An Investigation into the Components of Triricinoleic Acid Production in the Developing Castor Bean Endoplasmic Reticulum

Gadd, Stephen M.

An Investigation into the Components of Triricinoleic Acid Production in the Developing Castor Bean Endoplasmic Reticulum

Stephen M. Gadd

A Thesis submitted for the degree of Doctor of Philosophy

School of Biological and Biomedical Sciences

Durham University

February 2009

Acknowledgements

I would like to take this opportunity to thank those who, in various ways, have supported me through my PhD programme at the University of Durham.

Firstly, I would like to express my gratitude to my supervisor Professor Toni Slabas for his support throughout my doctorate and for giving me the opportunity to study for a PhD in his laboratory. I would like to thank Dr. Bill Simon for his training, advice and support in proteomics and mass spectrometry. I would also like to thank Dr. John Gatehouse for guiding me through the final stages of the process.

Dr. Daniel Maltman was a continued friend and mentor throughout my time in Durham. He generously shared his knowledge in proteomics and Castor cell biology, and was a pleasure to work with. Dr. Adrian Brown was always willing to provide advice on numerous aspects of my research, especially in lipid analysis. He also became a good friend. Dr. Johan Kroon tutored me through different yeast techniques and we shared many conversations and laughs about our laboratory life. Dr. Steve Chivasa was an inspiration, kind and generous of his time and with his knowledge. Many other people too, Dr. John Rowland, Dr. John Hamilton, Dr. John Hall, contributed to making my time in Durham enjoyable and rewarding.

Finally, I would like to thank my family and loved ones. Gemma Cook, for her patience and steady support and my Mum and Dad, for all their love and kindness through this time.

Abstract

Ricinoleic acid (12-hydroxyoleic acid) has a wide range of industrial uses. Its current source is the castor plant (*Ricinus communis*) which contains up to 90% ricinoleic acid in its seed storage lipid. *R. communis* has significant limitations as an agricultural source of ricinoleic acid; it produces potent allergens, requires hand harvesting and only grows effectively in limited climatic zones. A solution to these limitations is to identify the components of the seed storage lipid biosynthetic pathway and transfer them to an agronomic host such as oil seed rape. The developing seed ER is the major compartment of storage oil biosynthesis, whereas during germination these storage compounds are broken down to support the germinating seedling. In this study, a quantitative gel-based proteomic approach has been used to identify the proteins elevated in the developing seed ER compared to the germinating seed ER. On identification of the protein components of storage lipid biosynthesis in the developing seed, their influence on oil quality will be assessed. The use of yeast may be useful in this regard as the influence of transformed gene products on oil production can be measured within days of transformation. A protocol for the analysis of lipid production, including triricinolein production, in a yeast model has been established.

Developing ER preparations were made from seed material harvested between 25 to 30 days after flowering; a stage where lipid biosynthesis is at its maximum. ER samples prepared from 3-day germinated seed were used in a differential screen to identify a subset of proteins elevated in the developing seed. Four independent replicates of developing and germinating ER were prepared and analysed using 2-Dimensional Difference In-Gel Electrophoresis (2D DIGE). Spots elevated by $\geq$10%, with the criteria that they were present in all four gels with a student t-test value of $p<0.02$, were deemed significant and selected for picking and mass spectrometry analysis. Protein sequence data and peptide mass fingerprints were generated and used to search a complete *R. communis* protein database. Prior to the 2D DIGE analysis, all stages of the proteomic methodology were validated and if necessary optimised for the *R. communis* samples, from sample preparation through fluorescent labelling, isoelectric focussing, reproducibility of the analytical gels to the ability to identify and effectively pick spots from high loading preparative gels.

91 protein spots were identified as significantly elevated in the developing preparations and 15 spots as significantly elevated in the germinating preparations. On analysis with mass spectrometry a total of 54 developing spots and 10 germinating spots gave confident identities. The majority of the developing spots identified were protein chaperones, folding proteins, and storage proteins. No components of lipid biosynthesis were identified in the 2D DIGE analysis, likely due to their membrane bound nature and the loss of these proteins due to poor solubility in 2D electrophoresis buffers or their precipitation during isoelectric focussing.

The oleaginous yeast *Yarrowia lipolytica* was identified as a suitable candidate for in-vivo assays of the effect of *R. communis* lipid biosynthesis gene products on oil composition. This yeast can utilise hydrophobic substrates for growth including ricinoleic acid, makes significant amounts of storage oil, has a complete genome sequence available

and mature genetic tools facilitating its transformation. A protocol for its growth on hydrophobic substrates, lipid extraction and analysis of triricinolein production has been established.

2D DIGE provides a statistically rigourous method for identifying and quantifying elevated proteins in differential screens of plant seed ER. For the identification of the components of lipid biosynthesis in the developing ER, attention now turns to the membranes. Gel-free mass spectrometry based approaches provide the best chance of identifying these proteins and will complement the proteomic analysis presented here.

Contents

List of Figures

List of Tables

List of Abbreviations

βKR β-ketoacyl-ACP reductase

1-OH DAG 1-hydroxy diacylglycerol

1-OH TAG monoricinolein diolein

2-OH DAG 2-hydroxy diacylglycerol

2-OH TAG diricinolein monoolein

2D DIGE Two Dimensional In-Gel Electrophoresis

2DE two dimensional electrophoresis

3-OH TAG triricinolein

R_f retention factor

ACAT acyl-CoA : cholesterol acyl transferase

ACCase acetyl-CoA carboxylase

ACP acyl carrier protein

ACS acyl-CoA synthetase

BSA bovine serum albumin

BSTFA N,O -Bis(trimethylsilyl)trifluoroacetamide

BVA biological variance analysis

CID collision induced dissociation

CyDyes cyanine dyes

DAF days after flowering

DAG diacylglycerol

DAGAT diacylglycerol acyltransferase

DH β-ketohydroxyacyl-ACP dehydrase

DHAP dihydroxyacetone phosphate

DIA differential in-gel analysis

DMF N,N-dimethylformamide

EM electron micrograph

ENR enoyl reductase

ER endoplasmic reticulum

ESI electrospray ionisation

EST expressed sequenced tag

FAD fatty acid desaturase

FAME fatty acid methyl ester

FAS fatty acid synthase

Fe-O-Fe di-iron oxo bridge

FFA free fatty acid

G3P glycerol-3-phosphate

GC gas chomatography

GPAT glycerol-3-phosphate:acyl CoA acyltranferase

H:D:A hexane : diethyl ether : acetic acid

HDA hexane : diethyl ether : acetic acid

HEAR high erucic acid rapeseed

ICAT isotope coded affinity tags

IEF isoelectric focusing

IPG immobilised pH gradient

IT ion trap

iTRAQ isobaric tag for relative and absolute quantification

KAS β-ketoacyl-ACP synthase

kDa kilodalton

LC-MS liquid chromatography mass spectrometry

LCAT lecithin : cholesterol acyltransferase

LPA lysophosphatidic acid

LPAT lysophosphatidic acid acyltransferase

LPC lysophosphatidylcholine

LPCAT lysophosphatidylcholine acyltransferase

MAG monoacylglycerol

MALDI TOF matrix assisted laser desorption/ionisation time-of-flight

MCAT malonyl-CoA:ACP transacylase

MO methyl oleate

MOWSE Molecular Weight Search

MR methyl ricinoleate

MRD derivatised methyl ricinoleate

MS mass spectrometry

MudPIT multidimensional protein identification technology

NBS-LRR nucleotide-binding site leucine-rich repeat

NCBI*nr* National Centre for Biotechnology Information *non-redundant* database

ND non-derivatised

NEPHGE non-equilibrium pH gradient electrophoresis

NLS neutral lipid standard

OD optical density

OH-FAME hydroxy fatty acid methyl ester

OH-FFA hydroxy free fatty acid

OH-MAG hydroxy monoacylglycerol

OMP outer member protein

PA phosphatidic acid

PAP phosphatidic acid phosphatase

PC phosphatidylcholine

PDAT phospholipid:diacylglycerol acyltransferase

PDI protein disulphide isomerase

PE:D:A petroleum ether : diethyl ether : acetic acid

pI isoelectric point

PMF peptide mass fingerprint

PMT photomultiplier tube

PTM post-translational modification

Q1-3 quadrupoles 1-3

QTOF quadrupole time-of-flight

RP reverse phase

Sam50 sorting and assembly machinery protein 50

SCA synthetic carrier ampholytes

SCD1 stearoyl CoA desaturase

SCX strong cation exchange

SDS PAGE sodium dodecyl sulphate polyacrylamide gel electrophoresis

SERCA sarcoplasmic / endoplasmic reticulum Ca^{2+} ATPase

SF solvent front

SP signal peptide

SRP signal recognition particle

TAG triacylglycerol

TE thioesterase

TIGR The Institute of Genomic Research

TLC thin layer chromatography

TMCS trimethylchlorosilane

TQ triple quadrupole

TSP total soluble protein

YNB yeast nitrogen base

YPD yeast extract/peptone/dextrose

Chapter 1

Introduction

1.1 Background

1.1.1 *R. communis* Seed Oil Has Significant Industrial Applications

The seeds of castor bean (*Ricinus communis*) contain up to 90% ricinoleic acid, the majority of which is in the form of storage oil; esterified to all three positions of the glycerol backbone of the triacylglycerol molecule (triricinolein). Whilst most vegetable oils are produced for food and animal feed uses, vegetables oils are also used for industrial uses. For example, up to 15% of soya oil and up to 100% of certain commodity oils are used for industrial applications (McKeon, 2005). Commodity oils containing unusual fatty acids, while having no nutritive value, have chemical properties that make them amenable to industry. For example, ricinoleic acid has a mid-chain hydroxyl group (Figure 1.1) that enhances its viscous properties making it useful in the formulation lubricants, grease and coatings. The reactivity of the mid-chain hydroxyl group also allows the production of a range of thermopolymers and plastics (Caupin, 1997). It is the principle ingredient in the synthesis of Nylon 11, an important polymer for the automotive industry due to its superior heat and solvent resistant properties. Castor oil is an important ingredient in the cosmetics industry, where it is found in a diverse array of products, primarily as skin-conditioning agents, emulsion stabilisers and surfactants. The highest reported concentration of castor oil in a cosmetic product is a lipstick which contained 81% castor oil (Elmore, 2007). In the United States, 40 to 45 thousand tonnes of castor oil are imported annually to supply its entire domestic needs (Brigham, 1993). The European Union is similarly dependent on imported stocks, consuming 40% of world castor oil production (Laureti et al., 1998). Its importance to the industries of both regions mean it is considered to be of strategic importance, and in the case of the United States it is stockpiled should imports be disrupted in times of global turbulence (Brigham, 1993; Roetheli et al., 1991).

Historically, castor oil has been used as an engine lubricant but mineral oil based

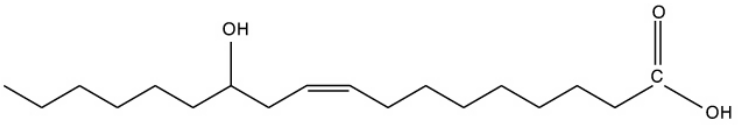


Figure 1.1: Structure of ricinoleic acid (12-hydroxyoleic acid).

products have generally replaced this role. The rise in mineral oil prices and the increasing will to reduce carbon emissions makes alternative vegetable oil-based engine lubricants more attractive as a product. Vegetable oil-based engine lubricant formulations exist and the inclusion of 10% ricinoleic acid is essential for their performance (James and Johnson, 1999). The inherently distributed nature of biological systems, that is, their ability to produce oil locally and on a scale appropriate for local demands, reduces the requirement to ship huge volumes of mineral oil around the world and the major associated environmental risks of spillage.

1.1.2 Castor Oil Production: Problems and Alternatives

R. communis is currently the only commercial source of triricinolein and its derivatives, yet the global scale of its production is restricted due to reasons which will be outlined here. *R. communis* is highly susceptible to divergent climatic variations, restricting its reliable growth to sub-tropical climatic regions of the world. Seed development occurs at different rates on a single *R. communis* raceme (i.e. it is a non-determinant plant) which makes optimum harvesting difficult to achieve. It produces toxic compounds which remain in the seed meal after oil pressing: the mildy toxic pyridine alkaloid ricinine and the highly toxic (7 times more toxic than cobra venom) ribosome inactivating protein ricin. *R. communis* also produces potent allergens including 2S albumin (Bashir et al., 1998) which can result in sensitisation of the harvester to the crop.

One approach being taken is to circumvent some of the plant's difficulties within the agricultural setting by identifying and removing or replacing the genes involved

in toxin and allergen biosynthesis (McKeon et al., 2002). If this approach works, and the nutritional content of the seed meal (which is currently heat-inactivated and used as an animal feed) can be maintained, the agronomic difficulties of restricted growing locations and its non-determinant seed development would remain. An alternative approach is to identify the constituent enzymes of the high triricinolein production biochemical pathway and ultimately re-build it in a plant such as *Brassica napus*. This may allow the production of high percentage triricinolein in a plant suited to agriculture, especially growth in temperate climatic regions, without the production of toxins and allergens associated with its traditonal source. Such a development would be compatible with the political will of large power blocs such the United States, as outlined in legal orders to fast track the creation of novel renewable plant sources of energy, chemicals and pharmaceuticals (Executive Order 13134 - Developing and Promoting Biobased Products and Bioenergy, 1999) (Clinton, 1999).

1.2 Plant Lipid Biochemistry

Metabolic pathways leading to the formation of neutral lipids, and hydrolytic reactions catalysing their mobilisation, are similar in the different kingdoms of life (Daum et al., 2007). The biochemistry of fatty acid and storage oil production in plants is examined here, presented alongside the biochemistry of storage oil production in yeast. Special attention is given to the relevant organisms *R. communis* and *Y. lipolytica* where applicable.

1.2.1 Plastidial Fatty Acid Biosynthesis

De novo fatty acid biosynthesis occurs in the plastid, and in developing oil seed virtually all *de novo* synthesised fatty acids are incorporated into storage oil (Murphy 1993). *De novo* fatty acid biosynthesis requires the concerted activity of acetyl-CoA carboxylase

(ACCase) and fatty acid synthase (FAS). ACCase catalyses the carboxylation of acetyl CoA to form malonyl CoA in an ATP and bicarbonate dependent reaction. ACCase is a key regulatory enzyme in fatty acid biosynthesis, as it is the first committed step of the pathway. Its regulation determines the level of fatty acid biosynthesis (Sasaki and Nagano, 2004; Ohlrogge and Jaworski, 1997).

The second fatty acid biosynthesis enzyme is FAS, which is found in two distinct arrangements in biology. Mammals, fungi and certain bacteria have a Type I FAS enzyme, a very large multi-domain enzyme where each reaction step is catalysed at a particular active site (Rock and Cronan, 1996). In contrast, plants and many bacteria including *Eschericia coli* have discrete fatty acid synthases consisting of at least eight separate proteins that are together termed Type II FAS (Shimakata and Stumpf, 1982).

Following acetyl CoAs conversion to malonyl CoA by ACCase, the malonyl CoA molecule can enter the FAS cycle. Malonyl-CoA:ACP transacylase (MCAT) converts malonyl CoA to malonyl acyl carrier protein (ACP) which is the carbon donor molecule for acyl chain elongation (Slabas et al., 2001). Elongation begins with the condensation of malonyl ACP with a molecule of acetyl CoA by the enzyme β-ketoacyl-ACP synthase (KAS) III. This forms β-ketoacyl-ACP, which is subsequently reduced by β-ketoacyl-ACP reductase (βKR) in an NADPH dependent reaction to form β-hydroxyacyl-ACP. This resulting compound then has a water molecule removed by β-ketohydroxyacyl-ACP dehydrase (DH) to give *trans* 2-enoyl acyl-ACP which is reduced by enoyl reductase (ENR) utilising NADH as the reductant. This initial sequence of reactions produces butyryl ACP, a four carbon fatty acid. Butyryl ACP can then replace acetyl CoA as the co-substrate (with malonyl CoA) for further elongation. However, another KAS enzyme is required for this to occur: KAS I. Unlike KAS III, KAS I uses acyl ACPs as a substrate rather than acetyl CoA. KAS I continues to feed elongated acyl ACPs into the reduction / dehydration / reduction pathway, increasing the acyl length by 2 carbons at a time, until the chain length reaches the 16 carbon (C16:0) product of

palmitoyl ACP (Stymne and Stobart, 1987). The remaining condensing enzyme, KAS II, can further elongate palmitoyl ACP to the 18 carbon (C18:0) product stearoyl ACP. The activity of KAS II is especially relevant as it determines the C16 / C18 ratio of the resulting acyl ACP pool, which in turn influences the degree of unsaturation of the final oil product as stearoyl ACP is almost entirely desaturated (Salas et al., 2000). The reactions of fatty acid biosynthesis are described in Table 1.1.

Table 1.1: Enzymes of plant Type II FAS. After transacylation of malonyl CoA to malonyl ACP, reactions occur in a cycle of condensation, reduction, dehydration and reduction. Three types of KAS enzyme condense increasing carbon chain length molecules to acetyl CoA with each cycle of the FAS reactions. Malonyl ACP (2 carbon) is initially condensed by KAS III to acetyl CoA. KAS I feeds C4:0 to C14:0 acyl ACPs into the cycle. The final condensation step is catalysed by KAS II, which feeds C16:0 palmitoyl ACP into a further round condensation, reduction, dehydration and reduction resulting in the 18 carbon (C18:0) stearoyl ACP product

Reaction type	Enzyme name		Substrate(s)	Product
	Full name	Short hand		
Transacylation	malonyl-CoA:ACP transacylase	MCAT	malonyl CoA	malonyl ACP
Condensation (Initial)	β-ketoacyl-ACP synthase III	KAS III	malonyl ACP (C2) + acetyl CoA	β-ketoacyl-ACP
Reduction	β-ketoacyl-ACP reductase [a]	βKR	β-ketoacyl-ACP	β-hydroxyacyl-ACP
Dehydration	β-ketohydroxyacyl-ACP dehydrase [c]	DH	β-hydroxyacyl-ACP	*trans*-2 enoyl acyl ACP
Reduction	enoyl reductase [b]	ENR	*trans*-2 enoyl acyl ACP	butyryl ACP (C4:0)
Condensation (2nd to penultimate)	β-ketoacyl-ACP synthase I	KAS I	C4:0 $\Rightarrow$ C14:0 ACP + acetyl CoA	β-ketoacyl-ACP (C6:0 $\Rightarrow$ C16:0)
Condensation (final)	β-ketoacyl-ACP synthase II	KAS II	palmitoyl (C16:0) ACP + acetyl CoA	β-ketoacyl-ACP (C18:0)

[a] NADPH dependent

[b] NADH dependent

[c] Releases H_2O molecule

The plastid synthesised acyl ACPs are either substrates for incorporation into glycerolipids such as monogalactosyldiacylglycerol and digalactosyldiacylglycerol by acyl transferases, or they can be hydrolysed by acyl-ACP thioesterases to produce ACP and free fatty acid. The released fatty acids are subsequently exported to the cytosol from the plastid, although the exact mechanism of transport is not yet certain. The substrate selectivities of thioesterases determines the pattern of fatty acid products synthesised by FAS. Hellyer et al. (1982) determined the substrate selectivity of *B. napus* acyl-ACP thioesterase for C16:0 and C18:1 fatty acid species. They found *B. napus* has considerable preference for C18:1 explaining why only small amounts of C16:0 are exported from the plastid (Hellyer et al., 1992). Cocoa butter from the cocoa tree (*Theobroma cacoa*) has an unusually high content of stearic acid (C18:0). It was hypothesised that this resulted from a thioesterase with a high preference for C18:0 fatty acid species rather than the usual C18:1 desaturated oleoyl-ACP substrate. Griffiths *et al* found that *T. cacoa* thioesterase had the usual preference for oleoyl-ACP, but it also showed four times as much activity for stearoyl-ACP compared to other plastid acyl-ACP thioesterases (Griffiths et al., 1993). Although it is thought a low activity of *T. cacoa* stearoyl-ACP $\triangle$9-desaturase plays a part in its unusually high levels of 18:0 species it is likely that its thioesterase substrate preference contributes to the high levels of stearic acid, indicating the roles thioesterases can play in final fatty acid profiles. California bay laurel (*Umbellularia californica*) (Pollard et al., 1991) is another example of the influence of thioesterase specificity on fatty acid makeup. It accumulates C10:0 and C12:0 fatty acids as the principle reserve of fatty acyl groups due to the presence of highly specific thioesterases.

1.2.2 Fatty Acid Modification and Hydroxylation

Desaturation is the introduction of double bonds at specific points along the acyl chain of a fatty acid. The main products of fatty acid biosynthesis are the C16:0 or C18:0

saturated acyl chains. However, most plant oils are rich in desaturated fatty acid species, such as oleic acid (C18:1) and linoleic acid (C18:2). For example, oil seed rape (*B. napus*) contains $\sim 60\%$ oleic acid in its seed. Plants must then have efficient desaturase capabilities. In the plastid stroma there is a highly active soluble stearoyl-ACP $\triangle 9$ desaturase which introduces a double bond at the $\triangle 9$-10 position of the acyl chain. This highly active desaturase results in oleate as the main product of fatty acid formation in most plant species (Shanklin and Somerville, 1991). Knutzon et al. (1992) used *B. napus* antisense stearoyl-ACP $\triangle 9$ desaturase sequences to manipulate the activity of this enzyme in an early genetic engineering of plant oils, significantly reducing the unsaturated fatty acid pool in *B. napus* seed oils and producing a high stearate oil (Knutzon et al., 1992). Spectroscopic analysis of the enyzme revealed evidence of a diiron-oxo centre in the desaturase (Fox et al., 1994), a powerful reactive group capable of hydrogen abstraction in the formation of double bonds and well conserved within this class of enzyme (Browse, 1996). The crystal structure of *R. communis* stearoyl-ACP $\triangle 9$ desaturase has been published (Lindqvist et al., 1996) which revealed a tunnel into the protein, lined by hydrophobic residues and of sufficient size for an 18 carbon substrate. The crystal structure also revealed the presence of the diiron structure within the hydrophobic tunnel. A diverse array of acyl chain modifications occur in plants, including other desaturation reactions. However, the majority of these occur in the endoplasmic reticulum (ER) (see Section 1.2.2) after plastid export.

Oleate produced through plastidial fatty acid biosynthesis and stearoyl-ACP $\triangle 9$ desaturation can be further desaturated to form polyunsaturated fatty acids. This can take place within the plastids or the ER, via what are respectively known as the prokaryotic or eukaryotic desaturation pathways (Heinz and Roughan, 1983). With the eukaryotic pathway, oleate is exported out of the plastid where it is subsequently activated to its CoA form by acyl-CoA synthetase (ACS). The acyl-CoA intermediates are incorporated into phosphatidylcholine (PC) where the PC-esterified acyl chain is

then a suitable substrate for ER resident desaturases. For example, oleate can be desaturated to linoleate (C18:2$^{\triangle 12}$) by the FAD2 (fatty acid desaturase) oleate $\triangle$12 desaturase, an enzymatic reaction initially characterised in 1972 (Vijay and Stumpf, 1972). Deficiencies in the activity of this enzyme caused by mutation to Asp150 to Arg150 produced a high oleic acid variety of *Arachis hypogaea*, the SunOleic 95R peanut (Gorbet and Knauft, 1997). Classical breeding techniques produced mutant *B. napus* lines with $\sim$ 80% oleic acid content. Analysis of these mutant lines indicated that the further increase in oleic acid resulted in a decrease in cold tolerance during seed germination. This was ascribed to a reduction in the polyunsaturated fatty acid content of membrane lipid (Kinney, 1994; Miquel and Browse, 1994). Indeed, the extent of desaturation of membrane fatty acids is known to play a significant role in cold tolerance, with the upregulation of fatty acid desaturases in response to cold shock reported in cyanobacteria *Synechocystis* sp. PCC6803 (Suzuki et al., 2000). When a $\triangle$9 desaturase cloned from the cyanobacterium *Anacystis nidulans* was introduced into tobacco, the transgenic plants exhibited significant chilling resistance (Ishizaki-Nishizawa et al., 1996).

Further desaturation of linoleate to form linolenate (C18:3$^{\triangle 12,15}$) is catalysed by $\triangle$15 desaturase (FAD3). Plastid localised $\triangle$12 and $\triangle$15 desaturases have been found, but their contribution to overall fatty acid desaturation is insignificant compared to the ER resident desaturases; instead they are primarily utilised for the production of specialised thylakoid membrane lipid (Harwood, 1996).

Oleate $\triangle$12 hydroxylase is a membrane bound enzyme that catalyses the hydroxylation of oleic acid (C18:1) to give ricinoleic acid (C18:1-OH). *In vivo* labelling studies on developing castor endosperm have shown that the reaction proceeds via a direct hydroxyl substitution with the oleic acid moiety without requiring an intermediate step such as desaturation (Morris, 1967). Hydroxylase activity was localised to the ER (Galliard and Stumpf, 1996) and this was confirmed by Maltman et al. (2002) where

a purified $NaCl/NaCO_3$ washed ER membrane fraction was separated by SDS PAGE and the bands excised and identified by matrix assisted laser desorption/ionisation time-of-flight (MALDI TOF) mass spectrometry and peptide mass fingerprinting.

Bafor et al. (1991) used developing castor microsome preparations to investigate the hydroxylation of oleic acid. Using [^{14}C]-oleoyl CoA they found it was readily incorporated into PC at the *sn*-2 position, but very rarely at the *sn*-1 position. When NADH was included in the assay radioactive ricinoleate was synthesised from the [^{14}C]-oleate and this was primarily recovered in the PC fraction and as free fatty acid (FFA). They were unable to find any evidence that oleoyl CoA is a direct substrate for the oleate $\triangle$12 hydroxylase, rather that the action of oleate $\triangle$12 hydroxylase is dependent on oleate being incorporated into PC (Bafor et al., 1991). The group also reported that after the introduction of NADH to the reaction mixture, radioactivity in the PC fraction decreased and a corresponding increase was observed in the free fatty acid pool. This suggests that as the [^{14}C]oleoyl-PC is hydroxylated to form [^{14}C]ricinoleoyl-PC it is selectively removed by phospholipase A (Bafor et al., 1991). Hydroxylase is NADH dependent as this provides the reducing power for the monooxygenase (incorporation of a single atom of oxygen from water into the substrate) reaction. In the presence of Mg^{2+}, ATP and CoA they found that [^{14}C]-ricinoleate released from the PC fraction was activated by ACS to form ricinoleoyl CoA which was then readily incorporated into TAG by the acylation of glycerol-3-phosphate (G3P). A year later this group reported the role of cytochrome b5 as an electron transporter from NAD(P)H to the site of hydroxylation. They raised antibodies against cytochrome b5 which when added to the reaction mixture shut down the action of both oleate $\triangle$12 hydroxylase and oleate $\triangle$12 desaturase (Smith et al., 1992). Plant fatty acyl desaturases contain two histidine rich motifs that form a di-iron oxo (Fe-O-Fe) bridge which are a critical component of the catalytic activity of the enzyme. Raman spectroscopy has been used to show that the soluble ubiquitous plant desaturase stearoyl $\triangle$9 desaturase from *R. communis*

contains this di-iron oxo bridge (Fox et al., 1994). The Fe-O-Fe cofactor is also found in the bacterial hydroxylase enzyme methane monooxygenase hydroxylase. Van de Loo et al. (1995) hypothesised that the *R. communis* oleate △12 hydroxylase would have these histidine rich motifs for the binding of the Fe-O-Fe co-factor. They also hypothesised that as *R. communis* leaves do not produce ricinoleic acid that there would not be oleate △12 hydroxylase gene expression in these tissues. They prepared a cDNA library from developing *R. communis* endosperm which was then differentially subtracted for expressed leaf cDNAs. They obtained 468 clones with this differential screening method of which three were found to have the two desaturase/hydroxylase histidine rich regions. They found that these clones were all of a single gene which was very strongly expressed in developing endosperm but not expressed at all in leaf (Van de Loo et al., 1995). Transgenic tobacco lines were developed containing the putative *R. communis* hydroxylase/desaturase gene and these were found to contain 0.1% ricinoleic acid in total seed fatty acids confirming the oleate △12 hydroxylase activity of this gene. *Lesquerella fendleri* also accumulates ricinoleic acid in its seed oil, along with lesquerolic and densipolic acid. The oleate △12 hydroxylase from *L. fendleri* was isolated on the basis of sequence similarity with the *R. communis* oleate △12 hydroxylase (Broun et al., 1998a). Interestingly, they presented evidence that the *L. fendleri* oleate △12 hydroxylase is bifunctional, having both hydroxylase and desaturase activities. Expression of the gene in an *A. thaliana fad2* mutant (which is deficient in oleate △12 desaturase activity) resulted in a partial suppression of the mutant phenotype in the roots suggesting desaturase activity. Further investigations into the activity of *L. fendleri* oleate △12 hydroxylase confirmed its bifunctionality and also identified it in *R. communis* oleate △12 hydroxylase (Smith et al., 2003). This group made use of an *A. thaliana* mutant line containing lesions in the △12 desaturase (*FAD2*) and C18:1 elongase (*FAE1*) genes. It produces a seed oil containing over 80% oleic acid (C18:1) but no linoleic acid (C18:2). Using seed specific promoters they

transformed separate *fad2/fae1* mutants with either the *L. fendleri* or *R. communis* oleate $\triangle 12$ hydroxylase. Fatty acid analysis of these transformed plants showed they both produced ricinoleic acid (C18:1-OH) in their seed oil but also the polyunsaturated fatty acid linoleic acid (C18:2) (Smith et al., 2003). The reported bifunctionality and significant homology of oleate $\triangle 12$ hydroxylase and oleate $\triangle 12$ desaturase indicates very little evolutionary divergence. Indeed, the *R. communis* oleate $\triangle 12$ desaturase has been converted into an oleate $\triangle 12$ hydroxylase and vice versa by site-directed mutagenesis (Broun et al., 1998b).

Some storage lipid contains acyl residues with a chain length greater than 18 carbons. Elongases, complexes localised to the membrane of the ER, catalyse elongation in a similar manner to fatty acid synthesis, although different enzymes are involved and acyl and malonyl residues are activated as acyl-CoA esters. The elongase complex cycles through four successive reactions of condensation (3-ketoacyl CoA synthetase), reduction (3-ketoacyl reductase) dehydration (3-OH acyl CoA dehydratase) and a further reduction (enoyl reductase) (Domergue et al., 2000). For example, C22:1 is the product of the elongation of oleoyl CoA and malonyl CoA, and is the major fatty acid component of the natural high erucic acid rapeseed (HEAR) variety which accumulates $\sim 50\%$ C22:1.

1.2.3 PA and DAG Production

Plant energy reserves in the form of fatty acids are, with the exception of jojoba (*Simmondsia chinensis*), esterified to a glycerol backbone to form triacylglycerol (TAG) molecules (in *S. chinensis* fatty acids are esterified to long chain alcohols to form waxes (Yermanos, 1975)). The chemical properties of TAG molecules make them excellent energy stores. They store more energy per gram than any other cell component, giving 2.5 times the ATP yield of glycogen and, being hydrophobic, require less water for hydration than polysaccharide and thus are far less bulky per gram to store.

The acyl chains produced in the plastid by the FAS enzymes are transferred to a glycerol backbone by a series of esterification reactions catalysed by the acyltransferase enzymes in the endoplasmic reticulum. The classical pathway of reactions resulting in TAG is known as the Kennedy pathway (Kennedy, 1961). It consists of a series of reactions sequentially acylating a G3P molecule and producing the intermediates of monoacylglycerol (MAG), phosphatidic acid (PA) and diacylglycerol (DAG). The final step in the enzymatic production of TAG is known to occur through two broad alternative routes, the acyl-CoA dependent and independent pathways. These are considered separately in Sections 1.2.4.1 and 1.2.4.2 respectively.

G3P is initially acylated with an acyl-CoA molecule by the action of glycerol-3-phosphate:acyl CoA acyltranferase (GPAT, EC 2.3.1.15) at the *sn*-1 position of the glycerol backbone to create lysophosphatidic acid (LPA, see Figure 1.2, step A). This enzyme resisted cloning for a considerable period, although more was known about the soluble chloroplast GPAT which had been cloned and its structure determined in squash (*Cucurbita moschata*) (Slabas et al., 2000). GPAT has been cloned in *A. thaliana* (AtGPAT) where there are at least 8 GPAT isoforms, the majority of which are ER localised (Zheng et al., 2003). Substrate selectivities of GPAT isoforms 1 and 5 have been studied in this organism, although neither have been reported to play a clear role in TAG biosynthesis (Zheng et al., 2003; Beisson et al., 2007).

A further acylation of LPA is catalysed by lysophosphatidate acyltransferase (LPAT, EC 2.3.1.51) at the *sn*-2 position of LPA resulting in PA (see Figure 1.2, step B). Early insight into the substrate selectivities of LPAT enzymes was gained when Ichihara (1984) reported that *Carthamus tinctorius* (Safflower) microsomes show a preference for palmitate (C16:0) in the formation of LPA. LPA was reported as the the major product when microsomal membranes were incubated with [^{14}C]acyl-CoA and G3P (Ichihara, 1984). However, this observation was only made when *C. tinctorius* microsomes were incubated with saturated acyl-CoAs. The experiment was repeated with linoleoyl CoA

and G3P and both *sn*-1 and *sn*-2 positions were readily acylated in the formation of PA (Griffiths et al., 1985). This supported the observation that no saturated fatty acids were present in the *sn*-2 position of *C. tinctorius* TAG (Ichihara and Noda, 1980) and indicates the selectivity *C. tinctorius* LPAT has for non-saturated acyl-CoAs. Analysis of the substrate selectivity of LPAT from the high erucic acid rapeseed (HEAR) variety of *B. napus* found that, despite the enrichment of erucic acid in the seed oil of the plant, the LPAT does not utilise erucic acid (C22:1) at all (Cao et al., 1990). Earlier positional analysis of the oil from a HEAR variety of *B. napus* found that erucic acid is a very minor component at the *sn*-2 position of the seed oil (Brockerhoff, 1971), see Table 1.2. Two LPATs were cloned from *Limnanthes douglassi* one of which was found to utilise erucic acid (Brown et al., 1995). When introduced into a HEAR *B. napus* plant, they demonstrated a dramatic increase in the C22:1 content at the *sn*-2 position of the *B. napus* seed oil. There are five predicted isoforms of LPAT in *A. thaliana* which have been reported to be ER localised (Kim et al., 2005). They are likely to play differing roles in the formation of TAG in this organism.

Phosphatidic acid phosphatase (PAP, EC 3.1.3.4) catalyses the removal of the phosphate group from the *sn*-3 position of PA and the formation of DAG (see Figure 1.2, step C). It is the formation of DAG that allows the final acylation in the formation of TAG or alternatively for the molecule to be converted into the membrane lipid PC (Slack et al., 1985; Browse and Somerville, 1991), catalysed by the reversible CDP-choline:cholinephosphotransferase. PA is also a precursor for the synthesis of phosphatidylinositol, phosphatidylserine, phosphatidylethanolamine and phosphatidylglycerol. Therefore, the flux between DAG and PA catalysed by the action of PAP is highly significant in the relative directions of lipid metabolism in the cell. PAP, identified in 1955 (Kates, 1955), was first purified to homogeneity in *Persea americana* (avocado) in 1998 (Pearce and Slabas, 1998). The enzyme, a monomeric protein of 51 kDa, displayed a wide substrate selectivity, dephosphorylating not only PA

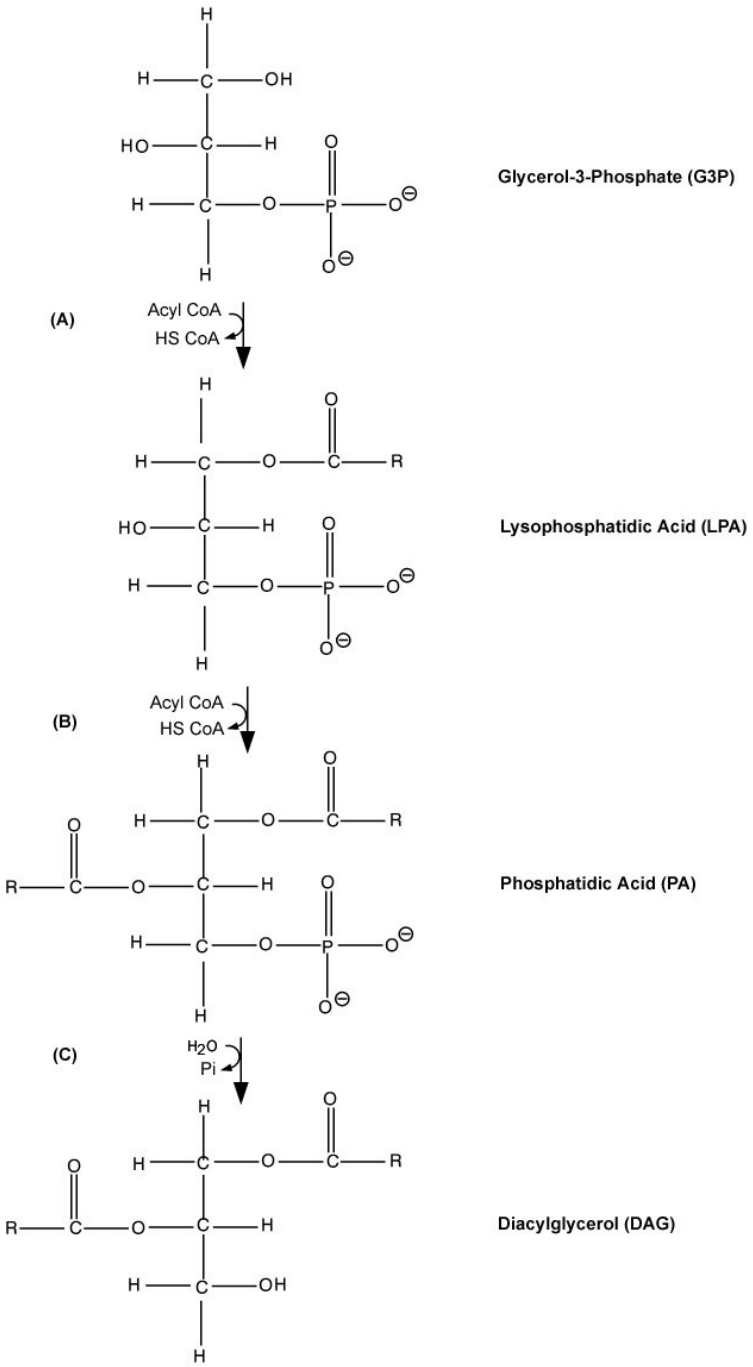


Figure 1.2: Route of PA and DAG Biosynthesis. G3P is acylated by an acyl-CoA at the *sn*-1 position by glycerol-3-phosphate acyltransferase (GPAT, A) to form LPA. This molecule can then be acylated at the *sn*-2 position by another acyl-CoA and lysophosphatidic acid acyltransferase (LPAT, B) to form PA. Dephosphorylation of PA is catalysed by phosphatidic acid phosphatase (PAP, C) to yield DAG and inorganic phosphate.

but also *sn*-1 LPA, *sn*-2 LPA and ceramide-1-phosphate. Interestingly, *sn*-1 LPA was demonstrated to be a competitive inhibitor of the dephosphorylation of PA. It was proposed that this may hinder the formation of DAG and that to prevent this there may be mechanisms for substrate channelling between enzymes of the TAG biosynthetic pathway to prevent competitive interactions such as would be seen with *sn*-1 LPA with PAP (Pearce and Slabas, 1998). The first plant PAP was cloned from *A. thaliana* in 2001. It was putatively identified as a PAP based on its homology to both mammalian and yeast PAP genes. Heterologous expression of the gene (*AtLPP1*) in yeast and *in vitro* biochemical characterisation identified phosphatidic acid phosphatase activity, confirming its identity (Pierrugues et al., 2001). An additional PAP-encoding gene was reported in this study (*AtLPP2*), identified by its homology to *AtLPP1*, which was found to have a different pattern of expression and a different substrate preference for phosphatidic acid when compared to *AtLPP1*. Subsequently, additional PAP homologues have been cloned in *A. thaliana* including plastid specific PAPs (Nakamura et al., 2007).

1.2.4 TAG Production

The final step in the production of storage oil is the acylation of the *sn*-3 position of a DAG molecule. This occurs through two broad routes, one of which utilises acyl chains attached to CoA molecules, the other utilises acyl chains that are independent of CoA molecules. The result is the same, a glycerol-derived molecule with acyl groups at all three positions of its backbone (Figure 1.3).

Analysis of the acyl chains at each *sn* position of the TAG molecules from the seeds of different oilseed crops revealed significantly different compositions of acyl chains, both between the different oil seeds and within a TAG molecule (see Table 1.2). The existence of acyl substrate selectivities of GPAT and LPAT enzymes was described in the previous section. Table 1.2 describes the different compositions of acyl chains

Table 1.2: Positional analysis of TAG acyl chains in different oil seeds. Analysis reveals acyl chain heterogeneity between *sn* positions, plant species and plant varieties.

		Fatty acids (%)						
Species	*sn* Position	16:0	18:0	18:1	18:2	18:3	20:1	22:1
B. napus (Rape) high erucic acid [a]	1	4	2	23	11	6	16	35
	2	1	0	37	36	20	2	4
	3	4	3	17	4	3	17	51
B. napus (Rape) low erucic acid [b]	1	6	2	65	16	7	2	1
	2	Trace	Trace	53	31	16	Trace	Trace
	3	8	2	71	10	5	3	1
Glycine max (Soya) [c]	1	14	6	23	48	9	-	-
	2	1	Trace	22	70	7	-	-
	3	13	6	28	45	8	-	-
C. tinctorius (Safflower) [d]	1	9	2	8	81	-	-	-
	2	0	0	8	92	-	-	-
	3	4	1	7	88	-	-	-

[a] (Brockerhoff, 1971)
[b] (Kaplan et al., 1975)
[c] (Gunstone, 1979)
[d] (Ichihara and Noda, 1980)

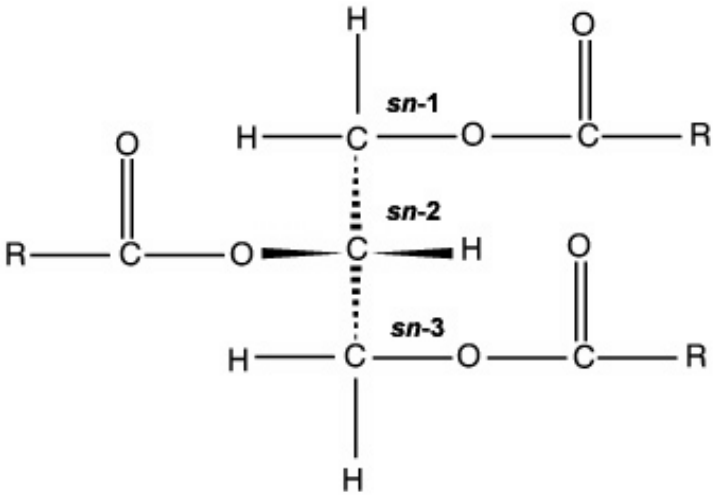


Figure 1.3: Structure of the triacylglycerol molecule. The R denotes acyl side chains. The glycerol molecule does not have rotational symmetry, allowing the individual carbon atoms of the glycerol backbone to be distinguished from one another. Stereochemical numbering is denoted by *sn*.

present at the *sn*-3 position of TAG, indicating the range of selectivities of the enzymes that catalyse *sn*-3 position acyl transfer. For example, 51% of the acyl groups esterified to the *sn*-3 position of TAG in a HEAR variety of *B. napus* is erucic acid (C22:1) (Brockerhoff, 1971), yet in a non-HEAR variety no erucic acid could be observed at all (Kaplan et al., 1975). The relative roles of the two routes to TAG formation in plants are described below.

1.2.4.1 Acyl-CoA Dependent Pathway

Diacylglycerol acyltransferase (DAGAT) catalyses the final stage in TAG formation, acylating DAG to form TAG. It has been cloned in a diverse range of organisms including *A. thaliana* (Hobbs et al., 1999) and *R. communis* (He et al., 2004). The activity of *B. napus* DAGAT has been reduced through mutation, and under conditions of high oil accumulation such as embryo development a significant build up of DAG intermediates was reported (Katavic et al., 1995) indicating that the conversion of DAG

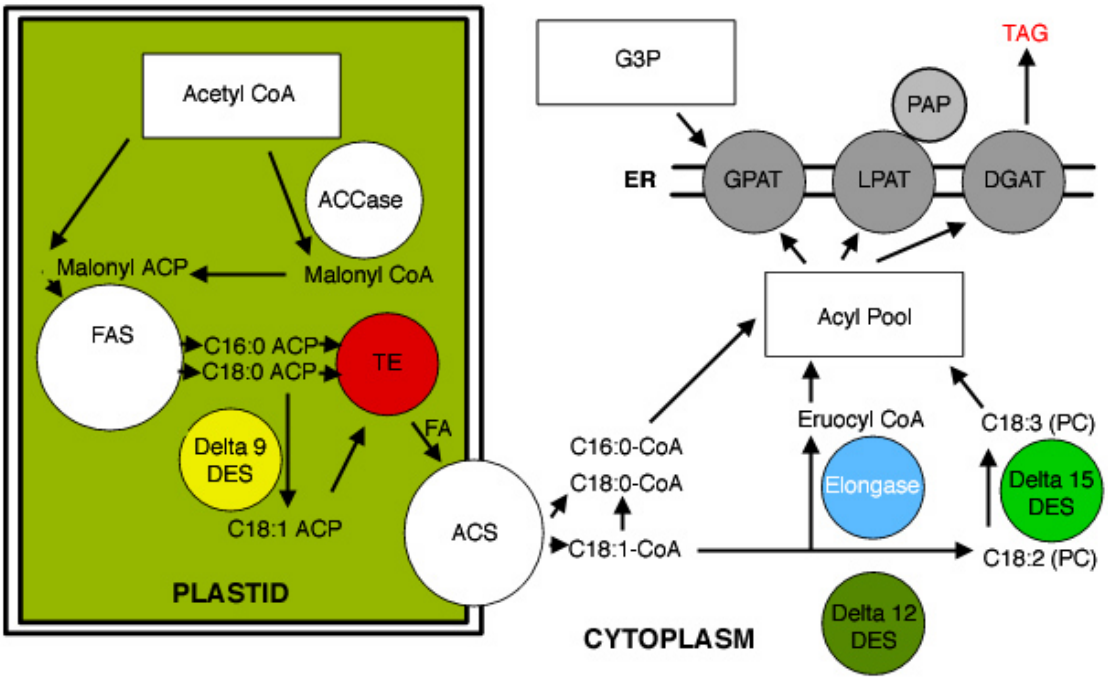


Figure 1.4: Schematic of fatty acid biosynthesis and complex lipid assembly. ACCase = acetyl-CoA carboxylase, ACS = acyl-CoA synthetase, Delta 9 DES = stearoyl-ACP $\triangle$9-desaturase, Delta 12 DES = $\triangle$12-desaturase, Delta 15 DES = $\triangle$15-desaturase, DGAT = diacylglycerol acyltransferase, FA = fatty acid, FAS = fatty acid synthase, G3P = glycerol-3-phosphate, GPAT = glycerol-3-phosphate acyltransferase, LPAT = lysophosphatidic acid acyltransferase, TAG = triacylglycerol, TE = thioesterase, PC = phosphatidylcholine. Figure adapted from Topfer *et al* (1995).

to TAG is highly dependent on DAGAT activity in this plant. This is supported by a study where seed specific over-expression of a DAGAT cDNA resulted in an increase in oil deposition and average seed weight, which correlated with DAGAT transcript levels (Jako et al., 2001). Further recent evidence suggests that, although there exists at least one alternative route to DAG acylation in plants and yeast, the only quantitatively important route to TAG formation in *A. thaliana*, olive and oil palm is through DAGAT (Mhaske et al., 2005; Ståhl et al., 2004; Ramli et al., 2005).

Characterisation of the oil palm DAGAT in microsomes identified substrate selectivity for oleoyl CoA and palmitoyl CoA. Interestingly, the assays identified two discreet peaks of activity which the authors suggested may be due to two isoforms of DAGAT in oil palm (Oo et al., 1989). A new DAGAT gene family was identified in the fungus *Mortierella ramanniana* (Lardizabal et al., 2001), subsequently named DAGAT2 and identified in mouse (Cases et al., 1998) and *S. cerevisiae* (Sorger and Daum, 2002). Kroon et al. (2006), of the laboratory of the author, cloned the DAGAT2 gene in *R. communis*. Studying its expression pattern they found that it is 18-fold more highly expressed in developing seed than in leaf, and shows temporal specific expression during seed development (Kroon et al., 2006). In the same year DAGAT2 was cloned from tung tree (*Vernicia fordii*) (Shockey et al., 2006), which accumulate 80% levels of unusual conjugated fatty acids in their seeds. Similar to the results of Kroon et al. (2006) they identified DAGAT2 as the biochemically significant enzyme in TAG production in seeds, compared to DAGAT1. Interestingly, they also found that DAGAT1 and DAGAT2 localise to different subdomains within the ER. This suggests there are distinct sites in the ER of *V. fordii* that are dedicated to TAG biosynthesis, and raises the possibility that subdomains exist in other plants including *R. communis*. Two lipid biosynthesis enzymes of close functional relationship were shown to have close physical relationship in mouse. DAGAT2 and stearoyl CoA desaturase (SCD1) were found to locate very closely to each other when studied with confocal microscopy,

co-immunoprecipitation and fluorescence resonance energy transfer (Man et al., 2006). Mice lacking SCD1 had a significant decrease in the tissue content of TAG, and the authors hypothesised that substrate channelling occurs between SCD1 and DAGAT2. It is possible that mouse DAGAT2 has a selectivity for monounsaturated acyl-CoAs, and the co-localising SCD1 creates a local pool of monodesaturated CoAs for it to use. A membrane subcompartment with enriched TAG biosynthetic activity has been reported in *B. napus* (Lacey and Hills, 1999). Although the authors described a subcompartment that was ER derived rather than a subdomain within the ER, it supports the model of enzyme distribution heterogeneity within membranes and the existence of enzyme complexes. It is quite feasible that enzyme complexes and substrate channelling mechanisms exist within the ER of *R. communis* which may play an important role in obtaining high triricinolein content in transgenic crops.

1.2.4.2 Acyl-CoA Independent Pathway

Evidence of acyl-CoA independent routes to TAG formation was published in different in papers in 2000. Fraser *et al* (2000) produced *Helianthus annuus* (sunflower) microsomal membranes with accumulated PA due to the inactivation of PAP in the presence of EDTA. PAP requires Mg^{2+} and the metal chelator EDTA chelates free Mg^{2+} thus inactivating PAP. As the incubated microsomal membranes contained inactivated PAP, the usual conversion of PA to DAG was suppressed and so PA accumulated. When these PA-rich microsomal membranes were washed and transferred into a buffer containing Mg^{2+} (reactivating PAP and the conversion of PA to DAG) and either acyl-CoA or no extraneous acyl-CoA molecules, they found no difference in the amount of TAG produced. They proposed that enzymes other than acyl-CoA dependent DAGAT were catalysing the formation of TAG, including a diacylglycerol:diacylglycerol transacylase. They also suggested that acyl-CoA independent routes to TAG formation do have a quantitatively important role in some plants (Fraser et al., 2000).

Dahlqvist *et al* (2000) identified a new acyl-CoA independent route to TAG formation, that uses PC as the acyl donor, and DAG as an acceptor. The enzyme that catalyses the reaction, phospholipid:diacylglycerol acyltransferase (PDAT) was identified in the microsomal preparation from three different oil seeds: *H. annuus*, *Crepis palaestina* and *R. communis*. PDAT activity and the PDAT-encoding gene *LRO1* were identified in the budding yeast *S. cerevisae* which was shown to have homology with animal lecithin : cholesterol acyltransferase (LCAT) (Dahlqvist et al., 2000). PDAT was subsequently identified in *A. thaliana* by homology to the yeast PDAT, sharing 28% identity (Ståhl et al., 2004).

Dalhqvist *et al* (2000) incubated lyophilised microsomal preparations from *H. annuus*, *C. palaestina* and *R. communis* with PC molecules synthesised to contain either [^{14}C]-oleoyl, [^{14}C]-ricinoleoyl or [^{14}C]-vernoloyl acyl residues at the *sn*-2 position. On analysis of the neutral lipid fraction they found although radioactivity incorporation into the neutral lipid fraction increased linearly over 4 hours for each plant species, there were significant differences in the recovery of radio-label between the different plant species and the different acyl residues.

The total amount of radiolabel recovered in the neutral lipid (i.e. removal of labelled acyl residue from the *sn*-2 position of PC to free fatty acid, DAG or TAG) was found to vary significantly between species. *H. annuus* and *C. palaestina* removed similar total amounts of oleoyl, ricinoleoyl and vernoloyl acyl residues from PC to the neutral lipid fraction (between 10 - 20% and 8 - 18% respectively) whilst *R. communis* removed between 30% (microsomes incubated with [^{14}C]-oleoyl-PC) to 60% (microsomes incubated with [^{14}C]-ricinoleoyl PC) of labelled acyl residues to its neutral lipid. This indicates more efficient mechanisms for fatty acid removal from PC in *R. communis* than in *H. annuus* or *C. palaestina*.

The microsomes of *H. annuus*, a species that only has common fatty acids in its seed oil, incorporated most of the label into DAG regardless of the type of acyl group on the

PC in the incubation mixture (see Figure 1.5). A small amount of radioactivity was recovered in the TAG from the ricinoleoyl PC and vernoloyl PC incubated sunflower microsomes, indicating an acyl-CoA independent mechanism to TAG production is utilised in the presence of these two unusual fatty acids in its PC fraction.

This pattern of radioactivity recovery was significantly different from that seen with *R. communis* microsomes. Here, the greatest recovery of [^{14}C]-acyl residues was identified in the TAG fraction, with about 45% recovery of total radioactivity. There was significant preference for the removal of both unusual fatty acids (ricinoleic and vernolic acid) into TAG indicating an efficient acyl-CoA independent route to TAG biosynthesis in this organism. Phospholipase activity was also significant in [^{14}C]-ricinoleic acid removal from the PC in *R. communis*, with $\sim$ 15% of total radioactivity recovered in the FFA. Some recovery of [^{14}C]-ricinoleic acid was identified in DAG, indicating that *R. communis* microsomal CDP-choline:cholinephosphotransferase is also able to use [^{14}C]-ricinoleoyl PC as substrate.

This study indicates significant substrate selectivity differences of the phospholipase, PDAT and cholinephosphotransferase enzymes in the microsomes of the different plant species assayed. It also suggests a significant role for PDAT in *R. communis* and, along with the phospholipase, make these enzymes likely useful components in engineering high triricinoleate *B. napus* seed oil.

Recent evaluation of PDAT mutants in *A. thaliana* have found PDAT to be quantitatively insignificant to overall TAG production in this organism (Mhaske et al., 2005). There appears to be some redundancy in the routes available for transiting fatty acids out of PC. It is plausible that PDAT evolved as an enzyme capable of nullifying the effects of potentially toxic unusual fatty acids by efficiently removing them from membrane lipid into TAG; where unusual fatty acids conveyed benefits to the plant, such as the emetic qualities of castor oil which would likely discourage animal predation, the PDAT enzyme has become highly tuned to the transfer of the unusual

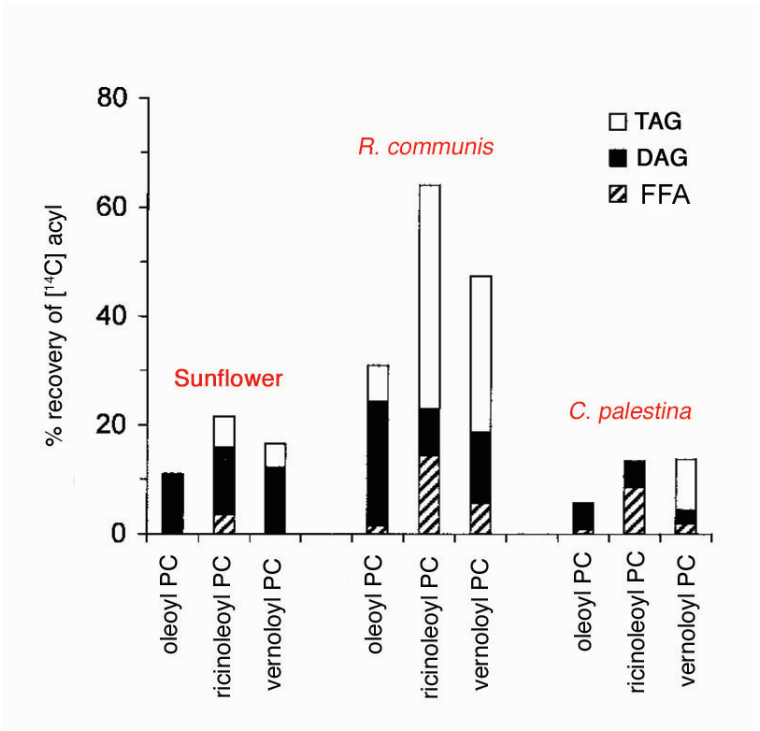


Figure 1.5: Metabolism of ^{14}C-labelled PC into FFA, DAG and TAG in plant microsomes. Microsomes from sunflower, *R. communis* and *C. palestina* were incubated with ^{14}C-oleoyl PC, ^{14}C-ricinoleoyl PC or ^{14}C-vernoleoyl PC. After incubation, the recovery of radioactivity in either free fatty acid (FFA), diacylglycerol (DAG) or triacylglycerol (TAG) was quantified for each combination of microsome and labelled PC. The data indicates significant differences in metabolism of unusual fatty acids in purified microsomes from the different plant species. In *R. communis* an acyl-CoA independent route to the removal of ricinoleic or vernoleic acid was quantitatively the most significant, with 45% of radioactivity recovered in the TAG fraction. The action of phospholipase was also significant in *R. communis* with ∼15% radioactivity recovered in FFA. Figure adapted from Dhalqvist et al. (2000).

fatty acid. The diversity of membrane lipid is likely to be greater in yeast cells that can assimilate a range of fatty acids for food. The ability of a yeast cell to efficiently transfer unusual and potentially toxic lipid species incorporated into its membrane from the environment into neutral lipid stores by the action of PDAT would give it a distinct evolutionary advantage.

An alternative acyl-CoA independent mechanism of TAG production is the formation of TAG from two molecules of DAG in a DAG:DAG transacylation (Stobart et al., 1997). The importance of DAG:DAG transacylation in the formation of TAG in *R. communis* was evaluated by Dalhqvist et al. (2000). The group incubated *R. communis* microsomes and an exogenous DAG molecule to act as an acceptor were incubated with *sn*-1 oleoyl *sn*-2 [^{14}C]-ricinoleoyl DAG and the formation of TAG examined. The experiment was repeated but this time the microsomes and exogenous DAG acceptor were incubated with *sn*-1 oleoyl *sn*-2 [^{14}C]-ricinoleoyl PC. Microsomes incubated with [^{14}C]-PC produced a significantly higher amount of [^{14}C]-TAG than microsomes incubated with [^{14}C]-DAG, indicating the likelihood that PC plays the greater role in TAG formation via acyl-CoA independent routes.

1.3 Production of Unusual Seed Oils in Agronomic Plants

Attempts to engineer model or crop plants that produce triacylglycerols containing unusual fatty acids at high levels have so far met only limited success. Divergent FAD2 (fatty acid desaturase) enzymes modify acyl groups attached to PC to produce unusual fatty acids, and have been the focus of transgenic efforts to produce industrially valuable oils in crop plants (reviewed in Cahoon and Kinney (2004)). FAD2 catalyses the introduction of a second double bond at the 12-carbon position of oleoyl groups esterfied to PC. Divergent enzymes have been identified in a variety of plants that catalyse diverse reactions such as hydroxylation, epoxygenation and double bond conjugation

(Cahoon and Kinney, 2005). Two examples of FAD2-derived enzymes are the oleate $\triangle 12$ hydroxylase , which introduces a hydroxyl group at the 12-carbon position of oleoyl PC (Bafor et al., 1991) and was identified in the seed ER of *R. communis* (Maltman et al., 2002), and a conjugase from pot marigold (*Calendula officinalis*).

