation was exon 1, which displayed almost the expected frequency of CGs. By contrast, the GC dinucleotide distribution (b) had several different features. Where the DNA had a percentage composition of greater than 35% G+C, the observed frequency exceeded that predicted and whilst the non-protein coding gene regions (I1, I2, 5' and 3') all had nearly the expected frequency of GC dinucleotides, three of the exons had a higher than was expected frequency of GCs.

The results obtained for the CA and TG dinucleotides were also interesting. The percentage composition of the DNA fell into a very narrow range (between 40-60%). For both of these nucleotides, for the relatively lower percentage compositions, the observed frequency of the dinucleotide was lower than was expected, but when the percentage composition exceeded 45%, the observed frequency exceeded that predicted. For the CA dinuleotide (d), all of the exons and the three introns had a higher than was expected frequency of the dinucleotide,

dinucleotide in exon 1 and intron 3 was very different to the Leg A situation (d). The CA frequency distribution for Leg D (q) was also similar to that described for Leg A. However, exon 1 and intron 2 had a very different CA frequency to that observed for Leg A.

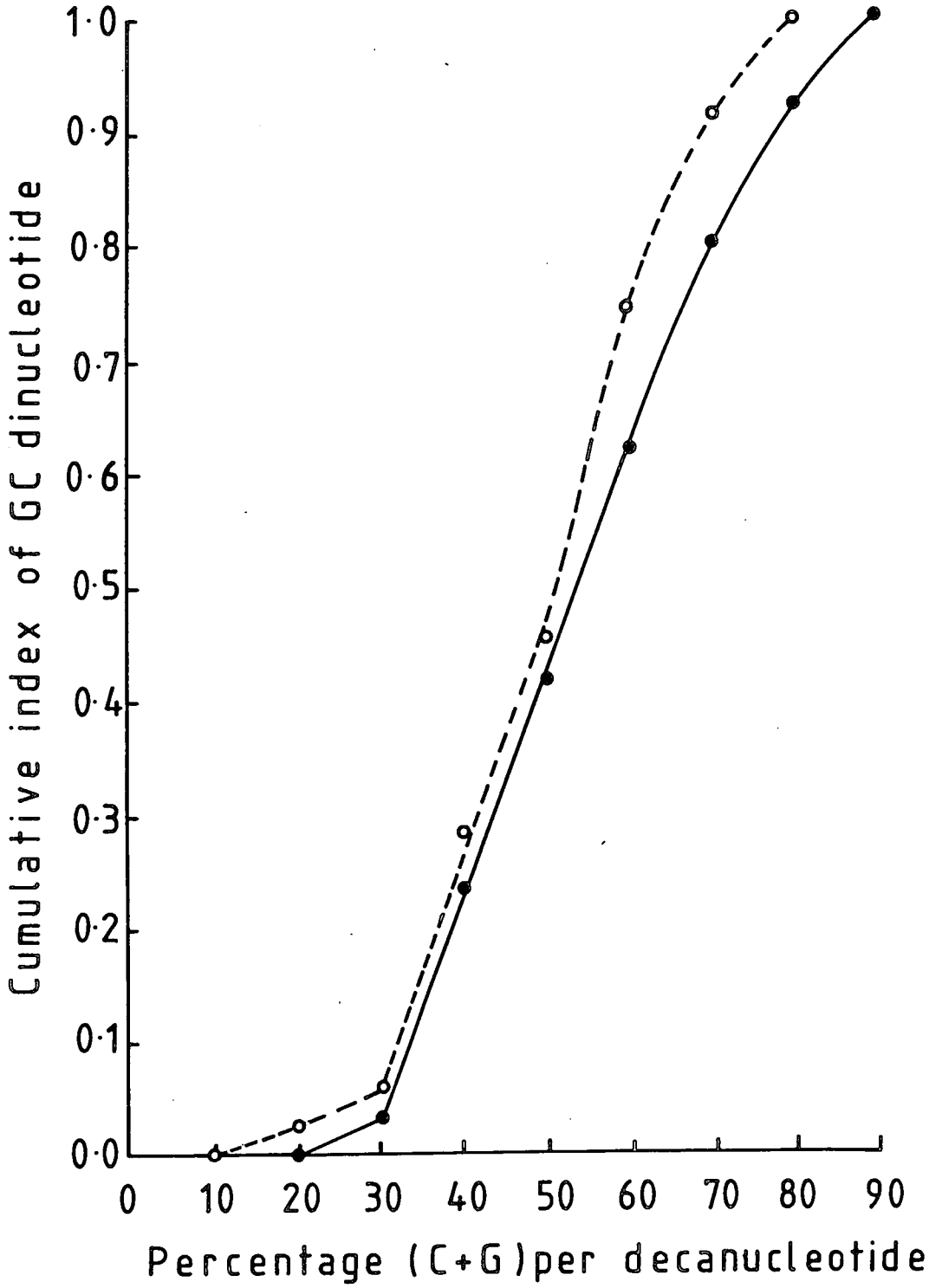
When the microenvironment around each CG or GC dinucleotide was examined (table 10 and figures 49 and 50), it was found that most sites occurred in decanucleotides which were between 40 and 70% C+G rich. These percentages were much higher than was observed for the larger stretches of DNA (figure 48a and b, m and n), where it was rare to find a percentage composition of over 50% C+G. When the microenvironment of these sites was assessed (figures 49 and 50) no marked difference was found between either of these genes or the dinucleotides, although there was a tendency for the GC dinucleotide to occur in a microenvironment with a slightly lower C+G percentage composition.

Figure 51 (a and b) illustrates the effect on dinucleotide frequencies, of the deamination of a 5-methylcytosine residue in a CG dinucleotide. Figures 52 and 53 show the relationship between the CG ratio and the CA+TG ratio for Leg A. Figure 52 shows the correlation between the CA+TG excess and CG deficiency in Leg A. Exon 4 and intron 3 were shown to be in general very CA+TG rich compared with their CG deficiency, although a few individual 100 nucleotide segments had both a CG deficiency and a CA+TG deficiency.

When the data was interpreted in terms of the observed to calculated ratios, it was found that when the overall gene regions were examined (figure 53a :I), the CA+TG ratio was found to be greater

Table 10: Microenvironment of CG and GC dinucleotides

Percentage composition (C+G) of decanucleotide	Number of sites occurring		
	Legumin gene A		Leg D
	CG	GC	CG
20	0	1	1
30	3	17	1
40	10	32	8
50	10	37	6
60	11	34	10
70	10	13	6
80	7	6	3
90	4	6	0
100	0	0	0
Total	55	146	35


Cumulative index of GC dinucleotide
1·0
0·9
0·8
0·7
0·6
0·5
0·4
0·3
0·2
0·1
0·0
0
10
20
30
40
50
60
70
80
90
Percentage (C+G) per decanucleotide

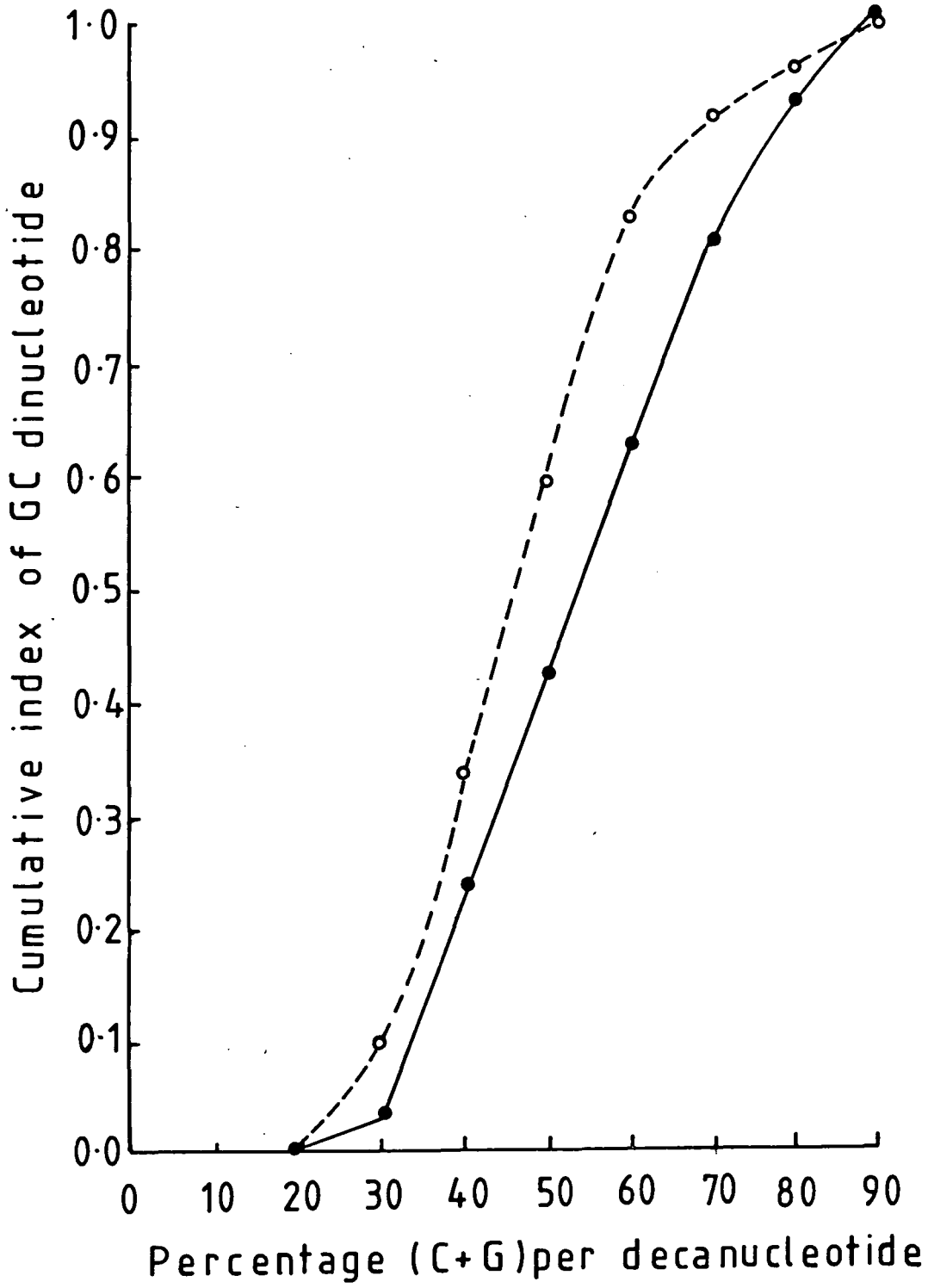
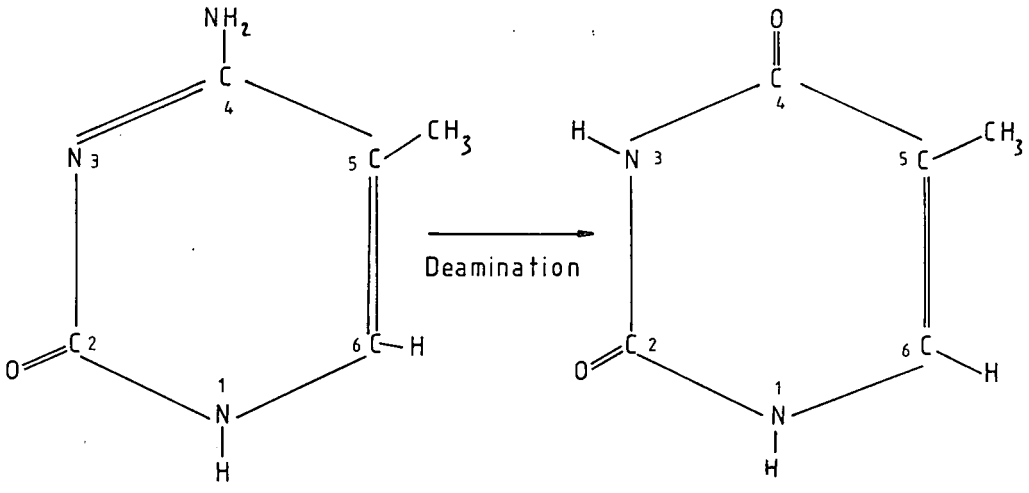

1·0
0·9
0·8
0·7
0·6
0·5
0·4
0·3
0·2
0·1
0·0
Cumulative index of GC dinucleotide
0
10
20
30
40
50
60
70
80
90
Percentage (C+G)per decanucleotide

Figure 51a:- Deamination of 5-methylcytosine to thymine.

Figure 51b:- Effect of deamination in a CG dinucleotide pair

a.



5-methylcytosine

Thymine

b.

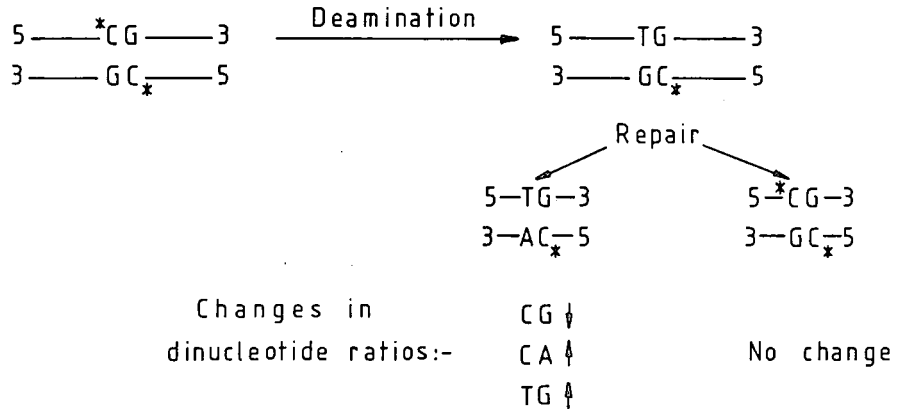


Figure 52:- Correlation between CG deficiency and CA+TG excess in Leg A (pDUB 24).

The symbols used are: O = 100 nucleotide segments and ● = gene regions (calculated on a 100 nucleotide basis).

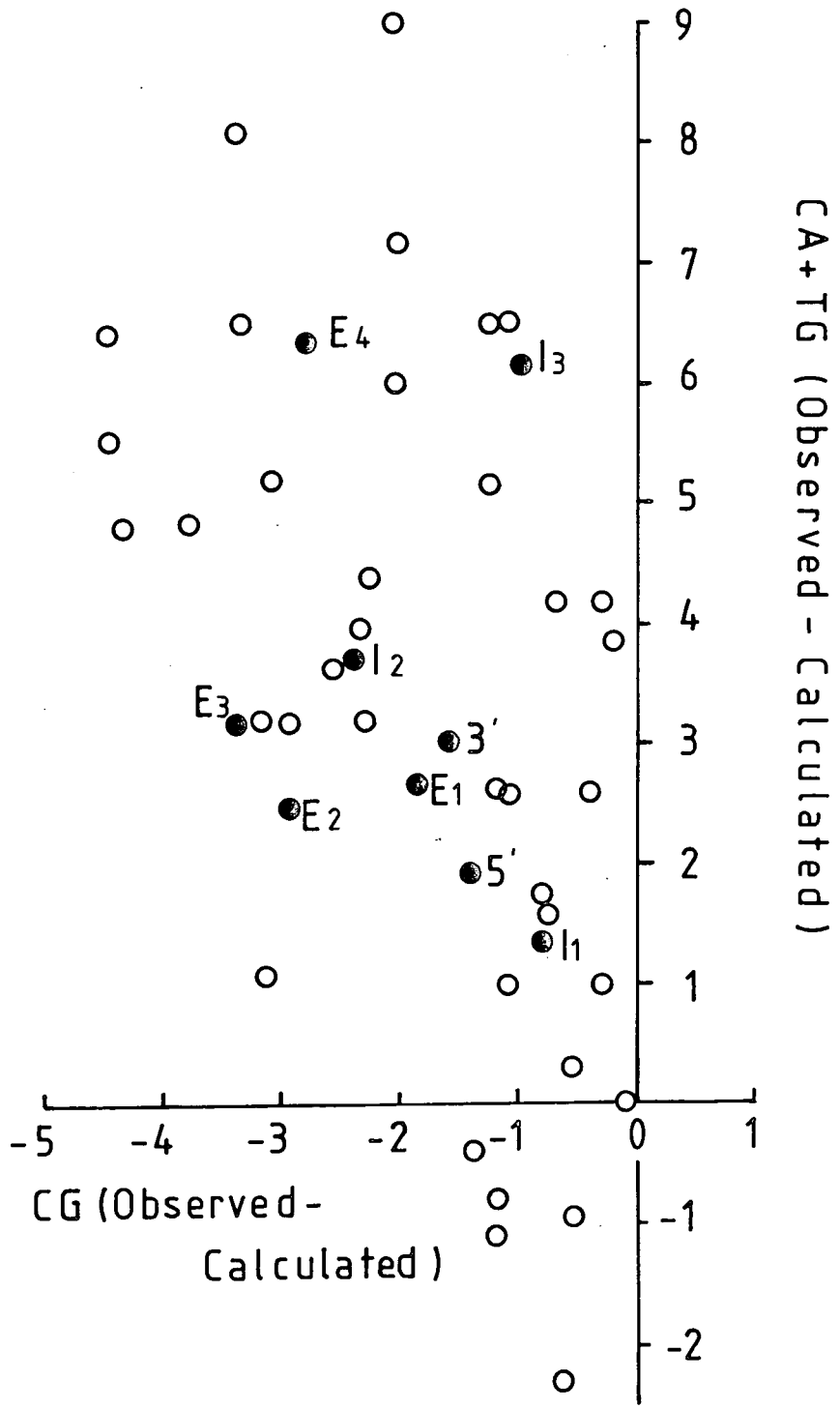

E4
I3
I2
E3
3'
E1
E2
5'
I1
9
8
7
6
5
4
3
2
1
-1
-2
-5
-4
-3
-2
-1
0
1
CA+TG (Observed - Calculated)
CG (Observed - Calculated)

Figure 53:- Comparison of the observed to calculated ratios for CG and CA + TG dinucleotides in legumin gene A (pDUB 24).

Figure 53a:- I Ratio of observed number (CG and CA+TG) to the number calculated for each gene region of Leg A.

II Map showing the coding and non-coding regions of Leg A.

Figure 53b:- I Ratio of observed number (CG and CA+TG) to the number calculated for 100 nucleotide segments of Leg A.

II Map showing the coding and non-coding regions of Leg A.

b.

than 1.0, throughout the gene, whilst that of CG was generally less than 0.5. There was no obvious correlation, in any gene region, between an increased CA+TG ratio and a particularly decreased CG ratio.

When the relationship between CA+TG and CG frequencies was examined in terms of 100 nucleotide segments (figure 53b :I), there was one noticable area, where a correlation did seem to exist. In this region (+ 100 to + 300 bp from the start of the protein coding sequence), there was an apparent relief from CG-suppression (ie the CG dinucleotide occurred at almost the frequency expected) and there was a corresponding expected occurrence of the CA+TG ratio (ie observed to calculated ratio was about 1.0). This region corresponded to the cluster of CG dinucleotides noted in figure 45, in exon 2.

When the sequence data for Leg D was analysed in a similar manner, there was again no obvious correlation in the ratios obtained for the gene regions (figure 54a :I). When the sequence data was analysed in 100 nucleotide segments a correlation was observed in exon 1 (+ 100 to + 300bp), similar to that described for Leg A (figure 53b).

If the cause of the CG suppression effect is the mutability of 5-methylcytosine, then in an analysis of plant gene sequence data, a correlation should also be observed between the methylated trinucleotides 5'-CAG-3' and 5'-CTG-3' and their mutation products, 5'-TAG-3' + 5'-CTA-3' and 5'-TTG-3' + 5'-CAA-3', respectively (table 11). No apparent correlation was evident in this table. In fact, the CAG trinucleotide did not seem to be suppressed at all, whilst the CTG trinucleotide was only suppressed in the flanking regions and introns

Figure 54 :- Comparison of the observed to calculated ratios for CG and CA+TG dinucleotides in legumin gene D.

Figure 54a:-

I Ratio of observed number (CG and CA+TG) to the number calculated for each gene region.

II Map showing the coding and non-coding regions of Leg D.

Figure 54b:-

I Ratio of observed number (CG and CA+TG) to the number calculated for 100 nucleotide segments of Leg D.

II Map showing the coding and non-coding regions of Leg D.

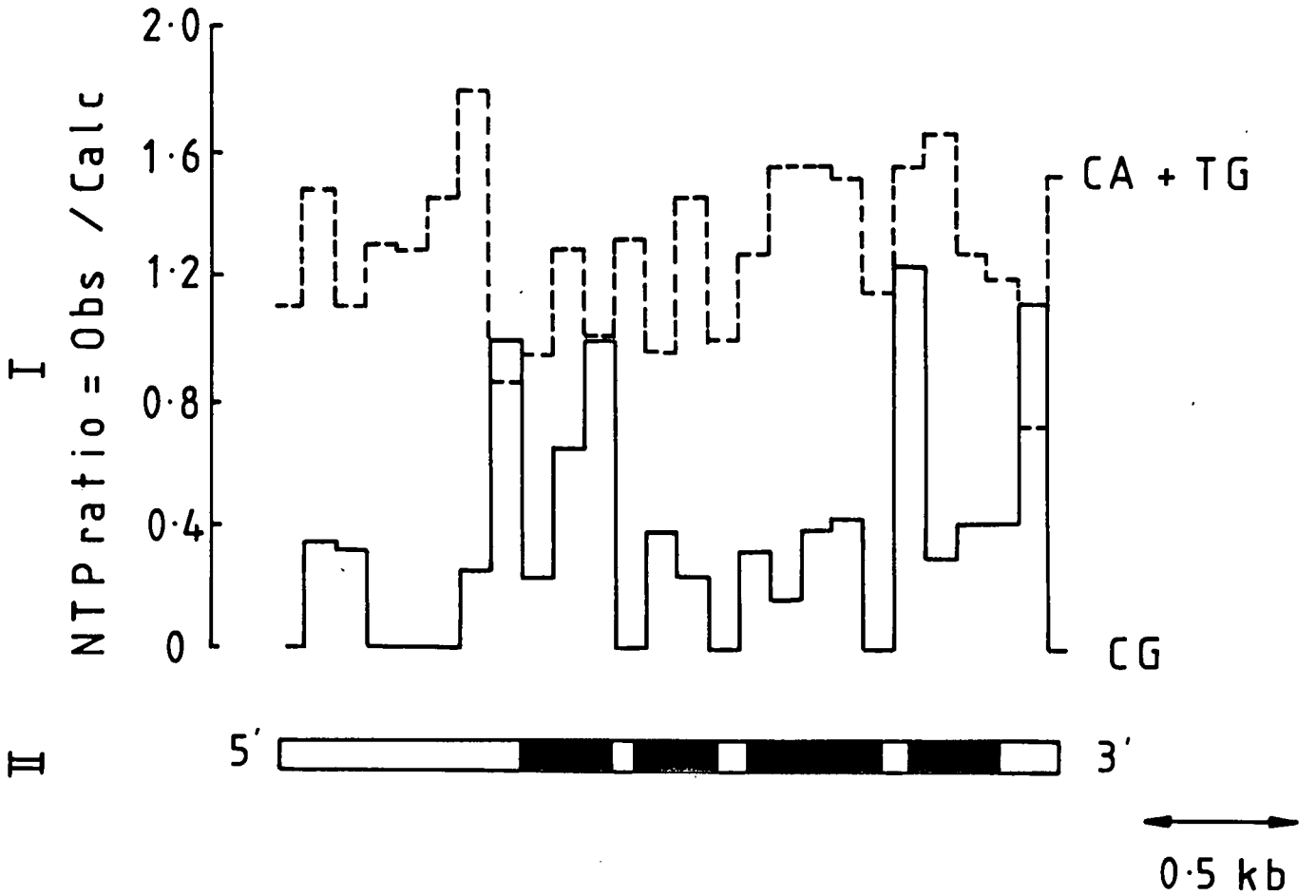

I
II
NTP ratio = Obs / Calc
2·0
1·6
1·2
0·8
0·4
0
CA + TG
CG
5′
3′
0·5 kb

Table 11: Frequency ratios (observed/calculated for selected di and trinucleotides in Leg A

Length (bp)	Total 3692	5'F 1231	E1 286	I1 89	E2 251	I2 89	E3 626	I3 99	E4 391	3'F 630
Sequence (5'→3')										
CG	0.44	0.36	0.68	0.58	0.40	0	0.41	0.51	0.45	0.33
CA	1.27	1.21	1.12	1.36	1.31	1.35	1.34	2.05	1.51	1.26
TG	1.13	1.16	1.33	.82	1.09	1.32	1.14	1.27	1.52	1.30
CA + TG	1.22	1.18	1.22	1.14	1.20	1.34	1.27	1.62	1.53	1.28
CAG	1.36	1.01	1.01	0	2.18	2.50	1.57	1.73	1.79	.52
TAG	0.65	0.81	0.52	1.81	0.44	0.71	0.32	0.62	0.87	0.76
CTA	0.76	0.53	0.94	1.44	0.50	1.61	0.72	1.07	1.29	0.94
TAG + CTA	0.70	0.66	0.76	1.57	0.47	1.23	0.49	0.86	1.07	0.83
CTG	0.82	0.53	0.90	0	1.25	1.48	1.14	0	1.67	0.44
TTG	1.17	1.04	1.39	0.68	1.99	0.84	1.89	0.86	1.49	1.15
CAA	1.28	1.12	2.50	1.92	1.11	0.57	1.31	1.55	2.15	0.82
TTG + CAA	1.24	1.08	1.98	1.32	1.52	0.68	1.41	1.10	1.88	0.97
CCG	0.74	0.72	0.44	0	1.17	0	0.53	0	0.70	1.00
CTG	0.82	0.53	0.90	0	1.24	1.47	1.14	0	1.66	0.44
TTG	1.17	1.04	1.39	0.68	1.99	0.84	1.89	0.86	1.49	1.15
CTG + TTG	1.04	0.88	1.14	0.46	1.66	1.07	1.45	0.63	1.57	0.93
CCG	0.74	0.72	0.44	0	1.17	0	0.53	0	0.70	1.00
TCG	0.39	0.32	1.13	0	0	0	0.19	1.20	0.42	0.44
TTG	1.17	1.04	1.39	0.68	1.99	0.84	1.89	0.86	1.49	1.15
TCG + TTG	.74	0.82	1.23	0.46	1.11	0.54	0.89	0.95	0.88	0.93
CGG	0.44	0.49	0.44	0	0.34	0	0.20	0	0.42	0.72
CAG	1.36	1.01	1.01	0	2.18	2.50	1.57	1.73	1.79	0.52
CAA	1.28	1.12	2.50	1.92	1.11	0.57	1.31	1.55	2.15	0.82
CAG + CAA	1.34	1.09	1.83	1.44	1.59	1.18	1.42	1.61	1.99	0.94
CGG	0.44	0.49	0.44	0	0.34	0	0.20	0	0.42	0.72
CGA	0.32	0.20	0.76	0	0.55	0	0.30	1.73	0.16	0.17
CAA	1.28	1.12	2.50	1.93	1.11	0.57	1.31	1.55	2.15	0.82
CGA + CAA	0.93	0.85	1.71	1.44	0.86	0.39	0.87	1.61	1.28	0.61

1 and 3. It is possible that a correlation does exist at a specific location but was not observed because of the size of gene regions examined and therefore, the data was also analysed in terms of 100 nucleotide segments (figures 55 and 56). However, even in these smaller sections no apparent correlation was observed between either (i) a trinucleotide suppression and an elevated level of the two trinucleotide mutation products or (ii) the relief from suppression (eg in a cluster region) and an expected frequency of the mutation product.

