PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Differential Expression of Genes During Legume Seed Development

J. A. Gatehouse, I. M. Evans, R. R. D. Croy and D. Boulter

Phil. Trans. R. Soc. Lond. B 1986 **314**, 367-384
doi: 10.1098/rstb.1986.0059

Phil. Trans. R. Soc. Lond. B **314**, 367–384 (1986)
Printed in Great Britain

Differential expression of genes during legume seed development

By J. A. Gatehouse, I. M. Evans, R. R. D. Croy and D. Boulter
Department of Botany, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, U.K.

The accumulation of certain proteins specific to those tissues in the developing seeds of legumes represents a system of academic and practical significance in the study of differential gene expression. Besides the simple distinction between 'seed-specific' and 'non-seed-specific' expression of genes, further controls are present in determining the level of expression of a particular gene, and the variations in its expression with cell type, developmental stage and environmental perturbation. There are also genetic factors that lead to variations in the expression of homologous genes between lines or species.

Gene expression can be assayed at the levels of synthesis of specific proteins, level of mRNA species, and transcription of specific genes, and the results of all these assays lead to a broad correlation between events at the level of the gene and protein deposition in the developing seed. This correlation is strong at earlier stages of seed development, but is weaker at later stages. Evidence is presented that control of gene expression occurs both at transcription and by post-transcriptional processes.

Seed protein genes have conserved sequences in their 5′ flanking regions that are specific to gene families, and these are suggested to be involved in transcriptional control of the expression of these genes. Although such sequences are unlikely to be solely responsible for transcription control, there is no strong evidence for changes in DNA methylation or in chromatin conformation being causally related to expression of seed protein genes.

Control of gene expression in developing seeds is considered in terms of a genetically determined, conserved developmental programme, the aim of which is to produce a viable embryo. This programme will allow considerable plasticity in gene expression within constraints prescribed by seed viability. Although it may be possible to understand the immediate controls of seed protein gene expression, present systems are not adequate to study the genes that control the developmental programme. More fundamental investigations will be assisted by mutants that possess altered seed development patterns.

0. Introduction

Although man's major interest in seeds is as a nutritional resource, there are good reasons for using the developing seed to study differential gene expression in plants. The seed contains proteins expressed solely in these tissues, besides proteins common to most or all other plant tissues, and thus its development must involve a high degree of specific control of the expression of well-defined genes. The changes in expression of these genes during seed development can be assayed because seed tissues are easily obtained in large amounts at different developmental stages from plants grown under uniform conditions. Legume seeds have an advantage in that the lower copy numbers of the genes encoding legume seed proteins compared with cereal seed proteins make them easier to analyse at the molecular level. As a consequence, seed protein

genes and their expression have been studied in detail in four agronomically important legumes: *Pisum sativum* (garden pea), *Vicia faba* (field bean), *Glycine max* (soybean) and *Phaseolus vulgaris* (french, kidney or haricot bean). This paper will concentrate on results obtained for *Pisum sativum* reflecting the work carried out at Durham; these results are in general applicable to the other species. A general review of seed protein genes in legumes and cereals, with a full bibliography is given by Croy & Gatehouse (1985); the present paper does not attempt to give a thorough coverage of the extensive literature on the subject.

An academic interest in the ways in which gene expression is regulated in seeds should not, however, be regarded as the only reason for studying this system. The final composition of seed protein is determined by the differential expression of genes during seed development, and thus the nutritional and functional properties of that seed protein are a direct consequence of events at the gene level. In pea, over 80% of the seed protein in a mature seed is the result of the expression of about 20 genes. Attempts to manipulate seed protein composition via genetic engineering, and possibly also via plant breeding programmes, will therefore depend on gaining some insight into why certain genes are expressed in seeds at such high levels.

1. A definition of differential gene expression

Differential gene expression involves a control of the rate at which a gene product is synthesized. In this paper, 'gene expression' will be taken to indicate the synthesis of a protein gene product; the synthesis of RNA from the DNA template of the gene will be referred to as 'transcription'. This initial RNA transcript is then 'processed' to give the cytoplasmic mRNA species, which is 'translated' on the ribosomes to give a polypeptide. All controls occurring after transcription are referred to as 'post-transcriptional'; control at translation is further distinguished as 'translational'. Control of expression can either be expressed as a 'switch' – that is, whether the gene is expressed or not expressed at all – or as a 'regulator' – that is, whether the gene product is synthesized at a high or low rate. This distinction is not logically necessary, but considering control as a dual-level process is a useful model, because the conditions that turn a gene 'on' may not necessarily be the same as those that lead to its expression at a high level. Differential expression of a single gene may be considered as consequence of its cellular environment, which in turn is a consequence of the lineage and spatial position of plant cells. It may also be affected by temporal factors – that is, a developmental control – and by environmental factors such as light, nutrient supply, or hormones. Besides comparing the expression of a single gene under different circumstances, differential gene expression is also understood to refer to a comparison of the expression of different genes under the same conditions; 'relative levels of expression' of those different genes assumes a direct comparison.

2. The accumulation of gene products in developing seeds

(i) *The stages of seed development*

In pea, as in other legumes, seed development goes through six physiological stages (Marinos 1970) after fertilization. From the aspect of seed protein synthesis, it is only the last two of these stages that are of importance, corresponding to the cessation of cell division in the seed tissue and the expansion of the cotyledons, although seed storage proteins have been detected in

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

embryos at stages before the occurrence of differentiation to form cotyledons (Domoney *et al.* 1980). The first half of the cotyledon expansion phase is one of rapid starch synthesis, with the onset of rapid seed protein synthesis lagging behind slightly. However, the synthesis of many seed proteins continues to the end of the active development period (i.e. the period of dry mass gain), whereas starch synthesis is decreased to low levels after two thirds of the active development period. In an oilseed like *Glycine max*, lipid synthesis at the expense of starch takes place during the latter part of seed development. The accumulation of a typical seed specific protein, the seed lectin, in *Pisum sativum* is given in figure 1.

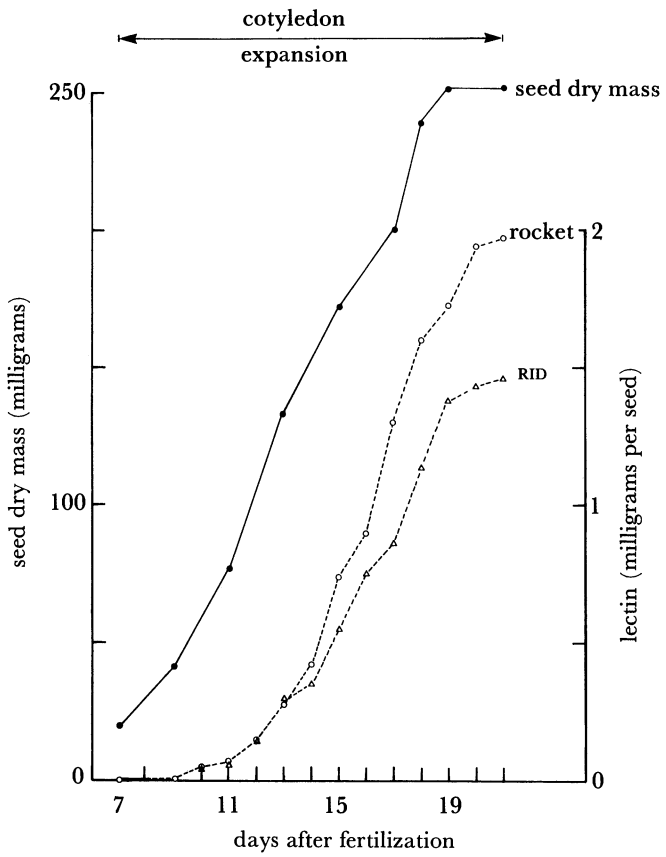

FIGURE 1. Accumulation of a 'typical' legume seed protein, the seed lectin, in developing seeds of *Pisum sativum* var. 'Feltham First'. Plants were grown under controlled environment conditions (Evans *et al.* 1979). ●, Seed dry mass; ○, △, lectin content, determined by rocket immunoelectrophoresis assay (○) or single radial immunodiffusion (△). Rocket immunoelectrophoresis was done at pH 5.0 in TEMED-acetate buffer with carbamylated rabbit (anti-lectin) IgG in the gel. Radial immunodiffusion was done at pH 8.0 in Tris–EDTA–borate buffer with rabbit (anti-lectin) IgG (100 μg ml^{-1}) in the gel. Both systems were calibrated by using purified pea seed lectin as standard. Freeze-dried seed meal samples were extracted at 20 mg ml^{-1} buffer (radial immunodiffusion) or 200 mg ml^{-1} buffer (rocket immunoelectrophoresis).

(ii) *Seed-specific genes*

Those genes that are expressed in the seeds and not in other tissues are designated seed-specific. Such genes are usually exemplified by those encoding polypeptides of the major storage proteins expressed at high levels. This class of genes also possibly includes further genes encoding other seed-specific proteins, e.g. seed lectins and protease inhibitors, which may

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

accumulate to levels similar to or within an order of magnitude of those of storage proteins. However, these highly expressed genes may be atypical of seed-specific genes as a whole, since a further and probably much larger set of genes will also be expressed only in developing seeds, but at relatively low levels, e.g. the cotyledonary α-amylase in pea (Al-Helal 1985). Developing seeds will also contain gene products from genes expressed in these and other tissues, which will be concerned with general metabolic pathways. These genes encode the so-called 'housekeeping' proteins (Boulter 1981) and although many will be expressed at relatively low levels, some at least must be expressed at levels comparable with 'high level' seed-specific genes, e.g. the genes encoding ribulose bisphosphate carboxylase small subunit and other nuclear-encoded chloroplast proteins, and ribosomal proteins. Some of these genes may show higher levels of expression in seeds than in other tissues. However, this not considered to represent 'seed-specific' expression, but is an aspect of control of gene expression by overall metabolic activity; for example, the expression of the ribosomal protein genes will be regulated by the protein synthesis requirements of different tissues.