The *R. communis* oleate $\triangle 12$ hydroxylase has been expressed in *A. thaliana* under the control of a *B. napus* seed specific promoter (Broun and Somerville, 1997). This resulted in the accumulation of 17% ricinoleic acid in the seed fatty acid along with two novel hydroxy-fatty acids which were identified as lesquerolic (C20:1-OH) and densipolic acid (C18:2-OH). Traces of auricolic acid (C20:2-OH) were also observed. It was hypothesised that the presence of lesquerolic and densipolic acid along with ricinoleic acid was either due to the action of oleate $\triangle 12$ hydroxylase on *sn*-2 oleoyl-PC (18:1-PC) and *sn*-2 eicosenoyl-PC (20:1-PC), or it was due to the presence of an elongase capable of using 18:1-OH as a substrate. Smith et al. (2003) were able to confirm it was due to the action of an elongase by expressing *R. communis* oleate $\triangle 12$ hydroxylase in an *A. thaliana* knockout lacking *FAE1*, an *A. thaliana* elongase (Smith et al., 2003). They found that lesquerolic acid (C20:1-OH) and auricolic acid (C20:2-OH) were entirely absent from this plant, the only hydroxylated fatty acids present were 18 carbon species. The ER resident desaturase *FAD3* catalyses the desaturation of C18:$1^{(\triangle 9)}$ to C18:$2^{(\triangle 9,15)}$ (Browse et al., 1992). Smith et al. (2003) expressed the *R. communis* oleate $\triangle 12$ hydroxylase gene in an *A. thaliana* dual mutant line lacking both FAD3 desaturase and FAE1 elongase activity. This transformant produced ricinoleic acid as the only hydroxy fatty acid, i.e. densipolic (18:2-OH) and auricolic (20:2-OH) fatty acids were absent, but at only 7% of the total seed oil. The maximum amount of total hydroxy fatty acids they obtained in the seed oil was 19.2%, of which 8.7% was ricinoleic acid. The maximum amount of ricinoleic acid obtained was 16.2%, in a *FAD3* knockout line producing a total of 18.7% hydroxy fatty acid in the seed oil. These levels are far less that the 90% ricinoleic acid present in the seed oil of *R. communis* however.

Broun and Somerville's (1997) initial work in expressing the oleate △12 hydroxylase in *A. thaliana* raised an interesting observation: a dramatic increase in oleic acid and a corresponding reduction in polyunsaturated fatty acids. The presence of the oleate △12 hydroxylase seemed to have an inhibitory effect on the oleate △12 desaturase. They were unable to find evidence of homology-dependent gene silencing as there was no decrease in mRNA levels for oleate △12 desaturase in the transformed plants; this was supported by extensive investigations reported in a further paper by the group (Singh et al., 2001). They concluded that there is a post-translational inhibition of oleate △12 desaturase in oleate △12 hydroxylase transformants. Smith et al. (2003) also noted a significant increase in oleic acid and a corresponding decrease of linoleic acid (C18:2) in their *R. communis* oleate △12 hydroxylase transformed *A. thaliana* lines. They identified a clear positive linear correlation between the percentage of hydroxylated fatty acids and the percentage of oleic acid in the TAG. Conversely, a negative linear correlation between the percentage of hydroxylated fatty acids and the percentage of linoleic acid in TAG was reported. This indicates a relationship between *sn*-2 oleoyl-PC modification by oleate △12 hydroxylase and oleate △ desaturase, and transit of oleoyl CoA to the TAG, either via the PC membrane fraction or through direct acylation. It was hypothesised that the increase in oleic acid, decrease in FAD2-desaturated fatty acids and low percentages of hydroxy fatty acids in the TAG of oleate △12 hydroxylase transformants was due to the build up of unusual fatty acids in the PC membrane fraction. If the transformed plant is unable to efficiently remove unusual fatty acids from the PC, they may begin to accumulate resulting in reduced oleoyl-PC substrate for FAD2 oleate △ desaturase. As the PC fraction becomes a bottleneck due to reduced flux of hydroxy fatty acids out of the PC, the availability of PC substrate with free *sn*-2 positions for the transfer of oleoyl CoA catalysed by lysophosphatidylcholine acyltransferase (LPCAT) is reduced. Similarly, the PC $\rightleftharpoons$ DAG transfer route catalysed by reversible CDP-choline:cholinephosphotransferase route may be effected.

This hypothesis is supported by an earlier analysis of *A. thaliana* expressing genes encoding the divergent FAD2 enzymes *Crepis alpine* $\triangle$12-acetylenase, a *C. palaestina* linoleate $\triangle$12 epoxygenase or *L. fendleri* oleate $\triangle$12 hydroxylase. The transformants were found to contain significant amounts of the unusual fatty acids (crepenynic, vernolic and ricinoleic acid respectively) in their PC membrane lipid fractions during TAG accumulation and 5-35 times less crepenynic, vernolic and ricinoleic acid in the TAG of the *A. thaliana* transformants than is found in the natural sources of these unusual fatty acids (Thomaeus et al., 2001). Further evidence for this hypothesis came in a recent study where the FAD2 conjugase enzyme from *C. officinalis* was transferred to *G. max* and *A. thaliana* , which were found to produce a seed oil containing up to 20% unusual conjugated fatty acid calendic acid (C18:3$\triangle^{8trans,10trans,12cis}$) (Cahoon et al., 2006). This compares to the 55% of calendic acid produced in the natural source, the seed oil of *C. officinalis*. However, on quantification of the fatty acid components of the PC membrane lipid fraction, they observed conjugated fatty acids that were at least equal to those sequestered in the TAG fraction. Analysis of the PC fraction of the seeds from 5 varieties of *C. officinalis* that naturally produce high calendic acid seed oil, the percentage of conjugated fatty acids observed was less that 1.5% of total fatty acids (Cahoon et al., 2006).

The known routes of fatty acid removal from PC are shown in Figure 1.6. Understanding the respective contributions of these enzymes (i.e. reversible LPCAT, phospholipase A_2, reversible cholinephosphotransferase and PDAT) to the flux of fatty acids out of PC in *R. communis* and *B. napus* is likely to be a crucial step in the pathway toward engineering a crop plant that produces high levels of ricinoleic acid in its TAG.

Engineering a crop plant such as *B. napus* that is capable of first producing unusual fatty acids and then removing them from the PC is unlikely to produce a seed oil as high in the unusual fatty acid as the natural source of that oil. As discussed

in Sections 1.2.3 and 1.2.4, the differing substrate selectivities of the plant Kennedy pathway acyl transferase enzymes introduce significant heterogeneity of acyl groups in TAG molecules. Understanding the contribution of the different *R. communis* acyltransferase enzymes in catalysing the production of high triricinoleate castor oil and transferring the appropriate complement of enzymes to the engineered plant is likely to be crucial. This is highlighted in *A. thaliana* seeds expressing *R. communis* oleate $\triangle 12$ hydroxylase, which display futile cycling of ricinoleic acid through peroxisomal β-oxidation (Moire et al., 2004). This is possibly caused by the build up of ricinoleoyl CoA, due to a bottle neck in the flux of the unusual CoAs from the CoA pool to TAG.

1.4 Emerging Evidence of Lipid Biosynthesis Complexity in Plants

There is growing evidence of significant lipid biosynthetic gene heterogeneity in plants. Two potentially important enzymes in directing the transit of unusual fatty acids from PC and into the acyl-CoA pool are phospholipase A_2 and acyl-CoA synthetase respectively. In a 2003 census of an *A. thaliana* EST database a large family of phospholipase-encoding genes were identified (Beisson et al., 2003) and plants are known to contain very large families of acyl-CoA synthetases (Shockey et al., 2002; Shockey et al., 2003). Experimental evidence suggested *A. thaliana* has 5 isoforms of LPAT (Kim et al., 2005). Based on homology searches, *A. thaliana* may contain up to 11 LPAT genes (Beisson et al., 2003). Post-genomic technologies such as microarrays or proteomics allows the identification of transcribed or translated genes to be identified in a particular tissue and at particular time points. Proteomic analysis of the ER of the developing seed of *R. communis* and at the correct developing seed should allow identification of the translated gene products involved in complex lipid biosynthesis, so identifying the biochemically relevant enzymes. Another important method by which

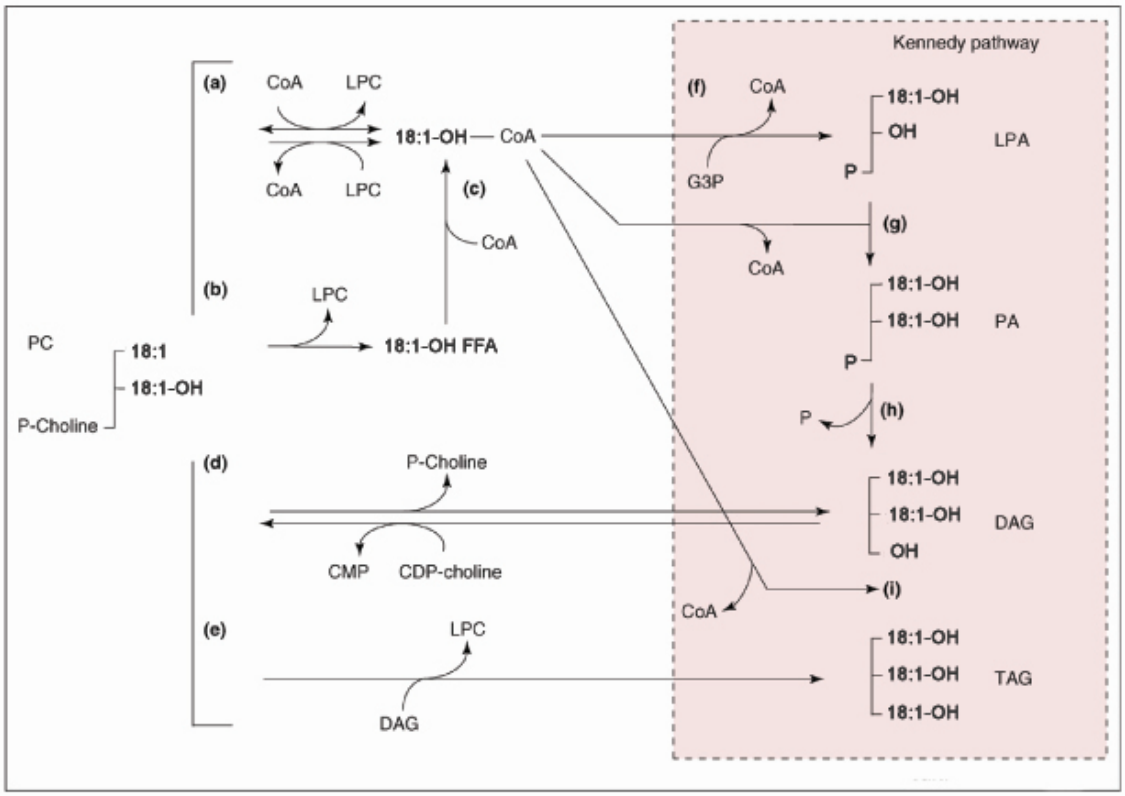


Figure 1.6: Potential pathways of hydroxy fatty acid flux from the PC to TAG. Multiple routes exist for the removal of fatty acids from PC. An *sn*-2 ricinoleoyl *sn*-1 oleoyl PC molecule is depicted at the left of the figure. Pathway **(a)** is catalysed by reversible lysophosphatidylcholine acyltransferase (LPCAT) which transfers an acyl group to CoA, producing acyl-CoA and lysophosphatidylcholine (LPC). Pathway **(b)** depicts phospholipase A_2, which releases free fatty acid (18:1-OH FFA) from PC substrates leaving LPC. Free fatty acids such as 18:1-OH FFA can be activated by acyl-CoA synthetases (ACS), denoted by **(c)** in the figure. Pathway **(d)** is catalysed by reversible CDP-choline:cholinephosphotransferase which removes the phosphocholine group from PC to yield DAG. Phosphatidylcholine acyltransferase (PDAT) catalyses the direct transfer of a fatty acid from PC to DAG in an acyl-CoA independent route to TAG formation **(e)**. Reaction **(f)** is catalysed by the Kennedy pathway enzyme glycerol-3-phosphate acyltransferase (GPAT), which utilises glycerol-3-phosphate (G3P) and an acyl-CoA to form lysophosphatidic acid (LPA). A further acylation is catalysed by lysophosphatidic acid acyltransferase (LPAT, **(g)**) which again uses acyl-CoA. The removal of the *sn*-3 phosphate group, catalyses by phosphatidic acid phosphatase (PAP, **(h)**) is required to form diacylglycerol (DAG) which then serves as a substrate for acyl-CoA utilising diacylglycerol acyltransferase (DAGAT, **(i)** or PDAT. Figure adapted from Cahoon et al. (2007)

the most important isoforms of enzymes in triricinolein production can be identified is by assay, which is discussed in the next section.

1.5 Strategies for identifying and validating the components of high triricinolein production in *R. communis*

There are two distinct components of research presented in this thesis. The first concerns the use of proteomics to identify components of lipid biosynthesis in *R. communis* whilst the second concerns the development of methodologies for the yeast-based *in vivo* assay of *R. communis* lipid biosynthesis components identified in the proteomic component of the work. These two distinct components are described in more detail in the following two sections (Sections 1.5.1 and 1.5.2), followed by an exploration of the proteomic technologies used in the study (Section 1.6.1) and background to yeast lipid biochemistry (Section 1.8).

1.5.1 The ER as the Target Organelle for Proteomic Analysis

As has been described in the preceeding section which reviewed the biochemistry of lipid and triricinolein production in *R. communis*, all the components of hydroxylation and complex lipid assembly are located in the endoplasmic reticulum of developing seed. It is known that the developing seed is actively involved in storage oil production and in *R. communis* key components of fatty acid and storage oil biosynthesis are at their peak between 25 and 30 days after flowering (DAF)(Simcox et al., 1979; Kroon et al., 2006). Therefore, by identifying the proteins of the developing seed ER at this optimum stage (25-30 DAF) the secrets of triricinolein production may be unlocked. A dual approach has been taken by the laboratory for the proteomic analysis of the developing ER: one focused on the soluble fraction and utilising 2D electrophoresis (2DE) as the primary analysis technology; the other focused on the membrane-bound components

and utilising mass spectrometry (MS) approaches such as isotope coded affinity tags (ICAT). The nature of the author's industrial funding limited his investigation to the 2DE analysis of the ER and it is this analysis which is presented in Chapters 3 and 4. Technologies relevant to the soluble proteomic analysis of this thesis are reviewed in Section 1.6.1.

The majority of known complex lipid biosynthesis components are membrane bound (for example, DAGAT2 (Kroon et al., 2006)) and thus are very unlikely to be soluble in 2DE lysis buffer and present in a 2DE analysis. However, as a method of identifying components of complex lipid biosynthesis, 2DE is still valid for two reasons: firstly, it may identify previously unknown components of complex lipid biosynthesis which are 2DE buffer soluble. Secondly, there are already known components of complex lipid biosynthesis which are only peripherally associated with the membrane (for example, PAP (Pearce and Slabas, 1998)) and this and other peripherally associated components may be identified on the 2DE gels.

1.5.2 *Yarrowia lipolytica* as a Potential *R. communis* Lipid Gene Assay Vehicle

The previous section explored lipid biosynthesis in plants, with particular reference to *R. communis* and the biosynthesis of high levels of triricinolein. It also reported emerging evidence of problems with flux of unusual fatty acids out of the PC into TAG in transformed agronomic plants and large heterogeneity of genes for a single lipid biosynthesis enzymatic step. This points to the need for sophisticated tools to not just identify the components of lipid biosynthesis in *R. communis* but also to understand the set of proteins required to obtain high levels of triricinolein production in transformed plant models. The ability to transform multiple *R. communis* genes into an *in-vivo* model allows an iterative examination of their influence on lipid biosynthesis until ultimately all the necessary components for high levels of ricinoleic acid biosynthesis

have been built into the foreign host. Such an *in vivo* host might be a plant model, for example *A. thaliana* or *B. napus*. Although this allows direct transformation of *R. communis* genes into an appropriate agronomic organism (in this case, *B. napus*), it has the disadvantage of an extended period of time between transformation and evaluation of the effect of the transgene on the lipid profile (typically 2 to 3 months for the development of mature seed). The use of microorganisms, such as yeast, provide an attractive alternative. Yeast strains which produce large amounts of storage oil (oleaginous yeasts) have similar lipid biosynthesis biochemistry to plants but have the advantage that they are able to grow rapidly. Thus they may provide an invaluable research tool for the assay of *R. communis* lipid biochemistry genes. *Y. lipolytica* was identified as a potential yeast strain for this purpose.

Y. lipolytica is a dimorphic (that is, it exists in a unicellular yeast state or a multicellular hyphae state) ascomycete. It is an obligate aerobe, providing an easy means for its removal from foodstuffs (McKay, 1992), and is considered non-pathogenic, in part because its growth temperature seldom exceeds 32-34 °C (Barth and Gaillardin, 1996). It is oleaginous, making it a potential system for the assay of *R. communis* lipid genes as it is well adapted to producing and storing lipid. It has the key requirements of having well-established molecular biology tools in place and having its genome sequenced (Dujon et al., 2004) and publicly available. The molecular biology tools include a range of auxotrophic strains available, and protocols for transformation (including homologous recombination) (Barth and Gaillardin, 1996). The availability of the genome sequence means it is possible to utilise homologous recomination to directly target *R. communis* lipid genes into specific areas of the *Y. lipolytica* genome. This allows the researcher to replace an existing gene with an equivalent *R. communis* transgene. It has a rapid growth rate with strains reaching lipid-accumulating stationary phase within 20-30 hours of culture inoculation (Mlickova et al., 2004). It is able to grow on standard media such as yeast nitrogen base (YNB) or yeast extract / peptone / dextrose (YPD)

(Barth and Gaillardin, 1996). A significant lipid biochemistry research base exists for the organism although this is mostly focused on its efficient lipid catabolism capability (Waché et al., 2000; Waché et al., 2001; Waché et al., 2003; Fickers et al., 2005). It is capable of growing on hydrophobic substrates, such as alkanes and fatty acids, including ricinoleic acid.

Once the growth, lipid extraction and lipid analysis procedures are in place, the components of the storage oil in wildtype cells can then be evaluated which can serve as a baseline by which the TAG analyses of transformed strains can be compared. Such analyses would be performed on *Y. lipolytica* cells fed on non-hydroxy (e.g. oleic acid) or hydroxy (e.g. ricinoleic acid) lipid to understand the extent of which wildtype cells can incorporate ricinoleic acid into their TAG.

A review of the literature concerning lipid biosynthesis in yeast and specifically *Y. lipolytica* is presented in the Section 1.8.1.

1.6 Proteomic Analysis of the Soluble *R. communis* ER

1.6.1 Two dimensional Electrophoresis (2DE)

1.6.1.1 Sample preparation

No single method of sample preparation can be applied universally to biological samples because of their diverse nature. However, the desired end result is generally the same: a sample of solubilised peptides, at a concentration appropriate for proteomic analysis and lacking molecular compounds that interfere with the 2DE process. It can be thought of as a two step process, (1) biological material preparation and (2) sample solubilisation. These processes may occur separately or at the same time.

1.6.1.2 Biological material preparation

Soluble, aqueous samples such as serum, plasma or cerebrospinal fluid are often analysed with comparatively little pre-treatment. Soluble samples containing high concentrations of protein may be diluted with the appropriate solubilisation buffer (see Section 1.6.1.3), but dilute samples may require concentrating through lyophilisation or precipitation. Trichloroacetic acid with acetone is a precipitation method commonly used for this purpose and is also useful in the inactivation of proteases and removal of isoelectric focusing (IEF)-interfering compounds such as salts, lipids and nucleotides (Görg et al., 1997).

Where protein needs to be solubilised from solid tissue, physical disruption of the material is often required. Freezing samples, including at liquid N_2 temperatures, often aids this process. Freezing can disrupt the integrity of subcellular components, altering the buoyant densities of organelles. Therefore, it is not a suitable option when intact organelles are required, as is the case in this study.

Fractionation of biological material may be employed at this step. It can be useful in reducing gel profile complexity, or in targeting the biologically relevant subsection of the tissue or cells. The disadvantage of subcellular fractionation is you may selectively lose biochemically important components or subfractions of an organelle and miss significant cellular machinery by focusing on a subset proteome. Therefore, sample preparation methodologies are a balance between reducing complexity of the resultant 2DE profile and the risk of selective samples losses resulting in an incomplete organelle proteome.

1.6.1.3 Sample solubilisation

The aim of sample solubilisation is to break the inter- and intra-molecular hydrogen and disulphide bonds in the protein sample, and to maintain these breakages through the 2DE process. Samples are solubilised in a combination of chaotropes, surfactants and reducing agents. In the first description of the 2DE technique (O'Farrel, 1975), very

high concentrations (9.5 M) of the chaotrope urea, combined with 4% (w/v) of the non-ionic surfactant NP40, 2% (w/v) synthetic carrier ampholytes (SCA) and 1% (w/v) of the thiol-containing reducing agent DTT were used. Urea overcomes hydrogen bonding within the protein molecules leading to unfolding and denaturation. After denaturation the hydrophobic domains of the amino acid monomers are exposed. This risks protein loss through aggregation and adsorption (Molloy, 2000). The surfactant interacts with the exposed hydrophobic domains and prevents this. DTT breaks the intra- and intermolecular disulphide protein bonds and carrier ampholytes are beneficial in reducing protein interactions within the IEF strip matrix. Improvements to this original procedure were made with the replacement of the surfactant with 2% CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate), which was found to improve membrane protein solubilisation compared to NP40 (Perdew et al., 1983; Santoni et al., 2000), and the additional chaotrope thiourea at 2 M concentration (Rabilloud et al., 1997; Molloy et al., 1998). Also crucial was the development of immobilised pH gradients (IPGs) (Bjellqvist et al., 1982), which did not suffer the difficulties of SCAs in tube gels such as cathodic drift resulting in pH gradient instability (see Section 1.6.1.4).

Problems with 2DE remain however. Certain proteins can be under-represented on gels or refuse to solubilise in the urea buffer. The most elusive proteins in 2DE gel analysis are membrane bound proteins. By their nature these proteins have very hydrophobic domains and are either insoluble in 2DE lysis buffer or are lost during IEF. This is of particular concern to the characterisation of ER proteins concerned with fatty acid modification and the Kennedy pathway as the majority of the known proteins are membrane bound. For proteomic projects in general, characterisation of membrane proteins is of great biological and pharmacological importance as they represent the 'doorbells and doorways' of the cell (Molloy, 2000). The analysis of membrane proteins can be aided by fractionation of the sample. Molloy et al. (1998) extracted *E. coli* proteins sequentially, first highly soluble proteins were extracted in Tris-base, followed

by extraction of more hydrophobic proteins using classical lysis buffer (8 M urea, 4% CHAPS, DTT and carrier ampholytes) and finally very hydrophobic proteins using thiourea, tributyl phosphine and sulphobentaines. The protein sample was separated across three gels revealing more proteins than is possible on a single gel using standard lysis buffer. Although there was overlap across the three gels, the final gel revealed several abundant protein spots that were absent on the single gel method, subsequent identification revealing them to be integral outer membrane proteins (Molloy et al., 1998). A further method for improving the solubilisation of membrane proteins for 2DE relies on the use of different detergents supplementing or replacing the typical detergent CHAPS. Molloy et al. (2000) investigated the solubilisation of *E. coli* outer membrane proteins (OMPs) with altered detergent lysis buffers. Sodium carbonate-isolated membranes were solubilised with 7M urea, 2M thiourea, 1% w/v ASB14, 40 mM Tris, 2 mM tributyl phosphine and 2% ampholytes. On analysis of the resultant 2DE gel, 80% of the *E. coli* SWISSPROT database predicated OMPs were represented on the gel. However, this group did not compare the modified ASB14 lysis buffer with the standard 9 M urea, 2 M thiourea, 4% CHAPS lysis buffer so actual improvements in the solubilisation of E. coli OMPs with this method could not be made. An alternative lysis buffer containing the detergent C8Ø was found to effectively solubilise plant plasma membrane proteins including H^+-ATPase proteins and water channel proteins (Chevallet et al., 1998). These proteins were not seen when they attempted to solubilise the same preparation with a standard 9 M urea, 2 M thiourea and 4% CHAPS lysis buffer.

As MS technologies have advanced and new methods of quantifying complex mixtures of proteins including membrane proteins have been developed, the requirement to effectively solubilise and separate membrane proteins by 2DE has reduced (see Section 1.6.4).

1.6.1.4 First dimension Isoelectric Focusing (IEF)

Isoelectric focusing is based on the principle that at a given peptide's isoelectric point (pI), whereby its net charge is zero, it ceases to migrate within an electric field. In a heterogeneous protein sample, a range of isoelectric points will be represented, each the sum of the charge profile of the peptides' amino acid constituents.

Initial IEF procedures relied on small molecule synthetic carrier ampholytes (SCAs), each molecule having a different pK value (O'Farrel, 1975). Carrier ampholytes were added to an acrylamide mixture and after polymerisation subjected to a pre-IEF separation to provide a linear distribution of SCAs within the acrylamide matrix. However, this technique is prone to pH gradient instabilities as the SCAs readily migrate by a process of electroendosmosis to the cathode (cathodic drift). A levelling off of pH gradient linearity can also occur in the central region of the IEF strip (plateau effect). These problems, combined with the batch to batch variability of SCAs, cause significant problems in establishing reproducibility of gel profiles and the ability to compare profiles between laboratories. An alternative procedure, termed non-equilibrium pH gradient electrophoresis (NEPHGE), was developed to improve upon these difficulties in the separation of basic proteins (O'Farrel et al., 1977). Reproducibility between gels remains a significant problem however.

These difficulties were overcome with the development of pH buffering molecules immobilised to the acrylamide matrix, termed Immobilised pH Gradients (IPGs) (Bjellqvist et al., 1982). The technology has developed into the first choice for 2DE based proteomic studies (Görg et al., 1998a; Görg et al., 2000). IPG IEF gels are prepared using Immobilines (GE Healthcare, Bucks UK), acrylamide derived molecules with the common formula:

$$CH_2{=}CH{-}CO{-}NH{-}R$$

where the R group is either a carboxyl (weakly acidic) or one of 7 tertiary amino

groups (basic). Immobilines are added to the IEF strip mixture in much the same way as SCAs but are covalently attached to the polyacrylamide backbone. This linkage prevents the effect of electroendosmosis. Strips are generally prepared upon an acetate backing for ease of handling before being cut into thin strips. This method has allowed the generation of broad, narrow and custom pH ranges (Görg et al., 2000) including linear and non-linear pH distributions (Bjellqvist et al., 1993).

Choice of IPG strip is often influenced by the familiarity with the sample. Initial proteomic investigations into a sample may favour a broad range (e.g. pH 3-10) IPG strip allowing a wide examination of protein pI distribution. As broad range IPG strips can reduce resolution of proteins, especially in the pH 4-7 range where many proteins' pIs occur, if spots of interest do not fall outside this narrower range a pH 4-7 IPG strip makes a more appropriate choice. Alternatively, 'zoom gels', multiple overlapping gels of narrow pH ranges, can be used to allow maximum separation of protein and help to overcome spatial limitations of broad range IPG strips with complex proteomes (Wildgruber et al., 2000).

There are multiple methods of protein sample introduction into the IPG strip. Paper bridge loading, whereby the protein sample is absorbed into the paper wick used to connect the electrode arm to the IPG strip, allows the sample to be drawn from the wick into the IPG strip when electrical current is passed through the strip (Sabounchi-Schutt et al., 2000). Alternatively, a small plastic cup is placed upon the surface of the IPG strip into which the protein sample, solubilised in lysis buffer, is pipetted. A small window within the cup allows the passage of the protein sample into the IPG strip when a current is passed across it (Görg et al., 1998b). A third method of sample introduction is termed in-gel rehydration. With this method a desiccated IPG strip is rehydrated with lysis buffer containing the protein sample. It has the advantage of allowing much larger volumes of protein sample containing-lysis buffer to enter the strip, useful for 2DE analysis of dilute protein samples. The alternative methods of cup

/ paper bridge loading and in-gel rehydration are compared in the schematic diagram Figure 1.7.

1.6.1.5 IPG Strip Equilibration

Following IEF, IPG strips must be equilibrated before second dimension electrophoresis separation. Equilibration can be performed immediately following IEF or the IPG strips can be frozen for up to 2 months. Equilibration is a two step process in which the IPG strip is first incubated for 15 minutes in a buffer containing 50 mM Tris pH 8.8, 2% (w/v) SDS, 30% (v/v) glycerol and 1% (w/v) DTT. The presence of the SDS is critical in incorporating a charge to the pI separated proteins, allowing their migration through the second dimension gel on application of an electric current. The glycerol reduces electroendosmotic effects which would otherwise reduce transfer from the IPG strip into the second dimension gel (Görg et al., 1998a). DTT is used to prevent re-oxidation of cysteine residues and the formation of disulphide bonds. Following the first incubation the IPG strip is incubated for a further 15 minutes in a new buffer identical to first except it contains 4.8% iodoacetamide instead of DTT. Iodoacetamide alkylates any remaining free DTT associated with the strip which otherwise leads to point streaking in the second dimension (Dunn and Görg, 2001).

1.6.1.6 Second Dimension Electrophoresis

Strips are separated in the second dimension by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE), utilising a discontinuous buffer system (Laemmli, 1970a). Gels that are either a single acrylamide concentration (for example, 12%) or are either linear or non linear gradients, allow control of migration of proteins in the second dimension.

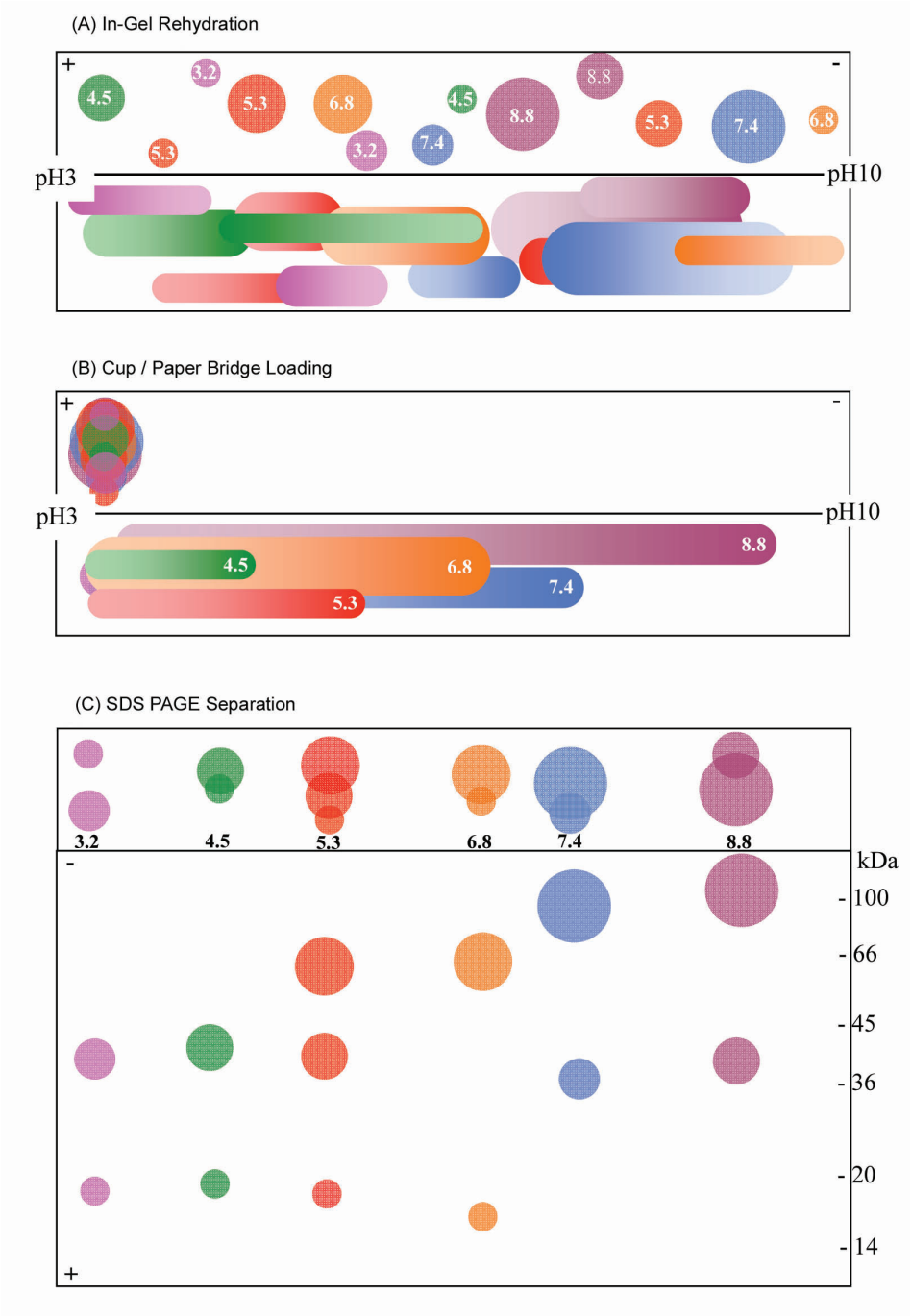


Figure 1.7: Principle of IEF and SDS PAGE separation of proteins. Two methods of sample introduction are in-gel rehydration **(A)** and cup / paper bridge loading **(B)**. On application of a current across the IPG strip during IEF proteins migrate to their isoelectric point (where they have no overall charge). The pattern of migration is dependent on where the proteins were located prior to IEF. Following IEF proteins are separated by molecular weight in an SDS PAGE gel **(C)**, with smaller proteins travelling further through the acrylamide matrix than larger proteins. Figure adapted from (Hall, 2004).

1.6.1.7 Gel Visualisation

Perhaps the most widely used method in visualising proteins in gels is the silver stain. This stain offers a high level of sensitivity of between 1-10 ng/SDS PAGE protein band although it cannot offer accurate quantification. As a photographic stain it suffers from saturation and also its linearity is protein dependent. Silver stain does offer a protein concentration / spot volume linearity over a 40 - 50 fold concentration range from 0.04 ng/mm^2 to 2.0 ng/mm^2. However, above this limit the intensity of stain to protein abundance becomes non-linear as the staining reaches saturation (Righetti, 1990) and quantification breaks down. Silver stain is also incompatible with mass spectrometry unless the procedure is modified by leaving out the protein cross-linker glutaraldehyde (Shevchenko et al., 1996a). Unfortunately modified silver stains suffer from reduced sensitivity. Silver staining is also more labour intensive than other stains, requiring a number of solutions and hours. It also requires the presence of a skilled person to monitor and finally stop the development of the stain. This can be especially problematic with multiple gels. However, for a highly sensitive qualitative investigation of a protein sample that can be visualised without the aid of transilluminators or laser scanners it is still very useful.

Another classical and widely used visible stain is Commassie brilliant blue. This stain is simple and quick to use and is compatible with down-stream mass spectrometry analysis. It does however suffer from low sensitivity (between 30-100 ng/SDS PAGE protein band). It should be noted that Tal et al. (1985) reported that the response of the dye to protein abundance, although linear, is protein dependent governed by the percentage of basic amino acids in the protein in question. The group published protein/abundance curves showing significant variation depending on what protein had been stained (Tal et al., 1985).

In recent years the use of fluorescent stains has become more widespread. This has been aided by the development of more sensitive and easy-to-use stains such as

SYPRO Ruby and the improvement of imaging technologies and software. Steinberg et al. (1996a) first reported the development of the fluorescent SYPRO Red and SYPRO Orange dyes. They described these new stains as having a level of sensitivity comparable to silver stain (i.e. 1-10 ng/SDS PAGE protein band) but without the lengthy staining procedure of silver stain or the incompatibilities of silver with mass spectrometry (Steinberg et al., 1996a; Steinberg et al., 1996b). SYPRO Red and SYPRO Orange fluorescent stains interact with the sodium dodecyl sulphate (SDS)-protein complex and thus do not require prior fixing. This has the benefits of ease of use and allows the visualisation of gels prior to Western blotting procedures. The disadvantage of not having a fixing step is that after a period of time an image can no longer be captured as the proteins, especially low molecular weight and less abundant proteins, would diffuse out of the gel. SYPRO Ruby stain is a commercial metal-chelate stain which offers a broad linear dynamic range not seen in silver or Coomassie stains (Lopez et al., 2000). Unlike unmodified silver staining it is also compatible with mass spectrometric analysis. A drawback of SYPRO Ruby was reported by Chevalier et al. (2004) who observed precipitants collecting on the surface of SYPRO stained gels leading to intense spikes of noise effecting quantification (Chevalier et al., 2004). Where these spikes fall in the same region as a protein spot the contributing fluorescence of the spike would be included in the spot abundance for that protein, thus overestimating its actual abundance. Unlike SYPRO Red or Orange, SYPRO Ruby is compatible with 40% (v/v) ethanol / 10% (v/v) acetic acid fixation suggesting the stain does not interact with SDS in the same way that SYPRO Red or Orange does. This means that SYPRO Ruby gels can be kept indefinitely, and can be restained if required. This is important for preparative gels where lengthy analysis might have to be carried out prior to being in a position to excise protein spots for mass spectrometry.

Deep Purple (GE Healthcare, Bucks, UK) commercialises the fluorescent dye epicocconone from the fungus *Epicoccum nigrum*. This dye was found to have a sensitivity

down to 2 ng/SDS PAGE protein band (Bell and Karuso, 2003), a similar sensitivity to SYPRO Ruby. Chevalier et al. (2004) examined Deep Purples application to proteomics and found it to not suffer from the background noise of SYPRO Ruby protein stains. They also investigated its sensitivity in comparison to silver, Coomassie brilliant blue and SYPRO Ruby and found SYPRO Ruby, silver and Deep Purple to have comparable sensitivity. They investigated the linearity of response of 1D band volume to concentration for both a high abundance and low abundance band. They found a linear relationship for Coomassie and SYPRO Ruby but as expected a high propensity for saturation for the higher abundance spots with silver stain. They also observed saturation with Deep Purple which was not present at that protein abundance for Coomasie or SYPRO Ruby suggesting it has a smaller dynamic range than these stains. This limited dynamic range of Deep Purple was observed again at the 2D level, where more abundant protein spots were saturating before their equivalent partners stained with Coomassie and SYPRO Ruby. In a real world example exposure times or Photomultiplier Tube (PMT) voltages would be reduced until the point where saturation was no longer observed, allowing quantification to still be accurate. This could mean however a dropping off of low abundance protein spots as exposure or voltage settings were altered to include all the spots without saturation.

1.6.2 Two dimensional In-Gel Electrophoresis (2D DIGE)

Fluorescence Two Dimensional In-Gel Electrophoresis (2D DIGE) is a development of the conventional 2DE platform, utilising the pre-electrophoretic labelling of protein samples with cyanine dyes (CyDyes) (Unlu et al., 1997; Tonge et al., 2001a). Three spectrally distinct CyDye fluorophores exist, each have an NHS-ester reactive group, which covalently attaches to the epsilon amino group of lysine residues within a protein. The concentration of dye within the protein/dye reaction mixture is such that the reaction is limiting. Approximately 1-2% of all available lysines are labelled, and of

those residues only one lysine amino acid per protein molecule. This is important as further labelling could affect quantification due to lysine rich proteins being over-represented in the spot map. CyDyes carry a single positive charge, so their effect on pI of labelled proteins and thus migration during IEF is minimal. Proteins labelled with CyDye exhibit an increase in mass of approximately 500 Daltons, so the migration of proteins in the second dimension SDS PAGE separation is not significantly altered (Unlu et al., 1997; Tonge et al., 2001a). This means CyDye labelled spot maps can be matched easily to spot maps of SYPRO stained sample. However, low molecular weight proteins see slight molecular weight shifts when compared to SYPRO (an effect which would be increased if extended labelling were allowed to occur). If the CyDye labelled gel is to be picked to provide protein material for MS analysis, it can be post-stained with SYRPO to ensure low molecular weight proteins are picked accurately.

Central to the technology is the existence of three spectrally distinct dyes allowing different labelled samples to be run on a single gel. Detecting protein differences with traditional 2DE methodology requires the comparison of images from at least two gels. This introduces variations in the 2D pattern not necessarily due to sample differences but to the intrinsic variation between 2D PAGE gels due to technological and/or experimenter limitations. With CyDye technology the differential analysis is performed on images from a single gel which is made possible by the distinct excitation wavelengths of each of the CyDyes. A further advantage is that the inaccuracies in quantification due to the effect of system (gel-to-gel) variation on spot abundance is reduced as much as possible by normalising against the Cy2 internal standard.

A traditional 2D DIGE experiment would involve the labelling of a control, treated and standard sample with a separate dye, each of which would then be run on a single gel. This removes the effect of gel to gel variation meaning changes seen within the resulting images have to be the result of sample variation.

A further advantage of this technology is the high level of sensitivity the dyes offer,

comparable to disruptive silver staining.

1.6.3 2D Gel Analysis Technologies for the Quantification of Abundance Change

The charting of physiological and pathological processes in biological systems is often aided by measuring changes in gene expression. Quantifying changes at the DNA or mRNA level cannot offer a complete picture as there is rarely a direct correlation between levels of DNA transcription and the levels of protein in cells. This was shown by Anderson and Seilhamer (1997) where they compared selected mRNAs and the resulting protein abundances in human liver and found little correlation between the two. Quantitative proteomic analysis measures the end product of protein synthesis which can then take into account processes such as protein degradation and post-translational modification (PTM).

1.6.4 Protein Identification Technologies

2DE techniques allows the separation and quantification of proteins within a mixture, and to chart how these proteins may change in response to stimuli or over time. Obtaining the identities of proteins of interest allows the assignment of the proteins' functional status.

Rapidly accumulating genomic and protein sequence data available to researchers in both public and private databases has revolutionised protein identification. To capitalise on the abundance of sequence data, a new wave of developments were made with the aim to provide reliable, accurate and high throughput protein identification technologies. Mass spectrometry accurately measures the mass of a molecule. MS has been adapted for the analysis of proteins. An MS system can only measure the mass of ionised molecules transferred into a vacuum. The development of 'soft' ionisation methods: electrospray ionisation (ESI) (Fenn et al., 1989) and matrix assisted laser

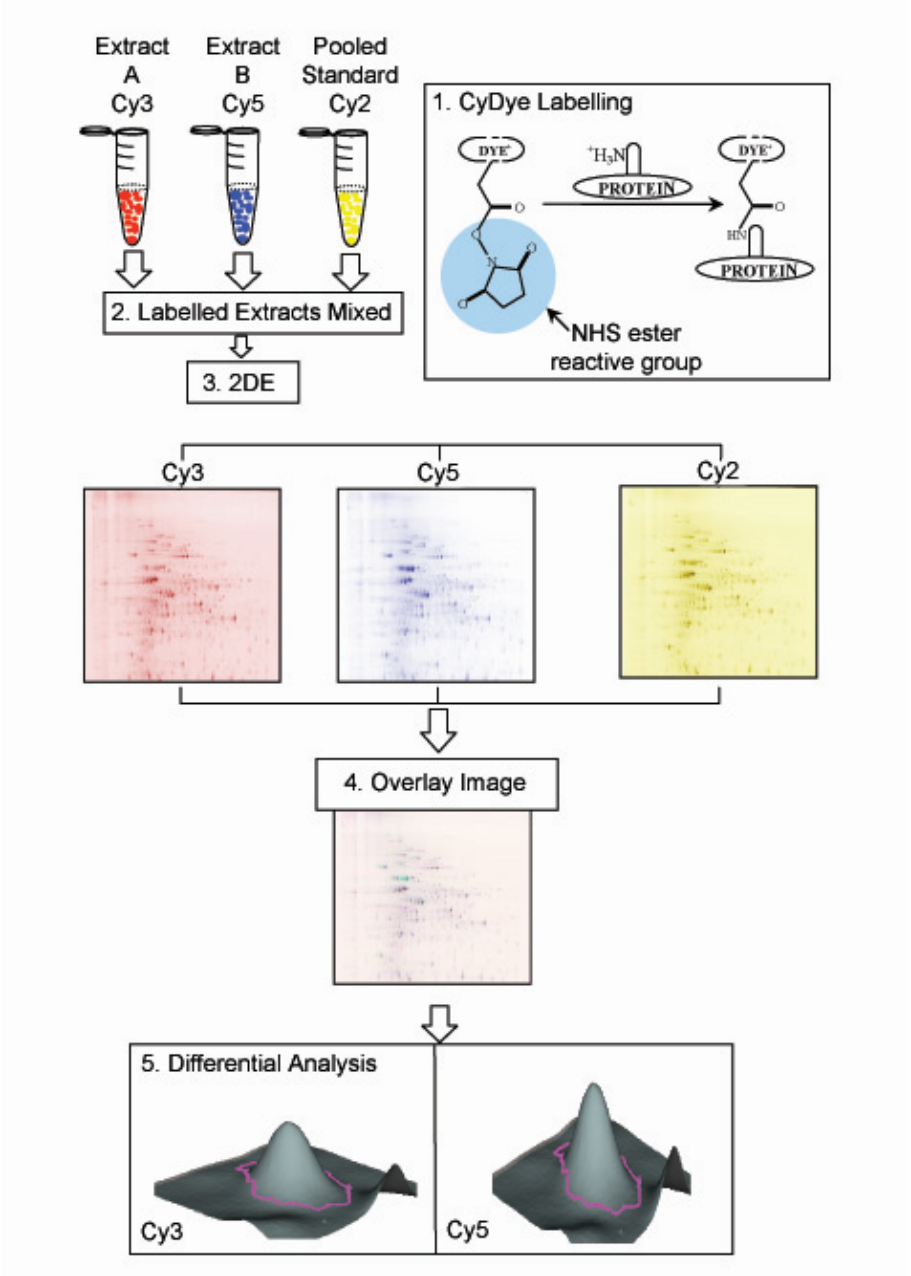


Figure 1.8: Schematic of 2D DIGE process. **1.** Protein extracts representing for example control and treated samples, or time points at differing stages of plant development, are labelled with Cy3 and Cy5 dyes. A pooled standard containing an equal amount of each protein sample (control and treated) within the experiment is labelled with Cy2. **2.** Samples are mixed in equal amounts and **3.** separated together by 2DE. Gels are scanned by a fluorescent imager such as the Typhoon Variable Mode Imager (GE Healthcare, Bucks UK). **4.**The fluorescent images can then be overlayed and **5.** subjected to computational analysis to identify significant abundance changes between sample groups, normalised against the Cy2 standard. Figure adapted from (Rowland, 2006)

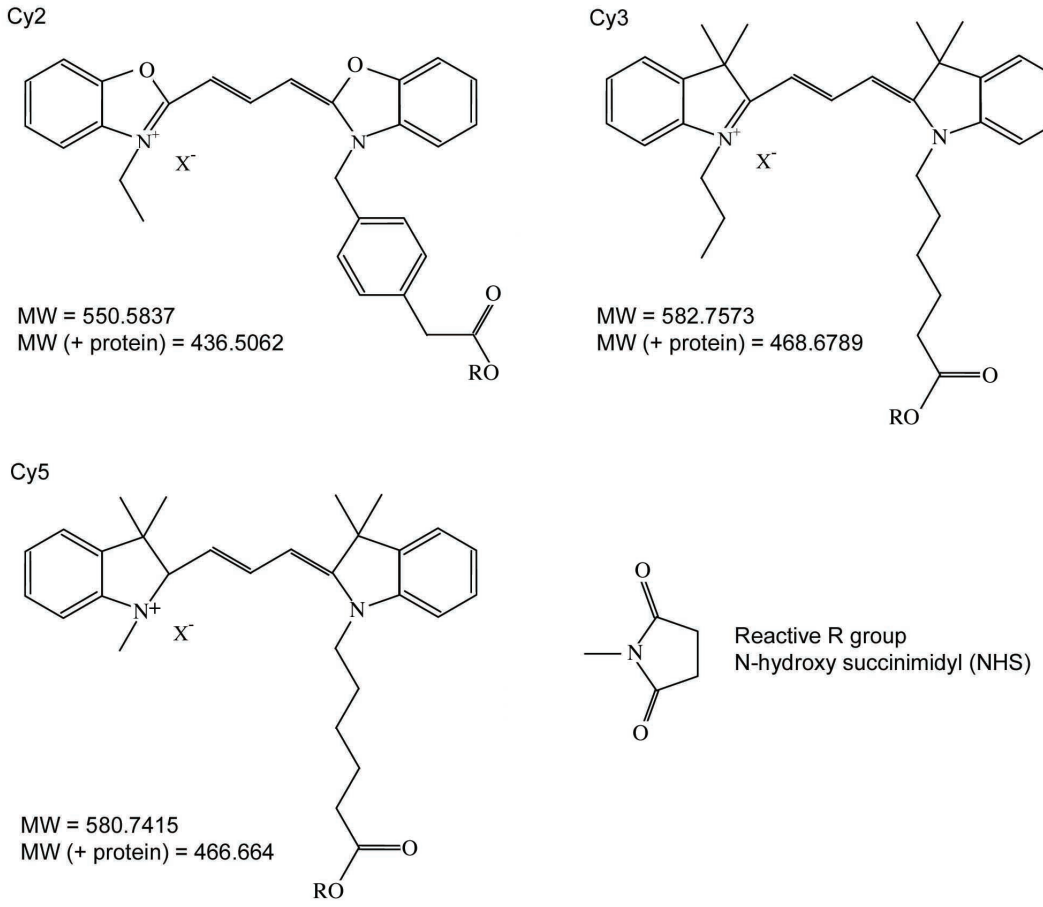


Figure 1.9: Cyanine dye structures used in minimal labelling

desorption ionisation (MALDI) (Karas and Hillenkamp, 1988) have allowed ionisation of large biopolymers, such as peptides and proteins, with minimal degradation. Mass spectrometers measure the mass of ionisable molecules by measuring their mass to charge (m/z) ratio. There are a variety of MS permutations, but all can be described as having three basic components: 1) an ionisation source, 2) a mass analyser and 3) a detector. Two different methods of ionisation, MALDI and ESI, have already been mentioned. Instruments utilising these ionisation methods are the most common in proteomic applications, and are described in more detail in the following sections.

1.6.4.1 Matrix Assisted Laser Desorption Ionisation Time-of-Flight (MALDI TOF)

With this MS method, molecules are ionised by MALDI and the m/z ratio of the ion is measured in a time-of-flight mass analyser. The analytes of interest are mixed with a low molecular weight matrix compound, usually α-cyano-4-hydroxy cinnamic acid or 2,5-dihydroxybenzoic acid. The matrix is typically dissolved in an acidic organic solvent. The matrix/solvent/analyte mixture is pipetted, often robotically if justified by the number of samples, onto a metallic target plate. Evaporation of the solvent in air results in the formation of analyte : matrix co-crystals (Patterson et al., 2001). The matrix molecules have absorption maxima at the 337 nm wavelength of the mass spectrometer's UV laser (Beavis and Chait, 1989). On application of the UV laser to the analyte / matrix on the plate, emission of the absorbed energy as heat results in sublimation of matrix crystals and analyte into the gas phase, where ionisation occurs. Ions are accelerated through the time-of-flight mass analyser, which is terminated by a detector. The flight time of the ions is recorded by comparing the laser pulse time to the time of the ion reaching the detector. As the velocity of ions is inversely proportional to the m/z ratio, the arrival time of the ions allows the m/z ratio to be calculated (Patterson et al., 2001).

MALDI TOF MS typically employs peptide mass fingerprinting as a means of protein identification. With this method protein and peptide molecules are first digested with an amino acid specific protease such as trypsin which cleaves at the arginine (R) and lysine (K) residues. The peptide fragments are then analysed by MS, giving a profile of peaks (the peptide mass fingerprint). The peptide mass fingerprint (PMF) data is then used to interrogate a database on theoretical *in silico* digested sequences. A common PMF searching algorithm, Molecular Weight Search (MOWSE), provides a confidence score alongside matched hits. The confidence score takes into account the molecular weight range of the identified protein, the protease used for digestion and the number of the peptides matched to the theoretical PMF (Pappin et al., 1993). Multiple hits are often obtained, probably due to homology between different proteins, and where this occurs the protein with the highest MOWSE score is generally assumed to be the correct identification.

Limitations with PMF based MS exist. Where mixtures are present within a protein sample, there is no way to discern which peptide peak is derived from which protein. This may cause attempts at identification to fail, by the presence of extra peptide peaks lowering the MOWSE score to below confidence limits, or may even cause a confident but erroneous match. PMF requires the existence of a sequenced genome with full length sequences. Databases of expressed sequence tags (ESTs) generally do not suffice, as the EST is less likely to incorporate enough full stretches between lysine and arginine-encoding codons to provide enough sequence coverage to obtain many confident database matches. A further problem is the detection of peptide peaks which do not match any corresponding entry in an available protein database. This may be due to incomplete digestion resulting in peptides longer than an *in-silico* digest would predict. The presence of post translational modifications alters the m/z ratios and this information is not always known or conveyed in sequence databases. The dust contaminant keratin can be ubiquitous in poorly handled samples and its abundance

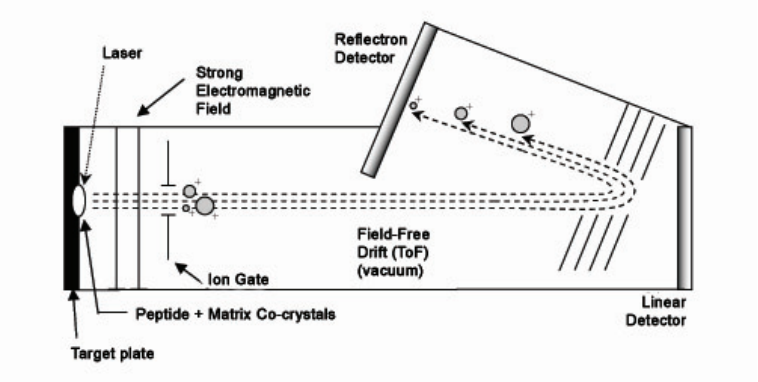


Figure 1.10: Schematic of MALDI TOF MS Instrument. Peptide / matrix co-crystals are ionised by UV laser before their transferral into the evacuated field-free TOF tube. The time-of-flight for each ion is measured by the detector and the m/z ratio calculated. Figure adapted from Patterson, et al. (2001).

can mask the experimental protein sample (Jensen et al., 1997; Eriksson et al., 2000). Many of these problems can be circumvented through the choice of suitable database parameters, for example mass units equivalents to post translational modifications can be removed from the peptide peaks before the database is interrogated.

1.6.4.2 MS/MS

For the purposes of identifying proteins in sequence databases, the use of actual sequence data provides a far more constraining approach than peptide mass fingerprinting (Zubarev et al., 1996). Tandem MS/MS utilises the specific ion selection capabilities and subsequent collision induced fragmentation of ions to yield overlapping small ion products which can then be used to directly search protein sequence databases or build sequence data. Tandem MS/MS typically employs triple quadrupole (TQ) (Lee et al., 1998), ion trap (IT) (Davis and Lee, 1997) or quadrupole time-of-flight (QTOF) MS (Borchers et al., 2000) instruments. Either ESI or MALDI may be used as an ion source. The use of ESI triple quadrupole system for the purposes of amino acid sequencing will now be described.

A schematic of a TQ MS instrument with ESI is shown in Figure 1.11. With ESI, a

solution containing the analyte (peptide/proteins) is passed through a fine needle held at high electrical potential (+1000 - 5000 V). The high potential causes the generation of positively charged ions and the dispersal of the analyte solution as a fine spray of highly charged droplets. The charged analyte solution is introduced into the evacuated MS through an entry orifice, held at lower potential (+100-1000 V). On entry into the MS, the analyte droplets are desolvated through the directed flow of inert, heated gas until only highly charged ion species remain (Kerbarle and Peshke, 2000; Kerbarle, 2000). A quadrupole is a series of four parallel rods, connected electrically and carrying DC and AC voltages of opposite polarity. It is through the fine control of quadrupole voltages that ions of particular m/z ratios pass through the quadrupole-containing cell, whilst ions having different m/z ratios are diverted from the flight path and are thus removed from further analysis. Thus the quadrupoles can be described as ion gates. As the name suggests, a triple quadrupole instrument contains three quadrupoles, ordered so that ions may pass sequentially through each cell. Operational modes of each quadrupole can be selected, allowing the instrument to carry different peptide analysis functions (see Table 1.3). For example, varying voltage of the quadrupole over time allows the selection of ions across the entire mass range of the instrument, in a 'scanning' operation mode. Alternatively, ions may be allowed to pass freely through a quadrupole by operating it in radio frequency (RF) mode. In the tandem MS/MS operating mode peptide ions are introduced into the first quadrupole (Q1) which selects specific ions based on their m/z ratios. Selected ions are allowed to enter the second quadrupole (Q2), in which collision with an inert gas such as argon causes collision induced dissociation (CID). The masses of the collision products are then determined by the third quadrupole (Q3) operating in scanning mode.

Table 1.3: Different operating modes of a TQ MS. MS/MS mode is most commonly used in proteomic applications

	Operation	Cells			Description
	Mode	Q1	Q2	Q3	
1	MS	RF	RF [a]	Scanning	Measures unfragmented ion masses. Alternatively, Q1 could scan masses, with Q3 in RF mode.
2	MS/MS	Specified *m/z*	CID	Scanning	The most common mode of operation in proteomic studies. Selected ions are subjected to CID and the m/z of resultant fragment ions are determined by scanning Q3. Resultant fragment ion mass spectrum is correlated to the m/z of the parent ion selected in Q1.
3	Neutral loss scan	Scanning	CID	Scanning [b]	Q1 and Q3 scan in a synchronous fashion, but are offset by the value a neutral fragment ion. Therefore, only ions which have lost the fragment ion group are transmitted through to the detector.
4	Precursor ion scan	Scanning	CID	Specified *m/z*	Q3 identifies specific groups which are cleaved off by CID, and act as reported ions. As the m/z of the parent ion is recorded, the presence of a reporter ion can be correlated back to the parent ion.

[a] Q2 is used without collision gas in this mode

[b] Scanning is offset by the value of a neutral group of interest which is lost during CID. Eg. 50 *m/z* corresponding to the loss of H_3PO_4 (Covey et al., 1991) allows identification of phosphorylated peptides.

ESI ionisation typically produces multiply-charged ion species (M + nH^{n+}). For example, trypsin digested peptides give rise to species with charges at the N-terminus of the peptide and the amino side chain of lysine or arginine residues, i.e. M + 2H^{2+} ions. These doubly-charged species are selected for fragmentation as both products are detectable in the instrument. Fragmentation of singly charged species results in only one fragmentation product carrying a charge, the uncharged species being undetectable to the Q3 mass analyser. Fragmentation of peptide ions commonly occurs at peptide bonds in a sequence-specific manner. Fragmentation products are classified by the type of fragmentation that occurred. Where the positive charge of the peptide ion remains on the N-terminus, the ion is termed a b series ion. A subscript following the b ion indicator denotes the number of amino acid residues in the fragmentation product, counting from the N-terminus. Where the fragmentation process has removed the N-terminus, and the charge remains on the C-terminus, they are classified as y series ions (Biemann, 1990). Similarly, a subscript number denotes the number of residues in the fragmented ion species, counting from the C-terminus. Sequential fragmentation results in the formation of b or y ions series, where the mass difference between successive peaks allows the determination of the amino acid residues that have been lost from the parent ion.

Non peptide-bond fragmentation results in the generation of other ion species. For example, the removal of carbon monoxide groups from b-series ions (-28 mass units) results in the formation of a series ions (see Figure 1.12). In addition, if a specific peptide ion undergoes multiple fragmentation events (Loo et al., 1993) internal acyl ions, containing 2 or more amino acids; or immonium ions, containing a single amino acid R group, are generated. As immonium ions represent individual amino acids they provide partial amino acid composition of the peptide.