The third methylated trinucleotide found in plant DNA is 5'-CCG-3'. The relationship between this trinucleotide and its mutation products (5'-CCG-3'→5'-CTG-3'→5'-TTG-3' or 5'-CCG-3'→5'-TCG-3'→5'-TTG-3') is shown in figures 57a and b and table 11. A rough correlation was observed in exons 3 and 4, between CCG suppression and the elevation of the mutation products CTG+TTG.

In a similar manner, the mutation of the CCG trinucleotide on the other non-sequenced DNA strand (the RNA template strand) was monitored by the comparison of the complementary trinucleotides on the sequenced strand (ie 5'-CCG-3'→5'-CTG-3'→5'TTG-3' on the RNA template strand corresponds to 5'-CGG-3'→5'-CAG-3'→5'-CAA-3' on the complementary, sequenced strand and 5'-CCG-3'→5'TCG-3'→5'TTG-3' on the RNA template strand corresponds to 5'-CGG-3'→5'-CGA-3'→5'-CAA-3' on the complementary, sequenced strand) (table 11; figure 58a and b). This time, a stronger correlation was observed, particularly in figure 58a, between the trinucleotide suppression and the mutation products.

The information in table 12 is presented in a form which allows

Figure 55:- Comparison of observed to calculated values for the trinucleotides CAG and TAG + CTA in legumin gene A (pDUB 24).

I Ratio of observed frequency (CAG and TAG + CTA) to the predicted frequency calculated for 100 nucleotide segments of Leg A.

II Position of CAG trinucleotides in Leg A.

III Location of protein coding sequences in Leg A.

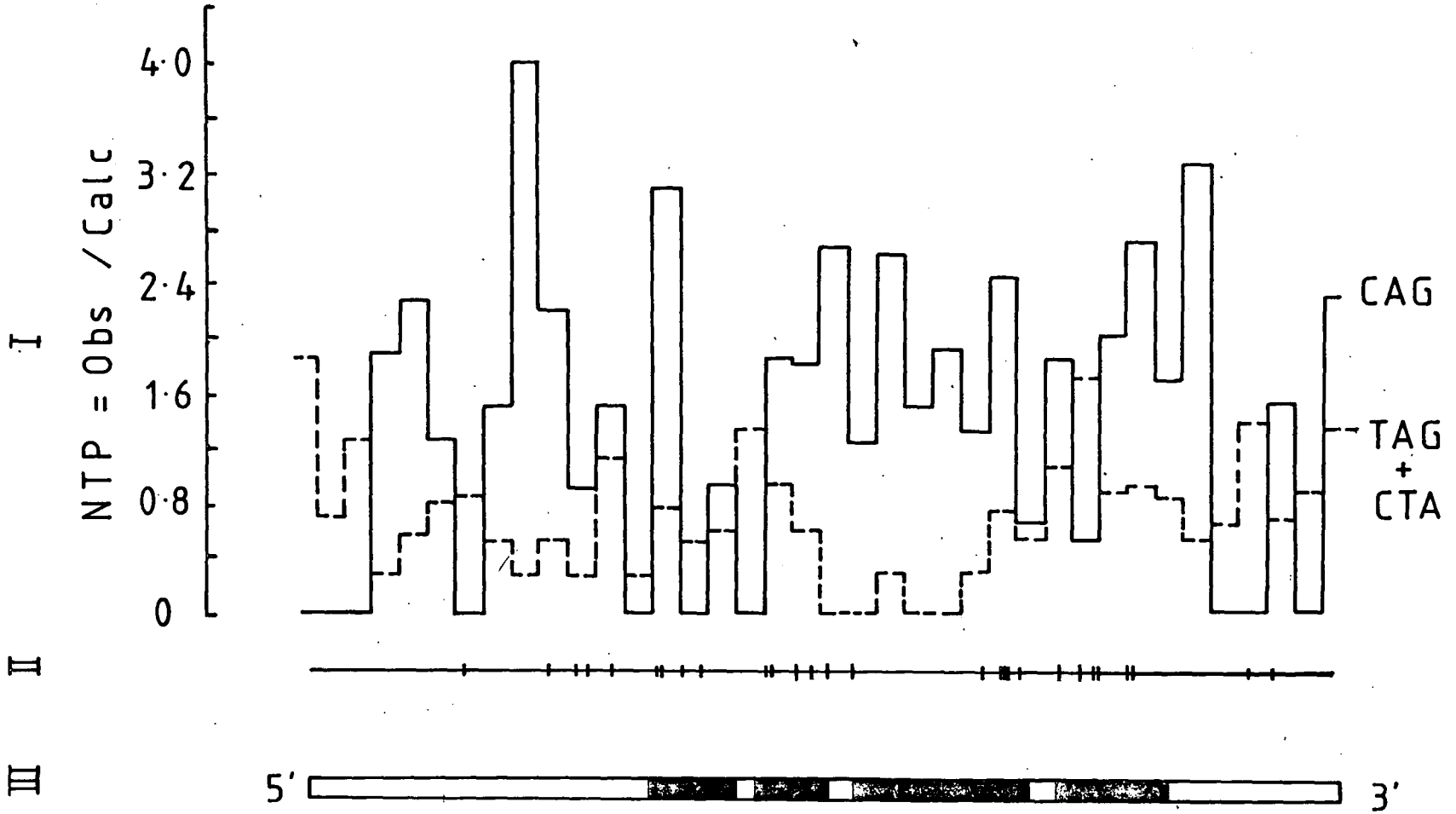

I
II
III
NTP = Obs / Calc
4·0
3·2
2·4
1·6
0·8
0
CAG
TAG
+
CTA
5'
3'

Figure 56:- Comparison of observed to calculated values for the trinucleotides (CTG and TTG+CAA) in legumin gene A (pDUB 24).

I Ratio of observed frequency (CTG and TTG+CAA) to the predicted frequency calculated for 100 nucleotide segments of Leg A.

II Position of CTG trinucleotides in Leg A.

III Position of protein coding sequences in Leg A.

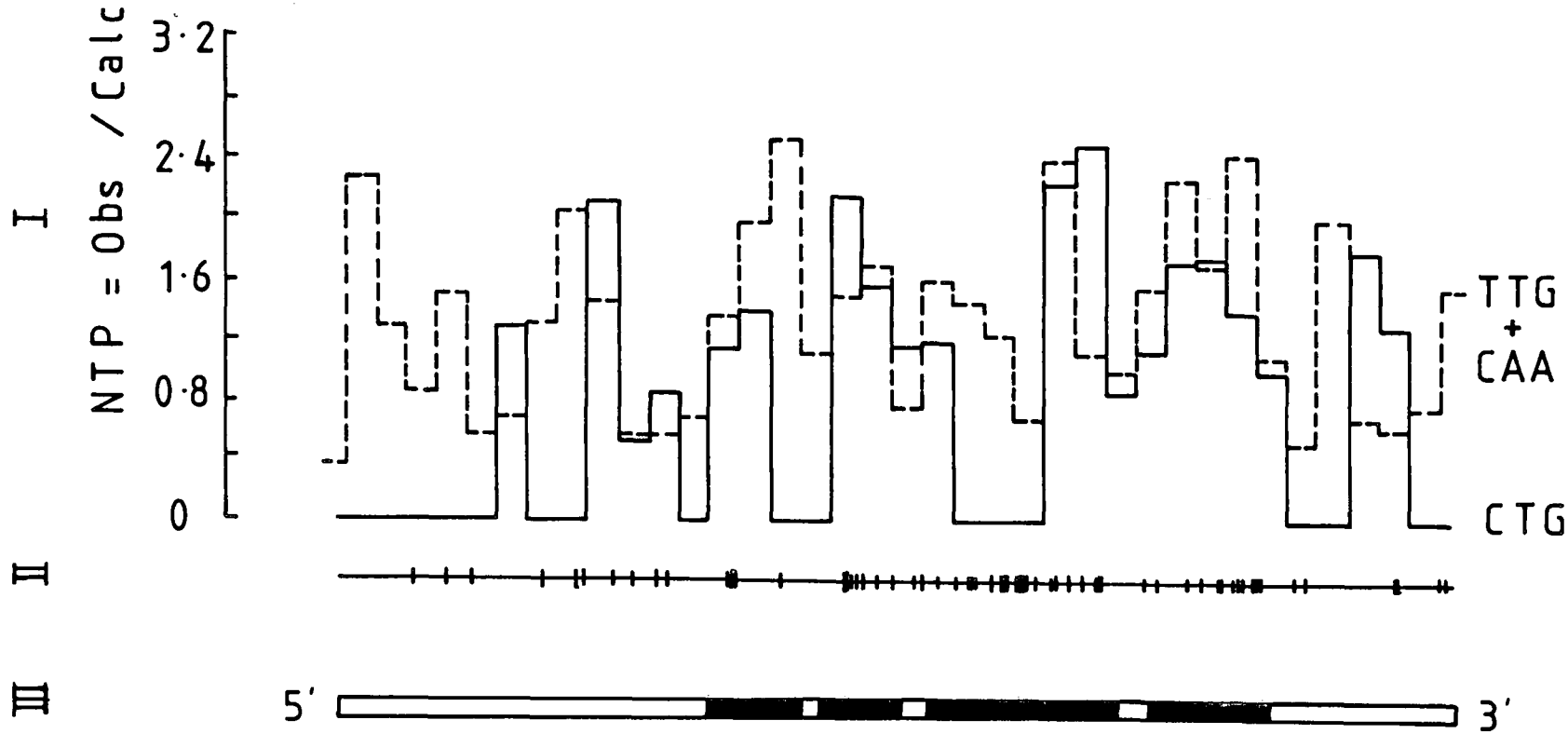

I
II
III
NTP = Obs / Calc
3·2
2·4
1·6
0·8
0
TTG
+
CAA
CTG
5'
3'

Figure 57:- Comparison of observed to calculated values for the trinucleotides CCG and CTG+TTG and TCG+TTG in legumin gene A (pDUB 24).

Figure 57a:- I Ratio of observed frequency (CCG and CTG+TTG) to the predicted frequency calculated for the gene regions of Leg A.

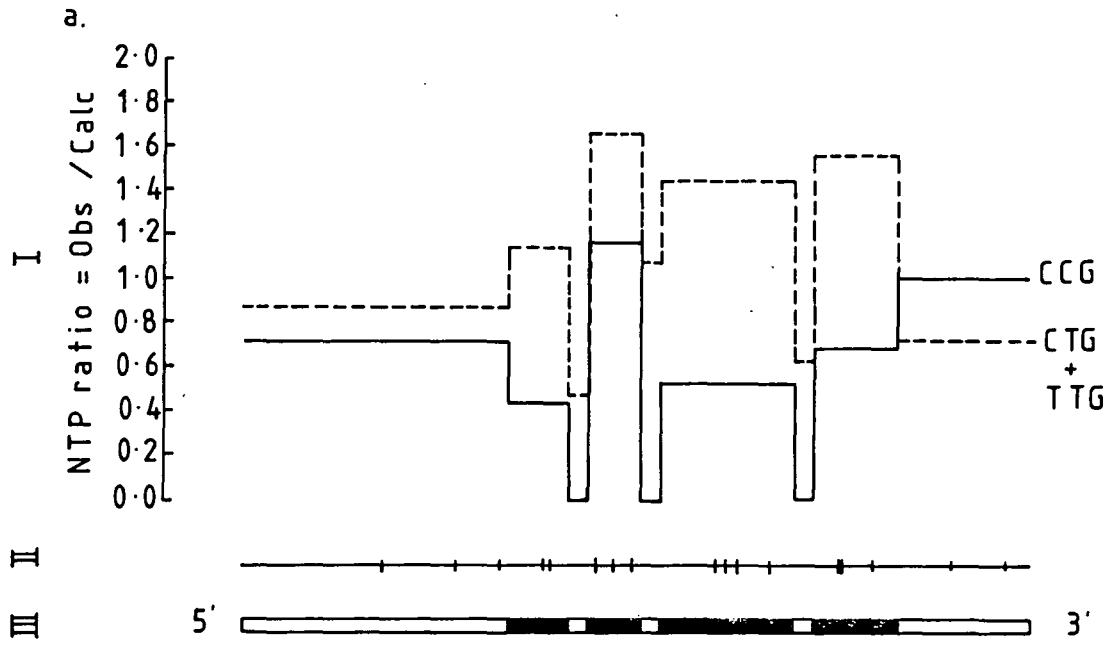
II Position of CCG trinucleotides.

III Location of protein coding sequences in Leg A.

Figure 57b:- I Ratio of observed frequency (CCG and TCG+TTG) to the predicted frequency calculated for the gene regions of Leg A.

II Position of CCG trinucleotides.

III Location of protein coding sequences in Leg A.


a.
I
II
III
NTP ratio = Obs / Calc
2·0
1·8
1·6
1·4
1·2
1·0
0·8
0·6
0·4
0·2
0·0
CCG
CTG
+
TTG
5'
3'

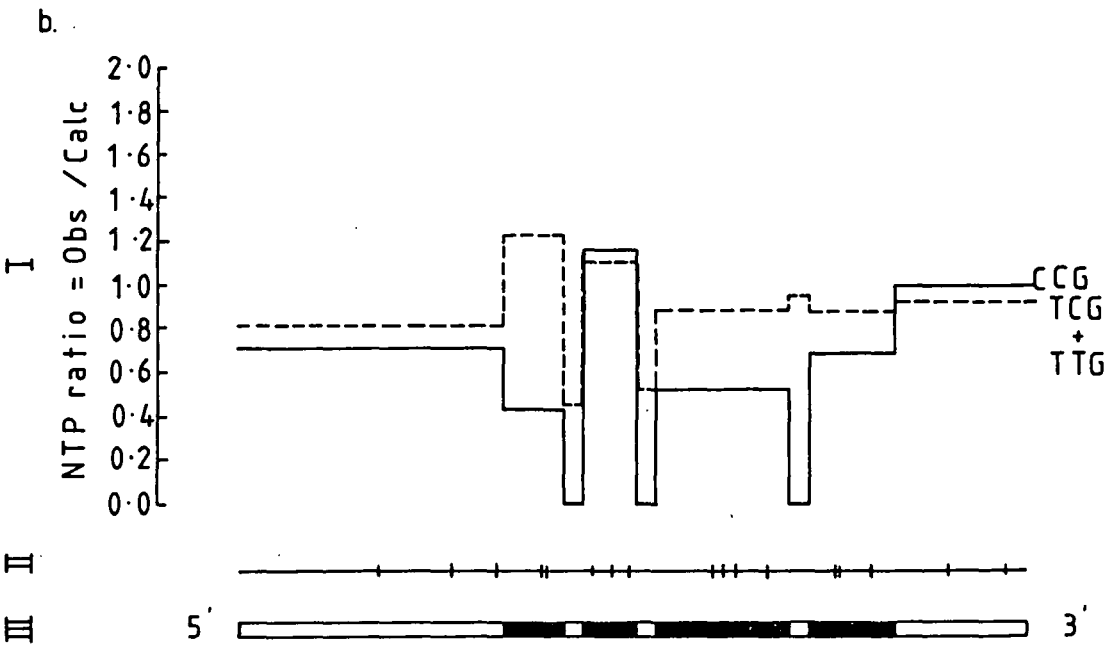

b.
I
II
III
NTP ratio = Obs / Calc
2·0
1·8
1·6
1·4
1·2
1·0
0·8
0·6
0·4
0·2
0·0
CCG
TCG
+
TTG
5'
3'

Figure 58:- Comparison of observed to calculated values for the trinucleotides CGG and CAG+CAA and CGA+CAA in legumin gene A (pDUB 24).

Figure 58a:- I Ratio observed frequency (CGG and CAG+CAA) to the predicted frequency calculated for the gene regions of Leg A.

II Position of CGG trinucleotides.

III Location of protein coding sequences in Leg A.

Figure 58b:- I Ratio observed frequency (CGG and CGA+CAA) to the predicted frequency calculated for the gene regions of Leg A.

II Position of CGG trinucleotides.

III Location of protein coding sequences in Leg A.

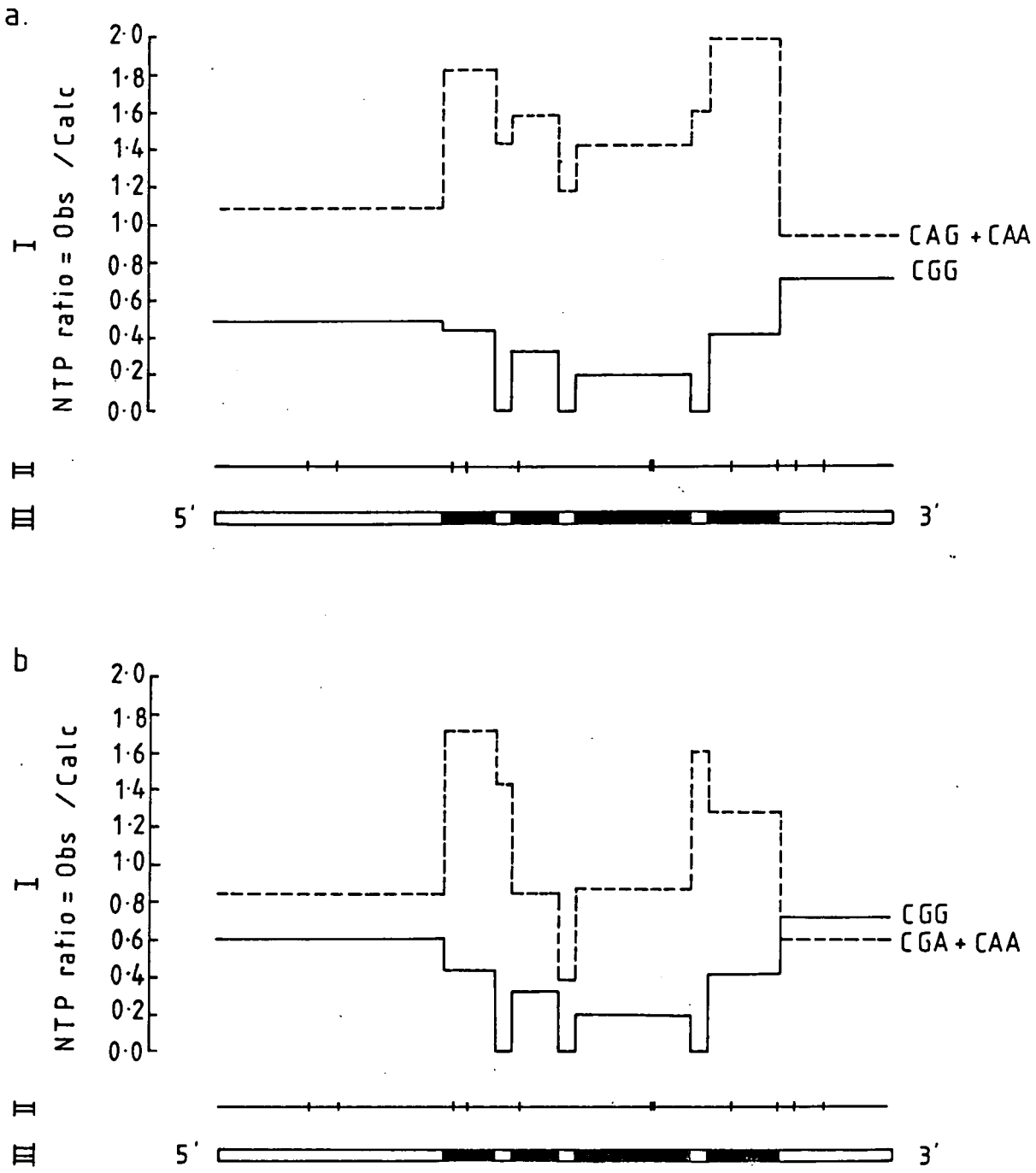

a.
I
NTP ratio = Obs /Calc
2·0
1·8
1·6
1·4
1·2
1·0
0·8
0·6
0·4
0·2
0·0
CAG + CAA
CGG
II
III
5'
3'
b
I
NTP ratio = Obs /Calc
2·0
1·8
1·6
1·4
1·2
1·0
0·8
0·6
0·4
0·2
0·0
CGG
CGA + CAA
II
III
5'
3'

Table 12: Summary of occurrence of selected di and trinucleotides in the protein and non-protein coding sequence of Leg A

Nucleotide Sequence	NON-PROTEIN CODING			PROTEIN CODING		
	obs	calc	%	obs.	calc	%
CA	145	114.3	126.9	145	108.5	133.7
TG	132	109.7	120.7	101	79.2	127.6
CG	17	48.5	35.1	40	85.0	47.1
CCG	5	7.2	69.8	12	19.0	63.3
CGG	4	7.6	53.0	7	21.2	33.1
CAG	16	17.6	90.8	43	27.1	158.6
CTG	8	16.2	49.5	23	17.7	129.7
CTA	29	38.3	75.8	20	22.5	89.1
TAG	32	39.8	80.4	12	24.7	48.7
TTG	38	36.6	103.7	29	17.4	166.6
CAA	43	40.9	105.1	57	34.8	163.9
TCG	6	16.2	37.1	8	17.7	45.1
CGA	4	17.6	22.7	10	27.1	36.8
CA + TG	277	224.0	123.7	246	187.6	131.1
CTG + CAG	24	33.8	71.1	66	44.9	147.2

the effects of protein coding requirements on the 5mC mutation rate, to be assessed. The major feature of this table is the marked difference between the protein coding and non-protein coding region, percentage occurrence of the CAG and CTG trinucleotides. Both of these trinucleotides were found to be relatively deficient in the non-protein coding regions, but were enhanced in the protein coding regions. The TTG and the CCA trinucleotides were also shown to occur far more frequently than was predicted in the protein coding regions. The CGG trinucleotide was shown to be suppressed to a greater extent in the protein coding region than in the non-protein coding regions.

When the codons, in which the cytosine of each CG dinucleotide were assessed (tables 13 and 14), it was found that overall, 40% of the cytosines (16 out of 40) occurred in the first codon position, 17.5% (7 out of 40) in the second position and 42.5% in the third codon position. The first exon was found to have a higher percentage (54.5%) of cytosines in the first codon position, whilst in exon 2 no cytosines, which were part of a CG dinucleotide, were found in the second codon position.

The codon position of the guanine base in a CG dinucleotide (table 15), reflects the location of the complementary cytosine in a CG dinucleotide on the RNA template strand. Only 17.5% of these guanines were located in the third codon position. Similarily, no external guanines in the 5'-CGG-3' sequence (complementary to 5'-CCG-3' on the template strand) and 28.6% of the internal guanines, were found in the third codon position. In contrast, over 64% of the thymines in TG dinucleotides and just under 70% of the thymines in the CTG

Table 13: Identification of codons in which the cytosine of the CG dinucleotides, in legumin gene A, occur

EXON 1							
Codon	GGC	CGC	CTC	CTC	CGT	TCG	CGA
Amino Acid	G	R	L	L	R	S	R
Codon Position	3	1	3	3	1	2	1
Codon	CGT	CGC	AAC	CGC			
Amino Acid	R	R	N	R			
Codon Position	1	1	3	1			
EXON 2							
Codon	CCC	CGC	CGA	AAC	GCC		
Amino Acid	P	R	R	N	A		
Codon Position	3	1	1	3	3		
EXON 3							
Codon	CAC	AAC	GAC	CCC	GCG	CGC	CCG
Amino Acid	H	N	D	P	A	R	P
Codon Position	3	3	3	3	2	1	2
Codon	CGC	CCG	CGT	CGC	GGC	CGA	CCG
Amino Acid	R	P	R	R	G	R	P
Codon Position	1	2	1	1+3	3	1	2
EXON 4							
Codon	TAC	CGT	ACC	GCC	CGT	AGC	CCG
Amino Acid	Y	R	T	A	R	S	P
Codon Position	3	1	3	3	1	3	2
Codon	CGT	TCG					
Amino Acid	R	S					
Codon Postion	1	2					

Table 14: Summary of codon position of cytosine residues which are part of a CG dinucleotide (legumin gene A data)

	CODON POSITION 1	2	3	Number of CGs per exon
Exon 1				
Frequency	6	1	4	11
Percentage	54.5	9.1	36.4	
Exon 2				
Frequency	2	0	3	5
Percentage	40.0	0	60.0	
Exon 3				
Freqency	5	4	6	15
Percentage	33.3	22.2	44.4	
Exon 4				
Frequency	3	2	4	9
Percentage	33.3	22.2	44.4	
Total				
Frequency	16	7	17	40
Percentage	40.0	17.5	42.5	

Table 16: Arginine codon usage in legumin gene A

EXON	CGT	CGC	CGA	CGG	AGA	AGG
1	2	3	1	0	2	0
2	0	1	1	0	5	1
3	1	3	1	0	10	7
4	3	0	0	0	3	4

Table 15: Summary of codon position of specified nucleotides in selected di and trinucleotide sequences in Leg A.