Some gene families have been shown to contain some seed-specific genes and other non-seed-specific genes, e.g. *Glycine max* lectin, where two genes encode very similar proteins, but where one gene is 'seed-specific' and expressed at fairly high levels, and a second gene is expressed in roots at low level (Vodkin *et al.* 1983). A similar situation may exist in pea.

Homologous genes in different species normally show the same seed-specificity in expression (though not enough data have been assembled to show whether this is invariably true), which suggests that seed-specificity in gene expression evolved with the first seeds and that seed-specific proteins have (or had) a functional role.

(iii) *Relative levels of accumulation of different gene products*

Although there are many factors affecting protein accumulation in developing seeds (*q.v.*), within a single line of a single species, under defined growth conditions, the relative amounts of different seed proteins that accumulate are determined genetically (reviewed in Croy & Gatehouse 1985). Variation even between lines is, however, marked; for example, in *Pisum sativum* the ratio of vicilin to legumin can vary from 0.5 to 4.0 (Casey *et al.* 1982). This variation is even more marked in a species such as *Phaseolus vulgaris*, where the seed lectin can be a major component, or absent, and where lines are known where the major storage protein (vicilin or phaseolin) is a minor component (see figure 2). Each species, has, however, a typical pattern of protein accumulation, e.g. *Glycine max* normally accumulates more legumin than vicilin, whereas *Phaseolus vulgaris* accumulates mainly vicilin. It is apparent from the foregoing that attempting to elucidate the genetic basis for quantitative gene expression across species, or even across lines, is likely to be difficult. In isolated cases, however, simple answers to explain variation may be forthcoming, e.g. inactivation of genes by insertion (Vodkin *et al.* 1983) or deletion of genes (Kreis *et al.* 1983).

A more broadly defined question such as 'What is the difference between genes encoding products accumulated to a relatively low level and genes encoding products normally accumulated to a high level?', seems worth exploring in this context.

(iv) *Tissue specificity in seed-specific expression*

The developing legume embryo contains cotyledons, axis and (in its early stages at least) endosperm, and the pattern of protein accumulation in these tissues differs. The endosperm

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

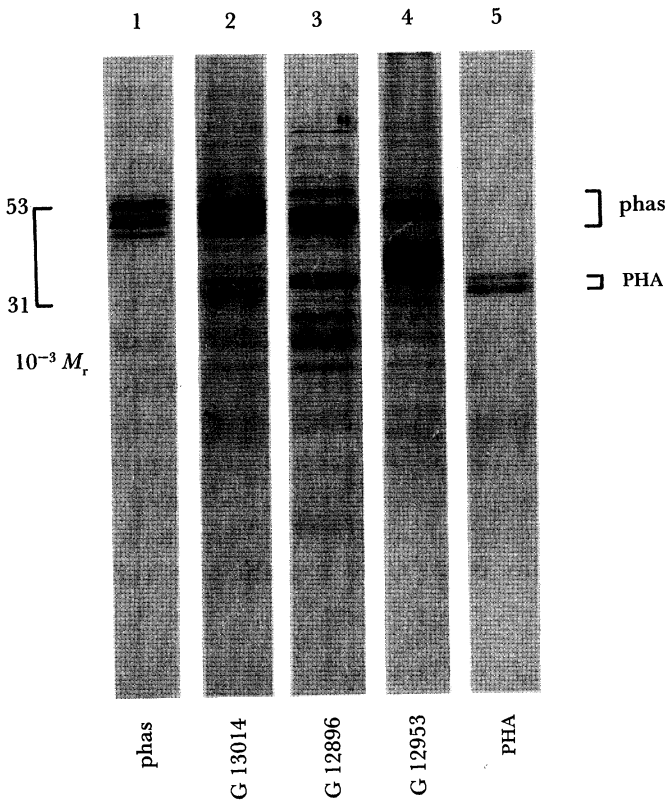

FIGURE 2. Variation in seed protein accumulation among lines of *Phaseolus vulgaris*. Track 1, purified *Phaseolus vulgaris* vicilin (phaseolin, phas); track 5, purified *Phaseolus vulgaris* seed lectin (PHA). Tracks 2–4, total protein extracts from the seeds of three lines of *Phaseolus vulgaris*. Track 3 shows a PHA^- line, track 4 a line with low levels of phaseolin.

is difficult to examine (although this is possibly the site of synthesis of storage proteins detected at very early developmental stages (Domoney *et al.* 1980; Gatehouse *et al.* 1982)) and thus most comparisons have been made between axes and cotyledons. The major seed storage proteins are expressed at lower levels in axes than cotyledons (Meinke *et al.* 1981) and a different pattern of relative expression of the different genes is observed. Unfortunately, the genetic basis of this observation is not clear and presumably reflects the differing environments of embryonic axis and cotyledon cells. Both embryonic axis and cotyledons thus undergo seed-specific protein synthesis during seed development, and then subsequently on germination express different sets of genes.

Within the cotyledons themselves, not all cells accumulate storage proteins to the same level, and the vascular tissues in particular show lower levels of accumulation of storage proteins (N. Harris, personal communication).

(v) *Temporal effects on accumulation*

The accumulation of different gene products with stage of seed development shows differences from gene family to gene family, and even from gene to gene. Although the sigmoid curve of the accumulation pattern in figure 1 is typical for seed-specific proteins, reflecting an increasing rate of synthesis for each gene product at earlier stages of development, which reaches a maximum and then decreases in later stages of development, different gene products

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

show differing times of onset of rapid synthesis, times and periods of maximum synthesis, and times of decline in synthesis. There is thus a differential expression of the encoding genes within the seed-specific expression determined by the developmental programme. For example, in pea, vicilin synthesis increases more rapidly than legumin synthesis in the earlier stages of seed development, but legumin synthesis does not decrease as much as vicilin synthesis towards the end of active seed development. Within the vicilin gene family, certain polypeptides accumulate mainly in the later stages of development (as do the related convicilin polypeptides), whereas others are mainly accumulated in the mid-stages of development.

(vi) *Environmental effects of gene expression*

Growth of plants under zero-sulphur régimes leads to drastic, but well defined changes in gene expression in that the synthesis of sulphur-containing seed proteins such as legumin is very much decreased (Chandler *et al.* 1983). This effect on gene expression is a differential one, since the accumulation of other proteins such as vicilin is unaffected or even increased. Changes in mineral nutrients, osmotic strength of nutrients, or hormone levels have also been shown to affect gene expression (Higgins 1984; Croy & Gatehouse 1985) although not as yet gene transcription.

3. Investigation of the bases for differential gene expression

(i) *Assays of protein synthesis*

Protein deposition can be followed qualitatively by gel analysis of the polypeptides accumulated by developing seeds. This crude approach is sufficient in pea to show clearly the accumulation of vicilin before detectable legumin, and of convicilin with or after legumin (Gatehouse *et al.* 1982). However, an accurate reflection of protein synthesis *in vivo* can only be obtained by pulse-labelling techniques, as exemplified by the experiments of Higgins and Spencer's group (Chrispeels *et al.* 1982*a*, *b*), followed by gel analysis or immunoprecipitation. In our experience, translation of mRNA isolated from developing seeds does not give a totally accurate reflection of protein synthesis *in vivo*, possibly owing to differing stabilities of mRNA species during the purification procedure.

The rates of accumulation of seed proteins are considerable: from various data it may be estimated that in *Pisum sativum* at a stage midway through the synthesis of storage protein, legumin is accumulating at approximately 2.2 mg per seed per day, vicilin at 2.0 mg per seed per day and the seed lectin at 0.05 mg per seed per day. However, these is still a need for accurate quantitative determinations of the rates of synthesis of individual polypeptides during seed development, because data such as those of Gatehouse *et al.* (1982), which measured the accumulation of legumin and vicilin in developing peas by rocket immunoelectrophoresis, used polyclonal antisera that clearly reacted with several different gene products.

These data are needed to correlate with assays of levels of mRNA species to assess the possibility of control of expression at the translational level.

(ii) *Assays of mRNA species*

Assays of mRNA species in developing seed tissue by Northern blot or dot-blot have been used to give semi-quantitative estimates of mRNA levels. There is a clear correlation over developmental stages between mRNA level and polypeptide synthesis for a number of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

polypeptides and probe copy DNA (cDNA) species (see, for example, Chandler *et al.* 1984), although the interpretation of results with storage proteins is complicated by cross-hybridization of cDNAs to more than one mRNA species, the difficulties of distinguishing individual polypeptides discussed earlier, and matching cDNA species to individual polypeptides. Despite these objections the overall interpretation of the results is sound. These assays also show the effects of environmental changes at the mRNA level; for example, under conditions of sulphur deficiency the synthesis of legumin in seeds of *Pisum sativum* is much depressed and the level of legumin mRNA is also drastically decreased to less than 10% of its normal level (Chandler *et al.* 1983). Taken together these data show that in this situation translation level control on the ribosomes is not a determining factor in seed protein gene expression.

