Molecular biology of the Oleosin gene family of Brassica napus

Edwards, Eira Wyn

Molecular Biology of the Oleosin Gene Family of *Brassica Napus*

by

Eira Wyn Edwards

This thesis was submitted to

The Department of Biological Sciences

University of Durham

In accordance with the requirements for the degree of

Doctor of Philosophy

June 1991

DECLARATION

Unless otherwise stated, I declare that all the experiments described in this thesis are my own original work, which has not been submitted for a degree at any other university.

This is what the Lord says:

"Let not the wise man boast of his wisdom
or the strong man boast of his strength
or the rich man boast of his riches,
but let him who boasts boast about this:
That he understands and knows me.
that I am the Lord, who exercises kindness,
justice and righteousness on earth
for in these I delight," declares the Lord

Jeremiah 9:23,24

For God so loved the world that He gave His one and only Son,
that whoever believes in Him shall not perish but have eternal life.

John 3:16

Acknowledgements

Many thanks to:

My supervisors Dr Charlie Shaw and Dr Denis Murphy.

Mr Terry Gibbons for encouragement and expert technical help.

Dr Andy Ryan for screening of the λgt11 cDNA library and construction of the λEMBL3 genomic library - without whom this project would not have been possible.

Dr Kieran Elbrough for technical advice and help.

Dr Jane O'Sullivan for construction of the λgt11 cDNA library.

Mr Ian Cummins for supplying polyclonal antibodies to oleosin.

Mr James Keddie for subcloninig and sequencing the Nap 3 oleosin genomic clone.

Miss Julia Bryden for help with DNA sequencing.

To my parents and friends who have kept me going with help and encouragement. Special thanks Michael Patterson.

Abbreviations

ABA	-	Abscisic acid
bp	-	Base pairs
Kb	-	Kilobase pairs
3' UTR	-	3' Untranslated region
SDS PAGE	-	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Molecular Biology of the Oleosin Gene Family of *Brassica napus*

by

Eira Wyn Edwards

Abstract

Oilseed rape (*Brassica napus*) is a valuable oil producing crop plant. Oil is produced by developing seeds and constitutes 45% of the dry seed weight. Storage oil is deposited as triacylglycerol in spherical structures called oil bodies. These are coated by a phospholipid monolayer interspersed with proteins called oleosins, which serve as emulsifiers, retaining the integrity of individual oil bodies.

In this project the oleosin gene family of *B. napus* was studied. Sequence information was obtained on two oleosin cDNA clones and a genomic clone was also isolated. The sequences for oleosins from *B. napus* were compared with those isolated from other plant species.

Preliminary expression studies were carried out, relating the timing of oleosin mRNA and protein production to the stages of embryo development. These results give an indication of the importance of transcriptional and post-transcriptional processes in the control of oleosin expression.

CONTENTS

CHAPTER 2 - MATERIALS AND METHODS

CHAPTER 3 - RESULTS

LIST OF FIGURES

LIST OF TABLES

RESULTS

CHAPTER 1 - INTRODUCTION

1.1 OLEOSINS IN DEVELOPING SEEDS

1.1.1 SEED DEVELOPMENT

Seeds contain everything needed to start a new generation: the genetic information and potential for growth within the embryo tissue and the storage reserves to begin the process. The processes involved in seed formation take place over a series of many weeks and are preceeded by flowering, anthesis and pollination. In many species the final event in embryogenesis is the dessication of the seed tissues. The dry seed in its protective coat can remain viable, in some cases for many years. Embryo development takes place simultaneously with the deposition of storage reserves.

The three main storage reserves are protein, lipid and starch. Lipid and starch are necessary for gluconeogenesis. Storage lipid consists of triacylglycerol which is synthesised by the Kennedy pathway (Gurr, 1980). The control of the fatty acid composition of storage lipid is dependent on three pathways:

1] Fatty acid synthesis in the chloroplast by repeated condensation of malonyl CoA molecules with an acyl group attached to acyl carrier protein. The enzymes responsible for fatty acid synthesis are collectively called the fatty acid synthase system (Stumpf, 1987).

2] Fatty acid desaturation which is carried out on fatty acids esterified to phosphatidyl choline (Stymne and Appelqvist, 1978). A process of acyl exchange returns the fatty acids to the acyl CoA pool following desaturation (Stymne, et al., 1983).

3] Condensation of fatty acids to a glycerol backbone to form triacylglycerol. These reactions are catalysed by a series of acyltransferase enzymes.

Enzyme	Position
Glycerol-3-phosphate acyltransferase	sn-1 (Ichihara, 1984)
Lysophosphatidic acid acyltransferase	sn-2 (Stymne and Stobart, 1987)
Diacylglycerol acyltransferase	sn-3 (Ichihara and Noda, 1982)

The fatty acid preference of the individual acyltransferase enzymes together with the availability of fatty acids in the cellular pool determines the fatty acid composition of the storage triacylglycerol.

Lipid is stored in the form of triacylglycerol within oil bodies. These are enclosed by a phospholipid monolayer. The most abundant protein inserted into this monolayer is called oleosin. Several isoforms of oleosin exist in the plant species investigated so far. Oil bodies observed under the EM range from 0.1-1μm in diameter.

All seeds contain storage protein which serves as a source of amino acids for synthesis of cell proteins during the initial stages of seedling growth. Storage protein is deposited in protein bodies, which consist of a core of closely packed storage protein surrounded by a single bilayer membrane. The organelles have been observed using electron microscopy and range from 1-20μm in diameter .

Seed storage proteins have been classified according to the ease with which they can be extracted in water based solvents (Higgins, 1984):

Albumins:	water
Globulins:	salt solutions
Prolamines:	aqueous alcohols
Glutelins:	acid or alkali

The storage proteins of *B. napus* are napin and cruciferin. In mature seeds of *B. napus*, 45% of the seed dry weight consists of storage triacylglycerol. Protein constitutes 25% of the dry weight.

The proportions of the different proteins in the seed are as follows (Murphy and Cummins, 1989):

Protein	% of total protein
Cruciferin	40-50
Napin	20
Oleosin	20

Napin is a 2S protein (Josefsson, et al., 1987) and cruciferin a 12S protein (Simon, et al., 1985). Napin is expressed by a gene family consisting of at least 16 genes (Scofield and Crouch, 1987). Cruciferin is a 12S globulin for which several cDNA clones have been isolated (Rödin, et al., 1990), as well as genomic clones (Ryan, et al., 1989).

1.1.2 POSSIBLE FUNCTIONS OF OLEOSINS

Oleosins are found in a wide variety of seeds where oil is a storage product (Qu, et al., 1986); the protein products have heterogeneous molecular weights (Tzen, et al., 1990). Oleosins are located on the surface of oil body membranes and have a role in maintaining the integrity of the oil body within the seed.

Derived amino acid sequences of the KD16 oleosin cDNA from maize have enabled structure predictions to be carried out (Vance and Huang, 1987). The protein is anchored in the oil body via the central hydrophobic domain, which is thought to protrude into the triacylglycerol core. The hydrophilic N-terminal region must extend outside the oil body into the more hydrophilic surroundings. At the C-terminus, the amino acid sequence has the ability to form an amphipathic α-helix, one face of which is hydrophobic and the other hydrophilic. This amphipathic helix could rest on the surface of the oil body with the hydrophobic face in contact with the triacylglycerol core and the hydrophilic face interacting with the polar headgroups of the phospholipid monolayer.

The presence of oleosins on the surface of oil body membranes may serve several functions:

Stabilization of the oil body - The presence of the amphipathic helices coating the surface of the oil body may enable the oil body to retain its integrity when the dielectric constant of the cell falls during dessication. During the later stages of embryogenesis, mRNA's known as Late Embryogenesis Abundant (LEA) mRNA's accumulate. These have been discovered in several species for example cotton (Barker, et al., 1988), wheat (Marcotte, et al., 1989) and rapeseed (Harada, et al., 1989). Sequence comparison using the derived amino acid sequence showed all LEA's have a high content of hydrophilic amino acids. A possible function for proteins encoded by LEA's may be to stabilize cell components after dessication has occurred. LEA mRNAs also accumulate in response to ABA, which is produced under conditions of water stress.

Act as a binding site for lipases - In *B. napus*, the mobilisation of storage reserves in germinating seedlings after imbibation occurs in several stages (Murphy, et al., 1989a) :

1) Lag phase - (10-15 hours.after imbibation)

2) Napin and cruciferin degradation (12 hours - 3 days).

3) Breakdown of storage oil and oil body membranes - (2 - 7 days).

Lipases digest the triacylglycerol from within oil bodies during germination and seedling growth. Sucrose density centrifugation of the membranes extracted from seedlings just after germination showed that the oleosin protein is found in the light membrane fraction (Murphy, et al., 1989b). Triacylglycerol degradation from within the oil bodies occurs, leaving the oleosin proteins associated with the phospholipid monolayer in the form of "empty" oil body ghosts. These results indicate that the lipid is somehow mobilised from within an intact oil body membrane. Hence the necessity for binding of lipase to the membrane.

It has been suggested that oleosins may provide a binding site for lipases, enabling subsequent triacylglycerol mobilisation from within the oil bodies (Somerville and Browse, 1991). Indirect evidence for the involvement of oleosin in

the regulation of oil mobilization comes from studies carried out on developing soybean seedlings (Herman, et al., 1990). A hydrophilic section of 10 amino acids is thought to be removed from the N-terminal region of the 34kDa oleosin. This cleavage event occurs simultaneously with the initiation of glyoxysome mediated oil mobilization. It is possible that this N-terminal region is necessary for recognition or activation of an enzyme involved with oil mobilization.

1.1.3 SYNTHESIS OF OLEOSIN AND INSERTION INTO OIL BODY MEMBRANES

Little is known about the processing and insertion of oleosin proteins into the oil body membrane. Here comparison will be made with storage protein synthesis and deposition.

Storage proteins are synthesised on the rough endoplasmic reticulum (ER) (Higgins, 1984). Synthesis begins on free ribosomes and a hydrophobic series of amino acids called the signal sequence is produced. A cytosolic protein called the signal recognition particle (SRP) binds to the signal peptide as soon as it is synthesised and causes a pause in translation. The SRP-ribosome complex then binds to the SRP receptor in the endopasmic reticulum. Once the partially synthesised polypeptide has made contact with the ER membrane, synthesis continues resulting in the polypeptide chain being directed to the interior of the ER via a protein channel which spans the membrane.

Storage protein accumulates within the ER and eventually specific regions "bud" off from the ER to form discrete organelles called protein bodies. N-linked glycosylation of the protein occurs within the ER together with proteolytic cleavage reactions which include removal of the signal sequence. Glycosylation of the protein occurs as soon as the polypeptide enters the lumen of the ER. The oligosaccharide is attached to an asparagine residue which is in one of the following sequences:

Asn-X-Ser

Asn-X-Thr Where X may be any amino acid except proline.

The main seed storage proteins of *B. napus* - napin and cruciferin, are cotranslationally transferred into the ER as already described. Napin is synthesised as a 178 amino acid precursor, which is processed by proteolytic cleavages resulting in the loss of the N-terminal signal sequence in addition to two short peptides. The final product consists of two peptides of 86 and 29 amino acids, which are linked together by disulphide bonds (Ericson, et al., 1986). Cruciferin is synthesised as a 488 amino acid precursor which is cleaved resulting in the loss of the N-terminal signal sequence and the formation of two polypeptides, α and β, which are again linked by disulphide bonds (Rödin, et al., 1990).

The enzymes responsible for lipid synthesis are found in the ER. Studies carried out on safflower microsomes show that triacylglycerol synthesis occurs *in vitro* to produce lipid droplets which are not associated with proteins (Stobart, et al., 1986). Electron microscopic studies of developing embryos have shown that oleosins coat the surface of the oil bodies in maize (Fernandez, et al., 1988), soybean (Herman, 1987) and rapeseed (Murphy, et al., 1989b). Immunogold labelled sections of developing seeds in the very early stages of embryogenesis show that no oleosin protein is associated with oil bodies at this stage of development.

Some possible explanations for the existence of protein-deficient oil bodies in developing embryos have been proposed (Qu, et al., 1986):

-Deposition of triacylglycerol between the two layers of the phospholipid bilayer in specific regions of the ER which bud off to form oil bodies deficient of oleosin.

-Release of free triacylglycerol molecules into the cytosol which coalesce to form lipid droplets.

Detailed studies of oil body formation in developing seeds are needed to anwser these questions.

The next question is how is the oil body coated with oleosin? Oleosin genomic clones isolated from maize (Qu and Huang, 1990) and carrot (Hatzopoulos, et al.,

1990) possess no detectable signal sequence. *In vitro* synthesis of the 24kDa oleosin (P24) from soybean produced a protein with the same molecular weight as *in vivo* (Herman, 1987).

In *B.napus,* EM studies of immunogold-labelled developing embryos showed oleosin protein was found only to be present in the oil body fraction and not in the ER fraction (Murphy, et al., 1989b). This was supported by sucrose density fractionation of the various membrane fractions. Similar results were obtained in the soybean system (Herman, 1987).

How oleosin becomes associated with the oil body membrane is unknown. Ribosomes may associate with the oil body and direct the synthesis of oleosin co-translationally into the oil body. Alternatively the polypeptide may be released into the cytosol, followed by post-translational insertion into the oil body membrane - possibly directed by the existence of a hydrophobic domain in the protein. Pulse-chase experiments on embryo tissue may be used to determine which theory is correct.

None of the oleosin proteins so far identified have been shown to be glycosylated or proteolytically cleaved prior to insertion into the oil-body membrane. Proteolytic cleavage of oleosin may however have a role in the regulation of oil-body degradation during germination (Herman, et al., 1990).

The timing of oleosin synthesis, relative to oil and storage protein synthesis in *B. napus* is shown in Figure 1.1 (Murphy and Cummins, 1989). Oil deposition begins several weeks before oleosin is synthesised. This is consistent with the appearance of naked oil bodies during the early stages of embryogenesis. It is interesting to note that the storage proteins napin and cruciferin are produced earlier than oleosin, indicating that they are under the control of different regulatory factors.

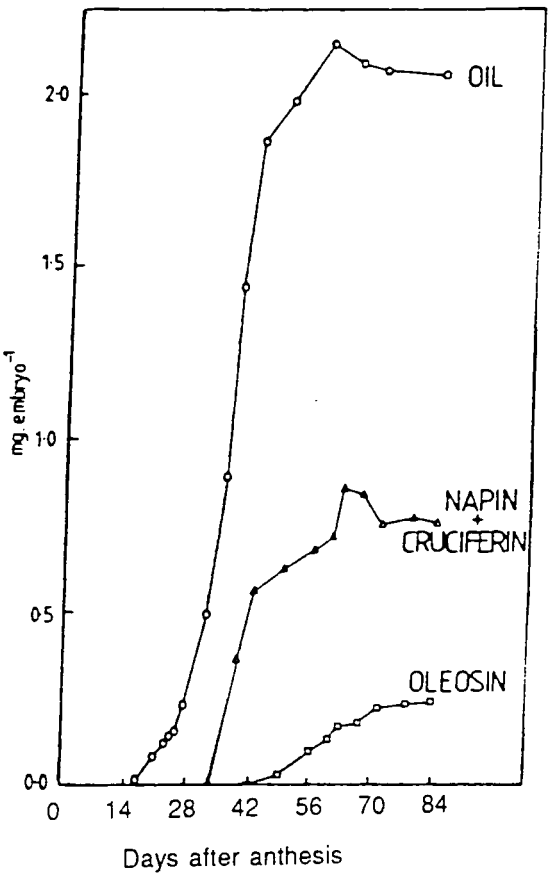


FIGURE 1.1

Oleosin, storage protein and oil production in *B. napus* embryos (Murphy and Cummins, 1989)

Triacylglycerol was quantified by extraction of total lipid from a seed homogenate with organic solvents followed by weighing on an accurate microbalance. Total seed protein was analysed by SDS PAGE. Relative proportions of the mass in napin, cruciferin and oleosin were estimated following scanning of the stained gel with a laser densitometer. Total protein was quantified by a modified Lowry assay.

Figure 1.1 indicates that three distinct temporal phases of gene expression exist during embryogenesis.

1] Accumulation of storage triacylglycerol. The enzymes required for triacylglycerol production are found in the ER. These include the enzymes for the fatty acid synthase complex, desaturases and acyltransferases (see section 1.1.1).

2] Storage protein synthesis. Napin and cruciferin accumulate later in development than storage lipid. The storage protein genes are therefore expressed later than the genes encoding enzymes involved in lipid synthesis.

3] Oleosin protein accumulation is observed last.

Isolation of the regulatory elements from the 5' upstream regions of genes expressed during the three different temporal stages would provide useful tools for engineering tissue specific and temporal expression of foreign genes in *B. napus*.

1.2 OLEOSIN IS ENCODED BY A GENE FAMILY

1.2.1 HOW ARE GENE FAMILIES FORMED?

Many different proteins are encoded by gene families, the family members being identified by their sequence similarity. This section deals with the advantages of gene families, over single genes and the possible significance of the existence of an oleosin gene family.

Gene families arise by duplication of a gene originally present as a single copy. Mechanisms for the evolution of gene families have been reviewed, (Maeda and Smithies, 1986). The initial event is rare and unpredictable and thought to occur by one of the following routes:-

- Non-homologous chromosome breakage and reunion.
- Homologous recombination, resulting in unequal crossing-over of chromosomes.

Rearrangement of genes within the genome can then proceed after gene amplification has produced several copies. Individual genes may diverge by

accumulation of mutations. Unequal crossing-over and gene conversion increase the sequence disparity between the members of a gene family. Mutations may result in the formation of pseudogenes which have no functional gene product.

1.2.2 WHY ARE GENE FAMILIES IMPORTANT?

The basic feature of a gene family is that it enables several genes encoding proteins with similar or identical functions to exist in a single genome (Smith, 1990). Some advantages of gene families over single copy genes are as follows:

- Isoforms may be localized in different subcellular compartments.
- If the coding sequence of the individual members has diverged sufficiently, the proteins may also differ slightly in function. In fat metabolism, the cofactor, acyl-carrier-protein has two isoforms which bind different fatty acids (Ohlrogge, 1987).
- Multiple genes encoding proteins with identical functions make it possible for the expression of each gene to be independently controlled so that different members of the family respond in a unique way to the cellular factors responsible for the control of gene expression.

Individual members of the gene family may be expressed in different plant tissues, or at different times during development. Such tissue specific and developmental control has been observed in gene families encoding a variety of different proteins. Each gene must therefore have its own independent regulatory sequences, allowing flexible expression of a particular gene product in a constant genotype. Examples include the chalcone synthase gene family from pea (Harker, et al., 1990) and the ribulose bisphosphate carboxylase small subunit gene family from tomato (Sugita and Gruissem, 1987).

The potential importance of the existence of several oleosin gene family members is as follows:

1] Subcellular localization - different genes may be expressed in different subcellular compartments.

2] Functional differences - the variable N- and C-termini of the protein may have different functions in the different oleosin proteins.

3] Temporal regulation - different oleosin genes may be expressed at different times during embryogenesis.

1.2.3 PROTEIN PRODUCTS OF THE OLEOSIN GENE FAMILY

SDS PAGE of oil bodies from various plant species reveals significant heterogeneity of oleosins, as shown in Figure 1.2 (Qu, et al., 1986). The major bands represent the oleosins. The minor bands of higher molecular weight, may represent proteins with enzymatic function such as lipases, or they may consist of aggregates of oleosin proteins.

The different maize oleosins have been classified according to the scheme in Table 1.1 (Vance and Huang, 1987).

TABLE 1.1 Oleosin proteins from maize

PROTEIN	SIZE(kDa)	Relative proportions
L1	19	0.1
L2 (KD18)	18	0.2
L3 (KD16)	16	0.5
H	40	0.2

A cDNA clone encoding the L3 (KD16) protein (Vance and Huang, 1987) and a genomic clone for the L2 (KD18) protein (Qu and Huang, 1990) have been isolated.

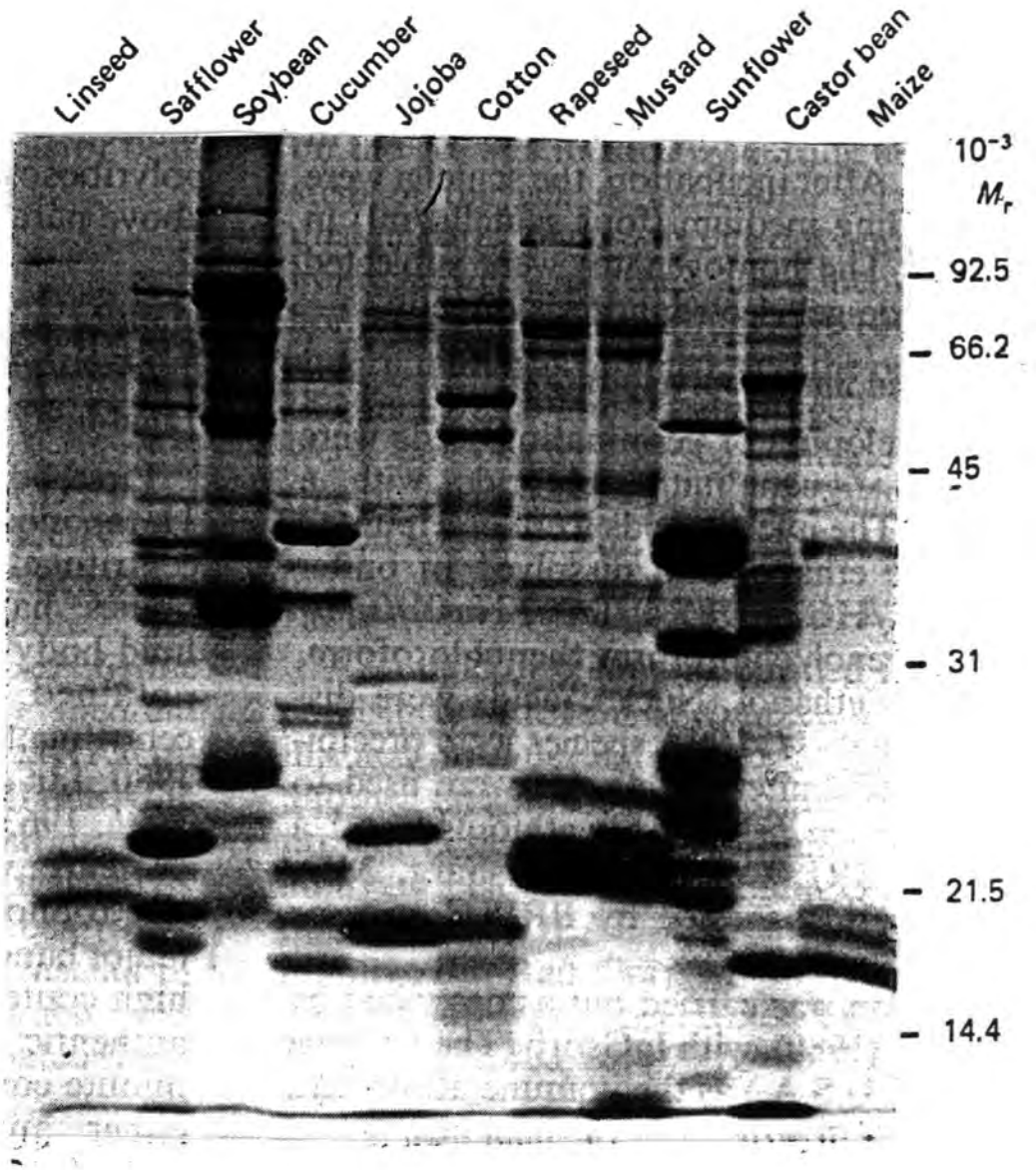


FIGURE 1.2

Oil body proteins from a range of plant species (Qu, et al., 1986)

Oil bodies were isolated from homogenised embryo tissue and purified by flotation centrifugation. The triacylglycerol was extracted with diethyl ether and the remaining oil body membranes analysed by SDS PAGE. For all species, several major bands and numerous small bands were observed. The oleosin proteins are believed to be around 30kDa or less.

Four oleosin proteins have been identified in soybean, having molecular weights of 34, 24, 18 and 17 kDa (Herman, 1987). Very little is known about the sequences of these proteins. The abundance of the protein products of the various types of oleosin in seeds is variable. In soybean seeds the major component of the oil bodies is the 24kDa protein. The 34, 18 and 17kDa proteins are present in relatively small amounts (Herman, 1987). In maize the major component is the 16 kDa protein (Vance and Huang, 1987).

An immunological study has been carried out on the oleosin isoforms from a range of seeds (Tzen, et al., 1990). The results showed that oleosins from distantly related plant species share common antigenic determinants. Two immunologically distinct forms of oleosin exist. Maize, soybean and rapeseed oleosins were studied in detail and classified as shown in Table 1.2. Oleosins within the low and high molecular weight classes are immunologically related, however antibodies raised to low molecular weight oleosins do not cross-react to high molecular weight oleosins and *vice versa*.

Table 1.2 Classification of oleosin proteins (Tzen, et al., 1990)

Plant	Low M_r class	High M_r class
Maize	16kDa	18, 19kDa
Soybean	17, 18kDa	24, (34)*kDa
rapeseed	20kDa	22, 24kDa

*Note : The 34kDa protein from soybean is not detected by either antibodies specific to the low or high molecular weight proteins.

The results shown in Table 1.2 indicate that oleosins diverged into two immunologically distinct isoforms early in the evolutionary pathway, even before the distinction between monocots and dicots. The existence of two isoforms may have some

functional significance. The N-terminal region has the least sequence conservation and it is this region of the 34kDa oleosin from soybean which has been implicated in regulation of oil mobilization (Herman, et al., 1990).