Nucleotide sequence	FREQUENCY			PERCENTAGE		
	Codon Position			Codon Position		
	1	2	3	1	2	3
CG	16	7	17	40.0	17.5	42.5
CG	17	16	7	42.5	40.0	17.5
TG	11	25	65	10.9	24.8	64.4
CA	65	48	32	44.8	33.1	22.1
CCG	4	5	3	33.3	41.7	25.0
CCG	3	4	5	25.0	33.3	41.7
CGG	2	5	0	28.6	71.4	0
CGG	5	0	2	71.4	0	28.6
CTG	3	4	16	13.0	17.4	69.6
CTG	4	16	3	17.4	69.6	13.0
CAG	19	13	11	44.2	30.2	25.6
CAG	13	11	19	30.2	25.6	44.2
CAG	11	19	13	25.6	44.2	30.2
TTG	7	18	4	24.1	62.5	13.8
TTG	4	7	18	13.8	24.1	62.5
CAA	6	28	23	10.5	49.1	40.4
CAA	28	23	6	49.1	40.4	10.5
TAG	0	8	4	0	66.7	33.3
TAG	4	0	8	33.3	0	66.7
CTA	7	5	8	35.0	25.0	40.0
CTA	5	8	7	25.0	40.0	35.0
TCG	2	2	4	25.0	25.0	50.0
TCG	4	2	2	50.0	25.0	25.0
CGA	0	7	3	0	70.0	30.0
CGA	7	3	0	70.0	30.0	0

The codon position refers to the position of the underlined base.

trinucleotide (the potential products of 5mC mutations) were found in the third codon position.

It was also noted that 6 cytosine residues, out of the 11 CG dinucleotides in exon 1 were located in an arginine codon and overall 16 cytosines out of a total of 40 CG dinucleotides were located in arginine codons. The arginine codon usage in Leg A, was also examined (table 16). In general, the arginine codons AGA and AGG were preferred (32 codons out of a possible 47 were of this nature), whilst in exon 1, the CG containing arginine codons were preferred.

4: DISCUSSION

4.1 Distribution of 5mC in Pea Genomic DNA

Both the activity of the isoschizomers and their specificity of cleavage were confirmed using unmethylated SV40 DNA. This control experiment, which is presented in detail in the ribosomal section (4:2), demonstrated that Msp I and Hpa II cleaved the unmethylated DNA molecule to produce the DNA fragments of the sizes predicted from published sequence data.

Figure 8a (tracks a, b and c) shows the electrophoretic fragment distribution of unrestricted, Hpa II digested and Msp I digested 9 d.a.f. genomic DNA. The unrestricted DNA, was incubated in Msp I buffer at 37°C and acted as a control to ensure that the observed degradation of the DNA observed in track c, was due to enzymatic digestion and not to buffer- or heat-induced, non-specific degradation of the DNA. The bulk of the DNA in the unrestricted track a, comprises DNA fragments of greater than 30kb in length.

After digestion with Msp I (track c), no DNA remained in the very high molecular weight range. The bulk of the DNA was located in fragments of between 20kb and 0.5kb. Following Hpa II digestion (track b), the bulk of the DNA remained in the very high molecular weight range but some DNA was cleaved to yield fragments of between 500bp to 30kb. This implied that the genomic DNA consisted of a very highly methylated fraction (ie the DNA in the very high molecular weight range, 25-30kb, which was essentially resistant to cleavage by Hpa II) and a minor, less highly methylated or 'unmethylated' fraction, ie that which had been cleaved by Hpa II. Similar results were observed at later stages of cotyledon development and for leaf DNA.

The extent of methylation observed in eukaryotes is in contrast to that observed in prokaryotes. In *E. coli* , all of the methylatable sites are methylated ($G^{m}ATC$ and $C^{m}C^{A}_{T}GG$)(Razin *et al* , 1980). In eukaryotes, methylation is incomplete and the 5mC is non-randomly distributed in the DNA. It has a far greater abundance in repetitive DNA than in unique copy DNA and the DNA associated with nucleosomes is also significantly more methylated than that associated with the spacer regions between the nucleosome cores (Razin, 1984).

Bird and Taggart (1980) suggested that the DNA could be divided into two categories or 'domains', the methylated (m+) and the unmethylated (m-) fractions. At present, it is not definitely known whether the unmethylated regions of DNA are found in similar chromosomal positions (or domains) or whether they are interspersed with the unmethylated DNA regions, as suggested by Molitor *et al* , (1976). In vertebrates, the (m+) fraction was predominant over the (m-) fraction (see section 1:4:4). In pea the (m-) fraction appeared to be very small and this view is supported by the fact that there was no region of overlap between either the Hpa II and Msp I or the Bst NI and Eco RII distribution curves (figures 10-14). Had the unmethylated fraction been larger then the unmethylated sites would have been cleaved by both enzymes in each isoschizomeric pair and the same DNA fragments would have been visible in comparative tracks. Thus when the distribution of the DNA was analysed, these unmethylated DNA fragments would have given similar distribution curves for both enzymes and a distinct area of overlap of the curves would have been seen. This was not observed and therefore the 'unmethylated' fraction in pea genomic DNA must be very small.

The shape of the Hpa II and Eco RII molecular size distribution curves (figures 10b, 11b, 14 and 15) also indicated the presence in pea DNA, of multiple methylation states and not a gradually changing level of methylation. The latter situation, ie a continuous distribution of methylation states ranging from completely unmethylated to fully methylated, would have produced smooth distribution curves, following digestion by either Hpa II or Eco RII (Kunnath and Locker, 1982a). Instead, distinct breaks in the distributions were observed at about 20kb (with minor breaks at about 4.0 and 10.0kb) for Hpa II digested cotyledon DNA (indicated in figure 10b) and at about 17.5kb (with a minor break at about 10.0kb) for Eco RII digested DNA (indicated in figure 14). In figure 10b, these breaks separate fractions of DNA with different levels of methylation at the internal cytosine of the Hpa II recognition sequence (ie $C^{m}CGG$).

There was also a marked organ-specific difference between the leaf and developing cotyledon distribution curves for both Msp I and Hpa II. This was particularly noticeable when the leaf + Msp I curve (figure 10a) was superimposed on the cotyledon and leaf + Hpa II curves (figure 10b). The complex nature of the leaf + Msp I distribution (figure 10a) clearly indicated that in leaf DNA, there were additional methylation states at the external cytosine of the Msp I recongition sequence ($^{m}CCGG$ not cleaved by Msp I). However, since the digestion of leaf DNA by Hpa II (which can cleave $^{m}CCGG$) was also rather limited, these new methylations, which were resistant to Msp I cleavage, were probably due to conversion of the partially methylated $C^{m}CGG$ (Hpa II resistant, Msp I sensitive) to the fully methylated $^{m}C^{m}CGG$ (Hpa II and Msp I resistant).

It was not possible to determine from these results whether all of these different methylation states were present in all cells or whether they result from the fact that cotyledons consist of a diverse mixture of cells ie there is cellular heterogeneity, In cotyledon tissue this would be four to five types for example storage parenchyma, xylem parenchyma, phloem (ie phloem parenchyma, sieve elements and companion cells) and epidermal cells. The proportion and transcriptional status of each cell type is not known.

Naveh-Many and Cedar (1982) found in several kinds of cells that overall the genomic DNA was 70% methylated, whilst the DNA in active chromatin (with increased DNAse I sensitivity) and the DNA sequences complementary to poly(A) RNA were only 40% and 20-30% methylated respectively. Kunnath and Locker (1982a) suggested that a developmental or tissue-specific change in total methylation arose by the methylation of the DNA shifting from one state to another, rather than by a gradual changing between states.

A high degree of methylation of plant DNA has previously been reported in wheat, tobacco, cauliflower, mustard and mistletoe (Naveh-Many and Cedar 1982; Wyatt, 1981; Wagner, 1981; Vanyushin *et al* , 1971). All pea cotyledon DNAs were found to be methylated. The values ranged from 11% to 27% methylation (table 2) and initially appeared to represent only a low level of methylation. However, when the Msp I average molecular size (M_n = 2.9 - 4.2kb)(table 2) was compared with the average expected molecular size (440bp) of Msp I or Hpa II digested wheat germ DNA (based on the results of nearest neighbour analyses by Gruenbaum and co-workers (1981)), it is

immediately apparent that these results obtained for pea DNA are 7 to 10 times greater than expected. Msp I has a 4-base recognition sequence and the overall base composition of pea and wheat genomic DNA would not be expected to be significantly different (table 1). Thus, the average size for Msp I digested pea genomic DNA would be expected to be less than 1kb and this is supported by the sizes of fragments obtained when λLeg 2 is restricted with Msp I (figure 29b). The larger than expected, average molecular sizes (M_n), obtained for Msp I digested pea genomic DNA (2.9 to 4.2kb), plus the indication of multiple Msp I methylation states in leaf DNA (figure 10a), therefore imply that many Msp I sites in pea DNA must be present in the double methylated form and are resistant to cleavage by Msp I (and Hpa II).

An alternative explanation for the larger than expected average size of Msp I fragment is that the genomic DNA contains ^{m}CCGG sites (Msp I resistant, Hpa II sensitive) but if this were so, then the bands of repetitive DNA, best seen in the leaf digests (figure 22a: tracks k and l), should be larger following Msp I digestion than after Hpa II digestion and this was not found.

The percentage methylation figures obtained (table 2), therefore reflect the degree of methylation of CCGG sites, methylated only at their internal cytosine and they do not take into account those Msp I sites which are methylated at both cytosines (ie the percentages do not accurately represent the overall methylation of all Msp I/Hpa II CG dinucleotides, in pea genomic DNA). Over 80% of all CG dinucleotides in eukaryotes DNA contain 5mC (eg Naveh-Many and Cedar, 1982) and to date, there is no reason to believe that the subset of CG

dinucleotides that are part of Msp I/Hpa II recognition sequences are treated differently with respect to their methylation, from the general population of CGs (Bird *et al* , 1985).

In vertebrates, changes of less than 10% in the overall methylation have been observed (Razin and Riggs, 1980). A change in total methylation can result in one of two ways: either as a general overall random reduction or as a sequence specific reduction, for example, satellite DNA accounts for only 8-10% of the mouse genome yet it has two to four times the number of methyl groups per μg DNA, than does single copy DNA. A complete loss of all the methyl groups in the satellite DNA could thus account for a 10% - 40% decrease in the percentage of DNA methylation (Reilly *et al* , 1982; Gruenbaum *et al* , 1981c).

The figures obtained for overall percentage methylation (table 2), indicated that there was an apparent decrease in methylation during the middle stages of cotyledon development (9 to 15 d.a.f.). These figures relected the gradually decreasing size of Hpa II fragments (9 to 11 d.a.f.) and the increasing size of the Msp I fragments (11 to 15 d.a.f.). These trends are emphasised in the 'r' values. A change in the value of 'r' indicates a change in the shape of the distribution (Kunnath and Locker, 1982). Therefore, it appears that during the middle stages of cotyledon development, there is a loss of methyl groups form some Msp I/Hpa II sites which had previously been methylated at their internal cytosines. This demethylation has the effect of decreasing the average size of the Hpa II fragments whilst not altering the size of the Msp I fragments

obtained.

During the middle to late stages of cotyledon development, some unmethylated sites become either partially (C^mCGG) or completely methylated, $^mC^mCGG$, (indicated by the increasing size of the Hpa II fragments of 3.99 to 4.23kb), whilst other sites (either CCGG or C^mCGG), become fully methylated (indicated by the increasing size of the Msp I fragments from 2.87 to 3.53kb).

In leaf DNA the average fragment sizes were larger than were detected in any cotyledon DNA (tables 2 and 3) and the ratio 'r' was also much lower for the leaf digests. Taken together, these results confirmed that the leaf DNA distributions obtained, following Msp I or Hpa II digestions, were different from the cotyledon distributions, due to an increased mumber of doubly methylated sites ($^mC^mCGG$).

In the early stages of cotyledon development, DNA synthesis occurs in association with cell division. After 11-12 d.a.f. (under the growth conditions untilised in this study), cell division ceases but DNA synthesis continues (ie endoreduplication). EM studies have shown that most cotyledon cells are synchronised to within 1 day. Marks and Davies (1979) demonstrated that both polyploid and polytene nuclei were present in the cells of developing *Pisum sativum* cotyledons. Polytene cells were about one third as frequent as polyploid cells.

Methylation has been related to new DNA synthesis (Bird, 1978). Therefore, because the level of DNA methylase is not thought to be in excess (Razin and Szyf, 1984), if the rate of DNA synthesis during endoreduplication was faster than that observed during normal DNA

synthesis, hypomethylation of the DNA might be observed. In Feltham First the fastest rate of DNA synthesis was found to be between 15, and 20 d.a.f. (Boulter, 1981). However, no DNA hypomethylation was observed during the endoreduplicating phase of pea seed development. Instead the size of the Hpa II fragment increased (M_n at 11 d.a.f. = 3.36kb and at 21 d.a.f. = 4.12kb) and because the size of the Msp I fragment remained at about 3.5kb during the later stages of cotyledon development (15 - 21 d.a.f.), the overall methylation of the DNA was only seen to slightly increase from 11.1% at 15 d.a.f. to 14.7% at 21 d.a.f.

These results confirm that DNA methylation must occur within a very short time of DNA replication and they show that the methylation of DNA can occur in the absence of cell division. A lag period between DNA synthesis and DNA methylation would have led to a sequential hypomethylation. Bird (1978) showed that replicating DNA underwent methylation at or near the replicating fork. Kappler (1970) and Szyf *et al* , (1982) have concluded that in eukaryotes most DNA methylation occurs within 2 minutes of DNA replication. However, Woodcock *et al* , (1983) reported that some methylation was delayed for several hours and could occur after the S-phase. Razin (1984) has suggested that this may represent the methylation of sites which were particularly resistant and Szyf *et al* , (1984) has stated that methylatable sites have different affinities for the methylase and sites with the lowest affinity will be methylate last (see section 1:4:6). Therefore, in a situation where the methylation quotient (see section 1:4:5) fell below 1.0 (for example rapidly replicating DNA),

there would be a failure to methylated some of these sites and a sequential hypomethylation would result. This was not observed during the period of endoreduplication in pea.

It may be that endoreduplication itself induces an increase in methylase activity (ie a situation analagous to that observed in the proliferating rat liver cells, following a partial hepatectomy (Razin, 1984)). Or does the presence of doubly methylated sites in pea DNA, indicate that the activity of methylase is higher in plants than in vertebrates? It should also be borne in mind that the failure to methylate a few specific sites would not have been detected by this kind of analysis, which can only measure gross changes.

Thus whilst there was an indication that there was some form of demethylation during the early stages of cotyledon development, when DNA synthesis and cell division were occurring, this was not observed to continue into the later stages of cotyledon development, when DNA synthesis occurred without cell division.

In higher plants, additional methyl sequences are located in the trinucleotide sequence C-X-G, where X can be C, A or T (Gruenbaum, 1981b). The previous experiments using Msp I and Hpa II confirmed that in pea DNA the sequence C-C-G could be methylated at both cytosines and the results obtained following Bst NI and Eco RII digestion of pea DNA confirmed that the sequence C-$\overset{A}{\underset{T}{}}$-G was highly methylated.

The percentage methylation figures obtained (table 4), indicated that this trinucleotide sequence was between 67 to 84% methylated and this is in agreement with the value of 80% obtained for wheat germ DNA, by Gruenbaum *et al* (1981b). The pea number average molecular

weights of Bst NI derived fragments were again larger than those predicted for wheat germ DNA (530bp)(*ibid* ; Naveh-Many and Cedar, 1982). There are two possible reasons for this. The first is because of residual fluorescence in the 10-30kb region of the Bst NI tracks. This residual fluorescence appears to be caused by the passage of a large number of genomic DNA fragments through the gel, during electrophoresis and therefore the problem is most pronounced following Bst NI digestion which produces smaller fragments, which migrate further during electrophoresis.

The second reason for an increased average size of DNA fragments may be the method by which the negatives were scanned. Whereas the Msp I/Hpa II negatives were scanned by a laser densitometer, which scans less than 1mm track width of negative, the Bst NI/Eco RII negatives were scanned using a Pye Unicam SP500 spectrometer, which scans a larger area of the negative and therefore the resolution was not as good.

4:2 Methylation of the Genes for Ribosomal RNA

4:2:1 Methylation of rDNA at Msp I/Hpa II sites

In experiments which depended on the differential digestion of DNA, it was first necessary to demonstrate that neither of the enzymes used was being inhibited by contaminants present in the genomic preparations. This was done by monitoring the digestion of an un-methylated control DNA (SV40). included in the genomic digestion. This SV40 was shown to be cleaved by both Msp I and Hpa II, at its one CCGG

sequence (figure 18a). Thus, any apparent difference in the extent of digestion of genomic DNA, by either Msp I and Hpa II could not be due to the presence of contaminating substances and had to be ascribed to the methylation of the DNA.

The ribosomal RNA genes are generally arranged in tandem arrays of repeating units, in the nucleolar organizing regions of chromosomes. The repeating units consist of the coding sequences for the 25S, 18S and 5.8S rRNAs, together with internal and external transcribed spacer regions. These transcribed regions are separated from one another by the non-transcribed spacer (NTS). (For reviews see Mandal, 1984; Leaver, 1979; Long and Dawid, 1980.) Heterogeneity in the NTS has been reported but the coding regions of the ribosomal genes are usually conserved. In higher plants between 70-90% of the ribosomal transcription unit is conserved (Leaver, 1979).

A cloned pea ribosomal gene, pHA 1, hybridized to a series of cotyledon Hpa II derived DNA fragments, which ranged in size from 10kb-0.93kb, although fragments of less than 2.09kb hybridized only weakly (figure 17b, eg track d and figure 18b, eg track f). In contrast the rDNA from the leaf was more highly methylated and the ribosomal probe hybridized to fragments of >20kb and no discrete Hpa II hybridization bands were observed (figure 17b, track u; figure 18b, track g). Thus it can be seen that there was a distinct organ specific difference in the extent of methylation of the ribosomal genes and the rDNA from the developing cotyledons was hypomethylated with respect to that derived from the leaf. The significance of these

discussed in sections 4:2:2 and 4:2:3.

The low average size (<1kb) of the pHA 1 + Msp I fragments (figure 18b, tracks j and k), compared to the length of the ribosomal repeat unit (9.3kb), indicated that there were many sites for Msp I in the cloned ribosomal RNA gene. (It should be noted that this collection of bands also includes those derived from the pACYC 184 vector). Although the number of Msp I sites in the pea rRNA gene is not known, Scott *et al* (1984) have estimated from the G+C content of cucurbit rDNA, that there should be about 20-30 sites for Msp I within a rRNA gene and a similar figure was obtained for flax rDNA by Ellis *et al* (1983). Because the ribosomal genes are highly conserved, there is no reason to believe that the pea ribosomal gene should differ widely from the cucurbit and flax values.

The multiplicity of sequences which cross hybridized to the pHA 1 probe (eg figure 17b, eg tracks c and d), complicated the analysis of the methylation of the pea ribosomal genes. The total lack of hybridization of the probe to the genomic Hpa II fragments of less than 1kb (figure 18b, tracks b, d, f and g) in positions equivalent to those observed in the marker pHA 1 tracks (figure 18b, track j), clearly indicated that no completely unmethylated rRNA genes were present in either cotyledon or leaf tissue. The weak hybridization of the probe to bands of less than 1kb in the Msp I tracks (figure 18b, tracks a, c, e and h), indicated that there were relatively few rRNA genes which contained a mixture of only unmethylated and partially methylated ($C^{m}CGG$) sites.

The far greater hybridization of the probe to the larger Msp I

fragments (1kb to 10kb) (figure 17b: tracks d, h, l, p and t), implied that the majority of the ribosomal repeat units contained a mixture of fully methylated ($^{m}C^{m}CGG$) and partially methylated ($C^{m}CGG$) sites, whilst a minority, ie those detected by Hpa II (figure 17b, tracks e, i, m, q and u), contained a mixture of partially or fully methylated sites and some unmethylated sites.

The overall similarity in size and relative intensities within a track, of the Msp I and Hpa II hybridization bands indicated a) that the ribosomal repeats were divided between a highly methylated region or 'domain' and a less highly methylated region; b) that certain sites, in association with the rDNA, have a greater probability of being undermethylated (ie unmethylated in the repeats from the less highly methylated domain and partially methylated in the ribosomal repeats in the highly methylated domain). However, two sites - those cleaved by Msp I to produce the 2.83 and 3.45kb bands, could exist in a partially methylated form but were not detected in the unmethylated form (figure 17b, eg tracks d and e).

The three levels of hybridization of the ribosomal probe to the Msp I digests (eg figure 17b, track d) is potentially a very interesting feature of these blots. It is possible to envisage two situations which would result in the hybridization patterns observed in pea. The first situation would be one where, overall, the rDNA comprises fully methylated sites with a few specific sites having a greater probability of being undermethylated, ie either partially methylated (Msp I sensitive) or unmethylated (Msp I and Hpa II sensitive). These sites would perhaps have a lower affinity for the

methylase.

The second situation would be one where the rDNA is divided into two unequal regions or 'domains'. The minor 'domain' contains rDNA which comprises a mixture of many partially methylated sites and a few unmethylated sites whilst the major 'domain' contains rDNA which comprises a mixture of many fully methylated sites and a few partially methylated sites. In the latter situation, the hybridization fragments of the intermediate intensity range (0.93 to 2.09kb) and the small fragments (<0.93kb) which hybridize only weakly to the probe, would be derived from those fragments which were detected as distinct bands in the Hpa II digests. Thus the weakly hybridizing fraction would represent the few repeat units in which all CCGG sites can be cleaved by Msp I (ie a mixture of unmethylated and partially methylated sites), the intermediate fraction would comprise those ribosomal repeat units with many partially methylated and unmethylated sites and a few fully methylated sites and finally, the strongly hybridizing Msp I fraction (2.29 to 10.0kb) would comprise rDNA with many fully methylated sites and a few partially methylated sites.

One way to test this second hypothesis would be to restrict cotyledon genomic DNA with Hpa II, to run it on an agarose gel and to isolate the highly methylated fraction DNA (>15kb). This highly methylated fraction when subsequently restricted with Msp I, electrophoresed, Southern blotted and hybridized to pHA 1 should then produce a band pattern in the 2.29-10.0kb range but no hybridization should be detected to rDNA fragments of less than 2.29kb.

Keshet and Cedar (1983) and Busslinger *et al* (1983a) have both

reported that the specificity of cleavage of Msp I was affected by certain nucleotide sequences immediately 5' or 3' to the CCGG sequence. In particular, these inhibitory sequences were 5'-CCGGCC-3' and 5'-GGCCGG-3'. In these two sequences, contrary to its normal specification, Msp I is inhibited by the methylation of the internal cytosine of its recognition sequence.

In order to investigate whether such sequences could be partly responsible for the large overall size of Msp I derived rDNA fragments, the hybridization of pHA 1 to Hae III digested genomic DNA was examined. Hae III has the recognition sequence GGCC and will also cleave GGC^{m}C but is unable to cut GG^{m}CC (Mann and Smith, 1977; Roberts, 1981; McClelland, 1983; McClelland and Nelson, 1985). Therefore where there are overlapping Hae III and Msp I sites, neither enzyme will be able to cleave GG^{m}CCGG but Hae III can cut GGC^{m}CGG whilst Msp I (and Hpa II) cannot (Busslinger *et al* ,1983).

The smallness of the size of Hae III rDNA hybridization fragments (figure 17b, tracks c, g, k, o and s) therefore indicated: a) that the sequence GG^{m}CC did not occur very often in the ribosomal repeat units (which is as expected because over 90% of 5mC in plant DNA occurs in 5'-CG-3' sequences and not 5'-GC-3' sequences); b) that GGC^{m}C may or may not exist; c) that GGC^{m}CGG could not exist to any great extent because if it did the Hpa II tracks would have to be of similar intensities to those of the Msp I tracks. Therefore, if the greater intensity of the Msp I tracks is due to the presence of a large number of doubly methylated sites (^{m}C^{m}CGG) and a few partially methylated sites (C^{m}CGG), there can be very few overlapping Msp I/Hae

III sequences because the methylation of both cytosines would also inhibit Hae III.

Whilst these results do not preclude the existence of one or more GGC^{m}CGG sequences, it must be concluded that they do not occur to any great extent in the pea ribosomal repeat units.

It was also not possible from these results to determine the location of the highly methylated and less highly methylated fractions of rDNA. It is not known whether all the repeat units, which contain a higher proportion of unmethylated and/or partially methylated sites, are occurring at a similar chromosomal position. However, most ribosomal repeat units must occur in areas of fairly highly methylated DNA because the average Hpa II rDNA fragment size is greater then 20kb (ie larger than the length of one ribosomal RNA repeat). Detailed restriction mapping of the cloned rRNA gene is needed to identify the locations of the undermethylated Msp I/Hpa II sites. It may then be possible to relate areas of hypomethlyation with distinct sections of the repeat units.