The combination of the rich information generated from the fragmentation of a selected parent ion and the spectra of the parent ion itself, constitute a peptide sequence

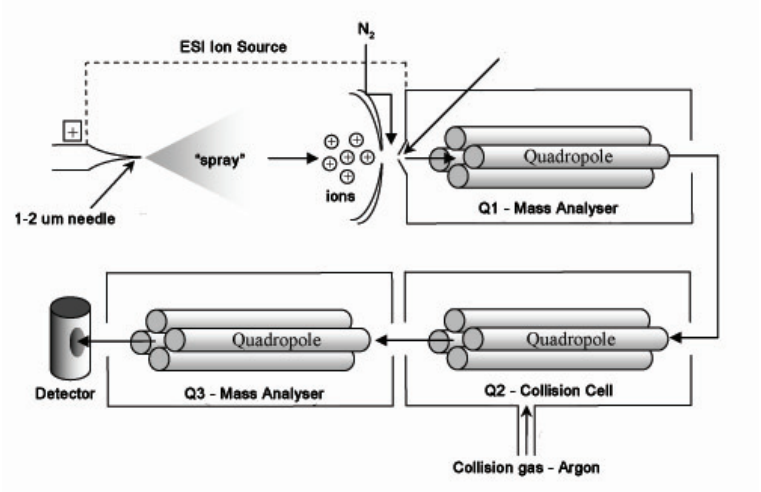


Figure 1.11: Schematic of a Triple Quadrupole MS/MS Instrument. Ejection of the peptide solution from a fine needle held at high voltage results in the formation of highly charged ions dispersed in a fine spray. N_2 gas desolvates the sample leaving peptide ions. In MS/MS mode, ions are selected for collision in the Q1 mass analyser. CID by argon gas in the Q2 cell, and the masses of peptide fragments are measured in the Q3 mass analyser. Figure adapted from Hall, JJ (2004).

tag which can be used to search protein databases (Mann and Wilm, 1994). Automated algorithms have been developed such as SEQUEST (Eng et al., 1994), although manual interpretation is sometimes required where the parentage of a particular ion series is unclear (Pardo et al., 2000). Algorithms for the automated identification of PTMs, such as phosphorylation and glycosylation, have also been developed (Annan and Carr, 1997). Together these technologies provide a highly sophisticated package for the sequencing of proteins, identification of PTMs and identification of the proteins, including to translated EST sequences, in protein databases.

1.7 Plant Organelle Research in the Post Genomic Era

The techniques and methodological developments in the areas of proteomics with relevance to this research project, namely sample preparation, 2D gel-based technologies and MS, have been described and discussed (see Sections 1.6.1 to 1.6.4.2). The

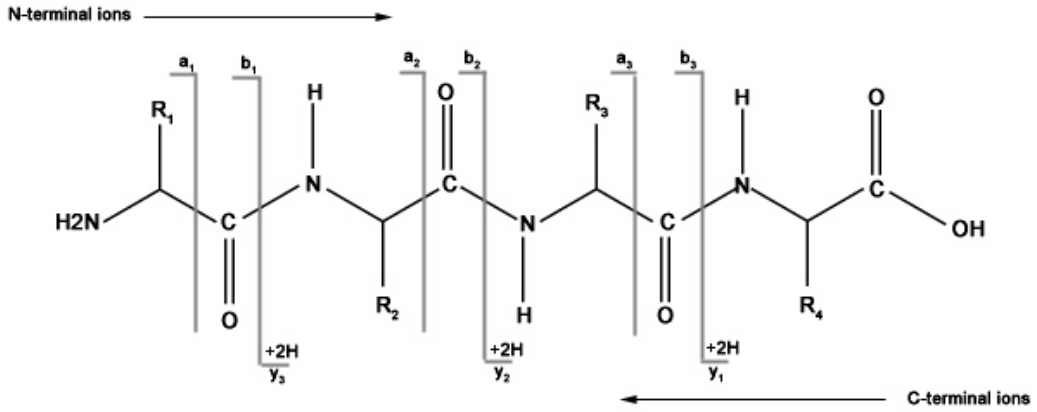


Figure 1.12: Fragment ion nomenclature (Biemann, 1990) for the *N*- and *C*-terminal ions. *y* ions result from *N*-terminal deletions, *b* ions from *C*-terminal deletions. *a* ions are the product of fragmentation of carbon monoxide from the peptide ion.

contribution of these technologies along with other proteomic techniques (i.e. blue native polyacrylamide gel electrophoresis, blue native DIGE, ICAT, iTRAQ, MudPIT) to the field of plant biology are discussed in this section.

Proteomics has been defined as the systematic analysis of the protein population in a tissue, cell or subcellular compartment, and the systematic analysis of the protein complement of the genome (Pandey and Mann, 2000; Patterson and Aebersold, 2003). The term was first coined in a research article which described the use of 2D electrophoresis to analyse the proteins extracted from *Mycoplasma genitalium*, the smallest known self-replicating organism, followed by downstream MS and Edman sequencing to identify proteins based on their amino acid sequence homology to known sequences from other organisms (Wasinger et al., 1995b).

The exponential rise in genome sequencing projects, starting with the sequencing of the first DNA-based genome (Sanger et al., 1977) through budding yeast (Goffeau et al., 1996), *Arabidopsis* (Theologis et al., 2000; Lin et al., 1999; Salanoubat et al., 2000; Mayer et al., 1999; Tabata et al., 2000), *Homo sapiens* (Lander et al., 2001; Venter et al., 2001) and recently Castor bean (http://castorbean.tigr.org/), has provided the genomic data to support a concomitant increase in proteomic projects. Proteomics,

that is - the analysis of all proteins expressed by the genome, provides a significant additional perspective into the functioning of biological systems. Unlike analyses at the DNA level, it can provide information on protein localisation within the cell, the temporal nature of gene expression, post-translational modifications, protein-protein associations and accurate quantification of protein amounts. Although quantification of gene expression is obtained with transcriptomic tools such as microarrays or real time PCR, differences in rates of protein and mRNA turnover mean it is not always possible to correlate quantification of mRNA transcripts to their respective protein products (Gygi et al., 1999). Thus the various methods and technologies of proteomic analysis have become significant components in a plant biologists experimental toolkit.

The technology traditionally associated with proteomic analysis is 2DE, initially reported by O'Farrell in 1975 (O'Farrel, 1975) (Please see Section 1.6.1 for a discussion on the historical development and limitations of this technology). The use of this technology in the analysis of plant protein samples occurred soon after, with an attempt to separate chloroplast membrane proteins (Novak-Hofer and Siegenthaler, 1977). Since then and to the present day, 2DE has been applied to a diverse array of plant protein samples. Different approaches have been taken, including analysing proteins extracted from plant cell culture suspensions (Lei et al., 2005), whole tissues (Wan and Liu, 2008) and fractionated organelles (Maltman, 2002). More recently, pre-electrophoretic labelled of protein samples rather than the use of post-stains has brought about high sensitivity, greater reproducibility and enhanced quantitative accuracy. It has been used in a diverse array of 2DE analyses including an investigation into the effects of salinity and hyperosmotic stress on *A. thaliana* cell proteins (Ndimba et al., 2005). In that study, a large total number of spots (2,949) were identified and 266 were found to be significantly changed in the stressed cells. Identifying such a large number of spots can itself present problems. 2DE proteomic analysis of subfractions rather than a total cellular protein analysis provides a significant benefit of reducing spotmap complexity

which can be a significant problem with many samples. It increases the likelihood of a single spot feature being an individual spot and thus increases the accuracy of the quantification of that spot and its subsequent analysis by MS. A greater benefit of organellar analysis though, and one universal to all proteomic analysis techniques, is it augments identification of unknown proteins because of their localisation within a known organelle. As biochemical function is to some degree separated by organelles, it gives potential clues to the identity and function of unknown or novel proteins. It also helps in the search for homologous proteins involved in biochemical pathways of interest and identified in other organisms. In the work by Maltman et al. (2002), the first published proteomic analysis of plant ER, a combined SDS PAGE and 2DE approach was taken to analyse soluble and membrane bound components of the *R. communis* ER proteome. In this work, chaperones and folding proteins of the ER were identified, along with the key ricinoleic acid biosynthetic enzyme oleate $\triangle 12$ hydroxylase. One of the challenges of this method is in obtaining pure organelle fractions. Some degree of confidence in a samples purity can be obtained by carrying out an initial survey of the sample, assaying for marker enzymes of different potential organelle, membrane and cystolic contaminants. A significant challenge to successful gel-based analysis of the proteome is the separation of membrane located proteins. Membrane proteins represent almost one-third of eukaryotic genomes, yet due to their poor solubility in the non-detergent isoelectric focussing buffer they tend to precipitate at their isoelectric point (Ephritikhine et al., 2004). Membrane proteins can be integral to the membrane, that is they contain at least one hydrophobic stretch of amino acids which cross the lipid layer of a membrane where they are resident, or membrane-associated where the protein is associated to the lipid membrane or to another associated or integral protein. Attempts to analyse membrane-associated proteins by 2D gels has met greater success than integral membrane protein analyses due to their comparatively greater hydrophilicity. The interaction of associated proteins can be selectively disrupted

by NaCl washes which with differential centrifugation can generate a useful subset of associated proteins for proteomic analysis. This technique was used by Maltman et al. (2002) in an analysis of proteins associated with the endoplasmic reticulum of germinating and developing *R. communis* endosperm. Sample complexity was reduced by separating the ER into lumenal, peripheral and membrane bound fractions. Proteomic studies focussing specifically on analysis of associated proteins in plants have been published. An analysis of rice plasma membrane associated proteins and their response to the plasma membrane defence elicitors chitooligosaccharides has been published which identified eight protein spots which were predicted to be membrane-associated and were up or down-regulated in response to the elicitor treatment (Chen et al., 2007). Attempts to identify membrane bound proteins using traditional 2DE methodologies in early proteomic studies were problematic. Plasma membrane proteins from *A. thaliana* were analysed by 2DE but on sequencing of spots the proteome was found to be significantly enriched for plasma membrane-associated proteins with few putative membrane proteins (Santoni et al., 1999). Alternative detergents such as C8Ø (Chevallet et al., 1998) and UTC8 (Santoni et al., 1999) have been used to replace CHAPS in the 2DE lysis buffer. These attempts met with some success, for example in the case of UTC8 there was 40% spot difference in the protein profile of *A. thaliana* plasma membrane proteins compared to the CHAPS 2DE lysis buffer control, and when these new spots were sequenced water channel proteins and H^+-ATPase were detected (Santoni et al., 1999). Generally, the availability of alternative non-gel-based proteomic technologies has superceded their use in the analysis of membrane proteins.

Whilst mass spectrometry has long been a central component in gel-based proteomic studies, first complementing and then often displacing earlier Edman sequencing technologies, it has recently developed the analytical and quantitative power to do away with gel-based steps for the purposes of measuring amounts of and fractionating proteins. MudPIT (multidimensional protein identification technology) utilises a

combination of liquid chromatography to provide multidimensional fractionation of a complex peptide mixture and MS (LC-MS) to analyse and identify the fractionated peptides (Washburn et al., 2001). The protein sample is first digested in the presence of the protease enzyme trypsin (Han et al., 2001; Goshe et al., 2003) or chemically with CNBr (Washburn et al., 2001) to yield a complex peptide mixture. The sample is then collected on a strong cation exchange (SCX) column, and multiple washes of increasingly concentrated salt solutions release successive peptide fractions, depending on their pI. Each fraction enters a second chromatography column, this time a reverse phase (RP) column, which provides a second dimension of separation. The RP column elutes peptides by hydrophobicity, through a gradient of increasingly concentrated organic solvent washes. Peptides then enter the tandem mass spectrometer for sequencing. MudPIT has been utilised in a number of plant proteomic studies including a comparison of rice (*Oryza sativa*) leaf, root and seed tissue which, along with a complimentary 2DE analysis, idenitified 2,528 non-redundant proteins (Koller et al., 2002). This analysis concentrated on the soluble portions of the three tissue types analysed, to allow a comparison of the effectiveness of 2DE and MudPIT. The MudPIT analysis identified a significantly larger number of unique proteins than identified via the 2DE method (2,363 versus 556). However, 165 non-redundant proteins were identified via the 2DE method which were not identified by MudPIT, indicating the complementary nature of the technologies in this study. Interestingly, this study identified only minor overlap of proteins between the three analysed tissues, with only 189 proteins being common to each tissue type. This indicates the highly specialised nature of each tissue type, and perhaps a relatively minor genetic commitment required to support central metabolic processes in a cell (Whitelegge, 2002). Although this Koller paper analysed soluble plant proteins, one of the most significant benefits to a MudPIT approach is its ability to analyse membrane proteins. MudPIT has been used to identify seed filling in soya bean endosperm, analysing five sequential stages of

endosperm development (2, 3, 4, 5 and 6 weeks after flowering) (Agrawal et al., 2008). Similarly to the Koller (2002) study, it utilised a complementary MudPIT and 2DE approach. 478 non-redundant proteins were identified, slightly more by MudPIT than by 2DE. Interestingly, just 70 proteins were common to both datasets, again indicating the usefulness of a combination of 2DE and MudPIT analysis methods. The MudPIT analysis was three-fold enriched for proteins predicted to be membrane resident due to the presence of hydrophobic amino acid stretches (Agrawal et al., 2008).

The ability to quantify levels of different proteins in a tissue or cell type is crucial if the researcher wants to compare the relative abundances of proteins within a proteome, or understand how protein abundance changes over time, between different tissues or in response to disease or treatment. Although highly sensitive quantification of gene expression can be obtained by analysing mRNA levels through microarray techniques, their abundances cannot confidently infer the abundances of their respective proteins. Although high throughput identification of proteins in complex mixtures is possible with MudPIT MS analysis, for quantification of proteins researchers were reliant on traditional 2DE proteomic techniques. As already discussed 2DE has inherent limitations with respect to the analysis of membrane proteins and proteins of low abundance, low molecular weight or extremes of pI. It is also labour intensive, often taking many days if not weeks to complete the experimentation and analysis of the 2DE gels, with only limited scope for automation. In the last ten years, techniques have become available to quantify proteins amounts within complex mixtures by mass spectrometry. Isotope coded affinity tags (ICAT) (Gygi et al., 2003) relies on a pair of cysteine-residue binding tags, consisting of three elements: a thiol-specific reactive group, a biotin group, and a central linker portion which either contains eight ^{1}H atoms (the light reagent), or 8 ^{2}H deuterium atoms (the heavy reagent, which is 8 Da heavier). Two protein samples to be compared are incubated with either the light and heavy reagents. Proteins are then enzymatically cleaved and those peptides with ICAT-tags (i.e. cysteine containing) are

separated by a biotin-binding column, before LC-MS/MS analysis. Peptides common to both samples co-elute into the mass spectrometer, as the two tags are chemically identical. The peptides have separate m/z ratios due to the mass difference of their tags, so their relative abundances can then be calculated from the mass spectra. A significant limitation of ICAT is it excludes all proteins that do not contain cysteine residues. Cysteine is not a particularly common amino acid, for example around 8% of proteins in *Saccharomyces cerevisiae* do not contain this residue (Gygi and Aebersold, 2000). Also, with only two tags available only pairwise comparisons can be made between samples. ICAT has been used in plant studies including a recent analysis of the chloroplast proteome of *A. thaliana*, comparing wildtype and thylakoid sorting protein mutant strains (Rutschow et al., 2008)). ICAT identified changes in light harvesting complex composition and changes in the relative abundances of photosystems I and II between the mutant and wildtype strains. The ICAT analysis in the study was complemented by another MS-based quantitative proteomic analysis technology, iTRAQ. iTRAQ (isobaric tag for relative and absolute quantification) (Ross et al., 2004) is a conceptually similar to the ICAT technology. This method utilises four isobaric (i.e. equal mass) tags which label the free amines of a digested complex protein sample, thus unlike ICAT all peptides in a sample are labelled. The labelled peptide samples are mixed and analysed by LC-MS/MS. As the tags are isobaric they affect elution of common peptides equally. In tandem mass spectrometry mode (peptides are isolated and fragmented, see Section 1.6.4.2) each iTRAQ tag generates a unique reporter ion. By comparing the relative intensities of the reporter ion on the mass spectra quantification is achieved. The four tags available and the complete labelling of peptides provide significant advantages over ICAT.

Protein mass spectrometry of sufficient sensitivity can identify post translational modifications to proteins, such as phosphorlyation or glycosylation. Ubiquitination, the conjugation of a protein with ubiquitin, is important in protein abundance, localisation

and activity regulation in cells. MudPIT was used in the first analysis of the plant ubiquitinated protein sub-proteome (Maor et al., 2007). An ubiquitin-binding affinity chromatography step captured ubiquitinated proteins from a total cellular protein extract from *A. thaliana*. The separated ubiquitinated protein subset was then digested with trypsin and analysed by a MudPIT system. 294 proteins were identified, 56 of which contained ubiquitinated lysine residues. This analysis indicates the potential to augment standard MudPIT or LC-MS chromatography separations to enrich a complex protein mixture to focus in on areas of interest. Another example would be to use immobilised antibodies to retain proteins of interest prior to MS analysis. Under suitable protein sample preparation conditions, protein-protein interactions can be retained to gain an understanding of protein complexes. Protein complexes are likely to play a significant role in biochemistry through for example substrate channelling, and thus proteomic methods conserving native protein interactions is desirable. Blue native electrophoresis is one such technique (Schagger and von Jagow, 1991). It utilises Coomassie Blue G-250 to provide a net negative charge to proteins in a first dimension separation, following by an SDS-PAGE separation in the second dimension. It has been successfully used to identify interacting components on the mitochondrial respiratory chain (Heazlewood et al., 2003) and photosystems in chloroplasts (Heinemeyer et al., 2004). It is possible that protein complexes play an important role in lipid biochemistry, and such a technique may be useful in identifying biochemically important lipid biosynthesis isoforms or even novel enzymes.

1.8 Yeast and *Yarrowia lipolytica* Lipid Biochemistry

1.8.1 Lipid Substrate Utilisation

A number of yeasts, such as *Candida maltosa* (Mauersberger et al., 1996), *Pichia guilliermondii* (Wickerham, 1966) and *Yarrowia lipolytica* (Barth and Gaillardin, 1996),

are able to utilise hydrophobic substrates such as alkanes as a carbon source for growth. The oleaginous yeast *Y. lipolytica* is able to readily utilise a diverse array of other hydrophobic substrates including alkenes, fatty acids (including ricinoleic acid), fats and oils (Daum et al., 2007). It can often be isolated from substrates rich in protein and lipids such as dairy products, cheese and meats. Indeed, it is one of the predominant species in Camembert and blue cheeses (Roostita and Fleet, 1996).

There are two hypothetical mechanisms by which hydrophobic substrates enter yeast cells. The substrates can be emulsified through the secretion of surfactants by the organism, in a process termed surfactant-mediated transport (Mauersberger et al., 1996). Alternatively, there is a direct interaction between the hydrophobic substrate and the cell wall, termed direct interfacial transport. There is evidence of direct interfacial transport in *Y. lipolytica*. Cell surface protrusions have been reported on cells growing on oleic acid, which are not present in cells growing on glucose media (Mlickova et al., 2004). Interaction of hydrophobic substrate and the *Y. lipolytica* cell surface has been found to correlate with a decrease in cell wall polarity (Kim et al., 2000) and it is hypothesised that this may be linked to protrusion formation. There is also evidence of surfactant mediated transport in *Y. lipolytica*, where the active secretion of surfactants into the culture media has been reported from cells growing on hydrophobic substrates (Kinjarde and Pant, 2002). It is possible that in *Y. lipolytica*, the two mechanisms work in concert to cause the efficient utilisation of hydrophobic substrates. The secretion of surfactants increases the surface area of accessible lipid within the culture media through the formation of micelles. Protrusion formation allows the formation of comparatively hydrophobic microenvironments at the surface of the cells, potentially aided by protrusion-targeted proteins with a high density of hydrophobic -R groups presenting at the cell surface. Electron micrograph (EM) figures of *Y. lipolytica* growing on oleic acid reveal interaction between the protrusions and the lipid micelles, and tantalisingly, transmission EM figures suggest a channel from

the protrusions into the ER (Mlickova et al., 2004).

Y. lipolytica is capable of utilising TAG molecules as the sole carbon substrate, through the secretion of extracellular and cell surface lipases (Barth and Gaillardin, 1996). The mechanism by which fatty acids, either free within the media or hydrolysed from TAG, then enter the cells is controversial (Fickers et al., 2005) although there is evidence that free diffusion of fatty acids into cells occurs when their concentration in the media is $\geq 1\mu$M (Kohlwein and Paltauf, 1984).

1.8.2 Storage Oil Production

De novo formation of the DAG precursor PA proceeds via the same step-wise acylation of G3P in yeast as described in plants (see Section 1.2.4.1). An alternative route to PA formation utilises dihydroxyacetone phosphate (DHAP) acyltransferase to acylate DHAP and form 1-acyl DHAP. This molecule is subsequently reduced by the action of 1-acyl DHAP reductase to form lysophosphatidic acid, which can be acylated at the *sn*-2 position of the glycerol backbone to form PA. PA can also be formed through the action of phospholipase D, or by the phosphorylation of DAG by DAG kinase (Müllner and Daum, 2004).

Dephosphorylation of PA yields DAG, the immediate precursor to TAG formation. As in plants, there are two major routes to TAG formation: the esterification of fatty acid from acyl-CoA (acyl-CoA dependent route) or the esterification of a fatty acid from phospholipid (acyl-CoA independent route). Genes expressing enzymes critical to both these routes have been identified in *Y. lipolytica* (Beopoulos et al., 2008).

1.8.2.1 Acyl-CoA Dependent Pathway

Acyl-CoA dependent formation of TAG is catalysed by DAGAT (Lehner and Kuksis, 1996), whose encoding gene, related to mammalian acyl-CoA : cholesterol acyl transferase (ACAT), has been cloned in a diverse range of organisms including humans

(Oelkers et al., 1998), mice (Cases et al., 1998) and plants (Hobbs et al., 1999; Zou et al., 1999). Although DAGAT activity had been localised to oil bodies *in vitro* in *S. cerevisiae* as early as 1978 (Christiansen, 1978; Christiansen, 1979), the cloning of a DAGAT gene in any yeast remained elusive. In 2001, two DAGAT genes unrelated to the ACAT gene family were cloned from the oleaginous fungus *M. ramanniana* (Lardizabal et al., 2001). Subsequently, homologues to the *M. ramanniana* DAGAT genes were identified in a range of organisms, leading to a proposed new DAGAT gene family, DAGAT2.

DAGAT2 has been identified in *S. cerevisiae* with 44% amino acid sequence homology (Sorger and Daum, 2002). Localisation studies of the *S. cerevisiae* DAGAT enzyme (termed Dga1p in this organism) using [^{14}C] labelled DAG and acyl-CoA found 70-fold enrichment of DAGAT activity in the oil bodies compared to crude homogenate, and a 2-3 fold enrichment in the microsomes (Sorger and Daum, 2002). *S. cerevisiae* Dga1p activity was found to be strongly dependent on Mg^{2+} and K^{+}. DAGAT activity in the microsomes of a dga1$\triangle$ mutant strain was only slightly decreased compared to wildtype control. However, DAGAT activity in the oil bodies was found to be just 5% of wildtype control. This indicates that Dga1p is likely the only or major DAGAT in oil bodies, that Dga1p is also localised to the ER alongside other DAGAT activity (Sorger and Daum, 2002). Despite the significant enrichment of DAGAT activity in the oil bodies reported by Sorger and Daum (2002), in another study published in the same year DAGAT activity was only identified in the ER (Oeklers et al., 2002).

DAGAT2 has been identified in *Y. lipolytica* as an oil body-resident enzyme (Athenstaedt et al., 2006) in oil bodies extracted from cells growing on oleic acid but not glucose. It has 33% amino acid homology with the *S. cerervisiae* DGAT2 (Beopoulos et al., 2008).

Yeast DAGAT displayed significant substrate selectivity *in vitro*, with oleoyl CoA and palmitoyl CoA the preferred substrates. Myristoyl CoA, stearoyl CoA, arachidinoyl

CoA and linoleoyl CoA were used at a significantly lower rate (Oeklers et al., 2002).

1.8.2.2 Acyl-CoA Independent Pathway

In mammalian cells, the synthesis of cholesterol esters occurs mainly through the transfer of acyl-CoAs by the DAGAT related enzyme ACAT (acyl-CoA cholesterol acyltransferase). An alternative route to cholesterol ester formation occurs through an acyl-CoA independent route, utilising acyl groups at the *sn*-2 position of PC, by the mammalian blood enzyme LCAT (Glomset, 1968). The discovery of an acyl-CoA independent route to TAG formation was reported for both plants and yeast in 2000 (Dahlqvist et al., 2000; Oeklers et al., 2000). Sequence homology between DAGAT and LCAT raised the possibility that an LCAT related enzyme could be responsible for PDAT activity observed in yeast microsomes (Dahlqvist et al., 2000). One *S. cerevisiae* open reading frame (YNR008w) was identified by homology search, and was found to share 27% overall identity with the human lecithin cholesterol acyltransfease (LCAT). Now termed *LRO1* (Lecithin Related Open reading frame), key catalytic residues are conserved between the yeast PDAT and the mammalian LCAT. Ser181 and Asp345, two members of the LCAT catalytic triads, are conserved in the yeast gene. As reported for *R. communis*, *C. palaestina* and sunflower, PDAT in yeast utilises acyl groups at the *sn*-2 position of PC and PE. Examination of a *lro1*$\triangle$ mutant strain found that PDAT activity was removed, but no change in the production of TAG using acyl-CoA as a substrate could be observed (Oeklers et al., 2000; Sorger and Daum, 2002), indicating the existence of at least two distinct TAG formation pathways.

In vitro assays with yeast microsomes found that yeast PDAT incorporates *sn*-2 ricinoleoyl CoA into TAG with 2.5 times high efficiency compared to an oleoyl group at the same position (Dahlqvist et al., 2000). In the study, no preference for *sn*-2 vernoloyl groups over *sn*-2 oleoyl was identified. Diricinoleoyl DAG and divernoloyl DAG were superior acyl acceptor molecules than dioleoyl DAG. Interestingly, despite its homology

to human LCAT, yeast PDAT is unable to catalyse removal of ergosterol from PC.

Lro1p activity could not be detected in the oil bodies of *S. cerevisiae*, but was localised exclusively to the ER (Dahlqvist et al., 2000; Oeklers et al., 2000). This is opposite to what was observed for Dga1p, whose activity was found to be highly enriched in the oil bodies (Sorger and Daum, 2002).

A *Y. lipolytica* orthologue to PDAT has been identified, which has 45% amino acid homology (Beopoulos et al., 2008).

1.8.2.3 Temporal Separation of DAGAT and PDAT Activity

The relative contributions of Dga1p and Lro1p to TAG synthesis in *S. cerevisiae* were studied in a *dga1*△ and *lro1*△ mutant strains, and revealed distinct temporal separation of Dga1p and Lro1p activity. Oeklers et al. (2002) compared TAG production at exponential growth phases compared to the stationary growth phase of yeast, and found a less pronounced decrease in TAG in exponentially growing cells than those grown to stationary phase. This was confirmed Sandager et al. (2002) who compared *dga1*△ and *lro1*△ mutant strains and found *dga1*△ to be the only strain with significantly reduced TAG levels at stationary phase (Sandager et al., 2002). Conversely, significant decreases in TAG production in *lro1*△ strains were only observed during exponential growth phases (Oeklers et al., 2000). Dahlqvist et al. (2000) found that *LRO1* expression was slightly higher during exponential growth. Transcriptional upregulation has been found to occur in *S. cerevisiae* entry into stationary phase (Gasch et al., 2000). Together the expression data support the observations of TAG accumulation reported by Oeklers et al. (2002) and Sandager et al. (2002).

Double knockout *dga1*△/*lro1*△ strains showed an 80% reduction in TAG when the cells reached late logarithmic growth (Sorger and Daum, 2002). However, Oeklers et al. (2002) reported a 97% decrease in TAG production throughout growth. The discrepancies between the observations were though possibly due to the use of different

background strains. Both observations indicate a) Lro1p and Dga1p are the major route to TAG formation in *S. cerevisiae* and b) another, minor, route to TAG formation must exist in this yeast.

It is possible that temporal separation of DAGAT and PDAT activity exists in *Y. lipolytica* but there is currently no published literature to support this.

1.8.2.4 Minor DAGAT Activities of Are1p and Are2p

Mammalian DAGAT belongs to the same family as ACAT, which includes two mammalian ACATs (ACAT1 and ACAT2) and two yeast ACAT related enzymes (*ARE1* and *ARE2*) (Müllner and Daum, 2004). Are1p and Are2p both catalyse sterol esterification (Yang et al., 1996; Yu et al., 1996). Their relationship to a gene family bearing close homology to DAGAT led research efforts to investigate whether they are responsible for the residual TAG formation reported in the *dga1*△/*lro1*△ double knockout strains (Sorger and Daum, 2002; Oeklers et al., 2002). This was confirmed by comparing two triple mutant *S. cerevisiae* strains: *lro1/dga1/are1* and *lro1/dga1/are2*, one of which expressed *ARE1* and the other *ARE2* (Sandager et al., 2002). Both strains, expressing either *ARE1* or *ARE2*, were found to have DAGAT activity, but at very low levels. There is however some controversy as to the extent of the respective contribution of Are1p and Are2p, as an earlier study ascribed DAGAT activity to Are1p solely, with Are2p restricted to sterol metabolism (Sandager et al., 2000). The quadruple mutant *dga1*△/*lro1*△/*are1*△/*are2*△ was devoid of all TAG however (Sandager et al., 2002).

In *Y. lipolytica*, only one steryl ester synthase (*ARE*) has been identified, which has 30% amino acid homology to Are2p of *S. cerevisiae*. Its identify was confirmed by enzymatic analysis (Beopoulos et al., 2008).

1.9 Aims

The overarching general hypothesis behind this study is as follows: The enzymatic components key to high levels of triricinolein (TRO) production in *R. communis* can be identified through the use of a proteomic analysis targeting the site of castor oil biosynthesis (the developing seed) and the contribution of these enzymes to TRO production can be rapidly assessed in yeast prior to transformation in a plant host.

Out of this general hypothesis came two core groups of aims. Firstly, to employ state-of-the-art proteomic technology to characterise the differences between the endoplasmic reticulum of germinating and developing *R. communis* seed with a view to identifying components of lipid biosynthesis in the developing tissue. Specifically:

- to establish the preparation procedures for *R. communis* cv. 99N89I and to assess the reproducibility of the preparation method with this strain
- to establish the proteomic techniques for large format preparative and analytical analysis of protein samples, and
- to perform a large-scale multi-replicate quantitative 2D DIGE analysis of germinating versus developing *R. communis* ER and to obtain the identities of the significantly elevated proteins by MALDI-TOF and MS/MS sequencing.

Secondly, with an aim of aiding the establishment of an *in vivo* lipid gene assay system, to develop the protocols for the growth of the appropriate strains of the oleagnious yeast *Y. lipolytica* on lipid media, the extraction of lipid from this organism and the procedures for lipid analysis amenable to the lipid extracts. In particular,

- to establish growth of *Y. lipolytica* on a range of carbon substrates, to extract lipid and identify the TAG component by TLC and GC
- to establish TLC and quantitative GC produces suitable for routine and rapid analysis of lipid extracts from *Y. lipolytica*, including TRO.

Chapter 2

Materials and Methods

2.1 Materials

All chemicals used throughout this study, unless otherwise stated, were supplied by Sigma Chemical Company Ltd., Fancy Road, Poole, Dorest BH17 7HN UK. Methyl ricinoleate was supplied by Stearinerie Dubois, Rue du Dome, Boulogne, France.

2.2 Plant methods

2.2.1 Growth of *R. communis*

2.2.1.1 Developing seed

R. communis seed variety 99N89I was obtained from Arkema (Colombes, France). For the production of fresh developing seed, seeds were first imbibed overnight in a Buchner flask connected to running water. After imbibition seeds were planted in a 50:50 mixture of sand and compost in 30 cm diameter pots measuring approximately 2 cm beneath the surface of the compost. Pots were watered daily directly on to the compost surface. Plants were grown in a Sanyo Gallenkamp (Leicestershire, UK) growth room with a 16 hour photoperiod. The temperature of the growth room was set to 23 °C in the light and 18 °C in the dark. Developing *R. communis* seeds were harvested between 25 - 28 days after flowering.

2.2.1.2 Germinating seed

R. communis 99N89I was also used for production of germinating seed. Seeds were imbibed overnight in a Buchner flask connected to running water before planting a tray of moistened Vermiculite. Seeds were distributed across the surface of the moistened Vermiculite with approximately 5 cm spacings between each seed, then covered to a depth of 1 - 2 cm with more moistened Vermiculite. Seeds were germinated in a growth room, in the dark, at 30 °C for 3 days.

Table 2.1: Constituents of *R. communis* endosperm homogenisation buffer

Constituents	Stock solution	Volume	Final Concentration
Sucrose	1 M	25 ml	500 mM
KCl	4 M	125 μl	10 mM
EDTA	0.5 M	100 μl	1 mM
$MgCl_2$	1 M	50 μl	1 mM
DTT	1 M	100 μl	2 mM
Tricine/KOH pH 7.5	1 M	7.5 ml	150 mM
PMSF	1 M	5 μl	0.1 mM

2.2.2 Subcellular fractionation of *R. communis* endosperm

2.2.2.1 Developing seed

25 - 28 day after flowering seed pods were bisected with a clean razor blade to reveal the white endosperm of the developing *R. communis* seed. A staging method of seed development was used (Greenwood and Bewley, 1982) to select endosperm of similar developmental age (Stage V, see 3.3.1.1). Endosperm was removed from the seed pods with a small spatula and transferred to a glass Petri dish on ice, containing 30 ml of homogenisation buffer (Table 2.1). After the collection of endosperm from typically 30 seed pods, they were homogenised for exactly 20 minutes with a new razor blade to give a uniform endosperm 'slurry' (on ice). The homogenate was filtered through a 100 μm mesh into an F0650 centrifuge tube (Beckman Coulter, Fullerton, CA. USA). This filtrate was then centrifuged (3000 x *g*, 4 °C, 20 minutes) to yield a pellet of unbroken cells and cell debris and a supernatant containing liberated subcellular material. The resultant fat pad floating on top of the supernatant was carefully removed with a spatula and the supernatant tipped away from the pellet.

The supernatant was then layered upon a discontinuous sucrose gradient in a 38 ml SW28 (Beckman Coulter, Fullerton, CA. USA) centrifuge tube, set up as follows: 7 ml 30% w/w sucrose (containing 1 mM EDTA, 0.1 mM PMSF), 15 ml 20% w/w sucrose

(containing 1mM EDTA, trace PMSF), 13 ml endosperm preparation. All solutions and tubes were kept on ice during this process. Balanced tubes were then ultracentrifuged for 2 hours (100,000 x *g*, 2 °C). Following ultracentrifugation, a defined band of semi-purified ER is visible at the interface between the 30% and 20% sucrose solutions. This was removed by carefully piercing the wall of the centrifuge tube with a sterile needle and syringe and sweeping through the interface region until the band is captured and a clean interface is left. This was done for all tubes, the semi-purified ER samples pooled and an equal volume of 60% w/v sucrose added. A second purification step was then performed using floating centrifugation in a 3-step discontinous sucrose gradient. 15 ml SW41 (Beckman Coulter, Fullerton, CA. USA) tubes were used, containing the following solutions: 4 ml 60% sucrose-mixed semi-purified ER, 3 ml 40% sucrose, 3 ml 30% sucrose, 2 ml 20% sucrose. All sucrose solutions contained 1mM EDTA and trace PMSF. Balanced tubes were then ultracentrifuged for 22 hours (100,000 x *g*, 2 °C) resulting in a band of purified ER collected at the 20% / 30% interface. As before, this was removed with a sterile needle for each tube and the samples pooled. An equal volume of ice-cold ddH_2O was added and the purified ER sample pelleted at 100,000 x *g* for 1 hour. Finally, the supernatant was removed to leave a dark pellet of purified ER. This was re-suspended in either 10% glycerol or a 2D-electrophoresis lysis buffer (9M urea, 2M thiourea, 4% CHAPS) depending on the downstream intended usage of the sample.

2.2.2.2 Germinating seed

After 3 days of germination, *R. communis* seeds were removed from the vermiculite and the extent of germination assessed. Those seeds that had reached the '3 day' development, characterised by the separated seed husk, intact endosperm and root hairs, were selected for preparation (see 3.3.1.1). Their roots and seed husks were removed before washing in ice-cold water (from this point all solutions were kept on

ice). A razor blade was then used to bisect the endosperm allowing the cotyledon to be removed and discarded. Endosperm halves were transferred to a Petri dish on ice, containing 30 ml of homogenisation buffer (Table 2.1). The purification procedure was then identical to the developing ER purification protocol above.

2.3 Proteomic methods

2.3.1 Bradford protein assay

The Bradford protein assay (Bradford, 1976) is a colourimetric assay based on the absorbance shift of Coomassie dye to 595 nm on its binding to the protein. The increase of absorbance at 595 nm is proportional to the amount of bound dye and thus to the concentration of protein within a sample. Protein Assay Dye Reagent Concentrate (BioRad Laboratories, Hercules US) was used as the dye, as per manufacturers instructions. Bovine Serum Albumin (BSA) was used as a protein standard from which standard curves were created.

2.3.2 Modified Bradford assay

A limitation of the standard Bradford protein assay is that it is incompatible with the 2DE lysis buffer due to the presence of thiol containing groups and carrier ampholytes which interfere with the standard assay. A modified Bradford assay was developed to circumvent this issue (Ramagli et al., 1985). This required the acidification of protein samples before absorbance measurement and allowed a stable, linear relationship between protein and absorbance to be obtained. 1 ml reaction volumes were typically used for the assay procedure, containing 25 μl of either 1 - 20 μg bovine serum albumin (BSA) in 2DE lysis buffer or the unknown protein sample in a total volume of 25 μl 2DE lysis buffer. To this was added 10 μl 0.1 M HCl and 65 μl ddH_2O. Finally, 900 μl 25% (v/v) Protein Assay Dye Reagent Concentrate (BioRad Laboratories, Hercules US) was

added and the reaction mixture vortexed and incubated at room temperature for 10 minutes. Absorbances were measured immediately after incubation and unknown values plotted against the known BSA standard curve to allow comparison to the standard.

2.3.3 1D SDS PAGE gels

Mini 1D SDS-PAGE was performed using the Laemmli tris-glycine buffer system (Laemmli, 1970b). Bio-Rad Mini-Protean II gel kit was used with a Bio-Rad PowerPac 300 power supply (Bio-Rad Laboratories, Hercules, US).

2.3.3.1 1D SDS PAGE Sample preparation

For the purposes of 1D SDS PAGE, protein samples were solubilised in the presence of 5x SDS loading buffer (10% (w/v) SDS, 5% (w/v) DTT, 0.05% (w/v) bromophenol blue, 0.312 M Tris-HCl pH 6.8, 50% (v/v) glycerol) to give a final volume of 1% SDS loading buffer.

2.3.3.2 Gel casting

For the purposes of 1D SDS PAGE 0.75 mm thick discontinuous acrylamide gels were routinely poured, consisting of a 12% resolving gel and 5% stacking gel. Glass plates were sprayed with 70% ethanol and allowed to air dry before gel cassettes were assembled. The resolving solution consisted of 12% acrylamide (acrylamide:bis-acrylamide 37.5:1) (Bio-Rad Laboratories, Hercules, US), 375 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.05% (w/v) ammonium persulphate (Bio-Rad Laboratories, Hercules, US), and 0.2% (v/v) TEMED. Ammonium persulphate and TEMED were added just prior to pipetting the 12% gel solution into the gel cassette. Water saturated butan-1-ol was then pipetted on top of the 12% gel solution which facilitated the formation of a smooth, horizontal gel surface on polymerisation. Following this step the butan-1-ol was tipped away, the gel surface washed with ddH_2O and dried with blotting

paper. Layered upon the 12% resolving gel was poured a 5% stacking gel consisting of 5% acrylamide (acrylamide:bis-acrylamide 37.5:1) (Bio-Rad Laboratories, Hercules, US), 125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.05% (w/v) ammonium persulphate (Bio-Rad Laboratories, Hercules, US), and 0.2% (v/v) TEMED and was prepared in the same fashion as for the 12% solution. A 70% ethanol cleaned 0.75 mm comb was inserted into the still-liquid stacking gel before leaving to set. Combs were then removed gently under slowly running water and the gels used immediately or stored in moistened tissue paper at 4 °C.

2.3.3.3 Tank set up and sample loading

Polymerised gels were clipped into the Bio-Rad (Bio-Rad Laboratories, Hercules, US) tank frame and inserted into the gel tank. 800 ml of 1 x running buffer was then poured into the tank. Running buffer was made as a 10 x stock containing 14.4% (w/v) glycine, 3.0% (w/v) tris, 1.0% SDS.

Samples solubilised in SDS loading buffer were loaded into wells with a tapered pipette tip and care was taken not to draw air into the pipette tip when taking up samples to be loaded onto the gel. Air in the pipette tip could potentially risk cross contamination of sample wells as the bubbles rise up from well. Typically protein samples were run alongside 'SDS 7' molecular weight standards, containing proteins of 66, 45, 36, 29, 24, 20.1, 14.2 kDa (Dalton Mark V11-L. Sigma).

2.3.3.4 Electrophoresis conditions

Initially 100V was applied to the gel to bring the samples through the stacking gel and into the resolving gel, at which point the voltage was increased to 200V. When the bromophenol blue dye-front reached within 1 mm of the end of the gel, the gel kit was carefully dismantled and the gels removed. For protein visualisation Coomassie Brillilant Blue R250, silver or Sypro^{TM} stains were used.

2.3.4 Mini 2-dimensional SDS PAGE

2.3.4.1 Protein sample preparation

Protein samples for 2-dimensional SDS PAGE are solubilised in a highly concentrated lysis buffer containing the chaotropic agent urea, alongside surfactants CHAPS and thiourea (9 M urea, 2 M thiourea, 4% CHAPS). The high concentrations of the components in the lysis buffer means solubilisation can take some time. This was quickened by keeping the lysis buffer in a water bath heated to 30 °C by means of a heating plate of a stirrer, with gentle mixing. It is critical that the temperature of the lysis buffer does not rise above 35 °C as urea will decompose above this temperature.

2.3.4.2 Reswelling IPG strips

7 cm IPG strips were re-swelled with lysis buffer (9 M urea, 2 M thiourea, 4% CHAPS, 1% DTT (Melford Laboratories, Suffolk UK), 2% carrier ampholytes, bromophenol blue) containing the protein sample. The pH range of the carrier ampholytes was matched to the pH range of the IPG strip to be used. Lysis buffer containing the desired amount of protein sample was diluted to a final volume of 125 μl. An Immobiline^{TM} Drystrip re-swelling tray (GE Healthcare, Bucks UK) was used for the re-swelling process, 125 μl sample pipetted into the IPG strip channel and with a pair of forceps the plastic backing removed from an IPG strip before lowering onto the sample. To keep the IPG strip on top of the sample, and to prevent the strip drying out, 2 ml DryStrip Cover Fluid (GE Healthcare, Bucks UK) was pipetted over the strip. Strips were re-swelled for at least 6 hours or overnight.

2.3.4.3 First-dimension isoelectric focussing (IEF)

Isoelectric focussing of mini (7 cm) strips was performed with a Multiphor II IEF unit (GE Healthcare, Bucks UK). This unit contains a ceramic plate connected a circulating

Table 2.2: IEF running conditions for 7 cm pH 3-10 IPG strips

Step	Volts (V)	Current (μA / strip)	Power (W)	Time	Volt Hours (Vh)
1	200	50	5	0:01	-
2	3500	50	5	-	2800
3	3500	50	5	-	3700
Total					6500

water bath (Grant Instruments Ltd., Cambridgeshire UK) set to 20 °C cooling the IPG strip during focussing. IPG strips are secured in an electrode assembly consisting of a support frame, two electrode arms, and an IPG strip alignment guide to ensure correct placement of strips within the assembly. The electrode arms are connected to a power source allowing current to pass through the strips when the Multiphor II is set up.

Re-swelled strips were removed from the re-swelling tray and gently rinsed with ddH_2O to remove surface DryStrip Cover Fluid. The plastic side of the strip was dried with blotting paper and placed gel-side up on the alignment guide of the Multiphor II. Approximately 5 ml of DryStrip Cover Fluid was poured between the ceramic plate and the support frame, and the support frame and the alignment guide, to aid conductance of heat away from the strips. Electrode wicks were dampened with ddH_2O and blotted to remove the excess, before laying upon the IPG strips. The electrode arms were then attached to the support and the lid clipped in place. On closing the lid the circuit incorporating the IPG strips is completed. A programmable power unit was used set up with the programme described in Table 2.2, and the programme allowed to run to completion.

2.3.4.4 Equilibration

Equilibration of IPG strips is a two-step process and takes place between IEF and the second dimension SDS PAGE separation. It serves to incorporate SDS into the strip in preparation for the second dimension separation, as well as maintaining reducing conditions (DTT) and alkylating thiol groups (iodoacetamide) to prevent their re-oxidation during electrophoresis. Following IEF, strips are removed from the Multiphor II unit and rinsed with ddH_2O to remove surface DryStrip Cover Fluid. 2 ml of equilibration buffer was pipetted into a channel of the equilibration tray (GE Healthcare, Bucks UK). Equilibration buffer (30% (v/v) glycerol, 2% (w/v) SDS, 50mM Tris-HCl pH 8.8, 0.002% (w/v) bromophenol blue) was pre-made, stored at -20 °C, and on usage either 1% (w/v) DTT or 4.8% iodoacetamide was added. Strips were incubated under gentle agitation for 15 minutes at room temperature in 1% DTT equilibration buffer. After this time, the strip was transferred to a channel containing 4.8% iodoacetamide equilibration buffer and incubated under the same conditions for a further 15 minutes.

2.3.4.5 Second-dimension electrophoresis

Gels for mini 2-dimensional SDS PAGE were set up in a similar way to 1D SDS PAGE except they did not contain a stacking gel section and were 1 mm in thickness (see Chapter 2.3.3.2). 12% acrylamide gels were poured (12% acrylamide (acrylamide:bis-acrylamide 37.5:1) (Bio-Rad Laboratories, Hercules, US), 375 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.05% (w/v) ammonium persulphate (Bio-Rad Laboratories, Hercules, US), and 0.2% (v/v) TEMED) to approximately 0.8 mm from the top of the glass cassette. Butan-1-ol was pipetted on to the gel solution surface to provide a smooth horizontal surface. On polymerisation, butan-1-ol was tipped away and the gel surface washed with ddH_2O and dried with blotting paper. Further ddH_2O was then pipetted onto the gel up to the top of the glass cassette. The equilibrated IPG strip could then be lowered down upon the gel surface, taking care so not to introduce air

bubbles between the IPG strip and the acrylamide gel. Once the IPG strip was in place, the water was then gently tipped away so as not to disturb the IPG strip or introduce air bubbles. IPG strips were then sealed in place with molten agarose sealing solution (1% (w/v) low melting point agarose, 0.002% (w/v) bromophenol blue in Tris-glycine SDS electrophoresis buffer). This was then left to set before electrophoresis was performed in the same manner as for mini 1D SDS PAGE (Chapter 2.3.3.4).

2.3.5 2-Dimensional in-Gel Electrophoresis

2-Dimensional in-Gel Electrophoresis (2D DIGE) is a fluorescence-based multiplexed proteomics platform (Tonge et al., 2001b) commercialised by Amersham Bioscience (now GE Healthcare). It allows for highly accurate quantitation of protein levels and identification of statistically significant protein changes between control and treated groups of sample. Although there are broad similarities in proteomic methodology between 2D DIGE and traditional 2DE, in the laboratory a number of pieces of apparatus including IEF equipment and second dimension tanks are used specifically for 2D DIGE and as such will be considered separately in this section.

2.3.5.1 Sample preparation for CyDye labelling

For the successful labelling of proteins the pH of the lysis buffer solubilised sample should be between pH 8.0-9.0. Prior to labelling the pH of the protein sample was tested by aliquoting a small volume (0.5-1.0 μl) of sample onto pH indicator strips (pH 7-14). This was adjusted as required with 0.1M NaOH to bring the sample into the region of pH 8.0-9.0. *R. communis* ER protein samples are dominated by highly abundant acidic proteins so it was found that if adjustment was required the pH would need to be increased through the addition of 0.1M NaOH. This is a critical stage of the method, as the pH will strongly affect the success of CyDye labelling.

2.3.5.2 Creation of CyDye stock solutions

CyDye reagent comes in 10 nmol batches for each dye. This needs to be reconstituted in anhydrous N,N-dimethylformamide (DMF), 99.8% (Sigma Order No. 22705-6) to create working stock solutions. As normal practice, new bottles of DMF were used with each new batch of dye purchased. Two levels of CyDye stock solution were used: the primary stock and working stock. Vials containing the dyes to be reconstituted were removed from the freezer and incubated at room temperature for at least 5 minutes to warm up. The vials were briefly centrifuged to bring the powder to the bottom before opening. The contents of the vial was then resuspended in 10 μl of DMF to obtain the primary stock, at a concentration of 1 nmol/μl (1 mM). The primary stock is stable for a maximum of 2 months at -20 °C. The tubes were covered with parafilm, stored in a bottle containing desiccant and then heat-sealed in a plastic bag to minimise moisture exposure.

For the creation of working stocks, 2.4 μl of DMF was aliquoted into PCR tubes labelled Cy2, Cy3 and Cy5. 1.6 μl of dye was pipetted into the appropriate labelled tube to give a final volume of 4 μl and vortexed. This gave the working stock at a dilution of 400 pmol/μl (0.04 mM).

2.3.5.3 CyDye labelling

Protein samples adjusted to the correct pH were quantified with the modified Bradford procedure. 1 μl of working stock was added to 50 μg of protein sample at a minimum concentration of 1 μg/μl.

2.3.6 Large format 2DE

Two separate large format 2DE systems were used in this study, either a Multiphor II IEF unit and Hoefer Dalt second dimension electrophoresis tank or an IPGPhor IEF unit and DaltTwelve second dimension tank.

2.3.6.1 Sample loading and first dimension isoelectric focussing

For analytical gels cup loading is the recommended method of sample introduction. 18 cm pH 3-10 Immobiline DryStrips (GE Healthcare, Bucks UK) were re-swelled overnight in 350 μl 2DE lysis buffer (9M urea, 2M thiourea, 4% CHAPS, 2% pH 3-10 ampholytes (GE Healthcare, Bucks UK), 1% DTT, trace bromophenol blue). The technical method for re-swelling strips is as described in Section 2.3.4.2 except the re-swelling solution did not contain any protein sample.

After re-swelling strips, they were removed with forceps by gently peeling the strip up from the re-swelling tray. The plastic back of the strip was wiped dry of cover oil against blotting paper. The strip was then placed gel side-up in the ceramic IPG strip holder of the IPGPhor IEF apparatus. Electrode wicks (GE Healthcare, Bucks UK) were cut to 1.2 cm in length, submerged in ddH_2O briefly and the excess blotted away before being placed on each end of the strip, so that there was no revealing IPG at the end of the strip. The sample loading cup was placed upon the IPG strip at the anodic end. The cup features an invagination so that it sits tightly across the strip minimising the chance of sample leakage. Care was taken to ensure the bottom of the cup was true to the gel on the strip. The sample cup was then filled with 70 μl of DryStrip cover oil (GE Healthcare, Bucks UK) and the level of oil within the cup was monitored to ensure that there were no leaks. This is critical to minimise risk of sample loss. If no leaks have occurred after 5 minutes, the protein sample is pipetted into the cup by lowering the pipette tip beneath the surface of the oil. The sample displaces the oil which in turn prevents the samples crystallisation. A further 4 ml of DryStrip cover oil was then pipetted across the surface of IPG strip. Care was taken not to disturb the sample loading cup with the force of oil by pipetting some distance away and allowing the oil to run across the strip. Finally, electrode arms are clipped in place at the anode and cathode connecting the terminals on the IPGPhor to the electrode wicks sitting upon the gel, completing the circuit.

Table 2.3: IEF running conditions for 18 cm pH 3-10 IPG strip on the IPGPhor IEF apparatus

Step	Volts (V)	Volt hours (Vh)	Time [a]	Current
1	500	41	00:10	Gradient
2	1000	1000	01:20	Gradient
3	4000	4166	01:40	Gradient
4	6500	65000	10:00	Step-n-Hold
5	1000	-	60:00	Step-n-Hold

[a] Time is in excess, run length determined by volt hours (run complete at around 70,000Vh)

The IPGPhor was pre-programmed with the protocol for 18 cm pH 3-10 strips, which is detailed in Table 2.3. The length of time for focussing was determined by volt hours (Vh), the run being terminated after 65,000 - 70,000 Vh.

2.3.6.2 Large format gel casting

12% homogenous second dimension gels were cast using the Ettan Dalt gel casting apparatus. The gel caster was cleaned with Teepol^{TM}, rinsed thoroughly with ddH_2O to remove any polymerised acrylamide and dried. Glass plates were cleaned by 1) scrubbing then soaking in 1% Decon^{TM} (v/v) for 1 hour, 2) rinsing thoroughly in ddH_2O, 3) soaking in 1% HCl (v/v) for 1 hour, 4) thoroughly rinsing in ddH_2O before drying with lint-free Crew Wipes (Kimberley Clarke). Plate dimensions were 260 x 200 x 1 mm for low-fluorescent plates destined for imaging of CyDye fluors or 260 x 200 x 1.5 mm for preparative gels. Gel solutions were made up as per Table 2.4 without APS and TEMED. The gel caster was assembled containing the required number of glass cassettes and spacers and a plastic funnel was attached to the inlet of the gel casting tank for gel solution introduction. Displacing solution (375mM Tris-HCl pH 8.8, 50%

Table 2.4: Constituents of 12% homogenous large format acrylamide gels

Constituent	Volume (ml)	Final Concentration
40% (v/v) acrylamide [b]	300	12%
10% (v/v) SDS	10	0.1%
1.5 M Tris-HCl pH 8.8	250	375 mM
ddH_2O	434	-
10% APS [a]	5	0.05%
TEMED [a]	1	0.1%
TOTAL	1000	

[a] APS and TEMED added just prior to casting to ensure polymerisation doesn't begin until gel solution has been poured

[b] Acrylamide:bis-acrylamide = 37:1

glycerol, 0.002% (w/v) bromophenol blue) was added to clear the tubing of gel solution prior to polymerisation and water saturated butan-1-ol pipetted onto the surface of the gel solution to give a smooth horizontal surface and to prevent drying of the gel. Gels were left overnight to ensure the polymerisation reaction had completed.

2.3.6.3 Preparation of backed gels

Backed gels provided the advantage that the acrylamide gel remained adhered to the glass plate surface allowing for much easier handling but still allowing the gel to be stained post-electrophoretically when required and picked using an automated robot. Gels were cast as normal except one side of the glass cassette to contain a backed gel was smeared with 2-4 ml of Bind-Silane solution (80% (v/v) ethanol, 0.2% acetic acid, 0.01% (v/v) Plus One Bind-Silane (GE Healthcare, Bucks UK) before cassette assembly. An even distribution of the Bind-Silane solution was applied with a lint free

Crew wipe (Kimberley Clarke), which also served to reduce incorporating dust into the solution and on the plate.

2.3.6.4 Second dimension separation

Following focussing strips were equilibrated as described (Section 2.3.4.4) and separated using two alternative second dimension systems, the Hoefer Dalt large format tank or the Ettan DaltTwelve tank (GE Healthcare, Bucks UK). Equilibrated strips were placed upon the polymerised gel surface and sealed in place with molten agarose (1% (w/v) low melting point agarose, 0.002% (w/v) bromophenol blue in Tris-glycine SDS electrophoresis buffer). ddH_2O was pipetted onto the gel up to the top of the glass cassette. The equilibrated IPG strip could then be lowered down upon the gel surface, taking care not to introduce air bubbles between the IPG strip and the acrylamide gel. Once the IPG strip was in place, the water was then gently tipped away so as not to disturb the IPG strip or introduce air bubbles. Molten agarose was then pipetted onto the IPG strip and this was left to set, sealing the IPG strip in place.

Second dimension electrophoresis in the Hoefer Dalt tank required 20 litres of 1x SDS running buffer (14.4% (w/v) glycine, 3.0% (w/v) tris, 1.0% SDS). Temperature was maintained at 25 °C by a circulating water bath connected to the tank's cooling frame, and gels were separated for approximately 15 hours at 20 mA per gel. For the Ettan DaltTwelve tank, 7.5 litres of 1x SDS running buffer was required in the lower chamber. Gels were inserted into the lower chamber before a further 2.5 litres of 2x SDS running buffer was added. The temperature was maintained at 30 °C for 15 hour separations at 2 W per gel, or 25 °C for shortened separations, requiring 5 W per gel for 30 minutes followed by 17 W per gel for 4 hours. The electrophoresis separation was halted when the bromophenol blue dye front had reached the bottom of the gel.

Table 2.5: Composition of Coomassie Brilliant Blue R250 solutions

Substance	Coomassie			Destain
	1	2	3	
1.25% Coomassie Brilliant Blue Stock	2%	0.25%	0.25%	-
Propan-2-ol	25%	10%	-	-
Glacial acetic acid	10%	10%	10%	10%
Glycerol	-	-	-	10%

2.3.7 Protein stains

Gels were removed from glass plates taking care not to contaminate the gel surface with dust which can cause staining artefacts and contamination of mass spectra. Large format gels which had not been backed to the gel plate were handled with extreme care due to their highly fragile nature. Where a gel was destined for picking and mass spectrometry analysis transferral of gel to stain would be carried out in a laminar flow hood to reduce the risk of keratin contamination. Staining vessels were thoroughly cleaned before use.

2.3.7.1 Coomassie Brilliant Blue R250 stain

Gels were sequentially stained in solutions Coomassie I, Coomassie II and Coomassie III, leaving the gels in each stain for 20 minutes or 2 minutes when the staining solutions were heated in a microwave. Alternatively gels could be left overnight in Coomassie I. After staining in Coomassie III solution gels were transferred to destain for a further 20 minutes or 2 minutes with microwave heating, which increased the contrast between stained proteins and the gel background. Coomassie solutions were made up as follows:

A volume of 30 ml for each staining step was typically used for mini gels, or 200 ml for large format gels.

Table 2.6: Composition of disruptive silver stain solutions

Step	Constituents	Time
Fix	40% ethanol, 10% acetic acid	30 minutes
Sensitisation	30% ethanol, 6.8% sodium acetate, 0.13% glutaraldehyde, 0.2% sodium thiosulphate	30 minutes
Wash	distilled water	3 x 5 minutes
Silver	0.1% silver nitrate, 0.008% formaldehyde	40 minutes
Wash	distilled water	2 x 1 minutes
Development	2.5% sodium carbonate, 0.004% formaldehyde	5 - 15 minutes
Stop	1.46% EDTA	5 - 10 minutes
Wash	distilled water	2 x 10 minutes

2.3.7.2 Disruptive silver stain

Disruptive silver staining was performed using a method adapted from Heukeshoven and Dernik (Heukeshoven and Dernick, 1988). All solutions were made up as required (Table 2.6). Formaldehyde and glutaraldehyde were added just before use. Staining was carried out in glass tubs under gentle agitation on an orbital shaker, and extra care was taken to ensure tubs were completely clean before use. Tubs were first scrubbed with Teepol^{TM} (Kent, UK), then rinsed with multiple changes of distilled water, before washing with 70% ethanol and incubating the tub with 1% H_2SO_4 for 30 minutes. Without rigourous cleaning staining efficiency was sometimes reduced through a likely sequestering of stain components to the tub walls. Gels were incubated in the developing solution until protein spots were stained but the background remained clear. As soon as darkening of the gel background occurred the gel was transferred to stop solution to quench the reaction. Gels were finally extensively washed in distilled water before scanning and storage.