Estimates of the relative levels of different mRNA species were more difficult to make owing to varying specific activities of probes and differing efficiencies of hybridization on blots. These difficulties can be overcome by using solution hybridization assays (see, for example, Goldberg *et al.* 1981), but such assays are tedious and require large amounts of mRNA. Absolute quantitation of mRNA species is also difficult. These difficulties may be overcome by using an *in vitro* transcription system such as those employing SP6 or T7 phage polymerases (Melton *et al.* 1984; Davanloo *et al.* 1984) to produce known amounts of synthetic mRNA that can be used as internal standards on blots. Work in progress in our laboratory has given the data shown in figure 3, showing the levels of mRNA species (as a proportion of total RNA) detected by four different cDNA species (three vicilin, one legumin) throughout seed development in pea, by dot-blot assays. The peak levels of vicilin mRNA species all occur about halfway through the period of storage protein synthesis, and levels decline to varying degrees after this. Legumin mRNA levels show a completely different pattern, with the level of legumin mRNA increasing continuously until virtually the end of storage protein synthesis. The peak levels of mRNA species range from approximately 0.2 to 0.7% of total RNA. Although the difference in mRNA levels between legumin and vicilin over seed development is clear, different vicilin mRNA species behave in a broadly similar way. These data do not represent absolute amounts of mRNA species, but assays of total RNA in developing seeds can be used to give a true quantitation. The amount of total RNA in developing pea cotyledons increases almost linearly during the first half of cotyledon expansion (by *ca.* sevenfold from 8 to 14 days after fertilization (d.a.f.)) but remains approximately constant thereafter (at *ca.* 700 μg per cotyledon) until seed maturation (unpublished data); the peak values for mRNA levels therefore represent *ca.* 1.4–5 μg of mRNA species per cotyledon.

(iii) *Gene dosage assays*

The number of genes encoding a particular type of polypeptide accumulated in large amounts will clearly influence what rate of transcription is necessary from these genes to produce the observed gene product. Gene copy numbers have been assayed for a number of legume storage protein gene families, usually by hybridization of cDNA probes to genomic DNA and 'calibrating' the blots with known amounts of cDNA plasmids equivalent to known gene copy numbers. This approach suffers the drawback of variable cross-hybridization between members of gene families, but allows reasonable estimates of copy numbers (Croy *et al.* 1982; Domoney & Casey 1985). Whereas legume storage proteins are normally encoded by small gene families, containing up to ten members, there is no parallel with the cereals, where storage proteins are typically encoded by gene families containing approximately five times as

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

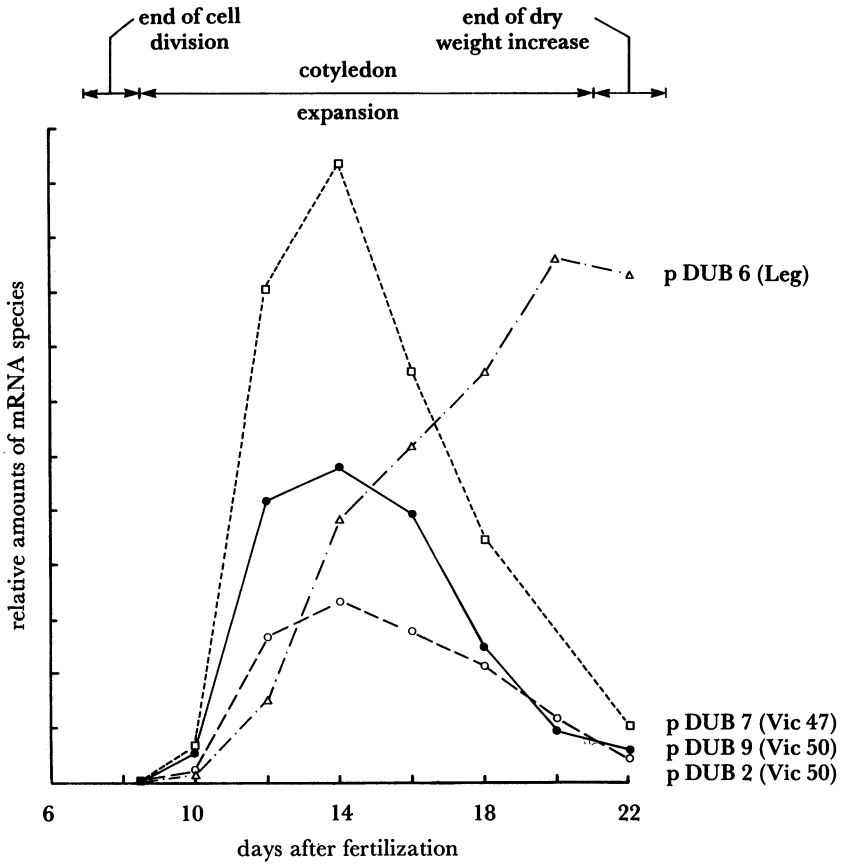

FIGURE 3. Quantitative assays of levels of mRNA species during seed development in *Pisum sativum*. Total RNA was isolated from cotyledons of developing seeds by the method of Chirgwin *et al.* (1979). Known amounts of total RNA samples were spotted on to nitrocellulose filters in a dot-blot apparatus and hybridized to cDNA plasmids as shown, labelled with ^{32}P by nick translation. The dot-blots were calibrated with known amounts of synthetic mRNA species produced by transcription of linearized SP6 plasmids containing the appropriate cDNA inserts with SP6 polymerase (Melton *et al.* 1984). All RNA samples were made up to 10 μg total RNA content by the addition of *E. coli* ribosomal RNA. Hybridization was quantitated by excision of the 'dots' and scintillation counting, after washing the dot-blots.

many genes. Although both polyteny and polyploidy occur during legume seed development (Smith 1971; Marks & Davies 1969), selective gene amplification of seed protein genes does not occur. It is possibly significant that there is a weak correlation between gene copy number and level of expression for seed proteins, e.g. in pea and soya there are only 2–3 lectin genes, whereas in *Phaseolus vulgaris*, where the lectin is a major seed protein, there are 5–6 (Vodkin *et al.* 1983; Hoffman 1984). However, gene copy number is clearly not solely responsible for high expression.

(iv) *Assays of transcription from isolated nuclei* in vitro

By incubating nuclei isolated from developing seeds in a medium containing radioactively labelled nucleotides (usually [α-32P]UTP), transcription to produce labelled RNA molecules can be shown to occur. Most of this RNA production is thought to be 'run-off' transcription with little or no reinitiation taking place, since the time course of incorporation of nucleotide into RNA reaches a plateau rapidly, typically after 15 min (Evans *et al.* 1984). Despite this

observation, transcription assays are often conducted over a much longer time period, typically 30 min to 1 h. It can be shown that the relative amounts of different mRNA species change over the extended assay period (Evans *et al.* 1984), although the significance of this observation is not clear. In our view, results from transcription assays that extend over these longer time periods should be interpreted with caution, as accurate representations of the initial 'run-off' transcription rate. Specific RNA transcripts in the total labelled transcript can be assayed by hybridization to cDNA species immobilized on nitrocellulose, either by hybridization to gel blots or to discs (Gallagher & Ellis 1982). Hybridization to discs offers the advantage of quantitation, although unless this system is very carefully calibrated such quantitation is virtually meaningless owing to background non-specific binding. Hybridization to cDNA (Southern) blots has the advantage of distinguishing specific from non-specific hybridization and does allow different transcripts to be compared with each other by densitometry, although quantitation as a proportion of total transcript is not possible (Evans *et al.* 1985*a*). These assays have confirmed that the transcription of storage protein genes in pea is indeed tissue-specific, because run-off transcripts were detected from nuclei isolated at several different stages of cotyledon expansion, but no transcripts could be detected from leaf nuclei. Typical assays are shown in figure 4. The broad correlation of run-off transcript levels and mRNA levels over seed development is discussed in more detail later, but it is worth noting that these techniques have been used to study gene expression under conditions of sulphur deficiency as well as development; the results show that effects at the transcriptional level are much less than effects on levels of mRNA species (and consequently on protein synthesis) (Evans *et al.* 1985*b*; Beach *et al.* 1985).

(v) *Conclusion and correlations on control of gene expression*

Although existing data point clearly to developmental control of gene expression in seeds being primarily at the transcriptional level, and are even sufficient to show that patterns of protein accumulation, mRNA levels, and transcription are all consistent with each other (Gatehouse *et al.* 1982; Evans *et al.* 1984; unpublished data), a number of problems concerning their detailed interpretation are still apparent.

1. Do levels of mRNA species correlate with rates of protein synthesis? This correlation may be tested by comparisons applied in two different ways: (i) comparing the rate of synthesis of a single protein over different stages of seed development; or (ii) comparing the rates of synthesis of different proteins at a single developmental stage. These comparisons are applied to the synthesis of legumin and vicilin in pea seeds.