4:2:2 The methylation of the CCGG sequence in leaf rDNA

An organ specific difference in the extent of rDNA methylation, as detected by Hpa II has been noted earlier. If the above second hypothesis, concerning the three levels of hybridization is true, then the appearance of the leaf + Msp I hybridization pattern (figure 17b, track t) implies that this organ-specific difference in the Hpa II tracks is due to the partial methylation of previously unmethylated sites plus some double methylation of formerly partially methylated

sites.

The rDNA hybridization pattern obtained for pea leaf DNA plus Hpa II, was very similar to that described by Scott *et al* (1984) for marrow and pumpkin fruit tissue and turnip tap root and it was also similar to that described by Uchimiya *et al* (1982) for mature tobacco leaves. Some similarity in the hybridization patterns obtained would be expected because the genes for ribosomal RNA are highly conserved (Mandal, 1984; Leaver, 1979). However, Ellis *et al* (1983) when using flax leaves, obtained a hybridization pattern of Hpa II rDNA fragments (20kb - 1.6kb) and Msp I rDNA fragments (about 5.0 - 1.5kb), which appeared to indicate that the ribosomal genes in flax were less methylated (at both cytosines of the CCGG sequence) than the pea, marrow, pumpkin, turnip and tobacco. It is however, possible that these apparent differences in the degree of methylation arise from differences in the ribosomal repeat number. The number of ribosomal RNA genes per haploid genome is as follows: flax - 990, pea - 3,900, marrow - 4,900 and pumpkin - 4,900 (Long and Dawid, 1980). Therefore, the differences in the extent of methylation may be related to ribosomal gene redundancy.

The extent of rDNA methylation is known to vary in different organisms. In invertebrates (eg *Drosophila*), rDNA is unmethylated (Rae and Steele, 1979), in amphibia it is found to be highly methylated (Bird and Taggart, 1980), whilst in vertebrates, eg mouse and rat, a combination of methylated and unmethylated rDNA is found (Bird *et al* , 1981; Kunnath and Locker, 1982b). The number of ribosomal genes per haploid genome for these organisms is as follows:

mouse - 100; *Drosophila* - 120; rat - 150; *Xenopus* - 600 (Long and Dawid, 1980). When the very large number of rRNA genes in higher plants is taken into consideration, eg 3900 rRNA genes per haploid complement in pea (Ingle and Sinclair, 1972; Long and Dawid, 1980), it is conceivable that only the few Hpa II sensitive rDNA sequences, which are presumed to be less highly methylated, are transcribed (Scott *et al*, 1984). Timmis *et al* (1972) observed that the transcription of less than 5% of pea rRNA genes was actually needed in actively growing pea roots. Ingle *et al* (1975) have proposed that in higher plants, the large numbers of rRNA genes is a compensatory mechanism for a reduced transcriptional activity (eg, due to methylation). In cucurbits, the percentage composition (C+G) of total DNA is 32.2%, whilst that of ribosomal DNA is 46.0%. Similarly, the percentage 5mC of total cytosine is 20.7% for total DNA and 36.4% for rDNA. Therefore the increased number of ribosomal genes in higher plant DNA, could be a consequence of the high degree of cytosine methylation present in the rDNA (Scott *et al*, 1984).

4:2:3 The hypomethylation of cotyledon rDNA

An organ-specific hypomethylation of the pea cotyledon rDNA has already been noted. The mature leaves are photosynthesising and are biochemically active, therefore a net protein synthesis (and consequential ribosome requirement), to account for enzyme turnover and membrane synthesis would be expected. A similar protein turnover level would perhaps be expected in the pumpkin and marrow fruit, turnip tap root and tobacco leaf. However in the pea cotyledon, a totally

different situation exists. This is an organ responsible for the synthesis of large amounts of storage protein (see figure 1). The major accumulation of vicilin and legumin occurs during the middle stages of cotyledon development (12 d.a.f. - 20 d.a.f.) although legumin synthesis has been detected as early as 10 d.a.f. and vicilin 9 d.a.f. (see section 1:2:2).

There was no obvious apparent change in the CG hypomethylation observed during the development of pea cotyledons (figure 17b, tracks e, i, m and q). The biological significance of endoreduplication is not known but it was initially thought that it might include the selective amplification of genes required for successful seed development. One suggestion was that it enabled the storage protein genes to be amplified but this was later disproved by Millerd and Whitfeld (1973). Another suggestion has been, that in a situation analogous to the amphibian oocyte, the rDNA would be selectively amplified. Ingle and Sinclair (1972) were not able to detect any gross amplification (or deletion) of the ribosomal genes during the various stages of plant development. However, Cullis and Davies (1975) did detect the selective amplification of the cotyledon rDNA in one pea variety but they were unable to detect any rDNA amplification in the cotyledon cells of three other varieties. Also, the one pea variety investigated, which apparently did amplify its rDNA, was shown to have a much lower proportion of rDNA in its meristematic cells and amplification of the rDNA increased the amount to a level comparable to that observed in other pea varieties, ie it is possible that the

low normal level of rDNA, whilst being adequate for most cellular requirements (eg those of the meristematic region of the root tip), was inadequate to satisfy the requirements for rDNA in the specialised cotyledon cells.

In the Feltham First cotyledon rDNA amplification could potentially take several forms. There could be either the amplification of the total rDNA or a selected amplification of the less highly methylated fraction of the rDNA. Both of these amplification events could have occurred under a situation where either a) the maintenance methylase was active and the methylation pattern was inherited, or b) where the action of the methylase was blocked and unmethylated rDNA copies would result. If any of these amplification events had occurred in Feltham First, they would show up as irregularities in the hybridization signal intensity for the Msp I and Hpa II tracks (figure 17b, tracks d and e, h and j, l and m, p and q). No such irregularity is observed and it must be concluded that selective amplification of the Feltham First rDNA does not occur during the endoreduplication phase of cotyledon development.

4:2:4 Changes in the CCGG methylation pattern in relation to plant development

From these results it was not possible to determine at what stage of development or under what stimuli the methylation of rDNA alters. There must be at least two stages in the life cycle of a plant where the rDNA undergoes a drastic change. Assuming that the situation observed in the leaf, reflects that of the rDNA in the plant in

general, then at some stages of the plant development there has to be an inhibition of the maintenance methylase to produce the hypomethylation observed in the developing cotyledon rDNA. It is not known, to what extent the rDNA in the pollen, ovules or embryonic axes is methylated. Similarily, because the mature cotyledon rDNA is still hypomethylated, the *de novo* methylation of the rDNA (necessary in order to produce the highly methylated leaf rDNA) must occur after the germination of the pea seed.

The maturation of the pea seed, at the end of the storage protein synthesis phase, did not appear to be the stimulus for rDNA *de novo* methylation. This may be either because there is a continued requirement for ribosomes and rDNA (perhaps in preparation for the demands of germination) but it is more likely to reflect an inability to alter any methylation pattern, due to the lack of DNA replication. (Maximum DNA content of the Feltham First seed is achieved 19-20 d.a.f. (Boulter, 1981)). If this is so, then it would be interesting to monitor the extent of rDNA methylation in the tissues of the germinating pea seed.

The *de novo* methylation of rDNA could be a gradual process, occurring over many cell divisions or it could be a major methylation event, occurring at a precisely determined time. Once established, the new pattern would maintain the gene in its new state of altered transcriptional competence during subsequent cell replications (Jaenisch and Jahner, 1984).

4:2:5 Methylation of pea rDNA at Hha I sites

The use of Msp I and Hpa II can only examine the extent of methylation at a particular subset of CG dinucleotides and whilst there is no evidence to indicate that this subset of CG dinucleotides is not representatives of the surrounding DNA (Bird, *et al* , 1985) it is useful to also investigate a different subset of CG dinucleotides using the restriction enzyme Hha I.

Hha I has the recognition sequence 5'-GCGC-3' but is unable to cut the sequence when either of the cytosines is methylated (eg McClelland and Nelson, 1985). Unfortunately there is no isoschizomer with a complementary specificity of cleavage (ie able to cut the methylated sequence) and in the absence of pea rDNA sequence data or detailed restiction mapping, this enzyme could only be used to monitor possible developmentally related or organ-specific changes, in the extent of methylation of the Hha I sites associated with the ribosomal RNA genes.

No rDNA Hha I fragments were observed to disappear during pea cotyledon development (figure 17b: tracks b,f,j, and n) and overall hybridization to discrete leaf rDNA Hha I fragments (track r) was less than had been observed for the cotyledon tracks (ie the Hha I sites in the cotyledon rDNA were hypomethylated with respect to those in the leaf rDNA). The similarity between this overall increased methylation of the Hha I sites in leaf rDNA and that previously noted for the Msp I and Hpa II sites in leaf rDNA, supports the idea that although these restriction enzymes monitor only a small proportion of the total number of genomic CG dinucleotides, those that they do monitor, are

representative of the CG dinucleotides in the surrounding DNA.

4:2:6 Methylation and transcription of the ribosomal genes

Even after allowing for the fact that the pea cotyledon rDNA is hypomethylated with respect to the leaf rDNA, it should be noted that the cotyledon rDNA is still considerably methylated. Isolated cotyledon nuclei are known to be more transcriptionally active than leaf nuclei and have been shown to produce more ribosomal transcripts. This implies that the ribosomal genes must be transcribed despite being highly methylated.

The *Xenopus laevis* rDNA is transcribed despite the fact that is methylated (Bird and Southern, 1978). However, the transcription may depend on an undermethylated region within the non-transcribed spacer between the ribosomal genes (Lindahl, 1981). Bird and Southern (1978) have demonstrated the presence of an Ava I/Hpa II 'hotspot' in the non-transcribed spacer between the repeat units in erythrocytic rDNA and a Hha I 'hotspot' in the 28S rRNA coding region. Both of these sites were undermethylated, despite being set amongst sequences which were fully methylated.

It is interesting to note that the spacer DNA in *Xenopus laevis* sperm DNA is highly methylated and is not normally transcribed, and yet it was efficiently transcribed after injection into *Xenopus* oocytes, although no concomitant loss of methylation was detected (Macleod and Bird, 1984). Also, whilst in *X. laevis* the spacer region is undermethylated in somatic rDNA but is methytlated in sperm rDNA, in *X borealis* the spacer region is hypomethylated in both somatic and

sperm rDNA. However, this undermethylation was insufficient to ensure either an active chromatin structure or efficient transcription of the *borealis* rDNA, in *borealis-laevis* hybrid tadpoles (Macleod and Bird, 1982).

This resistance to DNAse I, shown by the hypomethylated regions of *X. borealis* ,is in contrast to mouse rDNA, where it was demonstrated that only the hypomethylated rDNA was sensitive to DNAse I (Bird *et al* , 1981). La Volpe *et al* (1982) demonstrated that the hypomethylated Hpa II and Ava I sites, in the non-transcribed spacer region of *Xenopus laevis* and *Xenopus borealis* , were located in tandemly repeated conserved sequences 60bp long (NTS-60). Whilst they were unable to detect any apparent coupling of demethylation between the conserved sequences, they were able to detect coupling within the NTS-60, such that demethylation of the Hpa II site was usually associated with demethylation of the Ava I site, although demethylation of the Ava I site was not always accompanied by demethylation of the Hpa II site (ie hypomethylation was coupled in a polarized manner within a copy of NTS-60). The authors suggested that the binding of a protein to NTS-60, prevented the methylation of the central Ava I site whilst only partially interfering with the methylation at the Hpa II site, which was 16bp to the left of the Ava I site. Thus the hypomethylated areas were described as the 'footprints' of a protein that interacted with the NTS-60. Riggs and Jones (1983) have also proposed that 'determinator' proteins can influence the methylation pattern in a manner which is then somatically heritable, even in the absence of the determinator

protein.

In the developing rat liver, it has been demonstrated that the level of methylation of the rDNA increased during the liver development, from being mostly unmethylated at 14 days gestation to being 30% methylated by 18 days gestation (and it remained at this level in the adult)(Kunnath and Locker, 1982b). Such a trend was not apparent during the development of the pea cotyledon. Whilst there was a pronounced organ-specific difference in the extent of methylation, there was no apparent developmentally, correlated, incresed rDNA methylation in the cotyledon.

4:2:7 C-X-G methylation of pea rDNA

The smaller average size obtained for Bst NI derived rDNA hybridization fragments (eg figure 19b, track b) compared with those obtained after Eco RII digestion (eg figure 19b, track c), clearly demonstrated the marked methylation of the $CC\frac{A}{T}GG$ sequences in pea rDNA.

Two of the Bam HI derived genomic fragments, which were detected by the pHA 1 probe (2.92 and 1.27kb)(figure 20b eg track a), were larger than the equivalent pHA I + Bam HI fragments 2.60 and 1.22kb (figure 20b, track 1). This may reflect a slight variation in the ribosomal gene of the variety 'Feltham First' compared with 'Alaska' from which pHA I was derived (Cuellar, 1981). These two fragments do not correspond the the non-transcribed spacer (NTS) region of the ribosomal repeat unit (figure 16), where one might expect heterogeneity (Mandal, 1984)

The sum of the Bam HI fragments, detected by the pHA 1 probe (2.92 + 1.27 + 0.66) is 4.85kb. Therefore, the 4.95kb band in the Bam HI doublet (eg track a) may in fact indicate that the majority of the pea ribosomal gene copies contain Bam HI sites which are methylated in the regions coding for the structural genes. Bam HI has the recognition sequence 5'-GGATCC-3' but is unable to cleave the sequence 5'-GGAT^{m}CC-3' (eg McClelland and Nelson, 1985). Therefore, on average, in one in four cases this enzyme will be monitoring ^{m}CCG or ^{m}C^{m}CG methylation and the presence of this genomic Bam HI doublet, supports the evidence provided by the Msp I hybridization pattern, that in the ribosomal genes, the CCG trincucleotide can be found methylated at both cytosines. Gerlach and Bedbrook (1979) have reported that incomplete methylation (21-40%) occurs at two Bam HI sites in each of two different classes of barley rDNA gene repeats and Ellis *et al* (1983) have also reported the methylation of a Bam HI site in the flax ribosomal gene.

From the information available at present, it is not possible to identify the location of the larger 5.55kb Bam HI doublet band (eg track a) but it would appear that there must be one or more additional Bam HI sites in the Feltham First ribosomal repeat units. Microheterogeneity has been reported for the non-transcribed spacer regions and this is usually charactrised as a variation in the absolute length of the rDNA repeat unit, both within and between individuals. This length heterogeneity is due to the number of short repeated sequences in the NTS (Mandal, 1984). Length variation of the spacer region was also reported for the wheat and barley rDNA genes

(Gerlach and Bedbrook, 1979).

The Feltham First, 9.3kb Hind III monomer rDNA repeat unit length (figure 19b eg track a), appeared to be similar to that of pHA1 (track 1), despite the apparent anomolies in the sizes of the Bam HI rDNA fragments. The rDNA repeat unit length if other plant rDNA genes are 7.8kb for soybean (Varsanyi-Breiner *et al* , 1979), 8.8 and 9.0kb for wheat and 9.9 and 9.0kb for barley (Gerlach and Bedbrook, 1979), 7.8kb in rice (Oono and Sugiura, 1980) and 9.8kb in broad bean (Yakura and Tanifuji, 1981). It is possible that what appeared to be a broad, pea genomic Hind III rDNA band, is in fact two or more fragments, which were not resolved by the gel due to their similar mobilities. Length heterogeneity has been observed for wheat, barley and rice rRNA genes but not for soybean.

The restrictions with Bst NI and Eco RII indicated a) that all ribosomal Bam HI fragments contained $CC^{A}_{T}GG$ sequences, b) that these sites were cleaved by Bst NI but as far as could be detected, were not cleaved by Eco RII. When assessing these two experiments (figures 19 and 20) it is necessary to compare the activity of Eco RII on the SV40 internal control (figures 19c and 20c: tracks c,f,i, and k) with its activity on genomic DNA (figures 19b and 20b: tracks c, f, i and k). The Eco RII was shown to partially digest the SV40 but when the filters were probed with pHA 1, the Eco RII bands remained identical to those produced by Bam HI or Hind III. This demonstated that the Eco RII was functioning during the digestion and therefore its complete failure to cleave the rDNA must be at least partly attributable to the methylation of the DNA at some sequences ($C^{m}C^{A}_{T}GG$).

In order to quanitify the extent of the methylation at these sequences in genomic DNA it is essential to demonstrate Eco RII's ability to completely cleave unmethylated DNA. Eco RII from several different manufactureres was assessed, with and without B.S.A. and spermidine, which are reported to stabilize the DNA by binding to the major groove. The effect of reaction volume on Eco RII's ability to cleave the plasmid DNA was also investigated. In none of the reaction conditions used was it possible to demonstrate a similar digestion of unmethylated DNA, by both Bst NI and Eco RII (figure 21). In the absence of such a demonstration, this investigation into the methylation of $C^{A}_{T}G$ sequences, in pea genomic DNA, had to be stopped.

4:3 Methylation of the Pea Storage Protein Gene Families

4:3:1 Choice of hybridization conditions

The choice of hybridization system was varied according to the aims of a particular experiment. Generally hybridiizations were according to the Scheicher and Schull procedure, where 5 x Denhardts is used and the filters are washed in 0.1 x SSC; o.1% SDS at 50°C. At concentrations of less than 5 x Denhardts, the single stranded DNA probe is prevented from hybridizing to the nitrocellulose filter. At concentrations >5 x Denhardts, DNA hybridization may be hindered (Barinaga *et al* , 1981). This higher concentration of Denhardts (10x) was used for the Leg B flanking sequence experiments, in an attempt to reduce non-specific hybridization.

In the legumin experiments, the filters which were probed with

whole legumin gene probes, were washed at higher temperature than the experiments using isolated gene fragments (62 - 65°C compared with 50°C) but despite this, neither the relative intensities nor the number of hybridization bands appeared to be significantly altered by the high temperature washing.

The heparin system was also used during this work. This system was preferred because of its simple protocol. Heparin reduces non-specific adsorption of nuclei acid probes to nitrocellulose filters (Singh and Jones, 1984; Yuan *et al* , 1975). Pure double stranded DNA does not bind to nitrocellulose but the adsorption of DNA-protein complexes to the filter, is frequently the source of a high background signal. Heparin has been shown to inhibit certain specific enzymatic functions of some DNA binding proteins eg RNA polymerase (Zillig *et al* , 1970) and restriction endonucleases (Yuan *et al* , 1975), where inhibition was only observed if heparin was added to the reaction before but not after the DNA. Thus, in some ways, heparin may be considered to be a DNA analogue.

Vicilin cDNAs proved more difficult to hybridize than legumin cDNAs. Domoney and Casey (1985) have stated that when filters which had been hybridized with the two vicilin probes were washed at 50°C (compared with 65°C), the number of fragments hybridizing to the probe was not increased but the prominence of the fragments corresponding to the other vicilin probe was increased. During this work, when filters which had been hybridized with the vicilin cDNA probes were washed at higher temperatures (eg 65°C), substantial loss of signal was found. Therefore, in the vicilin experiments presented here the filters have

only been washed at 50°C.

4:3:2 Methylation of the legumin gene family

When a legumin cDNA probe was hybridized to Msp I and Hpa II digested genomic DNA (figure 22b), two distinct organ-specific, developmentally correlated demethylation events were observed. Thus: 1) a 6.5kb Hpa II fragment, detected fairly strongly in leaf (track l) and 10 d.a.f. cotyledon DNA (track c) and fairly weakly in the 15d.a.f. cotyledon DNA (track f) was not detected in 21 d.a.f. cotyledon DNA (track i); 2) an additional 4.19kb band was detected in the 15 d.a.f. cotyledon DNA (track f) and this is thought to have arisen by a CCGG sequence within the 6.5kb fragment, changing its methylation from $C^{m}CGG$ (Hpa II resistant) to CCGG (Hpa II sensitive). Between days 15 and 21 a further hypomethylation event must have occurred because the 4.19kb fragment was no longer detected. Instead the intensity of the 3.74kb band had increased. This is indicated in table 5b.

Before the possible origins of the Msp I and Hpa II fragments detected by the cDNA probe can be discussed (see section 4:3:2:6), it is necessary to describe the methylation at the specific CCGG sites of two legumin genes, Leg A and Leg B and also in their flanking regions.

4:3:2:2 Methylation of CCGG sequences in Leg A and Leg B

When an organ-specific demethylation event was correlated with the expression of the legumin genes in the pea cotyledon, it was decided to try and identify the site/sites concerned and thereby

relate the change in methylation to a specific legumin gene. The easiest way to achieve this aim was to restriction map the cloned legumin genes λLeg 1 and λLeg 2, identify all Msp I sites and to then use isolated fragments of this cloned DNA as specific probes, to monitor the genomic methylation at the specific CCGG sites contained in their sequence, during cotyledon development. The legumin genes are not actively transcribed in the leaf and the leaf digest therefore provided a useful control of the methylation status of a particular site, in the unexpressed gene.

In Leg A, sites M2 (in the second exon) and M3 (in the fourth exon) were both shown to be partially methylated, ie in both cases the genomic fragment was only partially cleaved by Hpa II (eg track c in figures 24b and 25b). This indicated that the sites were being detected in the CCGG and $C^{m}CGG$ forms. The M3 site was found to be on average 42% methylated whilst M2 was on average 25% methylated (table 6). At neither of these sites was there observed a demethylation which could account for the methylation change detected by the legumin cDNA probe. However, there were indications that the cotyledon M2 site became slightly less methylated as development proceeded (from 25.8%) and that the leaf Leg A gene was very slightly more methylated at both M2 and M3 than the cotyledon copies of Leg A. The M1 site, in the 5' flanking region of Leg A, was always detected in the unmethylated form (figure 26, tracks c, f, i and l).

Thus, in Leg A, an increasing level of methylation was detected, from being completely unmethylated in the 5' flanking region, to being 25% methylated in the 5' coding sequence and about 42% methylated in

the 3' coding sequence.

4:3:2:3 Comparison of the methylation of Leg A and Leg B

Restriction mapping had indicated that the coding sequence of three cloned legumin genes (Leg A, Leg B and Leg C) was very similar (figure 27). Leg B was found to contain sites M2 and M3 but the M1 site in the 5' flanking region was absent. When the extent of methylation at Leg B's sites M2 and M3 was analysed, the sites were again detected in the CCGG and $C^{m}CGG$ forms (eg track f in figures 28a and 28b). Leg B showed the same trend that had been indicated by the Leg A results in that the methylation in the coding sequence increased in the 5'⟶3' direction. However, the methylation at Leg B's site M2 was found to be less than that detected in Leg A (in cotyledon and leaf DNA, the average methylation at Leg B's M2 site was 14.1% compared with 24.7% in Leg A (table 6). There was also an indication that the imbibed cotyledon had a lower methylation at the two sites (6.1% at M2 and 32.6% at M3 (table 6)).

These trends were emphasised when the entire legumin genes were used as probes. The lower methylation of the M2 site in Leg B can be seen when the relative intensity of the 1.09 residual Taq 1 fragment in the Hpa II tracks was compared (eg track c in figures 34 and 35). The Southern blot to which the Leg B probe was hybridized, had previously been hybridized to the Leg A 5' flanking region (figure 26). Therefore, this blot provided a valuable check that the lower methylation at Leg B's M2 site (figure 28a) was real and was not an artifact created by a different batch of enzyme or DNA preparation

etc.

Another major feature illustrated in these two autoradiographs (figures 34 and 35) was the marked difference in the specificity of hybridization of the 5' flanking regions of the two genes. The 5' flanking region of Leg A (figure 34), hybridized to four distinct Taq I fragments (eg track a) of 1.57, 1.49, 1.40 and 1.25kb (see also figure 26). Of these, the 1.49kb fragment was cleaved by both Msp I and Hpa II (figures 34 and 26: eg tracks b and c). However, the Leg B probe hybridized to only one Taq I fragment, the 1.40kb band (figure 35 track a). Therefore, the Leg A 5' flanking region can cross hybridize to the Taq I, Leg B, 5' flanking region but the Leg B probe is unable to detect either the Leg A, Taq I, 5' flanking regions or the other two Taq I fragments, which were thought to have come from other legumin genes.

Both legumin gene probes detected the additional 2.75kb Taq I fragment (figure 34 and 35, eg track a). This Taq I fragment was detected by the 1.09kb Taq I probe of Leg A (figure 25b, track a) and Leg B (figure 28a, track a) and was also detected by the Leg A, 5' flanking region probe (figure 26, track a). These bands were thought to arise from the cross-hybridization of the probe with other legumin genes and were not thought to represent unrestricted Leg A or Leg B fragments. Taq I has a recognition sequence TCGA, but the larger fragments were not thought to arise from Taq I partial digestion of genomic DNA because Taq I is known to restrict TCGA sequences containing methylated cytosine (Streek, 1981; Gruenbaum *et al*, 1981a). Taq I, however, is inhibited by adenine methylation, ($TCG^{m}A$)

(McClelland, 1983a; Roberts, 1983) but methylated adenine has not been reported as being present in plant DNA to any great extent (Vanyushin *et al* , 1971). Therefore, it is thought that these high molecular weight Taq I fragments were due to the weak cross-hybridization of the probe to other legumin gene(s) and that such hybridization was detectable despite the fairly stringent washing conditions used (0.1 x SSC; 0.1% SDS; 50°C).

The final point to be made about the methylation of Leg A and Leg B is how relatively unmethylated they were, when compared to the ribosomal RNA genes. The rDNA had many fully and partially methylated sites and a few unmethylated sites. In contrast to this, some copies of Leg A and Leg B contained partially methylated sites (C^mCGG) but no completely methylated ($^mC^mCGG$) sites were detected.

4:3:2:4 Methylation of the 3' flanking sequence to Leg A

Because the Msp I site M1 in Leg A was completely unmethylated (ie always cleaved by Hpa II), this implied that if the Msp I/Hpa II site which during the development of the pea cotyledon became hypomethylated was associated with Leg A, then it had to be 3' of this gene.