Table 2.7: Composition of MS-compatible silver stain solutions

Step	Constituents	Time
Fix 1	40% (v/v) methanol, 10% (v/v) acetic acid	30 minutes
Fix 2	40% (v/v) methanol, 10% (v/v) acetic acid	30 minutes
Sensitisation	30% (v/v) methanol, 6.8% (w/v) sodium acetate, 0.2% (w/v) sodium thiosulphate	30 minutes
Wash	distilled water	3 x 5 minutes
Silver	0.5% (w/v) silver nitrate	40 minutes
Wash	distilled water	2 x 1 minutes
Development	2.5% (w/v) sodium carbonate, 0.008% (v/v) formaldehyde	5 - 15 minutes
Stop	1.46% EDTA (w/v)	5 - 10 minutes
Wash	distilled water	2 x 10 minutes

2.3.7.3 MS-compatible silver stain

Where gels were destined for mass spectrometry analysis an alternative MS-compatible silver stain was used which lacked the peptide cross-linking chemical glutaraldehyde. Unlike with disruptive silver staining there are two fixing steps performed with fresh methanol fixative and the silver incubation lacks formaldehyde but the developer contains twice its concentration. Other than these deviations, the method is the same as for disruptive silver staining. The method is similar to that of Shevchenko's modified silver stain (Shevchenko et al., 1996b).

2.3.7.4 SYPRO™ Ruby stain

SYPRO™ Ruby (Genomic Solutions Ltd., Cambridgeshire UK) is a commercial MS-compatible fluorescent stain. The staining procedure is as follows: gels are first incubated in fixative (40% methanol 10% acetic acid) for 30 minutes under gentle agitation on an orbital shaker. They are then transferred to the SYPRO™ staining solution, which is pre-filtered before use to remove any crystallised dye. Gels were stained overnight in light proof polycarbonate tubs under gentle agitation before rinsing in

distilled water and incubating in a destain solution (10% (v/v) methanol, 6% (v/v) acetic acid) for 2 hours. Gels were imaged as described (Chapter 2.3.8.1, 2.3.8.2) and stored (Chapter 2.3.8.3).

2.3.8 Gel documentation

2.3.8.1 ProXPRESS

The ProXPRESS imaging system provides 16 bit 100 μm resolution images via a cooled charge-coupled device (CCD) camera and allows imaging of fluorescent stains (SYPRO^{TM}, CyDye), visible stains (Coomassie, silver) and chemiluminescence. The imager collects a number of high resolution image sections before digitally 'stitching' the sections together to produce the final image. As a result the image can suffer from uneven illumination at the interface of the sections. This is corrected with a field image that is collected at regular intervals of ProXPRESS operation to account for drift in the imagers illumination and image capture hardware. For the imaging of fluorescent gels a flat field image was collected using a green acrylic flat field sheet, imaged at an emission wavelength of 620 nm, an excitation wavelength of 480 nm and a resolution of 100 μm. The exposure time was adjusted until a peak pixel value of the 40-50,000 pixels was obtained. The same imaging parameters were applied to the imaging of fluorescent gels, with the exposure time being adjusted to ensure good use of dynamic range without causing saturation to any part of the gel image.

For the imaging of visible stain gels a UV to white light converter plate was first installed. A flat field image was collected from the red acrylic flat field sheet using an emission wavelength of 530 nm and a resolution of 100 μm. This image was used to correct gel images collected under the same parameters and using a exposure time adjusted to ensure a peak pixel value between 40,000 to 50,000.

Gels were saved as .TIF files.

2.3.8.2 Typhoon 9200 Variable Mode Imager

For the aquisition of fluorescent wavelength gel images, a Typhoon 9200 Variable Mode imager was used. The scanner contains multiple emission filters and lasers for the imaging of a range of CyDye and fluorescent post-stain wavelengths, including SYPRO, and uses a photo multiplier tube (PMT) for image aquisition. The instrument was switched on 30 minutes before use to allow the scanner to warm up, until the instrument status light has changed from flashing to solid green. If Cy2 fluors were to be scanned the separate blue laser unit was switched on prior to the Typhoon instrument. The glass platen surface of the scanner was cleaned before scanning was performed to ensure that no dirt or dust would effect the obtained image. Generally ddH_2O and lint free tissue was used, although a combination of 70% ethanol followed by ddH_2O was used if this was not sufficient. Analytical gels were scanned within their low fluorescent glass plates minimising potential gel damage. Correct positioning of gels on the scanner was aided by the front location bar and grippers of the gel alignment guides. One edge of the dried glass cassette was placed against the front location bar, whilst the gripper was used to gently lower the glass cassette down upon the platen. The location of the gel cassette on the platen was noted by comparing against graduations on the platen edge, and this was inputted into the Scanner Control Software. The focal plane of the scanner could be set to 0 mm or 3 mm above the platen surface, allowing the scanning of loose gels or gels within glass cassettes respectively. Within the Scanner Control Software the image aquisition mode was set to fluoresence. Fluorescent scanning parameters were set up depending on the gel stain / dye type, using the combination of emission filters and lasers detailed in Table 2.3.8.2. For multiplex (CyDye) gels, multiple channel scans would typically be used. This allowed the sequential automated imaging of a variety of wavelengths to be performed from a single gel without further manual intervention.

Scans of new gels were initially obtained at 1000 μm. This low resolution scan allowed for quick preview images to be obtained. PMT voltages were adjusted so that

Table 2.8: Scanning parameters for the Typhoon Variable Mode Imager

Fluorophore	Emission (nm)	Laser
Cy2	520 BP 40	Blue2 (488)
Cy3	580 BP 30	Green (532)
Cy5	670 BP 30	Red (633)
SYPRO Ruby	610 BP 30	Blue1 (457)

the maximum pixel intensity on the preview image fell between 50,000 and 60,000 levels of grey out of 100,000. Once PMT values were obtained for each fluor to be scanned in a gel, a high resolution 100 μm scan could be carried out. Care was taken to ensure no image saturation occured by checking the peak pixel value of the obtained image. If saturation had occurred within the region of interest, the scan would be re-aquired with a lower PMT value. Images were saved as .gel files, a modified .tif file format allowing superior dynamic range, and cropped within the ImageQuant (GE Healthcare, Bucks UK) software.

2.3.8.3 Archival of 2DE gels

For the longer term storage of mini 2DE gels, they were dried between two sheets of pre-soaked cellophane in a frame positioned within an Easy Breeze^{TM} (Hoefer Scientific Instruments) drying chamber. Non-backed large format gels were bagged within heat sealed transparent plastic bags containing 5-10 ml ddH_2O.

2.3.9 Computational gel analysis

2.3.9.1 DeCyder

The DeCyder software allows the analysis of 2D DIGE datasets to identify statistically signficant differences between differentially labelled images. It consists of a Differential In-gel Analysis (DIA) module which performs the spot detection and quantitation of

gel images from the same gel, the Biological Variance Analysis (BVA) module which utilises the Cy2 standard to match images from different gels and performs statistical analyses and the Batch Processor which automates the gel analysis pipeline.

Gels images were imported into the DIA module either manually or via the batch processor. Spot detection was performed and for *R. communis* gels this was found to be optimum with a predicted spot number of 5000. After the spots on all gels were co-detected within the DIA module, normalisation and background subtraction were performed. Abundances for each spot were quantified and a log volume ratio figure calculated against the same spot on the Cy2 standard gel.

The DIA module was also used in isolation for the pairwise comparison of gel images from two co-migrated Cy labelled protein samples, allowing normalised fold change values to be produced for spots within a single gel.

For a complete 2D DIGE analysis, the BVA module was used for inter-gel matching, and the statistical comparison of spot volume ratios across gels to identify spots of interest. Spots maps and quantification data generated by the DIA module were imported into the BVA module, either manually or automatically with the Batch Processor. Spot maps within the experiment were assigned to different analysis groups, either Germinating, Developing or Standard. The gel with the most spots was automatically assigned the Master gel and the Cy2 standard spot maps of the other gels in the experiment were matched to the spot map of the Master Cy2 gel. This was performed with the automated BVA matching algorithm, after which every match was manually confirmed for accuracy and adjusted where required. The 'Split Spots' and 'Merge Spots' tools were used to achieve parity in the spot detection boundaries across the gels. The volume ratios generated in the DIA module were then compared as log volume ratios. An unpaired t-test analysis was used to identify those spots that were significantly upregulated by 10% in the germinating and developing gel sets, with a p=0.02 confidence value and the criteria that only those spots present in every gel were

included. This created a subset of spots within the germinating and developing gel sets that were signficantly different between the two states, allowing for their downstream analysis.

2.3.9.2 Phoretix Evolution

Phoretix Evolution (Nonlinear Dynamics, Northumberland, UK) was used for the routine analysis of 2DE gels that were not part of a complete 2D DIGE experiment. It allowed the analysis of a variety of gel formats and stains and incorporated warping algorithms to aid matching between different gels. Gels were imported into the software and spots were detected with the Evolution detection algorithm. For multiple gel analyses requiring matching across gels a reference spot map was created from the spot map of the most populous gel within the experiment. Spots on each gel were first matched to the reference spot map using the software's automated matching algorithms before assessment and manual adjustment as required. To improve the accuracy of quantification and allow the comparison of spot values between gels, adjustments were made to the spot volume data: background subtraction was performed with the 'mode of non-spot' algorithm to reduce the effects of differential background levels on spot volumes and spots were normalised using the total spot volume x 100 parameter. A pre-matching warping procedure was used to increase effectiveness of matching gels where standard matching had failed.

2.3.10 Spot excision

Protein spots were excised from 2DE gels using a ProPic spot picking robot (Genomic Solutions, Cambridgeshire, UK). The robot utilises in-built gel imaging equipment allowing spots to be selected within the robot's software or for co-ordinates from an existing analysis to be triangulated to the robot-generated image. The system was initially primed with ddH_2O for around 30 seconds until all air was removed from the

water supply lines. The gel was either placed directly onto the low fluorescence glass platen of the robot or if picking from a backed gel was affixed to the glass platen by means of adhesive tabs. The gel was then imaged, selecting the correct robot illumination mode for either fluorescent or white light gels. For picking more than just a few spots, triangulation was employed. The gel to be picked was given external triangulation landmark locations before its initial imaging on the Typhoon Variable Mode imager. On imaging the gels within the ProPic robot, the triangulation markers were easily identified, allowing them to be mapped to the same locations on the earlier Typhoon image and for pre-assigned spot pick locations selected within the Phoretix software to be automatically and accurately transposed.

2.4 Yeast methods

2.4.1 Growth and harvest of *Y. lipolytica* cells

The Yarrowia lipolytica yeast strain PO1G (MAT**a**; *leu2-270*; *xpr2-333*; *axp1-2*) was used throughout this experimentation. This strain is derived from the wildtype strain W29 but is a leucine auxotroph. 100 ml starter cultures were initiated from plate and grown on YPD complete media (10 g/litre yeast extract, 20 g/litre bactopeptone, 20 g/litre glucose) in 250 ml baffled conical flasks. Starter cultures were used to inoculate 500 ml experimental minimal media cultures in 2 litre baffled conical flasks. Minimal media contained 1.7 g/litre yeast nitrogen base (without amino acids) (Difco, Detroit, MI, USA), 5 g/litre NH_4Cl, 1 g/litre yeast extract, 50mM phosphate buffer pH 6.8. 0.1g/litre Uracil and 0.3g/litre leucine was added after autoclaving by filter sterilisation. Media were supplemented with either 2% glucose (from filter sterilised 20% w/v glucose stock) or 1% fatty acid from the relevant fatty acid stock solution. Fatty acid stock solutions were either 10% oleic acid, 10% methyl oleate or 10% methyl ricinoleate, sonicated in the presence of 1% Tween 80 using a Soniprep 150 preparative sonicator

(MSE Scientific Instruments, Crawley, UK) for 3x 1 minute at an amplitude of 22 microns on ice.

Cells were grown in an orbital shaker / incubator (Infors HT, Crewe UK) at 28°C, 140 rpm rotation. Harvests of well mixed cells from *Y. lipolytica* cultures were taken at 2, 3, 5 and 10 hours by removal of 50 ml of culture media and centrifugation at 3000 g in a Beckman JLite rotor. Following harvesting, cells grown in lipid media were gently resuspended in 100ml 0.5% BSA solution to remove any lipid retained on the surfaces of the cells or the centrifuge tubes. This step was repeated three times, followed by a final wash in 100 ml 0.5% BSA and 0.9% NaCl. Cells were resuspended in 50 ml ddH_2O as a final wash and transferred to pre-weighed and labelled Falcon tubes prior to pelleting in a Jouan bench top rotor (5 minutes, 3000 g, 4 °C), carefully pouring off the supernatant and recording the wet weight of the cells. Cells grown in glucose media were harvested in the same way accept they were washed once with ddH_2O. Cells were snap frozen in liquid nitrogen and stored at -20 °C so that lipid extractions could be performed together.

2.4.2 Measurement of growth

Growth was recorded by measuring the optical density (OD) of the cell culture using a spectrophotometer (GE Heathcare, Bucks, UK) at an absorbence of 600 nm. Cell culture was diluted to ensure that the measurement reading was within the dynamic range of the spectrophotometer (between 0-0.9). For the measurement of cell growth in cultures containing fed lipid, samples were first washed with 3x 0.5% BSA and 1x ddH_2O.

2.4.3 Extraction of lipid from *Y. lipolytica* cells

Frozen pellets of *Yarrowia* harvested cells were thawed gently on ice before resuspension in isopropanol to a concentration of 1g wet weight / 20 ml isopropanol. At this initial

stage of lipid extraction 0.3 mg of the TAG standard triheptadecanoic acid was added so that any losses of lipid could be accounted for and for quantification. An equal volume of chloroform was added to the isopropanol in each sample before homogenisation with a polytron (IKA, Staufen, Germany) for 10 seconds, followed by a 10 second rest on ice (repeated 3 times). Homogenised samples were then filtered through Whatman No. 1 (Whatman, Maidstone, UK) paper assembled within a Buchner funnel to remove cellular debris from the homogenate. The filtrate for each sample was then split between 6 round-bottomed Corex tubes and dried down under a stream of N_2 gas on a heating block maintained at 60 °C until a dry residue was obtained. The residue was then resuspended in 1 ml of 2:1 chloroform : methanol per tube and each sample type pooled to give a total of 6 ml per lipid extraction sample. This was filtered once more using Whatman No. 1 paper and a Buchner funnel and the final volume adjusted to 8 ml with more 2:1 chloroform : methanol. To this was added 2 ml of 0.88% KCl. The sample was inverted 5 times and span in a Sigma benchtop centrifuge for 5 minutes (2500 x *g*, room temperature) to obtain a phase partition, the upper being the aqueous portion. This upper phase was carefully removed with a Pasteur pipette leaving the organic lipid solvent. A further 1.4 ml of 0.88% KCl was added to the organic phase to aid separation of any remaining aqueous soluble components. This was again inverted 5 times and centrifuged under the same conditions as the first partitioning spin prior to removal of the upper aqueous phase.

2.4.4 Derivatisation of lipid samples

An aliquot of lipid sample was transferred to a clean glass tube and dried down under N_2 to obtain a pellet. The pellet was then resuspended in 250 μl pyridine. To this was added 250 μl *N,O* -Bis(trimethylsilyl)trifluoroacetamide (BSTFA) and 12.5 μl trimethylchlorosilane (TMCS). The derivatisation mixture was then heated for 30 minutes at 70 °C before cooling to room temperature and extracting silylated

compounds from the derivitisation mixture by the addition of 4 ml hexane, 1 ml 0.9% NaCl, 4 ml ddH_2O. Samples were dried down under N_2 and resuspended in hexane.

2.4.5 Thin layer chromatography of lipid samples

2.4.5.1 Basic thin layer chromatography method

Purified lipid was fractionated by thin layer chromatography (TLC) using Whatman (Maidstone, Kent) silica gel TLC plates with fluorescent indicator. Lipid sample was dotted onto the base of the plate, using typically 5 μl for analytical TLC plates and 50 μl for preparative TLC plates, drying the organic solvent away from the plate using a hair dryer to contain the lipid spot on the plate. A neutral lipid standard was dotted on to the plate containing triacylglycerol, monoacylglycerol and two diacylglycerol species. Plates were transferred to a TLC tank containing 100 ml of developing solvent. A range of TLC solvent systems were used in this study. The plates were removed from the tank when the solvent front neared the top two inches of the plate and this position was marked with a pencil. Analytical TLC plates were stained with phloxine B stain (0.001% phloxine B in 50% ethanol) and visualised under UV light. Preparative TLC plates were not stained with phloxine B but the TAG band was localised using iodine under a stream of N_2 gas. Where further gas chromatographic (GC) analysis of TAG components was required the localised TAG band on the TLC plate was removed by scraping the silica matrix away from the glass backing using a razor blade. Great care was taken not to scrape away any adjacent lipid components present on the TLC plate to minimise the risk of sample cross-contamination. The isolated TAG was extracted from the silica matrix by the addition of 1ml 2:1 chloroform : methanol, vortexing and removal of lipid containing solvent.

2.4.6 Conversion of silylated TAG components to fatty acid methyl esters

Following derivatisation of lipid samples, fractionation by preparative TLC, and removal and extraction of TAG bands from the TLC matrix, the solvent was dried down under N_2 gas on a heating block to obtain a lipid residue. This lipid sample was re-suspended in 1 ml 1% H_2SO_4 in methanol and heated overnight at 50 °C in an acid catalysed trans-esterification reaction. Following this reaction the methyl esters were extracted from the methylating reagent with the addition of 1 ml hexane and 0.5 ml 0.9 % NaCl. A phase partition was obtained and the upper lipid-containing organic phase removed. Finally the sample was transferred to a GC vial and dried down to a volume of 100 μl. Samples were immediately analysed by GC or stored at -20 °C in the dark until required.

2.4.7 Gas chromatography analysis

GC analysis was performed with a Shimadzu GC-14 gas chromatograph and a Shimadzu AOC-20s autosampler. The GC column was an Alltech Econo-Cap EC WAX 30 m x 0.25 mm x 0.25 μm column. The carrier gas used was nitrogen at a pressure of 1.0 kg/cm^2. The column was pre-heated to 160 °c and the injector and detector heated to 250 °C and 270 °C respectively. After the GC components had reached their correct operating temperatures a test run was performed by injecting 2 μl of hexane. This is to test that the GC is functioning correctly and that the column is clean (i.e. no peaks observed on the GC spectra). The GC was calibrated with a commercially available standard, GC96 (Nu-Chek, Elysian MN, US) which contains a range of 20 standard fatty acids incorporating myristic acid (C14:0) through palmitic (C16:0), palmitoleic (C16:1), stearic (C18:0), oleic (C18:1), linoleic (C18:2), linolenic (C18:2) to methyl nervonate (C24:1).

The identity of each peak was assigned by evaluating its percentage proportion

compared to the whole GC spectra, and comparing this to the known percentage proportion within the standard mixture as advised by the manufacturer. Further confidence in assignations of standard identity was obtained by examining the retention time and order of species as they came off the GC and again comparing this data to the guidelines supplied by the manufacturer. Following calibration of the GC with the GC96 standard experimental samples were analysed. 2 μl of each experimental sample were injected into the GC. A batch process was set up to automate the injection, and the GC96 standard re-acquired to measure any drift in retention time of the column toward lipid species over time.

Chapter 3

Establishment of growth, preparation and proteomic methodologies for the characterisation of the endoplasmic reticulum from *Ricininus communis* cv. 99N89I endosperm

3.1 Introduction

Proteomics was conceived as the identification of all proteins expressed by the genome in an organism, cell or tissue (Wasinger et al., 1995a) but is more accurately defined as the identification of all proteins expressed by the genome in an organism, cell or tissue at a given time point and under specified conditions. Improvements to 2DE such as commercially available immobilised pH gradients overcame earlier technical problems and allowed protein spot maps to be compared within and between laboratories (Corbett et al., 1994) facilitating its deployment as a reliable analytical procedure. The technology matured with the development of pre-electrophoretic CyDye labelling of proteins (Tonge et al., 2001b) and the availability of the 2D DIGE platform. Furthermore, improvements to downstream analysis procedures such as MALDI TOF and MS/MS mass spectrometry alongside growing protein databases and improved bioinformatic tools are allowing many more proteins to be identified.

Technical challenges remain however and analysing certain sample types by 2DE can present challenges. It is important to evaluate 2DE experimental procedures with the sample of interest as this can inform the selection of protocols most amenable to a successful analysis. For example, membrane and total soluble samples are liable to behave differently in a 2DE analysis and methodologies might need to be optimised to account for this.

Described in this chapter is a comprehensive study of the proteomic conditions for successful analysis of *R. communis* germinating and developing ER, with a view to maximising the data obtained in a multi-replicate 2D DIGE differential analysis of the two sample types.

3.2 Aims

The first aim of this Chapter was to establish routine growth, harvest and preparation procedures to create a dependable reproducible source of germinating and developing *R. communis* seed ER material. The second aim was to establish proteomic methodologies for the analysis of *R. communis* seed ER proteins, specifically the evaluation of CyDye labelling with seed ER and the establishment and evaluation of high loading gel picking strategies.

3.3 Results

3.3.1 Establishment of growth, harvest and sample preparation procedures with a view to creating a dependable, reproducible source of germinating and developing *R. communis* seed ER material

3.3.1.1 Evaluation of the germination and seed development characteristics of *R. communis* variety 99N891

A dependable supply of germinating and developing *R. communis* seed material was essential to the success of this investigation as it is this seed material which will be used to prepare ER samples; the ER being the location of TAG biosynthesis machinery in the developing seed. Initial proteomic studies previously performed in the laboratory utilised variety *R. communis* cv. Hale in which oleate $\triangle 12$ hydroxylase was identified as a membrane bound component of the ER (Maltman et al., 2002). This seed variety had become obsolete within the *R. communis* research community and was becoming difficult to obtain in sufficient quantities. Limited growth room facilities meant generation of large quantities of mature seed in-house was impractical, therefore an alternative variety was required that was amenable to research purposes: that is, was available in sufficient and reliable quantities and could be shipped on demand to

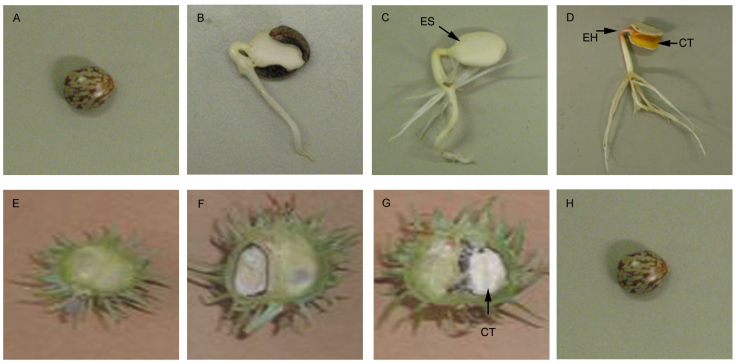


Figure 3.1: Endosperm development in germinating and developing *R. communis* seed. Figures A-D show germination at 0, 2, 3 and 5 days respectively. At 2 days (B) the endosperm is hard and solid, the seed coat is generally still attached and root growth is extensive but lacks the root hairs seen by Day 3. Day 3 of germination was the stage selected for endosperm ER preparation from cv. Hale. The endosperm has decreased in solidity from Day 2, but is still intact. The seed coat has generally been shed at this point. Day 5 of germination is characterised by endosperm separation and the emergence of an epicotyl hook (arrow EH) and cotyledons. The endosperm is thinner and depleted compared to Days 2 and 3. Figures E-H show different stages of seed development in *R. communis*. Figure E shows a sagittal section through a young fruit, at 5 DAF. By 10 DAF (Figure F) the seed coat is clearly visible but the seed lacks a defined endosperm or cotyledon at this point. Figure G shows a seed at 28 DAF, the upper point at which endosperm was harvested for ER preparations. At this stage a solid but wet endosperm almost reaches the seed coat and the cotyledon bisects the endosperm across its full extent (arrow CT). Figure H shows the mature seed. ES = endosperm, EH = epicotyl hook, CT = cotyledon.

the laboratory, that the seed would germinate efficiently and in a predictable uniform manner and that it would grow to produce new developing seed in a predictable pattern analogous to the staging descriptions for *R. communis* cv. Hale given by Greenwood and Bewley (Greenwood and Bewley, 1982). Critically, it must also be amenable to established ER preparation procedures (Maltman et al., 2002) giving suitable yields for downstream proteomic analysis.

R. communis cv. 99N89I was investigated as a potential replacement for cv. Hale. This variety, supplied by the research programme's industrial partner Arkema, was available in suitable and reliable quantities for projected duration of the research programme. The efficiency of *R. communis* cv. 99N89I seeds to germinate in growth

room conditions was investigated. Seeds were germinated under the same conditions as previously reported for cv. Hale (Maltman et al., 2002). Briefly, seeds were surface sterilised with 10% (v/v) hypochlorite and washed overnight in running tap water before planting in moistened vermiculite and transferred to a 30 °C growth room in the dark.

Figure 3.1 shows seedling germination at 2, 3 and 5 days of growth. Day 3 seedlings were characterised by an elongated but intact endosperm body that had not yet been separated by the emerging epicotyl hook and dicotyledon seen by day 5. Seeds were found to germinate with high efficiency and uniformity at 30 °C in vermiculite, with 88% of planted seeds reaching the same '3 day' stage of growth at 72 hours (see Figure 3.2).

Seeds grown to mature plants grew well under the conditions described (Chapter 2.2.1.1). At around 3 months all the plants had flowered with fruit appearing 5-10 days after flowering (DAF). Based on experience with cv. Hale, seed pods were selected for harvesting at 25-28 DAF. Pods were bisected to reveal the internal structure of the seed. Extent of development was measured by examining seed endosperm volume, liquidity, colour and the presence of the cotyledon. At 25-28 DAF cv. 99N89I seed endosperm was found to be solid, moist and white, but had not quite reached its full extent in making contact with the testa, equivalent to Stage V of Greenwood and Bewley's staging descriptions (Greenwood and Bewley, 1982) for cv. Hale. A morphologically distinct cotyledon first appears at Stage V of *R. communis* cv. Hale development allowing this to be a useful developmental marker. The development of the cotyledon in cv. 99N89I was found to follow the same pattern, reaching its full extent and fully bisecting the endosperm by 28-30 DAF (Figure 3.1, G). Studies into the biochemical aspects of *R. communis* seed development found that at 25-30 DAF peak lipid biosynthesis occurs at this stage, accompanied by a 6-fold increase in abundance of acetyl CoA carboxylase (Simcox et al., 1979). Evidence that optimal lipid

biosynthesis occurs at the same stage in 99N89I is supported by expression data from real time PCR experiments showing mRNA transcripts for the key Kennedy pathway TAG biosynthetic acyl transferase DAGAT2 genes are at their highest at this stage in this variety (Kroon et al., 2006). By 30 DAF endosperm fills the entire seed, described by Greenwood and Bewley as Stage VI. This stage is just after the optimal time for endosperm selection, i.e. when the endosperm has almost but not quite reached its full extent within the seed coat.

In summary, 99N89I seeds were found to germinate with high efficiency and uniformity. 99N89I seeds grown to mature plants produced developing fruit and examination of the fruit's endosperm found the pattern of developmental staging to closely follow that published previously for cv. Hale (Greenwood and Bewley, 1982). With the knowledge that raw endosperm material could be generated efficiently and predictably from 99N89I, the next question was whether endoplasmic reticulum could be purified from both developmental stages using the experimental procedure previously established in the laboratory.

3.3.1.2 Establishment of endosperm endoplasmic reticulum preparation procedure with 99N89I

Different strategies exist to reduce the complexity of proteomic analyses, such as fractionation of proteins into groups based on pI (Zuo and Speicher, 2002) or chromatography-based fractionations (Righetti et al., 2005). In this analysis the aim was to achieve proteome simplification through subcellular fractionation of the ER, purifying the component of the cell whose function is of direct biochemical relevance. A method for the purification of the ER from *R. communis* endosperm was first published by Coughlan *et al* (1996) in which a two-step density gradient fractionation of cell components was found to enrich ER membranes (based on ER enzyme marker assays for antimycin-insensitive NADH:cytochrome *c* reductase and CDPcholine:1,2-diacylglycerol choline phospho-

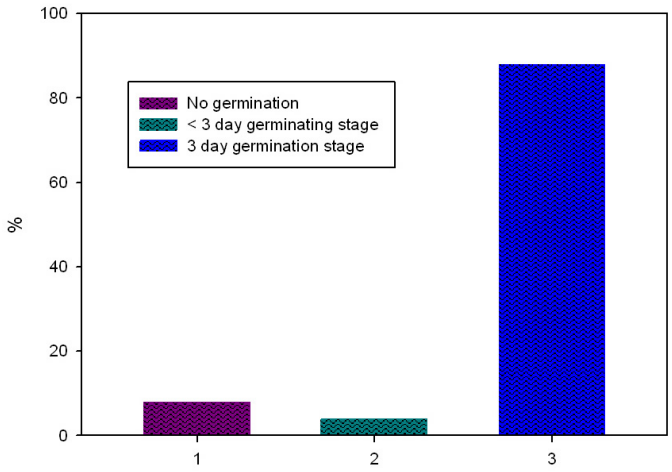


Figure 3.2: Graph to show efficiency of germination. Variety 99N89I was found to germinate efficiently in vermiculite at 30 °C. By 72 hours of growth, 88% of planted seedlings had reached the same 'Day 3' extent of germination, based on the criteria described in Figure 3.1. 8% of seeds did not germinate, 4% of seeds had germinated but had not reached the 'Day 3' stage.

transferase) at a density of 1.12 g/cm^3 (Coughlan et al., 1996). The preparation was devoid of contaminating endomembranes based on marker assays for Golgi (latent IDPase, glucan syntase I), mitochondria (fumarase, antimycin-sensitive NADH:cytochrome *c* reductase), glyoxysomes (catalase) and plastids (triosephosphate isomerase). Crucially, they employed an extended (24 hour) floatation centrifugation step to remove soluble proteins found by SDS PAGE to contaminate the ER after the first density gradient purification step. The two-step sucrose density gradient purification method was previously used in this laboratory to successfully isolate ER from *R. communis* cv. Hale developing and germinating seed endosperm (Maltman et al., 2002). In that study, the key ER-resident ricinoleic acid biosynthetic protein oleate $\triangle$12 hydroxylase was identified as a membrane component of the developing preparation. To test the suitability of *R. communis* cv. 99N89I for ER purification germinating and developing preparations were performed using the same sucrose density gradient fractionation procedure but using the new seed variety. As well as the ability to generate purified

ER, it was important that purified samples were of sufficient abundance for proteomic analysis.

Briefly, 25-30 seeds of developing endosperm at late Stage V (see Chapter 3.3.1.1) (Greenwood and Bewley, 1982) of development were chopped in homogenisation buffer for 20 minutes to produce a crude homogenate. Gentle centrifugation of the homogenate was performed to remove unbroken cells and a floating layer of fat which settled on the supernatant surface. The clarified homogenate was then fractionated through a sucrose step gradient by isopycnic centrifugation resulting in crude ER membranes visible as a cloudy band at the 30% (w/w) / 20% (w/w) sucrose interface (Figure 3.3). The presence of a cloudy band at this point was previously reported for preparations using the same method but with *R. communis* cv. Hale (Maltman et al., 2002) suggesting at this point that cv. 99N89I was yielding membranes to the homogenisation buffer with a similar buoyant density to those released in a cv. Hale ER preparation. Collected membranes were then removed with a syringe and subjected to an extended floatation centrifugation step previously found to remove contaminating soluble proteins (Coughlan et al., 1996). After the second centrifugation step a band was again collected from the 20% (w/w) / 30% (w/w) sucrose interface. This band was thinner than the band present at 2 hours of centrifugation but was clearly present (result not shown). The band was removed with a clean syringe, combined with an equal volume of ddH_2O and centrifuged at 100,000 x *g* to yield a membrane pellet which, after removal of supernatant, was re-suspended in 10% glycerol, snap frozen in liquid N_2 and stored at -80 °C until required. The same procedure was applied to germinating seed grown for 3 days in vermiculite at 30 °C in the dark. 30 processed 3 day seeds (see Chapter 2.2.2.2) were homogenised in 30 ml of homogenisation buffer, filtered and centrifuged to remove unbroken cells. A fat pad was obtained at this point but this was significantly reduced compared to the developing preparation fat pad, suggestive of the germinating seed's catabolism of carbon storage reserves. Bands were clearly visible at the 20% (w/w) / 30% (w/w) sucrose interfaces

after 2 hour and 22 hour centrifugation steps.

Assessment of the germinating and developing membrane preparations was then performed. To quantify the abundance of protein generated from the 99N89I endosperm, the protein concentration of the preparations was assessed with the colourimetric Bradford protein assay. Replicate aliquots of the membrane samples were assayed and the protein concentration estimated from the average. Table 3.1 shows protein concentration assay data of 3 germinating and 3 developing representative 99N89I preparations. Typical 30 seed germinating preparations were found to give between 100-400 μg of protein, whereas developing preparations consisting of 25-30 seeds gave higher protein yields of typically 400-800 μg. For proteomic analyses, a large format preparative gel destined for spot picking and MS analysis ideally contains 200 μg to 1 mg of protein. A large format analytical gel visualised with post electrophoretic silver or SYPRO^{TM} stain requires 50 μg, although this is reduced to 12.5 μg if the protein sample has been fluorescently labelled with CyDye for visualisation. Pre-analysis of samples including protein concentration assays, SDS PAGE and mini 2DE analysis presents further demands on protein sample amounts. The protein yields of 99N89I ER preparations are thus at the limits of suitability for proteomic studies with gel based technologies, although the use of highly sensitive CyDye labelling will alleviate this somewhat so the decision was made to continue evaluation of the variety.

To assess the protein profile of the purified *R. communis* germinating and developing membrane samples, 5 μg aliquots were separated by SDS PAGE on 12% acrylamide gels. Figure 3.4 shows an SDS PAGE gel of 3 cv. 99N89I germinating and developing ER preparations, revealing a high level of reproducibility between both the germinating and developing ER protein samples. It also reveals broad protein profile similarities between the two developmental states. The figure also shows an SDS PAGE profile of developing ER from the cv. Hale, as previously published (Maltman et al., 2002), with permission. Similarities between the varieties are evident. Both the Hale and 99N89I

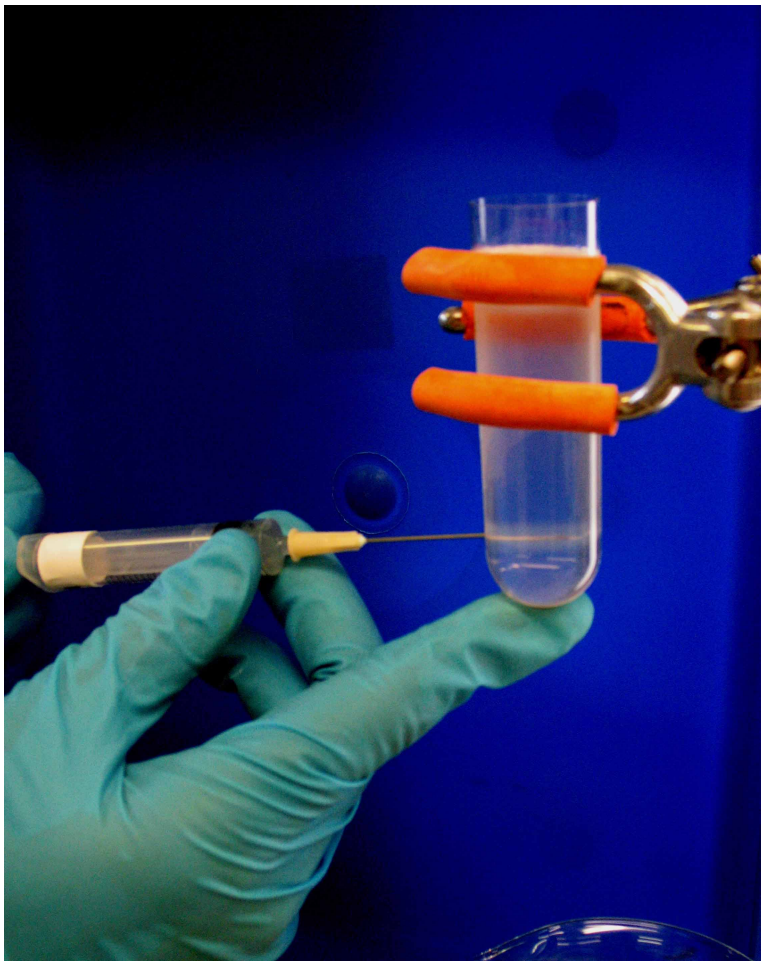


Figure 3.3: Partially purified ER is collected from the 30% (w/w) 20% (w/w) sucrose interface by isopycnic centrifugation. A band of cloudy white membranes is clearly visible at the 20% (w/w) / 30% (w/w) sucrose interface after 2 hours of isopycnic centrifugation. The appearance of this band was previously reported for cv. Hale (Maltman et al., 2002). A syringe was used to withdraw the membranes by puncturing the polycarbonate wall of the centrifuge tube. Membranes were subjected to an extended floatation centrifugation step to remove contaminating soluble proteins, before purified membranes were pelleted and stored until required.

Table 3.1: Bradford assay data of six representative *R. communis* cv. 99N89I germinating and developing ER preparations solubilised in 2DE lysis buffer. Two replicate assays were performed for each preparation to improve accuracy. Average absorbance readings were then compared against a linear absorbance standard curve generated from 1 - 20 μg BSA to give an estimation of protein concentration and thus total protein amount in each preparation.

Preparation	A595			Protein		Total
	Replicate 1	Replicate 2	Average	μg/2 μl	μg/μl	
Germinating						
16.03.04	0.192	0.248	0.220	5.00	2.50	375.0 μg
17.03.04	0.180	0.178	0.179	4.04	2.02	303.0 μg
20.03.05	0.099	0.091	0.095	2.15	1.075	161.0 μg
Developing						
3.06.03	0.124	0.114	0.119	2.59	1.295	647.5 μg
11.08.03	0.095	0.081	0.088	1.91	0.955	477.5 μg
14.10.03	0.084	0.089	0.087	1.88	0.94	946.0 μg

profiles are dominated by three high molecular weight (around 60 kDa) band clusters between 50 kDa and the top of the gel, marked by three arrows on the SDS PAGE profiles. Further bands between 24 and 45 kDa display a similar distribution pattern in the two developing variety types and the developing Hale sample. Both the Hale and 99N89I developing samples were separated on 12% acrylamide but were run separately so extent of separation varies. M_{r} marker positions are labelled for both gels.

In summary, when the previously established preparation procedure was applied to 99N89I endosperm, cloudy material floated at the 30% / 20% sucrose interfaces after 2 and 22 hour centrifugation steps as reported previously for cv. Hale (Maltman et al., 2002). On completion of the purification step, the protein concentration of the samples was found to be low, but suitable for proteomic methodologies, especially if 2D DIGE is used. When the purified membrane samples were analysed by SDS PAGE they were found to be broadly similar, to have significant similarities between the germinating and developing samples, and similarities to an SDS PAGE profile published previously

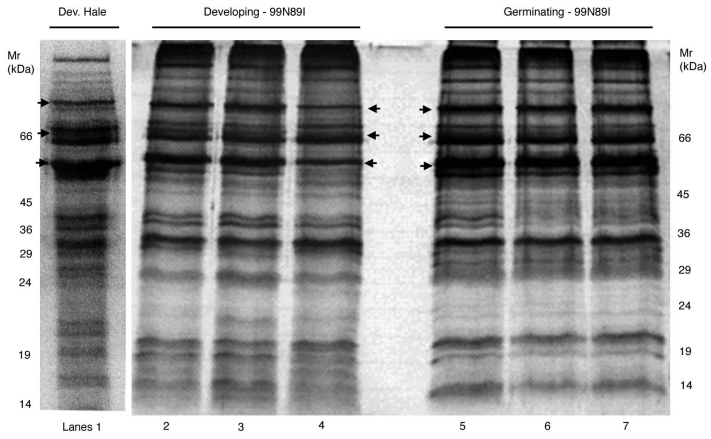


Figure 3.4: SDS PAGE profiles of developing and germinating endosperm ER. 99N89I preparations are labelled (lanes 2-6) and show broad reproducibility between themselves and similarities to each other. An SDS PAGE profile of *R. communis* cv. Hale (lane 1) developing ER has been reproduced for the purposes of comparison with the new variety, with permission (Maltman et al., 2002). Gels are both 12% acrylamide but were run separately so extent of separation varies between the two gels, as indicated by the M_{r} markers. The cv. Hale preparation was made at the same 25 DAF stage of seed development. Both gels were Coomassie stained. Arrows indicate 3 clusters of high molecular weight proteins which appear to be common to all samples.

for cv. Hale. Therefore, there is evidence that ER can be generated from 99N89I endosperm which is of sufficient concentration for proteomic analysis. Biochemical marker assays were not performed on the 99N89I preparations used in this study. However, the purification methodology replicated that previously reported to yield ER free of mitochondrial, glyoxysomal, cytosolic, Golgi and plasma membrane enzyme markers based on biochemical assay and soluble proteins based on SDS PAGE of sucrose gradient fractions (Coughlan et al., 1996). It was also identical to the preparation procedure used previously in the laboratory which identified the ER-resident enzyme oleate $\triangle 12$ hydroxylase as component of the preparation membranes. Evidence that the preparations used in the study are pure ER is supported by the proteomic analysis described in Chapter 4.

More research was needed to ascertain reproducibility of ER samples from germinating and developing endosperm.

3.3.1.3 Evaluation of sample reproducibility by 1D SDS PAGE

A critical aspect of scientific study is the replication of an experiment to increase the reseacher's confidence in the data obtained. This is true of proteomics where multi-biological replicate analyses are now expected.

To avoid any loss of sample from an additional precipitation step, ER pellets from new preparations were solubilised directly into 2DE lysis buffer. 5 μg aliquots of lysis buffer solubilised germinating and developing ER samples were then incubated with SDS PAGE loading buffer and separated on 12% SDS PAGE gels before Coomassie staining (Figure 3.5). Protein profiles of germinating and developing ER showed good reproducibility, indicating both a reproducible preparation method, consistent solubilisation of ER proteins in 2D lysis buffer and a reproducible separation method.

3.3.1.4 Evaluation of sample reproducibility by mini 2DE

Mini 2DE gels were used to critically assess reproducibility. A minimal 20 μg of lysis buffer solubilised protein was used for mini 2DE experiments to minimise use of *R. communis* sample. pH 3-10 IPG strips were used for the first dimension separation with 12% acrylamide second dimension gels. After separation in the second dimension protein gels were silver stained. Figure 3.6 shows three germinating and developing mini 2DE gels showing extent of reproducibility of *R. communis* samples and revealing differences between the germinating and developing states of endosperm ER. Reflecting the pattern seen on SDS PAGE gels, the 2D profile of both sample types is dominated by abundant proteins in the high molecular weight region of the gel, which were found to have a low pI. The gels reveal that the both the germinating and developing ER samples are broadly reproducible. Variability in the relative abundance of protein spots was identified in the germinating samples, the most significant example of which is indicated by an arrow in the figure. This might be reflective of active utilisation of storage proteins in the germinating seed and the importance of selecting germinating

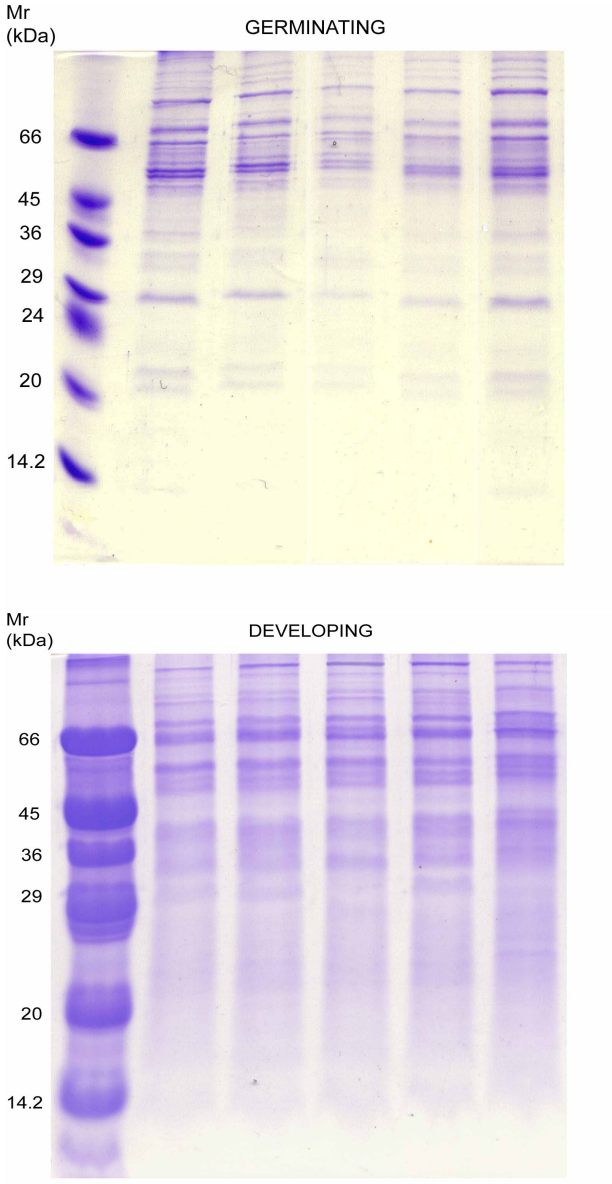


Figure 3.5: Reproducibility of 2DE lysis buffer solubilised germinating and developing *R. communis* ER proteins assessed by SDS PAGE. SDS PAGE analysis reveals developing and germinating ER samples are broadly reproducible, indicating both a reproducible preparation method and consistent solubilisation of ER proteins in the 2D lysis buffer. The 12% acrylamide gels were stained with Coomassie, 5 μg protein per lane.

endosperm closely matched in the size. Despite this variability, it was not deemed significant enough to warrant adaptation of the preparation procedure or the growth conditions for the germinating seed. Assessment of reproduciblity of ER samples by mini 2DE was independently validated by a colleague before the protein sample was approved for further analyses.

Mini 2DE is also useful for ascertaining the pI and molecular weight distribution of a protein sample, allowing the selection of the most suitable pH range of IPG strip and percentage of acrylamide gel. The gels show that *R. communis* endosperm ER sample contains proteins whose isoelectric points lie in both low and high pH ranges, indicating the requirement for broad range IPG strips with this sample type.

3.3.1.5 Summary

R. communis cv. 99N89I seed germinated with high efficiency and grew well under growth room conditions producing developing seed. Evaluation of the endosperm from germinating and developing seed found it to follow the staging descriptions published elsewhere (Greenwood and Bewley, 1982). ER was prepared from germinating and developing seed using an established procedure (Coughlan et al., 1996). The profile of cv. 99N89I developing ER was found to be broadly similar to that of cv. Hale developing ER published by the author's laboratory (Maltman et al., 2002). Assay of 99N89I ER sample protein yield revealed that protein amounts were suitable for downstream proteomic analysis, but a single ER preparation would not provide enough protein for more than one preparative and analytical gel. Therefore, care must be taken to conserve ER material and the use of the highly sensitive 2DE technology 2D DIGE is preferred[1]. Reproducibility of ER samples solubilised in 2DE lysis buffer was assessed by mini 1D gel and samples were found to be reproducible. Further

[1]A traditional quantitative analytical gel requires a minimum of 50 μg of protein when used with post-electrophoretic stains. CyDyes employed in the pre-electrophoretic labelling of 2D DIGE experiments provide high sensitivity protein visualisation with 12.5 μg of protein.

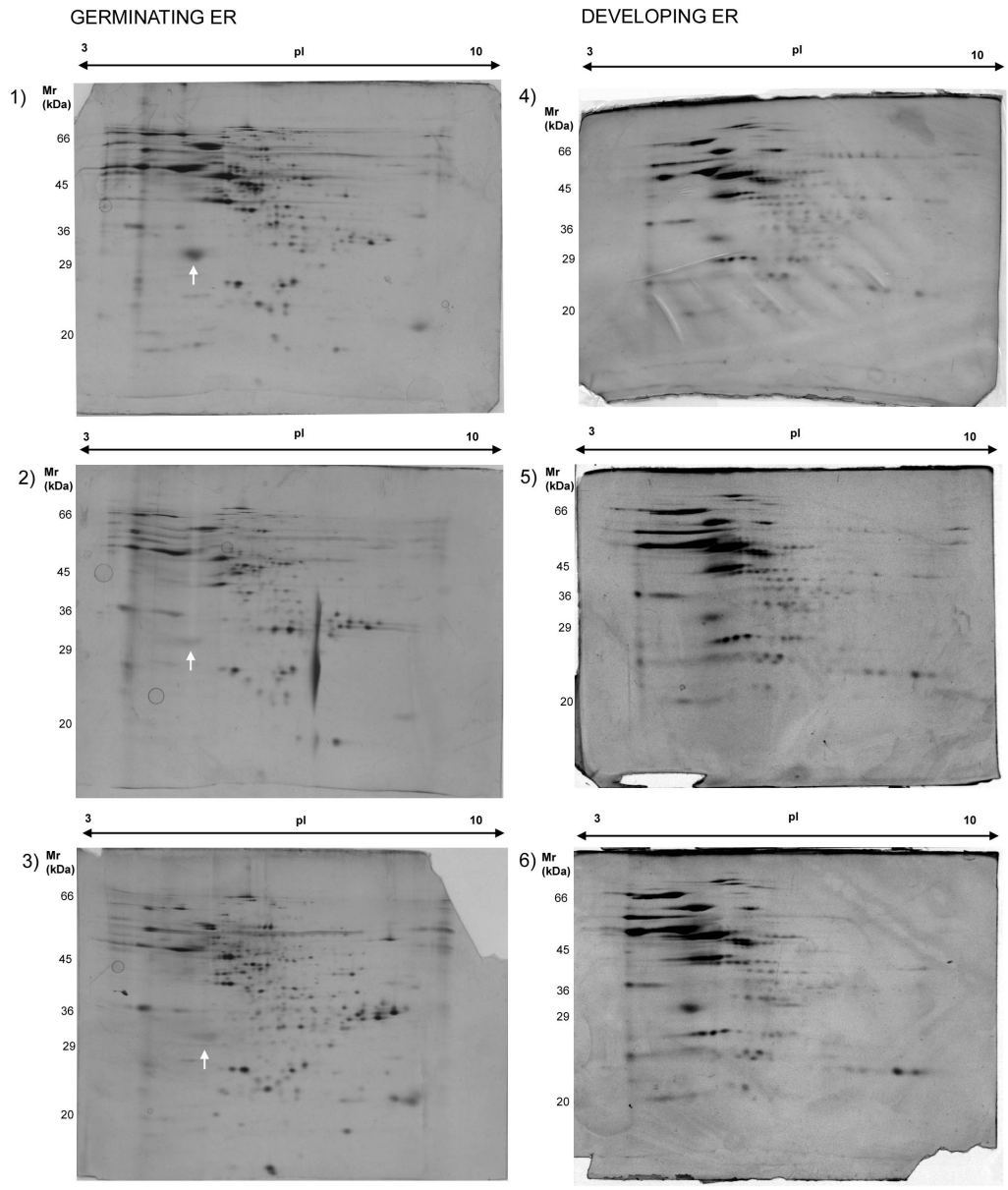


Figure 3.6: Reproducibility of *R. communis* germinating and developing ER preparations evaluated by mini 2DE. Evaluation of protein sample by mini 2DE revealed that both germinating and developing samples were broadly reproducible. Gel reproducibility was independently assessed by a colleague before samples were approved for downstream analysis. The arrow on the germinating gels indicates a spot which is more abundant in one sample than the others, but minor sample variability was tolerated. 7 cm pH 3-10 IPG strips were used to separate protein samples that were solubilised directly into 2DE lysis buffer. After focussing mini SDS PAGE was performed with 12% acrylamide gels. Gels were stained with disruptive silver.

evaluation of reproducibility by mini 2DE found samples to be broadly reproducible but some variability in protein abundance was identified, especially with germinating ER. Samples were not deemed to be sufficiently irreproducible to abandon the preparation or growth procedures; however as a result of this variability, prior to differential analysis of germinating and developing ER a thorough quantification of gel reproducibility will be made.

Following this work, the next stage was to evaluate CyDye labelling of *R. communis* ER samples and to establish and evaluate a preparative gel procedure that will maximise the chance of identifying and picking proteins of interest on high loading preparative gels.

3.3.2 Evaluation of CyDye labelling of dilute *R. communis* seed ER samples

Following the establishment a reproducible preparation procedure for *R. communis* 99N89I capable of producing enough purified ER for proteomic analyses, the next stage of the research was to evaluate the effectiveness of 2D DIGE with the ER samples. 2D DIGE (GE Healthcare, Bucks UK) is a sophisticated advancement of 2DE technologies. It utilises pre-electrophoretic covalent labelling of the protein sample's lysine residues with fluorescent cyanine dyes (CyDyes) rather than post-electrophoretic stains. Currently three spectrally distinct CyDyes exist, Cy3, Cy5 and Cy2, allowing for multiplexing of protein samples within a single gel. At the most basic level, this makes sample comparison simpler by removing 2DE profile variability introduced by gel-to-gel rather that biological variation. The incorporation of a (typically Cy2) labelled internal standard aids inter-gel matching and improves accuracy of quantitation by minimising the impact of gel-to-gel variation on quantification. The methodological background to the 2D DIGE process is described in Chapter 1.6.2 and the 2D DIGE protocols used are described in Chapter 2.3.5.

Before a complete 2D DIGE analysis of endosperm ER could occur the technical suitability of the recommended method, as set out in *Ettan* DIGE User Manual, to the *R. communis* endosperm ER samples was assessed.

The first description of the use cyanine dyes for proteomic applications set out the criteria for successful and reproducible labelling of proteins, requiring adherence to correct pH, reaction volume, reaction time and temperature (Tonge et al., 2001b). The methodology was developed in the commercialised Ettan DIGE^{TM} platform as set out in the *Ettan* DIGE User Manual, where it stated that the pH of the sample should fall between pH 8.0 and 9.0, that the working stock CyDye fluor stock solution should be at a concentration of 400 ρmol/μl and that the ratio of protein to dye in the labelling reaction should be maintained at 50 μg protein : 400 ρmol dye. It also

stated that the concentration of the protein sample should be at least 1 mg/ml, so that the labelling reaction occurs in a minimum 50 μl reaction volume for 50 μg protein. Further to this it was recommended that the optimum final reaction volume be 20 μl, i.e. a protein concentration of 2.5 mg/ml (B. Bacher, GE Healthcare, personal communication. February 2004).

R. communis samples destined for 2DE had been solubilised in 2DE lysis buffer at the final stage of sample preparation as this method prevented any potential loss of sample from a precipitation procedure and is compatible with both 1D SDS PAGE and traditional 2DE. Samples were solubilised in a volume of 2D lysis buffer suitable for in-gel rehydration of IPG strips followed by post-electrophoretic staining of gels. Dependant on the physical size of the pellet yielded by the preparation procedure, samples were solubilised in typically 200-400 μl of 2DE lysis buffer, generally giving protein concentrations of between 0.5 to 1.5 mg/ml. This posed a problem as it meant our ER protein stock was at a borderline level of concentration for the purposes of 2D DIGE labelling. The restricted time window in which developing preparations could be made, the large number of archived ER preparations and the time required to produce a new preparation meant generating a new set of stocks of *R. communis* ER at the recommended concentrations was undesirable. Instead, an experiment was designed to investigate whether the protein concentrations of the existing sample stock could be successfully labelled; without any effect on derivatisation of dye to protein sample.

To conserve ER protein, and to allow labelling at the optimal level of 2.5 mg/ml, abundant *A. thaliana* cell culture total soluble protein (TSP) (kindly supplied by Dr. S. Chivasa) was used in a comparison of CyDye labelling at (1) the recommended concentration and (2) the ER-sample equivalent concentration. *A. thaliana* sample in labelling buffer (pH adjusted 2DE lysis buffer, without DTT. See Chapter 2.3.5.1) was diluted to the two levels of protein concentration: the ideal labelling concentration of 2.5 mg/ml, and the concentration of the most dilute *R. communis* ER sample in store,

Table 3.2: Set up of ideal and non-ideal labelling conditions to investigate effects of protein concentration on label derivitisation. 50 μg of the abundant *A. thaliana* TSP sample was labelled under ideal (2.5 mg/ml, 1:20 dye volume : reaction volume ratio) and non-ideal (0.59 mg/ml, 1:84.7 dye volume : reaction volume ratio) concentrations, the non-ideal protein concentration replicating the lowest concentration *R. communis* ER sample in store. Apart from the non-ideal reaction volume, labelling conditions were performed normally (see Chapter 2.3.5).

Sample	Protein concentration	Protein amount labelled	400ρmol/μl dye stock	Dye amount	Dye	Reaction volume	Dye : Volume ratio
Recommended	2.5 mg/ml	50 μg	1 μl	400ρmol	Cy3	20 μl	1:20
ER-sample equivalent	0.59 mg/ml	50 μg	1 μl	400ρmol	Cy5	84.7 μl	1:84.7

0.59 mg/ml. Dilutions were made with labelling buffer. Dye concentrations, protein concentrations and labelling ratios used are presented in Table 3.2.

The two protein samples were then labelled with two spectrally distinct fluorophores so they could be multiplexed (i.e. run together) in a single gel. Samples were run under standard conditions: briefly, 12.5 μg of each sample was cup loaded onto a single rehydrated IPG strip and focussing performed under standard IPGPhor (GE Healthcare, Bucks UK) conditions for an 18 cm pH 3-10 Immobiline DryStrip (GE Healthcare, Bucks UK). Focussed strips were equilibrated and separated in the second dimension on a large format 1 mm 12% acylamide gel cast between low fluorescent glass plates (see 2.3.5). After the second dimension run was complete, Cy3 and Cy5 images were collected from the gel using the Typhoon variable mode imager (GE Healthcare, Bucks UK). Photomultiplier tube (PMT) values were adjusted to ensure neither image was saturated but made good use of the available dynamic range.

An initial visual check of the gels was made by overlaying the two profiles. As the Cy3 and Cy5 samples co-migrated within a single gel, and because the two dyes are charge and mass balanced, the location of identical proteins labelled with Cy3 and Cy5 on the gel are the same. Gel images were printed onto paper and transparency film allowing their overlay. Cy3 (magenta) and Cy5 (cyan) spots gave a composite (black) colouration when the print-out on the transparency (Cy3 gel image) and the print-out

on paper (Cy5 gel image) were aligned. On visual examination of the overlay image, no deviations from the composite colour of black could be seen (see Figure 3.7).