(i) For legumin, an increasing rate of protein synthesis over the period 8–13 d.a.f. is succeeded by an approximately constant rate over the period 13–21 d.a.f. The mRNA assays presented above and in previously cited references show a continuous increase for legumin mRNA level over the 8–13 d.a.f. period, in agreement with protein synthesis increasing in rate. The amount of legumin mRNA per cotyledon can be estimated at approximately 25 times as high at 12 d.a.f. as at 10 d.a.f. and 4 times as high at 14 d.a.f. as at 12 d.a.f., whereas the rate of protein synthesis increases by approximately sixfold and twofold. There is thus a reasonable correlation. Over the period of approximately constant synthesis of legumin, amounts of mRNA continue to increase (approximately twice as high at 20 d.a.f. as at 14 d.a.f.) and do not decrease significantly even when legumin synthesis decreases at the end of the dry mass increase in the seeds. These data imply a breakdown in the correlation between protein synthesis and mRNA levels at later developmental stages. The data for vicilin synthesis and mRNA levels

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

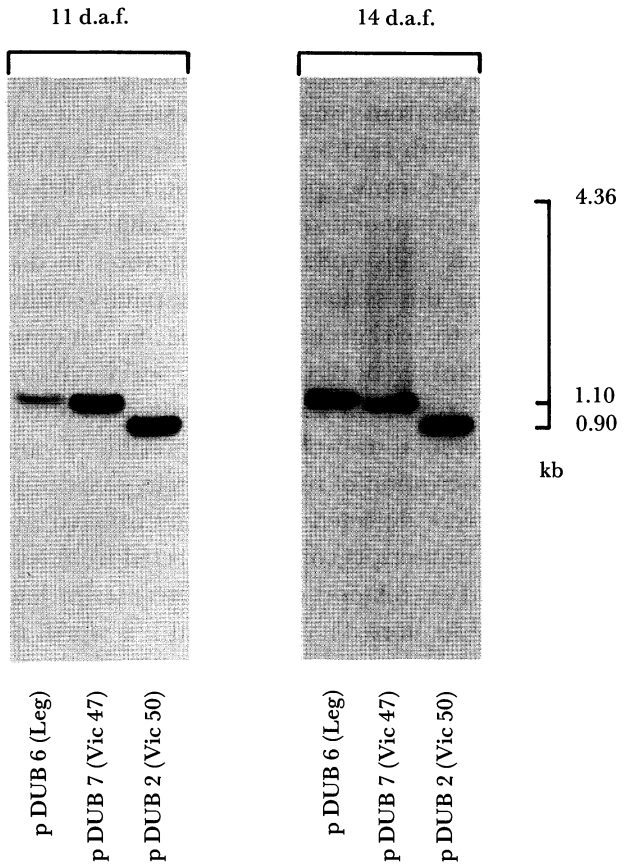

FIGURE 4. 'Run-off' transcription assays of seed storage protein genes in nuclei isolated from developing seeds of *Pisum sativum*. Nuclei isolated at the developmental stages shown were incubated in the presence of [^{32}P]UTP and the labelled RNA transcripts were purified (Evans *et al.* 1984). The transcripts were then hybridized to duplicate blots containing cDNA plasmids as shown (5 μg per track), restricted with the appropriate enzyme (*Bam*HI) to excise the insert (11 d.a.f., 7.9×10^7 c.p.m. transcripts; 14 d.a.f. 5.8×10^7 c.p.m. transcripts). The blots were washed and autoradiographed. Note that at 11 d.a.f. vicilin gene transcription is greater than that for legumin, whereas at 14 d.a.f. the transcriptional activity of the genes is similar. There is negligible background hybridization of the cDNA transcripts to the pBR322 vector DNA at 4.36 kb.

are more difficult to interpret owing to difficulties in measuring protein synthesis rates, because of immunological cross-reaction with convicilin (Croy *et al.* 1980), and in assessing whether the cDNA probes are detecting all or only most of the vicilin mRNA species. However, the increasing rate of vicilin synthesis over the period 8–13 d.a.f. correlates with increasing mRNA levels, and a reasonable qualititative correlation may be made (e.g. protein synthesis rate increases 5 times at 12 d.a.f. over 10 d.a.f., mRNA amount increases 25-fold). Vicilin mRNA levels peak at 14 d.a.f. and decrease subsequently; although this can be related to the disappearance of certain vicilin precursor polypeptides after 16 d.a.f. (Gatehouse *et al.* 1982), it does not correspond to an overall decrease in the rate of vicilin synthesis, which continues at rates comparable to the 12–14 d.a.f. period over the interval 17–20 d.a.f. The correlation between protein synthesis rate and mRNA level is thus again broken at these later developmental stages.

(ii) Comparison of the rates of synthesis of legumin and vicilin at a series of developmental stages may also be made. This shows a reasonable correlation both at early and late developmental stages; at early stages where vicilin synthesis predominates, vicilin mRNA levels

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

are higher (e.g. 10 d.a.f., protein synthesis vic:leg = 8, mRNA levels vic:leg = 9) and at late stages where legumin synthesis predominates, legumin mRNA levels are higher (e.g. 20 d.a.f., protein synthesis leg:vic = 2, mRNA levels leg:vic = 2.5). The correlation at intermediate stages is not as good but does not break down completely; the present data are not adequate to allow a detailed analysis.

2. Do levels of mRNA species correlate with transcription? Although quantitative statements cannot readily be made on this point, because transcription only gives a measure of RNA synthesis whereas the measured mRNA level is a balance between synthesis and degradation, there is agreement in that vicilin transcription is more active than legumin (more than threefold) before the mid-expansion phase corresponding to a faster increase in vicilin mRNA than legumin, whereas legumin transcription is more active (fivefold to tenfold) than vicilin in late expansion, correlating to legumin mRNA levels continuing to increase whereas vicilin mRNA levels decline. At mid-expansion (14 d.a.f.) levels of vicilin and legumin transcription are reproducibly similar, to within a factor of 2 (see figure 4). Equal transcription of vicilin and legumin genes at mid-development may seem anomalous, because the vicilin mRNA level is higher than that of legumin mRNA at this stage (figure 3), but legumin mRNA increases continuously as development proceeds whereas vicilin mRNA peaks at this stage. Similar gene transcription rates are therefore not inconsistent. A failure in complete correlation between transcription and mRNA levels is shown by the data on sulphur deprivation (Evans *et al.* 1985*b*; Beach *et al.* 1985), in that the ratio of legumin mRNA:vicilin mRNA is lower than would be predicted by relative transcriptional activities of the genes.

3. Are all the seed-specific genes switched on and off at the same times? The answer to this question depends on how 'on' and 'off' are interpreted. If we consider them to refer to the potential for expression, then a positive answer is possible, in that the expression of these genes is confined to a set period of embryo development. All seed-specific genes may thus gain the potential for expression at a similar stage in the developmental programme of the embryo, despite evidence showing that, for example, vicilin accumulation, mRNA levels, and run-off transcription all increase faster than legumin, because both gene families are expressed in seed cotyledons at all stages of their development. In this case it is necessary to assume that the genes respond differently to a 'regulator' control (possibly also common) that increases their level of expression. Similarly, the seed-specific genes are not expressed after seed maturation so a common mechanism that causes them all to lose the potential for expression is also possible. The decrease in expression of many seed-specific genes as development proceeds can then be assigned to differing responses to a 'regulator' control decreasing the level of expression. Unfortunately, the existing data can be interpreted equally well on the basis of independent controls for each gene (or gene family).

In conclusion, therefore, in normal seed development the first half of cotyledon expansion is dominated by transcription-driven increase in mRNA levels and protein synthesis, with nuclear events being primarily responsible for differential gene expression. Post-transcriptional events are implicated in modulating expression in so far as the rates of synthesis of seed proteins do not appear to increase as much as the amounts of their corresponding mRNA species. In the later stages of development, there is a transcriptional shut-down of some genes, but overall expression is determined to an increasing extent by post-transcriptional (possibly translational) controls. These latter factors are also the major determinants in alterations in expression produced by environmental perturbations. The data show that differential expression is therefore a complex process.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

4. The DNA sequence basis for control of differential gene expression

(i) *Is there a seed-specific DNA control sequence?*

There is perhaps still an unwritten assumption among many plant molecular biologists that genes expressed in a tissue-specific manner in a given tissue, for example the seed, will all contain somewhere in their 5′ flanking sequences the same short segment of DNA (preferably about 10–12 bases) which will be solely responsible for conferring tissue-specific expression. It is further assumed that this sequence is not species-specific, so that it will work in any plant, that it will work only if 5′ to the gene at the right sort of distance from the transcription start, and so on. The evidence for or against such sequences is at present tenuous, despite the experiments that show that foreign seed-specific genes are expressed in a seed-specific manner in transformed plants. Some evidence in favour of 'seed-specific' sequences will be given later, but the evidence against a single sequence's being applicable to all seed-specific genes may be stated here. Comparisons of 5′ flanking sequences of seed protein genes from different families show little homology exists between them apart from the 'TATA' box region and, in some cases, the regions of sequence near the transcription start. Such homology as is detected often turns out to be an A–T rich sequence, and because the 5′ flanking sequences of these genes are A–T rich, the significance of these homologous regions is dubious.

The long-standing suggestion that certain 5′ flanking sequence segments act as enhancers (Lycett *et al.* 1984), on the basis of sequence homology to known viral enhancer sequences, rests on even more tenuous evidence than 'seed-specific' sequences, owing to a lack of suitable functional assays. Such sequences do not appear to function when assayed by an *in vitro* translation system (Evans *et al.* 1985*a*).

(ii) *The 'legumin box', and similar sequences*

If a 'universal' seed-specific sequence is not present, can putative control sequences specific to gene families be recognized? One gene family where this can be attempted is legumin: 5′ flanking sequences for four pea genes, one *G. max* sequence and one *V. faba* sequence are available. These all show a conserved region of 28 bases (approximately 50% G+C) around −100 relative to the transcription start; the first 14 bases of the 28 are invariant in the genes studied, whereas the remainder has a maximum of five mismatches (figures 5 and 6). This region has been termed the 'legumin box'. Although this region does not affect transcription *in vitro* (Evans *et al.* 1985*a*), its strong conservation suggests a role *in vivo*. Because seed-specific expression of a pea legumin gene containing this sequence in tobacco (J. R. Ellis, unpublished results) has been achieved, its significance can be tested by mutagenesis.