A genomic clone for a carrot oleosin, DC59 has been isolated (Hatzopoulos, et al., 1990). Hydropathy plots of the protein sequences derived from carrot and maize oleosin clones show a similar pattern, with hydrophilic N- and C- terminal ends separated by a hydrophobic domain which penetrates the triacylglycerol core of the oil body. Comparison of the sequence from maize KD16 with the carrot sequence shows significant homology exists in the central hydrophobic domains of the two proteins, the hydrophilic regions being less conserved. Conservation of the central hydrophobic domain in these two distantly related species indicates that the central hydrophobic region is essential for the function of the oleosin proteins.

The central hydrophobic section of the oleosin proteins is buried in the triacylglycerol core of the oil body. Following extraction of the protein from its lipid environment in the oil body, the hydrophobic regions of the protein will associate in the aqueous solvent. For this reason the hydrophobic region will have a low antigenicity. Antibodies for the variable exposed surface regions will be more abundant. Sequence comparison between the low and high molecular weight forms of maize oleosins with molecular weights of 16 & 18kDa respectively, shows that great sequence similarity exists in the hydrophobic domains (see Figure 3.8). The immunological classification of oleosins into two isoforms is therefore dependent on the hydrophilic surface domains (Tzen, et al., 1990). More information is needed before the relative importance of the N- and C- temini can be established.

1.2.4 CONTROL OF OLEOSIN GENE EXPRESSION

The pattern of oleosin gene expression was investigated in maize as shown in Figure 1.3 (Qu, et al., 1990). The production of KD16 and KD18 oleosin RNAs was followed during embryogenesis. These appear to be under coordinate control. The 5' upstream region of a KD 18 genomic clone contains putative ABA responsive elements (Qu and Huang, 1990). These are 5'-TACGTGTC-3' at-62bp and 5'-ACGTACCC-3' at -83. It is possible that these sequences are involved in regulation of expression by ABA.

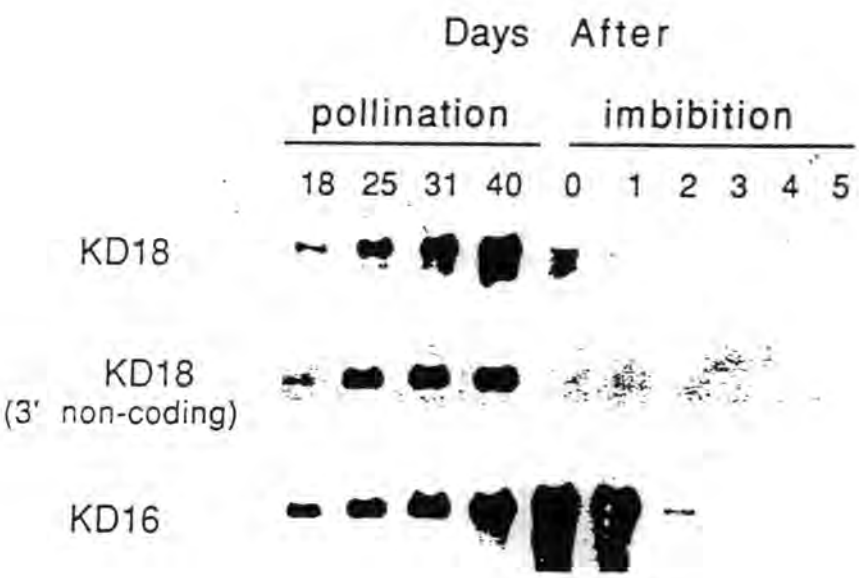


FIGURE 1.3

Steady state oleosin mRNA levels in maize embryos (Qu, et al., 1990)

Total RNA (10μg/lane) from maize embryos at different stages of embryogenesis and also from germinating seedlings was separated by electrophoresis, Northern blotted and probed using nick translated cDNA probes.for KD16 and KD18. Use of the KD18 3' non-coding region to probe the blot showed the greater sequence specificity of this probe compared with the coding region probe due to lower conservation of the 3' non-coding regions of the oleosin genes.

1.3 TRANSCRIPTIONAL vs POST-TRANSCRIPTIONAL CONTROL DURING EMBRYOGENESIS

1.3.1 EXAMPLES OF SYSTEMS WHERE TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL CONTROL MECHANISMS ARE IN OPERATION

Transcriptional and post-transcriptional control mechanisms regulate embryogenesis in soybean. Following fertilization, the expression of a specific set of genes is necessary for the formation of the embryo and the tissue surrounding it. These genes are usually not active in somatic cells. Expression of various genes during embryogenesis can be followed by monitoring the mRNA levels. This has been done for all the stages of embryogenesis in soybean (Goldberg, et al., 1989).

Figure 1.4 (Goldberg, et al., 1989) shows the times during embryo development when certain RNA classes accumulate and are degraded. RNAs encoding proteins involved in basic cell functions common to all plant cells are present throughout development. Seed protein mRNAs are produced early in development and degraded towards the end of embryogenesis, when the seed begins to dessicate. From Fig 1.4 it can be seen that some mRNA's are resistant to the enzymes involved in mRNA degradation and remain intact throughout the later stages of development, to be stored in the dry seed.

Similar patterns of expression have been observed for genes expressed in *B. napus* embryos and germinating seedlings. Some genes which encode seed storage proteins are expressed early in embryogenesis. mRNA levels for genes in this class peak during embryogenesis, then they fall as the seed begins to dessicate. Late embryogenesis abundant mRNA levels continue to increase throughout embryogenesis and are present at high concentrations in the dry seed, but disappear quickly after imbibition. mRNA levels for proteins which are active in the young seedling, such as the glyoxylate cycle enzymes isocitrate lyase and malate synthase accumulate late in embryogenesis and continue to accumulate after imbibition (Comai and Harada, 1990).

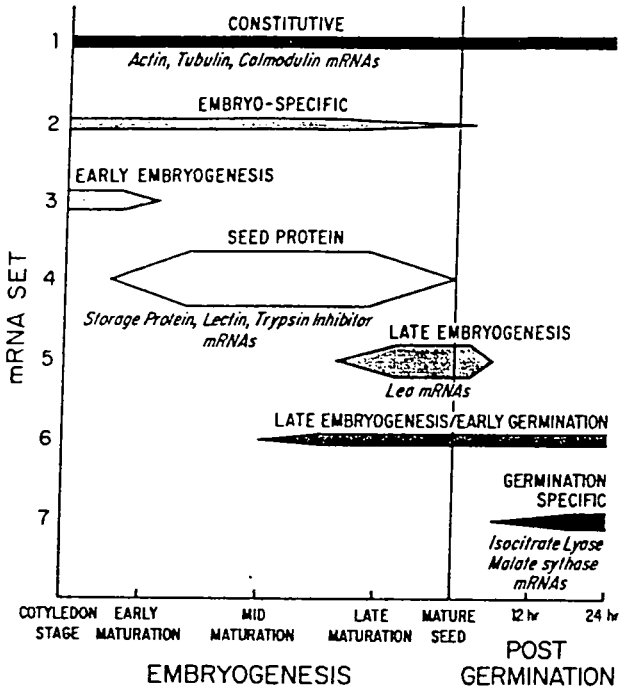


FIGURE 1.4

mRNA sets produced during embryogenesis in soybean (Goldberg, et al., 1989)

At various stages of development mRNA levels are dependent on the rate of transcription and processing versus the rate of degradation. Production of mRNAs requires a series of post-transcriptional steps. These include 5'-capping (Hamm and Mattaj, 1990), poly(A) tail addition (Wickens, 1990), intron removal (Goodall and Filipowiz, 1989) and transport from the nucleus. The rate of transcription of the mRNA is thought to be the main factor in the control of differential accumulation of individual RNA types.

Detailed studies have been carried out on the mRNA levels of various proteins produced in soybean seeds during embryogenesis (Walling, et al., 1986). Constitutively expressed genes that are necessary for basic cell function show relatively small changes in transcription rate; these changes being due to fluctuations

in the capacity of the system for gene expression at the different stages of embryogenesis. However there are large fluctuations in mRNA levels for seed specific proteins, such as seed storage proteins. This is mediated by activation and subsequent repression of transcription of these genes at different times during embryo development.

Using the storage proteins of soybean as a model system, the following basic principles can be derived:

- Seed protein genes are transcriptionally activated then repressed during embryogenesis.
- These genes are either inactive or transcribed at low levels in mature plants.
- mRNA's for seed storage proteins may be present in the seeds in greatly differing amounts, while the genes are often being transcribed at similar rates. The mRNA levels in this case are determined by post transcriptional events, which may include mRNA degradation.
- Transcriptionally active and inactive genes have indistinguishable methylation patterns.
- Seed storage protein genes are not selectively amplified during development.

Soybean seed storage proteins are β-conglycinin and glycinin. Glycinin is a hexamer encoded by a gene family with five members (Nielsen, et al., 1989). The β-conglycinin protein is a trimer made up of various combinations of two different subunit types, α and β (Bray and Beachy, 1985). These subunits are encoded by two groups of a gene family which has 15 members.

In this system, post transcriptional events are also important for the temporal regulation of gene expression during embryogenesis. The protein products of all five members of the glycinin gene family accumulate and decay at the same time, ie they show coordinate regulation (Walling, et al., 1986). This is not the case for the β-conglycinin gene family. The mRNA's encoding the α and β subunits are present in

different relative amounts during embryogenesis (Harada, et al., 1989). Transcription rates for the β-conglycinin genes were measured *in vitro*, and it was found that the genes for both subunit types are transcribed at similar rates at each stage of embryogenesis. Therefore the difference in mRNA levels for the subunits is dependent on post transcriptional events such as processing of hnRNA and degradation of mRNA molecules.

1.3.2 REGULATION OF NAPIN AND CRUCIFERIN EXPRESSION

Transcriptional and post-transcriptional control processes operate in the regulation of napin and cruciferin expression. This is clearly shown in Fig 1.5 (DeLisle and Crouch, 1989).

In this figure, transcription rates and steady state mRNA levels for napin and cruciferin are compared. For both napin and cruciferin the transcription rates increase to a peak value, then decrease in the later stages of embryogenesis. Since the timing of transcription for the two genes is so similar during embryogenesis it is likely that they share common upstream regulatory sequences. The decreased rate of transcription for napin was accompanied by a decrease in the steady-state mRNA level. The cruciferin mRNA level did not decrease significantly, even at the onset of dessication, when the rate of transcription was low. Cruciferin mRNA is therefore more resistant to degradation by cellular RNases. It is possible that napin is degraded at a faster rate due to the presence of specific endonuclease target sites in the 3' UTR.

Many processes are therefore operating in developing seeds to produce the observed pattern of gene expression. The relative contribution of transcriptional and post-transcriptional events must be taken into account in the study of oleosin gene expression.

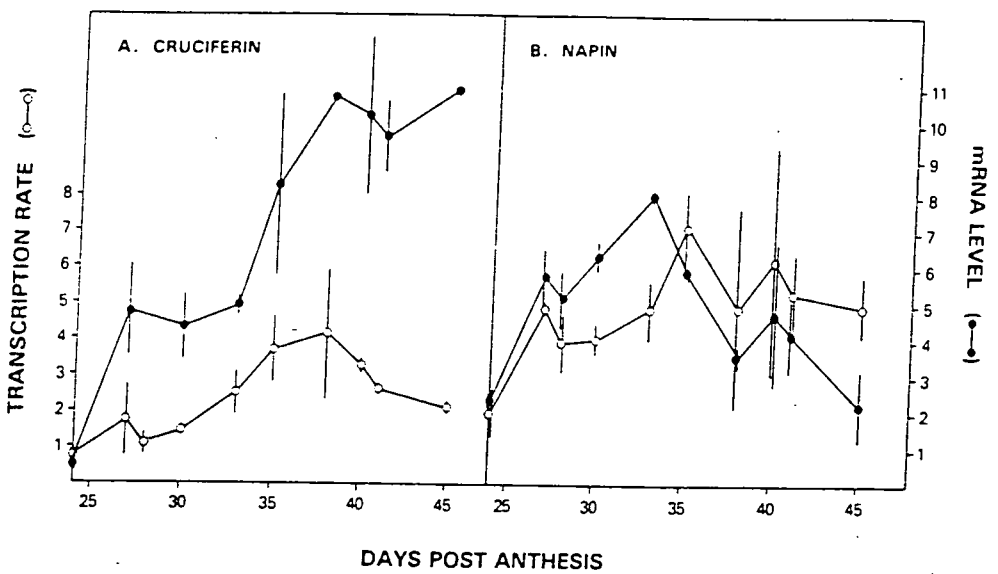


FIGURE 1.5

Napin and cruciferin mRNA levels and transcription rates during embryogenesis (DeLisle and Crouch, 1989)

Napin and cruciferin mRNA levels were measured relative to total mRNA by dot blotting, following extraction of total RNA from developing embryos at different stages of development. Relative transcription rates were measured by carrying out run-on transcription with nuclei from embryos at different stages of development. [^{32}P] GTP was used to label the transcription products.

1.4 FACTORS DETERMINING mRNA STABILITY

1.4.1 THE FORMATION OF THE POLY(A) TAIL

The most well characterised function of the 3' UTR of mRNA molecules is its role in the control of polyadenylation (Birnstiel, et al., 1985; Wickens, 1990). A highly conserved sequence 5'-AAUAAA-3', called the polyadenylation signal is required in the 3'UTR for accurate and efficient cleavage to occur. Cleavage at a minimum distance of 11 nucleotides downstream of the polyadenylation site provides a "free end" to which a poly(A) tail can be added. From a compilation study of the 3' untranslated regions of 15 different animal genes, the spacing between the site of poly(A) addition and the AATAAA polyadenylation signal was highly conserved. The spacing ranged from between 11-20bp, depending on the gene (Birnstiel, et al., 1985). For cleavage the polyadenylation signal is required, together with other UG rich sequences downstream of the poly(A) addition site. Many plant genes have the consensus sequence TgTgTTTT which resembles the consensus sequence YGTGTTYY found in mammalian genes approximately 30 bp downstream of the polyadenylation signal (Joshi, 1987b).

In a compilation study carried out on 46 different plant genes, a large proportion of the plant genes isolated were found to contain polyadenylation signals which closely resemble the AATAAA consensus in mammalian systems (Joshi, 1987b). The polyadenylation site ranged from 8-42bp downstream of the polyadenylation consensus signal. In many plant genes, multiple poly(A) addition sites exist, resulting in mRNA molecules which are heterogeneous with respect to the part of the 3'UTR to which the poly(A) tail is attached (Dean, et al., 1986). Processing and polyadenylation in plant genes seems to be more flexible than in animal genes. The potential for mRNA molecules to be polyadenylated at different sites may be involved in control of mRNA stability.

The processing mechanism appears to differ in monocots and dicots as there are signals in the 3' end of mRNA's from monocots and dicots which are preferred by the respective processing systems (Keith and Chua, 1986).

1.4.2 WHAT DETERMINES THE STABILITY OF mRNA MOLECULES?

Different mRNA molecules vary greatly in their stability within cells. Some mRNA's such as the mRNA for β-globin are stable and remain intact within cells for days whilst others such as c-*fos* mRNA are very unstable, with a half-life of only 15 minutes (Brawerman, 1987). Within a single cell there are mRNA's with widely differing stabilities. Some factor must direct the specificity for the degradation of some mRNA molecules more rapidly than others. A conserved AU rich sequence from the 3' UTR of the human lymphokine gene GM-CSF has been shown to be responsible for the selective degradation of the GM-CSF mRNA (Shaw and Kamen, 1986).

It is likely that two types of mechanism exist within cells for mRNA degradation (Brawerman, 1987; Raghow, 1987):

- Non specific 3'-5' exonuclease cleavage in which the enzymes responsible for degradation do not distinguish between different types of RNA molecule .
- Specific endonucleases cleave at defined sites within the 3' UTR of RNA molecules, as directed by cell components which are responsive to changes in the environment of the cell .

These two mechanisms act together. First cleavage at specific sites in the 3' UTR results in loss of the protective function conferred by the 3'-end of the molecule. The molecule is subsequently degraded by the action of non-specific exonucleases.

The mRNA of the *int* gene of bacteriophage λ is protected from 3' exonucleolytic digestion by the presence of a double stranded stem loop structure in the 3' UTR which is resistant to exonuclease digestion (Brawerman, 1987). The presence of a poly(A) tail on mRNA molecules may serve a similar function. For the c-*fos* mRNA, there is direct evidence that loss of the poly(A) tail results in rapid degradation of the RNA

(Wilson and Treisman, 1988). What property of the poly(A) sequence confers this resistance to exonuclease attack? Detailed study in *Xenopus* oocytes showed a protein called the poly(A) binding protein (PABP) associates with high affinity to poly(A) tails (Bernstein and Ross, 1989). It may protect the rest of the molecule from non-specific 3'-5' exonuclease digestion by sterically blocking the movement of the exonuclease along the RNA.

1.4.3 MECHANISM FOR THE PREFERENTIAL STABILIZATION OF SPECIFIC RNA MOLECULES

The stability of mRNA molecules is dependent on the presence of the poly(A) tail or a stem loop structure in the 3' UTR. Selective degradation of specific mRNA types therefore requires a mechanism for removal of the 3' end of the mRNA molecules. Endonuclease cleavage at specific sites in the 3' UTR of apolipoprotein II (apoII) mRNA destabilizes the RNA (Binder, et al., 1989). These sites have the following sequence: 5'-UAA-3' or 5'-AAU-3'. Cleavage only occurs at these sites if they are present in single stranded regions of the RNA molecule.

Proteins thought to be involved with mediating stability changes in the apoII mRNA associate to regions containing stem loop structures (Hwang, et al., 1989; Ratnasabapathy, et al., 1990). The single stranded regions within these stem loops in apoII contain the target sites for specific endonuclease cleavage. Recognition of specific regions in the RNA may depend on conserved nucleotide sequences or secondary structure motifs within the RNA (Draper, 1989).

Specific targetting of apoII mRNA's for degradation is under the control of hormones. ApoII mRNA is selectively destabilized in avian liver when levels of oestrogen drop in the tissue (Binder, et al., 1989). Some mechanisms controlling mRNA stability may function via the production of a second messenger. The second messenger, cAMP, destabilizes the mRNA for a protein called the 117 antigen in *Dictyostelium discoideum* (Juliani, et al., 1990).

1.4.4 STABILITY OF PLANT mRNA MOLECULES

In comparison to animal systems, little is known about the determination of stability in plant mRNA molecules. Evidence that the 3' end of plant mRNA molecules affects stability of the molecule comes from linking different 3' ends from plant mRNA's to the neomycin phosphotransferase (NPT) reporter system (Ingelbrecht, et al., 1989). The 3' ends from chalcone synthase (CHS), octopine synthase (OCS), ribulose bis phosphate carboxylase small subunit (rbc SS) gene, 2S seed protein gene and extensin were linked to the NPT reporter. When the constructs were expressed in protoplasts, all five produced a high level of NPT enzyme activity. Levels of expression in transgenic leaf tissue were very different for each construct and can be summarised as follows:

rbcSS > OCS > 2S > Extensin > CHS

The protoplast system was used as a control to prove that the constructs were not being transcribed at different rates. However in some genes, such as those produced as part of the plant wounding response, the 3' untranslated region of the gene has an effect on the transcription rate. A better control in this series of experiments would therefore be to carry out transcription run-on experiments with nuclei isolated from the differentiated plant material expressing the hybrid genes. The results obtained by Inglebrecht et al (1989) indicate that the levels of mRNA produced by the various constructs are dependent on differential processing and/or mRNA stability. Further control experiments are needed to prove the functional importance of the 3' end in post-transcriptional control.

Differential stability has been observed for the mRNAs of the KD16 and 18 oleosins of maize during seed germination (Qu, et al., 1990). After imbibation, KD18 mRNA was rapidly degraded in the seed, however KD16 mRNA remained stable for over 24 hours (see Figure 1.3). From this it can be seen that these two RNA moleclues

appeared to both be stable in the developing seed, however in the germinating seedling they showed different stabilities.

In *B.napus*, the seed storage proteins napin and cruciferin are coordinately expressed (DeLisle and Crouch, 1989). Their mRNAs however have different stabilities, which is apparent when transcription of the genes ceases late in embryogenesis. The level of napin mRNA decreases significantly during late embryogenesis, but the cruciferin mRNA level falls only slightly.

mRNA stability has a significant effect on the mRNA steady state levels. However the pattern of genes expressed in a particular tissue is dependent on the genes being actively transcribed. Many processes are therefore operating in developing seeds to produce the observed pattern of gene expression. The relative contribution of transcriptional and post-transcriptional events must be taken into account in the study of oleosin gene expression.

1.5 AIMS OF THE PROJECT

To establish whether oleosin is encoded by a gene family in *Brassica napus* and to characterise the members.

To observe the pattern of oleosin gene expression at the protein and RNA levels during embryogenesis. From this it is possible to determine if the control of expression is predominantly at the level of transcription or translation. Also to determine whether post-transcriptional processes have an effect on mRNA levels.

To isolate genomic clones for the oleosin gene to obtain sequence information for upstream regulatory sequences and enable introns to be located.

CHAPTER 2

MATERIALS AND METHODS

2.0 Materials and bacterial strains

Materials were obtained from the following suppliers:

BDH - Acetic acid, acrylamide, ammonium chloride, ammonium acetate, bis-acrylamide, calcium chloride, chloroform, DMSO (dimethylsulphoxide), formamide, glucose, glycerol, hydrochloric acid, potassium acetate, potassium hydroxide, sodium acetate, sodium hydroxide, sodium sulphite, sucrose, tri sodium citrate.

Sigma - Ammonium persulphate, ampicillin, bovine serum albumin, Coomassie Brilliant Blue, EDTA (ethylenediamine tetra acetic acid), formaldehyde, glycine, HEPES, lauryl sulphate (SDS), maltose, MOPS (3-[N-morpholino]propane sulphonic acid), polyethylene glycol 8,000 (PEG), sodium iodide, tetracycline, tetra sodium pyrophosphate, urea.

Oxoid - Agar (number 1), yeast extract.

BRL - Agarose

Beckton Dickinson - Trypticase peptone

Riedel de Haen - Sodium chloride

Amersham - α-[^{32}P]-dCTP, α-[^{35}S]-dATP, Hybond-C & Hybond-N

Schleicher & Schuell - Nitrocellulose filter circles

Biorad - 5-Bromo-4-chloro-3-indolyl phosphate, nitroblue tetrazolium, goat anti-rabbit IgG,

Boehringer Mannheim - T4 DNA ligase

Northumbria Biologicals - Restriction enzymes

Many of the basic molecular biological techniques have been taken from "Molecular Cloning - A Laboratory Manual" by J. Sambrook et al (1989).

***E. coli* strains used in subcloning and phage preparation:**

Strain	***Function/Phenotype***
DH5α	Transformation with plasmid vectors - unless stated otherwise *supE*-44 *Δlac*-U169 (Φ80 *lacZ* ΔM15) *hsd*-R17 *recA*1 *end*-A1 *gyr*-A96 *thi*-1 *rel*-A1
GM119	Obtaining unmethylated DNA *dcm*-6 *dam*-3 *met*-B1 *thi*-1 *lac*-Y1 *gal*-K2 *gal*-T22 *mtl*-2 *ton*-A2 *tsx*-78 *sut*-E44
TG2	Used when subcloning in M13 *supE hsdΔ*5 *thi Δ(lac-proAB) Δ(srl-recA)* 306::Tn10(*tet*^{r})
P2392	Growth of λEMBL 3 vectors F-*hsdR514* ($r_K^-m_K^+$) *supE*44 *supF*58 *lacY*1 or *Δ(lacIZY)6 galK*2 *gal*T22 *met B*1 *trpR*55 P2

2.1 *E. coli* plasmid Miniprep - Alkaline lysis method

A 5ml aliquot of L-broth containing the appropriate antibiotic was inoculated with a single colony and incubated at 37°C overnight with vigorous shaking. A suitable quantity of the culture was transferred to an Eppendorf tube, and the remainder stored at 4°C. The cells were pelleted at 13,000rpm for 2mins and the supernatant removed, leaving the pellet as dry as possible. The pellet was resuspended by vortexing in 100µl of an ice cold solution of 50mM glucose, 10mM EDTA, 25mM Tris-HCl (pH8.0) and stored for 5mins at room temperature. A 200µl aliquot of a freshly prepared solution of 0.2M NaOH, 1% SDS were added and this was mixed well (without vortexing). The tube was left on ice for 5mins, then 150µl of an ice-cold solution of potassium acetate ($3MK^+5MOAc^-$ pH4.8) was added. After mixing, this was stored on ice for 5 mins, then centrifuged 13,000rpm for 5mins at 4°C. The

supernatant was transferred to a fresh tube and an equal volume of phenol:chloroform (1:1) was added. This was left on ice for 5mins, centrifuged at 13,000rpm for 5mins to separate the two layers; the top aqueous layer was then extracted with chloroform:isoamyl alcohol 24:1. The DNA was precipitated by adding two volumes of ethanol and leaving at -20°C for 30mins, then pelleted at 10,000rpm for 10mins. After washing with 70% ethanol the DNA was dried under vacuum then resuspended in an appropriate volume of T.E. buffer.

Note - For DNA of a higher purity, acidified phenol was used in the phenol extraction step. This was prepared by equilibrating redistilled phenol at least three times with 50mM sodium acetate, pH 4.0.

Note - 13,000rpm spins were carried out in a microfuge.