Although a simple overlay of the print-outs revealed no differences in the uniformity of fluorescent labelling at the two levels of dilution, a quantitative examination of spot abundances would allow a more rigourous assessment of any difference in derivatisation of dye to protein at the two protein concentration levels. Both gel images were imported into DeCyder DIA (GE Healthcare, Bucks UK) for a pair-wise analysis. The DIA module is one component of the total DeCyder package, containing the spot detection algorithm and the ability to quantify protein levels within a single gel (for multiple gel analyses, DeCyder DIA is used in conjunction with DeCyder BVA). Gel images were imported into the DIA module and spots detected using the co-detection algorithm of the DeCyder software. Spot volumes were re-calculated following a background subtraction and normalisation procedure. The normalised volumes of matched spots (between the optimally labelled (Cy3) and dilute labelled (Cy5) gels) were expressed as a ratio; spots with identical normalised abundances would have a ratio of 1. To identify spots with different normalised abundances (potentially indicating a difference in derivatisation of CyDye to protein between the two levels of protein dilution), spots exhibiting a positive fold change of ≥ 1.1 were given the status 'increased'. Conversely those with a negative fold change of ≤ 1.1 were assigned the status 'decreased'. Spot volume ratios that were within a ± 1.1 fold change were assigned the status 'similar'. For the purposes of visualisation on the gel, similar spots were coloured green, increased spots were coloured red, and decreased spots blue. Figure 3.8 shows the gel images optimally and dilute labelled gel images (left and right respectively). The gel images show the spot outlines as detected by the DeCyder DIA software, coloured according to whether the fold change value is ≥ 1.1 in the optimally labelled gel (i.e. red), ≤ 1.1 in the optimally labelled gel (blue) or a fold change less than ± 1.1 (green). This analysis shows that the majority of the spots are green, i.e. their normalised abundances

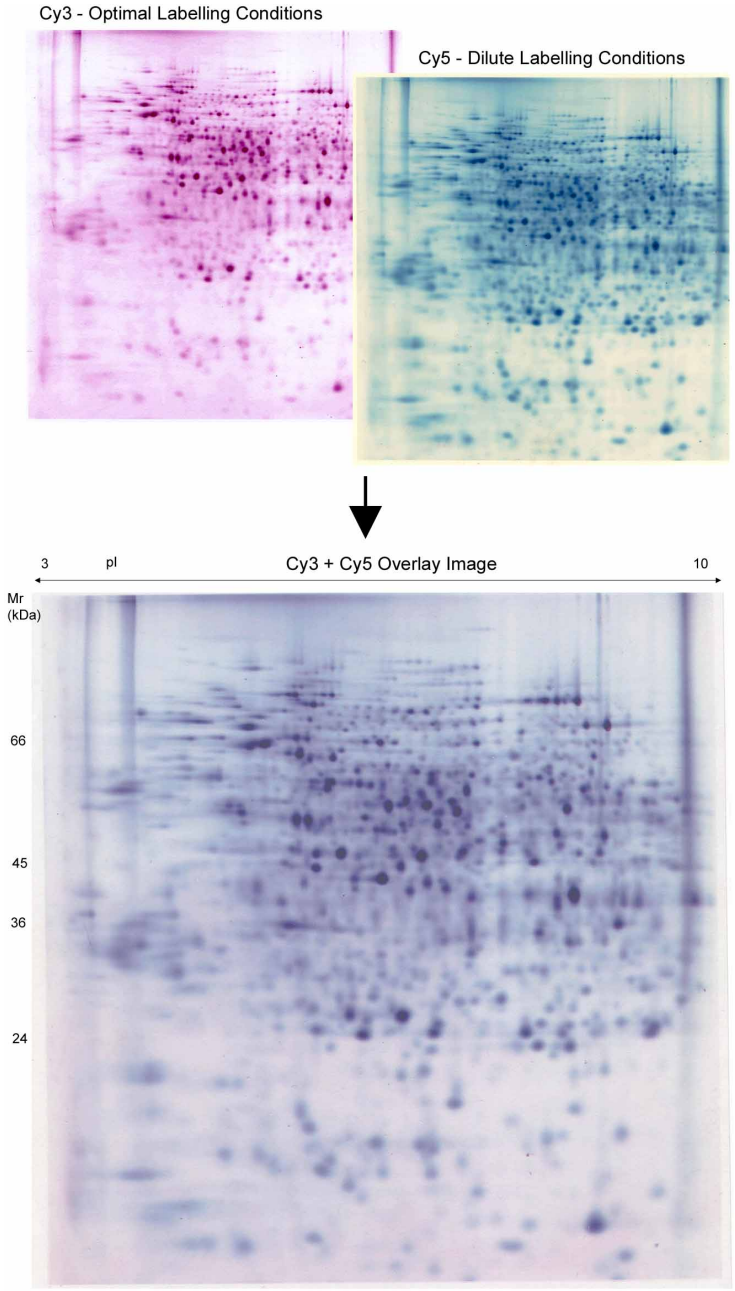


Figure 3.7: Overlay of optimal (Cy3) and dilute (Cy5) labelled *A. thaliana* protein on a 2DE gel. Physical overlay of gels was performed as a swift analysis of labelling, allowing significant differences to be revealed as deviations from the black overlay colour obtained from the overlay of similar sized and intensity Cy3 and Cy5 spots. It showed no visually obvious differences in labelling between the optimal and dilute labelled samples.

are similar (the abundance of the spots are within a fold change limit of ±1.1 in one labelling state compared to the other). There are areas of the optimally labelled gel with increased spot abundances compared to the dilute labelled gel image. By examining the gels with the green (similar) spot outlines switched off (bottom half of the figure), it is clear that all of these spots are noise or gel artefacts associated with the vertical streak from the cup. A spot located within the vertical streak associated with the cup was identified as increased in the optimally labelled gel, but its location within a gel artefact meant this result was discounted (green arrow). There are a smaller number of spots which are increased in abundance in the dilute labelled gel (blue outlines), but again for all but one of these outlines they are not detections of protein spots but noise or artefacts. The one incident where a spot does appear to be increased in abundance between the two labelling states is indicated by a white arrow on the bottom two gels. As this is the only example of a difference in abundance of a protein spot between the two labelling states, and as this is increased in the dilute labelled sample, it was not deemed a significant issue. In conclusion, with *A. thaliana* TSP, labelling 50 μg of protein with 400 ρmol dye in a volume of 84.7 μl (the dilution requirement for protein samples at 0.59 mg/ml) causes no reduction in derivitisation of dye to protein compared to the optimal reaction volume of 50 μg of protein : 400 ρmol dye in 20 μl.

Although *A. thaliana* TSP is not the same as the *R. communis* ER samples, this preliminary investigation gave evidence that labelling of CyDye to protein is unaffected when the concentration of protein in the labelling mix is 0.59 mg/ml rather than the recommended minimum or 1 mg/ml or the optimal concentration of 2.5 mg/ml. The decision was made to use ER samples at concentrations down to a lower limit of 0.59 mg/ml where necessary. ER samples continued to be generated during the study and efforts were made to keep their concentration above 1 mg/ml.

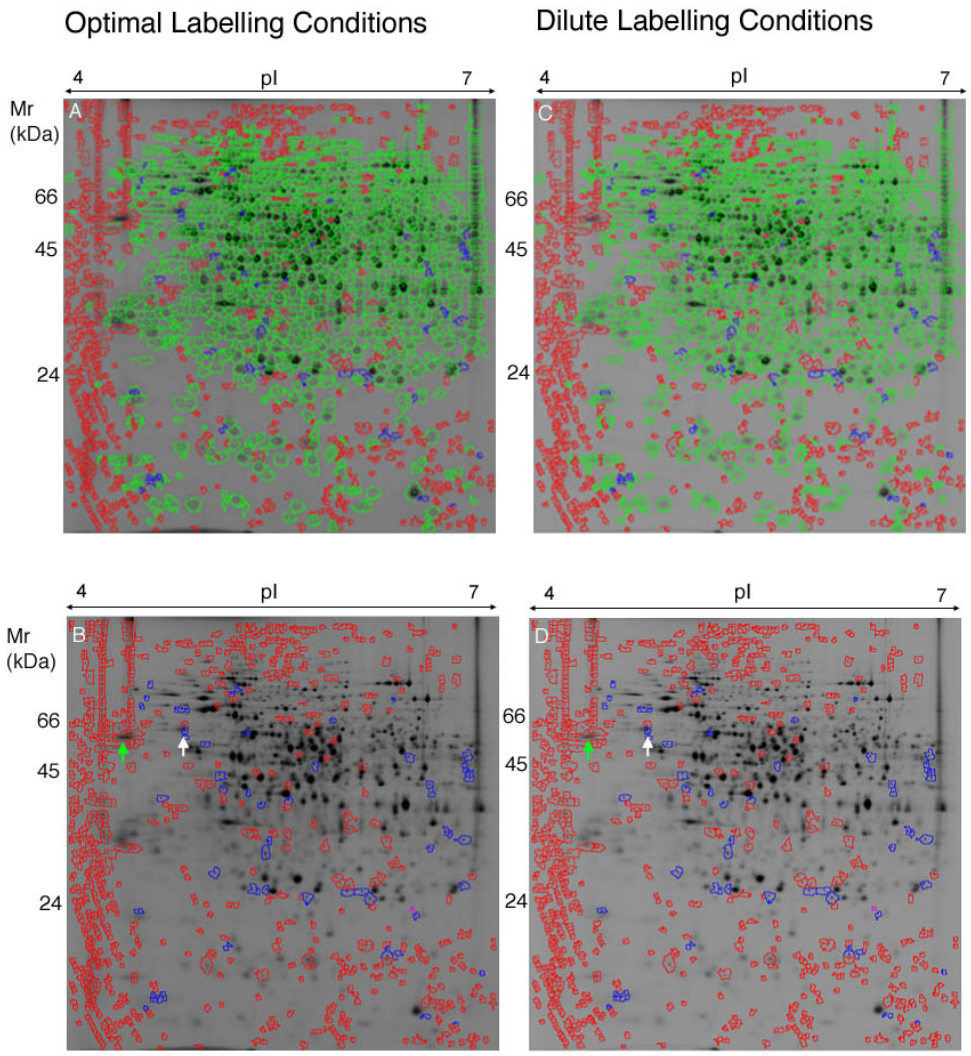


Figure 3.8: DeCyder DIA analysis of optimal (A+B) v dilute (C+D) CyDye labelled *A. thaliana* TSP. Spot outlines are shown in different colours to indicate whether they are increased (+1.1 fold change, red), decreased (-1.1 fold change, blue), or the same (< 1.1 fold change in either direction, green) in the gel of the optimally labelled sample compared to the gel of the dilute labelled sample. Gels B + D show the optimal and dilute gels with the green '*similar*' spot outlines switched off. All but one of the spots increased or decreased in abundance on the gels are detection noise or artefacts. The green arrows indicates a spot increased in the optimally labelled gel, but its presence within the vertical streak associated with the cup means this resulted was discounted. The single real spot which was identified as having a $\geq$1.1 fold change had increased in abundance in the dilute labelled gel, not the optimally labelled gel. As a single incident it was not deemed to reveal a significant deviation in labelling efficiency between the two states.

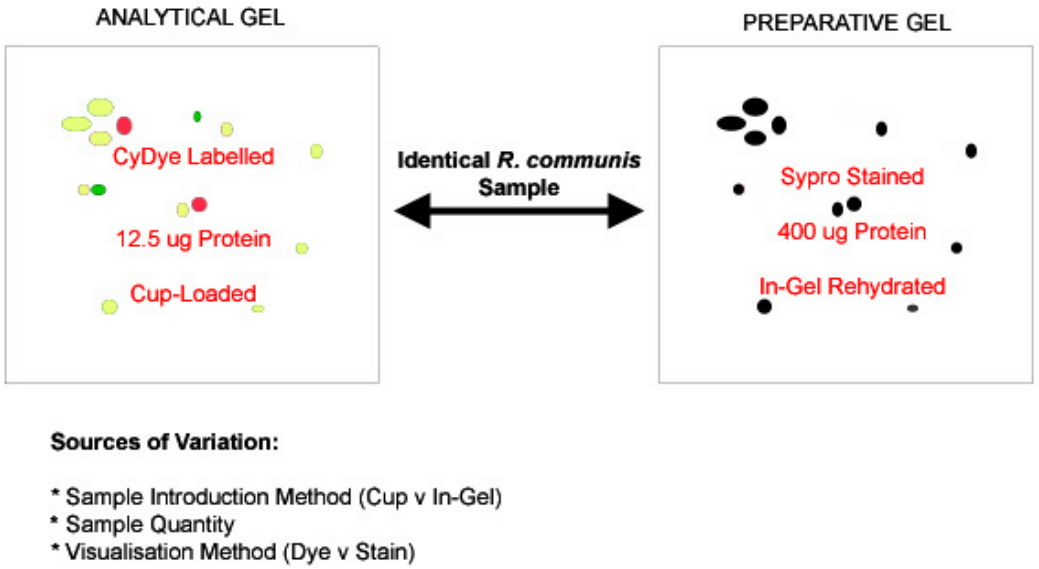


Figure 3.9: Sources of variability which might contribute to difficulties in matching between *R. communis* seed ER analytical and preparative gels.

3.3.3 Establishment of preparative gel procedure

Preparative gels are high loading gels for the purposes of MS analysis, run in addition to the analytical gel set. They generally contain ≥ 200 μg of protein although to maximise the number of protein spots giving MS data often between 400 μg and 1 mg of protein is used. Spots of interest identified on analytical gels are matched to the preparative gel before either manual or robotic excision followed by digestion and MS analysis. The ability to confidently match protein spots between analytical gels and preparative gels is of paramount importance in ensuring correct assignation of identity to analytical quantitation data. Methodological differences between preparative and analytical gels can affect the success of matching between the two gel types, for example spots can appear absent from two gels containing the same sample due to differential staining or labelling of proteins (Fey et al., 1997). This might be the case in a typical 2D DIGE experiment with an external preparative gel, where the analytical gel set is visualised with CyDye and the preparative gel visualised with a post-electrophoretic stain such as

SYPRO or Coomassie. However, despite the additional potential for gel-to-gel variation this brings, it is accepted as the alternative (labelling the preparative protein sample with CyDye) is not feasible due to the expense of the dye. As a rule though, it is preferable to remove unnecessary sources of technical variation wherever possible, for example ensuring that preparative and analytical gels are of the same thickness and that they are run together during second dimension electrophoresis.

In the case of the *R. communis* ER samples, another source of experimental variation was foreseen. The recommended 2D DIGE procedure for introducing labelled protein sample into the IPG strip is through cup loading (*Ettan* DIGE User Manual). However, the low concentration of the seed ER samples means the buffer volume required to provide a suitable quantity of protein for a preparative gel is too large for cup loading; instead the preparative gels have to be in-gel rehydrated. It was hypothesised that this could contribute to profile variation and confound matching of spots.

Potential sources of variation resulting from technical differences between analytical and preparative gels are highlighted in Figure 3.9. Experimentation exploring the contribution of these variables to 2D protein profile is presented in the following section.

3.3.3.1 Comparison of sample introduction methods with *A. thaliana* TSP

In 2DE there are two alternative methods for the introduction of protein sample into an IPG strip. The first of these is termed in-gel rehydration. With this method the Immobiline DryStrip is rehydrated with 2DE lysis buffer containing 1% DTT and 2% ampholytes. For rehydration of an 18 cm (large format) strip 350 μl 2DE lysis buffer is required, which is pipetted within a channel of an Immobiline DryStrip reswelling tray before carefully laying the strip upon the surface of the 2DE lysis buffer. Protein sample solubilised in 2DE lysis buffer makes up a proportion of the final 350 μl volume and thus is incorporated into the strip matrix during the overnight rehydration step.

After rehydration the IPG strip is ready for IEF (see Chapter 2.3.6).

The alternative method is termed anodic cup loading. In this method the IPG strip is re-swelled as for in-gel rehydration, but without the addition of protein sample (i.e. 350 μl 2DE lysis buffer containing 1% DTT and 2% ampholytes for an 18 cm strip). After rehydration the strip is set up for IEF in the same manner as in-gel rehydration except a small plastic cup is clipped upon the surface of the IPG strip at the anodic end of the strip. Up to 70 μl of protein sample solubilised in 2DE lysis buffer can be introduced into the strip through this cup, which is added before the start of IEF. Cup loading is the recommended method of sample introduction for CyDye labelled protein samples as it gives superior separation of protein compared to in-gel rehydration (*Ettan* DIGE User Manual). As stated, the cup method requires that the total volume of loaded sample is 70 μl. As the sample must also include 2% DTT and 2% ampholytes, the remaining volume available for the protein sample is 67.2 μl. The laboratory routinely uses 12.5 μg per labelled sample on a single gel, i.e. a total of 37.5 μg of protein for a typical analytical gel containing Cy2, Cy3 and Cy5-labelled protein samples. This amount of protein allows the use of protein sample concentrations down to a lower limit of 0.56 mg/ml. For preparative gels containing typically 200 - 400 μg of protein for the purposes of downstream MS analyses, the use of cup loading requires a protein concentration of 3.0 to 6.0 mg/ml respectively. This is achievable for abundant protein sources. For the comparatively low yielding *R. communis* ER preparations described in this study the small volumes required to achieve the necessary protein concentrations would make physical dispersal and solubilisation of the pellet difficult. Furthermore it is possible that at minimal volumes of 2DE lysis buffer, focussing issues caused by lipidic and salt contaminants would be exacerbated. Therefore, for *R. communis* preparative gels in-gel rehydration is the required method of sample introduction, but for analytical gels the preferred method of cup loading can still be used. The problem here is that by cup loading the analytical gel set and in-gel rehydrating the preparative gels this extra

variable may introduce differences leading to difficulties in matching spots between the analytical and preparative gels.

To investigate potential differences introduced by the new variable of in-gel rehydration on the preparative 2DE profile, compared to the standard method of cup loading for both analytical and preparative gels, the following experiment was devised. *A. thaliana* cell culture TSP (solubilised in 2DE lysis buffer) was used as a test sample due to its abundance and the preparation's high concentration which allowed preparative cup loaded gels to be run. A preparative protein loading of 200 μg was introduced into two 18 cm linear pH 4-7 IPG strips by the two loading methods: one using an anodic cup and the other in-gel rehydration. The two strips were focussed together on an IPGPhor (GE Healthcare, Bucks UK) using the standard 18 cm pH 4-7 IEF until a total of 70,000 Vh. Strips were then equilibrated as normal and separated in the second dimension for 17 hours at 2 W / gel in an Ettan Dalt (GE Healthcare, Bucks UK) electrophoresis tank (see Chapter 2.3.5). After second dimension electrophoresis, gels were fixed and stained overnight in the dark with SYPRO^{TM} Ruby, then imaged using the Typhoon variable mode imager (GE Healthcare, Bucks UK).

On evaluation of the gels (Figure 3.10) spot resolution is generally good for both methods. There are some low abundance spots visible in the lower pH and lower molecular weight regions of the in-gel hydrated gel which are absent or poorly resolved in the cup loaded gel. One dominant example of this is indicated by a red arrow on Figure 3.10. Comparing its location to the cup loaded gel it appears to be at the site of cup entry where there is a vertical streak associated with this point. Horizontal streaking in the high molecular weight region is apparent in both gels and is especially associated with trains of spots. This streaking is evidently worse in the in-gel rehydrated gel, especially in the acidic regions of the gel, indicated by the blue arrow. Vertical streaking is not an issue observed in either gel, except for a single incidence associated with an abundant spot in the basic region of both gels (see the zoom-in boxes, lower

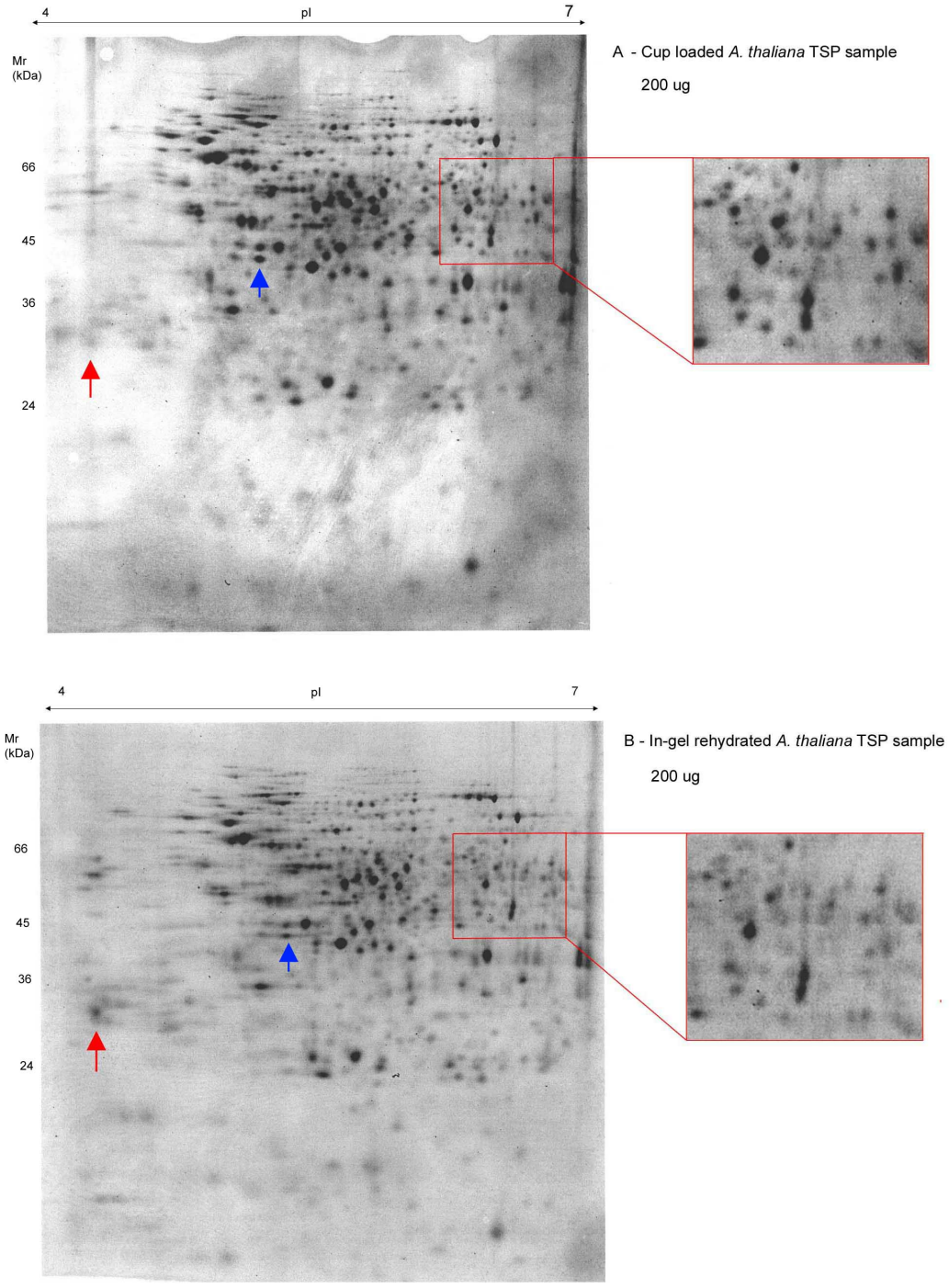


Figure 3.10: Comparison of cup-loading (Gel 'A') vs in-gel rehydration (Gel 'B') on 200 μg *A. thaliana* preparative 2DE profiles. With 200 μg of protein, profiles are broadly the same but the cup-loaded sample has superior resolution in the high molecular weight basic region of the gel (indicated by zoom-in boxes). However, low abundance, low pI spots have resolved less clearly on the cup-loaded gel, especially in the area associated with the vertical cup-loading streak. A dominant example of this is indicated by a red arrow on both gels. A blue arrow indicates an example of horizontal streaking observed in high molecular spots, which was observed to be worse with the in-gel rehydrated gel. Furthermore, protein spots appeared more abundant in the cup-loaded gel.

centre). Protein spots are more cleanly resolved in the high molecular weight and basic region of the cup-loaded gel as indicated by the zoom-in boxes of Figure 3.10.

To investigate whether the observations made for 200 μg gels were the same with higher protein loadings, the experiment was replicated in an identical manner except 400 μg of *A. thaliana* TSP was used. To reduce the contribution of experimental variation, the strips were focussed in the first dimension and separated in the second dimension alongside the 200 μg samples. The horizontal streaking reported for the 200 μg samples is significantly improved in both cup loading and in-gel rehydrated 400 μg gels (Figure 3.11). As observed for the 200 μg samples, spot resolution is improved in the cup loaded sample but the difference between the two sample introduction methods at this loading are negligible. A small amount of vertical streaking is observable on both gels, effecting the same spot as in the 200 μg samples. It is slightly worse in the cup loaded samples which might be due to the presence of slightly more abundant spots with this loading method. Studying the gels for differences in resolution of protein shows that they are broadly the same. The improvement in spot resolution in the cup loaded sample highlighted for the 200 μg is still apparent at the higher loadings, and again this is especially the case in the basic region of the gel (see zoom-in boxes, Figure 3.11). The improved resolution of low abundance, low molecular weight acidic spots seen in the in-gel rehydrated sample at 200 μg is not apparent here, but this may be masked by the staining / imaging artefacts present on the gels (Figure 3.11). As mentioned it appears that the cup loaded gel appears to contain a slightly higher amount of protein compared to the in-gel rehydrated sample. It might be the case that with the in-gel rehydation method there is a greater opportunity for protein loss as the sample is spread thinly across the re-swelling tray before the IPG strip is laid upon it, compared to the direct injection of protein sample for the cup loading method. This wasn't seen with the 200 μg samples however so experimental error cannot be discounted.

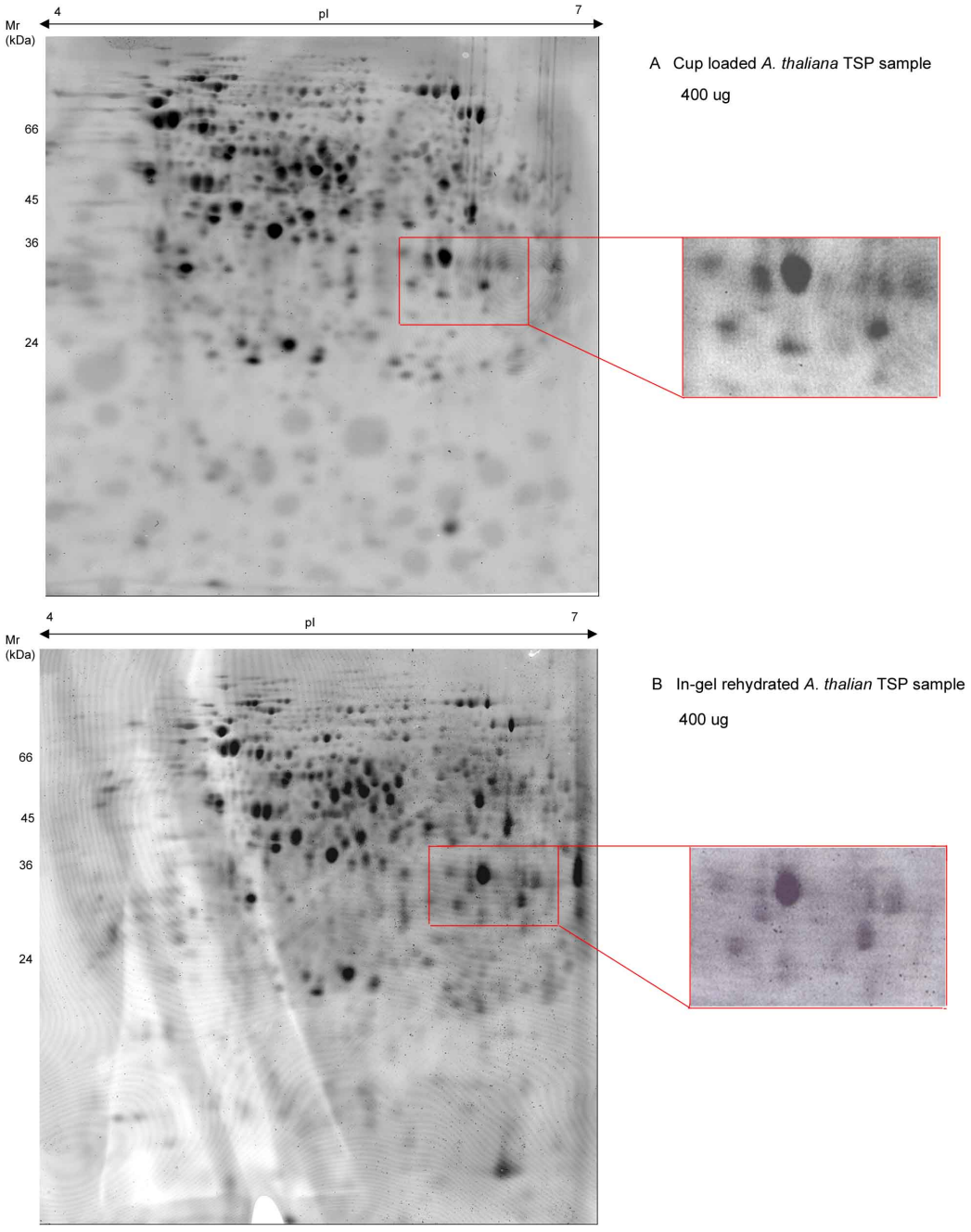


Figure 3.11: Comparison of cup-loading vs in-gel rehydration on 400 μg preparative *A. thaliana* 2DE profiles. Gel 'A' is a cup loaded gel, gel 'B' is an in-gel rehydrated gel. Zoom-in boxes highlight improved resolution of spots in the acidic region of the cup loaded gel, as was seen for the 200 μg cup loading and in-gel rehydration comparison (Figure 3.10).

The results here suggest cup loaded gels do provide superior resolution of spots and reduced streaking when compared to in-gel rehydrated gel spots, but with *A. thaliana* TSP the difference is restricted to either basic protein spots or high molecular weight proteins. Conversely, lower molecular weight (<36 kDa) and acidic proteins are more cleanly resolved in the in-gel rehydrated gels. Although there are differences in the gel profiles between the two sample introduction methods, they do not appear severe enough to cause difficulties in identifying the same protein spot on the two gel types so using both types of sample introduction method in a 2D DIGE experiment is unlikely to cause significant problems. At 400 μg loading, the differences in the protein profiles caused by the sample introduction methods are further reduced.

This information is important to a future proteomic analysis of *R. communis* seed ER as there is now evidence that using the two different methods of sample introduction in the same experiment does not alter the 2DE profile so much that they cannot be compared. The experiment does not provide evidence on the effect of 12.5 μg protein versus 400 μg protein on the 2DE profile, nor does it deal with actual *R. communis* ER samples which could behave differently, especially as they are membrane samples. The effect of sample introduction method on the protein profile of *R. communis* ER is reported in the following section, and the influence of protein loading on the 2DE profile, including the effect of 12.5 μg protein versus 400 μg protein, is reported in Section 3.3.3.3 (which also considers the use of CyDye labelled spikes run within preparative gels as a 'matching guide').

3.3.3.2 Comparison of sample introduction methods on *R. communis* ER 2DE profiles

With *A. thaliana* TSP no significant differences are introduced between the two loading methods, suggesting that matching between two gels will not be significantly more difficult because one has been cup loaded and the other in-gel rehydrated. The gels did

suggest that for the majority of spots cup loading provided superior resolution to in-gel rehydration with this sample type. However, the effect of cup loading compared to in-gel rehydration for *R. communis* membrane samples may differ from that found for *A. thaliana*. If it were the case that there was no difference between cup loading and in-gel rehydration, a more sensible approach might be to in-gel rehydrate the CyDye labelled analytical ER samples as well as the preparative samples. This would remove the issue of experimental variation caused by the use of two different sample introduction methods, a deviation from the recommended protocol because of the nature of the *R. communis* sample.

To investigate this question, a Cy2 labelled sample of germinating ER was used. The high sensitivity of CyDye labelling allowed a 12.5 μg of *R. communis* sample to be cup loaded but give sufficient protein to give a good quality 2D gel profile. 12.5 μg of the same sample was in-gel rehydrated and run alongside the cup loaded gel on an IPGPhor (GE Healthcare, Bucks UK). After equilibration and second dimension separation on a 12% acrylamide gel using the standard method (see Chapter 2.3.5), gels were visualised within their cassettes on a Typhoon variable mode imager (GE Healthcare, Bucks UK). Photo Multiplier Tube (PMT) values were adjusted to ensure optimal usage of the scanners dynamic range in the outputted gel files. As stated, the dilute nature of the *R. communis* ER samples means preparative amounts of protein cannot be cup loaded so a direct repeat of the *A. thaliana* could not be done. Instead, 12.5 μg Cy2 labelled ER sample was used in comparison of cup loading versus in-gel rehydration. This is still valid as it provides evidence on the effect of sample introduction method with the ER samples.

The obtained gel profiles are presented in Figure 3.12. On examination of the gel profiles although the general distribution of spots is the same it quickly becomes apparent that the cup loaded gel has multiple incidences of significantly superior spot resolution compared to the in-gel rehydrated gel. Throughout the cup loaded gel, spots

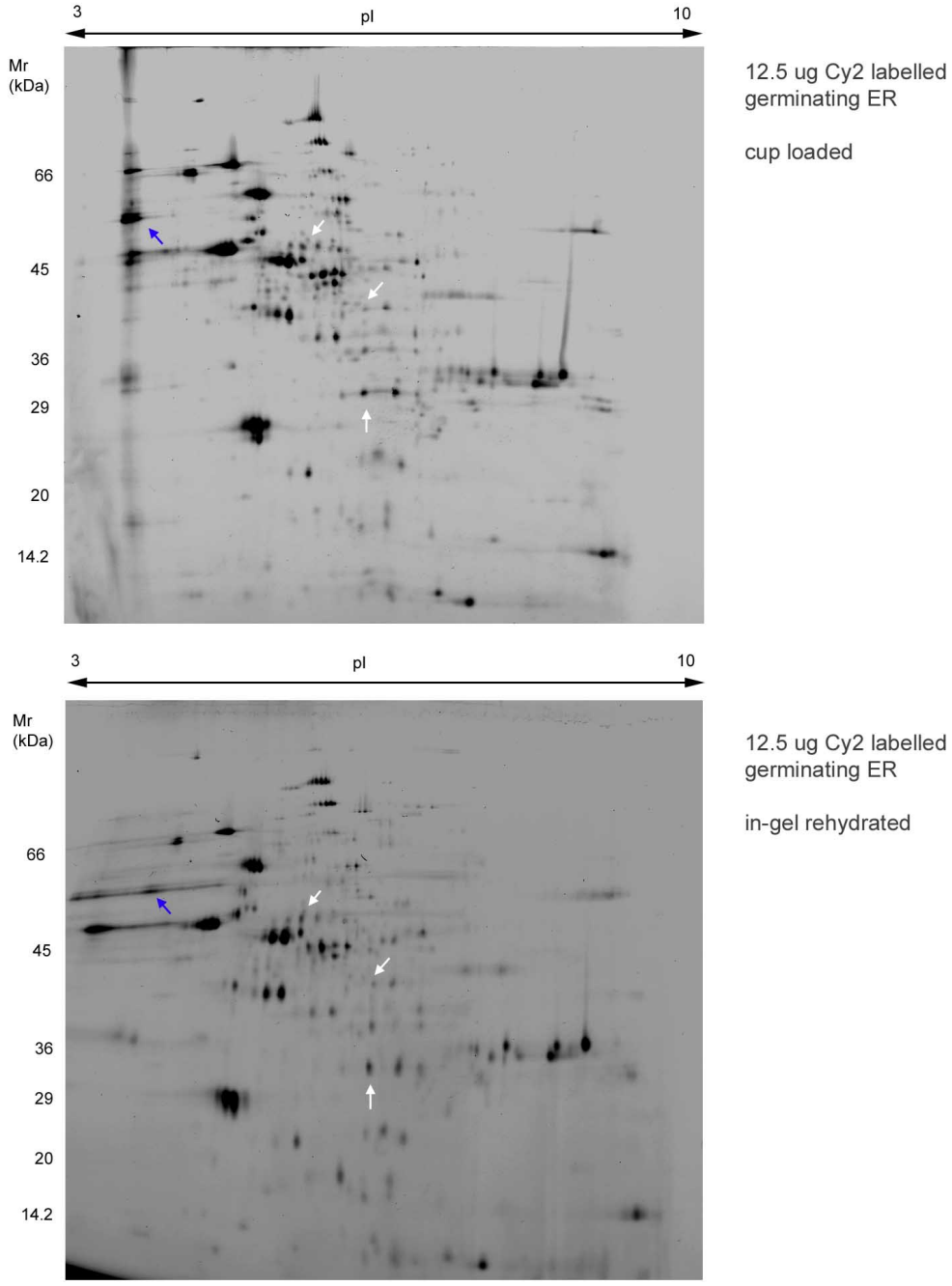


Figure 3.12: Comparison of cup loading versus in-gel rehydration with an identical Cy2 labelled 12.5 μg *R. communis* germinating ER sample. The cup loaded gel gave superior resolution compared to the in-gel rehydrated gel. White arrows indicate examples of the difference in spot resolution on the two gels. Migration of proteins within the region of the cup can be effected (blue arrow).

appear less diffuse than in the in-gel rehydrated gel. Some examples of these differences in resolution are indicated by arrows on Figure 3.12. Although 12.5 μg of ER sample was used for both loading methods, the cup loaded gel appears to contain a higher number of protein spots. This may be due to the superior resolution allowing more spots to be clearly identified, or it may be due to sample loss with the in-gel rehydrated sample, an observation made for the comparison of sample introduction method with *A. thaliana*. There is a large vertical streak present in the acidic region of the cup loaded gel, corresponding to the site of cup entry and seen previously in the *A. thaliana* gel (Chapter 3.10). Unlike in the *A. thaliana* gel there are no incidences of protein loss associated with the cup, although one protein has behaved differently between the two introduction methods (blue arrow).

The *R. communis* ER gels represented in Figure 3.12, and the *A. thaliana* TSP gels represented in Figures 3.10 and 3.11 suggest cup loading gives superior spot resolution to in-gel rehydration for the majority of spots on a 2D gel profile for both *A. thaliana* and *R. communis* samples. However, the differences to a 2DE gel profile introduced by the different sample loading methods does not appear to be so extensive that confidence in matching spots between gels of the same sample but separated by the two methods is significantly reduced.

Obtaining 2DE gels with the highest resolution and spot clarity is most critical for the analytical samples in a 2DE or 2D DIGE experiment. Based on the observations described here, cup loading is the method of choice for these gels. The observation that sample might be lost during in-gel rehydration confirms this conclusion, as this could adversely effect the inter-gel matching of different analytical samples and decrease the rigour of quantitative data in the experiment.

3.3.3.3 Evaluation of analytical spikes as preparative gel guides

Preparative samples are typically run on a separate gel to the analytical gels, which can introduce gel-to-gel variation due to, for example, an alternative method of sample introduction, different protein visualisation methods (e.g. CyDye and SYPRO^{TM}) and different protein amounts. These factors combined can introduce differences between the protein profiles of preparative and analytical gels, reducing confidence in matches made between the gel types. For a 2D DIGE analysis of *R. communis* samples, preparative gels needed to be in-gel rehydrated due to sample concentration constraints. Cup loading is the recommended method of sample introduction for CyDye labelled analytical samples (*Ettan* DIGE User Manual) and was found to provide superior resolution to in-gel rehydration with a Cy2 labelled analytical loading of *R. communis* germinating ER (Section 3.3.3.2). This provided evidence that cup loading is the preferred method of sample introduction for an ER 2D DIGE experiment. In addition, an investigation with *A. thaliana* into the gel profile differences caused by different sample introduction methods found that although the profiles are generally similar, differences to the profile are caused by the different sample introduction methods (Section 3.3.3.1).

The use of CyDye labelled spikes was investigated as an aid to matching by acting as a bridge between the analytical CyDye labelled gel set and the unlabelled high protein-loading preparative gels. It was anticipated that in circumstances where experimental variation between the analytical and preparative gels causes difficulty in confidently matching spot pairs, the CyDye labelled spike sample could act as an intermediary. The spike sample, being subjected to the same experimental variation as the preparative gel, would allow it to act as a superimposable 'map' to the preparative gel and allow differences to be traced. The CyDye labelled spike would be incorporated into the strip alongside the preparative sample at the same amount used in the analytical gel. Critically, the preparative gel, the CyDye labelled spike and one of the analytical gels

in the 2D DIGE analysis, would be from a single protein preparation so all differences can be ascribed to experimental variation rather than biological variation.

To test this method, 12.5 μg of Cy3 labelled *A. thaliana* TSP was combined with an unlabelled 400 μg aliquot of the same protein from the same parent sample. This was in-gel rehydrated overnight in a pH 4-7 linear 18 cm IPG strip. A 400 μg replicate containing the same CyDye labelled sample was cup-loaded at the same time to test whether labelled spike samples could be superimposed under different protein introduction conditions. Both strips were focussed with the standard pH 4-7 protocol on an IPGPhor, alongside a third strip analytical strip cup loaded with 12.5 μg of the same Cy3 labelled *A. thaliana* protein as used for the CyDye spike (Table 3.3). On completion of focussing, strips were separated in the second dimension with a 17 hour 2 W per gel programme. After the second dimension run, the two CyDye labelled spike sample images and the CyDye analytical image were collected from the three gels with the Typhoon Variable Mode Imager. These are shown in Figures 3.13 and 3.14 for the 400 μg in-gel rehydrated and cup loaded gels respectively, alongside their analytical partner. The two preparative gels were then removed from their glass cassettes, fixed and SYPRO^{TM} Ruby stained overnight (2.3.7.4). After the destain procedure, gels were imaged once again on the Typhoon under the emission and excitation wavelengths for the SYPRO^{TM} Ruby stain, this time collecting the fluorescently stained preparative component of the gel, plus the contribution of 12.5 μg of protein in spike.

On comparison of the gel images it is clear that technically the inclusion of CyDye labelled spike samples works well. Visualisation of CyDye samples run within preparative gels does not appear to be effected by the presence of substantially more unlabelled sample (either in-gel rehydrated or cup loaded *A. thaliana* TSP). SYPRO^{TM} stained spot maps were obtained after imaging showing that this post-electrophoretic stain can still be used with Cy3 labelled samples.

Figure 3.13 shows the 12.5 μg Cy3 cup loaded analytical gel image alongside the 400

Table 3.3: Gel configurations for investigation into the use of CyDye labelled spikes in analytical and preparative gel matching

Gels	Cy3 labelled	Unlabelled	Sample introduction [b]	
	A. thaliana TSP [a]	*A. thaliana* TSP [a]	Cup loaded	In-gel rehydration
1	12.5 μg	-	$\surd$	-
2	12.5 μg	400 μg	$\surd$	-
3	12.5 μg	400 μg	-	$\surd$

[a] Cy3 and unlabelled samples both derived from same protein preparation
[b] pH 4-7 linear 18 cm Immobiline Drystrips (GE Healthcare, Bucks UK) were used for all samples

μg in-gel rehydrated image and 12.5 μg Cy3 spike image. Comparing the cup loaded Cy3 analytical gel with the in-gel rehydrated 400 μg unlabelled preparative sample reveals that although the protein profiles are broadly similar there are differences in the protein profiles. This is especially true of low molecular weight, low abundance proteins which are less clearly defined on the preparative gel compared to the analytical gel. An arrow highlighting the poorly resolved low abundance protein spots has been drawn on the preparative gel image. Without extra information matching these regions confidently would prove difficult. However, the low molecular weight, low abundance protein spots have resolved with much greater clarity on the CyDye labelled spike image. This indicates their poor representation on the preparative sample is most likely due to staining rather than absence of those spots from the gel. The CyDye spike image can thus be used as a superimposable map to help identify the correct spots on the preparative SYPRO^{TM} gel image.

Comparing the cup loaded Cy3 analytical gel with the cup loaded 400 μg preparative SYPRO^{TM} stained gel, minor electrophoretic abnormalities have caused a wave effect on the preparative gel profile (Figure 3.14, red arrow). Although the distortion is not severe, it does decrease the confidence in matching spots in the effected regions

between the preparative and analytical gels. Turning to the Cy3 labelled spike sample, the protein profile is effected in the same area confirming it is a problem introduced during IEF or second dimension electrophoresis, and not associated with labelling or staining of protein. As the Cy3 spike and the Cy3 analytical images constitute the same protein sample, the spike image can be used to deconvolute the protein profile. Another difference introduced between the analytical and preparative gels is the presence of an extra spot on the 400 μg preparative gel (Figure 3.14, white arrow) which appears to be entirely absent on the analytical gel. With no further information available it would not be possible to confidently say whether its presence on the preparative gel is due to differential staining with SYPRO^{TM} or whether it is a vertical streaking artefact due to the high loading of protein on the gel. If it is due to differential staining then its absence from the analytical set would have implications for the comprehensiveness of the analysis. Turning to the Cy3 labelled spike image shows the extra spot is present on this gel also, meaning its absence cannot be ascribed to differential staining. There is vertical streaking present in the effected area on both the analytical and spike images suggesting this is the cause of the new spot and that it can be viewed as an artefact.

3.3.3.4 Development of a robotic spot picking procedure

Automated spot picking is highly desirable where you require more than a few spots to be extracted from a gel for MS analysis. Robotic spot picking brings significant advantages of speed and convenience over manual picking. It also has the advantage that by minimising gel handling the risk of keratin contamination is reduced, something which is critical in improving the quality of MS data obtained in an analysis. It also allows rapid picking of spots from SYPRO^{TM} Ruby stained gels, something which would otherwise require the undesirable use of a transilluminator for potentially extended periods.

Spot picking was performed with a ProPic^{TM} gel picking robot (Genomic Solutions

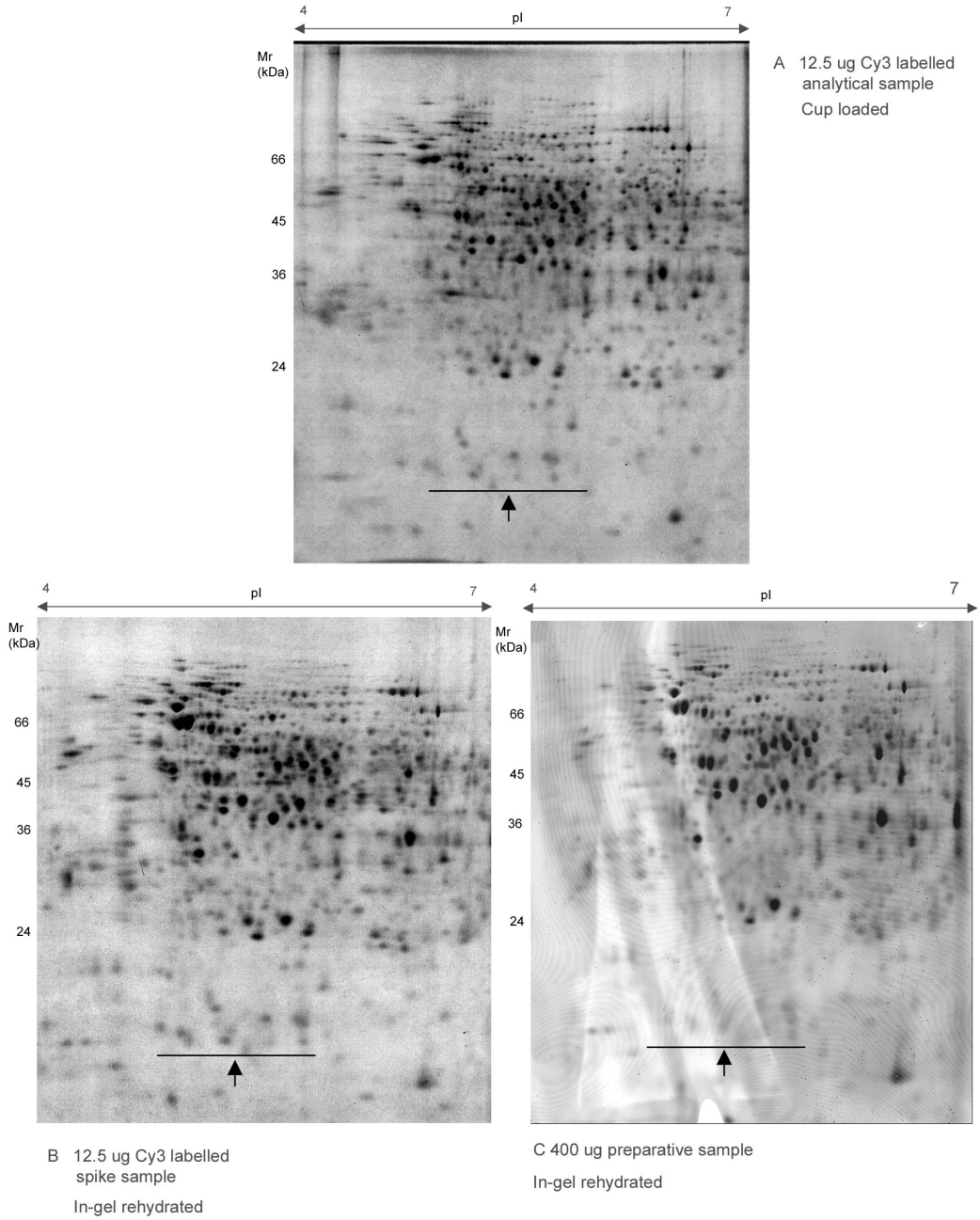


Figure 3.13: A 12.5 μg Cy3 spike image aids matching between a 12.5 μg cup loaded Cy3 analytical gel and an in-gel rehydrated 400 μg SYPRO^{TM} stained preparative gel. The Cy3 spike image (bottom left) co-migrated with the preparative gel (bottom right) and allowed differences between the preparative and analytical profile to be traced. For example, an arrow indicates a region of reduced complexity in the preparative gel compared to the analytical gel. However the Cy3 spike image has the same profile of spots in this region as the analytical gel indicating their absence is more likely due to post staining rather than electrophoretic differences. Gels are 12% acrylamide.

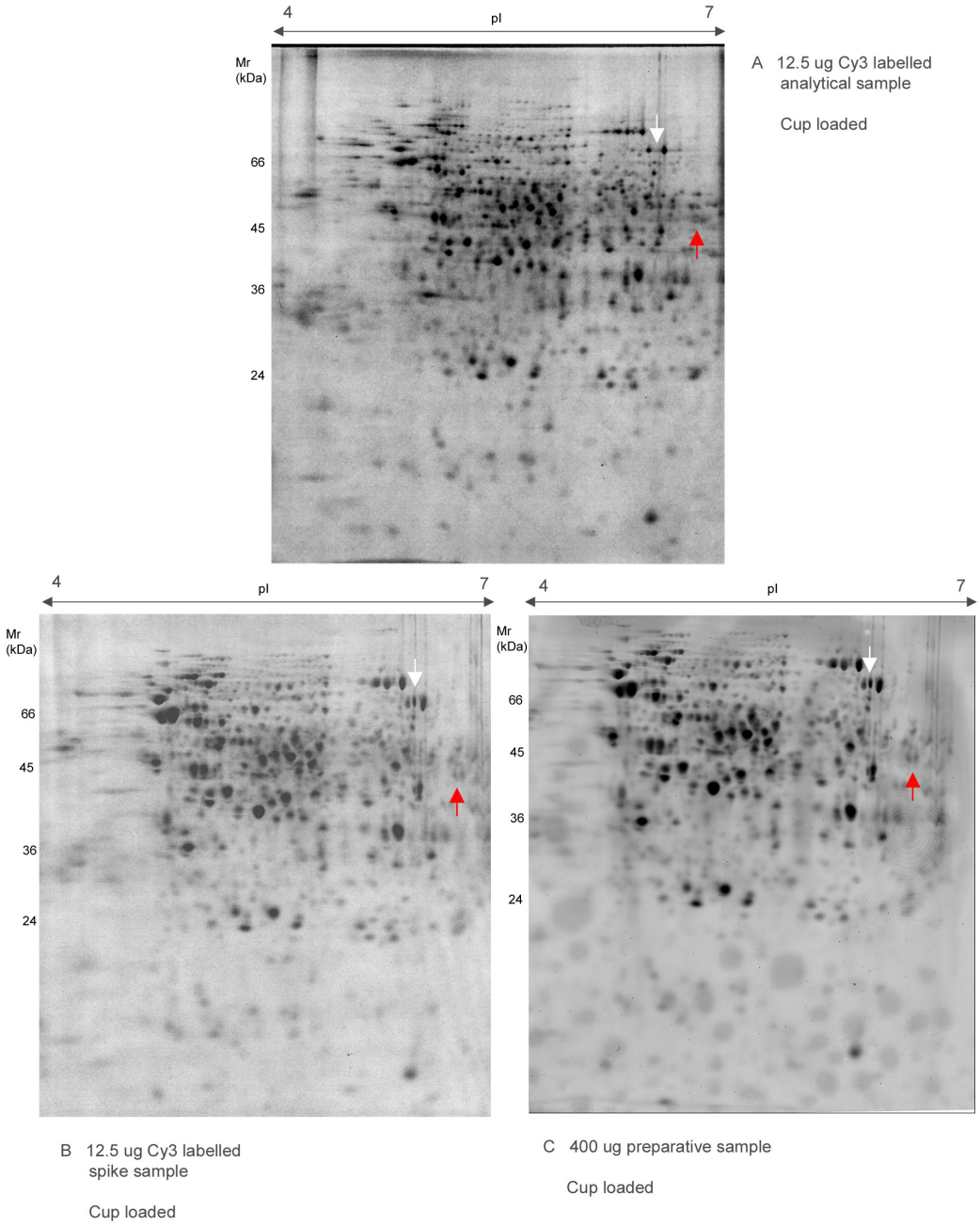


Figure 3.14: A 12.5 μg Cy3 spike image aids matching between a 12.5 μg cup loaded Cy3 analytical gel and a cup loaded 400 μg preparative gel. The Cy3 spike image (bottom left) co-migrated with the preparative gel (bottom right) and allowed differences between the preparative and analytical profile to be traced. For example, the white arrow indicates a spot present on the preparative gel but absent on the analytical gel. Its presence on the CyDye spike image indicates it is a streaking artefact rather than a protein unlabelled on the analytical gel. The red arrow indicates a region of the preparative gel which has run poorly. The CyDye spike image has been affected in the same way indicating it is a problem with electrophoresis rather than with the sample or staining and helps de-convolute the affected area.

Ltd. Cambridgeshire, UK) controlled by Phoretix Evolution software (Nonlinear Dynamics Ltd., Northumberland, UK). The ProPic^{TM} is currently incompatible with the DeCyder 2D DIGE analysis platform; a technology that has become the method of choice for quantitative gel based proteomics. As pick lists cannot be directly exported from DeCyder to the ProPic^{TM}, spots to be picked need to be re-selected in Phoretix Evolution which generates a list of co-ordinates which are exported to the picking robot.

For a small number of spots (sensibly, less than 15) the preparative gel to be picked is imaged by the ProPic robot, and spots of interest mapped to this image before picking. During this process, the gel remains in situ. As the process of mapping spots of interest to the gel image can be a lengthy process especially where an experiment has hundreds of spots to be picked, it is preferable for the preparative gel to remain in fixative during this process. This prevents drying and warping of the gel which could lead to picking inaccuracies. With the ProPic/Phoretix Evolution system, once the picking locations have been assigned to the gel image, the preparative gel is re-positioned in the ProPic robot and a secondary image captured. Spots are then triangulated from the gel image with the spots of interest assigned to the new secondary image. Co-ordinates are generated based on this triangulation and exported to the ProPic robot to initiate picking. Three or more distinct outlying spots are required to act as triangulation landmarks and these must be identifiable on both images for triangulation to occur.

A significant problem identified was the identification of suitable triangulation landmarks. As triangulation only occurs within the boundaries denoted by the landmarks, they had to be spots outlying all those to be picked. In addition they also had to be clear enough to be visualised on the low resolution images produced by the ProPic^{TM} robot. If triangulation landmarks were inaccurately identified this would cause misalignment of picking across the gel. Furthermore, speed is an important factor because whilst picking landmarks are being identified and vectors adjusted, the gel is drying out and shrinking exacerbating pick misalignment.

To improve upon this, new triangulation methodologies were developed. The problem of landmark selection was circumvented by the addition of artificial landmarks to the gel. Three artificial landmark methods were tried. Firstly, adhesive white paper tabs were placed upon the gel at locations outside of the spots to be picked. The tabs were first marked with cross hairs to act as targets for triangulation. Although it proved that paper tabs could be visualised on scanning with the Typhoon imager and the ProPic^{TM} robot camera, the method was discarded when some adhesive tabs came away from the gel surface whilst the gel was sitting in ddH_2O in its tub. The next method attempted involved boring out small holes within the gel at locations outside of the spots to be picked and pipetting SYPRO^{TM} Ruby stain into the holes, the idea being that these concentrated dots of stain would clearly fluoresce both under Typhoon and ProPic^{TM} imaging. In practice the staining solution did not fluoresce clearly enough to act as an easily identifiable landmark (result not shown). The permanence of the holes made in the gel was an improvement over the adhesive tabs and was combined with the use of white paper inserts that were found to be clearly identifiable both with Typhoon and ProPic^{TM} scanning. It is important to note that within the Phoretix software locations identified as triangulation markers are added to the pick list and are picked by the robot alongside other spots. To avoid damage to the robot's picking head, the paper markers should be removed before the picking run is initiated. An alternative method is to manually edit the picking co-ordinate file and deleting the entries corresponding to the triangulation markers. This has the advantage of not having empty 'samples' amongst the picked spots, and avoids wasting time and reagents in the downstream digestion and MS analysis. The process of linking 2D DIGE-based analysis in DeCyder to ProPic/Phoretix Evolution picking and the use of external triangulation markers is depicted in Figure 3.3.3.4.

It was found that in extended picking runs the accuracy of picking decreased the longer the picking run continued. On examination of the gel during the course of picking

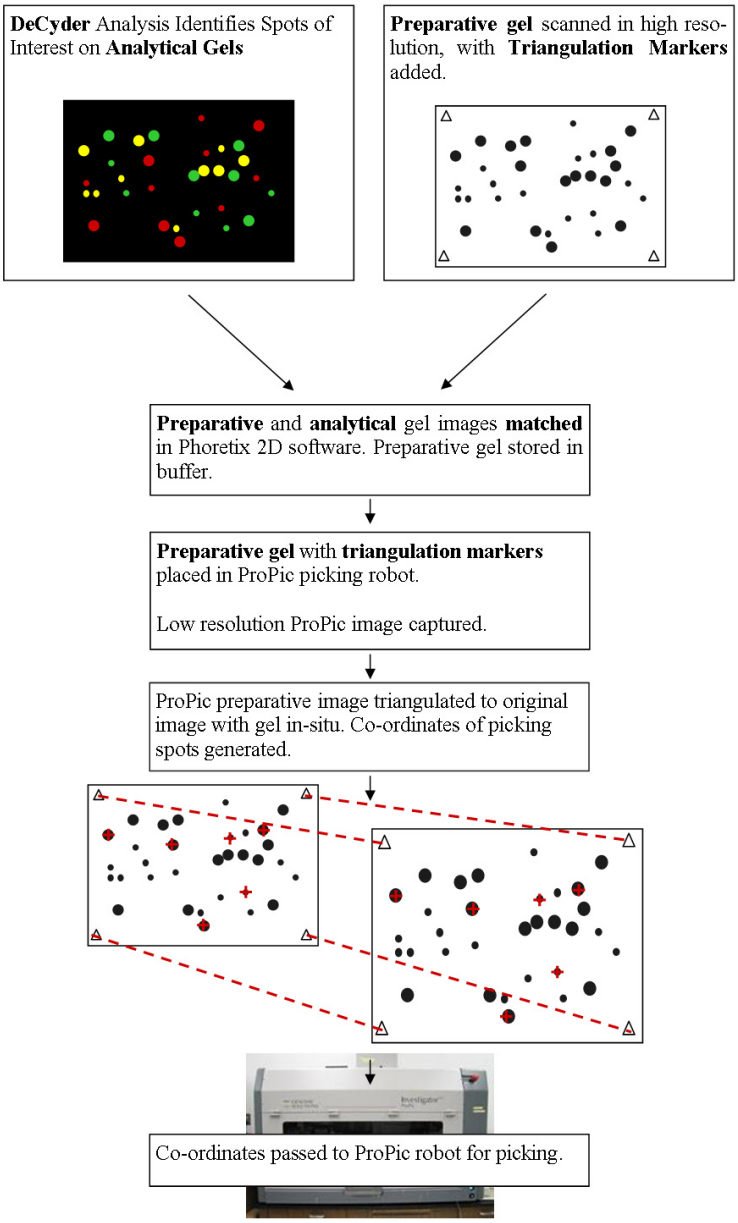


Figure 3.15: DeCyder to ProPic picking schema. The DeCyder analysis generates a list of spots of interest, to be picked from the external preparative gel. The external preparative gel is imaged with triangulation markers in place. In Phoretix Evolution spots of interest from an analytical gel are matched to the preparative gel. Once this is complete, the preparative gel is placed in the ProPic robot with triangulation markers in place, imaged once again, and spots to be picked triangulated to the new image.

it was found that slight movements to the gel could occur by the impact of the picking head into the gel surface. Furthermore gel drying and shrinkage during longer runs increased the deviation of predicted spot co-ordinates from actual spot locations. This was overcome by adhering one side of the gel to a glass plate by applying Bind Silane (γ-methacryloxypropyltrimethoxysilane, GE Healthcare, Bucks UK) to the gel cassette before casting. A firmware update to the robot was made by Genomic Solutions, allowing the robot's picking head to move not just on the vertical axis but also slightly on the horizontal axis in a movement termed a 'pick shift'. When picking the gel in pick shift mode, the picking head enters the gel at the dictated location but then moves horizontally for 0.7 mm breaking the adhesion and allowing excision of the gel plug from the gel. Over extended runs, accuracy of picking was maintained with this adjustment. A preparative gel alongside an image of the same gel post-picking is presented in Figure 3.16 and shows accuracy of pick targeting.