The next Msp I site 3' of Leg A was found to be in a 4.17kb Hind III fragment. Msp I was shown to partially cleave this genomic fragment (the 3.40kb cleaved fragment is indicated in figure 30, tracks b, e, h and k) whilst Hpa II was unable to cleave the genomic fragment (tracks c, f, i and l). This indicated that this site had

been detected in the $^{m}C^{m}CGG$ (Msp I and Hpa II resistant) and $C^{m}CGG$ (Msp I sensitive, Hpa II resistant) forms. The external cytosine was thought to be methylated in 40-50% of the copies of this genomic fragment.

Figure 30 also indicated that another Hind III fragment (2.90kb) (track a) was completely unmethylated at all stages of cotyledon development and in the leaf digest. Following both Msp I and Hpa II digestion (eg tracks b and c) two new fragments of 2.32 and 0.60kb were observed. It is not known whether this 2.90 kb Hind III fragment is associated with the flanking sequence of another legumin gene.

4:3:2:5 Methylation of the Leg B flanking sequences

The 5' flanking sequence of Leg B was used to probe the Hind III digested genomic DNA. It was not known how far upstream from the Bam HI site (used to clone pDUB 25) the next Hind III site was, neither was it known whether an Msp I site was contained in this region.

Two genomic Hind III bands (4.67 and 3.29kb) were cleaved by Msp I (figure 31, tracks g and h). The 3.29kb was almost completely cleaved by both Msp I and Hpa II, indicating that the majority of the copies of this genomic fragment were being detected in the unmethylated state. However, in the leaf digests (tracks b and c), the 3.29kb fragment had been only partially cleaved, indicating that this site was more highly methylated in leaf DNA. The 4.29kb band was partially cleaved by Msp I (eg track h) and was uncleaved by Hpa II (eg track i), which indicated that about half of the copies of this fragment was detected in the $C^{m}CGG$ form and half in the $^{m}C^{m}CGG$ form.

It is not known which of these two Hind III bands (if either) represents the 5' flanking region to Leg B. If the methylation of the Leg B flanking sequence resembles that observed for Leg A, then it might be expected that the 3.29kb Hind III band which was only very slightly methylated, represents the 5' flanking region to Leg B. If this is so then it would appear that in leaf DNA, a small proportion of the copies of this sequence (about 20%) are fully methylated and are resistant to both MspI and Hpa II.

The methylation of two Msp I sites 3' to the Leg B coding sequence was also investigated. These sites were contained in a 3.75kb Eco RI fragement and a 0.8kb Bam HI fragment. Both of these fragments were shown to be fully methylated ie resistant to Msp I and Hpa II (eg compare the intensity of the 3.75kb band in tracks a, b and c in figure 32 and the intensity of the 0.8kb fragment in tracks a, b and c of figure 33).

The interpretation of figure 32 was complicated by the wide range of Eco RI fragments which non-specifically hybridized to the probe, despite the stringent hybridization and washing procedures used. There are two possible reasons for this. First, the probe was prepared from a miniprep as there was insufficient time to do a maxi-prep to obtain the purified plasmid. Therefore the DNA probe was not as pure as it might have been. Second, the size of the probe was 3.75 kb which is fairly large for a probe, particularly when one is investigating non-protein coding DNA. However, there was insufficient DNA obtained from the miniprep, to do a detailed restriction mapping and isolation of a smaller probe fragment.

Whilst the 3.75kb genomic fragment was shown to be completely methylated a smaller 2.75kb fragment (figure 32, trakc a), the origin of which is not known, was found to be unmethylated at all stages of cotyledon development (eg tracks a and b). This is a situation similar to that observed for the 3' flanking sequence of Leg A (figure 30) and again it is not known whether this Eco RI fragment is associated with a legumin gene.

In figure 33, whilst it was apparent that CCGG site in the 0.8kb Bam HI fragment was completely methylated, three other features are worth noting. First, after Msp I digestion (tracks b, e, h and k) it was apparent that a 6.60kb Bam HI fragment had been cleaved and new bands were observed at 3.87, 1.25, 0.89 and 0.56kb. However, after Hpa II digestion (tracks c, f and i), although the 6.60kb had been cleaved the only new bands observed were 4.57 and 2.00kb. This possibly indicates that the 6.60kb Bam HI fragment contains one unmethylated site, which can be cleaved by Msp I. The 4.57 and 2.00kb Bam HI + Hpa II fragments must each contain one partially methylated site which can only be cleaved by Msp I (4.57$\rightarrow$3.87 + 0.56 and 2.0$\rightarrow$1.25 + 0.89). Once again, the origin of this 6.60kb Bam HI fragment is not known.

The second feature to note in this autoradiograph is that neither Bam HI site at the end of the 0.8kb fragment appeared to be methylated (Bam HI cannot cut GGAT^{m}CC).

Finally, although the 0.8kb Bam HI band was intense and did not appear to be a partial digest, it should be noted that the 0.8kb Bam HI + Hpa II band had a relatively increased intensity (for example compare this band in tracks c and d). It was also noted that in the

Bam HI + Hpa II tracks there was a decreased intensity of hybridization to fragments of 3.00kb. It is thought that this reflects the enzyme digestion conditions used. Whenever double digestion experiments were performed, the first enzyme was always the one with the lowest salt requirement. At the end of the first digestion, the first enzyme was inactivated, the buffer modified and the second restriction performed. Where Taq I and Rsa I have been used, they have the lowest salt requirement and so the genomic DNA was digested with Taq I or Rsa I first and Msp I or Hpa II second. However, when Bam HI, Hind III or Eco RI have been used these have a much higher salt requirement and restrictions have been performed with Msp I/Hpa II first and Bam HI etc. second. The appearance of the three tracks, Bam HI (figure 33, track a), Bam HI plus Msp I (track b), and Bam HI + Hpa II (track c) therefore indicates that the buffer conditions created by the conversion of the Hpa II buffer to a Bam HI buffer were better than those created by the conversion of the Msp I to a Bam HI buffer, or the Bam HI buffer on its own. Any difference between these three buffers are thought to have been very minor. An alternative method for double restricting the DNA would have been to perform the first restriction with Bam HI and to subsequently dilute the reaction volume to reduce the salt to the required level for Msp I and Hpa II. However, this can frequently result in having to load a large volume onto the gel and for this reason the other method was preferred.

4:3:2:6 Identification of Msp I and Hpa II fragments detected by the cDNA probe

In none of these experiments was there an obvious demethylation which correlated with the observed change when the cDNA probe was used. This implies that the site which underwent a demethylation must have been associated with a different legumin gene, ie either Leg C or the as yet uncloned postulated legumin gene 'Leg E'. This view is supported by the difficulty found in assigning the Hpa II fragments detected by the cDNA to the known Hpa II fragments of λLeg 1 and λLeg 2.

The locations of three fragments (4.3, 5.8 and 7.0kb) are indicated in figure 59. however, when one uses the information obtained by the flanking region experiments, it becomes apparent that it is unlikely that either the 6.5 or 5.37kb Hpa II genomic fragments detected by the cDNA probe (figure 22b, trakcs c, e and f), originate from these fragments predicted from restriction mapping of λLeg 2, since the site at the 3' end of both of these predicted fragments can never be cleaved in genomic DNA (figure 32).

Thus, at present, it is not possible to identify most of the Hpa II bands detected in figure 22b (tracks c, f, i and l). Those which can be identified are the 1.17kb band (M1→M2 in Leg A, B and C), the 0.51kb band is thought to represent the fragment from the M3 site in Leg B to the next 3' site and the 1.69kb is thought to be equivalent to the Leg B fragments from site M2 to this same 3' site ie the 1.69kb Hpa II fragment will be detected in the gene copies where the Leg B, M3 site is methylated.

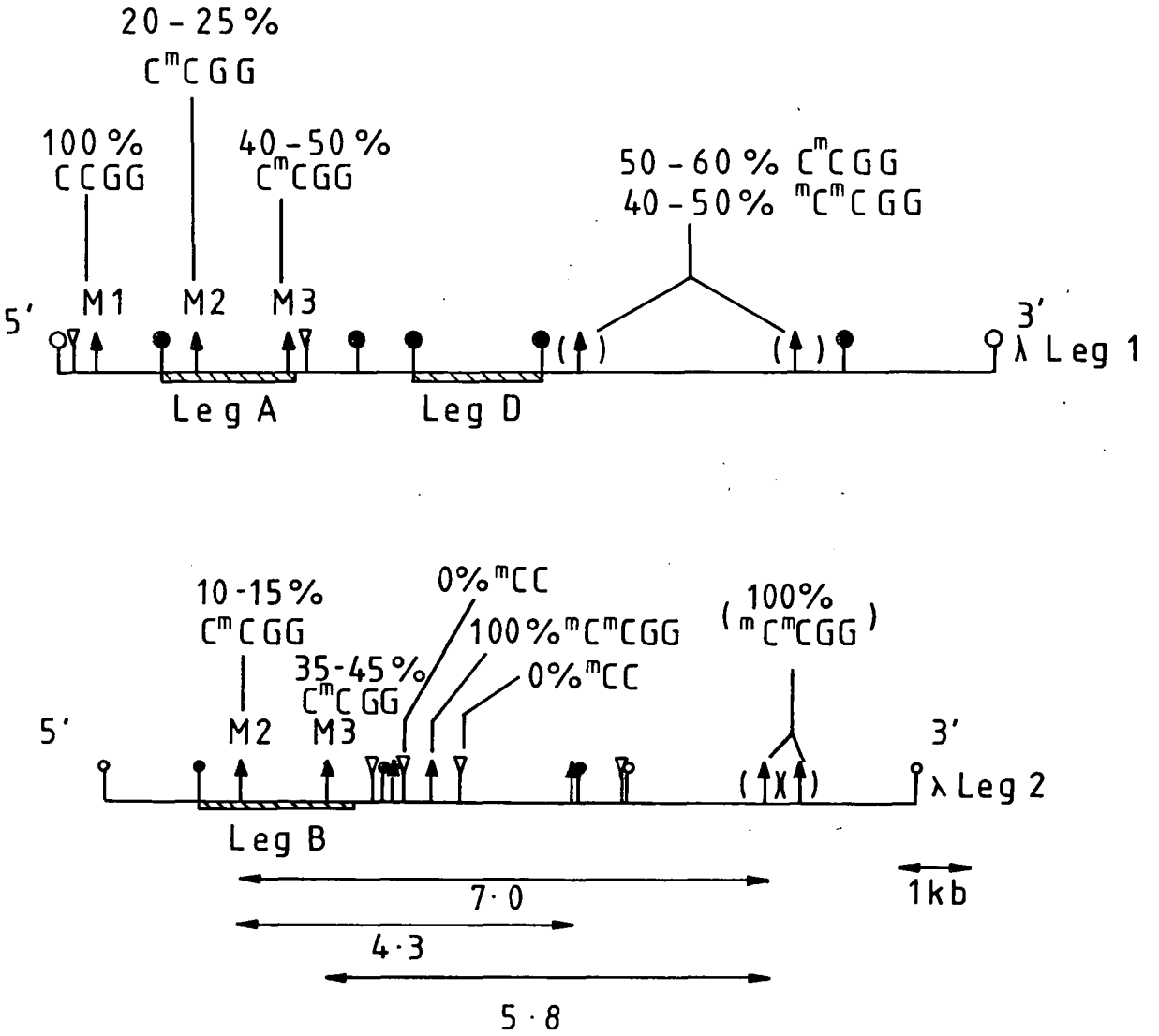


Fig 59. Summary of the genomic methylation at CCGG sites in the λLeg 1 and λLeg 2 sequences

A final attempt was made to identify the legumin gene associated with the demethylation event. In this experiment (figure 30) genomic DNA digested with Eco RI and MspI/Hpa II, was hybridized with a cDNA probe. This enabled the fragments equivalent to the cloned DNA (λ Leg 1, λLeg 2 and λLeg 3) to be identified but it was still not possible to link the demethylation event (ie the disappearance of the 6.24kb Eco RI plus Hpa II fragment from the 15 d.a.f. cotyledon digest (track 1), to any particular legumin gene.

4:3:3 Methylation of the vicilin gene family

Hybridization of a cDNA for the 47,000 -M_r vicilin polypeptide to genomic DNA (figure 38) again indicated that an organ-specific demethylation, in a region of DNA which was associated with the vicilin genes, ie an 11.61kb HpaII fragment was detected in 15 and 21 d.a.f. cotyledon DNA (tracks f and i) which was not detected in leaf or 10 d.a.f. cotyledon DNA (tracks l and c). This 11.61kb fragment (track f) has been derived from the DNA fragments of >20kb in 10 d.a.f. cotyledon DNA (track c), therefore either one or possibly two demethylations must have occurred (depending on whether the 11.61kb fragment came from the end or the middle of a >20kb DNA fragment).

The Msp I and Hae III hybridization pattern showed a different sequence of events. Three major Msp I fragments were detected by the cDNA probe at 11.40, 8.50 and 4.80kb (eg track b). The intensity of the 11.40kb fragment increased during cotyledon development (compare track h with track b), indicating that some sites which were cleaved by Msp I but not by Hpa II 10 d.a.f. (C^mCGG), became resistant to

Msp I cleavage at 21 d.a.f. ($^{m}C^{m}GG$). A similar observation was made following Hae III digestion. An 8.3kb fragment, which was not visible in the leaf digest (track 1) was just detected 10 d.a.f. (track a), was quite distinct in the 15 and 21 d.a.f. digests (tracks d and g) and the appearance of this fragment correlated with the disappearance of the 4.96kb Hae III fragment. This indicated that at least one Hae III recognition sequence became methylated ($GG^{m}CC$ or $GG^{m}C^{m}C$) and thus became resistant to Hae III. Therefore what both the Msp I and Hae III hybridization patterns were indicating was that during cotyledon development, one or more CCG or $C^{m}CG$ trinucleotides (Msp I sensitive and Hae III sensitive) were being methylated to the $^{m}C^{m}CG$ state, which was refractory to Msp I and Hae III. This could not be due to the inhibition of Msp I by a $C^{m}CGG$ sequence a part of an overlapping Hae III/Msp I sequence, $GGC^{m}CGG$ (see section 4:2:1), because such a sequence would be resistant to Msp I and Hpa II but sensitive to Hae III. It could, however, be due to overlapping Hae III/Msp I sites, which became methylated in the following forms, either $GG^{m}CCGG$ or $CC^{m}C^{m}CGG$. Alternatively, the hybridization patterns could indicate methylations at two separate sites, one a Hae III recognition sequence and the other an Msp I recognition sequence.

When the methylation of the genes for the 50,000-M_r vicilin polypeptide was examined (figure 39), a Hpa II hypomethylation event, similar to that detected with the 47,000-M_r probe was observed (note the 11.61kb Hpa II fragment in tracks b and d). The increase in intensity of the 11.40kb Msp I band was also observed (tracks a and c). However, this autoradiograph indicated an additional feature. The

11.40kb cotyledon Msp I band was distinctly smaller than the 11.61kb Hpa II band (compare tracks c and d) but in the leaf + Msp I digest (track e) the size of the Msp I band had increased to 11.61kb (compare tracks d and e). Therefore most of the sites which were cleaved by Msp I in the cotyledon digests (ie $C^{m}CGG$) were methylated in the leaf and the one (or two sites) which became unmethylated (Hpa II sensitive) during cotyledon development were partially methylated, $C^{m}CGG$, (Msp I sensitive; Hpa II resistant) in the leaf DNA.

4:3:4 Comparison of the methylation of the vicilin and legumin gene families

A comparison of the methylation patterns of the vicilin and legumin gene families, indicates several interesting features. The first was that in both gene families, a Hpa II band, which was not evident in the leaf or 10 d.a.f. cotyledon DNA, was apparant 15 d.a.f. in the cotyledon DNA. Thus, in both gene families a demethylation of one or more Hpa II sites has occurred. This demethylation was thought to occur within 4.2kb of a legumin gene but it could be up to 11.61kb away from a vicilin gene.

Second, some Leg A and Leg B gene copies were shown to contain two partially methylated, $C^{m}CGG$, sites in their coding sequences but the 1.17kb Hpa II band detected by the legumin cDNA (figure 22b track c) indicated that some copies of the legumin genes were unmethylated at both of these sites in their sequence. However, in comparison, no unmethylated copies of the 47,000-M_r vicilin gene were detected (figure 38), although the 1.82kb Hpa II and Msp I fragment (figure 39,

eg tracks c and d), may indicate that some copies of the 50,000-M_r genes were unmethylated at two Hpa II sites.

Third, the vicilin genes (both 47,000-M_r and 50,000-M_r) gave much larger Msp I fragments than those observed for the legumin gene family. Whilst the sizes of the vicilin genes' Msp I fragments are not yet known, this large average fragment size may indicate that most of the vicilin genes detected by these two probes, contain several fully methylated sites ($^{m}C^{m}CGG$). This view is supported by the observation that the level of ^{m}CCG and/or $^{m}C^{m}CG$ methylation, of the DNA associated with both vicilin gene families was shown to increase during cotyledon development. Thus a situation may exist in the vicilin gene family where one or two sites become specifically demethylated whilst other sites become additionally methylated (ie hypermethylated), but it is not known whether these two methylation changes occur in association with the same copy of a particular gene. However, because of the sizes of the fragments which are thought to represent the increased methylation (ie 8.3kb for Hae III and 11.40kb for Msp I), it is possible that these sites, which become methylated during cotyledon development, may in fact be sited some distance from the vicilin gene(s).

One vicilin gene in *Pisum sativum* has been mapped to a position very close to the R-locus on chromosome 7 and the legumin genes are sited about 15 map units away from the vicilin genes (J.A. Gatehouse - personal communication). The differences observed in the vicilin and legumin gene families indicated that the two gene families were located in regions of the chromosome which had very different levels

of methylation.

4:3:5 Methylation in relation to the expression of the storage protein genes

A change in the methylation of sites associated with both the vicilin and legumin gene families, during the development of pea cotyledons, was detected using cDNA probes. This demethylation was detectable in 15 d.a.f. but not 10 d.a.f. Other results (not presented) have indicated that no demethylation has occurred by 12 d.a.f.

Evans *et al* (1984) demonstrated that isolated nuclei from 9 d.a.f. cotyledons, produced legumin and vicilin transcripts after a 3 minute transcription labelling period. However, after a longer transcription time (45 minutes) only vicilin transcripts were detected, 9 d.a.f. Thus, whilst the legumin genes were undergoing a low level of expression 9 d.a.f. these legumin transcripts were being post-transcriptional processed. Vicilin transcripts (for both the 47,000-M_r and the 50,000-M_r polypeptide) were again detected at a higher concentration 11 d.a.f. but by 14 d.a.f. the levels were comparable. Therefore, although copies of both vicilin and legumin genes have previously been shown to be undergoing a low level of transcription 9 d.a.f., the demethylation of the genes reported here was not detected at 10 or 12 d.a.f. This implied that the organ specific demethylation observed in cotyledon DNA, may have been a consequence of the the transcription of the storage protein genes rather than the cause of it.

The hypomethylation of genes has frequently been correlated with the expression of eukaryotic and viral genes. However, this basic

question over the relative timing of the demethylation event with respect to the gene's transcription, is one of extreme importance and the answer to which will help to clarify the role of methylation in the control of gene expression.

Other examples of genes in which demethylation proceeds gene activation include the chicken δ-crystallin and vitellogenin genes. Grainger *et al* (1983) followed the methylation pattern of the δ-crystallin genes during the development of the chicken lens and found that one site underwent hypomethylation concurrent with the appearance of the protein whilst several other sites became demethylated significantly later. Similarly when 11-day chick embryos were treated with oestrogen (Burch and Weintraub, 1983; Meijlink *et al* , 1983), vitellogenin message was detected 16h after treatment but the demethylation of the vitellogenin gene was not detected until 72h after treatment (ie after two cell divisions). Also, Kunnath and Locker (1983) demonstrated that the rat albumin and α-fetoprotein genes were fully methylated in the 18-day foetus, where the genes were first expressed, but were found to be less methylated in the adult liver where the albumin (but not α-fetoprotein) gene was still expressed. Again this represents a demethylation of a gene after it has been 'turned on'.

At the moment there appear to be two possible reasons why a demethylation may be a post expression event. First, it requires two cycles of cell division, during which the maintenance methylase is prevented from methylating a particular site on the daughter DNA strand, before a site will become Hpa II sensitive. This prompted

~~strand, before a site will become Hpa II sensitive. This prompted~~ Yisraeli and Szyf (1984) to suggest that hemimethylation, ie the methylation of a cytosine on only one strand (see figure 6), may be sufficient to allow the activation of a gene, whilst a further round of replication (again without methylation) is required before the demethylation can be detected. However, this cannot be the rule because the chicken vitellogenin mRNA was detected 16h after oestrogen treatment (ie before one round of cell division) although the demethylation was observed after two rounds of cell division (Wilks *et al* , 1982).

The second reason why a demethylation might be delayed after the initiation of transcription is because methylation may be inhibited in regions of DNA which are undergoing massive transcription. Over several generations a gene-specific pattern of undermethylation would result (Razin and Cedar, 1984). Further, it has been proposed that the failure to detect demethylation following the activation of the muscle α-actin gene may be due to the fact that the gene is only activated after the muscle cells have ceased to divide and in the absence of replication, demethylation will not occur (Razin and Cedar, 1984). However, recently members of Razin's group have chemically induced leukemia cells to differentiate and have observed a 'genome-wide hypomethylation'. The cells lost methyl groups very quickly, which implied that these hypomethylations were by an active mechanism (Kolata, 1985).

Although post expression demethylation events were detected in the pea storage protein genes, it should be noted that this does not preclude other demethylation events occurring at other sites which

would not have been detected by Hpa II. Hpa II can examine about 6% (or one sixteenth) of the CG dinucleotides in a genome (Yisraeli and Szyf, 1984; Riggs and Jones, 1983; Cooper, 1983). In the case of Leg A, it has been shown that out of 57 CG dinucleotides, only three of them were in a CCGG sequence (figure 48). In order to examine the methylation of all the cytosines it would be necessary to use the Church and Gilbert (1984) method of genomic sequencing (see also Little, 1984).

The methylation maps obtained for the genomic sequences cloned in λLeg 1 and λLeg 2 (figure 57) are in reasonable agreement with Razin and Szyfs third paradigm (1984). However, it should be noted that a methylation map such as this, does not necessarily reflect the methylation state of the DNA in the cells which are active in storage protein synthesis, because of the cotyledon cellular heterogeneity.

The greatest correlation between undermethylation of a gene region and the expression of that gene has been found in the 5' flanking (promoter) regions (for reviews see Doerfler *et al* , 1984; Cedar, 1984; Razin and Szyf, 1984; Doerfler, 1983). The methylation pattern obtained for the Leg A is very similar to that described by McKeon *et al* (1982) for α2(1) collagen. These workers reported that the structural gene was heavily methylated, whilst the 5' region was unmethylated in all tissue sources, despite the fact that the activity of collagen was limited to the fibroblast cells. It has been suggested (Cedar, 1984) that this seemingly non-tissue-specific undermethylation of the 5' region may be related to the fact that the gene is actively expressed during the early stages of differentiation.

The strongest evidence which supports the view that a changing methylation in the 5' promoter region and the control of gene expression has been provided by two experiments. Kriegek and Doerfler (1983) used hybrid plasmids containing the chloramphenicol acetyl-transferase (CAT) gene and adenovirus type 12 promoters. When various regions of the promoter were *in vitro* methylated it was found that CCGG and GCGC sites upstream from the E1a promoter had a decisive regulatory function. Methylation of sites 300bp upstream from the TATA signal had no effect on the CAT gene expression and neither did the methylation of the Hpa II sites in the CAT gene sequence or in the plasmid vector sequence. Similarly, Busslinger *et al* (1983b) demonstrated that *in vitro* methylation of the human $^{A}\gamma$-globin gene between bases -760 to +100 prevented *in vivo* transcription of the gene in mouse cells. Finally, Vardimon *et al* (1982) have demonstrated that the *in vitro* methylation of non-CG sequences (eg GGCC) does not affect gene transcription *in vivo* .

However, in some genes gene expression is correlated to changes in the methylation of the structural gene and this has led Riggs and Jones (1983) to propose that this may represent two levels of control operating independently, with the methylation of the promoter sequence acting as a primary switching mechanism, whilst the methylation of the coding region and introns could provide fine tuning. Clearly this could be the situation in Leg A and Leg B. The M2 site in Leg B was found to be less methylated than the Leg A M2 site (although it should be noted that the relative transcription rates of these two genes are not known). In addition the methylation of the M2 site of both genes

appeared to slightly decrease during cotyledon development and this again could represent 'fine tuning'.

The *in vitro* methylation experiments by Doerfler and Busslinger indicate a causal link between methylation and gene expression. Whilst a 5', developmentally regulated, methylated/unmethylated Hpa II site was not detected in association with Leg A, this does not preclude the existence of such a Hpa II site either further upstream of a methylation change at a non-Hpa II CG dinucleotide. Whilst it may be inappropriate to extrapolate the findings from the vertebrate system to the plant system, particularly in view of the varying amounts of 5mC found in eukaryotes (insects 5mC virtually absent, higher vertebrates - low levels of 5mC, higher plants - high levels of 5mC), the fact that correlations have been detected for mammalian CG residues and some higher plant CG and C-X-G residues (Cedar, 1984), implies that the link between gene activity and DNA methylation must be a basic one.

Riggs and Jones (1983) have suggested that methylation can be considered as a locking device, with methylation locking the gene in an inactive state. Thus, in the cases of genes where no correlation between undermethylation and gene expression has been found (eg Leg A) it must be concluded that further regulatory factors are involved or as a useful analogy 'an unlocked door is not necessarily open' (*ibid*). Therefore in the control of gene expression there is a multifaceted regulatory system and undermethylation of the 5' region may be a necessary but not sufficient condition to ensure gene expression.