2.2 Plant DNA miniprep technique

Approximately 2g of leaf material was ground in a pestle and mortar containing liquid nitrogen. This was ground for a couple of minutes, to a coarse flour-like consistency, then transferred to a sterile corex tube containing sterile extraction buffer pH8.0 [100mM TrisHCl, 50mM EDTA, 500mM NaCl, 10mM ß-mercapto ethanol], plus 1ml 20% SDS. After mixing by vigorous shaking, the sample was incubated for 20mins at 65°C. Then 5ml of potassium acetate solution was added and after mixing, the sample was left on ice for 20mins to precipitate the protein. This was sedimented at 25,000g for 20mins @ 4°C. Any remaining particles were removed by filtration through a double layer of sterile cheese cloth into a fresh sterile 30ml corex tube.

The DNA was precipitated by adding 10ml of isopropanol and leaving for 30mins @-20°C. The DNA was pelleted at 20,000g for 15mins @ 4°C. The pellet was dried by inverting the tube for 10mins then resuspended in 700µl of resuspension buffer [50mM Tris HCl, 10mM EDTA pH8.0] at room temperature with occasional shaking (this was left for no longer than 30mins). Any material which did not resuspend, was spun down (10min, 13,000rpm). The supernatant containing the DNA

was then extracted several times with an equal volume of phenol:chloroform (1:1), until no residue was visible at the interface. This was followed by a single chloroform : isoamyl alcohol (24:1) extraction. The DNA was precipitated by adding 75µl of 3M sodium acetate pH5.2 and 500µl of isopropanol. After leaving at -20°C for 30mins, the DNA was pelleted (10 min, 13,000rpm). The pellet was washed with 70% ethanol and dried under vacuum. The DNA was resuspended in 200µl of TE.

2.3 DNA agarose gel electrophoresis & Southern Blotting

DNA was run in gels made to a final concentration of 0.5 - 2.0% agarose, depending on the size of the DNA. These were run in 1xTBE buffer (see section 2.25) containing ethidium bromide at a final concentration of 10µg/100ml. Gel loading buffer was mixed with the DNA samples before loading to 1x final concentration.

6x gel loading buffer-

0.25% bromophenol blue, 0.25% xylene cyanol, 40% sucrose, 0.1xTBE.

Large 200ml gels were run at 25V over night, whilst small 60ml gels were run at 60V for 2-3 hours. Before blotting, the DNA in the gel was denatured. For total genomic DNA first the gel had to be soaked in 1% HCl for 15mins, then washed twice in distilled water (dH_2O). The following procedure was then carried out. The gel was soaked for 45mins in denaturation buffer (1.5M NaCl, 0.5M NaOH), with occasional shaking. The denaturation buffer was poured off and the gel washed twice in dH_2O. The gel was then neutralized by soaking for 45mins in neutralization buffer (1.5M NaCl, 0.5M Tris-HCl pH 7.2, 0.001M EDTA) with occasional shaking. After rinsing in 20xSSC, the gel was blotted onto a Hybond membrane - either nylon or nitrocellulose.

2.4 DNA Fragment Isolation

(i) Freeze squeeze - After running the DNA sample on an agarose gel, the required band was cut out of the gel using a razor blade, and then soaked in the following solution for 15 mins at room temperature -

0.9ml dH_2O, 0.1ml 3M NaOAc pH5.2, 40µl 250mM EDTA.

A plug of sterile, siliconized glass wool was placed in a small Eppendorf tube, the base of which was pierced. The gel slice was blotted to remove excess liquid and then placed in the tube on top of the glass wool. This was then placed into a large Eppendorf, and frozen at -80°C for 15 mins. The tubes were spun at 13,000rpm @ 4°C for 15mins which transferred the DNA from the gel slice, in solution, into the large Eppendorf. To precipitate the DNA, the following were added - 5µl of 1M $MgCl_2$ in 10% acetic acid, 20µl of 3M NaOAc pH 5.2 and 1ml ethanol, and left at -80°C for >30mins. The DNA was pelleted for 5mins at 13,000rpm, then washed in 70% ethanol and resuspended in T.E. buffer.

(ii) Silica fines - The DNA was run in a 0.7% low melting point agarose gel at 40 V in 1xTAE buffer. The required band was excised and placed in an Eppendorf tube at 70°C for 10mins. This was cooled to 37°C and 1ml of sodium iodide (see note * for preparation) added. After mixing, 20µl of silica fines was added and the sample left at room temperature for 30mins for the DNA to adsorb. The fines were spun down by a very short spin at 13,000rpm, washed in 70% ethanol, resuspended 100µl TE and left for 10mins at 37°C to elute the DNA. The fines were spun down and the DNA recovered by ethanol precipitation.

Note - The sodium iodide was prepared by adding 90.8g sodium iodide and 1.5g sodium sulphite to 100ml distilled water. After dissolving, this solution was filter sterilised and 0.5g sodium sulphite added to saturate the solution - stored in the dark at 4°C.

2.5 Restriction of DNA

Restriction digests were carried out using enzymes supplied by Northumbrian Biologicals. Reaction conditions were as recommended for each individual enzyme.

Preparation of DNA size markers

The incubation mixture consisted of: 33µl λ DNA (10µg), 10µl appropriate buffer, 57µl distilled water, 60U restriction enzyme. Reaction was stopped with 18µl

of 6x stop buffer. A 10µl aliquot was loaded onto a gel. The size markers used were the fragments derived from *PstI* and *HindIII* digestion of phage λ.

Fragment sizes for the *PstI* digest were as follows:

11,501
4,507-5,077
2,838
2556-2443
2140
1986
1700
1158
1093
805
514
468, 448, 339

2.6 Ligation of foreign DNA into vectors

Ligations were carried out using T4 DNA ligase and the reaction volume was kept as small as possible - 9µl insert DNA, 1µl vector DNA (10-20ng), 1.3µl reaction buffer 2µl ligase (approximately 10 Units of ligase). DNA concentrations were varied until an optimal ratio of vector to insert DNA was obtained.

2.7 Plant RNA preparation (Prescott and Martin, 1987)

The tissue was ground in a pre-cooled pestle and mortar to a fine powder, ensuring that the sample remained frozen throughout. The ground tissue was transferred to 8 Eppendorfs containing 0.5ml of extraction buffer (50mM Tris-HCl, 150mM LiCl, 5mM EDTA, 5% SDS pH9). For embryo tissue, 0.5g of starting material was used. For other plant tissue - roots, stem, leaf and pod- 2g of tissue was used. The ground tissue was equally distributed between the tubes. To each tube, 0.5ml of water saturated phenol:chloroform (1:1) was added and they were then placed on ice for 5 minutes, shaking occasionally. The phases were separated by centrifuging at 13,000rpm for 10 minutes @ 4°C.The phenol : chloroform extraction was repeated as many times as required to remove all trace of protein from the sample. An extraction with an equal volume of chloroform : isoamyl alcohol (24:1) removed any residual phenol. Contaminating DNA was removed by precipitating the RNA with sterile 8M LiCl to a

final concentration of 2M at -80°C for 2 hours or at -20°C overnight. RNA was pelleted by spinning at 13,000rpm for 10 minutes @ 4°C. The pellet was resuspended in 200µl of sterile distilled water. The LiCl precipitation step was repeated until no more contaminating DNA was observed when an aliquot was run out in a mini-gel. The RNA concentration was determined from the absorbance of the sample at 260nm. The RNA was then divided into aliquots and precipitated with 0.1 vol 3M sodium acetate pH5.5 and 2.5 vol ethanol. The RNA was stored in this form at -80°C until required, when it was spun down and resuspended in sterile distilled water to a concentration of 15µg in 10µl.

2.8 RNA gel electrophoresis and Northern blotting

The RNA was run out on a 1% gel:

1g agarose in 62.5ml dH_2O - sterile

20ml 5x gel running buffer - sterile

17.5ml formaldehyde (12.3M)

The 5x running buffer (0.2M MOPS, 50mM NaOAc, 5mM EDTA) was diluted to 1x concentration before use. Once the gel had set, it was allowed to stand in the running buffer for at least 30 minutes before loading the samples, to allow any formaldehyde in the wells to disperse. RNA samples were prepared as follows:

10µl RNA (containing 15µg)

4µl 5x buffer

7µl formaldehyde

20µl formamide (deionised)

The samples were heated for 15 minutes @ 55°C to denature any secondary structure which would affect the running of the molecules in the gel. After cooling, 8µl of loading buffer was added (50% glycerol, 1mM EDTA pH8, 0.4% bromo-phenol-blue, 0.4% xylene cyanol). The samples were loaded onto the gel and run at 60V for 3-4 hours in a fume cupboard. The gel was stained for 2-5 minutes in a solution of 4xSSC containing

1μg/ml ethidium bromide. Destaining was done in sterile distilled water, with frequent changes, until the RNA bands could be seen. Blotting was done without prior treatment of the gel.

The RNA was transferred over night onto nylon (Hybond-N) by capillary transfer, using 20xSSPE. When transfer was complete, the nylon was rinsed in 2xSSPE and then baked at 80°C for 90 minutes under vacuum.

To remove probe from RNA bound nylon filters so that they could be re-probed with a different probe, the filter was washed as follows. A boiling solution of 0.1% SDS was poured onto the filter and allowed to cool to room temperature. The wash was repeated a second time and the blot stored damp, wrapped in cling film @ 4°C.

2.9 Preparation and Transformation of *E. coli* DH5α competent cells

A single colony was used to inoculate 5ml of L-broth, and the cells grown at 37°C over night. Then 1ml of this culture was used to inoculate a 50ml aliquot of L-broth. The cells were grown for 2.5 to 3 hours at 37°C. After pelleting at 5,000rpm in a bench centrifuge, the cells were resuspended in 25ml of ice-cold 0.1M $CaCl_2$. After this they were again pelleted and resuspended in 5ml 0.1M $CaCl_2$. Either they were used immediately or stored at -80°C, after the addition of 3ml 80% glycerol.

An aliquot of competent cells (stored at -80°C) was thawed out on ice. To the ligation mixture, 100μl of ice cold 0.1M $CaCl_2$ was added. The ligation mixture was then mixed with the cells and left on ice for 30mins. The cells were floated in a 42°C water bath for 2 mins, then again placed on ice for 15 mins. The cells were grown up in 1ml of L-broth at 37°C for 90mins (constant rotation). Then they were pelleted at 13,000rpm in a microfuge, and resuspended in 100μl L-broth. Aliquots were then spread onto selection plates containing antibiotic.

To transform TG2 cells a similar procedure was carried out. An aliquot of competent cells was left on ice for 30 minutes. The ligation mixture was then added and the cells placed at 42°C for 2 minutes. After cooling on ice the cells were grown up and plated out as above.

2.10 Preparation of plating bacteria

A single colony was inoculated into 5ml L-broth containing 0.2% (w/v) maltose and incubated at 37°C over night. The cells were pelleted at 5,000rpm for 10 minutes in a bench centrifuge and resuspended in 2ml of 10mM $MgSO_4$.

2.11 Plating out of phage λ

Appropriate dilutions of the phage were prepared in SM buffer and 100µl of the phage dilutions were added to 100µl (1.25ml) of plating cells. (Note - the first quantities shown are for 85mm diameter petri dishes and those in brackets are for 25x25cm plates). The phage were left to adsorb for 20 minutes @ 37°C. Then 3ml (37.5ml) of molten top agarose at 42°C was mixed with each phage dilution and poured evenly onto a pre-warmed agar plate. The top agarose was prepared at half the concentration of the bottom agar. After drying, the plates were incubated for between 8 to 15 hours at 37°C. It was ensured that the incubation time was long enough to allow relatively large plaques to form, but not long enough for them to become confluent.

2.12 Isolation of Phage λ DNA (Miller, 1987)

Approximately $2x10^9$ plating cells were incubated with 10^7 pfu of phage. Ratios of plating cells to pfu were varied to obtain the optimum multiplicity of infection. Phage were adsorbed to the plating cells for 20 minutes at 37°C then added to 100ml of pre-warmed L-broth containing 0.2% magnesium sulphate. This was incubated overnight at 37°C with constant shaking. If phage infection was efficient, lysis of the bacterial cells was observed. Lysis of cells was completed by adding 0.5ml of chloroform and incubating at room temperature for 30 minutes. The cell debris

were pelleted at 8,000rpm for 15mins at room temperature. The supernatant was removed, then NaCl was added to a final concentration of 1M and PEG (6,000) to a concentration of 10% w/v. These were dissolved slowly on ice with stirring, then left on ice for at least 2 hours until the phage were precipitated. The phage were pelleted at 10,000rpm for 10 minutes at 4°C. The pellet was resuspended in 2ml of SM buffer containing a drop of chloroform and left over night. An equal volume of chloroform was added and the two phases mixed thoroughly. The top aqueous phase was removed leaving the debris (such as PEG) at the interface. RNase and DNase were added to a final concentration of 10μg/ml and incubated for 30 minutes at room temperature. Proteinase K was added to a final concentration of 50μg/ml and SDS to a final concentration of 0.5% and incubated at 56°C for 1-2 hours. After cooling, protein extraction was carried out several times with phenol:chloroform followed by a single choloform extraction. The DNA was precipitated for 30 minutes at room temperature with 10% (v/v) sodium acetate and 2 volumes of ethanol. This was pelleted by centrifuging for 10mins at 13,000rpm and the DNA washed with 70% ethanol, air dried and resuspended in TE.

2.13 Plaque lifts

After incubation was complete, the plates were left in the fridge for at least one hour before taking the plaque lift. A Hybond filter (nitrocellulose or nylon) was cut to the size of the plate. This was placed carefully onto the plate working from the middle outwards to avoid air bubbles forming, and left in contact with the plate for 30-60 seconds. The filter was marked so that its orientation on the plate was known. It was removed carefully, using blunt forceps. A second filter was then placed on the plate and left for 1-2 minutes, then marked as above. The DNA on the filters was denatured by immersing in denaturation buffer, with the DNA side facing upwards. After 60 seconds the filter was removed to neutralization buffer and left for 5 minutes. The filter was rinsed in 2xSSPE, then left on Whatman 3MM to air dry and then baked at 80°C under vacuum for 90 minutes to fix the DNA onto the filter.

2.14 Radio-labelling of DNA using the "Multiprime" system

(according to protocol published by Amersham)

For the labelling reaction, 0.025-0.1µg of the required DNA was isolated by one of the fragment isolation procedures described in section 2.4. Labelling was carried out as instructed in the manual. The highest labelling efficiency was obtained with an overnight incubation at room temperature.

Oleosin probe - The 455bp probe, consisted of the 5'- region of the Nap4 cDNA up to the Sac I site (see Figure 1 of chapter 3). The poly(A) tail was removed from the cDNA region used to construct the probe, since the presence of a long poly(A) tract in the cDNA probe would interfere with the binding specificity.

Napin probe - The pN2 clone consisted of a 738bp cDNA cloned into the PstI site of the pUC8 polylinker. The 600bp probe, consisted of the 5'- region of the pN2 cDNA cut at the HincII site to remove the poly(A) tract. The cDNA was supplied by Dr Marti Crouch.

2.15 Filter probing

First the filter was pre-hybridised in pre-hybridisation solution, (5xSSC, 5xDenhardts, 0.1% SDS, 50µg/ml denatured salmon DNA, 0.1% tetra-sodium pyrophosphate), to block all the free sites to which the probe would be able to non-specifically bind. The filter was pre-hybridised for 1-2 hours at 65°C.

Hybridisation was carried out over night at 65°C with 50µCi of probe in 30-50ml of hybridisation solution (5xSSC, 2xDenhardts, 0.1% SDS). The probe was then poured off and stored at -20°C, ready for re-use. The filter was washed several times. The first wash was in 2xSSC, 0.1% SDS at 65°C for 20 minutes. Washing was repeated to the required stringency by lowering the salt concentration.

2.16 DNA sequencing (Sanger, et al., 1977)

Manual - The basic dideoxy sequencing technique was used. Details of the method can be found in the Bethesda Research Labs (BRL) "M13 Cloning/Dideoxy Sequencing instruction manual" for M13 sequencing and in the Stratagene "Bluescript Exo/Mung DNA sequencing system instruction manual" for sequencing in SK^+. Subcloning into M13 was carried out as described in Sambrook et al (1989) [section 4.33].

Automatic - This was carried out by Miss Julia Bryden using the Applied Biosystems 373A DNA sequencer.

Denaturation proceedure

Preparation of single stranded templates from plasmid DNA was necessary prior to sequencing. A 100µl aliquot of mini-prep plasmid DNA (prepared as described in section 2.1) was made up to a final volume of 200µl with TE buffer. To this 4µl of 20mg/ml RNase and 10µl of 2M NaOH were added and the solution kept at room temperature for 5 minutes. Then 40µl of 5M ammonium acetate pH6 and 500µl ethanol were added and left at -80°C for 15 minutes. The sample was spun in an Eppendorf for 10 minutes at 13,000rpm and the pellet rinsed in 70% ethanol then dried under vacuum. Just before use, the pellet was dissolved in 20µl of sterile distilled water.

2.17 Protein extraction

Embryos were removed from their coats and ground in ice cold extraction buffer (50mM HEPES pH 7.9, 1mM EDTA, 200mM KCl, 1mM DTT, 2.5mM PMSF) using a pestle and mortar. An aliquot of 40 seeds was extraced with 700µl of extraction buffer.

2.18 SDS Polyacrylamide gel electrophoresis (Sambrook, et al., 1989)

A 12% gel was prepared, together with a 6% stacker as described in Sambrook et al (1989) [section 18.47]. An equal volume of sample buffer was added to the protein sample. After boiling for 2 minutes the sample was allowed to cool and

loaded onto the gel. The gel was run on a Biorad mini-gel apparatus @250V for approximately 1 hour in a Tris HCl buffer system. The run was stopped when the dye front was about 1cm from the bottom of the gel. Staining was carried out using a Coomassie Brilliant Blue stain as described in Samrook et al (1989) [section 18.55].

2.19 Western Blotting and Antibody probing of Western blots

After running the protein samples, the gel was placed in 100ml of transfer buffer pH8.3 (25mM Tris, 192 mM glycine, 20% methanol) for 15 mins. This was necessary to remove salts and detergents and excess water prior to blotting. Two sheets of nitrocellulose were cut to the same size as the gel and soaked for 15-30 mins in transfer buffer. Four sections of 3MM paper were also cut to the size of the gel and pre-soaked in transfer buffer. The gel was then sandwiched between the two sheets of nitrocellulose, with the two pre-soaked 3MM papers either side. Blotting was carried out at 100V/150mA for 2 hours in a Sartorius "Sartoblot II" apparatus.

To probe the blot, the free sites on the nitrocellulose were blocked with blocking buffer (10% Marvel Milk Powder in PBS) in a sealable bag on a shaker for one hour. The blocking buffer was poured off and diluted primary antibody was added at a dilution of x0.002 in PBS with 5% BSA. This was left on a shaker for one hour. Any unbound antibody was removed by washing with 4x5ml PBS, 15 mins each wash. The secondary antibody was then added - 2ml of x0.002 goat anti-rabbit Ig-alkaline phosphatase in PBS. This was left on the shaker for one hour. Unbound antibody was removed by washing with PBS (as for primary antibody). To detect the antibody, the blot was placed in 10ml of alkaline phosphatase buffer (100mM Tris pH 9.5, 100mM NaCl, 50mM $MgCl_2$). To this was added 44µl of nitroblue tetrazolium (50mg/ml) and 33µl 5-bromo-4-chloro-3-indolyl phosphate (100mg/ml). After mixing well, the blot was left until colour developed (2-15mins). The reaction was stopped by adding acetic acid to a final concentration of 1%.

2.20 Media & Buffers

L-Broth and L-Agar were made up as described in Sambrook et al (1989). The required quantity of media was made up, then autoclaved. Antibiotics and indicator substances which are heat labile were added to agar after cooling to 50°C. X-gal was dissolved in DMF and 1mg added per 20ml of agar. Ampicillin was added to a final concentration of 50μg/ml and tetracycline to a concentration of 25μg/ml.

Buffers

The following buffers were made up as described in Sambrook et al (1989) [sections B 20 and B 23]:

SSPE, SSC (20x), TBE (10x), TAE (10x), TE (1x) and SM (pH 7.5)

2.21 Oligonucleotides

Oligo1

5'-AAGTATGCAACGGGAGAG-3'

This corresponds to the region spanning bases 358-375 of the Nap4 cDNA sequence.

Oligo 3

5'-AGCAATCTGTCTAGACTT-3'

This is the anti-sense strand which spans bases 121-138 of Nap4.

2.22 Computer programs

DNA sequences were analysed by the DNA Strider programme. Sequence alignments were carried out using the Gap and Bestfit algorithms in the UWGCG suite of programmes at SERC SEQNET Daresbury.

CHAPTER 3 - RESULTS

3.1 CHARACTERISATION OF THE OLEOSIN GENE FAMILY FROM *B. napus*

3.1.1 INTRODUCTION

The isolation of cDNA's was a necessary prelude to other molecular studies. For instance, the cDNA Nap 4 was used to make a radioactive cDNA probe. This enabled expression studies to be carried out, where mRNA levels were followed at various stages of embryogenesis (see section 3.3.3). This probe was also used to screen a genomic library for oleosin genomic clones (see section 3.2.2).

Two oleosin cDNA clones were isolated by screening a *B. napus* λgt11 expression library with polyclonal antibodies raised to oleosin. The antibodies were raised against oleosin purified by SDS PAGE.

This section describes the characterisation of the two oleosin cDNA clones - Nap 4 and Nap 5. Detailed analysis of the coding region was carried out to identify hydrophobic and hydrophilic domains within the oleosin proteins. Sequence comparisons were made with the highly characterised maize oleosins KD16 and KD18 (Qu and Huang, 1990; Vance and Huang, 1987).

Structural analysis of these proteins was carried out and basic structural motifs such as potential α helical forming regions were identified. The 3' UTR's of the cDNA sequences were analysed with regard to the importance of this region in controlling polyadenylation of the mRNA (see section 1.4.1). This region is also potentially important in the regulation of mRNA stability.

Nap 4 was then used to construct a radiolabelled probe. Initially the probe was used to determine sequence relatedness of the oleosin genes by genomic Southern blotting.

3.1.2 SUBCLONING AND SEQUENCING OF NAP 4 AND NAP 5

Four oleosin cDNA clones were isolated from a λgt11 expression library constructed using RNA isolated from 8-10 week old *B. napus* embryos. The library was previously constructed and screened in this laboratory. *Eco RI* linkers were attached to the ends of the cDNA molecules and subcloned into the *Eco RI* site of λgt11. Screening was carried out using antibodies raised in rabbit, against oleosin purified by preparative SDS PAGE. The gel system was not capable of separating the different oleosin isoforms, therefore the antiserum contained antibodies to all the isoforms of the protein. Further characterisation of these clones showed that there were only two sequence types represented. Nap 4 was a unique clone whereas the Nap 5 sequence was identified in the other three clones.

Nap 4 was cut into two sections with *SacI* and *EcoRI*. The 5' region was 455bp and the 3' region 360bp (see Figure 3.1). These fragments were subcloned into M13 (see section 2.16). Sequencing of the 5' region was carried out on an Applied Biosystems automatic sequencer (see section 2.16). The 3' region was only sequenced in one direction due to the long poly(A) tail which inhibits the progress of the sequencing reaction. To determine whether an extra section of coding region bounded by two *Sac I* sites existed in the central region of the cDNA, it was necessary to sequence across the site. For this purpose, a 614bp *EcoRI/BclI* fragment was subcloned into Bluescript plasmid vector SK^+ (see Figures 3.1 & 3.2). A synthetic oligonucleotide - oligo 1 (see section 2.21) was used to sequence across the *SacI* site.

Nap 5 was subcloned in three separate sections: The 296bp 5' *EcoRI/BclI* fragment; the central 378 bp *PstI/SmaI* fragment; and the 243bp 3' *SmaI/EcoRI* fragment (see Figures 3.1 & 3.2). These were all subcloned into the Bluescript vector SK^+ and sequenced in both directions using the forward and reverse primers. For DNA to be cut with BclI, it was necessary to use the methylation deficient host GM119.

DNA was prepared from SK^+ for sequencing by the miniprep method described in section 2.1 and cleaned up with acid phenol extractions. Single stranded DNA for sequencing was prepared as described in section 2.16. Sequencing was carried out both manually and using an automatic sequencer, as described in section 2.19. The sequencing strategy is shown in Figure 3.1.

From the nucleotide sequences, detailed restriction maps and the derived amino acid sequences were obtained. Figure 3.2a and 3.2b show the unique restriction sites within Nap 4 and Nap 5, derived from the nucleotide sequences.

Nap 4 and Nap 5 were subcloned as complete cDNA's into the *EcoRI* site of SK^+. The digests were then carried out on minipreped DNA (see section 2.1). Figure 3.2c shows selected restriction digests for these clones, which confirm some of the derived sites.