In summary, a combination of fluorescent markers included in the gel, the use of Bind Silane to fix the gel to a glass plate and the upgraded robot firmware allowing pick shifts has allowed highly accurate picking of multiple spots to be made over an extended period of time, with a decreased requirement for human involvement in the picking process.

3.4 Discussion

3.4.1 Development of a sample preparation methodology

The results presented here detail the assessment of *R. communis* cv. 99N89I as a variety suitable for laboratory research through the evaluation of its growth characteristics, the establishment of sample preparation procedures and the evaluation of the reproducibility of the ER protein sample.

99N89I was grown successfully from seed in the laboratory to produce 3 day ger-

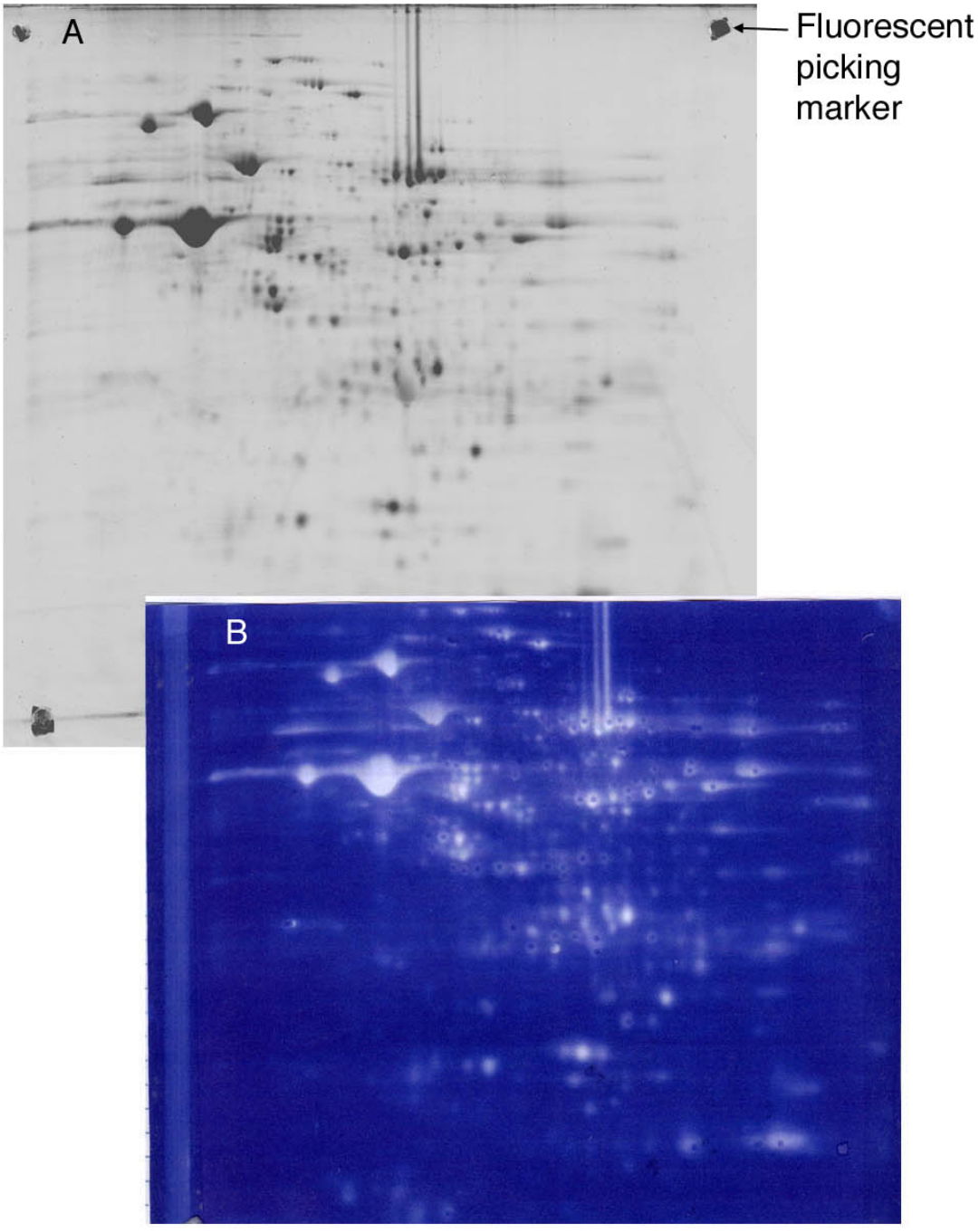


Figure 3.16: Optimisation of spot excision methodology. An example of a developing ER preparative pick gel with fluorescent triangulation markers, imaged on the Typhoon Variable Mode imager, is shown (A). The identification of clear external picking markers, along with Bind Silane backed acrylamide gels and a pick-shift enabled robot, allowed extended picking runs to excise spots with greater accuracy. (B) shows the same developing ER preparative gel post-pick with all spots picked within their boundaries.

minating seedlings and mature plants from which developing seed was obtained. Germination was found to be efficient for this variety allowing maximisation of seed in germinating preparations. The morphological characteristics of seed development were consistent with those published for cv. Hale (Greenwood and Bewley, 1982), allowing the harvesting of developing endosperm material of the stage shown to have peak mRNA levels of the Kennedy pathway enzyme DAGAT2 (Kroon et al., 2006) in this variety and peak abundance of the fatty acid biosynthetic enzyme ACCase (Simcox et al., 1979).

Preparation of this variety to yield purified ER using previously published protocols (Coughlan et al., 1996; Maltman et al., 2002) gave clearly defined bands at the 20% (w/w) / 30% (w/w) sucrose interface (density = 1.12 g/cm^3) after 2 hour and 22 hour centrifugation. On analysis of protein yields from the preparation it was found that the quantities obtainable were amenable to proteomic analyses, especially with the increased sensitivity offered by CyDye technology.

SDS PAGE analysis of germinating and developing 99N89I endosperm membrane samples suspended in 10% glycerol and solubilised in SDS PAGE running buffer showed high levels of reproducibility. On comparison of the developing sample with an SDS PAGE gel of developing ER from cv. Hale (Maltman et al., 2002) broad similarities were identified between the two profiles, with characteristic high abundant bands aligning with the same distribution pattern. In this earlier work with cv. Hale, the ER resident ricinoleic acid biosynthesis enzyme oleate $\triangle$12 hydroxylase was identified. The similarity of the Hale and 99N89I developing ER SDS PAGE profiles, along with the marker assay evidence of ER purity with samples generated by the same preparation methodology, gives strong evidence that the preparations described are endoplasmic reticulum.

3.4.2 Establishment of analytical procedures for 2D DIGE

A. thaliana protein labelled with CyDye under optimal reaction conditions of 50 μg protein : 400 ρmol of dye in reaction volume of 20 μl (i.e. 2.5 mg/ml) was compared to the same protein sample labelled at the same ratio of 50 μg protein : 400 ρmol of dye but in a reaction volume of 84.7 μl (0.59 mg/ml). The dilute reaction volume mimicked the lowest concentration of *R. communis* ER sample in store. On evaluation of 2DE gels of the two samples by simple overlay the profiles were found to be complete superimposable, indicating that the same complement of proteins were labelled at both levels of dilution. Also, there was no discernible deviation from the composite black colour of the overlaid blue/red images, indicating no significant difference in degree of labelling between the two dilution states.

The gels were also quantitatively analysed by the DIA component of the 2D DIGE software, which identified a single protein spot that had a $\geq$ 1.1 fold increase in abundance on the non-optimally labelled gel. All other spots were within a $\pm$ 1.1 fold change and were deemed similar.

3.4.3 Evaluation of preparative procedures to improve spot matching

The ability to confidently match the same spots on analytical gels and a preparative partner is of paramount importance to 2DE proteomic analyses using MS to obtain IDs. To reduce potential variation between the preparative and analytical gel profiles, variables between the two gel types should be minimised wherever possible. Ideally, the preparative gel would contain the same protein sample as an analytical gel, so biological variation cannot contribute to difficulties in matching between the two gels. Similarly, technical variables would be minimised where possible, for example the sample introduction method would be kept consistent between the analytical and preparative gels so this could not contribute to any variations between the protein profiles. As described in section 3.3.2, the comparatively dilute concentration of *R. communis* ER samples and

their pre-solubilisation in 2DE lysis buffer required adaptations to the ideal protocol for CyDye labelling of protein samples. The dilute nature of the *R. communis* ER samples also means preparative size (200 μg to 1 mg) protein samples cannot be introduced into an IPG strip with the same preferred method for analytical gels: cup loading. So the potential for variation in protein profile caused by different sample introduction methods could be understood before committing valuable *R. communis* ER material to a large scale proteomic study, a pre-evaluation with *A. thaliana* TSP was performed.

Cup loading was found to give an improved resolution of basic spots, but a slightly inferior resolution of low abundant acidic spots on the gel profile compared to in-gel rehydration, whether at 200 μg or 400 μg loadings. Also, in-gel rehydrated samples seemed more prone to horizontal streaking in the basic region of the gel, but this was alleviated somewhat at the higher protein loading. Overall, with *A. thaliana* TSP, the protein profiles were similar and although based on these observations cup loading would remain the method of choice for preparative gels, there was no significant profile differences introduced by in-gel rehydration that disqualified it as a technique. With *R. communis* sample superior resolution was again found for the cup loaded gel. The comparison of sample introduction method was made with CyDye labelled ER at analytical loadings. If no difference had been observed between analytical *R. communis* samples, the most suitable choice would have been to in-gel rehydrate both the analytical and preparative gel, as this would remove the sample introduction technical variable. As the resolution is superior for cup loading at analytical protein amounts the best choice would be to cup load the analytical gels and in-gel rehydrate the preparative gel.

The use of a small spike of CyDye labelled analytical sample run alongside the unlabelled preparative sample was investigated using *A. thaliana* TSP protein. Imaging of the spike sample before post-electrophoretic staining of the preparative sample was possible, giving a clear 2DE gel profile without interference from the unlabelled spots.

The CyDye spike image can be thought of as a map of gel-to-gel variation between the preparative and analytical gels. Thus, where the same sample is used for the analytical gel, the CyDye spike image and the preparative gel, it allows variation caused by preparative methodological differences such as protein staining and higher loadings to be identified. This may inform picking strategies where a key protein has not stained effectively with SYPRO^{TM} stain but is known to be present on the gel by its visualisation on the CyDye spike image. The use of triangulation would allow the picking of this 'absent' spot.

3.4.4 Development of a new spot excision methodology

A new method for accurate and rapid spot excision from preparative gels has been developed. This method utilises external triangulation landmarks which are easy to locate and always encompass the entire spot map. This offers substantial improvements over the standard method of triangulation as it was often difficult to find appropriate internal landmarks that met the criteria of both lying outside all spots to be picked and being sufficiently abundant to be identified on the low resolution images produced by the ProPic^{TM} camera. When the use of external landmarks was coupled with Bind Silane backed gels and the use of the firmware-upgraded ProPic that allows the pick head to move horizontally as well as vertically, highly accurate picking was obtained over extended picking runs.

3.5 Concluding remarks

R. communis cv. 99N89I has been established as a suitable variety for proteomics-based experimental research, allowing the routine growth, harvest and preparation of seed material for the purposes of endosperm ER purification. Methodological developments have been made to sample preparation techniques which provide the best combination

of sample preservation and 2DE compatibility.

Constraints specific to the *R. communis* samples imposed deviations from standard 2D DIGE proteomic methodology, specifically with regards to the labelling protocol and the preparative gel. Investigations were made using *A. thaliana* TSP and *R. communis* endosperm ER and the potential for problems caused by these deviations are now well understood. A thorough and rigourous experimental method is now in place, from plant material generation to spot picking, allowing *R. communis* endosperm ER samples to be committed to an extensive 2D DIGE analysis.

Chapter 4

A differential proteomic analysis of germinating and developing seed ER from *Ricinus communis*

4.1 Introduction

The process of endosperm development during seed formation induces temporally and spatially distinct stages of carbon deposition (Emes et al., 2003) whereby carbon sources are laid down to support the seedling on germination. *Ricinus communis* endosperm development is similarly concerned with high level carbon storage, including significant establishment of lipid reserves through plastidal fatty acid biosynthesis and subsequent esterification of fatty acids into TAG storage molecules in the ER. TAG molecules comprise up to 50% of *R. communis* seed weight and uniquely over 90% of this oil is comprised of ricinoleic acid (McKeon et al., 1997). Ricinoleic acid formation occurs in the ER, where oleolyl CoA is esterified to phosphatidyl choline and hydroxylated at its 12-carbon position. The reaction is catalysed by oleate $\triangle 12$ hydroxylase (FAH12), an NAD(P)H dependent enzyme cloned from *R. communis* (van de Loo et al., 1995). Expression of the gene encoding this enzyme in *Arabidopsis thaliana* resulted in a range of hydroxy fatty acid profiles in the seed lipid, however the maximum amount of hydroxy fatty acids in the TAG of any transformant was just 19.2% (Smith et al., 2003). One reason for this might be the substrate specificities of the acyl transferase enzymes involved in construction of TAG molecules from fatty acid precursors. For example, it is not unreasonable to expect *A. thaliana* to have evolved acyl transferases that efficiently incorporate its wildtype complement of acyl CoAs (specifically, 16:0-CoA, 18:0-CoA, 18:1-CoA, 18:2-CoA, 18:3-CoA, 20:0-CoA and 20:1-CoA) but that are less able to catalyse the transfer of unfamiliar hydroxy fatty acyl CoAs such as ricinoleoyl (18:1-OH) CoA. Conversely *R. communis* acyl transferase enzymes are likely to be well adapted at catalysing esterification of ricinoleoyl CoA into TAG molecules. Through the identification of further components of lipid biosynthesis in *R. communis* endosperm such as its complement of acyl transferase enzymes a clearer picture of the biosynthetic pathway can be constructed. As the key proteins involved in high ricinoleate TAG

production are identified, their respective genes can be cloned and transgenic studies in model organisms such as yeast and *A. thaliana* performed, moving toward the ultimate aim of the project : production of high triricinoleate-yielding transgenic crops.

With a view to identifying components of TAG biosynthesis in *R. communis* endosperm, developing and germinating seed ER was purified and used in a differential proteomic screen of urea soluble ER. 2-dimensional in-gel electrophoresis (2D DIGE) was employed whereby proteins were labelled with fluorescent CyDyes. This method offered significant advantages over traditional 2D electrophoresis (2DE) techniques, such as multiplexing of samples and experimental design centred around multireplicate statistical analysis. Where protein spots were significantly elevated in one developmental state compared to the other, they were robotically-picked from a high loading preparative gel, digested with trypsin and analysed by MALDI TOF MS and MS/MS. This investigation confirmed the previously reported presence of abundant chaperones and folding-proteins in this compartment (Maltman et al., 2002), added further insight into the processing and importance of storage proteins in the ER but did not reveal the presence of lipid biosynthesis enzymes in the urea-soluble fraction.

4.2 Aims

There were two aims of the work described in this chapter. Firstly to identify and quantify statistically significant differences between the urea-soluble fractions of germinating and developing seed ER by means of a differential proteomic analysis. Secondly to obtain identities of these significant differences by MALDI TOF PMF and MS/MS sequencing, with a view to locating components of lipid biosynthesis pathways in the 2DE lysis buffer soluble fraction.

4.3 Results

4.3.1 Sample preparation and experimental conditions for 2D DIGE

The establishment of reproducible ER preparations from *R. communis* endosperm and the perfection of 2D electrophoresis techniques facilitated a large scale differential analysis of germinating and developing seed ER. New developing and germinating preparations were made for this purpose. Purified ER membranes were centrifuged to yield a pellet which was then re-suspended into lysis buffer for 2DE. A minimal volume of lysis buffer was used (typically 100-400 μl depending on pellet size). It is recommended that for CyDye labelling the protein concentration should be ideally 2.5 mg / ml and not less than 1 mg / ml as this may effect the reaction kinetics of the dyes derivatisation to protein. Although there was no clear evidence that derivatisation of dye to protein was effected with protein concentrations less that 1 mg / ml with *A. thaliana* TSP (3.3.2), sample quantity constraints meant this was not tested with *R. communis* ER samples. The generation of new ER preparations meant that samples could be made at concentrations above the recommended minimum 1 mg / ml concentration.

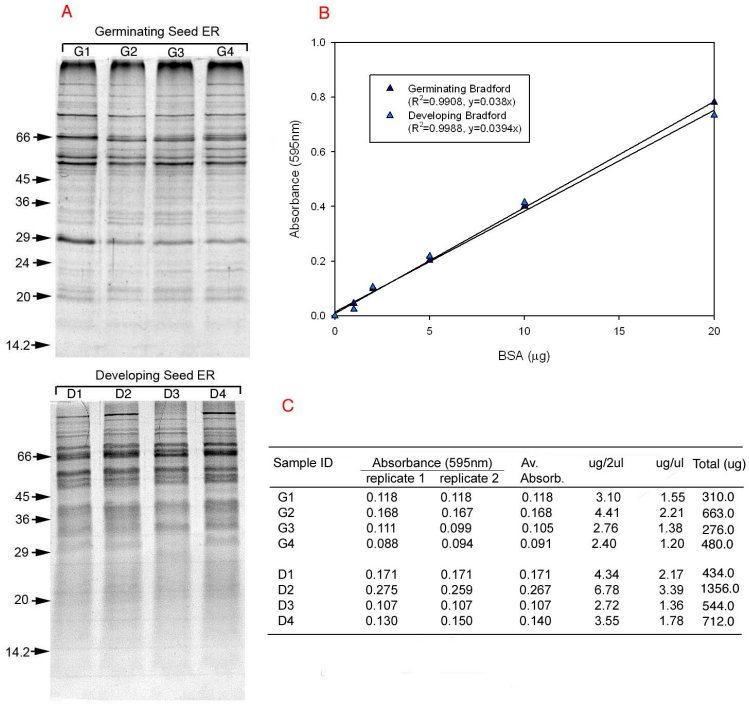
The concentration of protein within the ER samples was assayed using the detergent-compatible modified Bradford protein assay. The quantification of protein amount ensured all the samples were sufficiently concentrated for CyDye labelling and that the total amount of protein was enough for a 2D DIGE experiment. The assay was performed as described (Chapter 2.3.2) comparing the absorbances of ER samples to that of a standard curve (Figure 4.1, B) generated separately for the developing and germinating preparations. Figure 4.1 (C) shows the results of the modified Bradford assay of 8 new preparations destined for 2D DIGE analysis. All preparations were at a protein concentration $>$ 1 mg / ml, meaning they were able to be labelled under the recommended reaction conditions for 2D DIGE. Germinating sample G2 and developing

samples D2, D3 and D4, all contained greater than the minimum 500 μg of total protein required to provide 400 μg of protein for preparative gels in addition to the required 75 μg protein for the analytical gels (50 μg for the Cy3 / Cy5 analytical gel, 25 μg contribution to the Cy2 internal standard) and 25 μg for protein quantification, SDS PAGE validation and excess volume to prevent pipetting error.

Before committing CyDye label to our ER preparations it was critical to confirm sample reproducibility. 5 μg aliquots were taken from four germinating and four developing samples and separated by 1D SDS PAGE (Figure 4.1, A). All four germinating and developing samples displayed good preparation reproducibility and were prepared for labelling with CyDyes.

50 μg of protein from each sample was labelled with 400 pmol of CyDye at a concentration of 1 mg/ml (see Chapter 2.3.5). The 200 μg pooled standard sample was labelled with 1600 pmol of Cy2 (Table 4.1). The four germinating samples were labelled with Cy3 dye and the four developing samples were labelled with the spectrally distinct Cy5 dye. 25 μg from each germinating and developing sample were pooled and labelled with Cy2 to generate the internal standard. The success of the labelling reaction was tested by separation of 5 μg of each labelled protein sample (samples G1 to G4, D1 to D4 and Std) by 1D SDS PAGE. The gel was then visualised with the Typhoon 9200 Variable Mode Imager and total intensity for each protein lane checked with ImageQuant (GE Healthcare, Bucks UK) (Fig. 4.2). This provides an initial check for labelling failures by identifying significant differences in fluorescent intensity between samples and by comparing protein profiles. Germinating and developing sample types gave comparable protein profiles and figures for pixel intensity indicating that the efficiency of labelling was conserved across the 9 separate reactions (Fig. 4.2).

Following labelling, samples were prepared for large format 2D electrophoresis. 12.5 μg of a single germinating sample was paired to 12.5 μg of a single developing sample in the order described in Table 4.1. To each of these paired samples a further 12.5



C

Sample ID	Absorbance (595nm)		Av.	ug/2ul	ug/ul	Total (ug)
	replicate 1	replicate 2	Absorb.			
G1	0.118	0.118	0.118	3.10	1.55	310.0
G2	0.168	0.167	0.168	4.41	2.21	663.0
G3	0.111	0.099	0.105	2.76	1.38	276.0
G4	0.088	0.094	0.091	2.40	1.20	480.0
D1	0.171	0.171	0.171	4.34	2.17	434.0
D2	0.275	0.259	0.267	6.78	3.39	1356.0
D3	0.107	0.107	0.107	2.72	1.36	544.0
D4	0.130	0.150	0.140	3.55	1.78	712.0

Figure 4.1: Protein concentration measured by modified Bradford assay. Absorbances of germinating and developing samples were assayed independently against a BSA standard linear regression (1 - 20μg) **(B)**. Assay data is shown in the table **(C)**, which reveals all samples are at a concentration $>$ 1 mg / ml. Preparations were examined by 1D SDS PAGE on a 12% gel **(A)** to assess sample consistency, which were found to be highly reproducible.

Table 4.1: Set up of labelling reactions and multiplex gel configurations for 2D DIGE study

Sample ID	Sample Type	Dye	Protein		Gels
			concentration (μg/μl) [a]	amount (μg)	
G1	Germinating	Cy3	1.55	50	1
G2	Germinating	Cy3	2.21	50	2
G3	Germinating	Cy3	1.38	50	3
G4	Germinating	Cy3	1.20	50	4
D1	Developing	Cy5	2.17	50	1
D2	Developing	Cy5	3.39	50	2
D3	Developing	Cy5	1.36	50	3
D4	Developing	Cy5	1.78	50	4
Std	Internal Standard [b]	Cy2	1.88	200	1 to 4

[a] Protein concentration $>$ 1 mg/ml to ensure kinetics of labelling reaction are optimal.

[b] Internal standard sample comprised of 25 μg from samples 1 to 8 to give a total of 200 μg and containing equal representation of all germinating and developing spots in the analysis.

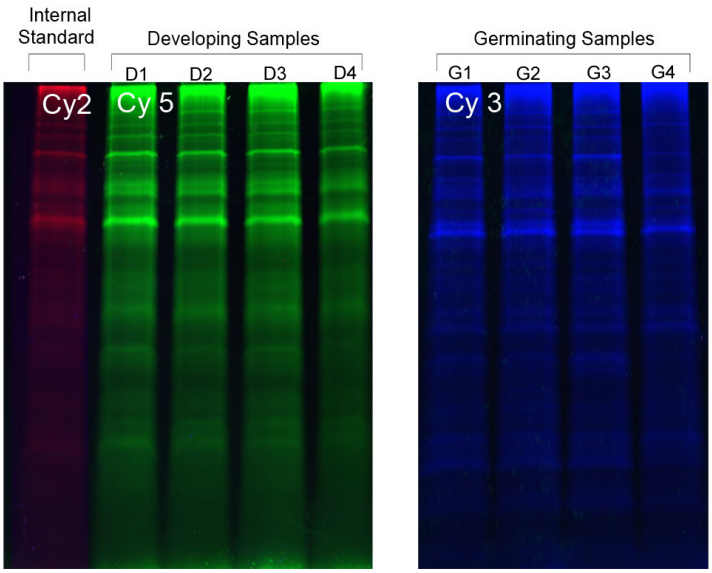


Figure 4.2: 1D SDS-PAGE gel of CyDye labelled protein samples. Labelling efficiency was compared for each protein lane using ImageQuant software to compare protein profiles and intensity, which revealed effective labelling of each sample.

μg of the Cy2 labelled internal standard was added. Finally, lysis buffer was added to a total volume of 70 μl containing 1% DTT and 2% pH 3-10 carrier ampholytes. Samples were cup-loaded at the anodic end of a re-swelled 18 cm pH 3-10 IEF strips and isoelectric focussing was performed for a total of 70 kVh (see Chapter 2.3.5). Strips were then equilibrated and proteins separated by their molecular weight on large format 12% acrylamide gels. Imaging of the gels was carried out immediately after the second dimension dye front had reached the bottom 2 mm of the gels. Gels were visualised using a Typhoon 9200 variable wavelength scanner, capable of imaging the Cy2, Cy3 and Cy5 fluors in an automated 3 phase scan. It was critical that the peak pixel intensity for each scanned image did not exceed 100,000 (on a scale of 0-100,000 greys). Above this value the image is at least partially saturated and cannot accurately be quantified. So as to make good use of the dynamic range of the extended 16-bit file format, PMT values were adjusted until a maximum pixel value of ~70,000 was

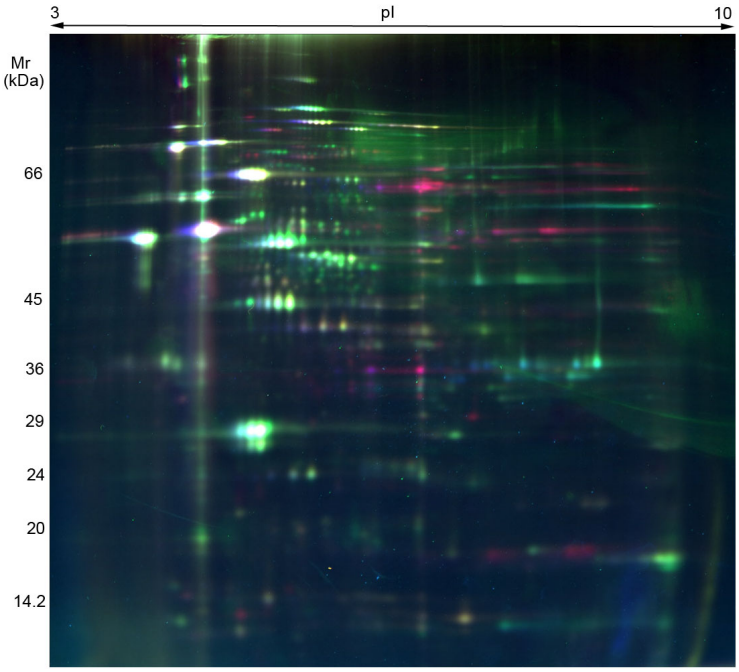


Figure 4.3: Representative image of 2D DIGE gel. Four replicates of germinating and developing seed ER were labelled with CyDyes and separated by 2D Electrophoresis. Imaging of the CyDyes was performed using the Typhoon Variable Mode Imager. Triple overlay image consists of Cy3 (developing ER, green), Cy5 (germinating ER, red), Cy2 (pooled standard, blue).

achieved. A 2DE gel is depicted in Figure 4.3. This is gel 1 (see Table 4.1 for gel list) containing the germinating seed ER sample G1, the developing seed ER sample D1 and the internal standard (Table 4.1). It is a representative triple overlay of each CyDye fluor.

4.3.2 Assessment of gel reproducibility

Before analysis of the analytical datasets to identify protein spots of interest, assessments were made of the reproducibility of the gels. As the sample preparation method utilised sucrose-gradient based subcellular fractionation of plant material rather than whole cell lysates (total soluble protein) the potential for differences between samples is increased. This is because to obtain ER of sufficient purity whole cell

extracts are fractionated through two separate sucrose gradients in a 36 hour process with a large degree of technical involvement; thus there is an increased potential for variability through sample losses, human error or sample degradation. Although sample reproducibility was good on 1D SDS PAGE gels and CyDye labelling was successful, problems with the IPG strip, focussing and sample entry into the second dimension can all effect the 2DE profile and reduce the quality of the data obtained. Therefore it was important to ensure that the protein samples had resolved well across the four germinating and developing gel sets. On visual examination of the gel images they appeared broadly reproducible and the protein spots had resolved well across the gels. However using the sophisticated gel analysis packages available a quantification of variance was made.

After collection of gel images with the Typhoon Variable Mode Imager the four developing ER gel files were imported into Phoretix Evolution gel analysis software. Spot features were detected with an automated detection algorithm and manually adjusted where required (see Chapter 2.3.9.2).

The first step in quantifying the reproducibility between the gels was in identifying the total number of unique spot features present on the 4 developing gels. This was done within Phoretix Evolution by creating a virtual reference gel from the spot map with the greatest number of spots. Where spots were present on the remaining three gels that were not already on the reference map they were manually added until the reference map was a complete composite of all 4 developing gels. All of the spot features present on each of the 4 developing gels could then be matched to the reference map. This then allowed the calculation of match frequency for every spot present across the developing gels. By counting how many spots on the reference map were able to match spots in all 4 of the developing gels a figure was obtained of the number of spots common to every gel in the developing set. Similarly, figures were obtained for the number of spots common to 3 of the 4 gels, 2 of the 4 gels and in just 1 gel. The numbers for each

Table 4.2: Reproducibility of the germinating and developing analytical gels assessed by Phoretix Evolution. The total number of unique spot features across all four gels in each developmental state was identified by the creation of a composite reference map. The number of these unique spot features found in every gel, in 3 / 4 gels, in 2 / 4 gels and just a single gel was then calculated. It was found that for the developing samples, 89% of all the identified spot features were present in each gel. In the germinating samples, 83% of spot features are common to every gel.

Developing gels					
Spots common to x / 4 gels	4/4	3/4	2/4	1/4	Total
No. of spots	348	26	10	5	389
Percentage of total	89%	7%	3%	1%	100%

Germinating gels					
Spots common to x / 4 gels	4/4	3/4	2/4	1/4	Total
No. of spots	261	28	20	6	315
Percentage of total	83%	8%	7%	6%	100%

of these categories could then be expressed as a percentage against the total number of spots on the reference gel. The composite reference gel was found to contain 389 spot features. Of these 389 distinct spots 348 were present in every gel, equivalent to 89% commonality (represented by arrows on Figure 4.4). 26 spots were identified across 3 gels (7% of total), 10 spots were identified across 2 gels (3% of total) and 5 spots were present in just a single gel (1% of total). The analysis procedure was repeated for the germinating dataset, where a total of 315 distinct spot features were identified. 83% of these spot features were found to be present in all 4 of the gel images, which represents 261 out 315 spots. 28 spots were found to be common to 3 out of the 4 gels (8% of total), 20 spots were identified in just 2 of the gel images (6%) and 6 spots (2%) were found to be present in just a single gel image (Table 4.2).

When analysing gels by the DeCyder 2D DIGE software one of the criteria commonly used for identifying statistically significant spots is that they are present in every gel. Therefore, to maximise the amount of data to be assessed in a statistically rigourous

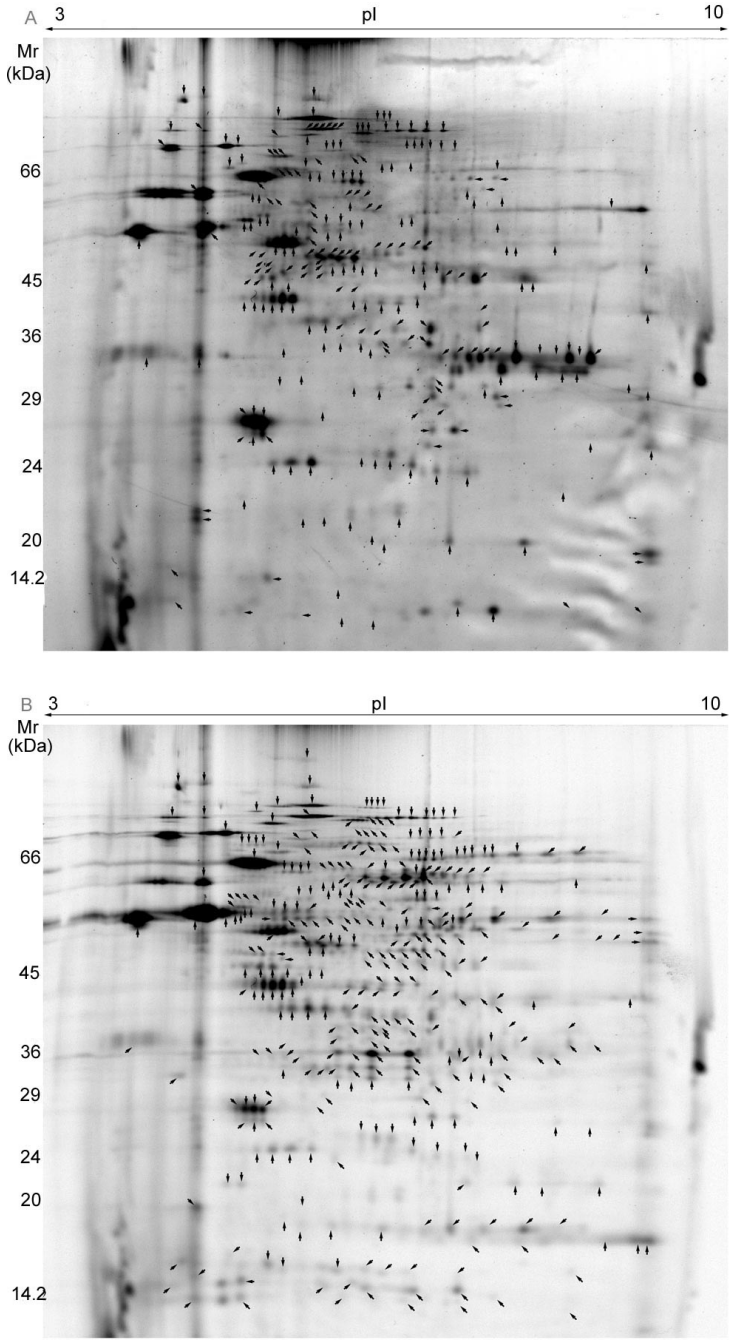


Figure 4.4: Assessment of 2D DIGE gel reproducibility using Phoretix Evolution (Nonlinear Dynamics, Northumbria, UK). Arrows indicate spots common to all gels. Gel A, top, is a germinating gel, arrows indicate 83% of spot features that are common to all germinating gels. Gel B, bottom, is a developing gel, arrows indicate 89% of spots common to all 4 gels

differential analysis, it is important to ensure the highest level of reproducibility between samples is obtained. The identification of 89% of spots common to developing gels and 83% of spots common to germinating gels was considered sufficiently reproducible to continue to a DeCyder analysis for the differential analysis of germinating and developing ER seed profiles. The analysis proved a useful tool in evaluating reproducibility within a sample set (i.e. the germinating or developing gels) and gave confidence in the quality of the dataset by quantifying the degree of commonality across the gels, before committing gels to DeCyder differential and downstream MS analysis.

4.3.3 Identification of differential protein levels between germinating and developing samples

Having made an assessment of the reproducibility of four germinating and developing seed ER preparations by 1D SDS PAGE and large format 2DE, gel images were analysed by DeCyder 2D DIGE software version 5.00.08 (GE Healthcare, Bucks, UK). The analysis was performed as described in Chapter 2.3.9. Briefly, spot detection, quantification of spot abundance and normalisation of these abundances were performed by the DeCyder DIA module. The DeCyder BVA module then attempted to match the spot features across the gels and the Batch Processor module was used to automate the process during an overnight run. Each automated match made by DeCyder BVA was validated manually and adjusted as required across all gels before being confirmed as an accurate match. The data at this stage consisted of normalised mean spot abundances for each spot matched across the developing, germinating and internal standard gel images. This data was then interrogated to identify differential protein levels between the germinating and developing seed ER samples. For each spot feature the mean spot intensity was compared between the two sample groups against the abundance of the same spot on the standard gel, giving a value of fold change. A subset of spot features was identified with fold change values of $\pm 10\%$ and a t-test analysis was used to give

a confidence value on whether the differences in average intensities were a result of chance. Spot features that were present in each biological replicate and gave a p value of $\leq$0.02 (i.e. 98% likelihood that the observed fold change was not random) were selected for picking and analysis by mass spectrometry.

Following this differential analysis of germinating and developing seed ER, a total of 91 spots were identified as significantly elevated in the developing samples and 15 spots as significantly elevated in the germinating samples. Figure 4.5 depicts a germinating and developing sample gel showing the location of spots that are at different levels in the two sample types. A developing gel is shown (B) with spots that are elevated in the developing ER circled green and those elevated in the germinating ER circled red. Similarly for a germinating gel (A), those spots elevated in the germinating state are circled red and those elevated in the developing state are circled green.

4.3.4 Identification and excision of spots of interest on preparative gels

Spots identified as having a $\pm$10% change between the germinating and developing seed ER gels, by the criteria of a t-test analysis (p=0.02) and their presence in every gel, were deemed significant changes and picked for MS analysis. A preparative loading gel containing 400 μg of protein was run for both the germinating and developing samples. To the 400 μg of protein sample was added a 12.5 μg aliquot of the same sample but from the CyDye labelled analytical stocks. The protein samples were made up to a final volume of 350 μl with lysis buffer containing 1% DTT and 2% pH 3-10 carrier ampholytes. The 2D electrophoresis procedure was carried out under the same conditions as used for the analytical 2D DIGE study (Chapter 4.3.1) apart from the IEF strips were rehydrated with lysis buffer containing the protein sample rather than loading the sample through an anodic cup. Bind-silane was used to bond the acrylamide gel to the low-fluorescence glass plate to prevent warping and gel movement during the

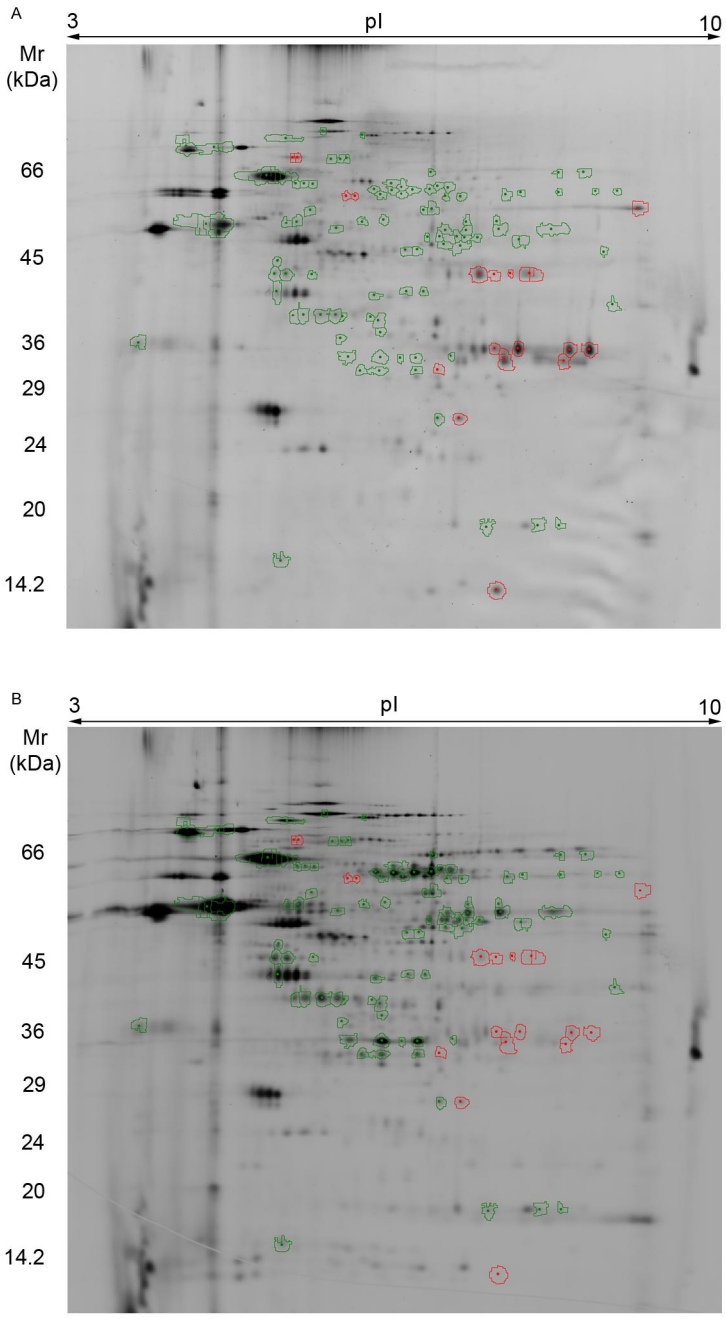


Figure 4.5: Identification of significant differences between germinating and developing ER samples. 2D DIGE analysis was performed identifying significant differences ($\geq$10% elevation, p value of $\leq$0.02) between the germinating and developing developmental states of endosperm ER. A germinating gel (A, top) is shown with 15 spots found to significantly elevated circled in red. Spots elevated in the developing state by the same criteria are depicted in green. Similarly a developing gel (B, bottom) has 91 spots found to significantly elevated in the developing state circled in green, those circled in red represent the positions of spots significantly elevated in the germinating state.

picking process. The CyDye labelled spike images were collected prior to SYPRO^{TM} staining for the preparative gel image. Also, triangulation markers were incorporated into the preparative gels prior to scanning the SYPRO^{TM} image as described previously (Chapter 3.3.3.4).

The inclusion of CyDye labelled spikes was to facilitate matching between the preparative SYPRO^{TM} stained gels and the CyDye labelled analytical set from the 2D DIGE analysis. As described in Chapter 3.3.3.1 the process of matching spots between preparative and analytical gels can be complicated by the use of in-gel hydration of protein sample into the IEF strip rather than the preferred technique of cup-loading for analytical size samples. It was found that the presence of a CyDye labelled spike of the same sample can aid in the matching of spots between analytical and preparative gels. Figure 4.6 shows the comparison between the CyDye spike image (top left, A) and the SYPRO^{TM} stained image (top right, B) of the preparative developing gel. The gel labelled C (bottom) is the original analytical gel from the 2D DIGE analysis. The CyDye labelled spike thus mediates between the analytical gel and the high-loading preparative gel. Examples of its usefulness are apparent in this figure. A large spot indicated by an arrow (1) shows a spot clearly visible in the CyDye labelled protein gels (A + C) but apparently absent from the SYPRO^{TM} gel. In an instance of a non-spiked preparative gel, you might postulate that the spot had not entered the gel. Its presence in the CyDye labelled spike indicates a likely differential staining effect between SYPRO^{TM} and CyDye. When this occurs for a spot of interest, the co-migration of the CyDye labelled spike aids its localisation for picking and downstream MS analysis.

In this experiment the preparative gels were run alongside the analytical set which means gel-to-gel variation was reduced. Where this cannot be done the presence of a CyDye labelled spike is desirable in minimising gel-to-gel variability problems.

The 91 developing and 15 germinating spots of interest from the 2D DIGE analysis were then picked with the ProPic^{TM} (Genomic Solutions Ltd. Cambridgeshire, UK)

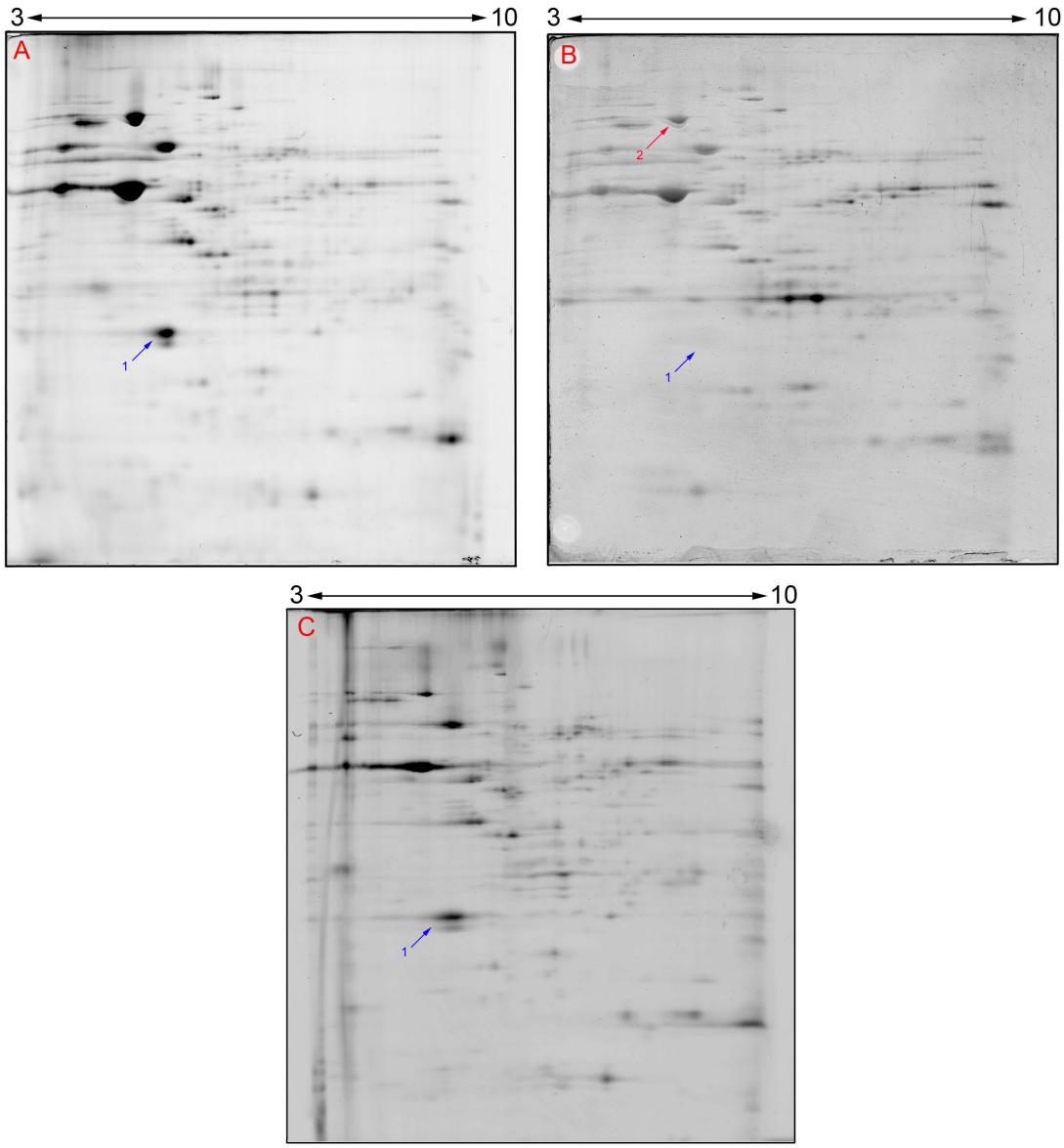


Figure 4.6: Cy spikes aid matching between analytical and preparative gels. The use of CyDye labelled sample spikes alongside a preparative gel can aid in spot matching between two separate gels, in this case the preparative gel and the analytical data gel. **(A)** shows the CyDye spike image generated from the sample incorporated into the **(B)** SYPRO^{TM} Ruby stained preparative gel. These images were obtained from a single gel so differences found cannot be ascribed to gel-to-gel variation. Arrow (1) indicates is use in identifying differential staining problems.

robot after triangulation within a Phoretix Evolution (Nonlinear Dynamics, Northumbria, UK) experiment (see Chapter 3.3.3.4, 2.3.10). Developing and germinating gel images annotated with spots of interest are shown in Figures 4.7 and 4.8. All spots were picked into a 96-well microtitre plate and trypsin digestion was performed using a Progest autodigestion robot (Genomic Solutions Ltd. Cambridgeshire, UK).

4.3.5 Protein identification by mass spectrometry

Tryptic-digested peptides were analysed by MALDI-ToF MS and MS/MS. Resultant PMF spectra and sequence data were initially used to interrogate the NCBI public database. This data is presented in Section 4.3.5.1 and published in Maltman et al. (2007). Subsequent to this publication, The Institute for Genomic Research (TIGR) released the complete genomic database for *R. communis*. The MALDI-TOF MS spectra and MS/MS sequence data were re-analysed against the complete TIGR genome database. This re-analysis is presented in Section 4.3.5.2.

4.3.5.1 Matches against the public NCBI protein database

Tryptic-digested peptides were analysed by MALDI-ToF MS and MS/MS. MALDI-ToF MS analysis generated PMF spectra which were used in a MASCOT database search (Perkins et al., 1999) of the National Centre for Biotechnology Information non-redundant (NCBI*nr*) database, interrogating all *Viridiplantae* sequences available. Matches with a Molecular Weight Search (MOWSE) score of >58 were considered statistically significant. The identity of spots of interest for developing gels is shown (Table 4.3). Figure 4.7 shows the location of each of these spots on the developing gel, annotated by a number which equates to the Spot ID number in Table 4.3. Table 4.4 lists the spots of interest from the germinating sample and Figure 4.8 is an annotated germinating gel showing the locations of these spots of interest. The exception to this for both the developing and germinating gels are 5 major spots that are labelled by

name on the gel images and the table, rather than an ID number.

For the MALDI-TOF analysis, out of the 91 developing spots 25 gave confident matches (27%) with the public database. Proteins belonging to the following families were observed: protein folding or chaperones, ricin and agglutinin, 2S albumin and legumin (Table 4.3). A PMF matching the *P. communis* (pear) H+ exporting ATPase catalytic subunit was also identified, with a MOWSE score of 99 and its protein spot exhibiting similar M_{r} and pI values to the predicted values from the database entry.

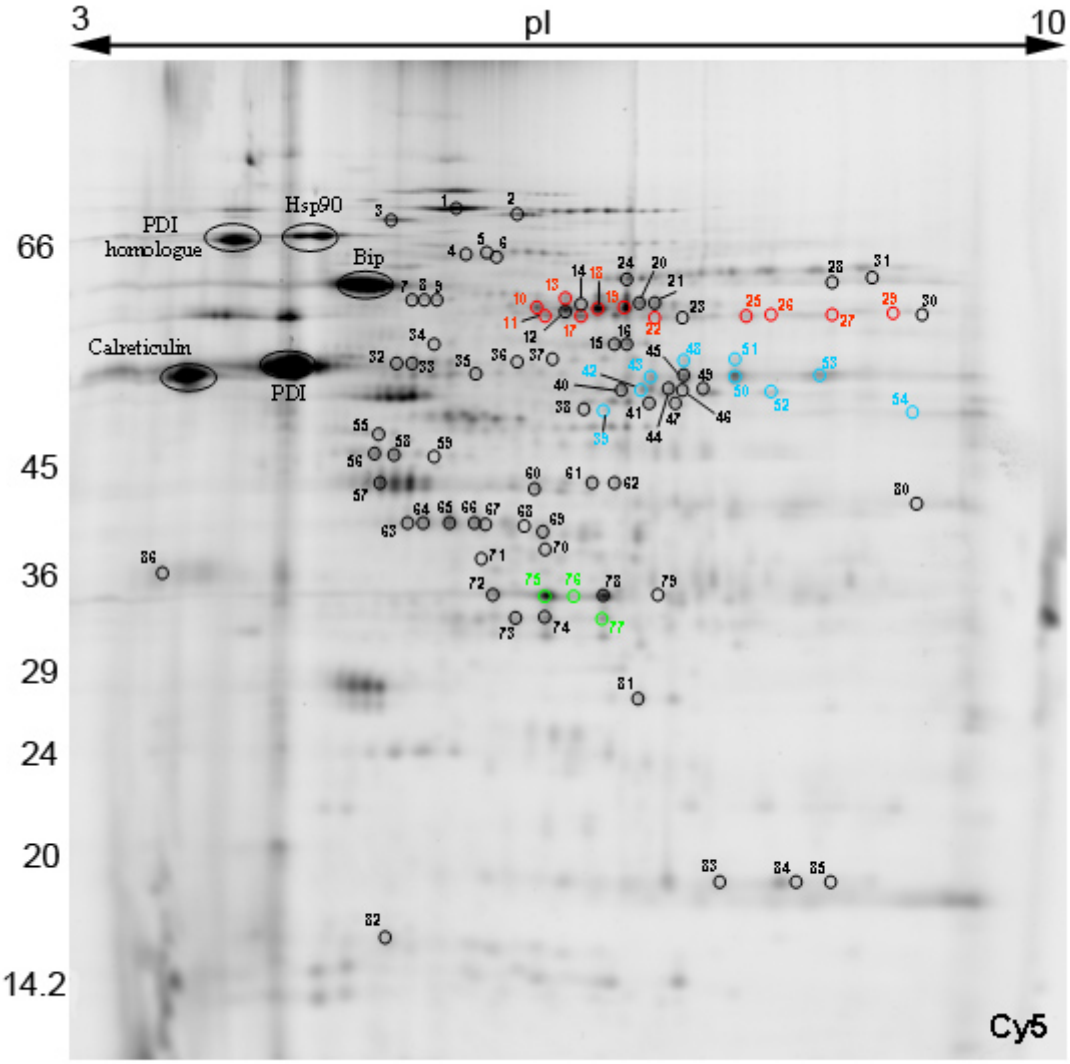


Figure 4.7: Analytical developing ER gel. The significant spots are annotated, numbers referring to those of Table 4.3. M_r and pI information is indicated. Spots identified in the NCBI*nr* database as seed storage proteins are coloured: ricin and RCA, red; legumin, blue; 2S albumin, green. Previously identified chaperones are annotated by name (Maltman et al., 2002).