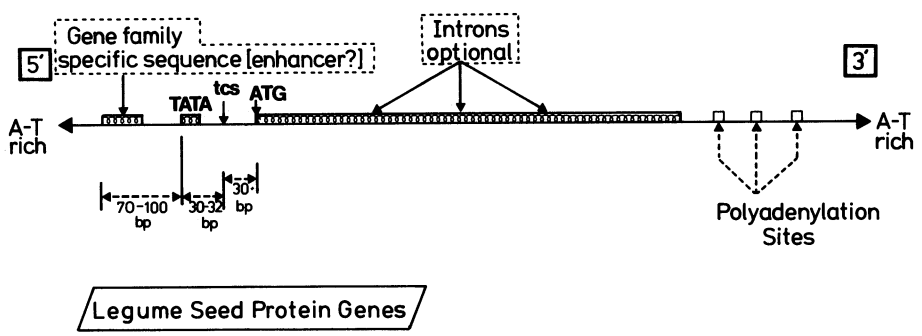

Figure 5. Diagrammatic representation of a legume seed protein gene, showing 5′ functional sequences; tcs, transcription start.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

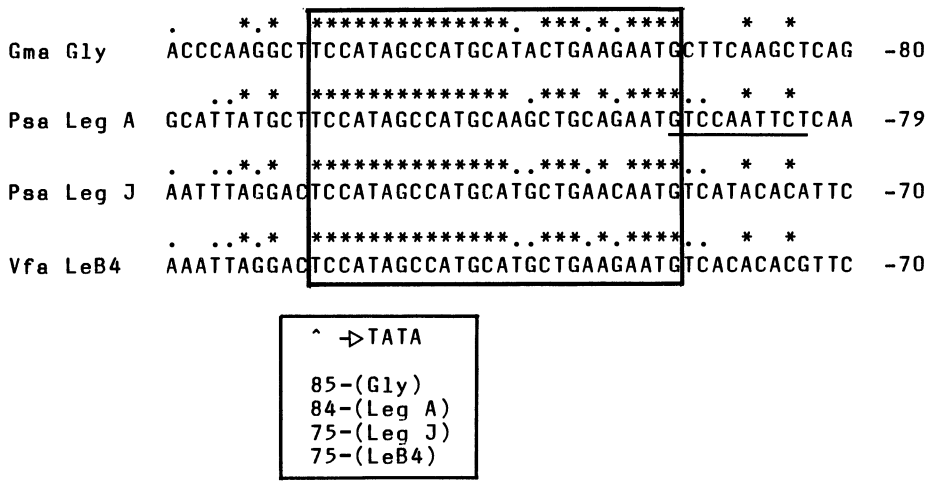

FIGURE 6. The 'legumin box'. The conserved sequence region in the 5′ flanking sequences of four legumin genes is boxed (Gma, *Glycine max*; Psa, *Pisum sativum*; Vfa, *Vicia faba*). The Leg A 'CAAT' box is underlined. Bases the same in all four sequences are marked with an asterisk, the same in three of four sequences with a dot. The conserved bases 5′ and 3′ to the 'legumin box' should be noted; they are possibly part of this putative regulatory element. Origin of sequences: Gma Gly, R. Goldberg (personal communication); Psa Leg A, Lycett *et al.* (1984); Psa Leg J, unpublished data; Vfa LeB4, Baumlein *et al.* (1986).

Comparison of the 5′ flanking sequence of three genes from the vicilin gene family, two from *Pisum sativum* and one from *Phaseolus vulgaris*, reveals the presence of a further conserved sequence region of 42 bases, in this case at approximately −120 to −130 bases relative to the transcription start (figure 7). This may be divided into two parts; a 5′ region with 13/13 conserved bases, which includes a sequence identified as homologous to a viral enhancer sequence (the adenovirus enhancer core element (Lycett *et al.* 1984)) in the pea legumin gene Leg A, and a 3′ region less strongly conserved region (19/27). This sequence is not homologous to the 'legumin box' and is likely to be another example of a conserved upstream sequence

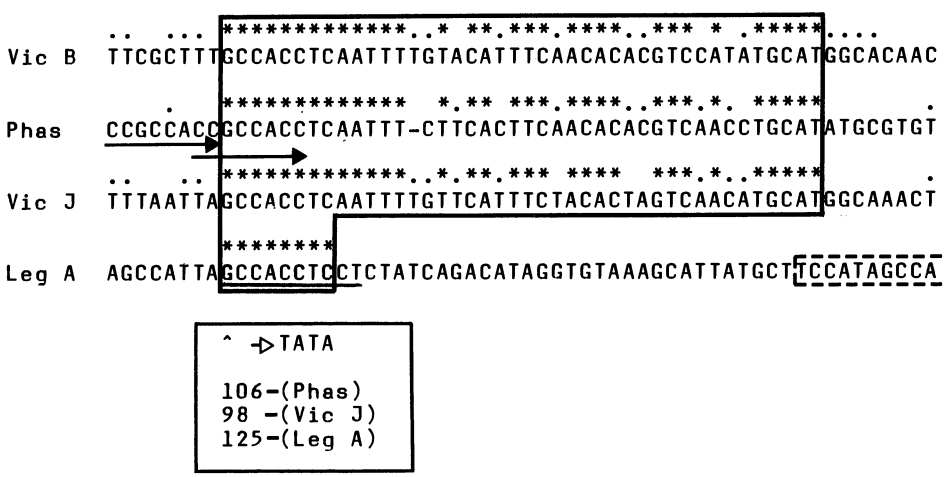

FIGURE 7. The 'vicilin box'. The conserved sequence region in the 5′ flanking sequences of three vicilin genes and a legumin gene is boxed. (Phas, phaseolin; Vic B, J, pea vicilin genes B and J; Leg A, pea legumin gene A). The 'legumin box' in Leg A is indicated by broken lines. Arrows show a repeat of the CCGCCACC motif in phaseolin. The sequence in Leg A identified as homologous to the adenovirus enhancer core element is underlined. Bases the same in the three vicilin sequences and the boxed region of Leg A are marked with an asterisk, the same in two of three vicilin sequences with a dot. Origin of sequences: Phas, Slightom *et al.* (1983); Leg A, Lycett *et al.* (1984); Vic B, J, unpublished data.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

specific to a gene family. Tissue-specific transcriptional enhancer sequences have been identified in the near 5′ flanking regions of other eukaryotic gene families (see, for example, Singh *et al.* 1986) and these plant sequences may be of this nature.

(iii) *Is there an activator (or repressor) protein?*

The existence of putative functionally significant 5′ flanking sequence segments presupposes that something – presumably a protein – interacts with them to affect the transcriptional activity of the associated gene. Few data to support this assertion are, as yet, available for seed protein genes.

(iv) *The role of DNA methylation*

Despite results with other eukaryotic genes (see, for example, Mandel & Chambon 1979) and plant rRNA genes (R. B. Flavell, personal communication), methylation of DNA has not been shown to play any role in control of seed protein gene expression (Waterhouse 1985; Walling *et al.* 1986). Although a tissue-specific demethylation event at an *Msp* I/*Hpa* II site associated with pea legumin genes has been observed, this site is not in the coding sequences of these genes or in the near (−500) 5′ flanking sequence (Waterhouse 1985). Methylation was shown to increase in legumin genes from essentially zero (5′ flanking) to 20% (5′ end of coding sequence) to 45% (3′ end of coding sequence). The assays carried out only measured methylation at CCGG sequences, but results obtained by genomic sequencing of a plant gene (Nick *et al.* 1986) were in agreement with the conclusions, namely that methylation did not show tissue-specific changes in parallel with gene expression.

(v) *The role of changes in chromatin conformation*

Plant genes appear to differ from other eukaryotic genes in their sensitivity to DNase, the method normally used to assay chromatin conformation. Plant genes show no evidence of hypersensitive sites, but active seed protein genes show a high overall DNase sensitivity (Murray & Kennard 1984). There is some evidence that the sensitivity is higher when the gene is expressed (Sawyer 1986). Although it is likely that changes in chromatin do play a role in plant seed gene expression, evidence as strong as that with hypersensitive sites in animal genes has not yet been presented to support this conclusion.

5. A model for control of gene expression

(i) *The developmental programme is of central significance*

The biological function of the developing seed is to provide a dormant embryo with a nutrient reserve which can, on the appropriate environmental stimuli, germinate to produce a new plant. Gene expression, and its control, in the developing seed must thus lead to this end. The expression of individual genes can vary greatly (see above) and many viable null mutants with particular seed proteins absent have been described, leading to the conclusion that it is not the expression of individual genes that is the overall 'aim' of seed development, but rather the production of a viable embryo. That embryo contains differentiated tissue containing cells committed to certain roles, which cannot spontaneously change to another role without going through a dedifferentiation stage; cotyledon cells, for example, cannot spontaneously start to form roots. The development of a plant embryo from the fertilized ovule thus involves a process of cell division and differentiation analogous to the animal embryo

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(and likely to be as complex) that is driven by a controlling genetic developmental programme. When looking at the expression of typical seed protein genes we are examining a result of the developmental programme, namely that in expanding cotyledon cells, when cell division is halted, and these genes are expressed at high level.

Unfortunately, although knowledge about seed protein genes is considerable, there is very little knowledge on the genes that contain the developmental programme for the embryo, and ultimately control the differential expression of seed protein genes.