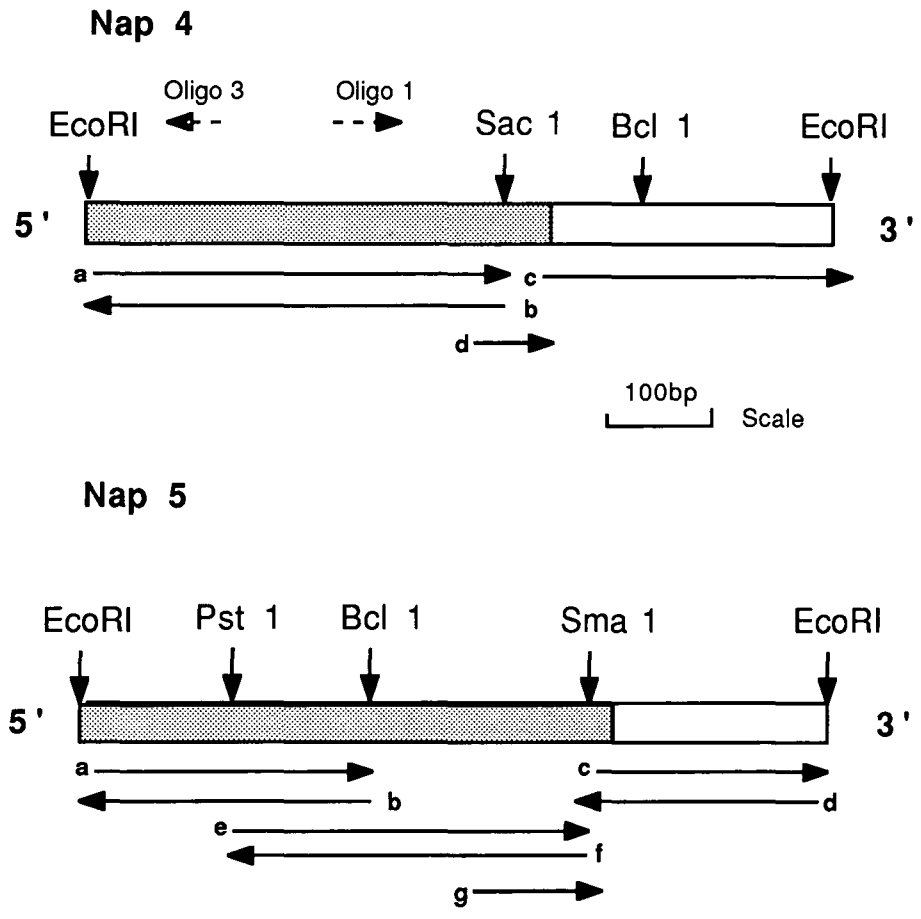
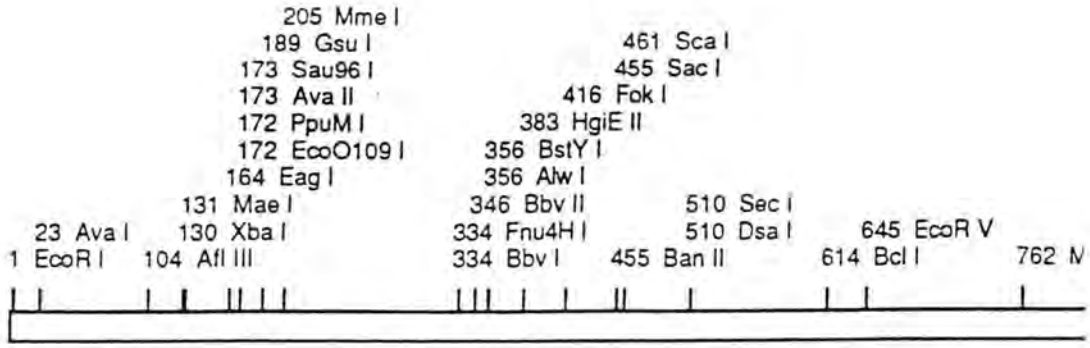


FIGURE 3.1

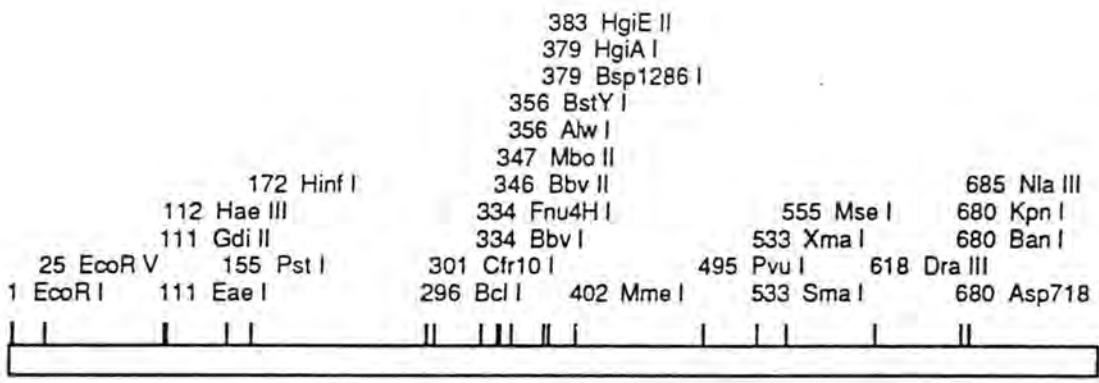
Sequencing strategies used to sequence Nap 4 and Nap 5

The sequence between the EcoRI sites covers the length of the original cDNA inserts in λgt11. The restriction sites used in subcloning are shown. The region sequenced in each sequencing run and the direction of sequencing are shown by horizontal arrows. The shaded sections represent the coding regions. The position of the synthetic oligonucleotide sequences, oligo 1 and oligo 3, derived from Nap 4 are indicated by dotted lines with arrows showing the orientations.

(a)


205 Mme I
189 Gsu I
173 Sau96 I
173 Ava II
172 PpuM I
172 EcoO109 I
164 Eag I
131 Mae I
23 Ava I
130 Xba I
1 EcoR I
104 Afl III
461 Sca I
455 Sac I
416 Fok I
383 HgiE II
356 BstY I
356 Alw I
346 Bbv II
334 Fnu4H I
334 Bbv I
510 Sec I
510 Dsa I
455 Ban II
645 EcoR V
614 Bcl I
762 N

(b)


383 HgiE II
379 HgiA I
379 Bsp1286 I
356 BstY I
356 Alw I
347 Mbo II
172 Hinf I
346 Bbv II
112 Hae III
334 Fnu4H I
111 Gdi II
334 Bbv I
25 EcoR V
155 Pst I
301 Cfr10 I
1 EcoR I
111 Eae I
296 Bcl I
402 Mme I
555 Mse I
533 Xma I
495 Pvu I
618 Dra III
533 Sma I
685 Nla III
680 Kpn I
680 Ban I
680 Asp718

(c)

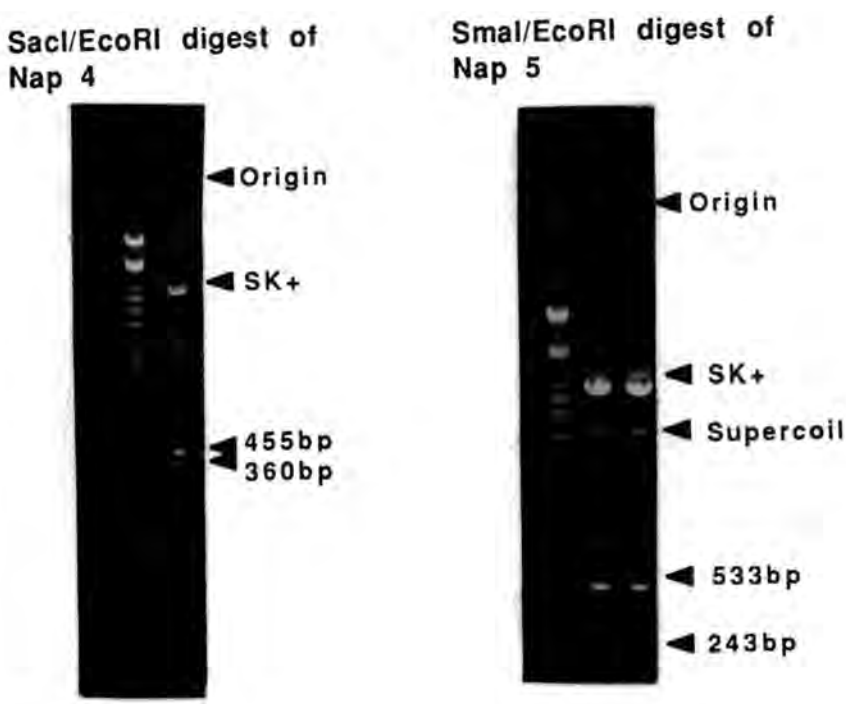

SacI/EcoRI digest of Nap 4
Origin
SK+
455bp
360bp
SmaI/EcoRI digest of Nap 5
Origin
SK+
Supercoil
533bp
243bp

FIGURE 3.2 Restriction maps and digests of the oleosin cDNA clones Nap 4 and Nap 5

3.2a Unique restriction sites within Nap 4

3.2b Unique restriction sites within Nap 5

3.2c Restriction digests of Nap 4 and Nap 5

Restriction enzyme digests were carried out with 100-200ng of miniprep DNA (prepared as described in section 2.1). Nap 4 was incubated with 10Units of *SacI,* followed by 10Units of *EcoRI* both digests being carried out for 2 hours at 37°C. Similarly, Nap 5 was digested with *SmaI* followed by *EcoRI.* The molecular weight markers are present in lane 1, and consist of phage λ DNA digested with *PstI* (prepared as described in section 2.5). The significant bands are labelled. The extra band (S) visible on the gel showing the digest of Nap 5 is probably an artefact due to supercoiling of the plasmid DNA.

```
CGT CGA GAT CAG TAT CCC CGA GAC CGA GAC CAG TAT TCT ATG ATC
Arg Arg Asp Gln Tyr Pro Arg Asp Arg Asp Gln Tyr Ser Met Ile 15

GGT CGA GAC CGA GAC AAG TAT TCC ATG ATT GGC CGA GAC CGA GAC
Gly Arg Asp Arg Asp Lys Tyr Ser Met Ile Gly Arg Asp Arg Asp 30

CAG TAC AAC ATG TAT GGT CGA GAC TAC TCC AAG TCT AGA CAG ATT
Gln Tyr Asn Met Tyr Gly Arg Asp Tyr Ser Lys Ser Arg Gln Ile 45

GCT AAG GCT GTT ACC GCA GTC ACG GCC GGT GGG TCC CTT CTT GTC
Ala Lys Ala Val Thr Ala Val Thr Ala Gly Gly Ser Leu Leu Val 60

CTC TCC AGT CTC ACC CTT GTC GGA ACT GTC ATT GCT CTG ACT GTT
Leu Ser Ser Leu Thr Leu Val Gly Thr Val Ile Ala Leu Thr Val 75

GCG ACT CCT CTG CTT GTT ATC TTT AGT CCA ATC CTT GTC CCT GCT
Ala Thr Pro Leu Leu Val Ile Phe Ser Pro Ile Leu Val Pro Ala 90

CTC ATC ACC GTT GCA TTG CTC ATC ACC GGC TTT CTC TCC TCT GGT
Leu Ile Thr Val Ala Leu Leu Ile Thr Gly Phe Leu Ser Ser Gly 105

GGC TTT GGC ATT GCA GCT ATA ACC GTC TTC TCT TGG ATC TAC AAG
Gly Phe Gly Ile Ala Ala Ile Thr Val Phe Ser Trp Ile Thr Lys 120

TAT GCA ACG GGA GAG CAC CCA CAA GGG TCA GAT AAA CTG GAC AGT
Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys Leu Asp Ser 135

GCA AGG ATG AAG CTG GGA GGC AAA GTT CAG GAT ATG AAG GAC AGA
Ala Arg Met Lys Leu Gly Gly Lys Val Gln Asp Met Lys Asp Arg 150

GCT CAG TAC TAT GGA CAA CAG CAA ACA GGT GGG GAA CAC GAC CGT
Ala Gln Tyr Tyr Gly Gln Gln Gln Thr Gly Gly Glu His Asp Arg 165

GAC CGT ACC CGT GGA ACC CAG CAC ACT ACC TAA
Asp Arg Thr Arg Gly Thr Gln His Thr Thr end

ATTACGCCATGACTATTTTCATAGTCCAATAAGGCTGATGTCGGGAGTCCAGTTTATGA

GCAATAAGGTGTTTAGAATTTGATCAATGTTTATAATAAAAGGGGGAAGATGATATCAC

AGTCTTTTTTTCTTTTTGGCTTTTGTTAAATTTGTGTGTTTCTATTTGTAAACCTTCTG

TATATGTTGTACTTCTTTCCCTTTTTAAGTGGTATCGTCTATATGGTAAAACGTTATGA

TTTGGTCTTTCCTTTTCTCTGTTTAGGATAAAAAAGACTGCATGTTT(A)76
```

FIGURE 3.3
Complete sequence of the Nap 4 oleosin cDNA

Internal repeat sequences in the N-terminal region are underlined. In both Nap 4 and Nap 5 these repeat sequences contain the basic sequence element N-Arg Asp Gln Tyr-C. The 3' UTRs of Nap 4 and Nap 5 are shown. Imperfect polyadenylation signals are overlined and perfect consensus sequences (AATAAA) are boxed.

```
CCA GCT AGA ACC CAT CAC GAT ATC ACC ACC CGA GAT CAG TAT CCA
Pro Ala Arg Thr His His Asp Ile Thr Thr Arg Asp Gln Tyr Pro15

TTG ATT AGC CGA GAC CGA GAC CAG TAT GGG ATG ATT GGT CGG GAC
Leu Ile Ser Arg Asp Arg Asp Gln Tyr Gly Met Ile Gly Arg Asp30

CAG TAC AAT ATG TCC GGC CAA AAC TAC TCC AAG TCT AGG CAG ATT
Gln Tyr Asn Met Ser Gly Gln Asn Tyr Ser Lys Ser Arg Gln Ile45

GCT AAA GCT ACC ACT GCA GTC ACC GCA GGG GAT TCT CTC CTT GTT
Ala Lys Ala Thr Thr Ala Val Thr Ala Gly Asp Ser Leu Leu Val60

CTC TCC AGT CTC ACC CTT GTG GGA ACG GTC ATT GCA TTG ATT GTT
Leu Ser Ser Leu Thr Leu Val Gly Thr Val Ile Ala Leu Ile Val75

GCC ACT CCA CTG CTT GTT ATC TTT AGC CCA ATC CTA GTC CCC GCT
Ala Thr Pro Leu Leu Val Ile Phe Ser Pro Ile Leu Val Pro Ala90

CTC ATC ACC GTC GCA CTG CTG ATC ACC GGC TTT CTC TCC TCT GGT
Leu Ile Thr Val Ala Leu Leu Ile Thr Gly Phe Leu Ser Ser Gly105

GCA TTT GGC ATT GCA GCT ATA ACG GTC TTC TCT TGG ATC TAC AAG
Ala Phe Gly Ile Ala Ala Ile Thr Val Phe Ser Trp Ile Tyr Lys120

TAT GCA ACG GGA GAG CAC CCA CAG GGG TCA GAT AAG TTG GAC AGT
Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys Leu Asp Ser135

GCA AGG ATG AAG CTG GGA AGC AAG GCT CAG GAT ATG AAG GAC AGA
Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp Met Lys Asp Arg150

GCT TAC TAC TAT GGA CAA CAA CAT ACA GGT GAG GAA CAC GAT CGG
Ala Tyr Tyr Tyr Gly Gln Gln His Thr Gly Glu Glu His Asp Arg165

GAC CGT GAC CAC CGC ACC GAC CGT GAC CGC ACC CGG GGA ACC CAA
Asp Arg Asp His Arg Thr Asp Arg Asp Arg Thr Arg Gly Thr Gln180

CAC ACT ACT TAA
His Thr Thr end

ATTACGCCAAGACTATAATAATTTGTTTAGGGTGGATAATAAAAGGGGGAAGCCGATCT

CACATAGTGTTTTTTTCCCTTTGTCATAAATAATGTGCAAGAGTGTGGTTCTCTTGTCA

CACGGTACCATGTTTTGTTGTGTTATTCTCAGCTTATGTTGTACGTTCTTTGCCTTGTT

TGTAAGTTGTTACGTCTATATGGGAAATGTCTTGTTTTACG(A)10
```

FIGURE 3.4
Complete sequence of the Nap 5 oleosin cDNA

3.1.3 SEQUENCE ANALYSIS OF NAP 4 AND NAP 5 cDNA CLONES

The nucleotide and derived amino acid sequences for Nap 4 and Nap 5 cDNAs are shown in Figures 3.3 and 3.4. The open reading frames are 525bp and 549bp long respectively. These are incomplete sequences due to the limitations of the method used to synthesise cDNA from mRNA molecules resulting in part of the N-terminal region being lost.

N-LINKED GLYCOSYLATION SITES

Potential N-linked glycosylation sites have been identified in both Nap 4 and Nap 5. The requirement for an N-linked glycosylation site is the sequence Asn -X-Ser/Thr where X is any amino acid except proline. The sequence Asn-Met-Tyr is found in Nap 4 at residue 33. Nap 5 has two potential glycosylation sites : Asn-Met-Ser at residue 33 and Asn-Tyr-Ser at residue 38.

Note : The cDNA clones are not full length since some of the sequence from the 5' region was lost during the process of reverse transcription from mRNA. Therefore the positions of the residues stated are not relative to the true N-termini of the RNAs. To determine the length of the complete reading frame, either the corresponding genomic clone would have to be isolated or the 5' region sequenced directly from the mRNA by primer extension (Culzone, et al., 1987) .

INTERNAL REPEATS AND INSERTED SEQUENCES

A series of internal direct repeats have been identified in the N-terminal regions of Nap 4 and Nap 5 as shown in Figures 3.3 and 3.4. The basic sequence element is N -Arg Asp Gln Tyr- C. Internal repeats are also present in the N-terminal region of Nap 3. No repeats are present in the maize oleosins KD 16 and 18 or the carrot oleosin DC 59 (see Figure 3.8). An 8 amino acid direct repeat is present in the C-terminal region of the Nap 3 genomic clone (indicated by arrows in Figure 3.6). This repeat is not present in the cDNA clones Nap 4 and Nap 5. Nap 5 has an 8 amino acid sequence

insert in the N-terminal region, which is absent from the Nap 3 and Nap 4 sequences. This sequence is not a repeat sequence.

POLYADENYLATION SIGNALS

The positions of putative polyadenylation signals in the 3' UTR's of the cDNA molecules Nap 3 and Nap 4 are shown in Figures 3.4 and 3.5 respectively. Both have the perfect polyadenylation consensus sequence AATAAA which is assumed to direct polyadenylation at a higher frequency than the imperfect sequences. A single plant gene may produce mRNA's polyadenylated at multiple sites within the cellular mRNA pool.

A summary of the polyadenylation signals positions relative to the poly(A) addition site is shown below:

TABLE 3.1
Polyadenylation signals in the 3' UTR's of Nap 4 & 5

	Sequence	Position (bp)
Nap 4	AATAAG	-250
	AATAAG	-216
	AATAAA	-184
	GATAAA	-15
Nap 5	AATAAT	-196
	AATAAA	-175
	AATAAT	-125

3.1.4 SEQUENCE COMPARISONS

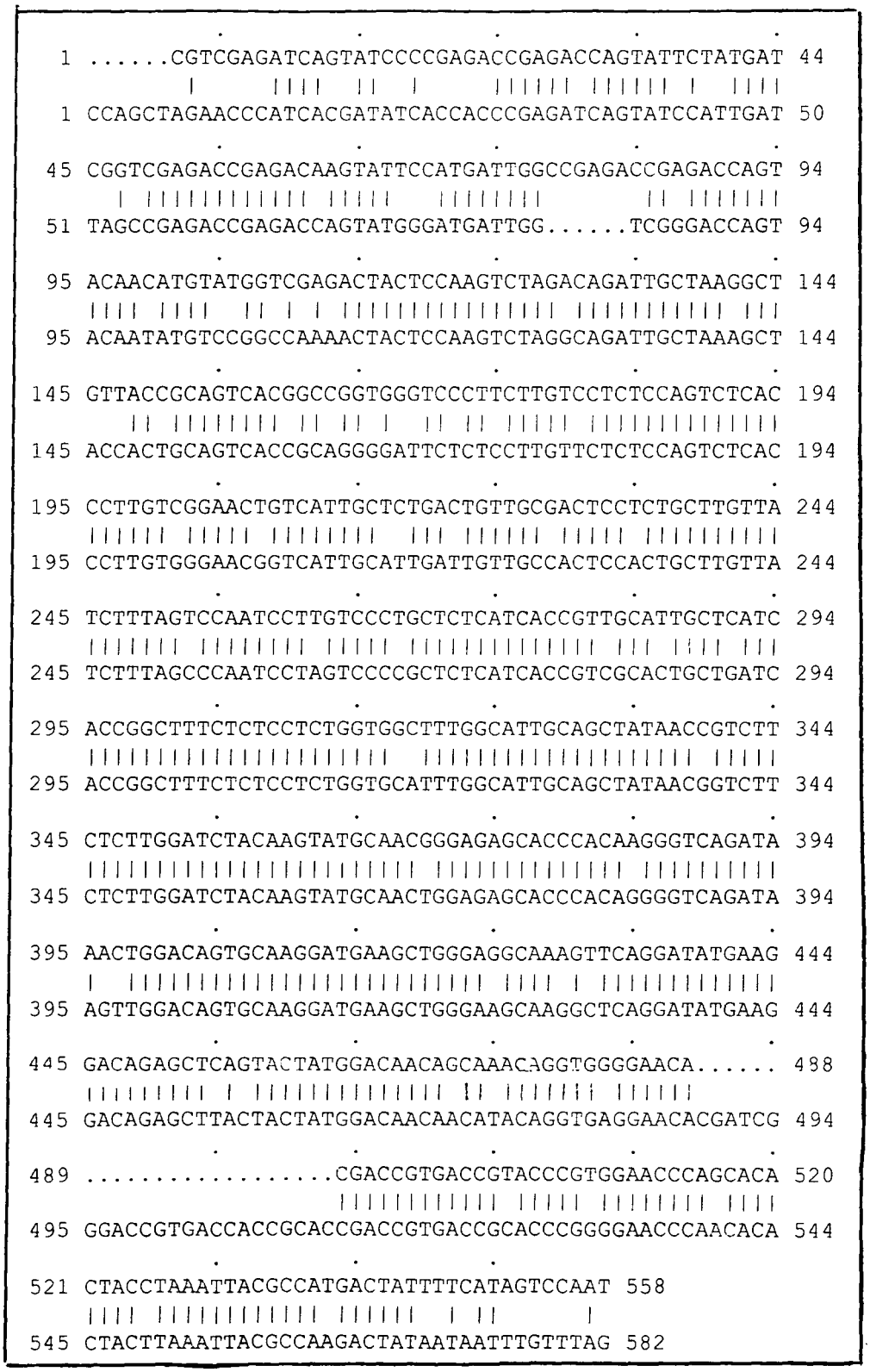
Sequence comparison of Nap 4 and Nap 5 at the nucleotide level are shown in Figure 3.5. The analysis was carried out using the Gap computer programme. Comparison for the values obtained for degree of sequence identity obtained using the Bestfit and Gap programmes is shown below:

	Bestfit	Gap
Coding Region	86.7%	83.5%
3' Untranslated Region	73.7%	56.4%

The Gap method of analysis has taken into account the whole of the sequence being analysed when calculating the sequence identity. The Bestfit analysis has only taken into account the regions which show maximum alignment. The results from the Gap analysis were considered the most reliable.

From the results of the Gap analysis, it can be seen that the coding region is highly conserved at the nucleotide level between Nap 4 and Nap 5. The 3' untranslated region is far less highly conserved. Much of the apparent sequence identity in this region is due to the characteristic high AT content present in all 3' untranslated regions.

(a)


1 ......CGTCGAGATCAGTATCCCCGAGACCGAGACCAGTATTCTATGAT 44
1 CCAGCTAGAACCCATCACGATATCACCACCCGAGATCAGTATCCATTGAT 50
45 CGGTCGAGACCGAGACAAGTATTCCATGATTGGCCGAGACCGAGACCAGT 94
51 TAGCCGAGACCGAGACCAGTATGGGATGATTGG......TCGGGACCAGT 94
95 ACAACATGTATGGTCGAGACTACTCCAAGTCTAGACAGATTGCTAAGGCT 144
95 ACAATATGTCCGGCCAAAACTACTCCAAGTCTAGGCAGATTGCTAAAGCT 144
145 GTTACCGCAGTCACGGCCGGTGGGTCCCTTCTTGTCCTCTCCAGTCTCAC 194
145 ACCACTGCAGTCACCGCAGGGGATTCTCTCCTTGTTCTCTCCAGTCTCAC 194
195 CCTTGTCGGAACTGTCATTGCTCTGACTGTTGCGACTCCTCTGCTTGTTA 244
195 CCTTGTGGGAACGGTCATTGCATTGATTGTTGCCACTCCACTGCTTGTTA 244
245 TCTTTAGTCCAATCCTTGTCCCTGCTCTCATCACCGTTGCATTGCTCATC 294
245 TCTTTAGCCCAATCCTAGTCCCCGCTCTCATCACCGTCGCACTGCTGATC 294
295 ACCGGCTTTCTCTCCTCTGGTGGCTTTGGCATTGCAGCTATAACCGTCTT 344
295 ACCGGCTTTCTCTCCTCTGGTGCATTTGGCATTGCAGCTATAACGGTCTT 344
345 CTCTTGGATCTACAAGTATGCAACGGGAGAGCACCCACAAGGGTCAGATA 394
345 CTCTTGGATCTACAAGTATGCAACTGGAGAGCACCCACAGGGGTCAGATA 394
395 AACTGGACAGTGCAAGGATGAAGCTGGGAGGCAAAGTTCAGGATATGAAG 444
395 AGTTGGACAGTGCAAGGATGAAGCTGGGAAGCAAGGCTCAGGATATGAAG 444
445 GACAGAGCTCAGTACTATGGACAACAGCAAACAGGTGGGGAACA..... 488
445 GACAGAGCTTACTACTATGGACAACAACATACAGGTGAGGAACACGATCG 494
489 .....CGACCGTGACCGTACCCGTGGAACCCAGCACA 520
495 GGACCGTGACCACCGCACCGACCGTGACCGCACCCGGGGAACCCAACACA 544
521 CTACCTAAATTACGCCATGACTATTTTCATAGTCCAAT 558
545 CTACTTAAATTACGCCAAGACTATAATAATTTGTTTAG 582

(b)

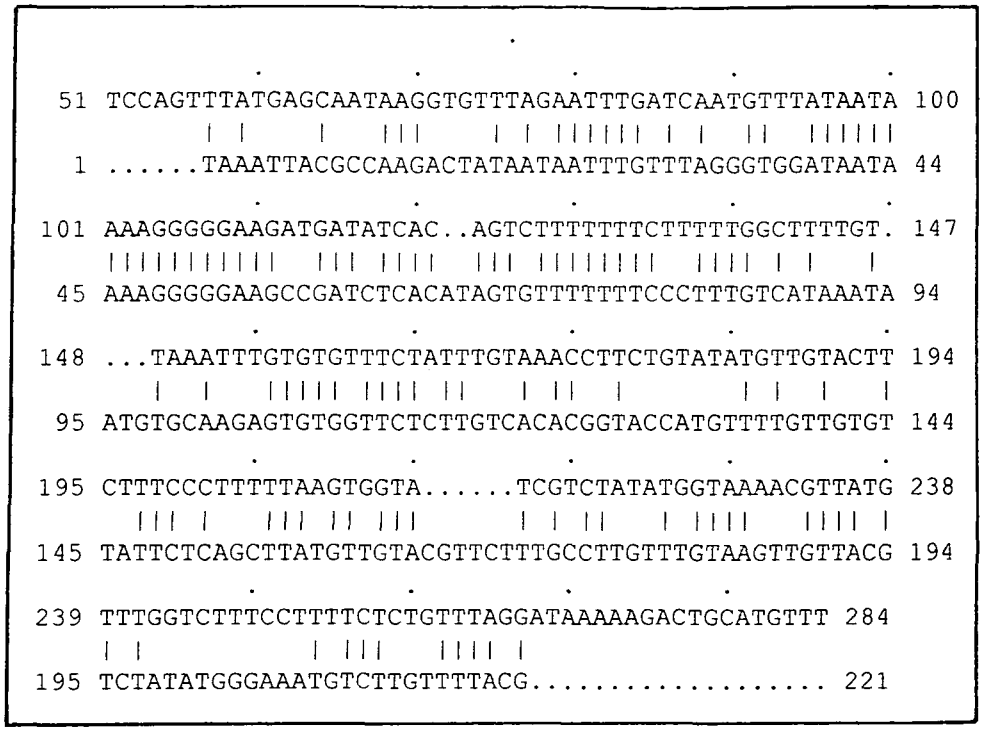


FIGURE 3.5

Alignment of the coding regions and the 3' untranslated regions of Nap 4 and Nap 5

Figure 3.5a shows the alignment of the nucleotide sequences of the coding regions of Nap 4 and Nap 5 using the Gap programme (see section 2.22). From this analysis, the coding regions were shown to have 83.5% identity over the whole coding region sequence.