Table 4.3: 2D DIGE and MS-based identification of spots significantly elevated (p=0.02, $\triangle$ change > 10%) in 2D gels of ER from developing endosperm. NI = not identified by MS

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold $\triangle$	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
Protein folding and chaperones														
BiP	gi|19813	Luminal binding protein (BiP) [*Nicotiana tabacum*]	Y		2.5	0.021	73870	5.08	65757	5.17	202	31	48%	
	gi|729623	Luminal-binding protein 5 precursor [*Nicotiana tabacum*]		Y			73744	5.08			144	10	17%	DYFDGKEPNK
														FEELNNDLFR
														EAEEFAEEDKK
														NQIDEIVLVGGSTR
														DAVVTVPAYFNDAQR
														EVEAVCNPIITAVYQR
														IKDAVVTVPAYFNDAQR
														VEIESLFDGVDFSEPLTR
														LKEVEAVCNPIITAVYQR
														GVNPDEAVAYGAAVQGGIL
Hsp90	gi|23477636	Grp94 (*Xerophyta viscosa*) Hsp90 homologue	Y		3.34	0.031	94148	4.92	72790	4.9	116	26	32%	
PDI	gi|11133775	Protein disulphide isomerase precursor (PDI) [*Ricinus communis*]	Y		7.55	0.001	55811	4.95	55446	4.64	343	35	71%	
PDI Homologue	gi|4678297	Protein disulphide-isomerase homologue protein [*Arabidopsis thaliana*]		Y	2.96	0.013	62735	4.72	72089	4.36	97	3	8	VIVGNNFDEIVLDESK
														AMGFPTLLFFPELK
														FSDPLTVDTDR

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
Calreticulin	gi|11131745	Calreticulin precursor [*Ricinus communis*]	Y		1.34	<0.02	47721	4.38	54449	4.09	182	23	53%	
32	gi|1134968	Protein disulphide isomerase [*Ricinus communis*]	Y		4.42	0.0079	55560	4.95	55580	5.31	174	18	35%	
33	gi|1134968	Protein disulphide isomerase [*Ricinus communis*]	Y		3.17	0.0078	55560	4.95	55850	5.41	91	14	30%	
				Y			55560	4.95			234	10	22%	FFNSPDAK GYPTVYFR SDADIVIAK QSGPASVEIK IFIVGVFPK SHDIPVVLAK SDYEFGHTLDAK SDYEFGHTLDAKK LDATANDIPSDTFDVR VDANEEANKELATQYDIK
1	gi|50912043	Putative growth regulator [*Oryza sativa*] (Hsp70 domains)		Y	5	0.0094	99370	5.29	77048	5.66	105	3	2%	SGIISLDR LQEFLGR ANLHFSLSR
61	gi|21593230	Unknown [*Arabidopsis thaliana*] (DnaJ domains)		Y	2.94	3.30E-05	39199	5.93	43689	6.46	110	2	4%	YGEEGLK SYYDVLQVPK
	gi|34897648	Hypothetical protein [*Oryza sativa*] (DnaJ domains)	Y				39276	6.43			110	2	4%	
Ricin and agglutinin														
10	gi|251808	Agglutinin I, proRCA I [*Ricinus communis*]	Y		15.56	0.00083	60168	7.6	61751	6.08	74	9	20%	
11	gi|2914589	Ricin A-chain (recombinant) [*Ricinus communis*]	Y		3.21	0.0013	29430	6.91	61559	6.21	102	9	46%	

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
13	gi|251808	Agglutinin I, proRCA I [*Ricinus communis*]	Y		6.07	1.20E-05	60168	7.6	63953	6.31	74	11	24%	
17	gi|21085	Pre-propolypeptide [*Ricinus communis*]	Y		25.85	1.70E-08	64090	6.34	61674	6.41	142	19	41%	
18	gi|251808	Agglutinin I, proRCA I [*Ricinus communis*]	Y		53.77	3.90E-06	60168	7.6	62575	6.52	77	12	30%	
19	gi|251808	Agglutinin I, proRCA [*Ricinus communis*] I	Y		44.95	6.70E-05	60168	7.6	63953	6.68	84	15	33%	
	gi|113504	Agglutinin precursor [*Ricinus communis*] (RCA)		Y			62851	6.66			382	8	12%	CLTISK
	gi|251808	Agglutinin precursor [*Ricinus communis*]		Y			60168	7.6			382	8	13%	+ Carbamidomethyl(C)
														VGLPISQR
														LEQLGGLR
														HEIPVLPNR
														FQYIEGEMR
														ILSCGPASSGQR
														+ Carbamidomethyl(C)
														DNCLTTDANIK
														+ Carbamidomethyl(C)
													(38)	VWLEDCTSEK
														+ Carbamidomethyl(C)
22	gi|113504	Agglutinin precursor [*Ricinus communis*] (RCA)		Y	8.23	0.00064	62851	6.66	61228	6.83	173	4	8%	LEQLGGLR
	gi|251808	Agglutinin precursor [*Ricinus communis*]					60168	7.6			173	4	8%	FQYIEGEMR
														ILSCGPASSGQR
														+ Carbamidomethyl(C)
														DNCLTTDANIK
														+ Carbamidomethyl(C)

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
25	gi|21085	Pre-propolypeptide [*Ricinus communis*]	Y		9.94	7.70E-05	64090	6.34	61525	7.4	123	14	32%	
26	gi|21085	Pre-propolypeptide [*Ricinus communis*]	Y		8.13	0.0015	64090	6.34	61973	7.55	117	14	32%	
27	gi|21085	Pre-propolypeptide [*Ricinus communis*]	Y		10.01	0.00087	64090	6.34	62048	7.91	126	17	38%	
29	gi|113504	Agglutinin precursor (RCA) [*Ricinus communis*]		Y	12	0.0033	62851	6.66	62273	8.26	79	3	7%	LEQLGGLR ILSCGPASSGQR
	gi|251808	Agglutinin I; proRCA I [*Ricinus communis*]					60168	7.6			79	3	7%	+ Carbamidomethyl(C) LSTAIQESNQGAFASPIQLQR
2S Albumin														
75	gi|21068	2S Albumin precursor [*Ricinus communis*]	Y		61.55	2.10E-05	29289	6.73	34593	6.21	240	20	70%	
76	gi|21068	2S Albumin precursor [*Ricinus communis*]	Y		18.23	9.10E-07	29289	6.73	34509	6.4	119	11	44%	
77	gi|21068	2S Albumin precursor [*Ricinus communis*]		Y	5.38	0.00021	29289	6.73	33039	6.55	95	3	13%	AGEIVSSCGVR + Carbamidomethyl(C) GQIQEQQNLR TTITTIEIDESKGER
Legumin														
39	gi|7269954	Putative protein [*A thaliana*]		Y	8.49	0.0063	28385	5.46	50698	6.56	58	1	3%	EILFEIVDR
	gi|7269956	Putative protein [*A thaliana*]					28224	5.35			58	1	3%	
	gi|21555885	Unknown [*A thaliana*]					28125	5.35			58	1	3%	
	gi|21593315	Defense-related protein [*A thaliana*]					28427	5.46			58	1	3%	
	gi|23505821	At4g30550/F17I23_110 [*A thaliana*]					28300	5.35			58	1	3%	

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
	gi|600108	Legumin A precursor [*Vicia narbonensis*]		Y			54667	7			49	1	1%	LDALEPDNR
	gi|8118510	Legumin-like [*Ricinus communis*]		Y			53668	8.65			46	2	3%	ENIADPSR SDVFVPEVGR
42	gi|8118510	Legumin-like [*Ricinus communis*]		Y	4.45	0.012	53668	8.65	53147	6.78	68	2	1%	VSIEEAR VSIEEARR
43	gi|8118512	Seed storage protein [*Ricinus communis*]	-	Y	6.5	0.00048	40099	9.29	54317	6.83	352	10	36%	INQLAGR YLQLSIQK DFFLAGNPQR QEVTLLSPGSR ASNEGLEWVSFK VTSVNSHNLPILR VIAESFNINTDLAR GIIVSVEHDLEMLAPQR VQIVNENGDSVFDGQVQR SMPEEVVANAFQVSVEDAR
48	gi|8118512	Seed storage protein [*Ricinus communis*]	Y		5.11	0.00033	40099	9.29	55917	7.03	127	14	42%	
	gi|8118512	Seed storage protein [*Ricinus communis*]		Y			40099	9.29			201	6	18%	INQLAGR YLQLSIQK DFFLAGNPQR QEVTLLSPGSR VTSVNSHNLPILR
													(42)	VQIVNENGDSVFDGQVQR
50	gi|8118512	Seed storage protein [*Ricinus communis*]	Y		22.22	1.40E-05	40099	9.29	54383	7.34	149	19	69%	

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
	gi|8118512	Seed storage protein [*Ricinus communis*]		Y			40099	9.29			452	12	35%	INQLAGR
														YLQLSIQK
														LRGENDLR
														EVQSQRGER
														DFFLAGNPQR
														QEVTLLSPGSR
50														VTSVNSHNLPILR
														DNRQEVTLLSPGSR
														HNINKPSEADIYNPR
														VQIVNENGDSVFDGQVQR
														STSTGSAHDNSGNVFSGMDER
													(74)	RSTSTGSAHDNSGNVFSGMDER
51	gi|8118512	Seed storage protein [*Ricinus communis*]	Y		7.02	0.0001	40099	9.29	56053	7.32	211	23	77%	
52	gi|8118510	Legumin-like protein [*Ricinus communis*]		Y	6.28	0.0016	53668	8.65	52445	7.53	99	3	3%	ENIADPSR
														FEYVAFK
														MKENIADPSR
53	gi|8118512	Seed storage protein [*Ricinus communis*]	Y		15.54	0.00038	40099	9.29	54383	7.84	227	26	81%	
54	gi|8118510	Legumin-like protein [*Ricinus communis*]	Y		5.62	0.00025	53668	8.65	50271	8.34	139	15	31%	
Other Proteins														
8	gi|60592632	H+ exporting ATPase catalytic subunit [*Pyrus communis*]	Y		6.07	0.0035	68757	5.43	63953	5.49	99	13	23%	
86	gi|4960154	Putative progesterone-binding protein [*Arabidopsis thaliana*]	Y		1.67	0.0011	28228	8.58	36440	3.9	46	2	9%	
	gi|7576227	Putative progesterone-binding protein homologue Atmp2 [*Arabidopsis thaliana*]		Y			25382	4.56			46	2	10%	GQIYDVSQSR

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
84	gi|28555917	NBS-LRR disease resistance protein [*Hordeum vulgare*]		Y	2.92	0.0012	160871	6.56	19100	7.71	54	2	1%	MFYGPGGPYALFAGK GLELVANK
85	gi|28555917	NBS-LRR disease resistance protein [*Hordeum vulgare*]		Y	7.29	0.0012	160871	6.56	19100	7.92	49	2	1%	LSSLKELR GLELVANK
3	gi|10177672	Glucosidase II alpha subunit [*Arabidopsis thaliana*]		Y	9.52	7.00E-05	104274	5.86	75116	5.31	73	2	1%	LSSLKELR FVFSK
	gi|34906342	Unnamed protein product [*Oryza sativa*]		Y										AGTIIPR
2	gi|34906342	Unnamed protein product [*Oryza sativa*]		Y	3.27	0.062	61904	7.16	76030	6.04	53	2	2%	LATGEPLR
23	gi|51535322	hypothetical protein [*Oryza sativa*]		Y	13.26	8.00E-05	20855	11.7	61154	6.97	62	2	6%	LATGEPLRLVSK LGPLVPR
73	gi|54660800	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit [*Cremastosperma sp. Pirie 71*]		Y	9.19	0.00011	50518	6.46	32999	6.02	49	2	2%	EIAMRR QGWEIIR DLARQGTEIIR
Unidentified Proteins														
4		NI			3.4	0.0031			69773	5.74				
5		NI			2.4	0.017			69857	5.83				
6		NI			2.27	0.018			69942	5.89				
7		NI			5.91	0.0019			63953	5.4				
9		NI			3.56	0.0059			63876	5.57				
12		NI			28.43	3.30E-06			62500	6.32				

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
14		NI		11.52	0.00025			63260	6.41				
15		NI		5.21	3.60E-05			58195	6.62				
16		NI		4.29	0.00046			58548	6.7				
20		NI		39.44	4.20E-05			63337	6.76				
21		NI		11.64	0.00015			63108	6.86				
24		NI		10.16	6.30E-05			66638	6.69				
28		NI		4.9	0.0061			66477	7.91				
30		NI		15.44	0.0065			62273	8.47				
31		NI		4.53	0.0082			66719	8.15				
34		NI		1.63	0.012			58195	5.54				
35		NI		2.22	0.00059			54647	5.79				
36		NI		3.48	0.00041			55850	6.04				
37		NI		4.05	0.0023			56257	6.24				
38		NI		3.99	0.015			50576	6.45				
40		NI		5.74	0.0065			53019	6.66				
41		NI		3.39	0.0069			51564	6.81				
44		NI		3.73	0.00052			53147	6.94				
45		NI		15.27	0.00017			54252	7.02				
46		NI		5.14	0.00056			52891	7.01				
47		NI		4.39	0.013			51377	6.97				
49		NI		6.4	0.0009			52827	7.13				
55		NI		2.81	0.0041			48656	5.23				
56		NI		3.4	0.00032			46526	5.2				
57		NI		2.9	0.02			43637	5.22				
58		NI		1.9	0.013			46470	5.31				
59		NI		1.96	0.003			46413	5.56				
60		NI		3.37	0.0026			43112	6.15				

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold △	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
62		NI			2.61	0.00026			43637	6.62				
63		NI			7.75	0.007			40142	5.38				
64		NI			6.27	0.0033			39983	5.48				
65		NI			4.11	0.00073			40239	5.64				
66		NI			2.56	0.0029			40191	5.78				
67		NI			3.02	0.0027			39997	5.85				
68		NI			2.53	0.0006			39900	6.12				
69		NI			3.4	0.00013			39468	6.19				
70		NI			3.04	0.012			37786	6.21				
71		NI			3.86	0.0072			37017	5.83				
72		NI			15.28	0.00097			34718	5.9				
74		NI			13.54	0.00059			32999	6.2				
78		NI			51.83	1.10E-05			34551	6.55				
79		NI			4.6	0.0067			34760	6.88				
80		NI			2.68	0.0031			41727	8.42				
81		NI			1.93	0.0073			27925	6.76				
82		NI			1.86	0.0076			16722	5.31				
83		NI			9.63	0.00061			18985	7.22				

From the germinating gel only a single protein was identified by MALDI TOF MS PMF, this being spot no. 15: malate synthase, matching a *R. communis* database entry (Table 4.4).

27% of spots giving data for the developing set and 7% of spots for the germinating set are low if we compare to 2D DIGE investigations by other groups (Rowland et al., 2007; Chivasa et al., 2006). This was due in part to the lack of a publicly available complete genomic database for *R. communis* at the time of the analysis. The entries that were in the *Viridiplantae* NCBI*nr* database were limited for this organism (320 *R. communis* NCBI*nr* database entries at the time of the analysis) so the potential for PMF-based identification is restricted. To obtain extra information from this proteomic study a further MS analysis was employed. LC-MS/MS analysis of the trypsin digested peptides was performed with an Applied Biosystems QSTAR PULSARi^{TM}. This instrument can select ions for fragmentation and peptide sequencing giving actual amino acid sequences for peptides successfully analysed. Having amino acid sequence data for a peptide is a major benefit over PMF-based MS and helps overcome the limitation of not having a complete genome database for *R. communis*. It can allow the putative identification of proteins based on their homology to other sequenced organisms (for example, the model plant *A. thaliana* whose genome database is available within the NCBI*nr* database). Having partial sequence data for a protein can also aid in the cloning of its gene.

The MS/MS analysis resulted in 21 individual spots giving identities after the analysis was submitted to the NCBI*nr* *Viridiplantae* database (matches with a MOWSE score of >43 were significant). Out of these 21 spots, 6 had given confident matches to database entries when their peptides were analysed by MALDI TOF MS (see Table 4.3). In all of these cases the identity obtained was the same. 15 spots which failed to give an identity by MALDI TOF MS gave positive hits after MS/MS sequencing. These included homologous peptide matches to protein families already observed in

the MALDI TOF study but from other organisms e.g. the chaperone-family protein disulphide isomerase homologue from *A. thaliana* (gi|4678297), or spot 39: the legumin-family homologue from *A. thaliana* (gi|7269954). It also included matches with proteins already observed by MALDI-TOF but from another near-by spot e.g. spot 22 : Agglutinin precursor RCA from *R. communis* (gi|113504) or spot 52 : the legumin-like protein from *R. communis* (gi|8118510). This observation reveals different sensitivities of the two MS technologies.

A number of proteins that don't fall into the category of storage proteins, chaperones or folding proteins gave hits by homology but had MOWSE scores close to the statistical cut off. These are listed under 'Other Proteins' in Table 4.4 and discussed in Chapter 4.4.4. They include matches to NBS-LRR disease resistance protein from *H. vulgare* (spot 84, 85), glucosidase II alpha subunit from *A. thaliana* (spot 3) and ribulose-1,5-bisphosphate carboxylase/oxygenase subunit fom *Cremastosperma sp. Piriie* 71 (spot 73).

In some cases more than one database entry gave a significant match to the MS/MS sequence data. For example, spot no. 19 gave a confident hit for Agglutinin I, proRCA - gi|251808 (*R. communis*) but also its precursor protein Agglutinin precursor (RCA) - gi|113504 (*R. communis*). This is to be expected where a protein is present in unprocessed and processed forms in the *Viridiplantae* database as the amino acid sequences will be the same for part of the protein.

The MS/MS analysis gave hits for a further 4 spots from the germinating gel (see Table 4.4). Spot 12 gave the highest-confidence MOWSE score (116) with a match to porin from *P. sativum* (gi|396819). No identification was made for the MALDI TOF MS identified malate synthase when analysed by MS/MS.

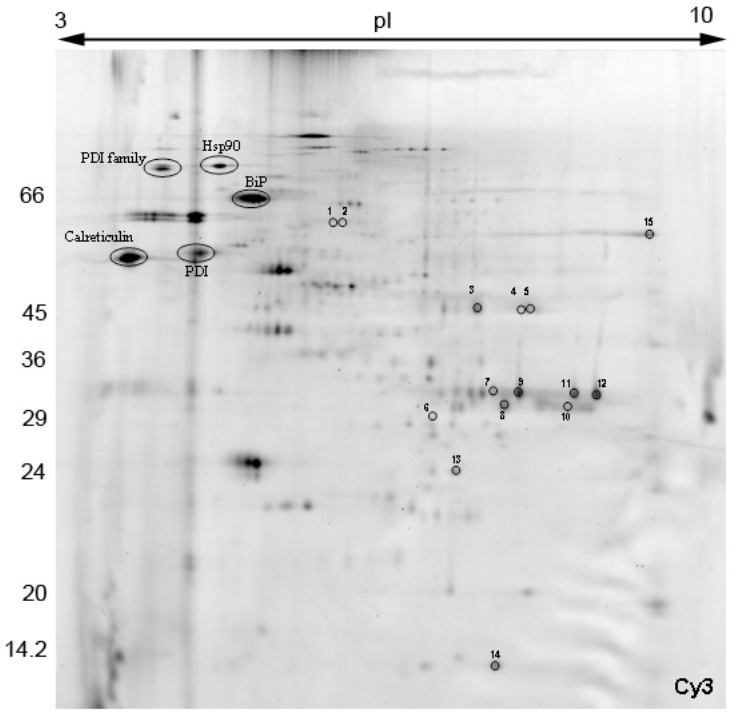


Figure 4.8: Analytical germinating ER gel. Significant spots are annotated, numbers referring to those of Table 4.4. M_{r} and pI information is indicated. Previously identified chaperones are annotated by name (Maltman et al., 2002).

Table 4.4: 2D DIGE and MS-based identification of spots significantly elevated (p=0.02, $\triangle$ change > 10%) in 2D gels of ER from germinating endosperm. NI = not identified by MS.

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	Fold $\triangle$	t-Test p value	M_r pred	pI pred	M_r gel	pI gel	MOWSE score	No. different matching peptides	Sequence coverage (% combined)	Matching MS/MS peptides
3	gi|61104883	Strictosidine synthase family protein [*Arabidopsis thaliana*]		Y	2.4	0.011	46653	6.55	46526	7.15	74	1	2%	GPYTGLADGR
4	gi|61104883	Strictosidine synthase family protein [*Arabidopsis thaliana*]		Y	2.64	0.015	46653	6.55	46639	7.45	50	1	2%	GPYTGLADGR
5	gi|37779748	NADH dehydrogenase subunit F [*Halodule wrightii*]		Y	3.61	0.0068	76908	8.1	47026	7.61	49	1	1%	XNNQLLKR
12	gi|396819	Porin [*Pisum sativum*]		Y	5.15	0.0023	29596	9.11	36325	8.24	116	4	8%	DLLYK NITTDIK GPGLYTDIGK GPGLYTDIGKK
15	gi|21076	Unnamed protein product [*Ricinus communis*] malate synthase	Y		5.19	0.0022	64262	8.53	58690	8.67	181	30	50%	
1		NI			2.27	0.019			61302	5.88				
2		NI			1.95	0.015			61302	5.97				
6		NI			1.54	0.011			33159	6.76				
7		NI			2.19	0.009			35698	7.3				
8		NI			2.28	0.0068			34426	7.39				
9		NI			3.67	0.004			35698	7.52				
10		NI			3.37	0.008			34218	7.96				
11		NI			2.71	0.0069			35569	8.01				
13		NI			1.82	0.0097			27925	6.96				
14		NI			7.35	0.00028			15197	7.31				

4.3.5.2 Re-analysis of MS data against a complete *R. communis* protein database

The availability of a complete *R. communis* genomic database, published by The Institute for Genomic Research, gave an opportunity for the MALDI-TOF and MS/MS data to be re-analysed to see if any new identities could be assigned to the dataset. The 86 spots with a number ID on the developing gel were re-analysed. For the MALDI-TOF data, 54 developing spots gave confident hits to the TIGR database (59%). This is a significant increase on the 27% of developing spots giving hits against the NCBI database in the original analysis. The MS/MS data was also analysed, and this provided a further 2 confident hits against the database. This data is presented in Table 4.5. Spots which previously gave identities have their Spot ID numbers in bold. Those spots that are suffixed with an asterisk have given a different identify in the TIGR analysis than in the original NCBI analysis.

The same broad protein families that were identified in the original analysis were identified in the TIGR re-analysis: protein folding and chaperones, ricin and agglutinin, 2S albumin and legumin. 16 proteins identified did not belong to these groups and are listed under 'Other Proteins'. 6 of these proteins previously gave matches to the NCBI database, however 3 of the TIGR hits are different types of protein to those published previously. For each of these 3 different proteins, they previously had low MOWSE scores (53, 62, 49). They are described in Section 4.4.4.

The MS data generated from the germinating spots was also re-analysed. A total of 10 spots gave positive hits (67%), again a significant increase on the 33% of total hits obtained in the original analysis, see Table 4.6. Spots ID numbers are in bold and asterisked as for the developing table. 5 spots identified in the TIGR analysis previously gave confident matches to the NCBI database; one of these spots has subsequently matched a different type of protein in the TIGR analysis (Spot 5, putative strictosidine synthase).

Table 4.5: Re-analysis of developing MS data against the TIGR *R. communis* database. Bold ID Numbers = Previously matched an entry in the NCBI database. Asterisk = Different identity obtained from TIGR analysis than identified in NCBI analysis

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	MOWSE score	No. different matching peptides	Matching MS/MS peptides
Protein folding and chaperones							
1	29724.m000869	Heat shock 70 kDa protein (putative)	Y		221	32	
32	29908.m006121	Protein disulfide isomerase (putative)	Y		174	19	
33	29908.m006121	Protein disulfide isomerase (putative)	Y		143	16	
57	30128.m008894	Protein disulfide isomerase (putative)	Y		58	9	
61	30076.m004681	Chaperone protein dnaJ (putative)		Y	79	1	FAEINNAYEVLSDSEKR
64	30128.m008894	Protein disulfide isomerase (putative)	Y		54	7	
65	30128.m008894	Protein disulfide isomerase (putative)	Y		80	9	
66	30128.m008894	Protein disulfide isomerase (putative)	Y		62	7	
68	30128.m008894	Protein disulfide isomerase (putative)	Y		106	12	
69	30128.m008894	Protein disulfide isomerase (putative)	Y		56	7	
71	30128.m008894	Protein disulfide isomerase (putative)	Y		95	10	
Ricin and agglutinin							
10	28274.m00033	Agglutinin precursor (RCA) [contains: Agglutinin A chain]	Y		81	11	
11	59679.m000011	Ricin precursor [Contains: Ricin A chain] (putative)	Y		80	9	
13	28274.m00033	Agglutinin precursor (RCA) [contains: Agglutinin A chain]	Y		101	14	
17	59679.m000011	Ricin precursor [Contains: Ricin A chain] (putative)	Y		118	16	
18	28274.m00033	Agglutinin precursor (RCA) [contains: Agglutinin A chain]	Y		92	14	
19	28274.m00033	Agglutinin precursor (RCA) [contains: Agglutinin A chain]	Y		126	20	
	28274.m00033	Agglutinin precursor (RCA) [contains: Agglutinin A chain]		Y	462	11	CLTISK + Carbamidomethyl (C) VGLPISQR LEQLGGLR SHLTTGGDVR

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	MOWSE score	No. different matching peptides	Matching MS/MS peptides
							FQYIEGEMR
							ILSCGPASSGQR
							+ Carbamidomethyl (C)
							DNCLTTDANIK
							+ Carbamidomethyl (C)
							VWLEDCTSEK
							+ Carbamidomethyl (C)
							TIINPTSGLVLAATSGNSGTK
							AEQQWALYADGSIRPQQNR
							LSTAIQESNQGAFASPIQLQR
22	29844.m003250	Vicilin GC72-A precursor (putative)	Y		189	25	
	29844.m003250	Vicilin GC72-A precursor (putative)		Y	444	24	GIENYR
							NFLAGQR
							GQAPLNLR
							CTISYVLR
							+ Carbamidomethyl (C)
							FSESSELLR
							NQQPLYSNR
							QKGQAPLNLR
							CEEEPIKER
							+ Carbamidomethyl (C)
							GSLMVPHYNSR
							NNPYYFHAQR
							VSYNLETGDVIK
							VEMACPHVASQK
							+ Carbamidomethyl (C)
							ESHFVAGPQQGQR
							ELSFNVPAELIEK
25	59679.m000011	Ricin precursor [Contains: Ricin A chain] (putative)	Y		119	13	
26	59679.m000011	Ricin precursor [Contains: Ricin A chain] (putative)	Y		102	12	
27	59679.m000011	Ricin precursor [Contains: Ricin A chain] (putative)	Y		126	16	

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	MS/MS	MOWSE score	No. different matching peptides	Matching MS/MS peptides
2S Albumin							
72	28166.m001081	Sweet protein mabinlin-1 Chain A	Y		30	3	
73*	28166.m001079	2S albumin precursor [contains 2S albumin small chain; 2S albumin large chain]	Y		122	16	
74	28166.m001079	2S albumin precursor [contains 2S albumin small chain; 2S albumin large chain]	Y		144	16	
75	28166.m001081	Sweet protein mabinlin-1 Chain A	Y		232	20	
76	28166.m001081	Sweet protein mabinlin-1 Chain A	Y		125	11	
77	28166.m001081	Sweet protein mabinlin-1 Chain A	Y		77	9	
78	28166.m001081	Sweet protein mabinlin-1 Chain A	Y		240	18	
79	28166.m001081	Sweet protein mabinlin-1 Chain A	Y		191	22	
82	28166.m001083	2S sulphur-rich seed storage protein precursor (Allergen Ber e 1)	Y		50	5	
Legumin							
39	29611.m000223	Legumin A precursor (Beta-globulin) (LEGA-C94) [Contains: Legumin A acidic chain; Legumin A basic chain]	Y		177	23	
	29611.m000223	Legumin A precursor (Beta-globulin) (LEGA-C94) [Contains: Legumin A acidic chain; Legumin A basic chain]		Y	418	11	VFDGNVK QETILTR LDALEPDNR NLFCGIDTR + Carbamidomethyl (C) VSTVNSNNLR ADVYVPEVGR IKENIADPSR IQVVDENGNR NFHLAGNPENEFQK

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	MOWSE score	No. different matching peptides	Matching MS/MS peptides
							EGQVLTVPQNFVVVK
							LLQLSASHVSLSNGAIR
41	29611.m000223	Legumin A precursor (Beta-globulin) (LEGA-C94) [Contains: Legumin A acidic chain; Legumin A basic chain]	Y		191	21	
42	30005.m001288	Legumin A precursor (Beta-globulin) (LEGA-C94) [Contains: Legumin A acidic chain; Legumin A basic chain]	Y		285	27	
43	30005.m001290	Legumin A precursor (Beta-globulin) (LEGA-C94) [Contains: Legumin A acidic chain; Legumin A basic chain]	Y		245	24	
48	29200.m000169	Glutelin type-A 3 precursor	Y		216	21	
50	29200.m000169	Glutelin type-A 3 precursor	Y		205	26	
	29200.m000169	Glutelin type-A 3 precursor		Y	337	7	INQLAGR
							YLQLSIQK
							GQQDQCQLNR
							+ Carbamidomethyl (C)
							GLLLPQYVNGPK
							VTSVNSHNLPILR
							ARFNGLEETFCTAR
							+ Carbamidomethyl (C)
							STSTGSAHDNSGNVFSGMDER
51	29200.m000169	Glutelin type-A 3 precursor	Y		297	31	
52	30005.m001290	Legumin A precursor (Beta-globulin) (LEGA-C94) [Contains: Legumin A acidic chain; Legumin A basic chain]	Y		239	27	
	30005.m001290	Legumin A precursor (Beta-globulin) (LEGA-C94) [Contains: Legumin A acidic chain; Legumin A basic chain]		Y	291	8	QESTFGR
							YSLSGDSER
							LNAFEPDNR
							ADLFVPEVGR
							MKENIADPSR
							AESDRFEYVAFK
							MSTVNSHNLPILR
							LIAEAFNINEQLAR

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	Identification MS/MS	MOWSE score	No. different matching peptides	Matching MS/MS peptides
53	29200.m000169	Glutelin type-A 3 precursor	Y		313	34	
	29200.m000169	Glutelin type-A 3 precursor		Y	403		INQLAGR
							YLQLSIQK
							QEVTLLSPGSR
							GQQDQCQLNR
							+ Carbamidomethyl (C)
							GLLLPQYVNGPK
							VTSVNSHNLPILR
							VQIVNENGDSVFDGQVQR
							STSTGSAHDNSGNVFSGMDER
54	29600.m000565	Legumin type B precursor	Y		141	15	
Other proteins							
2*	30221.m002257	strictosidine synthase (putative)	Y		159	20	
3	29687.m000597	neutral alpha-glucosidase ab precursor (putative)	Y		205	32	
4	28333.m000566	oligopeptidase A (putative)	Y		57	10	
8	30169.m006261	ATP synthase alpha subunit vacuolar (putative)	Y		172	18	
15	29864.m001501	9-cis-epoxycarotenoid dioxygenase (putative)	Y		177	21	
16	29864.m001501	9-cis-epoxycarotenoid dioxygenase (putative)	Y		80	10	
20	29703.m001491	nucleolar protein nop56 (putative)	Y		154	18	
23*	29703.m001491	nucleolar protein nop56 (putative)	Y		319	35	
24	29667.m000347	Cycloartenol synthase (putative)	Y		87	13	
31	29751.m001894	sorting and assembly machinery (sam50) protein (putative)	Y		210	24	
44	30170.m013972	cytochrome P450 (putative)	Y		73	8	
55	30128.m008712	subtilisin inhibitor-like 1	Y		118	14	
83	29994.m000460	conserved hypothetical protein	Y		44	14	
84*	29994.m000460	conserved hypothetical protein	Y		137	14	
85*	29994.m000460	conserved hypothetical protein		Y	138	3	VGDQSLGTK
							SISIGLLSR
							EAEVQPLLTGGASNAEFLK

Continued on Next Page...

Spot no.	Accession no.	Protein ID	Identification MALDI	MS/MS	MOWSE score	No. different matching peptides	Matching MS/MS peptides
86	30147.m013761	steroid binding protein (putative)	Y		55	9	

Table 4.6: Re-analysis of germinating MALDI data against the TIGR *R. communis* database. Bold ID Numbers = Previously matched an entry in the NCBI database. Asterisk = Different identity obtained from TIGR analysis than identified in NCBI analysis

Spot no.	Accession no.	Protein ID	Identification MALDI	MS/MS	MOWSE score	No. different matching peptides	Matching MS/MS peptides
1	30186.m001314	Protein SEY1 (Putative)	Y		96	19	
2	30186.m001314	Protein SEY1 (Putative)	Y		85	20	
3	30221.m002257	Strictosidine synthase (Putative)	Y		217	17	
4	30221.m002257	Strictosidine synthase (Putative)	Y		181	22	
5*	30221.m002257	Strictosidine synthase (putative)	Y		177	20	
10	29822.m003473	Voltage-dependent anion-selective channel (Putative)	Y		180	14	
12	29736.m002034	Voltage-dependent anion-selective channel (Putative)	Y		136	13	
13	29929.m004802	Cytochrome P450 (Putative)	Y		85	7	
14	30190.m010843	Nucleoside diphosphate kinase (Putative)	Y		60	6	
15	30147.m013773	maltate synthase (Putative)	Y		201	27	

4.3.6 Accurate assignment of M_r and pI values

Quantification of the molecular weight and pI of proteins on a 2D gel provides useful information to aid in the identification of proteins. It can give confidence to an MS identification when the predicted M_r and pI values match closely to those observed on a 2D gel. Deviations from the predicted values might indicate the presence of post-translational modifications. They are also commonly used in the comparison of spots between gels, allowing the localisation of the same area on different gels.

For the accurate identification of M_r and pI values 20 μg of commercially available M_r + pI markers (product code: M3411, Sigma-Aldrich) were labelled with Cy3 dye and separated by 2DE alongside a Cy2 standard (see Chapter 2.3.5). The purpose of labelling and running the markers within the same gel as the Cy2 standard is to increase the accuracy of M_r and pI assignations by utilising the co-migration benefits that CyDye labelled protein allows. The Cy2 (internal standard) sample was used because it contains both germinating and developing spots, allowing assignations of M_r and pI values to both samples types. After imaging, spot maps were imported into DeCyder DIA and M_r / pI values assigned to the visualised spots on the marker gel. The software was then able to assign accurate pI and M_r values to all spot features in the Cy2 gel by calibration based on the migration of the markers, using a linear regression algorithm for pI and log-linear regression for the M_r values.

Figure 4.9 depicts an overlay image of the Cy3 labelled M_r / pI standards and the Cy2 labelled *R. communis* seed ER sample. The protein spots of the Cy3 labelled M_r / pI standards (annotated 1-5) are clearly visible in red contrasting against the green-coloured Cy2 *R. communis* internal standard spot map. M_r / pI standards are amyloglucosidase (89 kDa and 70 kDa, 1 and 2 respectively), ovalbumin (45 kDa, 3), carbonic anhydrase (29 kDa, 4), horse heart myoglobin (17 kDa, 5).

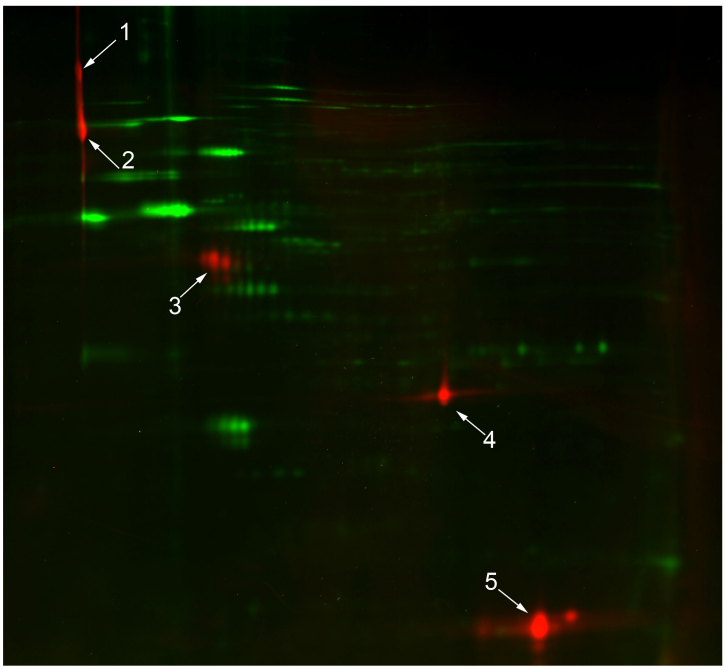


Figure 4.9: Multiplex M_r / pI and Cy2 *R. communis* gel. Protein spots belong to the commercial M_r / pI standard (red) have co-migrated with the Cy2 ER sample (green), allowing highly accurate calibration of M_r / pI values. M_r / pI standards are amyloglucosidase (89 kDa and 70 kDa, 1 and 2 respectively), ovalbumin (45 kDa, 3), carbonic anhydrase (29 kDa, 4), horse heart myoglobin (17 kDa, 5).

4.4 Discussion

4.4.1 Components of lipid biosynthesis were not identified in the urea-soluble fraction of ER

One of the major aims of this study was to obtain identities of developing ER proteins, significantly elevated in this developmental state, by MS. This is with a view to locating components of lipid biosynthesis pathways in the urea soluble fraction. Based on the hypothesis that lipid biosynthesis machinery is elevated in the developing seed ER compared to the germinating seed ER, the germinating proteome was used as a screen to reduce the number of targets for picking. Furthermore, with the advent of the 2D DIGE system the differential screen was supported by multiple replicate statistical criteria, giving a figure of confidence that the elevation in protein amount between the two states is real. 91 separate spots present across the developing set of gels were found to be significantly elevated compared to the germinating set. Of these, 40 spots gave MS data (either by MALDI TOF MS, MS/MS or both). No single peptide matched a database entry for a protein involved in lipid biosynthesis or metabolism, either in the NCBI*nr* or TIGR databases.

NCBI*nr* database entries for *R. communis* lipid metabolism proteins are limited. Database searches of the entire *Viridiplantae* database were made but for MALDI TOF MS the specific nature of PMF based MS searches and the sequence variability between different plant species means this technique is less likely to generate confident hits to non-*R. communis* entries. The sequence data generated by MS/MS allows for matches to be based on homology rather than just peptide mass, alleviating this issue to a degree. However, sequence coverage and similarity needs to be sufficiently high to obtain a confident match. Table 4.7 illustrates the limited nature of *R. communis* lipid metabolism related entries in the public protein database (current as of February 2009). Subcellular localisation information is shown and evidently the subset of ER lipid

metabolism proteins is restricted. The limited number of ER-resident lipid metabolism protein entries in the NCBI*nr* database might be a reason for not identifying any proteins of this type in the study. However, the MS data was re-analysed against a complete *R. communis* TIGR database and no lipid metabolism protein matches were made.

For those *R. communis* lipid metabolism related proteins present in the NCBI*nr* database and with known ER localisation, predicted molecular weight and pI information is shown. High molecular weight proteins (∼200 kDa), proteins with a pH close to or outside the pH 3 limit of the 3-10 IPG strip, or basic (pI >8) proteins (Yarmush and Jayaraman, 2002) are likely to be lost due to precipitation during focussing. The predicted M_{r} values suggest that molecular weights should not present a problem during focussing or IPG strip to gel transfer, although the very low molecular weight of acyl CoA binding proteins (gi|1938239) would be at risk of running off the bottom of the gel. pI values indicate a high pH for the DAGAT, FAD12, LPAT and FAH12 proteins (Table 4.7) although they should be resolvable as a pH 3-10 IPG strip was used.

A significant limitation of 2DE-based analyses is the poor solubility of hydrophobic proteins, especially membrane bound proteins, in 2D lysis buffer (Yarmush and Jayaraman, 2002). Where membrane bound proteins do solubilise, they can precipitate during the IEF stage of 2DE are lost from the analysis. This is likely to be the most significant reason as to why complex lipid biosynthesis enzymes were not identified in the analysis. Oleate $\triangle$12 hydroxylase was previously identified by SDS PAGE on NaCl and Na_2CO_3 treated *R. communis* ER preparations (Maltman et al., 2002), a treatment that causes extraction of associated and lumenal proteins and allows the separation of membranes and their integral proteins. Its localisation in this fraction confirms its membrane nature and suggests it is unlikely to be soluble in 2D lysis buffer. Furthermore as this report used the same ER purification technique as described here, it suggests that at least some components of lipid biosynthesis are present in the purified developing

ER samples. A comprehensive 2DE-based analysis of the whole developing seed of *B. napus* identified storage proteins as the dominant protein type within the developing seed (Hajduch et al., 2006). This study similarly utilised a combined MALDI-TOF and MS/MS approach. MS data were searched against the NCBI*nr* database and a *B. napus* tentative consensus database from The Institute of Genomic Research (TIGR), but not against a complete genomic database. Although some proteins related to lipid metabolism and fatty acid biosynthesis were identified, no complex lipid biosynthesis enzymes were found (full data at http://oilseedproteomics.missouri.edu). An earlier study examined changes of protein profiles during seed development in *G. max* by means of 2DE (Hajduch et al., 2005). Again, this study did not reveal any complex lipid biosynthesis enzymes. Assuming homology between *R. communis*, *B. napus* and *G. max* complex lipid biosynthesis genes[1], these studies support the hypothesis that complex lipid biosynthesis enzymes are not soluble in 2DE lysis buffer. Certain components of complex lipid biosynthesis such as phosphatidic acid phosphatase (PAP) are however known to be non-hydrophobic (Pearce and Slabas, 1998) and would be expected to be urea 2DE lysis buffer soluble, but were not seen in this study. It is possible that as associated proteins they were lost during subcellular fractionation.

Hydrophobicity plots are reported for four *R. communis* ER-resident enzymes involved in complex lipid biosynthesis or lipid modification (Figure 4.10). These enzymes were the only *R. communis* complex lipid biosynthesis database entries in the NCBI*nr* database at the time of the original analysis (Maltman et al., 2007) and thus provide supporting evidence as to why no proteins of this type were identified. Hydrophobicity analysis has been performed using the Kyte and Doolittle algorithm (Kyte and Doolittle, 1982). Hydrophobic regions are those with a K-D value > 0. Prediction of transmembrane domains was performed with the TMPred software at The Eukaryotic Linear Motif resource (http://elm.eu.org) and are boxed in red. These plots and

[1] *R. communis* and *B. napus* oleate $\triangle 12$ desaturase have 73% homology based on a ClustalW (Chenna et al., 2003) pairwise alignment

sequence motif predictions indicate the hydrophobic and membrane bound nature of the complex lipid biosynthesis enzymes in the NCBI database. This underlines the first point regarding solubility of complex lipid biosynthesis proteins in 2D lysis buffer being the likely cause of their absence from the analysis.

In summary, it is likely that the membrane bound nature of complex lipid biosynthesis proteins has prevented their solubilisation in 2DE lysis buffer and appearance in this analysis. Evidence of hydrophobicity and predicted transmembrane sites is presented for some of the components of *R. communis* complex lipid biosynthesis. Similar 2DE-based studies of developing seed which also had access to complete databases failed to identify complex lipid biosynthesis components. Although techniques exist to increase solubility of membrane proteins on 2DE gels such as alternative detergents (Rabilloud and Chevallet, 2000) a non-gel based proteomic analysis such as quantitative MS of integral membrane proteins will provide a more focussed analysis of the membrane fraction.

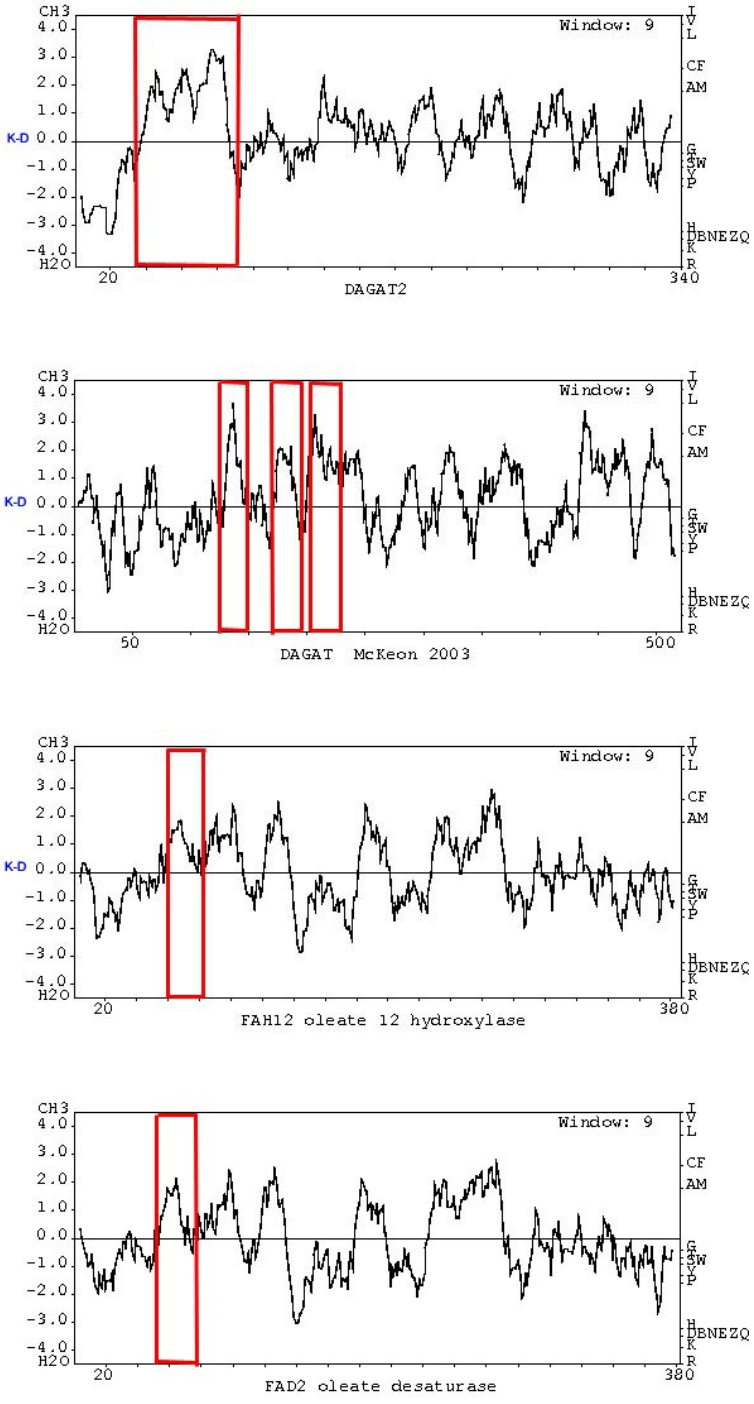


Figure 4.10: Hydrophobicity plots and predicted transmembrane domains of *R. communis* complex lipid biosynthesis proteins in the NCBI database. Hydrophobicity plots produced with the Kyte and Doolittle algorithm (Kyte and Doolittle, 1982). Predicted transmembrane domains boxed in red. This indicates these proteins are unlikely to be soluble in 2D lysis buffer or are likely to precipitate during IEF.

Table 4.7: NCBI*nr* *R. communis* database entries related to lipid metabolism (September 2007). Non-referenced "submission name + date" items were directly submitted to the NCBI database. Localisation information is shown with relevant references.

GI number	Submission author + date	Shorthand name	Full name	MR Pred.	pI Pred.	Localisation
ER localised protens:						
GI|114848908	Kroon 2006 (Kroon et al., 2006)	DGAT2	Diacylglycerol acyltransferase	38704	8.94	ER
GI|38146080	McKeon 2003	DGAT	Diacylglycerol acyltransferase	59894	8.39	ER
GI|117957296	Lu 2007 (Lu et al., 2007)	FAD2	oleate desaturase	43953	8.32	ER (Dyer and Mullen, 2001)
GI|1698844	Wang 1996 (Xu et al., 1996b)		phospholipase D	91992	5.44	vacuole, ER associated (Xu et al., 1996a)
GI|1938236	Erber 1997 (Erber et al., 1997)		acyl-CoA-binding protein	10051	5.83	endomembranes incl. ER (Li and Chye, 2003)
GI|1581593	Van de Loo 1995 (van de Loo et al., 1995)	FAH12	oleate 12-hydroxylase	44409	8.95	ER
GI|183211900	Fernandez-Garcia 2008	LPAT	1-acyl-sn-glycerol-3-phosphate acyltransferase	42696	8.92	ER
Non-ER localised protens:						
GI|1345969	Van de Loo 1994 (van de Loo and Somerville, 1994)		omega-3 fatty acid desaturase			plastid
GI|90110365	Liu 2006	LipRC1p	lipase			OB
GI|117957294	Lu 2007 (Lu et al., 2007)		acidic triacylglycerol lipase 2			OB
GI|118138597	Guy 2006 (Guy et al., 2006)		Stearoyl ACP Desaturase			plastid
GI|148791251	Sanchez-Garcia 2007	KASIII	3-keto-acyl-ACP synthase III			plastid
GI|148791249	Sanchez-Garcia 2007	KASI	3-keto-acyl-ACP synthase I			plastid
GI|148645269	Sanchez-Garcia 2007	KASII	3-keto-acyl-ACP synthase II			plastid
GI|294668	Genez 1993	KAS	β-ketoacyl ACP synthase precursor			plastid
GI|83320527	He 2005	ACS4	Acyl-Coenzyme A Synthetase 4			plastid
GI|83320525	He 2005	ACS2	Acyl-Coenzyme A Synthetase 2			plastid
GI|157417724	He 2005	ACS1	Acyl-Coenzyme A Synthetase 1			plastid
GI|170676820	He 2005	ACS1	Acyl-Coenzyme A Synthetase 1			plastid
GI|83320523	Burgal 2008	ACS4	Long chain ACS4			plastid
GI|62866924	Eastmond 2005		(castor endosperm) lipase			OB

Continued on Next Page...

GI number	Submission author + date	Shorthand name	Full name	MR Pred.	pI Pred.	Localisation
GI|1938236	Erber 1997 (Erber et al., 1997)		acyl-CoA-binding protein			endomembranes incl. ER (Li and Chye, 2003)
GI|55831356	Eastmond 2004		germinating oil-body associated lipase			OB
GI|38259660 (Eastmond, 2004)	Eastmond 2004		acid lipase			OB
GI|218023	Tsuboi 1991 (Tsuboi et al., 1991)		non specific lipid transfer protein-C			glyoxysomes, cytoplasm (cotyledons) (Weig and Komor, 1992)
GI|414732	Van de Loo 1993 (van de Loo and Somerville, 1994)		plastidial linoleoyl desaturase			plastid
GI|169711	Weig 1992		lipid transfer protein			glyoxysomes, cytoplasm (cotyledons) (Weig and Komor, 1992)
GI|169709	Weig 1992 (Weig and Komor, 1992)		lipid transfer protein			glyoxysomes, cytoplasm (cotyledons) (Weig and Komor, 1992)
GI|128380	Takishima 1988 (Takishima et al., 1988)	PLTP	Phospholipid transfer protein			microsomes, plastids (Dubacq et al., 1984)
GI|3334112	Erber 2005		Acyl-CoA-binding protein			
GI|128378	Takishima 1986 (Takishima et al., 1986)	PLTP	Phospholipid transfer protein			microsomes, plastids (Dubacq et al., 1984)
GI|224909	Takishima 1986 (Takishima et al., 1986)		Non-specific lipid transfer protein			glyoxysomes, cytoplasm (cotyledons) (Weig and Komor, 1992)

4.4.2 Chaperones and folding proteins dominate the soluble ER proteome

The proteomic study of *R. communis* seed ER described here and in Maltman et al. (2007) found chaperones and folding proteins to be the major component of the urea soluble ER in both germinating and developing seed (Figure 4.7, 4.8). BiP (the ER resident Hsp70), Hsp90, PDI, a PDI homologue and calreticulin were identified by MALDI TOF MS and MS/MS analysis. Chaperones and folding proteins are proteins that transiently interact with nascent proteins to catalyse the protein folding events and to prevent aggregates of partially formed proteins from occurring, or aid the refolding of damaged mature proteins (Gething and Sambrook, 1992). As the ER is the site of processing for all proteins of the secretory pathway, the cell surface and the assistant proteins of the endo- and exocytic pathways, there are substantial demands on protein folding and chaperone activity. Indeed, these components combined contain one third of the cellular protein in yeast (Ghaemmaghami et al., 2003). The ER thus has very specific protein folding demands and a high concentration of folding proteins and chaperones to assist this function.

4.4.2.1 BiP (Hsp70), DnaJ (Hsp40) proteins

BiP (Hsp70), so called because it was first identified in mammalian cells as a protein that 'Binds Immunoglobulin G Protein' (Hendershot et al., 1987) was identified by its PMF spectra matching the *N. tabacum* luminal binding protein (BiP). It was found to be elevated in the developing seed by 2.5 fold. A second protein (spot 1) was found to contain 3 peptides that were absolutely conserved with an *O. sativa* putative growth regulator with predicted Hsp70 domains, and matched Hsp70 in the TIGR database. This protein was elevated 5-fold in the developing seed. BiPs are known to bind short hydrophobic sequences of nascent amino acid chains that in the folded state would form the hydrophobic interiors of β-strands (Flynn et al., 1991) in a reversible ATP-

dependent manner. This binding by BiP prevents the nascent amino acid chain binding to itself or other hydrophobic components during translation.

Peptides generated from the 2.94 fold elevated developing spot 61 were sequenced by MS/MS and 2 peptides matched a hypothetical protein *O. sativa* containing DnaJ domains and a putative DnaJ in the TIGR database. The match had a good MOWSE score of 110 even though the sequence coverage was just 4% as the two peptides (YGEEGLK and SYYDVLQ) were absolutely conserved between the *R. communis* protein and the *O. sativa* hypothetical protein (Table 4.3). DnaJ domain-containing proteins belong to the Hsp40 family of chaperones and are known to aid BiP function by increasing its ability to hydrolyse ATP and thus bind to peptide chains (Bukau and Horwich, 1998). The degree of elevation of this protein in the developing state is comparable to BiP (2.94△ compared to 2.5△ for BiP) perhaps suggestive of their elevation in concert and its role as a co-chaperone.

4.4.2.2 GRP94 (Hsp90)

The final heat shock protein (Hsp) observed in this proteomic study is GRP-94, an ER resident member of the Hsp90 family. It is elevated 3.34 fold in the developing seed and was identified by MALDI TOF MS. Its presence in the ER has been previously reported (Melnick et al., 1994), indeed it is the only Hsp90 family member reported in this organelle. Like BiP it is a peptide binding protein, but thought to have a different subset of binding-peptides perhaps signalling an involvement in later stages of protein folding. Hsp90 is one of the most abundant proteins in the ER lumen of mammals (Koch et al., 1986) but is absent in yeast (Argon and Simen, 1999). It was first identified in plants by Walther-Larsen et al. (1993) as a highly elevated mRNA transcript during infection of *H. vulgare* by powdery mildew fungus (Walther-Larsen et al., 1993).

4.4.2.3 Protein disulphide isomerase

PDI and a PDI homologue were observed in the germinating and developing seed ER. PDI is a major component of the ER network of protein folding and chaperone enzymes. The strongly oxidising environment of the ER encourages the formation of disulphide bonds between the cysteine residues of nascent proteins (Hwang et al., 1992), providing stability and directing folding conformation of the growing protein. Although the formation of disulphide bonds can occur without the aid of protein catalysts PDI can alter disulphide arrangements to ensure the correct native state is achieved (Kainuma et al., 1995). PDI was identified in this study as the dominant gel spot (Figures 4.7 and 4.8) which was found to be up-regulated 7.55 fold in the developing seed (Table 4.3). The PDI homologue has a higher M_{r} than PDI and was identified by homology to a PDI homologue in *A. thaliana*. Its M_{r} was close to the predicted values for the *A. thaliana* protein. Three peptides were sequenced, of which one (VIVGNNFDEIVLDESK) was an absolute match with the other two exhibiting homology. Spots 32 and 33, exhibiting fold change values of 4.42 and 3.17 respectively, were also identified as PDI but were much less dominant within the profile than the major PDI and PDI homologue. These spots both matched the same database entry (gi|1134968, protein disulphide isomerase from *R. communis*) with good sequence coverage and MOWSE scores and may point to the post translationally modified PDI forms. On re-analysis of the MS data with the TIGR database, a further 7 spots were identified as PDI. The significant elevation and number of the PDI and PDI homologue proteins is indicative of the protein folding demands of the ER during seed development.

4.4.2.4 Glucosidase II

Glucosidase II alpha subunit was identified as a developing up-regulated high molecular weight protein by MS/MS sequencing (Spot 3 on Figure 4.7). Although sequence coverage was just 1% the MOWSE score is a reasonable 73 because the two sequenced

peptides are absolutely conserved with the *A. thaliana* protein. Furthermore it is a known resident ER enzyme (Saint-Jore-Dupas et al., 2006) and a Kyte and Doolittle hydropathy plot (Kyte and Doolittle, 1982) alongside TMPred transmembrane domain prediction (http://elm.eu.org/) of the alpha subunit suggests the *A. thaliana* protein is not membrane bound and thus is likely to be soluble in urea lysis buffer (Figure 4.11). Its identity was confirmed in the TIGR re-analysis where it was identified as a putative neutral alpha-glucosidase ab precursor with a MOWSE score of 205.

Glucosidase II removes the second and third glucose moieties from the *N*-glycans (Kornfeld and Kornfeld, 1985) added to asparagine (*N*) residues of nascent polypeptide chains by oligosaccharyltransferase, which forms part of the ER translocon complex (Johnson and van Waes, 1999). After the removal of the primary glucose it is the action of glucosidase II and its removal of the second glucose moiety which makes the *N*-glycosylated protein a substrate for the lectin class of chaperones (such as calreticulin) that target monoglucosylated glycoproteins (van Leeuwen and Kearse, 1996). Thus this is an essential enzyme in the activation of glycoproteins for chaperone action and further underlines the central role *R. communis* ER plays in ensuring correct protein folding.

4.4.2.5 Calreticulin

Calreticulin is a lectin chaperone identified as a major component of the developing and germinating urea soluble ER. This protein had a mean fold change value of 1.34 in the developing seed compared to the germinating seed but statistically the *p*-value for this change did not meet the 0.02 cut off, i.e. it cannot be confidently stated that this protein is increasing in developing gels. Its presence in the urea soluble fraction of the ER is expected as its mammalian equivalent is a lumenal protein (Peterson et al., 1995). Calreticulin forms one of a pair of lectin chaperones in mammals, the other being calnexin (Watanabe et al., 1994). Calnexin has been identified in *R. communis*

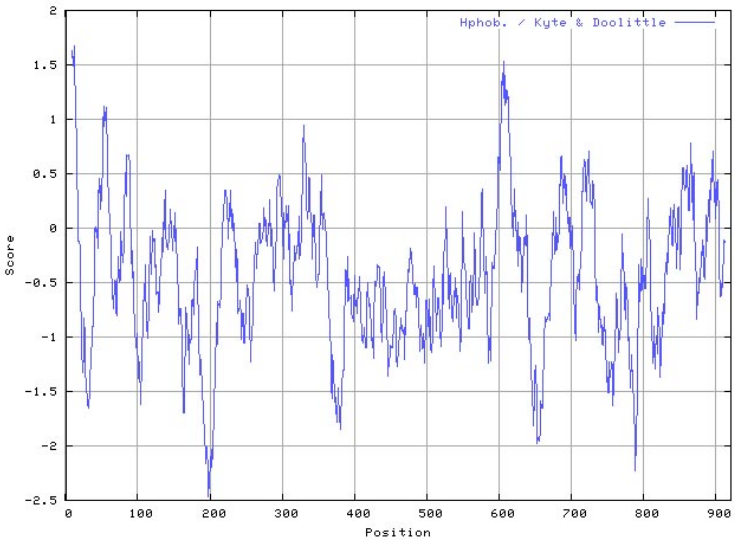


Figure 4.11: Kyte and Doolittle (Kyte and Doolittle, 1982) hydrophobicity plot of glucosidase II alpha subunit. TMPred analysis (http://elm.eu.org/) did not predict the presence of likely transmembrane domains within the sequence, indicating solubility of enzyme.

by immunoblot (Coughlan et al., 1997) but it is not present in the NCBI*nr* database at the time of writing. Although structurally and functionally broadly homologous to calreticulin, calnexin is a membrane anchored protein and this may explain why it was not identified by MS/MS-derived sequence homology in this study. Although originally thought to be primarily a Ca^{2+} binding ER protein (Michalak et al., 1992) it is now clear that calreticulin functions as a chaperone of specifically monoglucosylated lectins (Thomson and Williams, 2005). It's location in the ER is rationalised then as unlike the substrates for folding proteins outside the ER maturing secretory proteins are most often glycosylated (van Anken and Braakman, 2005). The observation that there is no confident decrease in the level of calreticulin in the germinating seed when other chaperones and folding proteins are seen to decrease may indicate its secondary role in Ca^{2+} binding (Meldolesi and Pozzan, 1998). The ER lumen has a Ca^{2+} concentration significantly higher than the cytosol and its abundance is attributed to a combination of sarcoplasmic / endoplasmic reticulum Ca^{2+} ATPase (SERCA) pumps and the large

amount of Ca^{2+} binding proteins in this compartment (Koch, 2005). Assuming that Ca^{2+} concentration is maintained in both germinating and developing ER you would expect the presence of Ca^{2+} binding proteins to be similar.

4.4.3 Seed storage proteins are highly elevated in the ER of developing seed

23 of the 40 spots giving matches to the NCBI*nr* database and 29 of the 56 spots giving matches to the TIGR database were seed storage proteins. This class of protein, a non-enzymatic store of amino acids, nitrogen and carbon for the germinating seedling, contributes the majority of the 10-50% protein that seeds of the major crop plants contain (Shotwell and Larkins, 1989). They are represented in this analysis by members of the 11S globulin and 2S albumin families.

4.4.3.1 11S Legumin

11S globulins are the major protein storage reserve in legumes and are well documented in the literature (Nielsen et al., 1989; Jain, 2004). They form large hexamer structures of 350 to 400 kDa, made up of six subunit pairs, each pair consisting of a basic 20 kDa polypeptide and an acidic 40 kDa polypeptide. The subunit pair is typically synthesised as a 60 kDa precursor protein, containing the *N*-terminal signal sequence, the acidic subunit, a linker peptide and the basic subunit. The typically 20-30 amino acid long signal peptide is removed co-translationally in the ER before possible glycosylation (Walburg and Larkins, 1983) . Further proteolytic processing occurs in protein bodies away from the ER. In soya glycinin a linker peptide is removed in at least three of the five known isoforms (Dickinson et al., 1989), although in other species including pea (Lycett et al., 1984) there is a single cleavage event, to give rise to the acidic and basic subunits link by a disulphide bond (Figure 4.12).

In this study, 10 individual spots were identified as belonging to the 11S legumin

class of seed storage proteins (9 in the NCBI*nr* analysis), which belong to the 11S globulin family (11S globulins are often given different trivial names depending on the plant species, whereas in *R. communis* and pea (*P. sativum*) they are called legumins, in soya bean (*G. max*) the 11S globulins are termed glycinins and in rapeseed (*B. napus*) they are termed cruciferins.) They were found to be elevated in the developing seed between 3.39 and 22.22-fold, and identified by both MALDI TOF MS and MS/MS sequencing. All but one of the spots identified as likely legumin storage proteins in the NCBI*nr* database matched either entry gi|8118510 *R. communis* legumin-like protein, or database entry gi|8118512 *R. communis* seed storage protein. The single spot which did not match these two *R. communis* entries gave MS/MS sequence data that matched with low confidence (MOWSE score of 58) to an *A. thaliana* putative protein (gi|7269954) of presently unknown function. This sequence also matched a legumin precursor from Mediterranean grain *V. narbonensis* (gi|600108) but with a MOWSE score of 49 it is not considered significant. On re-analysis with the TIGR database this spot was identified as a legumin A precursor, with a MOWSE score from the MS/MS peptides of 418 confirming its tentative identity from *A. thaliana* homology. On examination of the database amino acid sequences, gi|8118510 is 123 amino acids longer than gi|8118512. An alignment of the sequences with the MultiAlign sequence alignment package (Corpet, 1988) is shown (Figure 4.13). It is clear that gi|811510 has a 119 amino acid stretch at the *N*-terminal before alignment with gi|8118512 begins.

Disulphide bonds link the acidic and basic subunits and are formed between specific cysteine residues (Staswick et al., 1981). These cysteine residues were shown to be conserved across the five glycinin isoforms (Nielsen et al., 1989) and also between soya glycinin and peanut arachnin (Jain, 2004). On alignment of the five glycinin subunits with the *R. comminus* 11S legumin database entries matching peptides sequenced in this study it is clear the cysteine residues located within the basic subunit are conserved between soya glycinin and both *R. comminus* legumins (Figure 4.14). The known

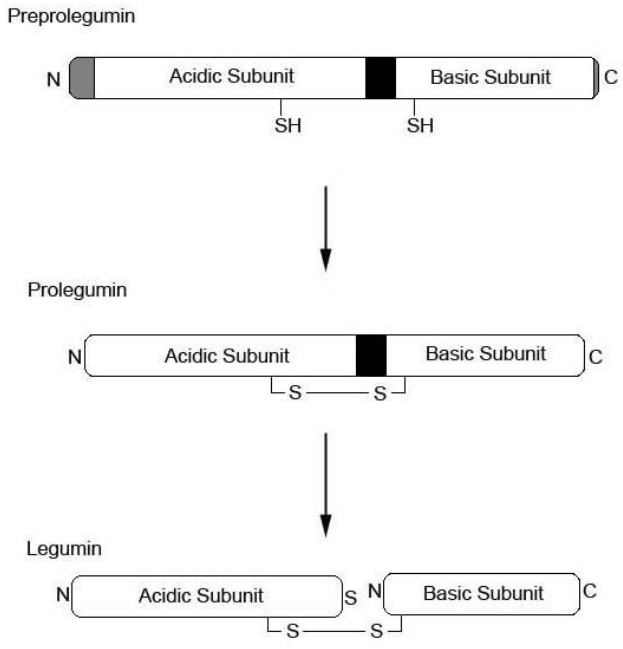


Figure 4.12: Processing of legumin precursor protein. Legumin is synthesised as an approx. 60 kDa two-subunit preprolegumin precursor protein including a *N*-terminal signal peptide (shaded grey) and linker peptide (black). The signal peptide is removed by a signal peptidase and a disulphide bridge is formed by PDI in the ER. Export of the prolegumin to protein bodies then occurs where the linker peptide is cleaved.