Doerfler *et al* (1984) has pointed out that genes can be subdivided into three classes: a) permanently inactivated genes; b) inactive genes that are occasionally reactivated; c) active genes. The authors pointed out that the cell- or tissue-specific patterns of DNA methylation may be superimposed on the DNA methylation patterns which reflect the inactivated states of genes. If this is true, then it is surprising that the pea leaf methylation patterns (where the legumin genes are permanently inactive) were the same as the 10 d.a.f. cotyledon methylation patterns (where the genes were becoming active). However, the methylation patterns obtained for the vicilin cDNA probes supported Doerfler's scheme. The Msp I patterns obtained for leaf and 10 d.a.f. cotyledon DNAs were very different (figure 38 and 39) and this does indicate that the methylation of a completely inactive leaf vicilin gene was different to that of a gene which was either being activated or was about to be activated. Doerfler has suggested that once a gene has been inactivated, additional methylations of that gene may be related to the organization of chromatin (or possibly to other factors not directly responsible for the regulation of gene expression).

Both the vicilin cDNAs detected distinct hypermethylations of Msp I and Hae III sites during the cotyledon development. In the mammalian system, the differentiation and development of a tissue (or an organ) is usually considered to lead to hypomethylation of the genes (ie genes are heavily methylated in the sperm, less heavily methylated in somatic tissue and undermethylated in the expressing tissue). This is supported by experiments involving 5-azacytidine

where the drug has been used to induce cells to differentiate and DNA demethylation has also been detected (eg Jones and Taylor, 1980; Jones *et al* , 1982). The drawbacks to experiments using 5-azactidine were discussed in section 1:4:5).

Bovine DNA provides one exception to the differtiation = hypomethylation rule. The level of methylation in bovine sperm is 66% that observed in somatic cells (Vanyushin *et al* , 1980). However, Sturm and Taylor (1981) have shown that this difference is chiefly due to the level of methylation of the major satellite DNA. In sperm satellite DNA less than 1% of the cytosines were methylated whilst in somatic tissue satellite DNA, 10% of the cytosine residues were found to be methylated. Therefore a significant *de novo* methylation had occurred during bovine development. Jahner and Jaenisch (1984) have shown that proviral genomes inserted into mouse cells become highly methylated in mouse embryos. It was subsequently found that viral genomes became methylated if the viral genome had integrated into the DNA of the preimplantation but not the post- implantation embryo (Jahner *et al* , 1982). Therefore both repetitive and unique sequence DNA can become *de novo* methylated during early embryonic development. Thus cells undergoing differentiation have the potential for either demethylation or *de novo* methylation of their DNA.

Jaenisch and Jahner (1984) suggest that in vertebrates, the *de novo* methylation of the embryonal genome may in fact represent the repressing of the genes which had been active during oogenesis but were no longer required or which may be deleterious for the development of the early embryo ie the *de novo* methylation serves to

'reset' the genetic programme. A similar argument could be made for the pea vicilin genes in that the storage protein genes, whose expression is essential for the development of a viable pea seed, are not rquired after the end of the storage protein phase (ie about 20-21 d.a.f.). In fact, it could be argued that the expression of these genes during germination may be deleterious to the survival of the seedling, whose limited resources are required for other biochemical pathways. Therefore, the increasing methylation of the vicilin genes, during the later stages of cotyledon development might be expected (figure 38). However, if this is so, then the same criteria do not appear to hold for the legumin genes which fail to show an increased methylation (at Msp I and Hae III sites) during cotyledon development (figure 22b). This may reflect the difference in their chromosomal position ie it is not the vicilin genes whose expression would be deleterious to the germinating seedling but some neighbouring gene. Therefore the increased vicilin gene methylation may be a consequence of a general increase in methylation of that region of DNA. This is supported by the view of Bird (1984), that the unit of *de novo* replication (as distinct from the unit of methyl replication) is a sequence domain, within which all CGs become methylated at once.

The possible mode of action of the post expression, demethylation events in the pea cotyledon, is an interesting question. This site is thought to be up to 4.2kb from a legumin gene and up to 11.61kb from a vicilin gene ie the demethylation may be exerting an influence over a considerable distance. This may imply that it is related to an enhancer sequence. Jahner and Jaenisch (1984) have suggested that in

vertbrates, single copy genes may be regulated by two different sets of control elements: an embryonic set which is active independent of methylation and a somatic set which is sensitive to methylation. The mouse α-amylase gene has been found to use two different promoters in the parotid gland and the liver (Schibler *et al* , 1983). It may be that in the developing pea cotyledon, a cotyledon-specific enhancer is active at a low level despite being methylated but once the storage protein genes are expressed, the enhancer becomes demethylated (perhaps through the intervention of a protein factor which binds to the enhancer preventing access for the maintenance methylase). Once demethylated, the enhancer may have a higher level of activity.

Alternatively, these methylation changes could be operating over considerable distances by a system similar to that described by Rich (1983) ie DNA-domain activation of chromatin (see section 1:4:6 and figure 7). In this case the Hpa II demethylation detected by the cDNA probes would have to be regarded as one Hpa II site which was typical of a larger 5mCG-rich region of Z-DNA (which could still be an enhancer region). Thus the decrease in methylation of this region would promote a change in the DNA from the Z-form to the B-form and this change would be detected some distance away, within the same loop or 'domain'.

4:3:6 Investigation of a vicilin mutant pea line (5478)

The final section of experimental work which investigated the relationship between cytosine methylation and gene expression, involved mutant pea line 5478. This pea line was known to produce

greatly decreased amounts of the 50,000-M_r vicilin polypeptide (J.A. Gatehouse, personal communication).

The 50,000-M_r vicilin polypeptide genes were detected in the mutant pea line by Southern blotting and the use of a cDNA probe (figure 41). The genes were analysed with 12 different restriction enzymes (with either a 4 or 6 mucleotide recognition sequence) but no difference in the genes was detected between the mutant and parental pea lines. (The differences in the total intensity of some tracks is due to a loading artifact and does not indicate differences in copy number. The relative intensities of the bands within a track, can be compared between tracks.) No gross deletion or insertion into the mutant vicilin genes was apparent.

Having confirmed that the 50,000-M_r vicilin genes were present in the mutant pea line, total RNA was isolated from developing cotyledons and analysed by Northern blotting (figure 43a, b and c). Identical amounts of total RNA were loaded in each track, therefore the signal intensity is a measure of the proportion of the total RNA contributed by each mRNA. The pattern of 47,000-M, vicilin message accumulation (figure 43a) was similar to that described by Gatehouse *et al* (1982). Similar amounts of this message were detected in parental and mutant pea lines at each stage of cotyledon development.

In the parental line, the pattern of accumulation of the 50,000-M, vicilin polypeptide message (figure 43b) was again similar to that described by Gatehouse *et al* (*loc cit*) except that the message was detected in higher amounts slightly earlier in the cotyledon development (ie day 11). The mutant pea line showed a

similar pattern of 50,000-M_r polypeptide message accumulation, however, the level detected at each stage of cotyledon development was approximately 10-fold lower. This indicated either that the genes in the mutant were not being transcribed at the normal rate or that the message was being degraded in the nucleus (ie a post-transcriptional error). Measurement of the transcription rate by isolated nuclei would help to differentiate between these two possibilities.

Interestingly, when the same filter was rehybridized (figure 43c) with a cDNA representing a different 50,000-Mr vicilin message (pDUB 9), relatively higher amounts of this message were detected in the mutant pea line. This cDNA is thought to represent a message which is synthesised at a later stage of cotyledon development (1. M. Evans personal communication), although this experiment indicated that the message accumulated over a similar time period to the other 50,000-M_r cDNA (pDUB 2; figure 43b). Restriction mapping has indicated that pDUB 9 and pDUB 2 are dissimilar (Delauney, 1984). The significance of the differences of these two messages for the 50,000-M_r polypeptide is not yet understood.

Having established that the 'error' in this mutant pea line may be at the transcriptional control level, it was important to ascertain whether the demethylation event, observed in the Feltham First vicilin genes, was also occuring in the Witham Wonder mutant pea line (figure 44). The 11.61kb Hpa II fragment was detected in the mutant pea line 15 d.a.f. (track b) and was clearly visible by 22 d.a.f. (track f), whilst being absent from the leaf DNA (track j). This indicated that the same demethylation event which was observed in Feltham First was

also happening in the mutant cotyledon DNA. The 1.82kb Msp I and Hpa II band was just visible in parental and mutant pea lines. As this is thought to represent unmethylated sites in the Feltham First 50,000-M_r polypeptide gene, this indicates that hypermethylation of these mutant genes is not the cause of their low level of expression. It is interesting to note that the hypermethylation of the Msp I sites observed in Feltham First leaf DNA (figure 39, track e) is not detected in either the Witham Wonder parental or mutant lines (figure 44, tracks g and i).

These results do not indicate whether the reason for the failure to detect the normal amounts of cytoplasmic, 50,000-M_r mRNA, is due to changes in transcription or post transcriptional processing but they do indicate that an incorrect change in the methylation pattern of developing cotyledon DNA, is not the cause. The most probable reason for the low level of expression is a base change in a critical sequence of the promoter region (see figure 3).

4:4 Analysis of Leg A and Leg D sequence data

4:4:1 Analysis for 'CG-suppression'

Bird (1980) demonstrated that the genomes of higher eukaryotes (but not insects) were deficient in the CG dinucleotide. This feature is normally called 'CG-suppression', although in some ways this is an incorrect term to use, since it implies an active mechanism to ensure the rarity of the CG-dinucleotide, whereas in actual fact the rarity of the dinucleotide is probably a consequence of the high mutability

of the 5mC residues. McClelland and Ivarie (1982) analysed the data for an 'average' mammalian gene (based on the sequence data for 15 mammalian genes) and found that overall the CG dinucleotide occured at 40% the frequency expected from the base composition. Similarily, when analysing data for a typical leguminous plant gene, McClelland, (1983b) found that overall the CG dinucleotide occured at a slightly higher frequency (49%). Analysis of Leg A and Leg D sequence data has confirmed that in these two genes, the dinucleotide is deficient (tables 7 and 8). Overall the CG dinucleotide was found to occur at 44% the frequency expected in Leg A and 71% the frequency expected in Leg D. It is not known whether this discrepancy is related to the fact that Leg A is expressed, whilst Leg D is silent (Bown *et al* , 1985).

In contrast to the CG frequencies, the GC dinucleotides occured with a frequency slightly higher than was expected (tables 7 and 8). McClelland and Ivarie (1982) and McClelland (1983b) also demonstrated that in both mammalian and plant genes the CG dinucleotide was asymmetrically distributed. However, the two distributions were markedly different. The mammalian type gene had a far greater suppression of the dinucleotide in the 3' flanking region than in the 5' flanking region (16% compared with 53%) whilst the plant gene showed the reverse trend (96% compared with 31%). This distribution in Leg A is shown in figure 46a:III and for Leg D in figure 47a:III (see also tables 7 and 8). Neither of these analyses confirmed the relief from suppression in the 3' flanking region reported by McClelland (1983b) The reason for this is probably due to the fact that McClelland's results are based on the data of 7 soybean genes, 2 pea

cDNAs, 1 French bean cDNA and 1 maize gene. Therefore, McClelland's distribution reflects very strongly the situation observed for the soybean genes, whilst these legumin gene flanking sequences ratios, reflect the very low C+G percentage composition of the pea legumin gene flanking sequences (figure 46b:I and 47b:I).

The distribution of the CG dinucleotide in the flanking region of Leg A is shown in figure 45c and figure 46b:III. It should be noted that the stretch of DNA from the translation start (0) to -500bp upstream is virtually devoid of CG residues. In other genes, this is the region of DNA which when methylated *in vitro* was found to prevent expression *in vivo* (eg Busslinger *et al* , 1983b) (see section 4:3:5). Therefore, it might be expected that one of these three CG residues will in the future, be shown to have a dramatic effect on gene expression.

It was suggested by Salser (1977) that this lower frequency of the CG dinucleotide may be due to the fixation of mutations generated by a high rate of 5mC deamination (figure 51). If this is so, then the dinucleotide mutation products should be detected at a higher frequency. Figure 52 gave the clearest demonstration of the correlation between CG deficiency and (CA+TG) excess. Bird (1980), using an analysis of this kind, demonstrated that insects had neither a CG deficiency nor a reproducable (CA+TG) excess, non-arthropod invertebrates had an intermediate level of CG deficiency and (CA+TG) excess and finally vertebrates had a marked CG deficiency and a marked (CA+TG) excess. The distribution of points obtained for Leg A (figure 52) were more widely spread than those of Bird (1980) and this may be

due to the difference in unit length of DNA analysed (100 nucleotides in this study, whereas Bird used a unit length of 1000 dinucleotides).

When the data was expressed in terms of gene position (figure 53) only one area of Leg A was shown to have a very close correlation between CG deficiency and (CA+TG) excess (figure 53b:I). This area was found to correspond to the cluster of CG dinucleotides, seen in exon 1, in figure 45c. This region of exon 1 showed a relief from CG suppression and a near normal occurrence of (CA+TG). When one considers the overall rarity of the CG dinucleotide, then a region in which the CGs are maintained at almost the expected frequency assumes an increasing significance.

The origin of localised CG-rich regions, is currently the subject of much controversy (Max, 1984; Adams and Eason, 1984a; Max, 1985). Tykcocinski and Max (1984) have suggested that regions of DNA which contain a CG cluster are maintained in a highly demethylated state in germline DNA, relative to regions of DNA which exhibit CG suppression. If this is so then the model predicts that CG-rich regions will not show the elevated TG+CA levels that result from 5mCG deamination and this was found. However, Adams and Eason (1984a and b) have pointed out that there are two possible reasons for a lack of CG suppression: a) a failure to methylate (as proposed by Max) or b) a failure to deaminate the methylated CG-rich regions. Adams and Eason argue that a high G+C content (>60%) of a region of DNA will confer an increased stability on the DNA helix and this will mitigate against the deamination reaction, which *in vitro* requires the DNA to be single stranded. Therefore, even if the region of DNA is heavily methylated,

deamination is less likely to occur.

When one examines the sequence data of Leg A and Leg D, in the light of these two differing theories, it is not possible to favour one theory or the other. The extent of methylation of the first exon in Leg A is not known but the M2 site, about 200bp downstream is known to be about 25% methylated. However, the M1 site in the 5' flanking region, is known to be unmethylated and yet this site was located in a region in which there was a marked suppression of the CG dinucleotide. Max supports his theory by pointing out that is the major histo-compatibility complex genes, CG clustering in observed in the 5' region and that the mouse dihydrofolate reductase gene (DHFR) and chicken α-2 collagen gene have both been shown to have a cluster of CG dinucleotides which are undermethylated relative to the 3' region (ie there is a germline demethylation of these regions).

When the CG cluster in Leg A and D is examined in terms of the C+G content of the DNA, it was found that the two 100 nucleotide segments in which the CG sites are located (+100 to +300bp) have a C+G content of 52% and 46% (figure 46b:I, see also figure 48a), ie considerably less than Adams and Eason's requirement of 60% and not very much higher than the figure of 40% below which the CG dinucleotide in vertebrate DNA is highly suppressed. It is possible that the C+G composition of a much smaller region surrounding a CC dinucleotide (ie microenvironment) will dictate the stability of the DNA helix in that region. Figure 49 and table 10 examine the microenvironment of the CG dinucleotides in Leg A and Leg D. As a control, the microenvironment around the Leg A GC dinucleotides was

examined (figure 50 and table 10). This cytosine residue will not normally be methylated (unless the next 3' base is another guanine ie 5'-GCG-3') and therefore if subject to deamination, the resultant base would be uracil which would be recognised as an incorrect base. The Leg A, CG dinucleotides were found to occur predominantly in microenvironment which were between 40-70% C+G rich (ie considerably higher than the overall DNA percentage composition), whereas the GC dinucleotides occurred predominantly in microenvironments with a slightly lower percentage composition (40-60%). Therefore, it is possible that the CG dinucleotides may be protected from deamination by the increased stability of a localised C+G rich region or domain but the data is not strong enough to state categorically that this is so.

When an apparent anomaly in the sequence data for the protein coding regions of a gene is observed, it is necessary to consider whether the anomaly is a consequence of the protein coding requirement. It is probable that protein coding requirements or anomalies in the codon usage, would show up in the preliminary analysis shown in table 7. Some oddities were observed (eg a high CA ratio in exon 4, a high CT ratio in exon 3 and low GT and TA ratios in exon 3) but the CG dinucleotide was the only one to show an abnormal frequency throughout the gene. This point is emphasised in figure 48. Most published papers which report on the CG-suppression phenomenon, only show details of the CG dinucleotide distribution but it is only when the CG distribution and those of its potential mutation products (TG and CA) are compared with the other distributions (eg CT, TC, GA,

AG, TA and AT), that the significance of this phenomenon is emphasised.

4:4:2 Analysis of Leg A sequence data for 'CCG-suppression'

The analysis of a plant gene sequence has an added advantage in that the methylation of the cytosine residue is not confined to the CG dinucleotide. In addition, the C-X-G trinucleotide (where X can be C, A or T) is known to be methylated in plant DNA (Gruenbaum *et al*, 1981b). Therefore, if the rarity of the CG dinucleotide is due to the mutation of the 5mC, then similar anomalies should be indicated by the analysis of the distributions of the C-X-G trinucleotides.

Gruenbaum *et al* (1981b) reported that 50% of the external cytosines and 80% of the internal cytosines were methylated in wheat germ DNA. Therefore it is likely that the internal cytosine has a higher probability of undergoing deamination. The trinucleotide product of such a deamination would be 5'-CTG-3' and the complementary sequence would be 5'-CAG-3'. Some of these deamination products will still be modified at their remaining cytosine residue. These trinucleotides still have 180° rotational symmetry and therefore the methylation of the external cytosine, will be maintained after DNA replication. Where this cytosine residue is methylated in the CAG/CTG trinucleotides it too will be susceptible to a deamination (ie CTG→TTG and CAG→TAG). However, it should also be noted that it is possible that the external cytosine of the methylated CCG trinucleotide could be deaminated first (ie CCG→TCG→TTG) (see table 11).

The equivalent mutations, on the complementary DNA strand (ie the strand which provides the template for the RNA message) can also be monitored. A 5'-CCG-3' sequence on the template strand can be detected as a 5'-CGG-3' sequence on the sequenced DNA strand. Thus, the sequence of events 5'-CCG-3'⟶5'-CTG-3'⟶5'-TTG-3' on the template strand can be monitored by the complementary 5'-CGG-3'⟶5'-CAG-3'⟶5'-CAA-3' on the sequenced strand whilst 5'-CCG-3'⟶5'TCG-3'⟶5' TTG-3' on the template strand can be monitored by the complementary 5'-CGG-3'⟶5'-CGA-3'⟶5'-CAA-3' on the sequenced strand (table 11).

McClelland (1983b), using the data for leguminous plant protein coding genes, performed a similar analysis for trinucleotide mutations, by comparing the overall ratios of CNG and TNG+CNA and CAG+CTG. The results obtained for Leg A are not in complete agreement with those of McClelland. In Leg A, the degenerate trinucleotide CCG and its complementary sequence 5'-CGG-3' were both found to occur at a frequency lower than expected (74% and 44% respectively) (McClelland obtained values of 114% and 88% respectively). However, to a certain extent the Leg A figures reflect the rarity of all CG-trinucleotides (table 9) and do not necessarily indicate that the degenerate trinucleotide has mutated to its transition products.

Figures 57 and 58 show the relationships between the CCG trinucleotides on both DNA strands and their deamination products. In both cases, a better correlation was found when the internal cytosine was assumed to mutate at a greater rate than the external cytosine (ie figures 57a and 58a). However, it is probable that the lack of correlation in 57b and 58b is a consequence of the rarity of all CG-containing trinucleotides (see

table 9).

4:4:3 Relationship between possible 5mC deamination and protein coding requirements

When considering the effects of the 5mC Leg A deamination, it must be borne in mind that not all mutations will be permitted if a functional gene is to survive. It is only possible to measure the rate at which they are fixed in the population (acceptance rate). It is necessary to consider the 'total genetic load' ie the proportion by which population fitness or survival is decreased in comparison with that of the optimal genetype (Kimura, 1968; Salser and Isaacson, 1976). The total genetic load is normally considered to be the sum of the 'substitution load' (ie the cost of evolution, of replacing all the original genotypes with genotypes carrying a new advantageous mutation) plus the 'mutation load' (ie the cost of keeping the genome accurate by eliminating deleterious mutations) (Kimura, 1960). King and Jukes (1969) have stated that most proteins contain regions where the substitution of many amino acids can be made without producing appreciable changes in protein function. Kimura (1968) has estimated that about 20% of nucleotide replacement caused by mutation, is estimated to be 'synonymous' or 'silent' (ie codes for the same amino acid). It should be noted that these silent mutations are not necessarily 'neutral' mutations because they could disrupt important base-pairing relationships or change the binding of control proteins to DNA or RNA signal sequences etc. (Salser and Isaacson, 1976).

In the rabbit β-globin gene base substitutions were found to be

concentrated in the 'silent' (ie third coding position) rather than in amino acid changing positions (Salser, 1977). However, it was also found that in certain regions of the β-globin structural gene there was a complete absence of base substitutions of any kind. Salser (1977) concluded that in the rabbit β-globin gene, mutations which caused amino acid substitution, appeared to have a fixation probability about fivefold less than that for silent muations within the same gene. This implied that most amino acid substitutions in the β-globin gene were sufficiently deleterious to be eliminated in the course of evolution (ie a neo-Darwinist rather than a neutralist's argument). Thus the true state of affairs regarding the mutation of single-copy gene, is probably that the major proportion of the changes in the genome are neutral and those that cause deleterious amino acid substitutions will be eliminated during evolution.

The codon position of selected Leg A nucleotides which were thought to be involved in the deamination of 5mC was investigated. Table 12 shows a comparison of the frequency of occurrence of selected di and trinucleotides in Leg A, between the protein coding and non-protein coding regions of Leg A. A major difference in the occurrence of certain trinucleotides would indicate that the structural gene imposed certain constraints on the deamination of 5mC residues. Salser and Isaacson (1946) have suggested that only small fraction of the genome is kept accurate (<2%), whilst most single-copy DNA is genetically drifting. Surprisingly, the major differences observed in table 12 concerned the distribution of the CAG, CTG, TTG and CAA trinucleotides, all of which had higher than expected

occurrence in the protein coding regions compared to the non-coding regions. In view of the fact that both CGG and CCG were depressed in the protein coding regions, these increased ratios may reflect the deamination of the CCG trinucleotides in the protein coding region. It should also be noted that the CA dinucleotide and CAG trinucleotide occur more frequently than the ATG dinucleotide and CTG trinucleotide. This may reflect either the DNA or RNA structural considerations or it may indicate a faster mutation rate on the RNA template strand than on the DNA sequenced strand.

The codon position of specified nucleotides was examined. This is not quite the same as looking for silent mutations, since a change in certain third coding positions will alter the amino acid encoded by the codon. In light of the previously discussed β-globin gene, which had a fivefold greater occurrence in nucleotide substitutions which were synonymous, then it might be expected that in Leg A, nucleotide substitutions due to 5mC deamination might be expected to be detected in the third coding position (table 15).

Whilst the frequency of CG dinucleotide cytosine residues in position 3 was not particularly low (42.5%), there was the expected increase in the number of TG dinucleotide thymine residues in position 3 (64.4%). It was also possible to examine the situation on the complementary RNA template strand ie the G in CG reflects the complementary cytosine and the number observed in the third coding position was found to be decreased (17.5%). However, the expected increase in the A of the CA dinucleotide was not observed (22.1%). Similarly, the number of Ts in CTG, in the third position was found to

be increased (69.6%), but there was no equivalent increase in the number of As in the CAG trinucleotide, in the third position (25.6). (The position of this A reflects the position of the T in the complementary CTG). This difference may reflect the increased number of CTG trinucleotides on the template strand (thought to be a consequence of an increased CCG mutation rate on the template strand), or it may reflect DNA or RNA structural requirements.

The codon position of the various nucleotides in the TTG trinucleotide was also examined (table 15). This trinucleotide is known to occur more frequently than was expected in the protein coding regions (table 12). Some of these TTG trinucleotides may have arisen from the deamination of the methylated CTG trinucleotides, some of which in turn could have been the deamination products of the CCG mutation. It was found that 62.5% of the internal Ts (T<u>T</u>G) were in the third codon postion. The TGs, which were a part of a TTG sequence, represented 29% of the total number of TGs in the protein coding region and this taken with the prevalence of the TTG trinucleotide may support the view that they represent the final mutation product of the CCG trinucleotide. However, if so then this means that a large proportion of the second mutations (CTG→TTG) must have occurred in the second codon position, an event which would result in a change in the amino acid encoded. This might indicate that the need to delete the 5mC is of a greater importance than the need to maintain the accuracy of the final protein.

If a situation existed, where the deamination of CCG, CTG and CAG trinucleotides were unrelated events, then it would be expected that a

greater proportion of the external T in TTG trinucleotides (CTG deamination products) and of the T in TAG trinucleotides (CAG deamination products), would be found in the third codon postion (ie a potential silent mutation). The fact that the middle nucleotide, in both TTG and TAG is predominatly located in the third codon position, supports the view that they are the final products of a CCG deamination.

However, the RNA template strand CTG trinucleotides do not show the same tendency to deaminate to form the TTG trinucleotide, as was observed for the CTG trinucleotides on the sequenced strand. Only 10.5% internal adenine residues (of the complementary CAA on the sequenced strand), were found in the third codon position whilst a higher percentage (40.4%) of the external adenines (CAA), were located in the third codon postion. Together the CAG and CAA trinucleotides accounted for almost 70% of all CA dinucleotides in the protein coding sequence and both trinucleotides had a higher than expected occurrence in the protein coding regions (table 12).

Taken as a whole the results tend to confirm that the deamination of 5mC residues occurs predominantly in the potentially 'silent' third codon position. However, from this limited data source, it is difficult to equate the stipulations which dictate whether or not a deamination will occur. It is also difficult to explain the apparent difference in the deamination rates of the two DNA strands.