Figure 3.5b shows the alignment of the 3' untranslated regions of Nap 4 and Nap 5 using the Gap programme. Over the whole of the 3' end there was only 56.4% identity at the nucleotide level.

Figure 3.6 compares the derived amino acid sequences of the genomic clone Nap3, the cDNA clones Nap 4 and 5 and the sequences for Nap 1 and 2 obtained by direct sequencing of polypeptide fragments (Murphy, et al., 1991c). The sequences for Nap 1 and 2 are incomplete due to problems encountered during sequencing such as N-terminal blockages. Nucleotide sequences were obtained for the Nap 3 genomic clone and the Nap 4 and Nap 5 cDNA clones.

The amino acids were aligned so as to give maximum sequence identity. Any mismatches are highlighted with open boxes. Three proline residues towards the middle of the sequence are completely conserved in Nap 1, 3, 4 & 5 (no sequence available for Nap 2) as indicated by shaded boxes. There is a high degree of sequence similarity between Nap 2, 3, 4 & 5, especially in the central hydrophobic domain. Nap 1 shows greater sequence divergence and may belong to a different isoform class to the other oleosins. It is not possible from this data to predict which of the clones corresponds to the various oleosin proteins isolated by SDS PAGE (see section 1.2.3).

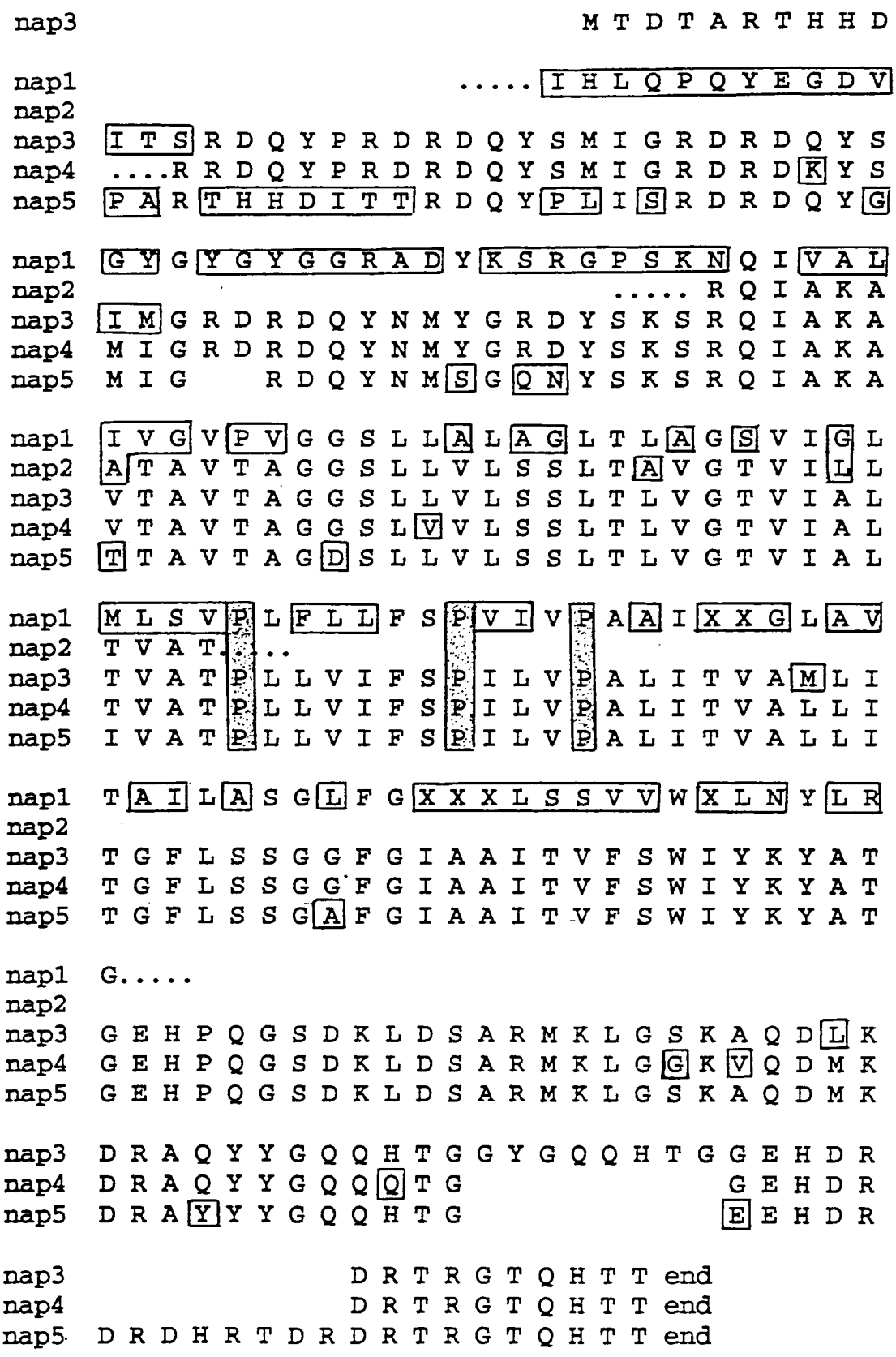
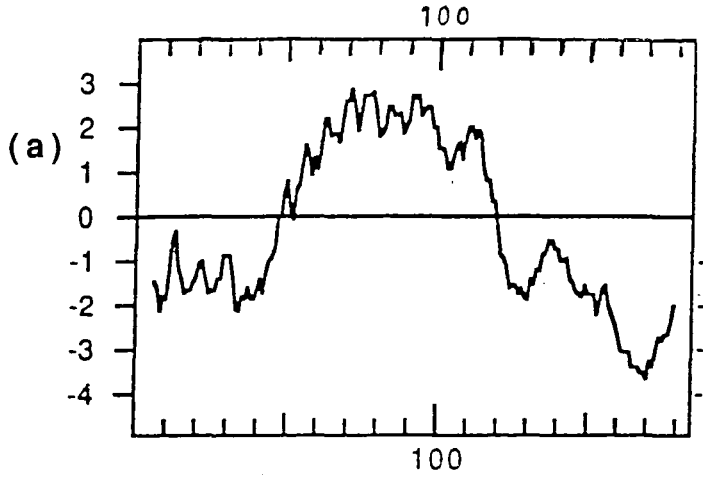

nap3 M T D T A R T H H D
nap1 ..... I H L Q P Q Y E G D V
nap2
nap3 I T S R D Q Y P R D R D Q Y S M I G R D R D Q Y S
nap4 ....R R D Q Y P R D R D Q Y S M I G R D R D K Y S
nap5 P A R T H H D I T T R D Q Y P L I S R D R D Q Y G
nap1 G Y G Y G Y G G R A D Y K S R G P S K N Q I V A L
nap2 ..... R Q I A K A
nap3 I M G R D R D Q Y N M Y G R D Y S K S R Q I A K A
nap4 M I G R D R D Q Y N M Y G R D Y S K S R Q I A K A
nap5 M I G R D Q Y N M S G Q N Y S K S R Q I A K A
nap1 I V G V P V G G S L L A L A G L T L A G S V I G L
nap2 A T A V T A G G S L L V L S S L T A V G T V I L L
nap3 V T A V T A G G S L L V L S S L T L V G T V I A L
nap4 V T A V T A G G S L V V L S S L T L V G T V I A L
nap5 T T A V T A G D S L L V L S S L T L V G T V I A L
nap1 M L S V P L F L L F S P V I V P A A I X X G L A V
nap2 T V A T. . ..
nap3 T V A T P L L V I F S P I L V P A L I T V A M L I
nap4 T V A T P L L V I F S P I L V P A L I T V A L L I
nap5 I V A T P L L V I F S P I L V P A L I T V A L L I
nap1 T A I L A S G L F G X X X L S S V V W X L N Y L R
nap2
nap3 T G F L S S G G F G I A A I T V F S W I Y K Y A T
nap4 T G F L S S G G F G I A A I T V F S W I Y K Y A T
nap5 T G F L S S G A F G I A A I T V F S W I Y K Y A T
nap1 G.....
nap2
nap3 G E H P Q G S D K L D S A R M K L G S K A Q D L K
nap4 G E H P Q G S D K L D S A R M K L G G K V Q D M K
nap5 G E H P Q G S D K L D S A R M K L G S K A Q D M K
nap3 D R A Q Y Y G Q Q H T G G Y G Q Q H T G G E H D R
nap4 D R A Q Y Y G Q Q Q T G G E H D R
nap5 D R A Y Y Y G Q Q H T G E E H D R
nap3 D R T R G T Q H T T end
nap4 D R T R G T Q H T T end
nap5 D R D H R T D R D R T R G T Q H T T end

FIGURE 3.6

Comparison of oleosin amino acid sequences from *B. napus*

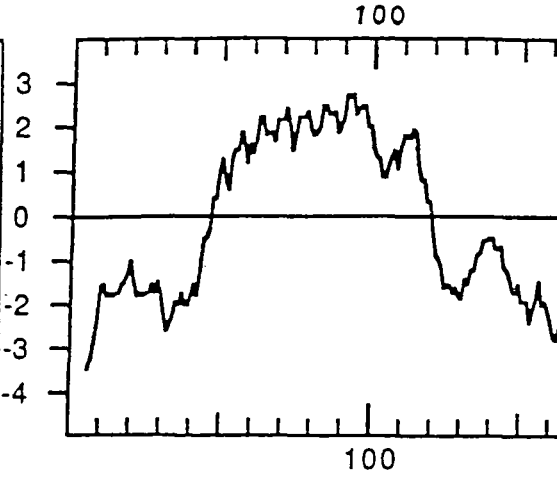
Nap 1 and 2 were obtained by direct sequencing of polypeptide fragments (Murphy, et al., 1991c). The Nap 3 sequence is derived from the nucleotide sequence of the corresponding genomic clone (see section 3.2). Nap 4 and 5 sequences are derived are derived from the corresponding cDNA clone sequences (see section 3.1.2). Sequence alignments were made to maximise matches. Non-conserved amino acids are highlighted by boxes. Shaded boxes represent highly conserved prolines. An internal repeat near the C-terminal of Nap 3 is underlined and the direction of the repeats indicated by horizontal arrows. A vertical arrow above the Nap 3 sequence shows the position of an intron in the genomic clone.

(a)



Residue number

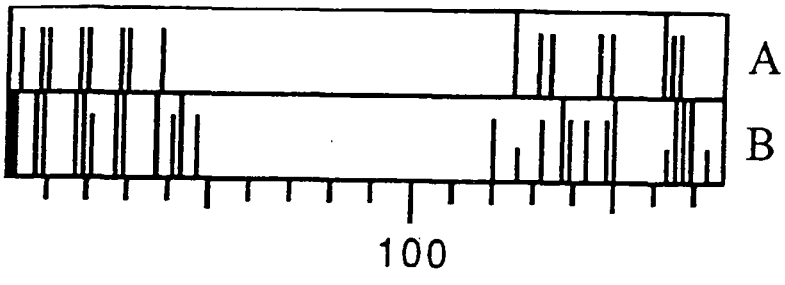
Nap 4



Residue number

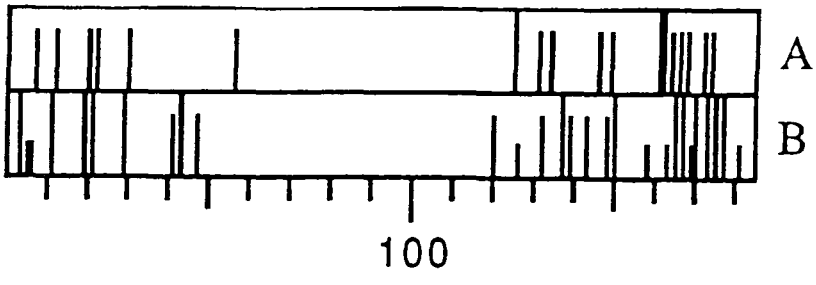
Nap 5

(b)



Residue number

Nap 4



Residue number

Nap 5

FIGURE 3.7a

Hydropathy plots of Nap 4 and Nap 5

Hydropathy plots for Nap 4 and 5 constructed by the method of Kyte and Doolittle (Kyte and Doolittle, 1982). The information was processed using the "DNA Strider" software package (see section 2.22). The hydropathy plots represent the range of hydrophobicity and hydrophilicity of the residues within a protein molecule. Hydrophobic regions are represented by positive values on the hydropathy plot and hydrophilic regions are represented by negative values.

FIGURE 3.7b

Acid/Basic maps of Nap 4 and Nap 5

Plots of the distribution of the acidic and basic amino acids were carried out on Nap 4 and Nap 5 using the "DNA Strider" software package. The distribution of the acidic amino acids aspartate and glutamate are shown in the boxed region A and the basic amino acids histidine, lysine and arginine are shown in the boxed region B.

The hydropathy plots show that the N- and C- termini of oleosin proteins are hydrophilic whilst the central region is hydrophobic (Figure 3.7a). The hydrophobic region spans from residue 44-119 in Nap 4 and from residue 48-115 in Nap 5. The results obtained from the acid/basic plots (Figure 3.7b) mirror those obtained from the hydropathy plots. There are no acidic or basic amino acids within the central hydrophobic region, however they are present in high abundance at the N- and C-terminal regions. This is to be expected since the acidic and basic amino acids are hydrophilic and highly charged. Based on the hydropathy plots in Figure 3.7a, the translated sequences of the cDNAs were subdivided into three regions: hydrophilic N- and C- terminal domains and a hydrophobic central domain. These subdivisions are shown in Table 3.2. The sequence identity of each of these regions was then compared using the Gap algorithm (see section 2.22).

TABLE 3.2

Sequence identity at the amino acid level of the domains of Nap 4 & 5

cDNA	N-Terminal Domain	Central Domain	C-Terminal Domain
Nap 4	1-43	44-119	120-176
Nap 5	1-47	48-115	116-183
Identity	75.6%	97.1%	96.4%

From the results shown in Table 3.2 the N-terminal hydrophilic region is the least highly conserved region. The central hydrophilic domain and the C-terminal domain are both highly conserved. The results in Figure 3.8 of oleosin sequences from carrot, maize and rapeseed show that there is significant sequence conservation between the oleosins from these distantly related species. The region having greatest sequence conservation is the central hydrophobic domain. A notable feature is the conservation of the three proline residues (highlighted by shaded boxes). The hydrophilic N- and C- terminal regions show great sequence divergence.

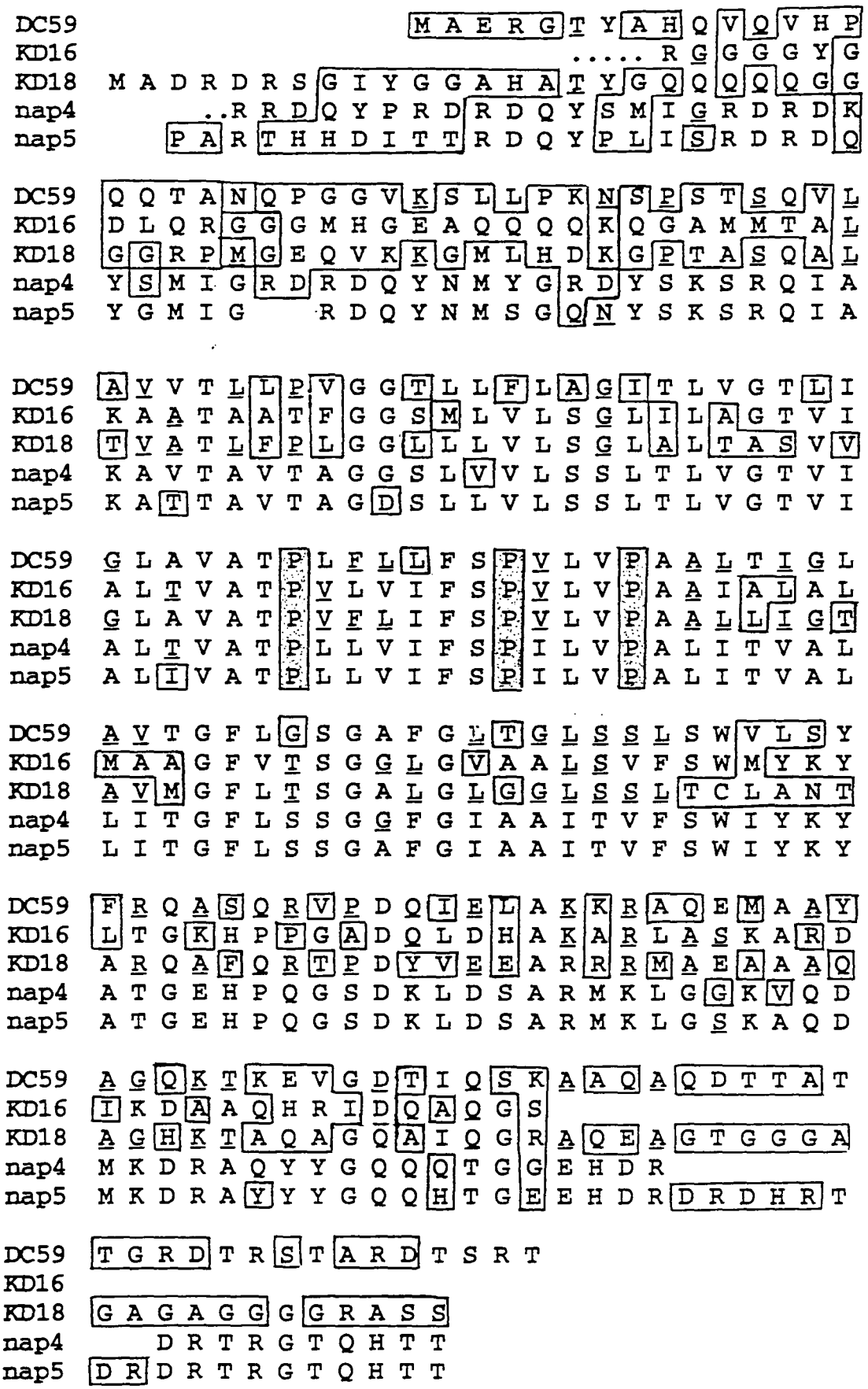

DC59 M A E R G T Y A H Q V Q V H P
KD16 ..... R G G G G Y G
KD18 M A D R D R S G I Y G G A H A T Y G Q Q Q Q Q G G
nap4 ..R R D Q Y P R D R D Q Y S M I G R D R D K
nap5 P A R T H H D I T T R D Q Y P L I S R D R D Q
DC59 Q Q T A N Q P G G V K S L L P K N S P S T S Q V L
KD16 D L Q R G G G M H G E A Q Q Q Q K Q G A M M T A L
KD18 G G R P M G E Q V K K G M L H D K G P T A S Q A L
nap4 Y S M I G R D R D Q Y N M Y G R D Y S K S R Q I A
nap5 Y G M I G R D Q Y N M S G Q N Y S K S R Q I A
DC59 A V V T L L P V G G T L L F L A G I T L V G T L I
KD16 K A A T A A T F G G S M L V L S G L I L A G T V I
KD18 T V A T L F P L G G L L L V L S G L A L T A S V V
nap4 K A V T A V T A G G S L V V L S S L T L V G T V I
nap5 K A T T A V T A G D S L L V L S S L T L V G T V I
DC59 G L A V A T P L F L L F S P V L V P A A L T I G L
KD16 A L T V A T P V L V I F S P V L V P A A I A L A L
KD18 G L A V A T P V F L I F S P V L V P A A L L I G T
nap4 A L T V A T P L L V I F S P I L V P A L I T V A L
nap5 A L I V A T P L L V I F S P I L V P A L I T V A L
DC59 A V T G F L G S G A F G L T G L S S L S W V L S Y
KD16 M A A G F V T S G G L G V A A L S V F S W M Y K Y
KD18 A V M G F L T S G A L G L G G L S S L T C L A N T
nap4 L I T G F L S S G G F G I A A I T V F S W I Y K Y
nap5 L I T G F L S S G A F G I A A I T V F S W I Y K Y
DC59 F R Q A S Q R V P D Q I E L A K K R A Q E M A A Y
KD16 L T G K H P P G A D Q L D H A K A R L A S K A R D
KD18 A R Q A F Q R T P D Y V E E A R R R M A E A A A Q
nap4 A T G E H P Q G S D K L D S A R M K L G G K V Q D
nap5 A T G E H P Q G S D K L D S A R M K L G S K A Q D
DC59 A G Q K T K E V G D T I Q S K A A Q A Q D T T A T
KD16 I K D A A Q H R I D Q A Q G S
KD18 A G H K T A Q A G Q A I Q G R A Q E A G T G G G A
nap4 M K D R A Q Y Y G Q Q Q T G G E H D R
nap5 M K D R A Y Y Y G Q Q H T G E E H D R D R D H R T
DC59 T G R D T R S T A R D T S R T
KD16
KD18 G A G A G G G G R A S S
nap4 D R T R G T Q H T T
nap5 D R D R T R G T Q H T T

FIGURE 3.8

Sequence comparisons of oleosin proteins from carrot (DC59), maize (KD 16 & 18) and *B. napus* (Nap 4 & 5).

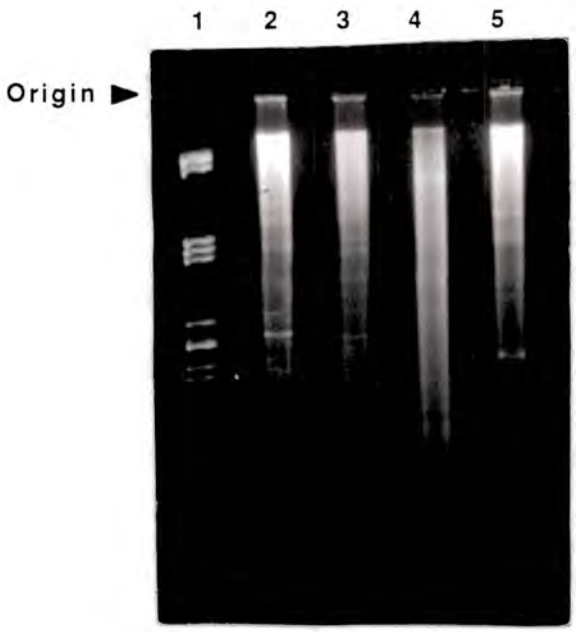
Alignment of the derived amino acid sequences of the oleosin proteins DC59 from carrot (Hatzopoulos, et al., 1990), KD16 (Vance and Huang, 1987) and KD18 (Qu and Huang, 1990) from maize and Nap 4 and 5 from rapeseed. Residues which are completely mismatched are boxed. Where there are two conserved amino acid types at one position, these are differentiated by underlining of one of the conserved sets. Shaded boxes represent conserved proline residues.

3.1.5 IDENTIFICATION OF OLEOSIN GENES FROM *Bam*HI FRAGMENTS OF A GENOMIC SOUTHERN

Figure 3.9 shows a Southern blot of *Brassica napus* DNA, used to estimate oleosin gene copy number. Washing was carried out at low (2xSSC) and high (0.2xSSC) stringency to determine whether sequence variation of the separate members of the family could be distinguished with this probe. After the low stringency wash, probe was still bound relatively non-specifically to the blot. This is apparent from the high background. After washing at higher stringency, the background was reduced. Seven radiolabelled bands were visible on the blot after the low stringency wash. Their sizes were as follows : 7.1, 5.0, 3.6, 3.2, 3.1, 2.7 & 2.5Kb.