(ii) *The plasticity of the developmental programme*

Given that the aim of the developmental programme is to produce a viable embryo, there is no reason why changes in the composition of proteins in that embryo cannot be tolerated within certain limits of viability. Thus there is no good biological reason why the ratios of different storage proteins cannot vary between species or varieties, or why different species should not accumulate different proteins. Further, it is reasonable that when environmental perturbations are applied to a developing seed it will respond in such a way to allow a viable embryo still to be produced; if, for example, it is deprived of sulphur, it will respond by cutting the production of sulphur-containing storage proteins without altering the overall process of seed development. The developmental programme must allow this plasticity to allow plants to cope with varying environments. The plasticity of the programme will also allow production of antimetabolic substances in the seeds for protection, or of mechanisms to ensure dormancy, and so on.

The developmental programme is also plastic in that it can vary in duration, and in some species will even allow the next stage in embryo development – germination — to begin before development would normally have finished, given the right conditions. This does not, however, invalidate the view that the function of seed development is to produce a viable embryo, nor is the developmental programme changed in so far as a certain sequence of events must occur. This developmental programme is expressed at the transcriptional level; it is possibly based on a cascade mechanism of gene control. The developmental programme has been conserved in evolution because it is essentially the same in all legumes, so far as can be determined from existing data.

The environmental effects on development that account for some of the observed plasticity occur mainly by modulating the products of transcription, i.e. either by post-transcriptional control or more likely by altering the stabilities of mRNA species in a differential manner. This modulation therefore affects gene expression, but has little or no direct effect on transcription. The result is to allow a developmental programme that cannot be altered by the environment (which would lead to non-viable embryos) but is plastic in those aspects that do not affect embryo viability.

(iii) *A simplistic view?*

Is it therefore necessary to understand development in order to understand differential expression of genes in developing seeds? Yes, if it is to be understood fully; but it is also true that a simplistic view, which attempts to find parts of seed protein genes and their flanking sequences that appear to be necessary for seed-specific expression in functional assays, will go some way towards understanding the basis of differential expression. Such a simplistic view may also allow seed composition to be manipulated in transformed plants to mankind's practical benefit.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

As has been indicated, differential gene expression in seed development is manifested not only as an 'on–off', but also in a range of subtle changes in the level of expression, which may not be explicable in terms of simple differences in DNA sequences. Thus variation in expression of specific genes between lines is often involved with a genetic locus distinct from the structural gene locus, and which has effects on many different aspects of metabolism, e.g. the r_a locus in pea, which affects legumin gene expression but has its primary effect on carbohydrate metabolism (Davies 1980; Domoney & Davies 1983). The basis for differential gene expression in this type of system is likely to be complex, although even in this case a simplistic approach at the level of the seed protein genes may account for the changes observed in their expression. This type of variation in gene expression may yield results that are more relevant to 'regulation' of gene expression than would a study of lines where genes are totally inactivated by insertion or deletion. The production and identification of natural or induced variant or mutant lines with quantitative changes in gene expression is thus likely to be of value, especially if the alteration is specific to small numbers of genes. An example of such a mutant is given in figure 8, which shows data for a vicilin-deficient pea line with a normal complement of vicilin genes but reduced expression of a subfamily of those genes.

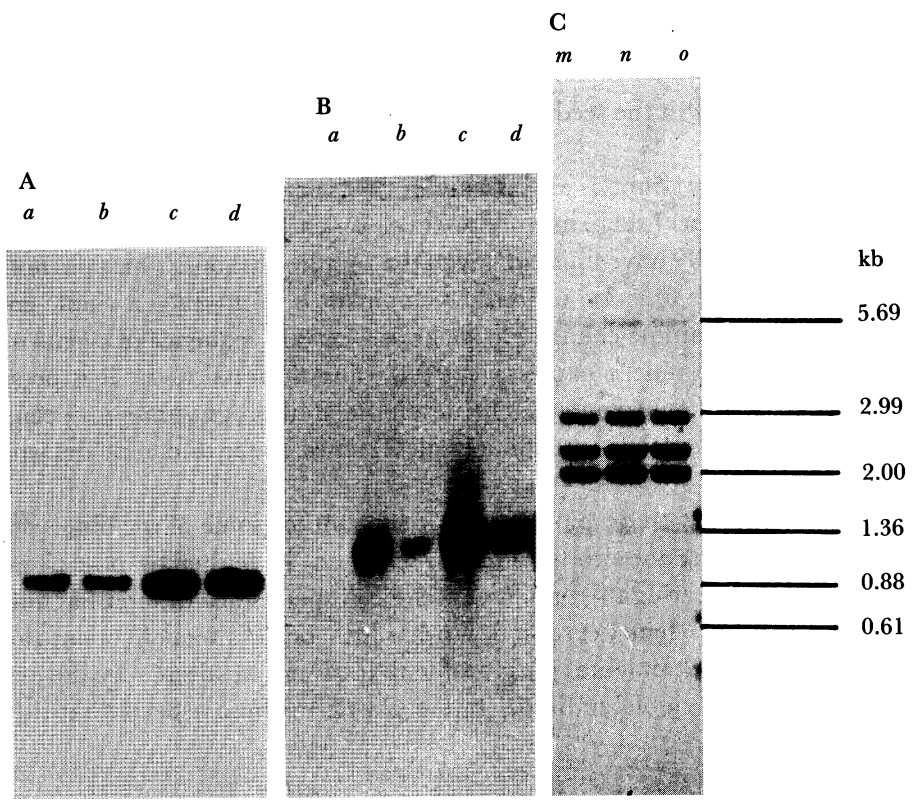

FIGURE 8. A mutant pea line showing altered expression of a subset of vicilin genes. (A, B) Tracks *a–d*: Northern blot of total RNA extracted from cotyledons of developing seeds of *Pisum sativum* parental line 200 (tracks *a* and *c*) and mutant line 5478 (tracks *b* and *d*) at 9 d.a.f. (tracks *a* and *b*) and 11 d.a.f. (tracks *c* and *d*). The blot was hybridized to labelled cDNA plasmid pDUB7 (vicilin 47000 M_r polypeptide) (A, *a–d*) or pDUB2 (vicilin 50000 M_r in polypeptide) (B, *a–d*). Levels of mRNA species detected by pDUB2 are lower in the mutant line than in the parent, but the levels of mRNA species detected by pDUB7 are similar. (C) Tracks *m–o*: Southern blot of genomic DNA, extracted from leaves of *Pisum sativum* lines 'Feltham First' (track *m*), 200 (track *n*) and 5478 (track *o*); DNA was restricted with *Taq*I and the blot was hybridized to labelled cDNA plasmid pDUB2. Note that despite the reduced levels of mRNA species detected by pDUB2 in line 5478 compared with line 200, the genome of the mutant contains an identical set of restriction fragments hybridizing to this probe, and, by implication, a similar number of genes (Waterhouse 1985).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The 'simplistic' view does not, however, lead to an understanding of the genes that control the developmental programme. Investigations at this more fundamental level will require the production of a different type of mutation to those described above, i.e. mutations that affect the developmental programme itself.

(iv) *Conclusions*

The differential expression of genes during legume seed development is well documented (although there is still a need for accurate quantitative assays of the expression of individual genes), but is not yet understood in any detail. There are DNA sequences in the 5′ flanking regions of seed protein genes that confer seed-specific expression, but it is not clear whether these sequences are specific to particular genes or whether consensus sequences applicable to many different genes exist. The A–T rich nature of the flanking regions and introns of seed protein genes suggests that DNA base composition has a role, but this is as yet an unproven suggestion. Changes in methylation of DNA do not appear to be causative in control of expression. Seed protein genes appear to be located in uncondensed regions of chromatin, but do not have hypersensitive sites for DNase. The differential expression of seed protein genes is under the primary control of a genetic developmental programme, which exerts its effects at the transcriptional level; expression can be modified by environmental factors, which operate mainly at post-transcriptional stages.

The authors thank their co-workers, who have provided material for this paper. These include R. N. Waterhouse, R. M. Sawyer, D. Bown, M. Levasseur, A. H. Shirsat and A. M. R. Gatehouse. They also express their gratitude to N. Ellis, R. Casey, and other members of the Seed Protein Group, John Innes Institute, who supplied genes Leg J, Vic B and Vic J. Support from the A.F.R.C. and S.E.R.C. is also gratefully acknowledged.

References

Al-Helal, A. 1985 The use of isoenzymes in the study of germination, development and breeding of legumes. Ph.D thesis, University of Durham.

Baumlein H., Wobus, U., Pustell, J. & Kafatos, F. C. 1986 The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. *Nucl. Acids Res.* **14**, 2707–2720.

Beach, L. R., Spencer, D., Randall, P. J. & Higgins, T. J. V. 1985 Transcriptional and post-transcriptional regulation of storage protein gene expression in sulphur-deficient pea seeds. *Nucl. Acids Res.* **13**, 999–1013.

Boulter, D. 1981 Biochemistry of storage protein synthesis and deposition in the developing legume seed. *Adv. bot. Res.* **9**, 1–31.

Casey, R., Sharman, J. E., Wright, D. J., Bacon, J. R. & Guldager, P. 1982 Quantitative variability in *Pisum* seed globulins; its assessment and significance. *Qual. Plant. Plant Fds hum. Nutr.* **31**, 333–346.

Chandler, P. M., Higgins, T. J. V., Randall, P. J. & Spencer, D. 1983 Regulation of legumin levels in developing pea seeds under conditions of sulphur deficiency. Rates of legumin synthesis and levels of legumin mRNA. *Pl. Physiol.* **71**, 47–54.