4:4:4 Arginine codon usage

40 codons specifying arginine were found in the Leg A sequence.

Of these, 16 were found to be CG containing codons. In general the arginine codons AGA and AGG are prefered over the other four CG-containing codons (Lycett *et al* , 1983a). However, in exon 1, it was found that 6 cytosines out of the 11 CG dinucleotides were located in a CG-containing arginine codon (table 16). Thus the arginine codon usage in exon 1 was very different to the rest of the Leg A gene. Subak-Sharpe *et al* (1966) have suggested that in mammalian DNA the CG shortage is of a magnitude which virtually precludes the use of CG for the general coding of amino acids. It is also interestng to note that mammalian cells with a low CG level probably use the CG-containg arginine codons infrequently and yet the larger mammalian virus, (eg *Herpes*) make frequent use of CG containing codons. To overcome this discrepancy the virus DNA also has to code for new tRNAs which will qualitatively modify the translation apparatus in the host cells *(ibid)* .

King and Jukes (1969) have compared the observed number of arginine residues in 53 vertebrate polypeptides, with the frequency expected from random permutations of the nucleotides. Arginine residues occurred at 40% of the frequency expected. The authors have suggested that one explanation may be that the amount of arginine that can be tolerated in animal protein is less than the amount which would result if all six arginine codons were present at the expected frequency. Therefore, the CG content of animal DNA has been lowered by natural selection.

It is not possible to determine whether the arginine codon usage observed in the Leg A exon 1 is a consequence of a high CG

dinucleotide occurrence or vice versa. In some ways the non-typical use of the argine codons in exon 1 supports the idea that the genes for polypeptides may have arisen by exon-shuffling. In 1978, Gilbert proposed that a gene was a mosaic of expressed sequences held in a matrix of silent DNA. Gilbert argued that such a structure would allow the recombination of different coding regions and thus speed up evolution. Blake (1978 and 1983) suggested that exons might code for discrete, stable regions of protein. Therefore, exon shuffling might assemble new proteins through the novel reassortment of stable substructures. Both theories predicted that introns would be found between DNA regions which coded for definable structural units of proteins (Lewin, 1982). In fact some genes eg chicken pyruvate kinase (Longerg and Gilbert, 1985) support this prediction whilst others (eg actin and myosin) conflicted with it.

Some authors eg Cavalier-Smith (1985) strongly discount the idea of exon shuffling for protein recombination. One of the main objections to the idea of exon shuffling is that it fails to adequately explain the origin (or co-evolution) of either RNA splicing or of introns. Cavalier-Smith (*loc cit*) proposes that instead of exon shuffling, the introns were inserted (ie transposon theory of introns or selfish DNA). Rogers (1985) has reported that the serine protease genes (which encode trypsin, blood clotting and complement factors), show a clear pattern of domains (ie support for exon shuffling) but in addition demonstrate that some new introns have been inserted at various times in evolution.

The CG frequency and arginine coson usage in Leg A's exon 1 is

sufficiently different from the other Leg A exons to warrant a further investigation of the legumin gene exon-intron relationship (eg a Go plot). It must also be appreciated that because there is such overwhelming evidence supporting the view that the CG dinucleotide is suppresses in eukaryotic genomic DNA, any region of DNA which contradicts this generalisation may be of great significance. It is possible that in the future such CG-cluster regions may be shown to have a special role in the control of gene expression (eg structural).

Finally, since the sequence data of Leg B and C is likely to be very similar to that of Leg A, an analysis for CG suppression in these two genes is not likely to yield very much new information. However, an analysis of the sequence data of two other legumin genes (J and K), which are known to be dissimilar to Leg A (ie they do not cross-hybridize), would prove very interesting.

5: BIBLIOGRAPHY

Bibliography

1. von ACKEN, U., SIMON, D., GRUNERT, F., DORING, H-P. and KROGER, H. (1979). Methylation of viral DNA *in vivo* and *in vitro* . Virology, 99, 152-157.

2. ADAMS, R.L.P., DAVIS, T., FULTON, J., KIRK, D., QURESHI, M. and BURDON, R.H. (1984). Eukaryotic DNA methylase-properties and action on native DNA and chromatin. Curr. Top. Microbiol. Immunol., 108, 143-156.

3. ADAMS, R.L.P. and EASON, R. (1984a). Increased G+C content stabilises methyl CpG dinucleotides. Nucl. Acids R., 12, 5869-5877.

4. ADAMS, R.L.P. and EASON, R. (1984b). DNA methylation and G+C rich DNA. Nature, Lond., 312, 407-408.

5. ADAMS, R.L.P., McKAY, E.L., CRAIG, L.M. and BURDON, R.H., (1979). Mouse DNA methylase: methylation of native DNA. Biochim. Biophys. Acta., 561, 345-357.

6. AUBIN, R.J., FRECHETTE, A., de MURCIA, G., MANDEL, D., LORD, A., GRONDIN, G. and POIRIER, G.G. (1983). Correlations between endogenous nucleosomal hyper (ADP-ribosyl) ation of histone M1 and the induction of chromatin relaxation. EMBO J., 2, 1685-1693.

7. BAIN, J.M. and MERCER, F.V. (1966). Subcellular organization of the developing cotyledon of *Pisum sativum* L.. Aust. J. Biol. Sci., 19, 49-68.

8. BAKER, C.C., HERISSE, J., COURTOIS, G., GALIBERT, F. and ZIFF, E. (1979). Messenger RNA for the Ad2 DNA binding protein: DNA sequence encoding the first leader and heterogeneity at the mRNA 5' end. Cell, 18, 569-580.

9. BANERJI, J., OLSON, L. and SCHAFFNER, W. (1983). A lymphocyte-specific cellular enhancer is located downsteam of the joining region in the immunoglobulin heavy chain genes. Cell, 33, 729-740.

10. BARINGA, M., FRANCO, R., MEINKOTH, J., ONG, E. and WAHL, G.M. (1981). Method for the transfer of DNA, RNA and Protein to Nitrocellulose and diazotized paper solid supports. Schleicher and Schull.

11. BEACH, L.R., SPENCER, D., RANDALL, P.J. and HIGGINS, T.J.V. (1985). Transcriptional and post-transcriptional regulation of storage protein gene expression in sulphur deficient pea seeds. Nucl. Acids R., 13, 999-1013.

12. BEHE, M. and FELSENFELD, G. (1981). Effects of methylation on a synthetic polynucleotide: the B-Z transition in poly (dG-m^5dC). poly (dG-m^5dC). Proc. Natl. Acad. Sci. U.S.A., 78, 1619-1623.

13. BENOIST, C., O'HARE, K., BREATHNACH, R. and CHAMBON, P. (1980). The ovalbumin gene-sequence of putative control regions. Nucl. Acids R., 8, 127-142.

14. BIRD, A.P. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: II The synthesis of methylated sites supports semi-conservative copying of the methylation pattern. J. Molec. Biol., 118, 49-60.

15. BIRD, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucl. Acids R., 8, 1499-1504.

16. BIRD, A.P. (1983). DNA modification In: MACLEAN, M., FLAVELL, R.A. (eds). Eukaryotic genes: their structure, activity and regulation. Butterworths, Sevenoaks, U.K. pp.53-67

17. BIRD, A.P. (1984). DNA methylation - how important in gene control. Nature, Lond., 307, 503-504.

18. BIRD, A.P. and SOUTHERN, E.M. (1984). Use of restriction enzymes to study eukaryotic DNA methylation: 1. The methylation pattern in ribosomal DNA for *X. laevis* . J. Molec. Biol., 118, 27-47.

19. BIRD, A.P. and TAGGART, M.B. (1980). Variable patterns of total DNA and rDNA methylation in animals. Nucl. Acids R., 8, 1485-1497.

20. BIRD, A.P., TAGGART, M.H., FROMMER, M., MILLER, O.J. and MACLEOD, D. (1985). A fraction of the mouse genome that is derived from islands of non-methylated, CpG-rich DNA. Cell, 40, 91-99.

21. BIRD, A.P., TAGGART, M.H. and GEHRING, C.A. (1981). Methylated and unmethylated ribosomal RNA genes in mouse. J. Molec. Biol., 152, 1-17.

22. BIRD, A.P., TAGGART, M.B. and SMITH, B.A. (1979). Methylated and unmethylated DNA compartments in the sea urchin genome. Cell, 17, 889-901.

23. BIRNBOIM, H.C., and DOLY, H.C. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids R., 7, 1513-1523.

24. BLAKE, C.C.F. (1978). Do genes-in-pieces imply proteins-in-pieces? Nature, Lond., 273, 267.

25. BLAKE, C.C.F. (1983). Exons-present from the beginning? Nature, Lond., 306, 535-537.

26. BOREK, E. and SRINIVASAN, P.R. (1966). The methylation of nucleic acids. A. Rev. Biochem., 35, 275-298.

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

28. BOWN, D., LEVASSEUR, M., CROY, R.R.D., BOULTER, D. and GATEHOUSE, J.A. (1985). Sequence of a pseudogene in the legumin gene family of pea (*Pisum sativum* L.). Nucl. Acids R., 13, 4527-4538.

29. BREATHNACH, R. and CHAMBON, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. A. Rev. Biochem., 50, 349-383.

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

31. BUSSLINGER, M., de BOER, E., WRIGHT, S., GROSVELD, F.G. and FLAVELL, R.H. (1983a). The sequence GGCCGG is resistant to Msp I cleavage. Nucl. Acids R., 11, 3559-3569.

32. BUSSLINGER, M., HURST, J. and FLAVELL, R.A. (1983b). DNA methylation and regulation of globin gene expression. Cell, 34, 197-206.

33. CASEY, R. and SANGER, E. (1980). Purification and some properties of a 7S seed storage protein from Pisum pea. Biochem. Soc. Trans., 8, 658.

34. CATE, R.L., CHICK, W. and GILBERT, W. (1983). Comparison of the methylation patterns of the two rat insulin genes. J. Biol. Chem., 258, 6655-6652.

35. CAVALIER-SMITH, T. (1985). Selfish DNA and the origin of introns. Nature, Lond., 315, 283-284.

36. CEDAR, H. (1984). DNA methylation and gene expression. In: Razin, A., Cedar, H. and Riggs, A.D. (eds). DNA methylation: biochemistry and biological significance. Springer-Verlag, New York, pp.147-164.

37. CEDAR, H., SOLAGE, A., GLASER, G. and RAZIN, A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme Msp I. Nucl. Acids R., 6, 2125-2132.

38. CHANDLER, P.M., HIGGINS, T.J.V., RANDALL, P.J. and SPENCER, D. (1985). Regulation of legumin levels in developing pea seeds under conditions of sulphur deficiency. Pl. Physiol., 71, 47-54.

39. CHAPMAN, V., FORRESTER, C., SANFORD, J., HASTIE, N. and ROSSANT, J. (1984). Cell lineage specific undermethylation of mouse repetitive DNA. Nature, Lond., 307, 284-286.

40. CHEN, C.W. and THOMAS, C.A. (1980). Recovery of DNA segments from agarose gels. Anal. Biochem., 101, 339-341.

41. CHARNAY, P., TREISMAN, R., MELLON, P., CHAO, M., AXEL, R., MANIATIS, T. (1984). Differences in human α and β-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell, 38, 251-263.

42. CHIRGWIN, J.M., PRZYBYLA, A.E., MACDONALD, R.J. and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonucleic acids. Biochemistry, 18, 5294-5299.

43. CHRISTMAN, J.K., PRICE, P., PEDRIMAN, L. and ACS,G. (1977). Correlation between hypomethylation of DNA and expression of globin genes in Friend erythroleukemia cells. Eur. J. Biochem., 81, 53-61.

44. CHURCH, G.M. and GILBERT, W. (1984). Genomic sequencing. Proc. Natl. Acad. Sci. U.S.A., 81, 1991-1995.

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

46. CONNER, B.N., TAKANO, T., TANAKA, S., ITAKURA, K., DICKERSON, R.E. (1982). The molecular structure of d(CpCpGpG), a fragment of right-handed double helical A-DNA. Nature, Lond., 295, 294-299.

47. COOPER, D.N. (1983). Eukaryotic DNA methylation. Hum. Genet., 64, 315-333.

48. COOPER, D.N., TAGGART, M.H. and BIRD, A.P. (1983). Unmethylated domains in vertebrate DNA. Nucl. Acids R., 11, 644-658.

49. CROUCH, M.L. and SUSSEX, I.M. (1981). Development and storage protein synthesis in *Brassica napus* L. embryos *in vivo* and *in vitro* . Planta, 153, 64-74.

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

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

52. CROY, R.R.D., GATEHOUSE, J.A., TYLER, M. and BOULTER, D. (1980c). The purification and characterization of a third storage protein (convicilin) from the seeds of pea (*Pisum sativum* L.). Biochem. J., 191, 509-516.

53. CROY, R.R.D.,LYCETT, G.W., GATEHOUSE, J.A., YARWOOD, J.N. and BOULTER, D. (1982). Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature, Lond., 295, 76-79.

54. CUELLAR, R.E. (1982). PhD Thesis. Stanford University, USA.

55. CULLIS, C.A. and DAVIES, D.R. (1975). Ribosomal DNA amounts in *Pisum sativum* . Genetics, 81, 485-492.

56. CUMMINGS, D.J., TAIT, A. and GODDARD, J.M. (1974). Methylated bases in DNA from *Paramecium aurelia* . Biochim. Biophys. Acta., 374, 1-11.

57. CUTHBERTSON, D. (1970). In: Lawrie, R.A. (edt), Proteins as human food. Butterworths, London. pp.xiii-xvii.

58. DANIELSSON, C.E. (1949). Seed globulins of the Gramineae and Leguminosae. 1. Seed globulins of the most common species of the Gramineae and their differentiation in the seed. Biochem. J., 44, 387-400.

59. DARNELL, J.E. (1982). Variety in the level of gene control in eukaryotic cells. Nature, Lond., 297, 365-371.

60. DEAN, D.C., KNOLL, B.J., RISER, M.E., O'MALLEY, B.W. (1983). 5' flanking sequences essential for progesterone regulation of an ovalbumin fusion gene. Nature, Lond., 305, 551-554.

61. DELAUNEY, A.J. (1984). Cloning and characterisation of cDNAs encoding the major, pea storage proteins, and expression of vicilin in *E.coli* . PhD Thesis, University of Durham.

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

63. DERBYSHIRE, E., WRIGHT, D.J. and BOULTER, D. (1976). Legumin and vicilin storage proteins of legume seeds. Phytochem., 15, 3-24.

64. DOERFLER, W. (1983). DNA methylation and gene activity. A. Rev. Biochem. 52, 93-124.

65. DOERFLER, W., LANGNER, K-D., DRUCZEK, I., VARDIMON, L. and RENZ, D. (1984). Specific promoter methylations cause gene inactivation. In: Razin, A., Cedar, H., and Riggs, A.D. (edt). DNA methylation: biochemistry and biological significance, Springer-Verlag, New York, pp.221-248.

66. DOMONEY, C. and CASEY, R. (1984). Storage protein precursor polypeptides in cotyledons of *Pisum sativum* L. Eur. J. Biochem., 139, 321-327.

67. DOMONEY, C. and CASEY, R. (1985). measurement of gene number for seed storage proteins in *Pisum* . Nucl. Acids R., 13, 687-699.

68. DOMONEY, C., DAVIES, D.R. and CASEY, R. (1980). The initiation of legumin synthesis in immature embryos of *P. sativum* L. grown *in vivo* and *in vitro* . Plant, 149, 454-460.

69. DRAHOVSKY, D., BOEHM, T.L.J. and KREIS, W. (1979). Distribution pattern and enzymic hypermethylation of inverted repetitive DNA sequences in P815 mastocytoma cells. Biochim. Biophys. Acta, 563, 28-35.

70. DRAHOVSKY, D. and MORRIS, N.R. (1971a). Mechanism of action of rat liver DNA methylase. I: Interaction with double-stranded methyl accepter DNA. J. Molec. Biol. 57, 475-489.

71. DRAHOVSKY, D. and MORRIS, N.R. (1971b). Mechanism of action of rat liver methylase. II: Interaction with single-stranded methyl acceptor DNA. J. Molec. Biol., 61, 343-356.

72. DRETZEN, G., BELLARD, M., SASSONE-CORSI, P. and CHAMBON, P. (1981). A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem., 112, 295-296.

73. DURE III, L.S. (1975). Seed formation. A. Rev. Pl. Physiol., 26, 259-278.

74. DURE III, L.S. and GALAU, G.A. (1981). Developmental biochemistry of cottonseed embryogenesis and germination. XIII: Regulation of biosynthesis of principal storage proteins. Pl. Physiol., 68, 180-86.

75. ECONOMIST, (1984a). Old MacDonald had a gene. Econ., 293, 87.

76. ECONOMIST, (1984b). Genetic millionairing: Britain turns to agriculture. Econ., 293, 87-88.

77. ECONOMIST, (1985a). Peasants rising. Econ., 294, 13-14.

78. ECONOMIST, (1985b). In praise of peasants. Econ., 294, 94-94.

79. ECONOMIST, (1985c). Giving plants new genes is easier than it looked. Econ., 295, 83-84.

80. EHRLICH, M. and WANG, R.Y-H. (1981). 5-methylcytosine in eukaryotic DNA. Science, 212, 1250-1357.

81. ELLIS, T.H.N., GOLDSBROUGH, P.B. and CASTLETON, J.A. (1983). Transcription and methylation of flax rDNA. Nucl. Acids R., 11, 3047-3064.

82. ENCYCLOPAEDIA BRITANNICA, (1974). Malthus, T.R. Macropaedia, 11, pp.394-395.

83. EVANS, I.M., CROY, R.R.D., HUTCHINGSON, P., BOULTER, D., PAYNE, P.I. and GORDON, M.E. (1979). Cell free synthesis of some storage protein subunits by polyribosomes and RNA isolated from developing seeds of pea (*Pisum sativum* L.). Planta, 144, 455-462.

84. EVANS I.M., GATEHOUSE, J.A.G., CROY, R.R.D. and BOULTER, D. (1984). Regulation of the transcription of storage-protein mRNAs in nuclei isolated from developing pea (*Pisum sativum* L.) cotyledons. Planta, 160, 559-568.

85. GATEHOUSE, J.A., EVANS, I.M., BOWN, D., CROY, R.R.D. and BOULTER, D. (1982). Control of storage-protein synthesis during seed development in pea (*Pisum sativum* L.). Biochem. J., 208, 119-127.

86. GATEHOUSE, J.A., LYCETT, G.W., DELAUNEY, A.J., CROY, R.R.D. and BOULTER, D. (1983). Sequence specificity of the post-translational proteolytic cleavage of vicilin a seed storage protein of pea (*Pisum sativum* L.). Biochem. J., 212, 427-432.

87. GELINAS, R., ENDLICH, B., PFEIFFER, C., YAGI, M. and STAMATOYANNOPOULOS, G. (1985). G to A substitution in the distal CCAATA box of $^{A}\gamma$-globin gene. Nature, Lond., 313, 323-326.

88. GERBER-HUBER, J., MAY, F.E.B., WESTLEY, B.R., FELBER, B.K., HOSBACH, H., ANDRES, A-C., RYFFEL, G.G. (1983). In contrast to other *Xenopus* genes the estrogen-inducible vitellogenin genes are expressed when totally methylated. Cell, 33, 43-51.

89. GERLACH, W.C. and BEDBROOK, J.R. (1979). Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucl. Acids R., 7, 1869-1885.

90. GIL, A. and PROUDFOOT, N.I. (1984). A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3'-end formation. Nature, Lond., 312, 473-474.

91. GILBERT, W. (1978). Why genes in pieces? Nature, Lond., 271, 501.

92. GILLIES, S.D., MORRISON, S.L., OI, V.T., TONEGAWA, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell, 33, 717-718.

93. GJERSET, R. and MARTIN, D.W. (1982). presence of a DNA demethylating activity in the nucleus of murine erythroleukemia cells. J. Biol. Chem., 257, 8581-8583.

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

95. GRAINGER, R.M., HAZARD-LEONARDS, R.M., HOUGAN, L.M. LESK, M.R. and THOMSEN. G.M. (1983). Is hypomethylation linked to activation of δ-crystallin genes during lens development. Nature, Lond., 306, 88-91.

96. GRONER, B., KENNEDY, N. and SKROCH, P., HYNES, N.E. and PONTA, H. (1984). DNA sequences involved in the regulation of gene expression by glucocerticoid hormones. Biochim. Biophys. Acta., 781, 1-6.

97. GROUDINE, M., EISENMAN, R. and WEINTRAUB, H. (1981a). Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. Nature, Lond., 292, 311-317.

98. GROUDINE, M. and WEINTRAUB, H. (1981b). Activation of globin genes during chicken development. Cell, 24, 393-401.

99. GRUENBAUM, Y., CEDAR, H. and RAZIN, A. (1981a). Restriction enzyme digestion of hemi-methylated DNA Nucl. Acids R., 9, 2509-2515.

100. GRUENBAUM, Y., NAVEH-MANY, T., CEDAR, H. and RAZIN, A. (1981b). Sequence specificity of methylation in higher plant DNA. Nature, Lond., 292, 860-862.

101. GRUENBAUM, Y., STEIN, R., CEDAR, H., and RAZIN, H. (1981c). Methylation of CpG sequences in eukaryotic DNA. FEBS. LETT., 124, 67-71.

102. GRUENBAUM, Y., CEDAR, H., and RAZIN, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. Nature, Lond., 295, 620-622.

103. HALL, T.C., MA, Y., BUCHBINDER, B.U., PYNE, U.W., SUN, S.M. and BLISS, F.A. (1978). Messenger RNA for GI protein of French bean seeds: cell free translation and product characterization. Proc. Natl. Acad. Sci. U.S.A., 75, 3196-3200.

104. HARLAND, R.M. (1982). Inheritance of DNA methylation in microinjected eggs of *Xenopus laevis* . Proc. Natl. Acad. Sci. U.S.A., 79, 2323-2321.

105. HIGGINS, T.J.V. (1984). Synthesis and regulation of major proteins in seeds. A. Rev. Biochem., 35, 191-221.

106. HIGGINS, T.J.V. and SPENCER, D. (1981). Precursor form of pea vicilin subunits. Pl. Physiol., 67 205-211.

107. HILL, B. (1975). In: Studies in the British economy: Britain's agricultural industry. Heinemann Educational Books, London, pp.89.

108. HOFER, E. and DARNELL, J.E. 91981). The primary transcription unit of the mouse -major globin gene. Cell, 23, 585-599.

109. HOLLIDAY, R. and PUGH, J.E. (1975). DNA modification mechanism and gene activity during development. Science, 187, 226-232.

110. HOTTA, Y. and HECHT. N. (1971). Methylation of *Lilum* DNA during the meiotic cycle. Biochem. Biophys. Acta., 238, 50-59.

111. IGO-KEMENES, T., HORZ, W. and ZACHAU, H.G. (1982). Chromatin. A. Rev. Biochem., 51, 89-121.

112. INGLE, J. and SINCLAIR, J. (1972). Ribosomal RNA genes and plant development. Nature, Lond., 235, 30-32.

113. INGLE, J., TIMMIS, J.N. and SINCLAIR, J. (1975). Relation between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy and genome size. Pl. Physiol., 55, 496-501.

114. ISENBERG, I. (1979). Histones. A. Rev. Biochem., 48, 159-191.

115. JAENISCH, R. and JAHNER, D. (1984). Methylation, expression and chromosomal position of genes in mammals. Biochim. Biophys. Acta., 782, 1-9.

116. JAHNER, D. and JAENISCH, R. (1984). DNA methylation in early mammalian development. In: Razin, A., Cedar, H. and Riggs, A.D. (edt), DNA methylation: biochemistry and biological significance. Springer-Verlag, New York, pp.189-220.

117. JAHNER, D., STUHLMANN, H., STEWART, C.L., HARBERS, K., LOHLER, J., SIMON, I. and JAENISCH, R. (1982). *De novo* methylation and expression of retroviral genomes during mouse embryogenesis. Nature, Lond., 298, 623-628.

118. JAVEHERIAN, L., LIU, L.F. and WANG, J.C. (1978). Non-histone proteins HMG, and HMG_2 change the DNA helical structure. Science, 199, 1345-1346.

119. JONES, P.A. and TAYLOR, S.M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell, 20, 85-93.

120. JONES, P.A. TAYLOR, S.M. MOHANDRAS, T. and SHAPIRO, L.J. (1982). Cell cycle specific reactivation of an inactive X-chromosome locus by 5-aza deoxycytidine. Proc. Natl. Acad. Sci. U.S.A., 79, 1215-1219.

121. KATZ, L., WILLIAMS, P.H., SATO, H., LEAVITT, R.W. and HELINSKY, D.R. (1977). Purification and characterisation of covalently closed replicative intermediates of ColE1 DNA from *E. Coli* . Biochemistry, 16, 1677-1683.

122. KAYE, A.M. and WINOCOUR, E. (1967). On the 5-methylcytosine in the DNA extracted from polyoma virus mouse kidney. J. Molec. Biol., 24, 475-478.

123. KESHET, E. and CEDAR, H. (1983). Effect of CpG methylation on Msp I. Nucl. Acids R., 11, 3571-3580.

124. KHOURY, G. and GRUSS, P. (1983). Enhancer elements. Cell, 33, 313-314.

125. KILPATRICK, M.W., WEI, C.F., GRAY, H.B. and WELLS, R.D. (1983). Bal 31 nuclease as a probe in concentrated salt for the B-Z DNA junction. Nucl. Acids R., 11, 3811-3822.