The intensity of probe bound to the various bands was significantly different at the two stringencies. The 5kb and 3.2kb bands showed little reduction in intensity at 0.2xSSC whilst the other bands showed a great reduction in intensity. This indicates that the oleosin probe used showed greater homology to the genes on the 5kb and 3.2 kb bands than to the other genes.

These results allowed prediction about the expected sizes of genomic fragments to be isolated during cloning (see section 3.2).



Selected restriction enzyme digests of DNA from *B. napus (var. Jet neuf)*

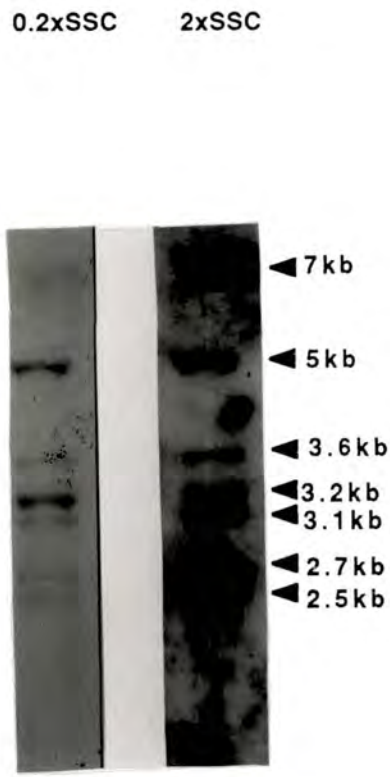


FIGURE 3.9

Genomic Southern

Genomic DNA was digested with *BamHI, EcoRI, HindIII and SacI.* The digests were carried out over night at 37°C using 15µg of DNA with 100 units of enzyme in a volume of 70µl. The fragments were separated on a 0.7% gel run in TBE over night at 30V. The gel was prepared and blotted as described in section 2.3. Some enzymes are less efficient at digesting genomic DNA due to contaminants in the relatively crude DNA preparation which inhibit enzyme action (see section 2.1). This explains the poor digestion of DNA in some of the tracks. The sizes of the radiolabelled bands are shown in the figure. These were estimated from the positions of the size markers on the gel photograph.

The blot was probed with the radiolabelled oleosin Nap 4 probe at 65°C over night (see sections 2.14 & 2.15) and washed at a low stringency - 2xSSC at 65°C, followed by a higher stringency - 0.2xSSC.at 65°C Washing was done at 65°C. Three 20 minute washes were carried out at 2xSSC after which the blot was autoradiographed at -70°C over night. After repeating the washing proceedure at 0.2xSSC the blot was again autoradiographed over night.

3.2 ISOLATION OF OLEOSIN GENOMIC CLONES

3.2.1 INTRODUCTION

This section describes the isolation and partial characterisation of two of oleosin genomic clones. The genomic library which was screened was previously constructed in this laboratory from a *BamHI* partial digest of the *Brassica napus* (var. jet neuf) genomic DNA in λ EMBL 3 (Ryan, et al., 1989). The library was screened using the Nap 4 cDNA.

To investigate the mechanism for control of transcription of the oleosin gene in *B. napus*, sequence 5' of the transcription start site was needed. cDNA sequences being derived from mRNA's will have at most, sequence beginning at the transcription start site. Isolation of oleosin genomic clones was therefore necessary for upstream sequence information and also the identification of intron regions.

Six genomic clones were characterised from the genomic library. A series of single and double digests were carried out to map the clones and the position of the oleosin coding regions located.

3.2.2 ISOLATION OF TWO OLEOSIN GENOMIC CLONES

The λEMBL 3 library was titrated and an aliquot containing 250,000 pfu was plated out as described in section 2.14. The phage were incubated for 6-8 hours until lysis was nearly confluent. Duplicate plaque lifts were taken for each plate (section 2.16) and these were probed with the oleosin probe (sections 2.17 and 2.18). After autoradiography, the positive signals on the two lifts were aligned to locate the potential oleosin genomic clones. Forty plaques corresponding were identified as

putative oleosin clones by positive signals on the autoradiograph. These plaques were cut out from the plate and the phage eluted in SM buffer. After this primary screen ten putative clones were chosen and a series of screens was carried out to purify the phage. Six of these were found to be true positives and large scale phage preparations were carried out to isolate the DNA (see section 2.15).

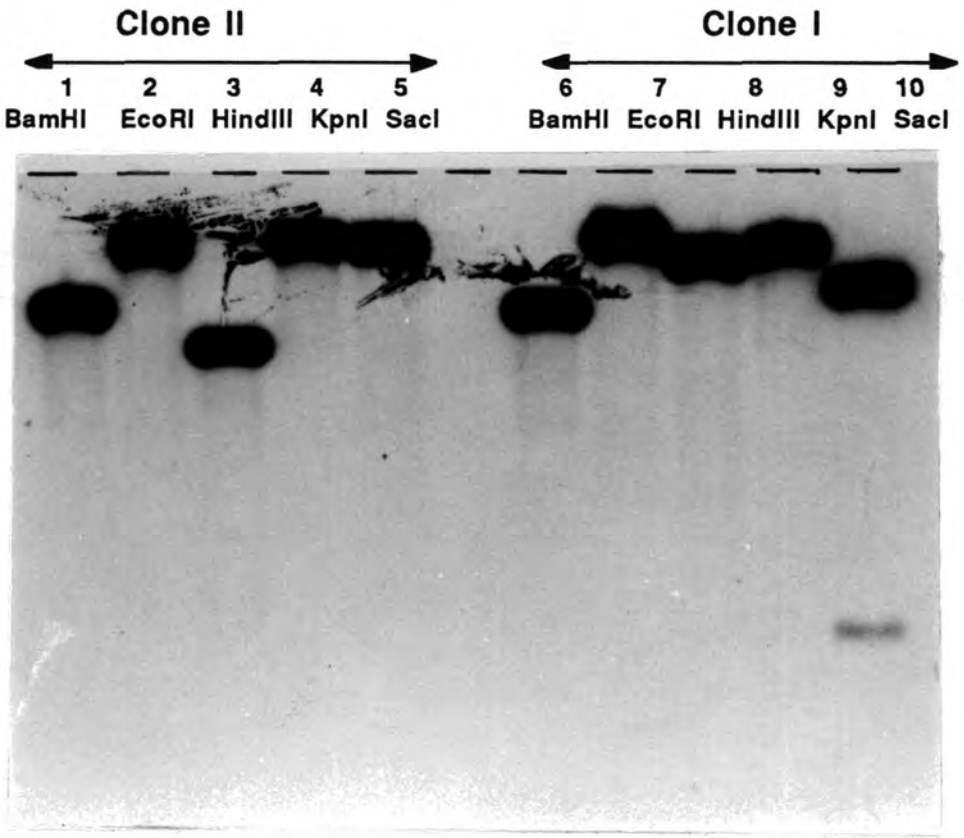
3.2.3 GENOMIC CLONE CHARACTERISATION

Two sets of three identical clones were identified by restriction mapping of the six genomic clones isolated. These were grouped as follows : λ A, F & G and λ E, B & J. Representative clones were chosen for further characterisation. λA was chosen from the first group and renamed Clone I and λJ was chosen from the second group and renamed Clone II. The 4.5Kb *HindIII/SacI* fragment from clone II, containing the oleosin coding sequence was subcloned in pUC18 by James Keddie at the John Innes Institute in Norwich. This genomic clone was then renamed Nap 3

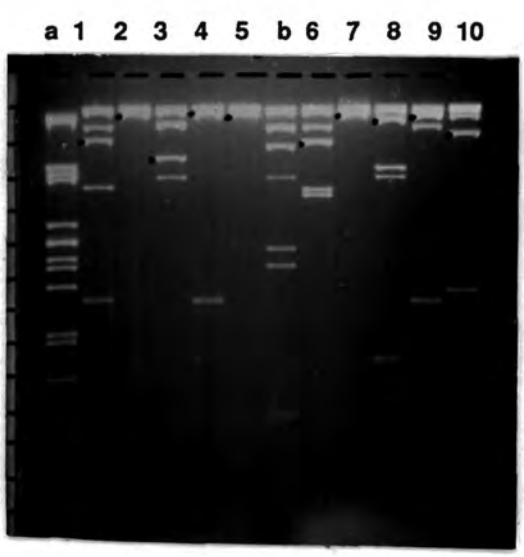
Sequencing of the clone has shown that an intron is present at the interface between the hydrophobic central domain and the C-terminal hydrophilic domain. The position of the intron is indicated by a vertical arrow (see Figure 3.6). The intron is approximately 400bp long and is rich in AT sequences. A putative ABA responsive element with the sequence ACGTGGC has been identified in the 5' upstream non-coding region at a position -154 nucleotides upstream of the translation start site.

A dense band can be seen in the *B. napus* genomic Southern blot after *BamHI* digestion (see section 3.1.4). The band is approximately 7Kb. This is only a rough estimate of the size of this band due to the poor resolution obtained for large fragments in the agarose gel system. This 7Kb band may contain several oleosin coding sequences on different *BamHI* fragments of similar size. Correlation can be made between the sizes of and the sizes of the BamHI fragments carrying the oleosin coding regions from the oleosin genomic clones Clone I and II which are 7.3Kb and 7.05Kb respectively.

(a)



(b)



Approximate fragment sizes

Lane	Size (kb)
1	7.1
2	-
3	5.9
4	-
5	-
6	7.3
7	-
8	-
9	-
10	8.2, 1.8

FIGURE 3.10

Restriction patterns for Clones I and II.

The bands containing the oleosin coding sequence were detected by Southern blotting followed by probing with the oleosin probe (sections 2.3, 2.14 & 2.15) as shown in Figure 3.10a. The labelled bands in Figure 3.10a are indicated in Figure 3.10b by dots at the left of the band. The size markers were the restriction products of phage λ digested with *PstI* (track a) and *HindIII* (track b).

Restriction digests were carried out using between 200ng and 1μg of λDNA per digest. The DNA was digested with 10Units of restriction enzyme for 2 hours at 37°C. DNA fragments were separated on a 0.6% agarose gel at 25V over night and the gel was blotted onto Hybond-N over night (see section 2.3). The DNA was bound irreversibly to the solid support by exposure to UV irradiation for 2 minutes. The blot was probed with radiolabelled oleosin Nap 4 cDNA (see section 2.14) at 65°C over night. Washing was carried out at 65°C as follows: 2xSSC for 20 minutes, followed by two washes with 1xSSC/0.1% SDS for 20 minutes.

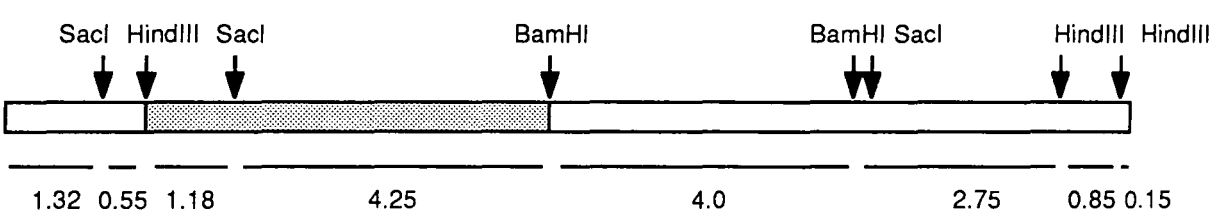
The sizes of the fragments obtained from single restriction digests could not be accurately determined due to the large size of the fragments. This gel merely serves to demonstrate the different restriction patterns observed for the two classes of genomic clone. The approximate sizes of some of the smaller labelled bands are shown in Figure 3.10b.

FIGURE 3.11

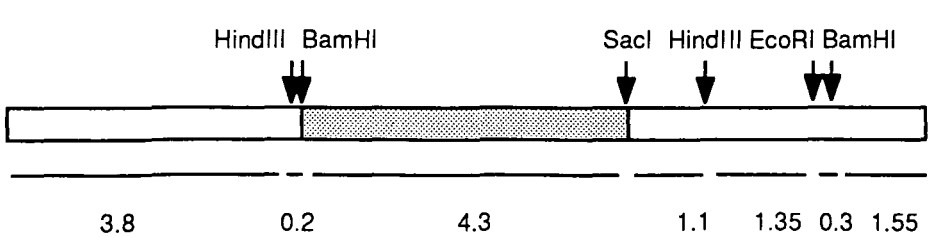
Restriction maps for oleosin genomic clones I & II

Mapping of the clones was carried out by digesting the DNA with a series of single and double restriction digests. Figure 3.11 shows preliminary restriction maps for the clones isolated. The map for Clone II has been confirmed by subcloning, however the map for clone I has yet to be confirmed. The shaded regions represent the restriction fragments which were radiolabelled with the oleosin probe. The 4.5Kb *HindIII/SacI* fragment of Clone II was isolated and subcloned into pUC18 and sequenced by James Keddie at the John Innes Institute in Norwich.

CLONE I



CLONE II



3.3 OLEOSIN EXPRESSION STUDIES

3.3.1 INTRODUCTION

In this section, the work done on oleosin expression during embryogenesis is described. Embryos were collected at five developmental stages throughout embryogenesis. Stage 1 embryos represent the earliest developmental stage and stage 5 embryos represent embryos excised from the dry seed (see Figure 3.12). All the experiments were carried out on tissue isolated from these five stages.

The purpose of these expression studies was to gather preliminary data on the relative importance of transcription and post-transcriptional events in the control of oleosin gene expression. The production of oleosin protein was followed using SDS PAGE and Western blotting techniques. A measure of mRNA accumulation was obtained by following mRNA steady state levels at each developmental stage. Since mRNA steady state levels are dependant on the relative rates of transcription and mRNA degradation, it is impossible to estimate transcription rates from steady state mRNA levels. Other techniques such as transcription run-on experiments are necessary to directly follow transcription rates.

Napin mRNA levels were also measured at each developmental stage. A direct comparison of napin and oleosin steady state mRNA levels of was then possible.

Oleosin is believed to be expressed in a tissue specific manner in *B. napus*, expression being predominantly within developing seeds. To test this mRNA was extracted from several different plant tissues and probed with the oleosin cDNA probe Nap 4. Although this test is limited by the sensitivity of the technique, the results still give an indication of the spatial pattern of expression within the plant.

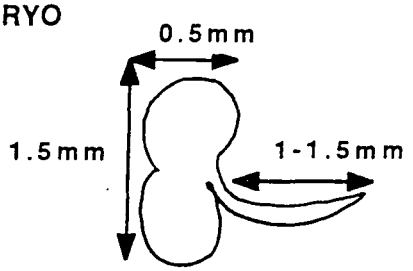
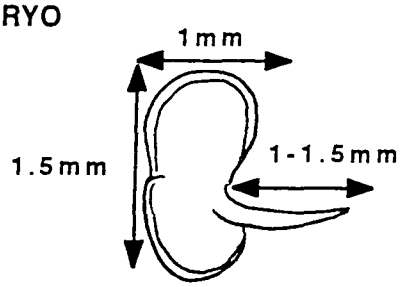
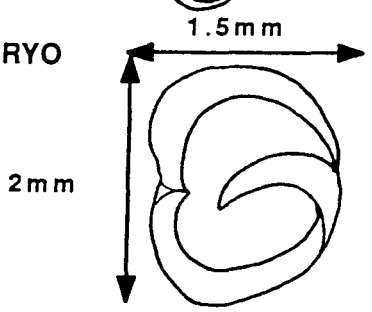
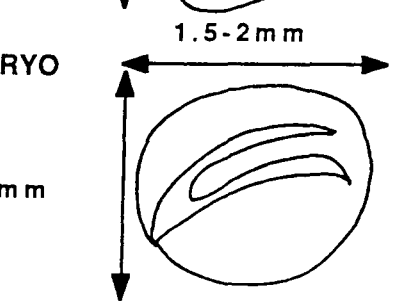
	VIEW OF EMBRYO FROM ABOVE	DESCRIPTION OF SEED IN SEED COAT
STAGE 1 EMBRYO		Pale green/yellow Translucent Soft
STAGE 2 EMBRYO		Green Soft
STAGE 3 EMBRYO		Dark green Firm
STAGE 4 EMBRYO		Seed becoming dessicated Green/Yellow Opaque Hard
STAGE 5 EMBRYO	DRY SEED	Completely dessicated Very hard

Figure 3.12

A diagramatic representation of embryo and seed development at the five stages of embryogenesis used in expression studies.

The reason for classifying embryos in this way was to overcome problems in determining ages of individual embryos. Determining the age of the seed pod is not an accurate way of classifying the developmental stage of the embryos as great variation in the developmental stages of different embryos is observed even within a single pod. The reason for this is that in field grown crops such as the one used in these studies, weather conditions affect the rate of embryo development. Dry conditions cause premature seed dessication, resulting in smaller seeds. The stages of development described here will then take place over a shorter time-scale than in a crop grown under wet weather conditions. This premature dessication may be an indirect effect of increased abscisic acid production by water stressed plants (Skriver and Mundy, 1990)

Below is an estimate of the age of the embryos at each stage:

Embryonic stage	Days post anthesis
1	28
2	42
3	56
4	70
5	Dry seed

These estimates are not considered to be accurate. To make the results for the oleosin mRNA steady state levels comparable to work already carried out on other aspects of storage product synthesis two controls were done. The first was a study on the pattern of storage protein synthesis throughout development by SDS PAGE (oleosin protein levels were specifically followed by Western blotting). The second was a study of napin mRNA steady state levels and these results were compared with work done by DeLisle and Crouch (1989).

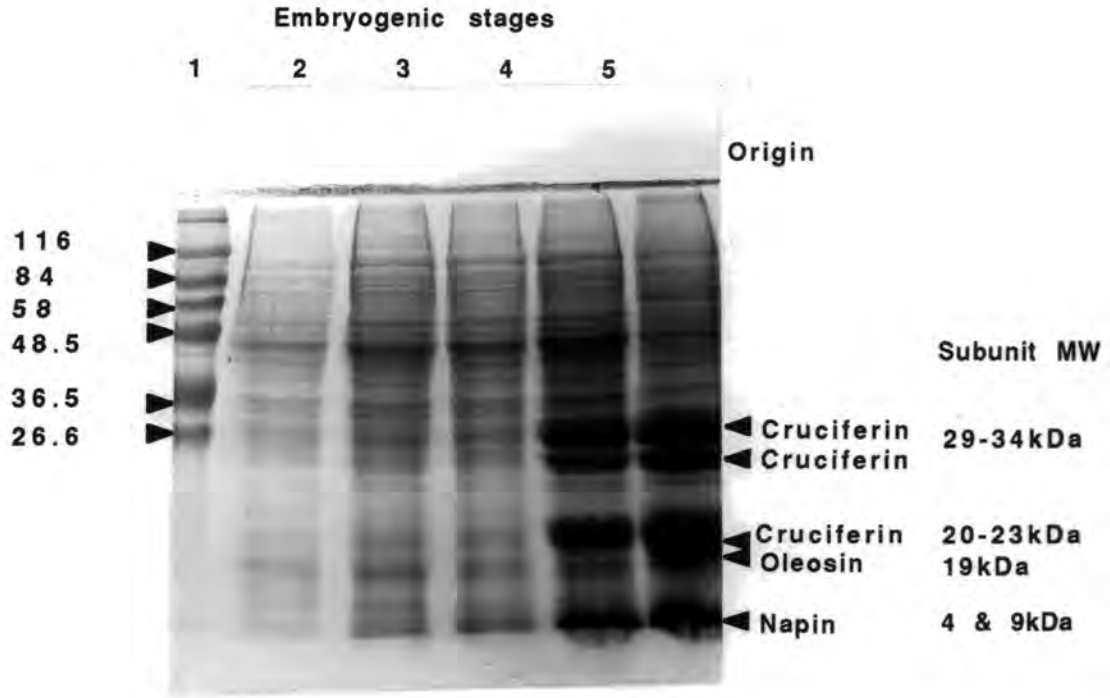


FIGURE 3.13

Protein production during embryogenesis in *B. napus*

In Figure 3.13, the pattern of protein production at five stages during the course of embryogenesis is shown. Protein was extracted from developing embryos at five different developmental stages (see Figure 3.12) as described in section 2.17. The constituent polypeptides were separated by SDS PAGE (see section 2.18). Loading was done on a per embryo basis, with protein equivalent to 5 embryos being loaded for each developmental stage (see note). Staining was carried out with Coomassie Brilliant Blue (see section 2.18).

Note : Since an equal amount of RNA was loaded for each developmental stage in the Northern blotting experiments, a more comparable result would have been obtained with the protein data if equal protein loading had been carried out.

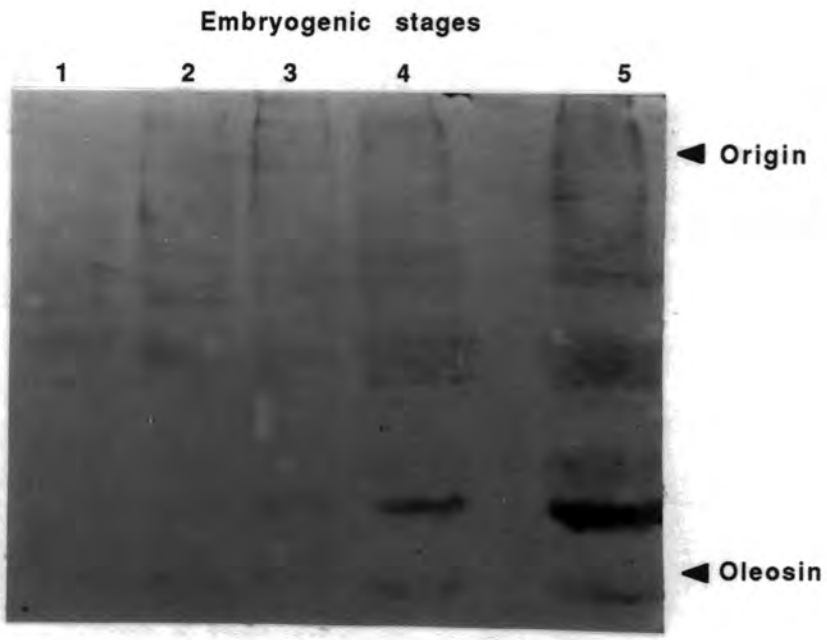


FIGURE 3.14

Western blot of the pattern of oleosin production at the various stages of embryogenesis.

The blot (a duplicate of Figure 3.13) was probed with a primary antibody raised against oleosin in rabbit. The position of bound antibody was detected by the assay system described in section 2.19. The colour was developed for 20 minutes to detect oleosin in some of the early stages of embryogenesis.

.

3.3.2 PROTEIN PRODUCTION DURING EMBRYOGENESIS

The SDS PAGE gel in Figure 3.13 shows the pattern of protein production during the five stages of embryogenesis described in Figure 3.12. From Figure 3.13 the constituent subunits of the two seed storage proteins from *B. napus* - cruciferin and napin, can be identified. Cruciferin is an oligomer consisting of six subunit pairs; each pair contains a heavy α and a light β chain. The α subunits have molecular weights between 29 and 34kDa, while the β subunits have molecular weights between 20 and 23 kDa (Rödin, et al., 1990). Napin is composed of two polypeptides of molecular weight 4 and 9kDa, which are linked together by disulphide bonds (Ericson, et al., 1986). This gel system was incapable of resolving these small polypeptides.

Oleosin is a protein of approximately 19kDa. It is therefore very close to the 20-23kDa cruciferin bands in this gel system. Cruciferin represents 40-50% of total seed protein while oleosin only accounts for about 20% (Murphy, et al., 1989b). Therefore the oleosin band cannot be distinguished from the cruciferin band. To clearly see the pattern of oleosin production, a Western blot probed specifically for oleosin was necessary.

A Western blot of oleosin production is shown in Figure 3.14. Oleosin is detectable at stage 3 of embryogenesis (see Figure 3.12), but is present at very low abundance until stage 4. After this the levels continue to rise until the seed undergoes dessication.