Chandler, P. M., Spencer, D., Randall, P. J. & Higgins, T. J. V. 1984 Influence of sulfur nutrition on developmental patterns of some major pea seed proteins and their mRNAs. *Pl. Physiol.* **75**, 651–657.

Chirgwin, J. M., Przybyla, A. E., Macdonald, R. J. & Rutter, W. J. 1979 Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry, Wash.* **18**, 5294–5299.

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

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

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

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

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

Croy, R. R. D., Lycett, G. W., Gatehouse, J. A., Yarwood, J. N. & Boulter, D. 1982 Cloning, characterization and sequence analysis of cDNAs encoding *Pisum sativum* L. (pea) storage protein precursors – legumin. *Nature, Lond.* **295**, 76–79.

Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. 1984 Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2035–2039.

Davies, D. R. 1980 The r_a locus and legumin synthesis in *Pisum sativum*. *Biochem. Genet.* **18**, 1207–1219.

Domoney, C. & Casey, R. 1985 Measurement of gene number for seed storage proteins in *Pisum*. *Nucl. Acids Res.* **13**, 687–699.

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

Domoney, C., & Davies, D. R. 1983 Storage proteins in round and wrinkled peas *in vivo* and *in vitro*. In *Perspectives for peas and lupins as protein crops* (ed. R. Thompson & R. Casey), pp. 256–271. Martinus. The Hague: Nijhoff.

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

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

Evans, I. M., Gatehouse, J. A. & Boulter, D. 1985*b* Regulation of storage-protein synthesis in pea (*Pisum sativum* L.) cotyledons under conditions of sulphur deficiency. *Biochem. J.* **232**, 261–265.

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

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

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

Goldberg, R. B., Hoschek, G., Ditta, G. S. & Breidenbach, R. W. 1981 Developmental regulation of cloned superabundant embryo mRNAs in soybean. *Devl. Biol.* **83**, 218–231.

Higgins, T. J. V. 1984 Synthesis and regulation of major proteins in seeds. *A. Rev. Pl. Physiol.* **35**, 191–221.

Hoffman, L. M. 1984 Structure of a chromosomal *Phaseolus vulgaris* lectin gene and its transcript. *J. molec. appl. Gen.* **2**, 447–453.

Kreis, M., Shewry, F. R., Forde, B. G., Rahman, S. & Miflin, B. J. 1983 Molecular analysis of a mutation conferring the high-lysine phenotype on the grain of barley (*Hordeum vulgare*). *Cell* **34**, 161–167.

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

Mandel, J. L. & Chambon, P. 1979 DNA methylation: organ-specific variations in the methylation pattern within and around ovalbumin and other chicken genes. *Nucl. Acids Res.* **7**, 2081–2103.

Marinos, N. G. 1970 Embryogenesis of the pea (*Pisum sativum*). I. The cytological environment of the developing embryo. *Protoplasma* **101**, 261–279.

Marks, G. E. & Davies, D. R. 1969 The cytology of cotyledon cells and the induction of giant polytene chromosomes in *Pisum sativum*. *Protoplasma* **101**, 73–80.

Meinke, D. W., Chen, J. & Beachy, R. N. 1981 Expression of storage protein genes during soybean (*Glycine max* cv. Provar) seed development. *Planta* **153**, 130–139.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. 1984 Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucl. Acids Res.* **12**, 7035–7056.

Murray, M. G. & Kennard, W. C. 1984 Altered chromatin conformation of higher plant gene phaseolin. *Biochemistry, Wash.* **23**, 4225–4232.

Nick, H., Bowen, B., Ferl, R. J. & Gilbert, W. 1986 Detection of cytosine methylation in the maize alcohol dehyrogenase gene by genomic sequencing. *Nature, Lond.* **319**, 243–246.

Sawyer, R. M. 1986 Isolation of a vicilin gene from pea (*Pisum sativum* L.) and nuclease sensitivity of seed storage protein genes in pea chromatin. Ph.D. thesis, University of Durham.

Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. 1986 A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature, Lond.* **319**, 154–158.

Slightom, J. L., Sun, S. M. & Hall, T. C. 1983 Complete nucleotide sequence of a french bean storage protein gene: phaseolin. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1897–1901.

Smith, D. L. 1971 Nuclear changes in the cotyledons of *Pisum arvense* L. during germination. *Ann. Bot.* **35**, 511–521.

Vodkin, L. O., Rhodes, P. R. & Goldberg, R. B. 1983 cA lectin gene has the structural features of a transposable element. *Cell* **34**, 1023–1031.

Walling, L., Drews, G. N. & Goldberg, R. B. 1986 Transcriptional and posttranscriptional regulation of soybean seed protein mRNA levels. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2123–2127.

Waterhouse, R. N. 1985 The control of expression of storage protein genes in *Pisum sativum* L. Ph.D. thesis, University of Durham.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

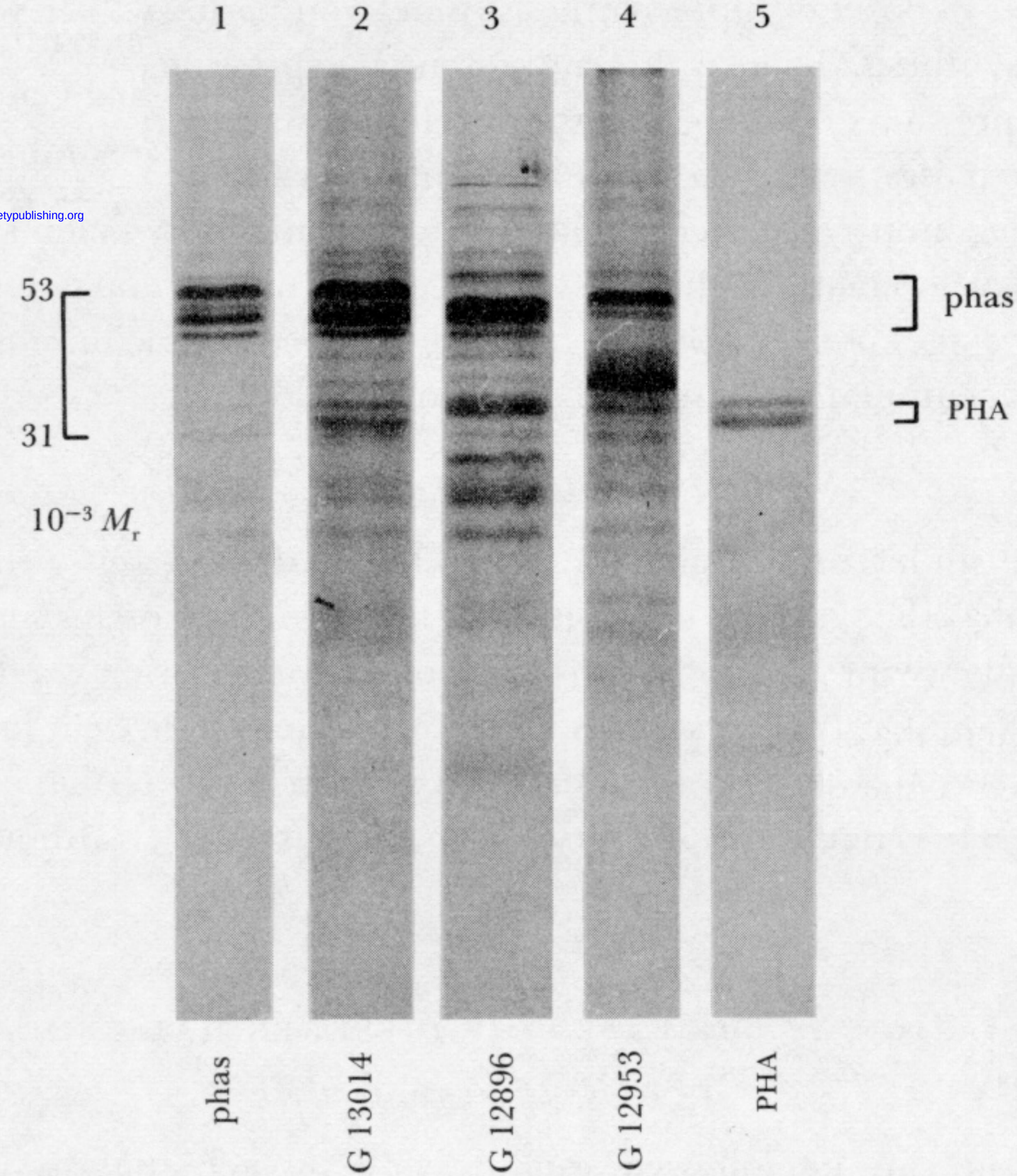

FIGURE 2. Variation in seed protein accumulation among lines of *Phaseolus vulgaris*. Track 1, purified *Phaseolus vulgaris* vicilin (phaseolin, phas); track 5, purified *Phaseolus vulgaris* seed lectin (PHA). Tracks 2–4, total protein extracts from the seeds of three lines of *Phaseolus vulgaris*. Track 3 shows a PHA^- line, track 4 a line with low levels of phaseolin.