126. KIMURA, M. (1960). Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. J. Genet., 57, 21-34.

127. KIMURA, M. (1968), Evolutionary rate at the molecular level. Nature, Lond., 217, 624-626.

128. KING, J.L. and JUKES, T.H. (1969). Non-Darwinian Evolution. Science, 164, 788-797.

129. KOLATA, G. (1983). Z-DNA moves towards 'Real Biology'. Science, 222, 495-496.

130. KOLATA, G. (1985). Fitting methylation into development. Science, 228, 1183-1184.

131. KRUCZEK, I. and DOERFLER, W. (1983). Expression of chloramphenicol acetyl transferase gene in mammalian cells under the control of adenovirus type 12 promoters: Effect of promoter methylation on gene expression. Proc. Natl. Acad. Sci. U.S.A. 80, 7586-7590.

132. KUNNATH, L. and LOCKER, J. (1982a). Characterization of DNA methylation in the rat. Biochem. Biophys. Acta, 699, 264-271.

133. KUNNATH, L. and LOCKER, J. (1982b). Variable methylation of the ribosomal RNA genes of the rat. Nucl. Acids R., 10, 3877-3892.

134. KUNNATH, L. and LOCKER, J. (1983). Developmental changes in the methylation of the rat albumin and α-fetoprotein genes. EMBO J., 2, 317-324.

135. LANGRIDGE, P., PINTOR-TORO, J.A. and FEIX, G. (1982). Zein precursor mRNAs from maize endosperm. Mol. Gen. Genet., 187, 432-438.

136. LEAVER, L.J. (1979). Ribosomal RNA of plants. In: Hall, T.C. and Davies, J.W. (edt), Nucleic acids in plants. CRC Press. Boca-Raton, Florida, U.S.A. pp.193-215.

137. LEVINGER, L. and VARSHAUSKY, A. (1982). Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the *Drosophila* genome. Cell, 28, 375-385.

138. LEWIN, B. (1982). Origin of introns. Science, 217, 921-922.

139. LEWIN, B. (1983). Promoters: the sites for imitating transcription. In: Genes. John Wiley and Sons, New York, pp.174-194.

140. LINDAHL, T. (1981). DNA methylation and control of gene expression. Nature, Lond., 290, 363-364.

141. LINK, G. (1984). DNA sequence requirements for the accurate transcription of a protein-coding plastid gene in a plastid *in vitro* system for mustard. EMBO J., 3, 1694-1704.

142. LONBERG, N. and Gilbert, W. (1985). Intron/exon structure of the chicken pyruvate kinase gene. Cell, 40, 81-90.

143. LONG, E.O. and DAWID, I.B. (1980). Repeated genes in eukaryotes. A. Rev. Biochem., 49, 727-764.

144. LYCETT, G.W., DELAUNEY, A.J. and CROY, R.R.D. (1983a). Are plant genes different? FEBS. LETT., 153, 43-46.

145. LYCETT, G.W. DELAUNEY, A.J., GATEHOUSE, J.A., GILROY, J., CROY, R.R.D. and BOULTER, D. (1983b). The vicilin gene family of pea (*Pisum sativum* L.): a complete cDNA coding sequence for preprovicilin. Nucl. Acids R., 11, 2367-2380.

146. LYCETT, G.W., CROY, R.R.D., SHIRSAT, A.H. and BOULTER, D. (1984). The complete nucleotide sequence of a legumin gene from pea (*Pisum sativum* L.). Nucl. Acids R., 12, 4493-4506.

147. MACLEOD, D. and BIRD, A. (1982). DNAse I sensitivity and methylation of active versus inactive rRNA genes in *Xenopus* species hybrids. Cell, 29, 211-218.

148. MACLEOD, D. and BIRD, A. (1983). Transcription in oocytes of highly methylated rDNA from *Xenopus laevis* sperm. Nature, Lond., 306, 200-203.

149. MADISON, J.T., THOMPSON, J.F. and MUENSTER, A.E. (1981). Turnover of storage proteins in seeds of soya bean and pea. Ann. Bot., 47, 65-73.

150. MANDAL, R.K. (1984). The organization and transcription of eukaryotic ribosomal RNA genes. Prog. Nucleic Acid Res., 31, 115-160.

151. MANDEL, J.L. and CHAMBON, P. (1979). DNA methylation: organ-specific variations in the methylation pattern within and around ovalbumin and other chicken genes. Nucl. Acids R., 7, 2081-2103.

152. MANIATIS, T., FRITSCH, E.G. and SAMBROOK, J. (1982). Molecular cloning: A laboratory manual. Cold Spring Harbour Laboratory, U.S.A.

153. MANN, M. and SMITH, O. (1977). Specificity of Hpa II and Hae III DNA methylases. Nucl. Acids R., 4, 4211-4221.

154. MARKS, G.E. and DAVIES, D.R. (1979). The cytology of cotyledon cells and the induction of giant polytene chromosomes in *Pisum sativum* . Protoplasma, 101, 73-80.

155. MATTA, N., GATEHOUSE, J.A. and BOULTER, D. (1981). The structure of legumin of *Vicia faba* L. - a reappraisal. J. Exp. Bot., 32, 183-197.

156. MAX, E.E. (1984). New twist to DNA methylation. Nature, Lond., 310, 100.

157. MAX, E.E. (1985). CG doublet difficulties in vertebrate DNA. Nature, Lond., 313, 437-438.

158. McCLELLAND, M. (1981). CG doublet difficulties in vertebrate DNA. Nature, Lond., 313, 437-438.

159. McCLELLAND, M. (1983a). The effect of site specific methylation on restriction endonuclease cleavage (update). Nucl. Acids R., 11, r169-r174.

160. McCLELLAND, M. (1983b). The frequency and distribution of methylatable DNA sequences in leguminous plant protein coding genes. J. Mol. Evol., 19, 346-354.

161. McCLELLAND, M. and IVARIE, R. (1982). Assymmetrical distribution of CpG in an "average" mammalian gene. Nucl. Acids R., 10, 7865-7877.

162. McCLELLAND, M. and NELSON, M. (1985). The effect of site specific methylation on restriction endonuclease digestion. Nucl. Acid R., 13, r201-r207.

163. McGHEE, J.D. and FELSENFELD, G. (1980). Nucleosome structure. A. Rev. Brochem., 49, 1115-1156.

164. McGHEE, J.D. and GINDER, G.D. (1979). Specific DNA methylation sites in the vicinity of the chicken β-globin genes. Nature, Lond., 280, 419-420.

165. McGHEE, J.D., RAU, D.C., CHARNEY, E. and FELSENFELD, G. (1980). Orientation of the nucleosome within the higher order structure of chromatin. Cell, 22, 87-96.

166. McKEE, H.S. ROBERTSON, R.N. and LEE, J.B. (1955). Physiology of pea fruits. 1: The developing fruit. Aust. J. Biol. Sci., 8, 137-163.

167. McKEON, C., OHKUBO, H., PASTON, I. and de CROMBRUGGHE, B. (1982). Unusual methylation pattern of the α2(1) collagen gene Cell, 29, 203-210.

168. McKNIGHT, S.L. and KINGSBURY, R. (1982). Transcriptional control signals of a eukaryotic protein coding gene. Science, 217, 316-324.

169. McMASTER, G.K. and CARMICHAEL, G.C. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. U.S.A., 74, 4835-4838.

170. MEIJILINK, F.C.P.W., PHILIPSEN,. J.N.J., GRUBER, M. and GEERT, A.B. (1983). Methylation of the chicken vitellogenin gene: influence of estradiol administration. Nucl. Acids R., 11, 1361-1373.

171. MERRILL, G.F., HALISCHKA, J.D. and McKNIGHT, S.C. (1984). The enzyme expression in differentiating muscle cells is regulated through an internal segment of the cellular gene. Mol. Cell. Biol., 4, 1777-1784.

172. MESSING, J., GERAGHTY, D., HEIDECKER, G., HU, N-T, KRIDL, J. and RUBENSTEIN, I. (1983). Plant gene structure. In: Kosuge, Meredith and Hollander (edt). Genetic engineering of plants - an agricultural perspective. Plenum. pp.211-227.

173. MILLERD, A. (1975). Biochemistry of legume seed proteins. A. Rev. Pl. Physiol., 26, 53-72.

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

175. MILLERD, A. and WHITFELD, P.R. (1973). Deoxyribonucleic acid and ribonucleic acid synthesis during the cell expansion phase of cotyledon development in *Vicia faba* L. Pl. Physiol., 51, 1005-1010.

176. MILNER, R.J., BLOOM, F.E., LAI, C., LERNER, R.A. and SUTCLIFFE, J.G. (1984). Brain specific genes have identifier sequences in their introns. Proc. Natl. Acad. Sci. U.S.A., 81, 713-717.

177. MINTY, A. and NEWMARK, P. (1983). Gene regulation: new, old and remote controls. Nature, Lond., 288, 210-211.

178. MOLITOR, H., DRAHOUSKY, D. and WALKER, A. (1976). Unmethylated DNA in mouse cells. 1: DNA fibre autoradiography. Biochim. Biophys. Acta., 432, 28-36.

179. MORTON, H., EVANS, I.M., GATEHOUSE, J.A. and BOULTER, D. (1983). Sequence complexity of messenger RNA in cotyedons of developing pea (*Pisum sativum*)seeds. Phytochem., 22, 807-812.

180. NAORA, H. and DEACON, N.J. (1982). Relationship between the total size of exons and introns in protein-coding genes of higher eukaryotes. Proc. Natl. Acad., Sci. U.S.A., 79, 6196-6200.

181. NAVEH-MANY, T. and CEDAR, H. (1982). Topographical distribution of 5-methylcytosine in animal and plant DNA. Mol. Cell Biol., 2, 758-762.

182. NICKOL, J., BEHE, M. and FELSENFELD, G. (1982). Effect of the B-Z

transition in poly (dG-m^{5}dC). Poly (dG-m^{5}dC) on nucleosome formation. Proc. Natl. Acad. Sci. U.S.A., 79, 1771-1775.

183. NORDHEIM, A., PARDUE, M.L., LAFER, E.M., MULLER, A., STOLLER, B.D. and RICH, A. (1981). Antibodies to left-handed Z-DNA bind to interband regions of *Drosophila* polytene chromosomes. Nature, Lond., 294, 417-422.

184. NORDHEIM, A. and RICH. A. (1983). Negatively supercoiled SV40 viral DNA containing Z-DNA segments within the transcriptional enhancer sequences. Nature, Lond., 306, 313-314.

185. OONO, K. and SUGIURA, M. (1980). Heterogeneity of the ribosomal RNA gene cluster in rice. Chromosoma, 76, 85-89.

186. OPIK, H. (1965). The form of nuclei in the storage cells of the cotyledons of germinating seeds of *Phaselous vulgaris* L. Expl. Cell Res., 38, 517-527.

187. OSBORNE, T.B. (1924). The Vegetable Proteins. 2nd ed. Longmans, Green and Co., London.

188. OTT, M-O., SPERLING, L., CASSIO, D., LEVILLIERS, J., SALA-TREPAT, J. and WEISS, M.C. (1982). Undermethylation at the 5' end of the albumin gene is necessary but not sufficient for albumin production by rat hepatoma cells in culture. Cell, 30, 825-833.

189. OTT, M-O., SPERLING, L., HERBOMEL, P., YANIV, M. and WEISS, M.C. (1984). Tissue specific expression is conferred by a sequence from the 5' end of the rat albumin gene. EMBO. J., 3, 2505-1510.

190. PARSLOW, T.G., BLAIR, D.L., MURPHY, W.J. and GRANNER, D.K. (1984). Structure of the 5' ends of the Ig genes: a novel conserved sequence. Proc. Natl. Acad. Scie. U.S.A., 81, 2650-2654.

191. PECH, M., STREECK, R.E. and ZACHAU, H.G. (1979). patchwork structure of a bovine satellite DNA. Cell, 18, 883-893.

192. PICARD, D. and SCHAFFNER, W. (1984). A lymphocyte-specific enhancer in the mouse immunoglobulin Kappa gene. Nature, Lond., 307, 80-82.

193. van der PLOEG, L.H.T. and FLAVELL, R.A. (1980). DNA methylation in the human γ δ β-globin locus in erythroid and non erythroid tissue. Cell, 19, 947-958.

194. POIRIER, G.G., de MARCIA, G., JONGSTRA-BILEN, J., NIEDERGANG, C. and MANDEL, P. (1982). Poly (ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure. Proc. Natl. Acad. Sci. U.S.A., 79, 3423-3427.

195. POLLACK, Y., STEIN, R., RAZIN, A. and CEDAR, H. (1980). methylation of foreing DNA sequences in eukaryotic cells. Proc. Natl. Acad. Sci. U.S.A., 77, 6463-6467.

196. PROUDFOOT, N. (1982). The end of the message. Nature, Lond., 298, 516-517.

197. QURESHI, M., ADAMS, R.L.P. and BURDON, R.H. (1982). Soluble and bound forms of DNA methylase in mouse cell nuclei. Trans. Biochem. Soc., 10, 455-456.

198. RAE, P.M.M. and STEELE, R.E. (1979). Absence of cytosine methylation at CCGG and GCGC sites in the rDNA coding regions and intervening sequences of *Drosophila* and the rDNA of other higher insects. Nucl. Acids R., 6, 2987-2995.

199. RANDALL, P.J., THOMPSON, J.A. and SCHROEDER, H.E. (1979). Cotyledonary storage proteins in *Pisum sativum* . IV: Effects of sulphur, phosphorus, potassium and magnesium deficiencies. Aust. J. Plant. Physiol., 6, 11-24.

200. RAZIN, A. (1984). DNA methylation patterns: formation and biological functions. In: Razin, A., Cedar, H. and Riggs, A.D. (edt), DNA methylation: biochemistry and biological significance. Springer-Gerlag, New York, pp.127-146.

201. RAZIN, A. and CEDAR, H. (1984). DNA methylation in eukaryotic cells. Int. Rev. Cytol., 92, 159-185.

202. RAZIN, A. and FRIEDMAN, J. (1981). DNA methylation and its possible biological role. Prog. Nucleic Acid Res., 25, 33-52.

203. RAZIN, A. and RIGGS, A.D. (1980). DNA methylation and gene function. Science, 210, 604-610.

204. RAZIN, A. and SZYF, M. (1984). DNA methylation patterns: formation and functions. Biochim. Biophys. Acta, 782, 331-342.

205. RAZIN, A., URIELI, S., POLLARK, Y., GRUENBAUM, Y. and GLASER, G. (1980). Studies on the biological role of DNA methylation. 4: mode of methylation of DNA in *E.coli* cells. Nucl. Acids R., 8, 1783-1792.

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

207. REILLY, J.G. and ARUP-SEN, T. (1982). DNA methylation in mouse cells in culture as measured by restriction enzymes. Biochim. Biophys. Acta, 697, 53-59.

208. REUDELHUBER, T. (1984). Upstream and dowsnstream control of eukaryotic genes. Nature, Lond., 312, 700-701.

209. RICH, A. (1983). Right-handed and left-handed DNA: conformational information in genetic material. Cold Spring Harb. Symp. Quant. Biol., 47, 1-12.

210. RIGGS, A.D. (1975). X-inactivation, differentiation and DNA methylation. Cytogenet. and Cell Genet., 14, 9-25.

211. RIGGS, A.D. and JONES, P.A. (1983). 5-methylcytosine, gene regulation and cancer. Adv. Cancer Res., 40, 1-30.

212. ROBERTS, R.J. (1981). Restriction and modification enzymes and their recognition sequences. Nucl. Acids R., 9, r75-r96.

213. ROGERS, J. (1985). Exon shuffling and intron insertion in serine protease genes. Nature, Lond., 315, 458-459.

214. RUSSELL, G.R., WALKER, P.M.B., ELTON, R.A. and SUBAK-HARPEL J.H. Doublet frequency analysis of fractionated vertebrate nuclear DNA. J. Molec. Biol., 108, 1-23.

215. SALSER, W. (1977). Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harb. Symp. Quant. Biol., 42, 985-1002.

216. SALSER, W. and ISAACSON, J.J. (1976). Mutation rates in globin genes: the genetic load and Haldane's dilemma. Prog. Nucleic Acids Res., 19, 205-220.

217. SANTORO, C., COSTANZO, F. NS CILIBERTO, G. (1984). Inhibition of eukaryotic ERNA transcription by potential Z-DNA sequences. EMBO., J., 3, 1553-1559.

218. SCHARPE, A. and van PARIJS, R. (1973). The formation of polyploid cells in ripening cotyledons of *Pisum sativum* L., in relation to ribosome and protein synthesis. J. Exp. Bot., 24, 216-222.

219. SCHIBLER, U., HAGENBUCHLE, O., WELLAUER, P.K. and PITTET, A.C. (1983). Two promoters of different strengths control the transcription of the mouse alpha-amylase gene Amy-1^a in the parotid gland and the liver. Cell, 33, 501-508.

220. SCOTT, N.S., KAUANAGH, T.A. and TIMMIS, J.N. (1984). Methylation of rRNA genes in some higher plants. Plant Sci. Lett., 35, 213-217.

221. SCRIMSHAW, N.S. TAYLOR, L. (1980). Food. Scient. Am., 243, 74-90.

222. SHAW, C.H., WATSON, M.D., CARTER, G.H. and SHAW, C.H. (1984). The right hand copy of the nopaline and Ti-plasmid 25bp repeat is required for tumour formation. Nucl. Acids R., 12, 6031-6041.

223. SHIELDS, R. (1984). Throwing light on plant genes. Nature, Lond., 310, 98.

224. SHIRSAT, A.H. (1984). A gene for legumin: a major storage protein of *Pisum sativum* L. PhD thesis. University of Durham.

225. SINGER, J., ROBERT-EMS, J. and RIGGS, A.D. (1979). Methylation of mouse liver DNA studied by means of restriction enzymes Hpa II and Msp I. Science, 203, 1019-1021.

226. SINGH, L. and JONES, K.W. (1984). The use of heparin as a simple cost-effective means of controlling background in nucleic acid hybridization procedures. Nucl. Acids R., 12, 5627-5638.

227. SKENE, K.G.M. (1970). The gibberellins of developing bean seeds. J. Exp. bot., 21, 236-246.

228. SMITH, D.L. (1971). Nuclear changes in the cotyledons of *Pisum arvense* L. during germination. Ann. Bot., 35, 511-521.

229. SNEIDER, T.W. (1980). The 5'-cytosine in CCGG is methylated in two eukaryotic DNAs and Msp I is sensitive to methylation at this site. Nucl. Acids R., 8, 3829-3840.

230. SOUTHERN, E.M. (1975). Gel electrophoresis of restriction fragments. Meth. Enzym. 68, 152-176.

231. SPENCER, D. and HIGGINS, T.J.V. (1982). Seed maturation and deposition of storage proteins. In: Smith, M. and Grierson, D. (edt). The molecular biology of plant development; Botanical monographs, 18, Blackwell Scientific Publications, pp.306-335.

232. STEIN, R., GRUENBAUM, Y., POLLACK, Y., RAZIN, A. and CEDAR, H. (1982). Clonal inheritance of the patttern of DNA methylation in mouse cells. Proc. Natl. Acad. Sci. U.S.A., 79, 61-65.

233. STURM. K.S. and TAYLOR, J.K. (1981). Distribution of 5-methylcytosine in DNA of somatic and germline cells from bovine tissues. Nucl. Acids R., 9, 4537-4546.

234. SUBAK-SHARPE, H., BURK, R.R., CRAWFORD, L.V. MORRISON, J.M., HAY, J. and KEIR, H.U. (1966). An approach to evolutionary relationships of mammalian DNA viruses through analysis of the pattern of nearest neighbour base sequences. Cold Spring Harb. Symp. Quant. Biol. 31, 737-748.

235. SUTTER, D. and DOERFLER, W. (1980). Methylation of integrated adenovirus type 12 DNA sequence in transformed cells is inversely correlated with viral gene expression. Proc. Natl. Acad. Sci. U.S.A., 77, 253-256.

236. SZYF, M., AVRAHAM-HAETZNI, K., REIFMAN, A., SHLOMAI, J., KAPLAIN, F., OPPENHEIM, A. and RAZIN, A. (1984). DNA methylation pattern is determined by the intracellular level of the methylase. Proc. Natl. Acad. Sci. U.S.A., 81, 3278-3282.

237. SZYF,M., GRUENBAUM, Y., URIELI-SHOVAL, S. and RAZIN, A. (1982). Studies on the biological role of DNA methylation. 5: the pattern of *E. coli* DNA methylation. Nucl. Acids, R., 10, 7247-7259.

238. TANFORD, C. (1961). Statistics of linear polymers. In: The physical chemistry of macromolecules. Wiley, New York, pp.138-150.

239. TAYLOR, J.H. (1979). Enzymatic methylation of DNA: patterns and possible regulatory roles. In: Molecular Genetics. part III: Chromosome structure. Academic Press, New York, pp.89-116.

240. TAYLOR, S.M. and JONES, P.A. (1979). Multiple new phenotypes induced in IOTI/2 and 3T3 cells treated with 5-azacytidine. Cell, 17, 771-779.

241. TAYLOR, S.M. and JONES, P.A. (1982). Mechanism of action of eukaryotic methyltransferase: use of 5-azacytidine containing DNA. J. Molec. Biol., 162, 679-692.

242. THOMAS. P.S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A., 77, 5201-5205.

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

244. TIMMIS, J.N., SINCLAIR, J. and INGLE, J. (1972). Ribosomal RNA genes in euploids and aneuploids of hyacinth. Cell Differ. 1, 335-339.

245. TURNBALL, J.F. and ADAMS, R.L.P. (1976). DNA methylase: purification from ascites cells and the effect of verious substrates on its activity. Nucl. Acids R., 3, 677-695.

246. TYKOCINSKI, M.,. and MAX, E.E. (1984). CG dinucleotide clusters in MHC genes and in 5' demethylated genes. Nucl. Acids R., 10, 4385-4396.

247. UCHIMIYA, H., KATO, H., OHGAWARA, T., HARADA, H. and SUGIURA, M. (1982). Sequence-specific methylation of ribosomal RNA genes contained in the nuclear DNA of tobacco. Pl. Cell Physiol., Tokyo, 23, 1129-1131.

248. VANYUSHIN, B.F. KADYROUA, D.K., KARIMOV, K.K. and BEOZERSKY, A.N. (1971). Minor bases in higher plant DNA. Biokhimiya, 36, 1251-1258.

249. VANYUSHIN, B.F., TKACHEVA, S.G. and BELOZERSKY, A.N. (1970). Rare bases in animal DNA. Nature, Lond., 225, 948-951.

250. VARDIMON, L., GUNTHERET, U. and DOERFLER, W. (1982). *In vitro* methylation of the Bsu RI (5'GGCC-3') sites in the E2a region of adenovirus type 2 DNA does not affect expression in *Xenopus laevis* oocytes. Mol. Cell Biol., 2, 1574-1580.

251. VELICH, A. and ZIFF, E. (1984). Repression of activators. Nature, Lond., 312, 594-595.

252. la VOLPE, A., TAGGART, M., MACLEOD, D. and Bird, A. (1982). Coupled demethylation of sites in a conserved sequence of *Xenopus* ribosomal DNA. Cold Spring Harb. Symp. Quant. Biol., 47, 585-592.

253. VOVIS, G.F., HORIUCHI, K. and ZINDER, N.D. (1974). Kinetics of methylation of DNA by a restriction endonuclease from *E. coli* B. Proc. Natl. Acad. Sci. U.S.A., 71, 3810-3813.

254. WAALWIJK, C. and FLAVELL, R.A. (1978). DNA methylation at a CCGG sequence in the large intron of the rabbit -globin gene: tissue specific variation. Nucl. Acids R., 5,, 4631-4641.

255. WAGNER, I. and CAPESIUS, I. (1981). Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography. Biochim. Biophys. Acta, 654, 52-56.

256. WIGLER, M.H. (1981). The inheritance of methylation patterns in vertebrates. Cell, 24, 285-386.

257. WIGLER, M., LEVY, D. and PERUCHO, M. (1981). The somatic replication of DNA methylation. Cell, 24, 33-40.

258. WILSON, V.L. and JONES, P.A. (1983). Inhibition of DNA methylation by chemical carcinogens *in vitro* . Cell, 32, 239-246.

259. WEISBROD, S. (1982). Active chromatin. Nature, Lond., 297, 289-295.

260. WILKS, A.F., COZENS, P.J., MATTAJ, I.W. and JOST, J-P. (1982). Estrogen induces a demethylation at the 5' end region of the chicken vitellogenin gene. Proc. Natl. Acad. Sci. U.S.A., 79, 4252-4255.

261. WOODCOCK, D.M., ADAMS, J.K., ALLAN, R.G. and COOPER, I.A. (1983). Effect of several inhibitors of ensymatic DNA methylation of different clases of DNA in cultured cell lines. Nucl. Acids R., 11, 489-499.

262. WRIGHT, S., ROSENTHAL, A., FLAVELL, R. and GROSVELD, F. (1984). DNA sequences required for regulated expression of ß-globin genes in murine erythroleukemia cells. Cell, 38, 265-273.

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

264. WYATT, G.R. (1951). Recognition and estimation of 5-methlcytosine in nucleic acids. Biochem. J., 48, 581-584.

265. YAKURA, K. and TANIFUJI, S. (1981). The organization of rRNA genes in *Vicia faba* . Pl. Cell Physiol., Tokyo, 22, 1105-1111.

266. YISRAELI, J. and SZYF,M. (1984). Gene methylation patterns and expression. In: Razin, A., Cedar, H. and Riggs, A.D. (edt). DNA methylation: biochemistry and biological significance. Springer-Verlag, New York, pp.353-378.

267. YUAN, R., BICKLE, T.A., EBBERS, W. and BRACK, C. (1975). Multiple steps in DNA recognition by restriction endonuclease from *E. coli* K. Nature, Lond., 256, 556-560.

268. ZILLIG, W., ZECHEL, K. RABUSSAY, D., SCHACHNER, M., SETHI, V.S., PALM, P., HEIL, A. and SEIFERT, W. (1970). On the role of different subunits of DNA-dependetn RNA polymerase from *E. Coli* in the transcription process. Cold Spring Harb. Symp. Quant. Biol., 35, 47-58.