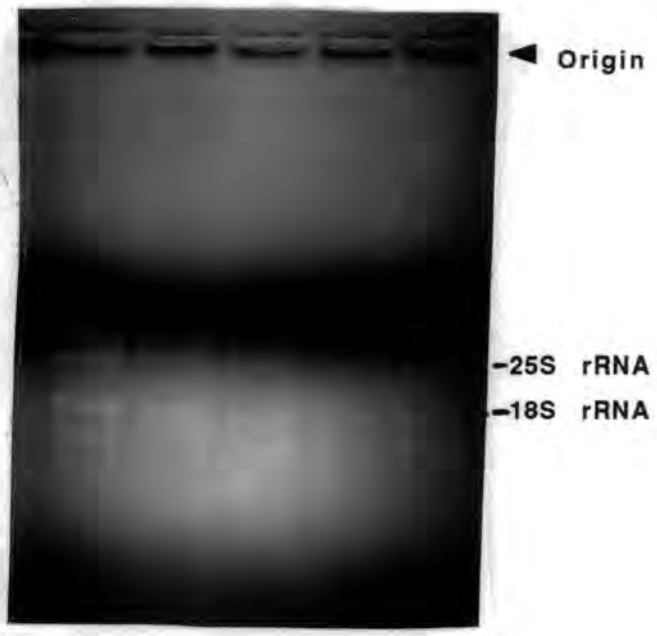


FIGURE 3.15

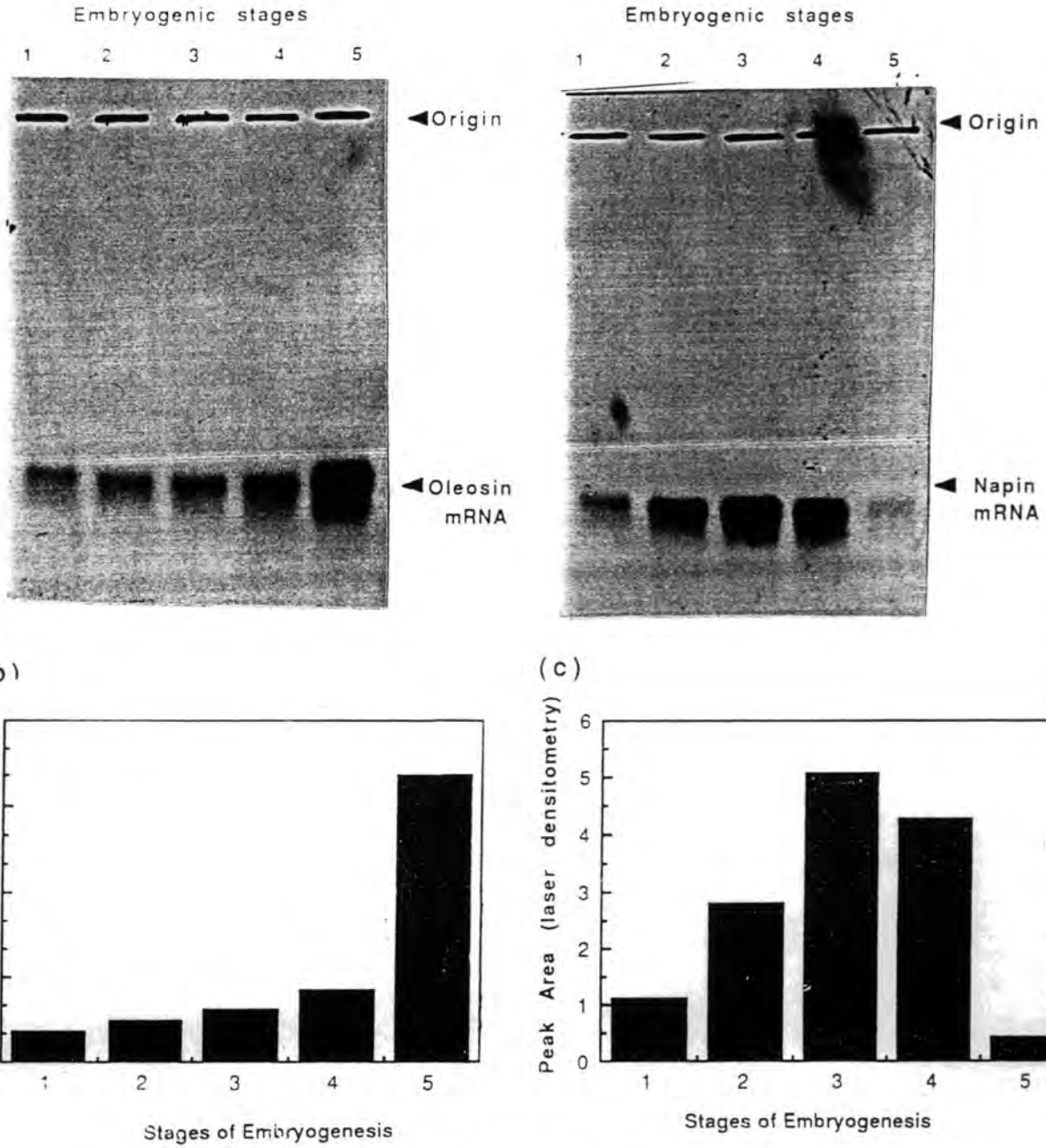
a) Stained gel of total RNA showing approximately equal laoding

b) Steady state levels of oleosin mRNA during embryogenesis

c) Steady state levels of napin mRNA during embryogenesis

Figure 3.15 shows the accumulation of oleosin and napin mRNAs in developing embryos. Total RNA was extracted from *Brassica napus* embryos (var. *Cobra*) at each stage of development. Embryos were prepared by removal of the testa, after which they were immediately frozen in liquid nitrogen. For each developmental stage, 0.5g of embryo tissue was extracted in 4ml of extraction buffer. The extraction proceedure is described in section 2.7. A 15μg sample of RNA for each developmental stage was run out on a gel and blotted onto nylon (section 2.8).

In panel (a) is a photograph of the gel of the seed RNA extracts showing the rRNA bands to indicate that approximately equal loading of RNA was achieved for each embryogenic stage. The gel was stained for 3 minutes in an ethidium bromide solution (5μg/ml) in sterile distilled water and destained by washing in sterile distilled water for 2-4 hours at 4°C with several changes of water. The 18S and 25S bands are indicated on the figure.



The blot was probed as described in section 2.18. The same blot was probed using the oleosin and napin probes (section 2.14). The initial probing was carried out with the oleosin cDNA (b). The blot was hybridised with probe over night at 65°C, then washed twice in 2xSSC/0.1%SDS at 65°C for 20 minutes. The blot was autoradiographed over night. Bound probe was removed from the blot (see section 2.8) and the same blot probed with the napin cDNA probe (c). The procedure previously described for the oleosin probe was then followed using the napin probe.Beneath panels b and c, are shown graphs derived from densitometer tracings of the intensities of the labelled bands at each stage of embryogenesis.

3.3.3 OLEOSIN & NAPIN STEADY STATE mRNA LEVELS DURING EMBRYOGENESIS

The steady state levels of napin and oleosin mRNAs at various stages of embryogenesis are shown in Figure 3.15. Beneath the Northern blots are plots of the intensity of the radiolabelled bands as estimated by laser densitometry.

A significant feature of the Northern blot (Figure 3.15) showing napin and oleosin mRNA steady state levels which must be explained is the close running of the two mRNAs on the gel. Napin is synthesised as a 178 amino acid polypeptide precursor and the major polyadenylated transcript is 850 nucleotides long (Josefsson, et al., 1987). The full length oleosin sequence of Nap 3 is 195 amino acids long. There is no information on the oleosin transcript size. However as the coding region of oleosin is significantly greater than that of napin, it would be expected that the oleosin mRNA would be larger. However the resolution of this gel system is insufficient to allow the napin and oleosin mRNAs to be resolved. The poor resolution provided by this gel system is seen from the relatively small separation between the 25S (3.8Kb) and 18S (2.1Kb) rRNAs.

Since the same blot was probed with the two different probes the steady state levels of oleosin and napin mRNA's are directly comparable. Oleosin mRNA accumulates later in embryogenesis than napin. Oleosin is detectable at stage 3 of embryogenesis, but is present in very low abundance until stage 4, which represents embryos 70 days post anthesis. Napin steady state level reaches a peak at stage 3 of embryogenesis, after which it decreases until levels are barely detectable in the dry seed. The oleosin steady state mRNA level continues to increase throughout embryogenesis and oleosin mRNA is present at high levels in the dry seed.

3.3.4 TISSUE SPECIFIC EXPRESSION OF OLEOSIN

To assess whether oleosin expression was tissue specific, a Northern blot of total RNA from various tissues was probed with the Nap 4 oleosin cDNA probe. The result is shown in Figure 3.16. Oleosin mRNA was only detectable in the embryo tissue. Some expression may occur in other tissues, but mRNA levels are too low to detect using this assay system. To improve the sensitivity of the system, a larger quantity of RNA would have to be loaded onto the gel. More than 15μg per well would overload this gel system. Purification of poly(A)$^+$ mRNA is a possible solution.

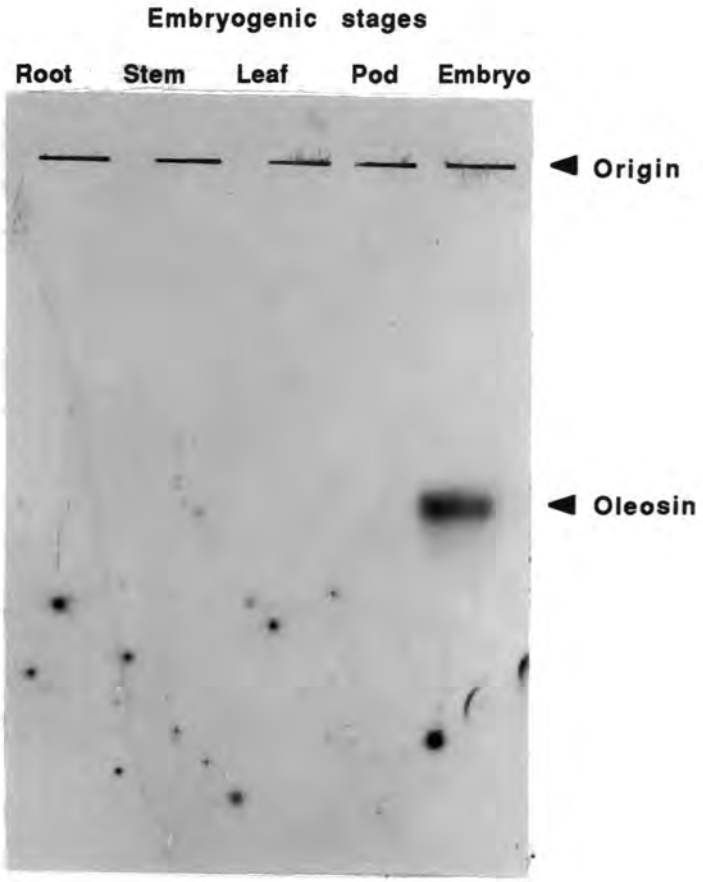


FIGURE 3.16

Distribution of oleosin mRNA in different plant tissues

Total RNA extracted from root (15 µg), stem (15 µg), pod (30 µg) and leaf (30 µg) tissue of young seedlings was electrophoresed, blotted and probed with the Nap 4 oleosin cDNA. Stage 3 embryo RNA (15 µg) was used as a positive control. The blot was probed with the oleosin cDNA probe at 65°C over night. The blot was washed twice in 2xSSC/0.1% SDS at 65°C for 20 minutes

CHAPTER4 DISCUSSION

4.1 FEATURES OF THE OLEOSIN NAP 4 AND NAP 5 cDNA CLONES

4.1.1 SEQUENCE ANALYSIS

Several important sequence elements have been identified in the Nap 4 and Nap 5 cDNA clones. These include internal repeats, glycosylation sites and polyadenylation signals.

INTERNAL REPEATS - Several internal repeat sequences have been identified in the N-terminal region of Nap 3, Nap 4 and Nap 5. No such repeats are present in the N-terminal regions of oleosin proteins from other plant species such as maize and carrot (see Figure 3.8). It is therefore unlikely that these sequences have a role in the basic function of oleosins, ie. in their role as stabilising elements in the oil bodies. However since they are conserved in the *B. napus* oleosin gene family they may have a specific function in this plant.

The C-terminal region of the proteins contains sequence inserts which are present only in certain members of the gene family. The presence of extra sequences in this region has been observed in the maize oleosins KD16 and 18, and is responsible for the size differences in the two proteins. KD18 has an extra 18 amino acid sequence in this domain. Figure 3.6 shows the presence of inserts in the C-termini of Nap 3 and Nap 5 . The insert in Nap 3 is an 8 amino acid direct repeat which may have arisen by an unequal cross over event in this region. Nap 5 also has an 8 amino acid insert which is absent from Nap 3 and Nap 4, but this does not appear anywhere else in the Nap 5 sequence.

GLYCOSYLATION SITES - Potential N-linked glycosylation sites have been identified in the sequences of Nap 4 and Nap 5. Core-glycosylation of polypeptides takes place co-translationally within the ER. Evidence from electron microscopy of immunogold

labelled developing embryos shows that no oleósin protein is associated with the rough ER. If the theories concerning oleosin synthesis are correct, then oleosin is formed on free ribosomes and does not enter the ER membrane system. The glycosylation signals within the oleosin coding sequence would then be redundant.

POLYADENYLATION SIGNALS - The positions of the potential polyadenylation signals from Nap 4 and Nap 5 are shown in Table 3.1. The perfect consensus AATAAA is the most effective polyadenylation signal and therefore polyadenylation of the highest proportion of mRNA's will be directed by this signal. A compilation study carried out by Joshi (1987) indicated that for most plant genes, polyadenylation occurs within 8-43bp downstream of the polyadenylation signal. Only one of the putative poly adenylation signals - the GATAAA sequence at -15bp, from Nap 4 is within this range. From the results of the positions of the polyadenylation signals in Nap 4 and Nap 5, it appears that polyadenylation is very flexible with regard to the spacing between the polyadenylation signal and the site of poly(A) addition.

SEQUENCE COMPARISONS

Three oleosin proteins have been identified by SDS PAGE in *B. napus,* with molecular weights of 20, 22 and 24kDa (Tzen, et al., 1990). The 20 kDa protein shows a very high abundance, which may mean that there are two oleosin proteins in *B. napus* with molecular weights close to 20kDa. Studies done by Tzen et al (1990) on oleosins from a wide range of plant species have shown that two antigenically distinct isoforms of oleosin exist, which have been classed as the high and low molecular weight forms (see section 1.2.3). In *B. napus* the 20kDa oleosin belongs to the low molecular weight class and the 22 and 24kDa protein belong to the high molecular weight class.

Direct sequence comparison of Nap 4 and 5 with carrot DC59 and the maize KD16 and KD18 in Figure 3.8 show the high amino acid sequence similarity which exists between the oleosins from these diverse species. A significant feature is the sequence conservation in the central hydrophobic domain in all five proteins.

Conservation of the three prolines which are highlighted indicates that these prolines have an important role in the function of oleosin.

KD16 and 18 from maize belong to the low and high molecular weight classes respectively. Direct sequence comparison of these proteins in Figure 3.8 shows significant sequence divergence exists between KD16 & 18 which may account for their lack of antigenic cross reactivity.

B. napus (2n=38, AACC) is an amphidiploid made up of the genomes of *B. campestris* (2n=20, AA) and *B. oleracea* (2n=18, CC). Also the protein sequences for Nap 1 and 2 were obtained from oleosins of a different rape variety to that used to isolate the cDNA clones, which in turn differed from that from which the genomic clone Nap 3 was derived. Therefore these five sequences do not necessarily represent five distinct genes but may be alleles of the same gene

The Southern blot shown in Figure 3.9 has seven radiolabelled bands when incubated with the oleosin probe. Washing the blot at two stringencies showed that the bands hybridised with different affinities to the probe. The 2.5, 2.7, 3.1, 3.6 & 7kb bands were significantly reduced in intensity compared with the 3.2 and 5.0kb bands after the high stringency wash. The 3.2 and 5.0kb bands have a higher affinity for the probe and must show greater sequence similarity to Nap 4 than the other four bands. Differential binding of the probe to the bands shows there are at least two sequence types represented in this blot. The exact copy number of the oleosin gene family in *B. napus* (var *Jet neuf*) cannot be determined unequivocally from these results. Gene copy reconstruction experiments are necessary for the gene copy number to be determined.

The sizes of these bands were compared to the sizes of the *BamHI* fragments from the genomic clones carrying the oleosin coding region (see section 3.2). The *BamHI* fragments from the genomic clones were 7.3 and 7.05Kb respectively for Clones I & II. It is therefore likely that the 7Kb band on the genomic Southern contains the BamHI fragments present in the genomic clones (see section 3.13).

4.1.2 STRUCTURE ANALYSIS

Analysis of some of the basic structural elements of the predicted amino acid sequences for Nap 4 and Nap 5 are described here. The hydropathy plots shown in Figure 3.7a for Nap 4 and 5 have a similar pattern to those of KD16 (Vance and Huang, 1987) and KD18 (Qu and Huang, 1990) from maize and DC59 from carrot (Hatzopoulos, et al., 1990). The N- and C- termini are hydrophilic, whilst the central region is hydrophobic in all these oleosins. The distribution of charged amino acids represented in the acid/basic plots shown in Figure 3.7b corresponds well to the distribution of amino acids observed in the hydropathy plots. The hydrophilic acidic and basic amino acids are clustered at the N- and C- termini. None are found in the central hydrophobic region.

Detailed structure predictions were carried out for the maize KD16 (Vance and Huang, 1987). The central hydrophobic region, consisting of 72 amino acids is expected to form a hairpin consisting of two anti parallel helices (either α or 3_{10}). The algorithms used to predict these secondary structure motifs are based on those developed for soluble globular proteins. Because soluble and hydrophobic proteins have differerent properties, the accuracy of these structure predictions for the oleosins is uncertain. The predicted helices are separated by a region of random folding containing the three conserved proline residues. These enable a bend to form between the helices.

The central hydrophobic regions in Nap 4 and 5 are 75 and 67 amino acids long respectively. The central hydrophobic domain is 97.1% conserved between Nap 4 and Nap 5 (see Table 3.2). This is the most highly conserved of the three domains. No structure predictions have been carried out on this region in Nap 4 and Nap 5. A high degree of sequence conservation exists in this region between Nap 4 & 5 and the maize KD16 sequence. Therefore the secondary structure of this domain in the *B. napus* protein is likely to be very similar to that predicted for KD16.

The hydrophobic domain of KD16 has been compared to the lipophilic transmembrane segments of integral membrane proteins (Vance and Huang, 1987).

The transmembrane seqments of these proteins consist of 20-33 uncharged, non-polar amino acids flanked by charged amino acids. The hydrophobic domain of oleosins is longer than in these proteins as it does a "U" turn in the hydrophobic core of the oil body. The presence of the three highly conserved prolines in the hydrophobic domain provides the flexibility for this bend to occur in the polypeptide backbone.

The hydrophilic N-terminal regions of Nap 4 and Nap 5 showed the greatest divergence of the three domains having only 75.6% identity (see Table 3.2). Comparison of the N-terminal sequences of oleosins from different plant species shows that this region has diverged significantly even between oleosin isoforms from the same species (see Figure 3.8). The function of the N-terminal domains of oleosins is unknown, although it may be involved in the regulation of triacylglycerol mobilization from within the oil body (see section 1.2.3). The hydrophilic N- terminal region of KD16 is believed to contain some helical structure.

The construction of a helical wheel plot of the C- terminal region of the maize KD 16 showed it to consist of an amphipathic α helix with a hydrophobic and a hydrophilic face. The helix is thought to rest on the surface of the oil body with the hydrophobic face embedded in the triacylglycerol core. Similar results were obtained with KD18 from maize, showing a C- terminal amphipathic α helix (Qu and Huang, 1990). To determine whether the C-terminal domains of Nap 4 and Nap 5 are capable of forming amphipathic α helices, helical wheel plots would have to be carried out.

Animal apolipoproteins have been shown to contain amphipathic α helices and the C- terminal domain of KD16 has 33-46% sequence similarity with conserved regions of various apolipoprotein genes (Vance and Huang, 1987). Animal apolipoproteins have a similar function to plant oleosins - they coat lipid containing organelles and provide structural support.

4.1.3 USE OF PROBES

The cDNA probe derived from Nap 4 which was used to probe the genomic Southern was 455bp long. This is a long sequence to use as a probe. Short probes of 20-200bp have advantages over long probes (Sporel and Kafatos, 1987). Long probes may contain several melting domains. Even if only one of these domains has high sequence similarity to the gene being probed, it will cause the probe to bind strongly to the DNA on the blot. Long probes are therefore not as good as short probes for differentiating between members of a gene family. A series of oligonucleotide probes complementary to unique sequences present in different members of the gene family could be synthesised. This would allow each individual gene to be identified on a Southern blot.

Oligonucleotide probes have advantages over cDNA probes for the estimation of gene copy number. Due to the length of cDNA's, they may span a region within the gene containing one of the restriction sites used in digestion of the genomic DNA. This is particularly likely to occur if a large intron exists within the gene. A single gene would then be represented by two labelled bands on a Southern. An oligonucleotide probe is much shorter and unlikely to span such a restiction site. Oligonucleotide probes therefore give a more accurate representation of gene copy number in Southern blot analysis. However preliminary experiments using radiolabelled oligonucleotides to probe genomic Southern blots did not label bands sufficiently to be detected by autoradiography (data not shown). This is due to low efficiency binding of these relatively short oligonucleotide sequences to the filter bound DNA.

A better alternative would be to use the 3' untranslated regions of cDNA sequences as sequence specific probes. Using the Gap algorithm to calculate the nucleotide sequence conservation between the Nap 4 and Nap 5 cDNA clones, the coding region was shown to have 83.5% identity between the two clones, whilst the 3' non-coding region had only 56.4% identity (see Figure 3.5). Since the 3' non-coding region

is far less conserved between different members of the oleosin gene family, it would make a better sequence specific probe than the Nap 4 coding region probe.

Use of the 3' untranslated region (UTR) of a gene as a sequence specific probe has been demonstrated with the maize oleosin KD18 (see Figure 1.3). Probing a Northern blot of total seed RNA with a probe made from the 3' UTR of KD 18 produced less diffuse bands of lower intensity than those produced after probing with the KD18 cDNA coding region. This indicates that the 3' UTR probe specifically hybridised to the KD18 mRNA. The KD18 cDNA probably hybridised weakly with other related mRNA species of different molecular weight - hence the higher intensity diffuse bands.

These results indicate that the 3' UTR of a cDNA molecule would be useful as a sequence specific probe. Similarly the 3' UTR of Nap 5 could be used as a DNA probe.

4.2 THE ISOLATION OF OLEOSIN GENOMIC CLONES

The pattern of transcription of a gene is dependent on the regulatory sequences that exist in the 5' untranslated region. Oleosin expression was shown to be seed specific. To characterise the sequences responsible for seed specific expression in oleosin from *B. napus* (var jet neuf), genomic clones were isolated with the aim of sequencing the 5' upstream region. Two clones were characterised and crude restriction maps were worked out.

The oleosin coding region, (identified by probing a genomic Southern with the Nap 4 cDNA) was found to reside on a 7.3Kb *BamHI* fragment in Clone I and a 7.05Kb fragment in Clone II. These bands are identified in Figure 3.10. The sizes of the *BamHI* fragments in these clones was compared with the sizes of the radiolabelled bands in the Southern blot of *B. Napus* (var jet neuf) shown in Figure 3.9. A dense band (which at first was thought to be some non specifically bound probe on the blot) is approimately 7Kb.

The plant hormone abscisic acid (ABA) has been shown to influence the expression of embryo specific genes (see section 4.4.4). Abscisic acid has been implicated in the regulation of oleosin expression during embryogenesis (see section 4.4.5). A sequence called the abscisic acid responsive element is found in the upstream region of genes in which ABA has been shown to enhance transcription. These include the *Rab* 21 gene from rice (Mundy, et al., 1990) and the E_m gene from wheat (Marcotte, et al., 1989). The ABA responsive element described in section 1.2.4 is found in the 5' upstream region of the maize oleosin protein KD18 (Qu and Huang, 1990).

Preliminary sequence data for the 5' upstream region of the genomic clone Nap3 shows the presence of a putative ABA responsive element -154 nucleotides upstream of the translation start site. The putative ABA responsive element from Nap 3 has the following sequence ACGTGGC. This sequence information will provide a means for investigating the mechanism of action of ABA on oleosin expression in this system.

Another feature of the Nap 3 genomic clone is the presence of an intron within the gene. It has the splice sites found in other plant introns : 5' -GT- INTRON -AG- 3' and also as is common for plant introns, there is a bias for A and T nucleotides within the intron sequence (Goodall and Filipowiz, 1989). The intron is approximately 400bp long and it is present at the interface between the central hydrophobic domain and the C-terminal hydrophilic domain of the protein. The position of the intron is indicated by an arrow in the Nap 3 sequence in Figure 3.6 between amino acid 132 and 133. The KD18 oleosin genomic clone isolated from maize has high homology with the *B. napus* oleosins (see Figure 3.8). However the maize genomic clone has no intron sequences.

4.3 OLEOSIN EXPRESSION STUDIES

The relative order of events in storage product synthesis is shown in Figure 1.1 (Murphy and Cummins, 1989). Oil synthesis starts early in embryogenesis, followed by synthesis of the storage proteins napin and cruciferin. Oleosin synthesis occurs last of all. Investigation of the mechanism of oil body formation using electron microscopy showed that oil droplets accumulate in the cytosol prior to oleosin synthesis (see section 1.1.3).

Figure 4.1 shows the pattern of protein production during the course of embryogenesis. The results followed the trend described in section 1.1.3. From Figure 1.1 (Murphy and Cummins, 1989) it can be seen that napin and cruciferin are synthesised before oleosin. The results in Figure 4.2 show that oleosin can be detected in stage 3 embryos, increasing in abundance in stage 4 and stage 5 embryos. The times assigned to the different stages of embryogenesis are only a rough estimate and are not absolute (see Figure 3.12). The studies on protein production (Figure 3.13) and napin mRNA production (Figure 3.15) during embryogenesis were used to relate the results of the oleosin expression studies to experiments carried out by other workers on storage product synthesis in *B. napus* during embryogenesis.

The results in Figure 3.15 for napin mRNA accumulation show that napin message becomes most abundant in embryos at stage 3 of embryogenesis and then the levels fall until there is none remaining in the dry seed. This follows the trend for napin mRNA steady state levels shown in Figure 1.5 where napin message rises to a peak concentration, then falls to a very low level. The times post anthesis at which these events occur do not correspond with the time estimates made for this present study, hence the need for controls. Figure 3.15 shows that oleosin mRNA is present early in embryogenesis. Levels increase gradually until the time interval between stages 4 and 5 when a rapid increase in oleosin mRNA levels is observed. At this time the seed is undergoing dessication.