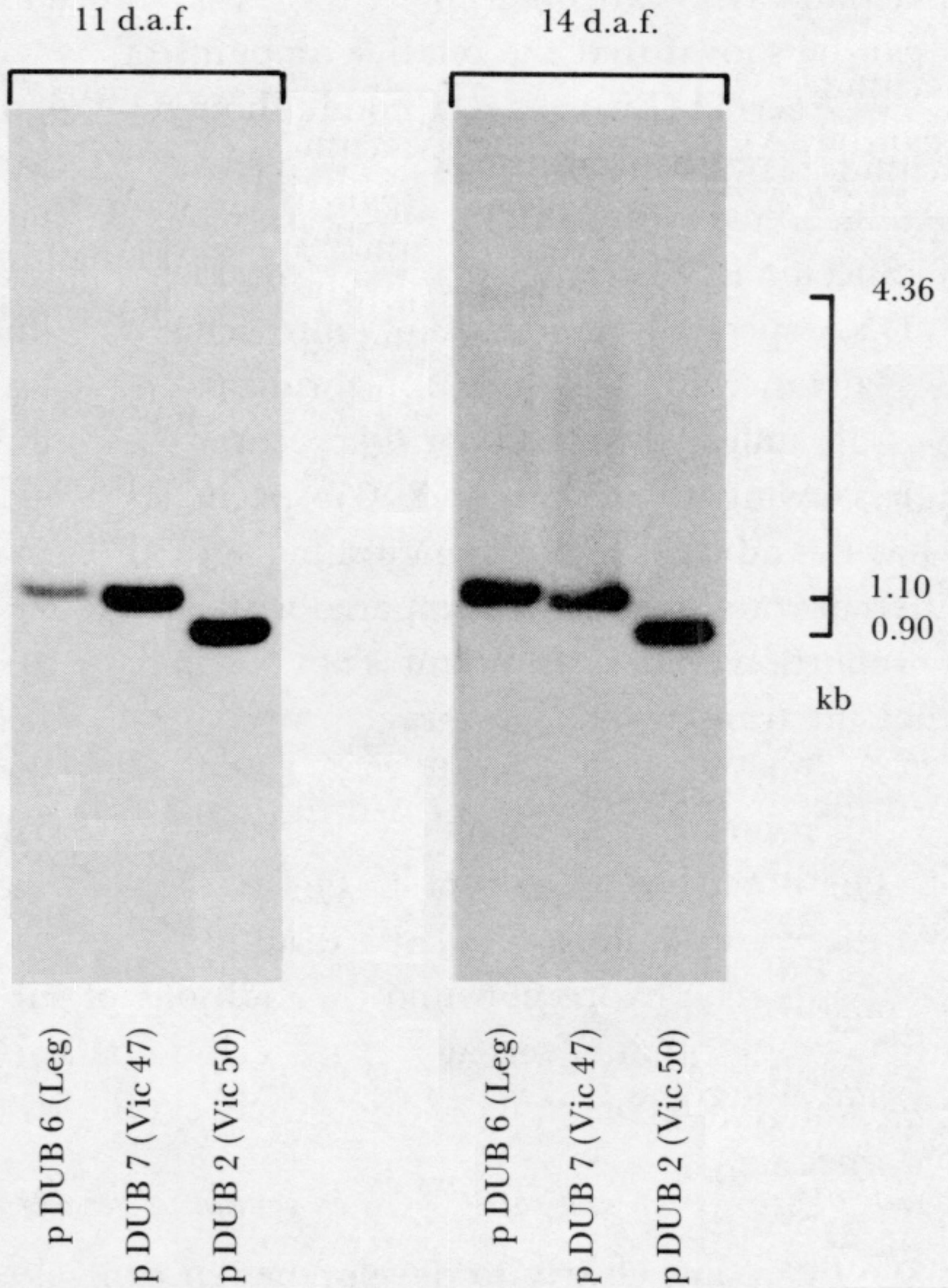

FIGURE 4. 'Run-off' transcription assays of seed storage protein genes in nuclei isolated from developing seeds of *Pisum sativum*. Nuclei isolated at the developmental stages shown were incubated in the presence of [^{32}P]UTP and the labelled RNA transcripts were purified (Evans *et al.* 1984). The transcripts were then hybridized to duplicate blots containing cDNA plasmids as shown (5 μg per track), restricted with the appropriate enzyme (*Bam*HI) to excise the insert (11 d.a.f., 7.9×10^7 c.p.m. transcripts; 14 d.a.f. 5.8×10^7 c.p.m. transcripts). The blots were washed and autoradiographed. Note that at 11 d.a.f. vicilin gene transcription is greater than that for legumin, whereas at 14 d.a.f. the transcriptional activity of the genes is similar. There is negligible background hybridization of the cDNA transcripts to the pBR322 vector DNA at 4.36 kb.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

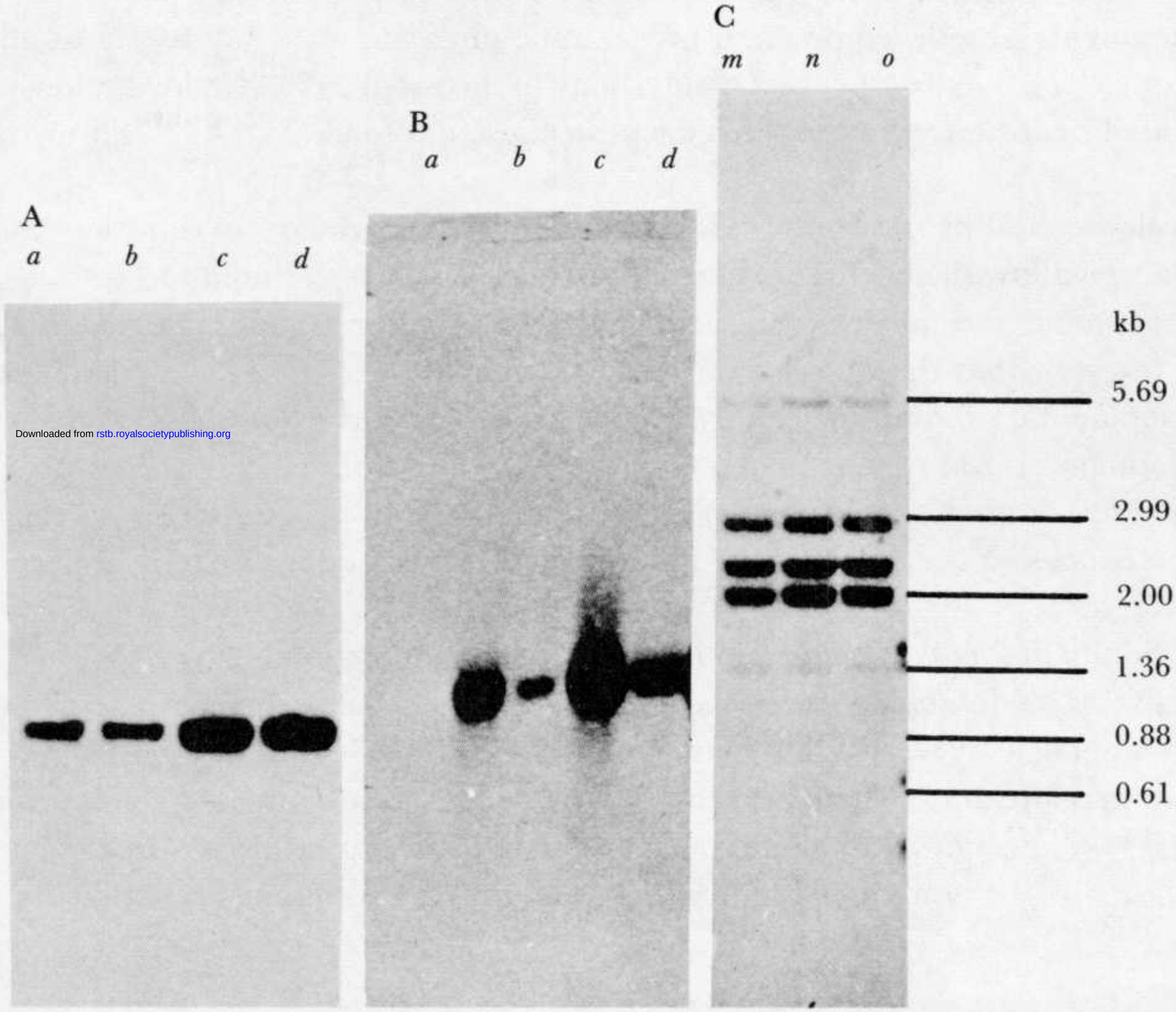

FIGURE 8. A mutant pea line showing altered expression of a subset of vicilin genes. (A, B) Tracks *a–d*: Northern blot of total RNA extracted from cotyledons of developing seeds of *Pisum sativum* parental line 200 (tracks *a* and *c*) and mutant line 5478 (tracks *b* and *d*) at 9 d.a.f. (tracks *a* and *b*) and 11 d.a.f. (tracks *c* and *d*). The blot was hybridized to labelled cDNA plasmid pDUB7 (vicilin 47000 M_r polypeptide) (A, *a–d*) or pDUB2 (vicilin 50000 M_r in polypeptide) (B, *a–d*). Levels of mRNA species detected by pDUB2 are lower in the mutant line than in the parent, but the levels of mRNA species detected by pDUB7 are similar. (C) Tracks *m–o*: Southern blot of genomic DNA, extracted from leaves of *Pisum sativum* lines 'Feltham First' (track *m*), 200 (track *n*) and 5478 (track *o*); DNA was restricted with *Taq*I and the blot was hybridized to labelled cDNA plasmid pDUB2. Note that despite the reduced levels of mRNA species detected by pDUB2 in line 5478 compared with line 200, the genome of the mutant contains an identical set of restriction fragments hybridizing to this probe, and, by implication, a similar number of genes (Waterhouse 1985).