The increase in oleosin mRNA abundance at each embryonic stage is determined by the relative rates of transcription and mRNA degradation. As in *B.napus*, oleosin mRNA levels in maize increase throughout embryogenesis (Qu, et al., 1990). KD16 and 18 mRNA's are present in low abundance early in embryogenesis, increasing until dessication of the seed begins. However the relative abundance of the two messages is different. Early in embryogenesis, KD16 mRNA is present in greater abundance than KD18 mRNA. As described in section 1.2.2 & 1.2.3 members of gene families often have different functions. It is possible that KD16 & 18 have different functions in the developing embryo and are therefore required in different amounts early in embryogenesis.

Oleosin proteins are encoded by a gene family and it would be interesting to investigate whether the different members of the family are transcribed at the same time and whether they contribute equally to the oleosin mRNA pool. Three cDNA clones representing the Nap 5 sequence were isolated and only one representing the Nap 4 sequence (see section 3.2.1). This may indicate that the mRNA for Nap 5 is more prevalent at the developmental stage when mRNA was isolated for cDNA synthesis , or that it is more easily isolated during cloning. To determine whether the members of the *B. napus* gene family show different transcription rates, probes which are specific for unique sequences in the different members of the family are needed. These could then be used to probe for expression of individual members of the gene family. In maize, expriments were carried out to follow the pattern of expression of two oleosin isoforms - KD 16 and KD18 (see Figure 1.3). The Northern blot was probed using KD 16 and KD18 cDNA probes.

Figure 3.15 shows that after reaching a peak concentration at stage 3 of embryogenesis, the levels of napin mRNA decrease. Oleosin mRNA however continues to accumulate throughout embryogenesis. As shown in Figure 1.4, the napin transcription rate is reduced around 34 days after anthesis, followed by a decrease in the napin mRNA levels (DeLisle and Crouch, 1989). However the observed decrease in napin mRNA levels is not only due to a reduction in the transcription rate. In the same figure,

the cruciferin transcription rate is shown to decrease around 38 days after anthesis. However the cruciferin mRNA levels remain relatively constant. There appears to be a specific mechanism operating which directs the degradation of the napin mRNA in preference to the cruciferin mRNA.

A general account of determinants of mRNA stability was given in sections 1.4.2 & 1.4.3. The importance of the 3' UTR in determining mRNA stability has been demonstrated in many systems. The basic principle of how this operates is the steric hindrance of non-specific 3'-5' random exonuclease action by the formation of secondary structure motifs in the 3' UTR. Stability of the mRNA is reduced by endonuclease cleavage at specific sites in the 3' UTR which result in loss of the region causing steric hindrance. The mRNA is then rapidly degraded.

To determine whether post-transcriptional processes are in operation in the control of oleosin mRNA levels, a study of the transcription rate of oleosin at various stages in embryogenesis is necessary. This could be done using transcription run-on experiments with nuclei isolated from embryos at different stages of embryogenesis. If the transcription rate is reduced during late embryogenesis as in the case of napin and cruciferin, the high levels of oleosin mRNA late in embryogenesis can be attributed to an inherent stability of the mRNA in this system.

To investigate whether the expression of oleosin is tissue specific in *B. napus*, oleosin mRNA levels were assayed in various tissues. From Figure 3.16 it can be seen that oleosin is found only in developing embryos, there is no detectable oleosin mRNA produced in the other tissues tested. This was not an exhaustive experiment as not all tissues were tested and the assay system is limited by its sensitivity. However this result is sufficient to indicate that oleosin expression is tissue specific. Expreiments in maize on the tissue specificity of KD16 and KD18 gene expression show that these oleosins are expressed only in developing embryos [Qu (1990)].

The isolation of the oleosin genomic clones will enable transcriptional regulation of the genes to be analysed in detail.

4.4 FUTURE PROSPECTS

4.4.1 INTRODUCTION

The study described in this thesis on various aspects of oleosin structure, function and expression gives an indication of the direction of future work. In this concluding discussion, the potential for further work on the molecular biology of the oleosin gene family of *B. napus* is outlined. Three basic areas are discussed:

1] Further characterisation of the members of the oleosin gene family.

2] Detailed experiments on the expression of different members of the oleosin gene family and studies on the regulation of oleosin expression during embryogenesis.

3] Isolation and characterisation of regulatory elements from the 5' upstream regions of oleosin genomic clones.

This discussion section begins with an introduction to plant seed specific regulatory elements from 5' upstream regions of seed specific protein genes, followed by an overview of the regulation of gene expression by abscisic acid. This chapter is concluded with suggestions of possible future experiments.

4.4.2 TRANSCRIPTIONAL REGULATION IN PLANT CELLS

The dependence of tissue specific and developmental gene expression on a regulatory region within the DNA has been demonstrated for the soybean seed lectin gene (Okamuro, et al., 1986). A 17.1kb fragment of soybean DNA was transferred to tobacco. This fragment contained a lectin gene which is expressed only in seeds and four non-seed protein genes expressed constitutively in all tissues. In tobacco, the lectin gene was found to be expressed in seeds during seed development, but not in any other tissues. The other proteins were expressed constitutively in all tissues.

Upstream regulatory sequences have been isolated from many plant genes. A compilation study of 79 plant genes revealed that a putative TATA box is a common feature of the genes analysed (Joshi, 1987a). In 84% of the genes analysed, there is

very close sequence identity to the consensus TATATA. The other genes have AT rich sequences at the site of the putative TATA box. The CAAT sequence, like the TATA box is common in prokaryotic as well as eukaryotic genes. However the CAAT sequence is not as highly conserved in genes as the TATA box.

4.4.3 REGULATORY SEQUENCES IN SEED PROTEIN GENES

To confer tissue specific expression on seed protein genes, sequences other than the basic TATA and CAAT sequences are required. The sequence element

5'-A T (T/A) A A T-3'

is found in the 5' upstream region of several seed protein genes (Jofuku, et al., 1987). The ones so far identified are soybean lectin, Kunitz trypsin inhibitor, glycinin and β-conglycinin genes. It is possible that this is an embryogenesis specific control sequence which reponds to a common signal resulting in coordinate regulation of these genes. Direct evidence has been provided for the importance of the 5' upstream region of the β-conglycinin in the regulation of transcription. A 200 nucleotide sequence from the upstream region of the gene was shown to enhance gene transcription when linked to the chloramphenicol acyltransferase reporter gene (Chen, et al., 1986).

4.4.4 ABA IS INVOLVED IN COORDINATING GENE EXPRESSION DURING EMBRYOGENESIS

To regulate so many genes simultaneously, there must be a control mechanism to trigger and coordinate the timing of their expression during embryogenesis (Skriver and Mundy, 1990). Abscisic acid (ABA) is produced by plants in response to drought stress and also in tissues of developing seeds. ABA has been shown to enhance the expression of certain seed protein genes. It is therefore possible that increased levels of ABA after fertilization act as the trigger resulting in the pattern of gene expression unique to embryogenesis.

The mechanism of action of ABA is unknown. Some studies done on wheat have implicated Ca^{2+} as a second messenger (Napier, et al., 1989). There is also some

evidence that cAMP may function as a second messenger for ABA (Mundy, et al., 1990). The function of a second messenger is to relay to the cell the increase in ABA levels and enable it to mount an appropriate response by altering the pattern of gene expression. The effect of ABA may be at the transcriptional or post-transcriptional level.

Oil synthesis begins early in embryogenesis, followed by synthesis of the main storage proteins, napin and cruciferin. Oleosin production does not begin until oil synthesis has nearly been completed. The pattern of oleosin expression during embryogenesis is shown in Fig 1.1 (Murphy, et al., 1989b).

4.4.5 THE ABSCISIC ACID RESPONSIVE ELEMENT

An important element in regulation of embryogenesis is the hormone abscisic acid (ABA). Rice plants produce a basic glycine-rich protein from the *Rab*21 gene (previously called *Rab*16), when they are exposed to water stress (Mundy and Chua, 1988). *Rab*21 is transcribed at a higher rate in the presence of ABA. A putative ABA responsive sequence 5'-TACGTGGC-3' is present in the 5' region of the *Rab*21 gene (Mundy, et al., 1990). Deletion of the region containing this sequence abolishes the ABA response.

The consensus sequence 5'-ACGTGccgC-3' has also been identified in the 5' upstream region of the wheat Em gene, which encodes a late embryogenesis abundant (LEA) message (Marcotte, et al., 1989). A second sequence 5'-CGAGCA-3' in the 5' region of the Em gene has also been implicated in the ABA response. Identical sequences are found in the 5' regions of the α' subunit of the β-conglycinin gene (Chen, et al., 1986) and the *Rab*21 gene (Mundy, et al., 1990).

The 5' upstream region of the maize oleosin genomic clone genomic clone KD 18 from contains putative ABA responsive elements (Qu and Huang, 1990). These are 5'-TACGTGTC-3' at-62bp and 5'-ACGTACCC-3' at -83. Germination of maize embryos in a solution containing ABA results in a significant increase in oleosin mRNA stability (Qu, et al., 1990). The expression of oleosin during embryogenesis in carrot has also been shown to be under the control of ABA (Hatzopoulos, et al., 1990).

4.4.6 FUTURE EXPERIMENTS

1] Characterisation of the oleosin gene family

Gene copy number determination can be carried out by gene copy reconstruction experiments. Three oleosin bands have been identified by SDS PAGE. These have molecular weights of 20, 22 and 24 KDa (see section 1.2.3), but each may include several proteins. A systematic study is needed to identify gene sequences corresponding to these proteins. Sequence information will enable the regions contributing to subdivision of these proteins into two immunologically distinct classes to be identified (see section 1.2.3).

2] Experiments using highly specific probes.

As already described probes made from the 3' untranslated regions of cDNAs are highly specific for the gene sequence of that particular oleosin isoform. These specific probes can then be used to isolate specific genomic clones and in subsequent studies on specific oleosin isoforms.

Rescreening the λ EMBL 3 genomic library with sequence specific cDNA probes for Nap 4 and Nap 5 under conditions of high stringency would enable the genomic clones corresponding to these cDNA clones to be isolated.

Studies on expression of specific mRNA isoforms are possible using these probes. This type of experiment has already been done in maize (see section 1.2.4).

3] Determination of the relative importance of transcriptional and post-transcriptional processes in the control of steady state mRNA levels.

Transcription run-on experiments are necessary to determine oleosin transcription rates. The relative importance of post-transcriptional processes such as mRNA degradation can then be determined.

4] Characterisation of 5' upstream regulatory regions from oleosin genomic clones.

This would involve experiments such as gel retardation and DNaseI footprinting. Isolation of the 5' regulatory region, mutagenesis and ligation to a reporter gene, would enable detailed analysis of the control elements to be carried out (Rosenthal, 1987).

The forty putative clones isolated in the primary screen represent a sublibrary. Repeated screening of this sublibrary at increased stringency would enable genes with decreasing sequence similarity to the probe to be identified (Sporel and Kafatos, 1987). The use of probes from the 3'UTRs of the oleosin Nap 4 and Nap 5 cDNA clones will enable the corresponding genomic clones to be isolated.

Further characterisation of the oleosin genomic clones will provide useful information on the regulation of oleosin expression and processing of gene transcripts.

4.5 SUMMARY

Oleosin belongs to a gene family. Oleosins have recently been sudivided into two molecular weight classes which are antigenically distinct forms (Tzen, et al., 1990). Preliminary classification of the cDNA and genomic clones described in this work has been possible from sequence information and molecular weight prediction.

The oleosins from *B. napus* were shown to have similar hydropathic plots to the highly characterised oleosins from maize. Sequence comparisons with oleosins from maize and carrot showed a high degree of sequence conservation.

Expression of the oleosin genes is tissue specific - the protein being expressed predominantly in developing embryos. There is evidence for differential expression of different members of the gene family. The oleosin genes are under transcriptional control, being activated during embryogenesis. The hormone abscisic acid may be involved in regulation of expression. Steady state oleosin mRNA levels may be influenced by differential stabilisation of oleosin mRNA's.

REFERENCES

Barker, J., Steele, C. and Dure, L. (1988). Sequence and characterisation of six LEA proteins and their genes from cotton. Plant Mol Biol *11*, 277-291.

Bernstein, P. and Ross, J. (1989). Poly (A), poly (A) binding protein and the regulation of mRNA stability. TIBS *14*, 373-377.

Binder, R., Hwang, S.-P. L., Ratnasabapathy, R. and Williams, D. L. (1989). Degradation of Apolipoprotein II mRNA occurs via endonucleolytic cleavage at 5'-AAU-3'/5'-UAA-3' elements in single stranded loop domains of the 3'-noncoding region. J. Biol. Chem. *264*, 910-918.

Birnstiel, M. L., Busslinger, M. and Strub, K. (1985). Transcription termination and 3' processing : The end is in site! Cell *41*, 349-359.

Brawerman, G. (1987). Determinants of mRNA stability. Cell *48*, 5-6.

Bray, E. A. and Beachy, R. N. (1985). Regulation by abscisic acid of beta-conglycinin expression in cultured developing soybean cotyledons. Plant Physiol *79*, 746-750.

Chen, Z.-L., Schuler, M. A. and Beachy, R. N. (1986). Functional analysis of regulatory elements in a plant embryo-specific gene. Proc. Nat. Acad. Sci. *83*, 8560-8564.

Comai, L. and Harada, J. J. (1990). Transcriptional activities in dry seed nuclei indicate the timing of the transition from embryogeny to germination. P.N.A.S. *87*, 2671-2674.

Culzone, F. J., Britten, R. J. and Davidson, E. H. (1987). Mapping gene transcripts. Methods in Enzymol *152*, 611-632.

Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986). mRNA transcripts of several higher plant genes are polyadenylated at multiple sites *in vivo*. Nucl. Acids Res. *14*, 2229-2240.

DeLisle, A. J. and Crouch, M. L. (1989). Seed storage protein transcription and mRNA levels in *Brassica napus* during development and in response to exogenous abscisic acid. Plant Physiol. *91*, 617-623.

Draper, D. E. (1989). How do proteins recognize specific RNA sites? New clues from autogenously regulated ribosomal proteins. T.I.B.S. *14*, 335-338.

Ericson, M. L., Rodin, J., Lenman, M., Glimelius, K., Josefsson, L.-G. and Rask, L. (1986). Structure of the rapeseed 1.7S storage protein, Napin, and its precursor. J. Biol. Chem. *261*, 14576-14581.

Fernandez, D. E., Qu, R. and Huang, A. C. (1988). Immunogold localization of the L3 protein of maize lipid bodies during germination and seedling growth. Plant Physiol. *86*, 277-274.

Goldberg, R. B., Barker, S. J. and Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. Cell *56*, 149-160.

Goodall, G. J. and Filipowiz, W. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNA`s are reuired for splicing. Cell *58*, 473-483.

Gurr, M. I. (1980). The biosynthesis of triacylglycerol. The Biochem of Plants *4*, 205-248.

Hamm, J. and Mattaj, I. W. (1990). Monomethylated cap structures facilitate RNA export from the nucleus. Cell *63*, 109-118.

Harada, J. J., Barker, S. J. and Goldberg, R. B. (1989). Soybean ß-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and post-transcriptional processes. The Plant Cell *1*, 415-425.

Harada, J. J., DeLisle, A. J., Baden, C. S. and Crouch, M. L. (1989). Unusual sequence of an abcisic acid-inducible mRNA which accumulates late in *Brassica napus* seed development. Plant Mol. Biol. *12*, 395-401.

Harker, C. L., Ellis, T. H. N. and Coen, E. S. (1990). Identification and genetic regulation of the chalcone synthase multigene family in pea. The Plant Cell *2*, 185-194.

Hatzopoulos, P., Franz, G., Choy, L. and Sung, R. Z. (1990). Interaction of nuclear factors with upstream sequences of a lipid body membrane protein gene from carrot. The Plant Cell *2*, 457-467.

Herman, E. M. (1987). Immunogold-localization and synthesis of an oil-body membrane protein in developing soybeans. Planta *172*, 336-345.

Herman, E. M., Melroy, D. L. and Buckhout, T. J. (1990). Apparent processing of a soybean oil body protein accompanies the onset of oil mobilization. Plant Physiol. *94*, 314-349.

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

Hwang, S.-P. L., Eisenberg, M., Binder, R., Shelness, G. S. and Williams, D. L. (1989). Predicted structures of apolipoprotein II mRNA constrained by nuclease and dimethyl sulphate reactivity; stable secondary structures occur predominantly in local domains via intraexonic base pairing. J. Biol. Chem. *264*, 8410-8418.

Ichihara, K. (1984). sn-Glycerol-3-phosphate acyltransferase in a particulate fraction from maturing safflower seeds. Archiv Biochem Biophys *232*, 685-698.

Ichihara, K. and Noda, M. (1982). Some properties of diacylglycerol acyltransferase in a particulate fraction from maturing safflower seeds. Phytochem *21*, 1898-1901.

Ingelbrecht, I. L. W., Herman, L. M. F., Dekeyser, R. A., Van Montagu, M. C. V. and Depicker, A. G. (1989). Different 3' end regions strongly influence the level of gene expression in plant cells. The Plant Cell *1*, 671-680.

Jofuku, K. D., Okamuro, J. K. and Goldberg, R. B. (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature *328*, 734-736.

Josefsson, L.-G., Lenman, M., Ericson, M. L. and Rask, L. (1987). Structure of a gene encoding the 1.7S storage protein, Napin, from *Brassica napus*. J. Biol. Chem. *262*, 12196-12201.

Joshi, C. P. (1987a). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl. Acids Res. *15*, 6643-6653.

Joshi, C. P. (1987b). Putative polyadenylation signals in nuclear genes of higher plants : a compilation analysis. Nucl. Acids. Res. *15*, 9627-9640.

Juliani, M. H., Soutla, G. M. and Klein, C. (1990). cAMP stimulation of *Dictyostelium discoideum* destabilizes the mRNA for 117 antigen. J. Biol. Chem. *265*, 9077-9082.

Keith, B. and Chua, N.-H. (1986). Monocot and dicot pre-mRNA's are processed with different efficiencies in transgenic tobacco. EMBO J. *5*, 2419-2425.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105-132.

Maeda, N. and Smithies, O. (1986). The evolution of multigene families : Human haptoglobin genes. Ann. Rev. Genet. *20*, 81-108.

Marcotte, W. R., Russell, S. H. and Quatrano, R. S. (1989). Abscisic acid-responsive sequences from the Em gene of wheat. The Plant Cell *1*, 969-976.

Miller, H. (1987). Preparing phage and plasmid DNA. Methods in Enzymol *152*, 145-170.

Mundy, J. and Chua, N.-H. (1988). Abscisic acid and water stress induce the expression of a novel rice gene. EMBO J. *7*, 2279-2286.

Mundy, J., Yamaguchi-Shinozaki, K. and Chua, N. H. (1990). Nuclear proteins bind conserved elements in the abscisic acid responsive promoter of a rice rab gene. P.N.A.S. *87*, 1406-1410.

Murphy, D. J. and Cummins, I. (1989). Biosynthesis of seed storage products during embryogenesis in rape seed Brassica napus. J. Plant Physiol. *135*, 63-69.

Murphy, D. J., Cummins, I. and Kang, A. S. (1989a). Immunocytochemical and biochemical studies of the mobilisation of storage oil-bodies and proteins in germinating cotyledons and oil seed rape, Brassica napus. J. Sci. Food Agric. *48*, 209-223.

Murphy, D. J., Cummins, I. and Kang, A. S. (1989b). Synthesis of the major oil-body membrane protein in developing rapeseed (Brassica napus) embryos. Biochem. J. *258*, 285-293.

Murphy, D. J., Keen, J. N., O'Sullivan, J. N., Au, M. Y., Edwards, E.-W., Jackson, P. J., Cummins, I., Gibbons, T., Shaw, C. H. and Ryan, A. J. (1991c). A class of amphipathic

proteins associated with lipid storage bodies in plants. Possible similarities with animal serum apolipoproteins. Biochem et Biophys Acta *1088*, 86-94.

Napier, J. A., Chapman, J. M. and Black, M. (1989). Calcium dependent induction of novel proteins by abscisic acid in wheat aleurone tissue of different developmental stages. Planta *179*, 156-164.

Nielsen, N. C., Dickinson, C. D., Cho, T.-J., Thanh, V. H., Scallon, B. J., Fischer, R. L., Sims, T. L., Drews, G. N. and Goldberg, R. B. (1989). Characterization of the glycinin gene family in soybean. The Plant Cell *1*, 313-328.

Ohlrogge, J. B. (1987). Biochemistry of Plant Acyl Carrier Protein. The Biochem. of Plants *9*, 137-157.

Okamuro, J. K., Jofuku, K. D. and Goldberg, R. B. (1986). Soybean seed lectin gene and flanking nonseed protein genes are developmentally regulated in transformed tobacco plants. P.N.A.S. *83*, 8240-8244.

Prescott, A. and Martin, C. (1987). A rapid method for the assessment of levels of specific mRNA's in plants. Plant Mol Biol Reporter *4*, 219-224.

Qu, R. and Huang, A. H. C. (1990). Oleosin KD 18 on the surface of oil bodies in maize. Genomic and cDNA sequences. J. Mol. Biol. *265*, 2238-2243.

Qu, R., Vance, V. B. and Huang, H. C. (1990). Expression of genes encoding oleosin isoforms in the embryos of maturing maize kernels. Plant Sci. *72*, 223-232.

Qu, R., Wang, S.-M., Lin, Y.-H., Vance, V. B. and Huang, H. C. (1986). Characteristics and biosynthesis of membrane proteins of lipid bodies in the scutella of maize (Zea mays). Biochem. J. *235*, 257-265.

Raghow, R. (1987). Regulation of messenger RNA turnover in eukaryotes. T.I.B.S. *12*, 358-360.

Ratnasabapathy, R., Hwang, S.-P. L. and Williams, D. L. (1990). The 3'-untranslated region of the apolipoprotein II mRNA contains two independent domains that bind distinct cytosolic factors. J. Biol. Chem. *265*, 14050-14055.

Rödin, J., Ericson, M. L., Josefsson, L. G. and Rask, L. (1990). Characterization of a cDNA clone encoding a *Brassica napus* 12S protein (cruciferin) subunit: Relationship between precursors and mature chains. J. Biol. Chem. *265*, 2720-2723.

Rosenthal, N. (1987). Identification of regulatory elements of cloned genes with functional assays. Methods in Enzymol *152*, 704-715.

Ryan, A. J., Royal, C. L., Hutchinson, J. and Shaw, C. H. (1989). Genomic sequence of a 12S seed storage protein from oilseed rape (*Brassica napus* var. *Jet neuf*). Nucl. Acids Res. *17*, 3584.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Cloning : A laboratory manual. Molecular Cloning : A laboratory manual

Sanger, F., Micklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. P.N.A.S. *74*, 5463-5467.

Scofield, S. R. and Crouch, M. L. (1987). Nucleotide sequence of a member of the napin storage protein family from Brassica napus. J. Biol. Chem. *262*, 12202-12208.

Shaw, G. and Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA, mediates selective mRNA degradation of *beta*-globin mRNA. Cell *46*, 659-667.

Simon, A. E., Tenbarge, K. M., Scofield, S. R., Finkelstein, R. R. and Crouch, M. L. (1985). Nucleotide sequence of a cDNA clone of *B. napus* 12S storage protein shows homology with legumin from *Pisum sativum*. Plant Mol. Biol. *5*, 191-201.

Skriver, K. and Mundy, J. (1990). Gene expression in response to abscisic acid and osmotic stress. The Plant Cell *2*, 503-512.

Smith, H. (1990). Signal perception, differential expression within multigene families and the molecular basis of phenotypic plasticity. Plant, Cell and Environment *13*, 585-594.

Somerville, C. and Browse, J. (1991). Plant lipids : Metabolism, mutants, and membranes. Science *252*, 80-87.

Sporel, N. A. and Kafatos, F. C. (1987). Identification of genomic sequences corresponding to cDNA clones. Methods in Enzymol *152*, 588-597.

Stobart, A. K., Stymne, S. and Hogland, S. (1986). Safflower microsomes catalyse oil accumulation *in vitro*: A model system. Planta *169*, 33-37.

Stumpf, P. K. (1987). The biosynthesis of saturated fatty acids. The Biochem of Plants *9*, 121-136.

Stymne, S. and Appelqvist, L. A. (1978). The biosynthesis of linoleate from oleoyl CoA via oleoyl phosphatidyl choline in microsomes of developing safflower seeds. Eur J Biochem *90*, 223-229.

Stymne, S. and Stobart, A. K. (1987). Triacylglycerol biosynthesis. The Biochem of Plants *9*, 175-214.

Stymne, S., Stobart, A. K. and Glad, G. (1983). The role of the acyl CoA pool in the synthesis of poly-unsaturated 18-carbon fatty acids and triacylglycerol production in the microsomes of developing safflower seeds. Biochem et Biophys Acta *752*, 198-208.

Sugita, M. and Gruissem, W. (1987). Developmental, organ-specific, and light dependent expression of the tomato ribulose 1,5-bisphosphate carboxylase small subunit gene family. P.N.A.S. *84*, 7104-7108.

Tzen, J. T. C., Lai, Y.-K., Chan, K.-L. and Huang, A. H. C. (1990). Oleosin isoforms of high and low molecular weights are present in the oil bodies of diverse seed species. Plant Physiol. *94*, 1282-1289.

Vance, V. B. and Huang, A. H. C. (1987). The major protein from lipid bodies of maize: Characterization and structure based on cDNA cloning. J. Biol. Chem. *262*, 11275-11279.

Walling, L., Drews, G. N. and Goldberg, R. B. (1986). Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. P.N.A.S. *83*, 2123-2127.

Wickens, M. (1990). How the messenger got its tail : Addition of poly(A) in the nucleus. T.I.B.S. *15*, 277-281.

Wilson, T. and Treisman, R. (1988). Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU rich sequences. Nature *336*, 396-398.