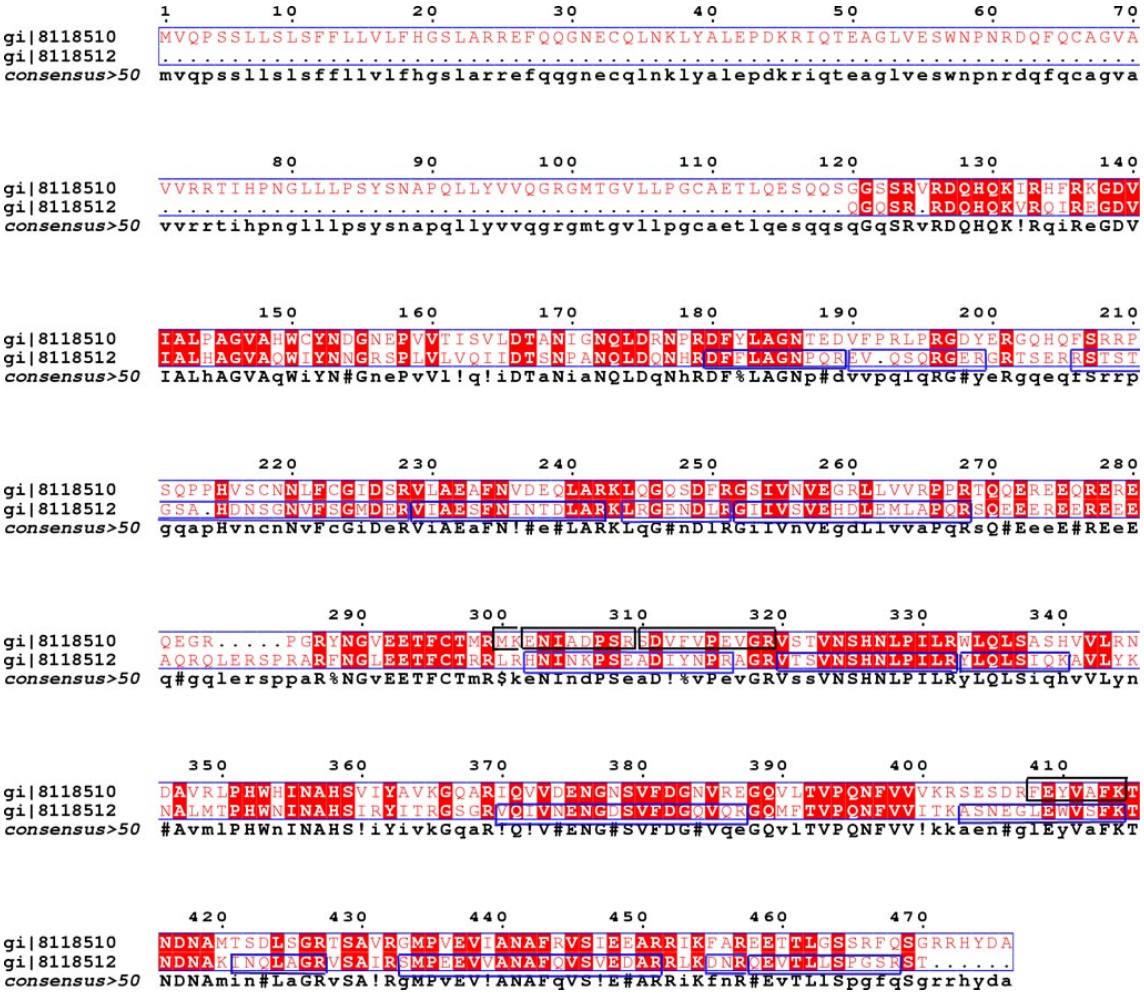


Figure 4.13: Alignment of gi|811510 and gi|8118512 *R. communis* legumin sequences. MS data generated from 8 significantly elevated developing spots matched either the gi|811510 or gi|8118512 sequences. The alignment shows that gi|8118512 to be 123 amino acids shorter than gi|811510. Boxes indicate matching peptides from the MS analysis.

disulphide bond-forming cysteine residue of the acidic subunit and two further cysteine residues identified as possibly involved in interchain linkages are also conserved between the glycinins and the gi|811510 476 amino acid database entry. In the short gi|8118512 sequence there are no cysteine residues present within the acidic subunit as this stretch of amino acid sequence is absent. This suggests that the sequence for this database entry is not full length and this is supported by the observation that the known N-terminal signal peptide is absent in gi|8118512. The observed M_{r} is similar to the predicted value for gi|811510 (for example Spot 42 the observed M_{r} is 53147 daltons, compared to a predicted value of 53668 daltons and for Spot 52 the observed M_{r} is 52445 daltons). Spot 54 also matched gi|811510 by MALDI MS, but the observed M_{r} for this spot is 50271 daltons. This may be explained by the presence of a legumin after removal of the signal peptide (20 amino acids each of an average of 110 daltons would give an M_{r} of 2310Da). For those spots giving sequence data matching gi|811512 the observed M_{r} is greater than is predicted (for example Spot 53, observed is 54383 daltons to a predicted mass of 40099 daltons). This is further evidence that the gi|8118512 database is curtailed. Unfortunately, no peptides were sequenced which did not fall within the gi|8118512 NCBI*nr* database entry which would have provided firm evidence that the sequence is not full length.

Another potential reason for the differences between observed and predicted molecular weight is post-translational modification. This is known to occur in the 7S globulins such as conglycinin in soya but thought to be less common in 11S globulins (Hurkman and Beevers, 1982). Classic glycosylation signals (triplets arginine-x-serine or arginine-x-threonine) have been seen in the peanut arachnin (Jain, 2004) however they are not seen in the two *R. communis* legumin sequences. Known cleavage points for the removal of the linker peptide are conserved across all the glycinins and the *R. communis* legumins. This linker peptide region is known to be a region of hypervariability as can be seen in Figure 4.14. This variability is also present between the two *R. communis*

legumins.

4.4.3.2 2S Albumin

Spots 75, 76 and 77 are developing seed ER proteins elevated up to 60-fold, identified in the NCBI*nr* database as 2S albumin precursors by both MALDI TOF MS and MS/MS sequencing. A further six spots were identified as belonging to the 2S albumin family after re-analysis against the TIGR database; in this database the sequence is assigned the name 'Sweet Protein Mabinlin-1 Chain A' but the sequence is identical to the gi|21068 entry. Spot 73 gave a confident hit to the TIGR 2S albumin precursor with a MOWSE score of 122. In the original analysis this gave a low confident match to RUBISCO based on MS/MS sequence homology; this underlines the caution that is required in making assignations based on homology where sequence coverage percentage and MOWSE scores are low.

The 2S albumins form the major allergen in *R. communis* (Thorpe et al., 1988) and were the first identified storage proteins in this organism (Youle and Huang, 1978). The sequence of 2S albumin was partly characterised by *N*-terminal sequencing and was shown to exist as a large (7kDa) and small (4kDa) subunit linked by a disulphide bond (Sharief and Li, 1982). This early effort allowed the cloning of the full length cDNA for a 29.3kDa precursor (Irwin et al., 1990). They confirmed the presence of Sharief and Li's 4 and 7kDa subunits but also proposed regions encoding a second large and small subunit within the precursor, similarly glutamine-rich and displaying a similar distribution of cysteine residues. At the *N*-terminus of the preproprotein they identified a signal peptide and gave a presumptive signal cleavage site. A schematic of the preproprotein structure is shown (Figure 4.15). Small and large subunits 2 (SS2, LS2) correspond to the sequences published by Sharief and Li (1982), and each subunit is separated by proposed intervening peptides. Processing of the preproprotein occurs in the vacuole after targeting of precursors to this compartment via precursor

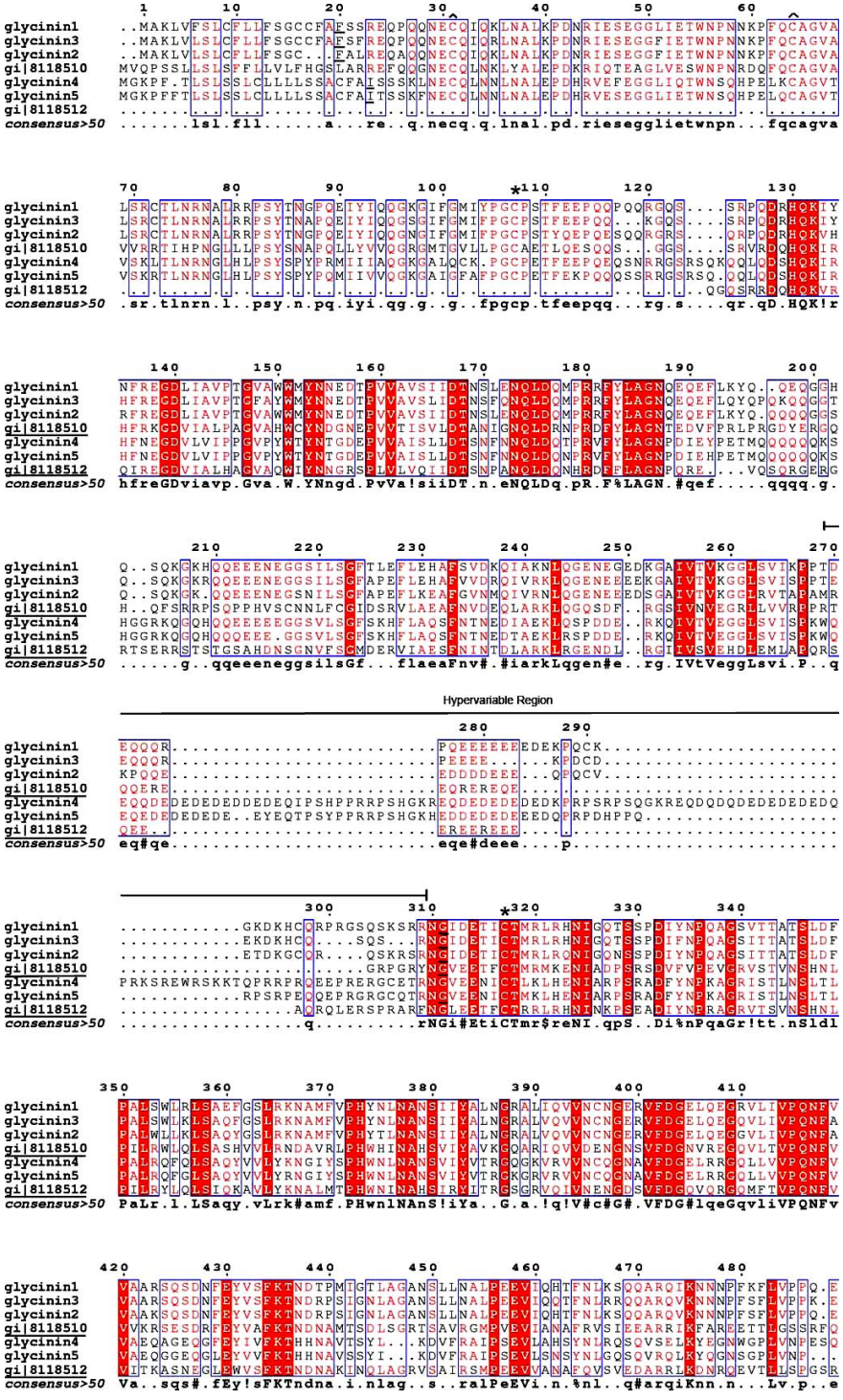


Figure 4.14: Alignment of soya glycinins (G1 to G5) and *R. communis* legumin (gi|8118512, gi|811510) sequences. * = site of disulphide bond formation in soya sequences. ∧ = conserved residues possibly involved with disulphide bond formation (Nielsen et al., 1989). Note homology with *R. communis* sequences in the acidic subunit, and to sequence gi|811510. There are no possible sites for disulphide linkage in the basic subunit of gi|811512 indicating it is likely a curtailed sequence. Underlined amino acids at residues 20/22 and 311 indicate likely cleavage sites for the signal peptide and acidic subunit respectively (Staswick et al., 1984). Note also the lack of signal peptide for sequence gi|8118512.

Figure 4.15: ***R. communis*** **2S albumin structure and MS sequence coverage**. **(A)** Structure of the 2S albumin preproprotein. Small and large subunits are labelled SS and LS respectively and designated 1 or 2 referring to the two 2S albumins of the mature protein. The signal peptide (blue) is co-translationally removed during entry into the ER lumen by a signal peptidase, the black arrow denoting the predicted cleavage location. **(B)** Amino acid sequence of *R. communis* 2S albumin preproprotein (gi|21068). Sequence coverage obtained in this study is indicated (solid underscore = MS/MS matches, dashed underscore = MALDI TOF MS). Adapted from Maltman et al. (2007)

accumulating vesicles (Hara-Nishimura et al., 1998; Brown et al., 2003). Three digested peptides from spot 77 were sequenced by MS/MS and their regions of alignment to the database entry gi|21068 are shown in Figure 4.15 (solid underlines). This shows that in our purified ER fractions the 2S albumin is present as a precursor supporting the hypothesis of vacuolar processing. Further evidence is given by the MS/MS peptide sequence starting TTIT which encompasses both a region of interchain linkage and the small subunit 1 (SS1) previously identified in the vacuole by N-terminal sequencing as a sequence beginning ESK.

4.4.3.3 Ricin and agglutinin

11 spots elevated in the developing ER gels were identified as either agglutinin or ricin. The fold change values compared to the germinating ER ranged from 3.21 (spot 11) to 53.77 (spot 18). The M_{r} values were all around 61-63kDa and the pI values ranged from 6-8. Ricin and agglutinin (RCA) are lectins with very similar amino acid sequences (Roberts et al., 1985), however ricin is cytotoxic and RCA relatively non-toxic. Both ricin and RCA are synthesised as preproproteins, ricin being a heterodimer composed of an A chain and B chain linked by a short linkage peptide. RCA is a tetramer consisting of two duplicated ricin-like A and B chains, indicating a duplication event at some point in its evolution. Processing of the preproprotein occurs in the ER, where the nascent protein's signal peptide is removed co-translationally after directing the translated sequence to the ER lumen. After PDI-catalysed disulphide bond formation in the ER the proricin is exported to protein storage vacuoles where final processing occurs; removing the interchain linkage to give the mature ricin. Figure 4.16 shows the amino acid sequence of database entry gi|21085 and the extent of sequence coverage by the MS analysis is shown. As the observed peptides include extents of the preproprotein covering the A-chain and B-chain; and the observed M_{r} value agrees with the predicted M_{r} for the complete preproprotein, it supports the observation that the mature ricin is synthesised outside of the ER. Assuming this is the case, it is encouraging that the abundance of these proteins in this study is not contamination from abundant developing seed protein storage vacuoles.

A significant heterogeneity of ricin has been reported in studies utilising SDS-PAGE, IEF and purification (Hedge and Podder, 1992). More recent work using advanced mass spectrometry analyses has identified glycosylation to be a major contributor to the heterogeneity seen, giving a pattern corresponding to successive additions of mannose sugar residues to the ricin lectin (McDonnell et al., 2000). In the analysis reported here glycosylated forms of ricin or RCA were not detected in the peptides analysed but it

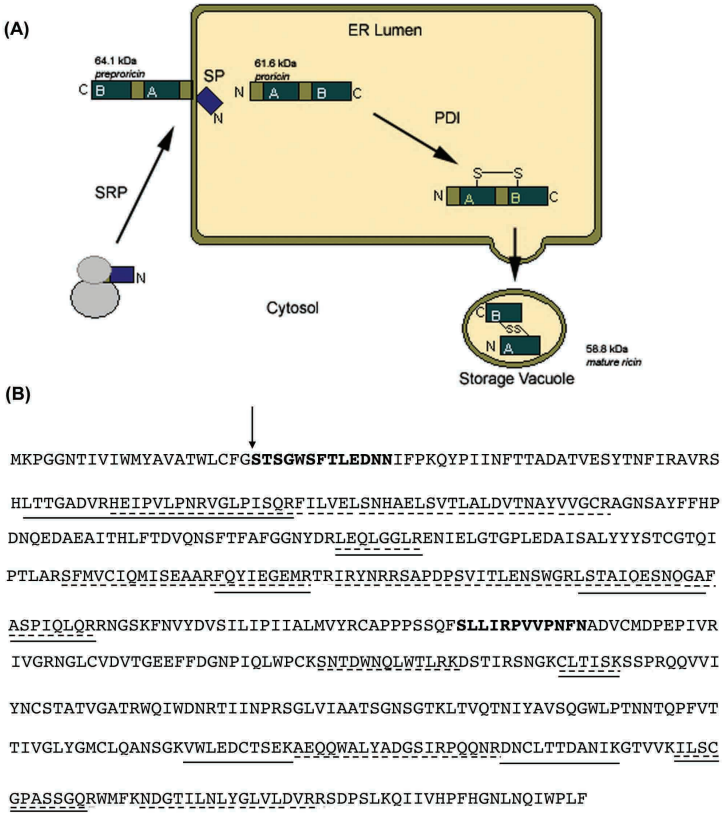


Figure 4.16: Processing of pre-proricinin and MS sequence coverage. Initial processing of the pre-proprotein occurs in the ER (**part A** of figure). The ricin signal peptide (SP) is bound by the signal recognition particle (SRP) and targeted to the ER where the SP is removed co-translationally. A disulphide bridge is formed by PDI, linking the two ricin subunits on final processing. On export the proricin is processed further to yield mature ricin consisting of A and B subunits. Pre-proricin sequence gi|21085 is shown (**part B** of figure), sequence coverage obtained in this analysis is underlined (MALDI TOF coverage shown with dashed lines, MS/MS with solid lines). Arrow indicates theoretical site of action for the signal peptidase. Sections of sequence shown in bold italic refer to the propeptides (between the signal peptide and the A chain, and between the A and B chains). This confirms that the ricin spots on the developing ER gel are of the proricin variety. Figure adapted from Maltman et al. (2007)

provides a likely hypothesis for the multiple ricin and RCA spots on the developing gel.

It has previously been reported that ricin is degraded by hydrolysis during the first few days after germination (McDonnell et al., 2000) and this is supported by the observation of a significant decrease in ricin in the germinating seed reported here and in Maltman et al. (2007).

4.4.4 Other identified proteins

16 elevated developing protein spots gave matches to TIGR database entries for proteins that do not fall into the other categories of folding proteins, chaperones and storage proteins (eight in the NCBI*nr* screen). Spot 8, a 6-fold elevated protein, gave a match in the NCBI*nr* database with reasonable confidence (MOWSE score of 99) to H^+ exporting ATPase. Interestingly, the TIGR analysis identified this as an ATP synthase alpha subunit (MOWSE score of 172) indicating sequence similarity between the two enzymes due to their similar (but reversed) biochemical function. Acidification of secretory pathway organelles through ATP catalysing enzymes has previously been identified, including in the ER (Okorokov et al., 2001). Spot 86 matched a putative progesterone binding protein from *A. thaliana* with both MALDI-TOF and MS/MS data. Further evidence for this assignation came from the TIGR database where the spot matched a putative steroid binding protein. Although not well characterised in plants they have been localised to the ER in mammalian systems (Falkenstein et al., 1998; Sakamotoa et al., 2004). Spots 84 and 85 were elevated 2.92 and 7.29-fold respectively and matched an NBS-LRR disease resistance protein from *H. vulgare*. NBS-LRR proteins are the largest group disease resistance proteins in plants, are ubiquitous in dicots and contain a characteristic nucleotide binding site and leucine rich repeats. A recent genome analysis in *A. thaliana* identified over 14 NBS-LRR disease resistance genes, the majority of which encode sequences homologous to known *R* (resistance) genes although they stated that there's limited evidence for their

involvment in other aspects of plant biology, such as development (Tan et al., 2007). In the TIGR re-analysis both of these protein spots matched a 'conserved hypothetical protein' suggesting the evidence for this protein being an NBS-LRR protein is not yet conclusive in *R. communis*. Spot 2 matched the key terpenoid biosynthesis enzyme strictosidine synthase. Terpenoids are ubiquitous in nature and play essential roles in membrane fluidity, electron transport, glycosylation of proteins and the regulation of cellular development (McCaskill and Croteau, 1998). A putative oligopeptidase A protein was identified in the TIGR database; upregulated 3.4 fold in the developing seed. Oligopeptidase B has been identified as a protein involved in protein mobilisation in the germinating wheat seedling (Tsuji et al., 2004) and it is possible the oligopeptidase here performs a similar function, perhaps in protein quality control associated with the high levels of protein translation and modification in this development state. It should be noted though that the MOWSE score for this assignation is low (57). Two spots were identified in the TIGR database analysis as putative 9-*cis*-epoxycarotenoid dioxygenase enzmyes (Spots 15+16, MOWSE scores of 177 and 80 respectively). This is a key regulatory enzyme in abscisic acid biosythesis. Both spots are modestly elevated in the developing seed (5.21 and 4.29 fold respectively). Abscisic acid, in the context of the developing seed, is important in controlling seed dormancy and preventing precocious germination. It has previously been shown to inhibit the synthesis of germinating specific proteins and allow the continued accumulation of seed storage proteins (Karssen et al., 1983; Vernieri et al., 1989). It is also important in regulating TAG biosynthesis (Rodriguez-Sotres and Black, 1993; Pacheco-Moises et al., 1997). Two spots (20 and 23) gave confident matches to the TIGR database entry 'nucleolar protein nop56', with MOWSE scores of 154 and 319 respectively. Nop56 is known to form part of the small nucleolar ribonucleoprotein complex the C/D box; through its action of methylating ribosomal RNA it is important in ribosomal maturation and ultimately protein translation (Matera et al., 2007). Thus its presence in the ER and especially

the developing ER is easy to rationalise. It has been identified in diverse kingdoms of life including Planta (Barneche et al., 2001; Minglin et al., 2005).

The key sterol biosynthethic enzyme cycloarternol synthase was identified in the TIGR screen. This enzyme catalyses the conversion of 2,3-oxidosqualene to cycloarternol. It was first identified in *A. thaliana* microsomes (Corey et al., 1993). Plant sterols are critical in plant development (Clouse, 2002; Lindsey et al., 2003).

Sorting and assembly machinery protein 50 (Sam50) was identified from a protein spot elevated 4.53-fold in the developing ER, with a significant MOWSE score of 210 and 24 matching peptides. Sam50 is known to be a mitochondrial outer membrane protein and is involved in protein import into the mitochondria (Becker et al., 2008). The protein can be characterised by its transmembrane beta-barrels, a conformation common in bacteria but also chloroplasts and mitochondria due to the endosymbiotic origins of these organelles. It is possible that its presence in the developing ER sample is due to the stable interactions which occur between mitochondria and the ER (Filippin et al., 2003). Its elevation in the developing state might be due to temporal alterations in mitochondrial / ER interaction. Alternatively, a related protein, porin, is known to readily translocate from the mitochondria to the ER (Sakaguchi et al., 1992), and this may be example of the same process for Sam50.

Cytochrome P450 is an ER-resident membrane protein (Szczesna-Skorupa et al., 1998) involved in electron transport and catalysing monooxygenase reactions. A Kyte and Doolittle plot of the hydrophobicity of *A. thaliana* protein sequence predicted one possible transmembrane domain (result not shown) (Kyte and Doolittle, 1982). This provides confirmation that some membrane bound proteins are separable in a 2DE system.

4.4.5 35 developing proteins gave no identity

35 protein spots elevated on the developing gel were sequenced but gave no identities after searching both NCBI*nr* and TIGR databases. The majority of these gave very weak MALDI spectra. It is possible that if peptides digested from a higher loading preparative gel were re-sequenced more data would be obtained.

4.4.6 Proteins identified as significantly increased in germinating seed

Fifteen spots were found to be significantly elevated in the germinating seed ER compared to the developing seed. Only one of these spots, number 15 (Table 4.4) gave an identity by MALDI TOF MS in the original NCBI*nr* analysis. This 5.19-fold elevated spot gave a significant (MOWSE score = 181, 50% sequence coverage, 30 matching peptides) hit to the *R. communis* malate synthase enzyme (gi|21076). In the subsequent TIGR analysis it was identified again as malate synthase, with a MOWSE score of 201. Malate synthase is an enzyme of the glyoxylate cycle, responsible for the condensation of acetyl CoA and glyoxylate to form malate before subsequent conversion to oxaloacetate and potentially glucose. Seedlings utilise the glyoxylate cycle for the conversion of stored lipid to sugar compounds during germination so its presence in this stage in an oilseed such as *R. communis* is not unexpected. As a peripherally-associated glyoxysomal enzyme (Huang and Beevers, 1973) it was initially a surprise to identify it within our ER preparations. However its association with the ER is well documented (Gonzalez and Beevers, 1976; Gonzalez, 1982; Bowden and Lord, 1976a). Gonzalez (1982) found that malate synthase activity overlapped that of the ER marker enzyme NADH-cytochrome *c* reductase when 3-day homogenised germinating *R. communis* seeds were fractionated through a continuous sucrose gradient (Gonzalez, 1982). It was also found that malate synthase activity remained associated with the ER in preparations made without $MgCl_2$, causing ribosome dissociation from the ER gradiant thus giving the organelle a reduced buoyant density, further suggesting linkage

between the ER and malate synthase activity. It was hypothesised that glyoxysomes emerge directly from dilated cisternae of ER and in early germinating seedlings there remains an association between these two organelles. This is supported by findings such as highly similar phospholipid profiles between the ER and the glyoxysomal membrances (Donaldson and Beevers, 1977), the morphological similarities between the two compartments (Vigil, 1973), and similarities in the polypeptide composition between the two compartments (Bowden and Lord, 1976b). The presence of malate synthase within our germinating preparations support these earlier observations.

MS/MS sequencing identified two spots in the NCBI*nr* database with a peptide sequence matching that of strictosidine synthase enzyme from *A. thaliana*. These spots were modestly elevated in germinating seed (2.4 and 2.64-fold) and displayed similar M_{r} and pI values to the *A. thaliana* predicted figures. These spots plus an additional spot residing in the same streak (previously identified as NADH dehydrogenase in the NCBI*nr* database) were also identified in TIGR re-analysis by PMF. Strictosidine synthase is a key enzyme in the biosynthetic pathway of terpenoid indole alkaloids and has been well-researched in *Catharanthus roseus* (Madagascan periwinkle) and *Rauvolfia serpentina* where the alkaloids have significant medicinal value as anti-tumour drugs. An enzyme immediately downstream of strictosidine synthase, strictosidine β-D-glucosidase, has been localised to the ER (Geerlings et al., 2000) and the data presented here provides further information of the cellular location of this pathway in *R. communis*.

4 peptides from a 5.15-fold elevated germinating spot (spot 12) matched pea (*P. sativum*) porin (gi|396819). Further evidence of its identity came from the TIGR analysis, where this spot and another adjacent spot were identified as putative voltage-dependent anion-selective channels. Porins are ubiquitous voltage-gated diffusion pores found in all eukaryotic kingdoms and first identified in plants in 1994 (Fischer et al., 1994). There is close interaction between the mitochondria and ER in eukaryotic cells

responsible for mediating protein transport, Ca^{2+} signalling, ATP and cell death signals (Szabadkai et al., 2006). It has previously been shown that mitochondrial porin readily translocates across ER membranes *in vitro* (Sakaguchi et al., 1992) and it's proposed that the intracellular transport of proteins, Ca^{2+}, ATP and other metabolites occurs via ER localised porins in rat cerebellum (Shoshan-Barmatz et al., 1994).

4.4.7 Concluding remarks

A differential proteomic analysis of germinating and developing *R. communis* seed ER has been performed. Levels of protein were accurately quantified allowing statistically significant differences between the two developmental states to be identified and analysed further. On analysis of these significant differences, developing up-regulated spots were found to be dominated by folding proteins, chaperones and storage proteins.

The identification of folding proteins and chaperones as the major component of the *R. communis* seed urea-soluble ER indicates the major role protein folding and thus protein production has in this compartment. Major known lumenal protein folding families of BiP, Hsp90, PDI, Hsp40 and calreticulin were represented revealing the extensive protein folding functionality of the *R. communis* ER. All of these identified proteins were at higher levels in the developing seed with the exception of calreticulin, indicating the role of protein synthesis in the developing ER. This is consistent with the storage reserve production activities of the developing seed, which includes storage proteins.

Proteins involved in complex lipid biosynthesis were not identified in the 2DE lysis buffer soluble fraction. The likely reasons for this have been discussed. The study presented here is currently being extended by an analysis of the germinating and developing *R. communis* ER integral membrane by iTRAQ. This component of the proteomic investigation was outside the remit of my industrial funding, but will hopefully overcome the limitations of 2DE and complement the work presented here.

Chapter 5

Establishment of the growth, harvest and lipid analysis methodologies for the use of *Yarrowia lipolytica* as an *in vivo* tool for the assay of *Ricinus communis* complex lipid biosynthesis genes.

5.1 Introduction

The previous two chapters examined the results of a differential 2D DIGE based screen of developing and germinating ER and the development of rigourous proteomic methodologies for its successful execution. This soluble proteomic analysis formed one component of the two pronged proteomic approaches to the discovery of lipid biosynthesis machinery in *R. communis* developing seed (the others being the mass spectrometry-based analysis of the ER membrane). Although no components of lipid biosynthesis were identified in the soluble proteomic study (Chapter 4), an in-house EST database has generated interesting candidates (for example, the first DAGAT2 in *R. communis* (Kroon et al., 2006)).

The ability to test the effects of these candidate enzymes on oil quality is crucial, to both validate their function and to quantitatively assess their contribution to high levels of triricinolein production in new plant vehicles. Although *in vitro* assays are useful for validating function, model organisms are used to assess the influence of a gene product within the physiological and biochemical context of the whole organism. A distinct disadvantage of plant models such as *A. thaliana* organisms is the time required to produce stable transformant lines and the long period of time (around 3 months) between planting new genetically modified seed and being able to analyse the lipid content of the storage oil in the transgenic developing seed. The ability to circumvent this with a model system capable of producing analysable lipid within just 24 hours of starting growth is thus highly desirable. *Yarrowia lipolytica* is a non-pathogenic yeast with strains available that are suitable for genetic experimentation (reviewed in Chapter 1). Crucially, it produces significant storage lipid when in the stationary phase of growth. This chapter details the methodological development of growth, harvest and lipid analysis methodologies required to use *Y. lipolytica* as a rapid lipid gene assay vehicle.

For the investigation of hydroxy-lipid incorporation, methyl ricinoleate was used in all analyses rather than ricinoleic acid. This is because high purity methyl ricinoleate was available in quantities sufficient for routine large scale growth: critical for generation of enough lipid material for analysis. The choice of methyl ricinoleate over alternatives is discussed in Section 5.4.4.

5.2 Aims

The characteristics of *Y. lipolytica*, that is, its capability of utilising hydrophobic substrates such as fatty acid methyl esters and its production of large quantities of triacylglycerol, make it a potential useful candidate for *in vivo* assay of *R. communis* lipid biosynthesis genes. The degree to which an untransformed strain of *Y. lipolytica* incorporates ricinoleoyl CoA into its storage oil compared to a strain transformed with a *R. communis* acyl transferase may provide useful evidence in identifying the importance of the *R. communis* enzyme in oil quality. Different combinations of *R. communis* lipid biosynthesis genes can be transformed to identify a strain with the highest level of triricinolein production, and thus this set of *R. communis* genes becomes a first choice for evaluation in *A. thaliana*. Prior to the stage whereby transformed and untransformed strains can be compared, rigourous methodology must be established so that experimental procedures are well understood and confidence can be had in the quantification data obtained. Therefore, the aim of this chapter is to establish the growth conditions, cell harvest method, lipid extraction methodologies and lipid analysis procedures needed to ultimately quantify the components of the triacylglycerol in *Y. lipolytica* when fed on hydrophobic substrates including methyl ricinoleate in both transformed and untransformed strains.

5.3 Results

5.3.1 Maintenance of yeast stocks

The *Y. lipolytica* variety P01G (MAT**a**; *leu2-270*; *xpr2-333*; *axp1-2*) was kindly received from Prof. JM Nicaud, Laboratoire de Génétique des Micro-Organismes, Thiverval-Grignon, France. Replicate cultures of the strain were initially grown on selective media with the addition of leucine to account for auxotrophy. The culture of *Y. lipolytica* was grown to exponential phase, at which point aliquots were archived in glycerol protected media at -80 °C (see Materials and Methods, Chapter 2.4, for information on liquid culture media, culture plates and glycerols). Culture plates were made with selective media containing leucine. All liquid cultures of *Y. lipolytica* or *S. cerevisiae* used in this study were initiated from plates as this allowed selection of single parent colonies and reduced the possibility of maintaining contaminants within growth media.

5.3.2 Evaluation of the growth characteristics of *Y. lipolytica* PO1G

The pattern of growth of the *Y. lipolytica* auxotrophic strain P01G on 2% glucose was evaluated, so that the timings of the different stages (i.e. lag phase, exponential phase, stationary phase) of growth in this *Y. lipolytica* strain could be determined. For comparison, the auxotrophic *S. cerevisiae* strain Y00000 (MAT**a**; *his3*$\triangle$1; *leu2*$\triangle$*0*; *met15*$\triangle$*0*; *ura3*$\triangle$*0*) (Brachmann et al., 1998) was used. Cells were grown separately in two 250 ml baffled conical flasks containing 50 ml YNB (minimal media) and 2% glucose, 1.7% yeast nitrogen base, 0.5% NH_4Cl and 50 mM phosphate buffer pH 6.8. Leucine, uracil, histidine and methionine were added as required. The two flasks were inoculated with *Y. lipolytica* and *S. cerevisiae* cells respectively to an initial OD of 0.1. Inoculations were made from a liquid YNB parent culture that was in exponential growth phase. Two independent replicates of *Y. lipolytica* and *S. cerevisiae* cells were grown and measured, and two aliquots of well mixed culture were removed from the

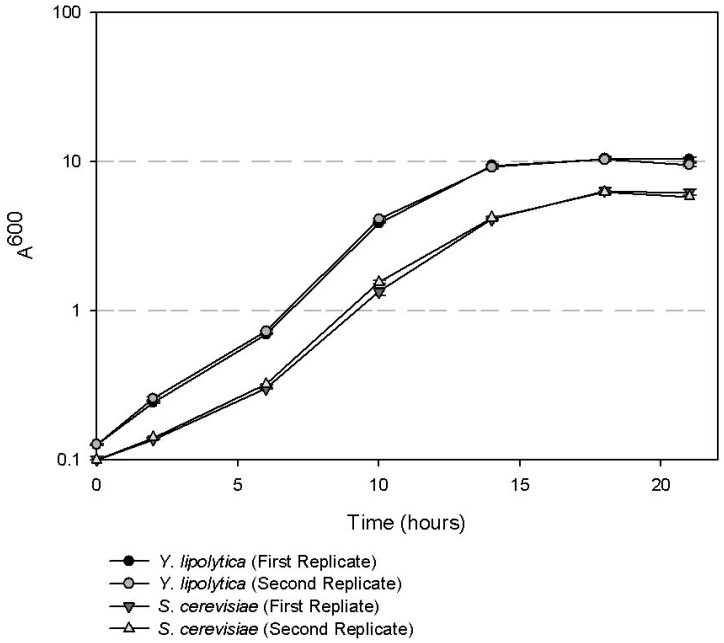


Figure 5.1: Growth curve of *S. cerevisiae* Y00000 versus *Y. lipolytica* P01G on minimal YNB media containing 2% glucose. The growth of two independent replicates for each yeast were measured, and for each replicate two OD readings were taken (the average ODs with standard deviations are plotted). *Y. lipolytica* P01G was found to grow more efficiently than *S. cerevisiae* Y00000 on YPD.

flasks at each time point. Growth was measured by recording the absorbance of the culture media at 600 nm in a spectrophotometer, which was blanked with fresh YNB from the same batch used for the growth of the cells. As spectrophotometers are unable to accurately measure absorbance of cultures above an OD of 1.5, culture media (and the respective blank sample) was diluted with ddH_2O to keep the absorbance reading below 1.00, and multiplied by the dilution factor to obtain the correct reading.

The average of the two pipetted aliquots from a single culture was plotted against time with standard deviation error bars. Two biological replicates were analysed for both yeasts (labelled as First Replicate and Second Replicate for *Y. lipolytica* and *S. cerevisiae* in the figure). Figure 5.1.

Y. lipolytica P01G grew faster than *S. cerevisiae* Y00000 on YNB containing 2% glucose. Both yeasts were in lag phase at the 6 hour time point, and by 10 hours, the early exponential growth phase. However, both the early exponential growth phase and the exponential growth component of the curve was steeper for *Y. lipolytica* than *S. cerevisiae* indicating a more efficient conversion of carbon substrate and other nutrients into new biomass. Between 13 and 18 hours the *Y. lipolytica* curve began to plateau indicating the entry of growth into a late exponential or early stationary phase. By 18 hours the *Y. lipolytica* culture had entered stationary phase, with an OD of 9.5 - 10.5. The two *Y. lipolytica* independent replicates displayed essentially identical patterns of growth, until the final time point at 22 hours where the OD of one of the replicates was decreased compared to the other. This may indicate the beginnings of cell senescence however the error bars indicate greater variance of spectrophotometer reading at this and the previous time point. This may suggest the difference between the readings is due to pipetting errors or increasing culture heterogeneity as the cells enter stationary phase; *Y. lipolytica* is a dimorphic yeast and the yeast state to hyphae transition is known to be effected by reduction in nutrient sources (Ruiz-Herrera and Sentandreu, 2002; Szabo, 1999). *S. cerevisiae* Y00000 had entered stationary phase by the 18 hour

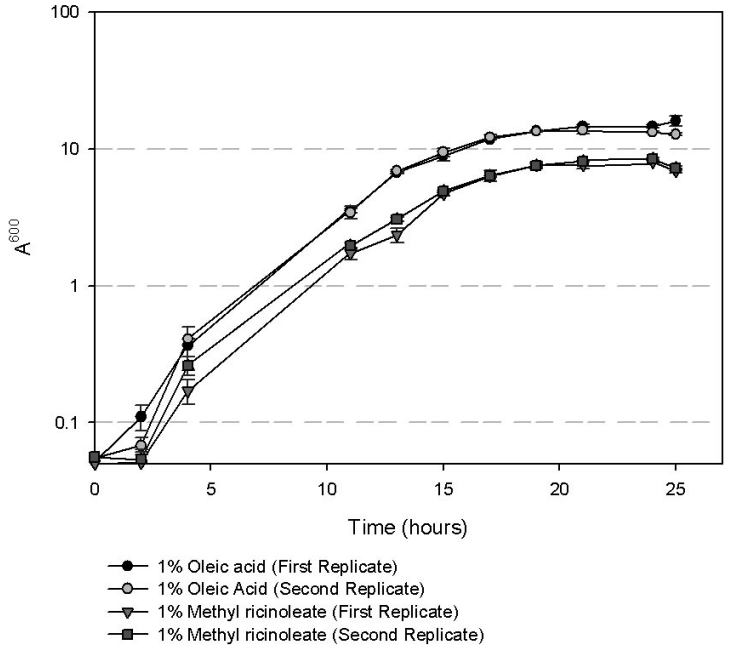


Figure 5.2: Comparison of the Growth Curve of *Y. lipolytica* P01G Growing on Methyl Ricinoleate or Oleic Acid. Cells were grown on YNB minimal media containing the carbon substrate and the required drop-in nutrient leucine to account for auxotrophy. The growth of two independent replicates for each carbon substrate were measured, and for each replicate two OD readings were taken (the average ODs with standard deviations are plotted). Duration of growth stages (i.e. lag phase, exponential phase) were similar on both media, but the total rate and biomass produced was higher on oleic acid media compared to methyl ricinoleate.

time point, with an OD of ~ 6.0, a substantially lower total biomass than *Y. lipolytica* at the same stage of growth. It is likely that the auxotrophic nature of the yeasts analysed influenced the rate and final amount of biomass produced.

5.3.3 Growth of *Y. lipolytica* on methyl ricinoleate

The growth pattern of *Y. lipolytica* P01G growing on 1% oleic acid or 1% methyl ricinoleate was compared. A single parent culture of *Y. lipolytica* grown on 2% glucose YNB minimal media was used to inoculate 50 ml YNB cultures in 250 ml baffled conical flasks, containing either oleic acid or methyl ricinoleate emulsified by sonication in the

presence of Tween 80 (see Chapter 2.4.1). Tween 80 is a surfactant and emulsifier derived from oleic acid and commonly used in the creation of lipid stocks for *Y. lipolytica* feeding experiments. It has previously been reported that its presence in the culture media at 10% (w/v) has negligible affect on cell growth and lipid accumulation (Papanikolaou et al., 2003). Culture media also contained leucine to account for the auxotrophic nature of the P01G strain. Cultures were inoculated to a starting OD of 0.1, and cells were washed with ddH_2O between transferral from the 2% glucose parent culture to the 1% lipid growth curve cultures. Although transferring from one carbon source type (glucose) to another (lipid) will be accompanied by metabolic changes in the yeast and possibly an increased lag phase, this method was used for all further experiments where *Y. lipolytica* is grown on lipid, so the growth phase timings are still valid. Cells were grown and the growth measured as described (Chapter 2.4.2). The presence of lipid in the media prevented accurate reading of OD by the spectrophotometer, so three washes of 0.5% (w/v) solution of the lipid sequestering protein BSA and a final wash of ddH_2O were employed (Mlícková et al., 2004).

Cultures remained in lag phase at 2 and 4 hours after inoculation. By 11 hours of growth the oleic acid cultures had reached an OD of $\sim$ 3.0 and the methyl ricinoleate cultures an OD of $\sim$ 2.0. Measurements of growth between 11 and 19 hours showed an exponential pattern in both oleic acid and methyl ricinoleate cultures, although the rate of growth was depressed in the methyl ricinoleate culture during the exponential phase. Between 21 and 23 hours both cultures exhibited a stationary pattern of growth with a total OD value for oleic acid fed cells between $\sim$ 13.5 - 14.8. In contrast the total amount of biomass for methyl ricinoleate fed cells is lower, with an OD value of between $\sim$ 7.5 to 8.0. Therefore, the timing of growth is similar for cells grown on both carbon substrates, however the rate and total biomass produced is lower in cells growing on methyl ricinoleate than oleic acid. This observation has been reported previously (Ratledge, 1984) and was hypothesised to relate to the direct cytotoxicity

of ricinoleic acid or the cytotoxicity of a product of its metabolism (see Discussion). Although the rate of growth and total conversion of nutrients to biomass was depressed in cells growing on methyl ricinoleate compared to oleic acid, the times of growth stage transition were the same for the cells growing on either carbon substrate. This is important as it simplifies harvest timings in experiments comparing oleic and methyl ricinoleate grown cells at a desired growth phase.

5.3.4 Characterisation of *Y. lipolytica* P01G auxotrophy

Y. lipolytica P01G is auxotrophic for the essential amino acid leucine, that is, it is unable to produce leucine itself and requires the external addition of the amino acid to the media. Auxotrophic strains are useful tools in microbial molecular biology. By growing cells transformed with vectors containing the absent metabolic gene in a minimal media (i.e. without supplementation) they are selected over those cells that do not contain the vector and additional genes. Auxotrophy can also be used to validate untransformed yeast strains.

To characterise the auxotrophic status of *Y. lipolytica* P01G, cells were streaked on minimal media agar plates containing either (1) no supplements : -/-, (2) uracil: U/-, (3) leucine : -/L, or (4) leucine and uracil: U/L. The agar media also contained a carbon substrate, either 1% (w/v) oleic acid or 2% (w/v) glucose. Cells were grown for 3 days at 28 °C in the dark, and the extent of growth evaluated for each plate.

Examining the 4 glucose plates (Figure 5.3), growth is clearly seen on plates containing uracil and leucine or just leucine. Where leucine was absent (i.e. -/-, or U/-) no growth was seen. The same pattern is seen on the oleic acid plates, although the overall extent of growth was less. This confirms that *Y. lipolytica* is dependent on leucine for growth, and that *Y. lipolytica* P01G is capable of utilising both glucose and oleic acid for its growth.

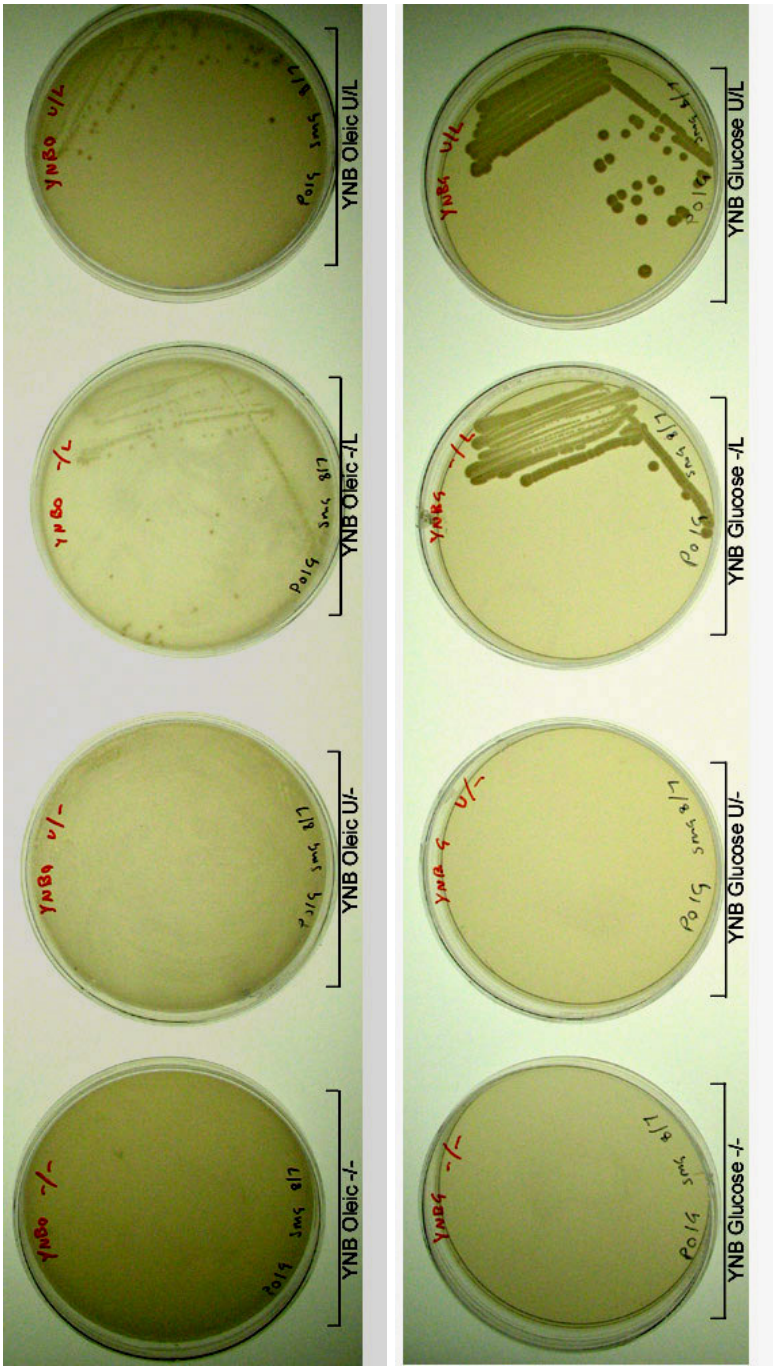


Figure 5.3: Characterisation of *Y. lipolytica* P01G (MATa; leu2-270) auxotrophy. Minimal media agar plates supplemented with: U/L = media contains uracil and leucine, U/- = media contains uracil, -/L = media contains leucine, -/- = no drop-in supplements. Plates confirm that P01G is leucine dependent.

5.3.5 Evaluation of derivatisation on the migration of ricinoleate-containing TAG on thin layer chromatography (TLC)

A common method for the separation of lipid extracts is thin layer chromatography (TLC). Using TLC and different solvent combinations, neutral lipids can be fractionated and identified based on their discreet migrations. Fractionated lipid can be removed from the TLC plate for further analysis, for example to identify the fatty acid components of triacylgycerol by gas chromatography (GC).

Migration of lipid species within a TLC plate is effected by the lipid's chemical properties. The polar nature of the free hydroxyl groups of triricinolein (3-OH TAG) and other hydroxy fatty acid-containing molecules such as ricinoleic acid or methyl ricinoleate results in a greater interaction between the lipid molecule and the TLC plate matrix, retarding their migration through the TLC plate compared to their non-hydroxy equivalents. If a mixture containing 3-OH TAG molecules and non-hydroxy TAG molecules are separated by TLC, the 3-OH TAG and non-hydroxy TAG will have separate positions on the TLC plate due to their different chemical properties. If the lipid mixture also contains monoricinolein diolein (1-OH TAG) and 2-OH TAG these will also have different migratory characteristics and thus different positions on the TLC plate. The presence of other hydroxy lipid molecules, such as hydroxy DAG or MAG molecules, further increases complexity. Although this might be an advantage for qualitative analyses to identify a broad range of molecules quickly, it makes quantifying total TAG in a sample problematic. The development of a lipid analysis procedure that will allow quantification of triricinolein production in *Y. lipolytica* cells fed on hydroxy fatty acid is of central importance to this chapter's aim.

To overcome the retardation of hydroxy-TAG on TLC plates, the use of silylating reagents was investigated as a means to convert the polar hydroxyl group into a non-polar functional group, thus conferring on the hydroxy-TAG molecule the same migration characteristics as non-hydroxy TAG molecules such as triolein. Silylation occurs

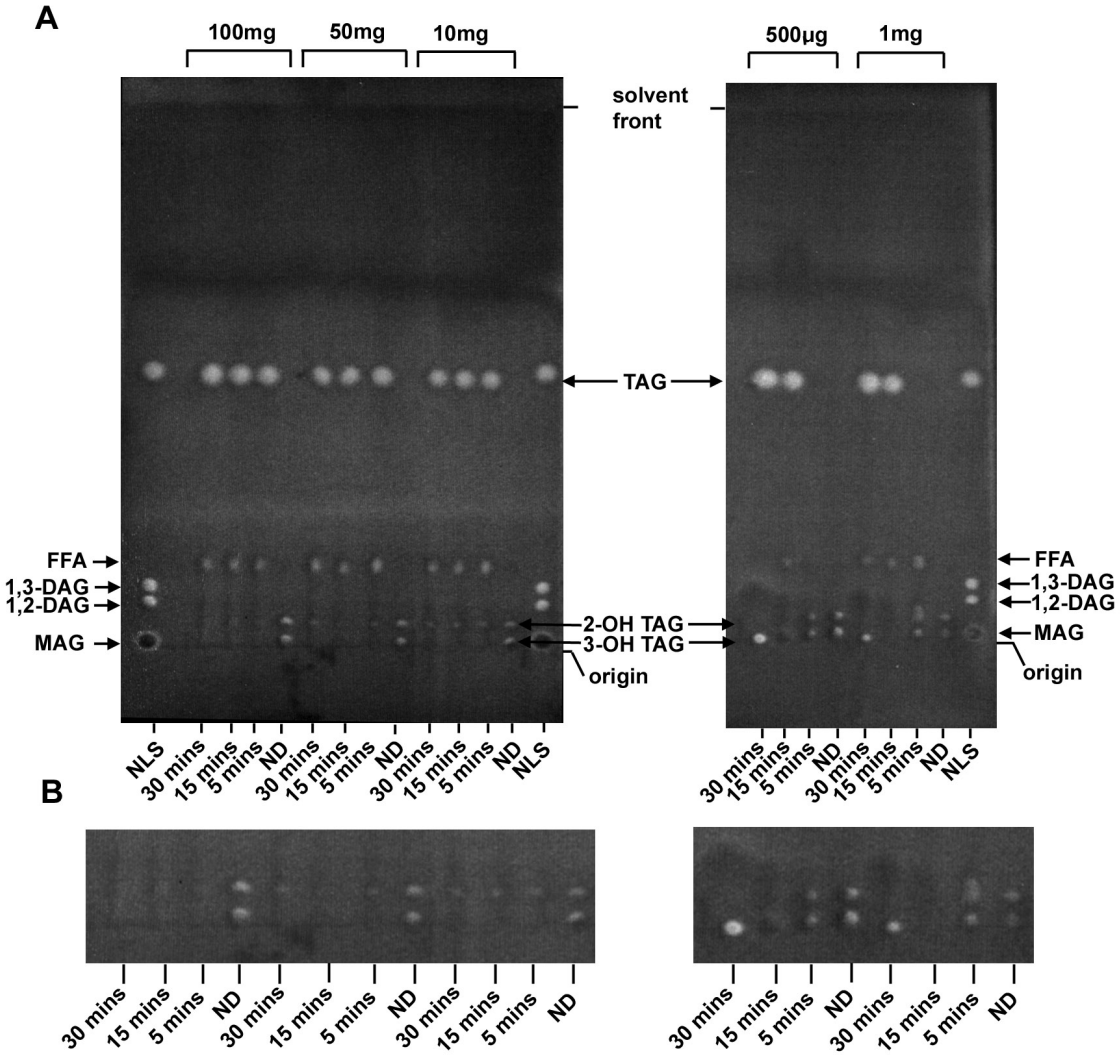


Figure 5.4: Evaluation of the effects of lipid concentration and reaction time on the derivatisation of castor oil and its migration on TLC.
500 μg - 100mg Castor oil samples were derivatised for either 5, 15 or 30 minutes at 70 °C. Following derivatisation reaction sample concentrations were equalised and evaluated by TLC. 500 μg and 1 mg quantities of castor oil were not effectively derivatised within 5 minutes, with clear spots aligning to the triricinolein (3-OH TAG) and diricinolein monoolein (2-OH TAG) positions of the underivatised sample (ND) and no spot aligning to the triolein (TAG) standard. Castor oil quantities $\geq$ 10 mg had derivatised even within 5 minutes, although residual 2-OH TAG is visible within some of the samples and required further derivatisation for its full conversion. **NLS=neutral lipid standard, ND = not derivatised, B = zoom-in of lower region of whole TLC plate (A, above) to show underivatised spots**.

through a nucleophilic attack, and reacts preferentially with water and alcohols. Thus to ensure the effectiveness of the silylating reagents it is critical to keep them moisture free. For this reason reagent bottles were stored in sealed containers containing desiccant. A combination of the silylating reagents *N,O* -Bis(trimethylsilyl)trifluoroacetamide (BSTFA) (Stalling et al., 1968) and trimethylchlorosilane (TMCS) was investigated, with the lipid sample solubilised in the aprotic (i.e. no available protons that would react with the silylating reagent) solvent pyridine. To investigate the extent of conversion of hydroxy-TAG molecules to their silylated equivalents, castor oil (containing ∼90% 3-OH TAG) was used. The effect of derivatisation reaction time on extent of derivatisation was investigated by terminating the reaction after either 5, 15 or 30 minutes at 70 °C. This was to ensure that full derivatisation of 3-OH TAG had occurred by 30 minutes in the castor oil. Also, a range of concentrations of castor oil were derivatised - either 500 μg, 1 mg, 10 mg, 50 mg or 100 mg, to investigate whether this influences extent of derivatisation. Following derivatisation, equal loadings of the derivatised castor oil samples were aliquoted on to Whatman Silica G TLC plates and separated in a hexane : diethyl ether : acetic acid (70:30:1) solvent system. The plate was then visualised with phloxine B. See Figure 5.4.

On evaluation of the TLC plates, it appears that derivatisation of hydroxy TAG compounds has not occurred effectively after 5 minutes of reaction time for the 500 μg and 1 mg samples (See Figure 5.4, right most plate). There is no visible non-hydroxy TAG spot for either of these 5 minutes samples, when compared to the TAG standard within the neutral lipid standard (NLS) marker. In the 500 μg and 1 mg samples, there are still two clear spots aligning to both the 3-OH TAG and 2-OH TAG positions of the underivatised castor oil sample (ND). A close-up of this region of the TLC plate is shown in section B of the figure. At concentrations of 10 mg, 50 mg and 100 mg, derivatisation has occurred by 5 minutes, with a strong TAG band aligning to the TAG standard in the NLS. There is not a significant difference in the size of TAG

band between 5, 15 and 30 minutes, and there is no 3-OH TAG specie visible in these samples either, indicating that the conversion of 3-OH TAG to its silylated equivalent has progressed to completion even after 5 minutes. Conversion of 2-OH TAG appears to occur more slowly than 3-OH TAG, and there is evidence of a minor amount of residual 2-OH TAG in the 10 mg and 50 mg samples. Thus in the *Y. lipolytica* lipid extracts fed on hydroxy fatty acid molecules, special attention must be paid to ensure no spots align to the 2-OH TAG position as this could indicate residual underivatised hydroxy fatty acid species that would result in an under-representation of hydroxy fatty acid in the TAG. However, derivatisation with TMCS and BSTFA in pyridine appears to be an effective way of altering migration of a range of concentrations of hydroxy TAG species on TLC plates, ensuring the reaction is carried out for at least 15 minutes and the TLC plate is critically analysed to ensure derivatisation has occurred effectively. As silylation reagents are highly moisture sensitive and their effectiveness can degrade, it is always important to check TLC plates before further analysis to ensure the derivatisation reaction has gone to completion.

5.3.6 Establishment of a lipid extraction procedure

The Folch procedure was employed to extract lipids from frozen *Y. lipolytica* cell pellets. Briefly, pellets were removed from -20 °C storage and thawed gently on ice. When thawed, cells were resuspended in 1:1 isopropanol and chloroform. At this initial stage triheptadecanoic acid was added to each sample undergoing extraction. Triheptadecanoic acid is a triacylglycerol molecule with three 17-carbon long fatty acid molecules esterified to the glycerol backbone. Heptadecanoic acid does not occur naturally thus the assumption can be made that its presence at the analysis stage is a result of its addition during the lipid extraction process. By adding the triheptadecanoic acid in a known amount it can serve as an internal standard allowing the quantification of other lipid species of unknown abundance against the known heptadecanoic acid.

Also, by adding triheptadecanoic acid at the most initial stage of extraction, any losses during the extraction procedure will effect the internal standard and *Y. lipolytica* lipid equally, and improve the accuracy of quantification.

Cells were disrupted with a Polytron probe in a 1:1 mixture of chilled chloroform : isopropanol on ice. Cell debris was separated from the solubilised contents by filtration and the filtrate dried down to a dry residue before re-suspension in 2 : 1 chloroform : methanol. It is on addition of a dilute salt solution (in this case, 0.88% KCl) that biphasic separation occurs; the mixture forming a lighter 'aqueous' upper phase (comprising chloroform : methanol : water in the proportions of 3 : 48 : 47) and a heavier 'organic' phase (chloroform : methanol : water in the proportions of 86 : 14 : 1). The hydrophobic property of lipids means they are readily soluble in the non-polar organic solvent.

The biphasic nature of the extraction method means that the aqueous upper phase is easily removed with a Pasteur pipette. This was discarded, before a second addition of 0.88% KCl to 20% total volume, phase separation and aqueous-phase removal. The remaining organic phase was then dried down under a stream of nitrogen at 50 °C to yield a total lipid sample. Finally this was re-suspended in a small volume of hexane which could either be stored at -20 °C or analysed.

To test the success of the extraction procedure on *Y. lipolytica* cells grown on either 2% glucose, 1% oleic acid and 1% methyl ricinoleate, 10 μl aliquots of total lipid extract were silylated (see Chapters 5.3.5, 2.4.4) and applied to Whatman Silica G TLC plates and separated in a hexane : diethyl ether : acetic acid (70:30:1) solvent system. Double replicate extractions were performed to evaluate reproducibility of the extraction procedure and extracts were made at different time points (2, 3, 5 and 10 hours of growth). Samples were run alongside a neutral lipid standard containing DAG and TAG molecules so the presence of these species could be identified in the *Y. lipolytica* extracts. When the solvent front reached the top quarter of the TLC plate,

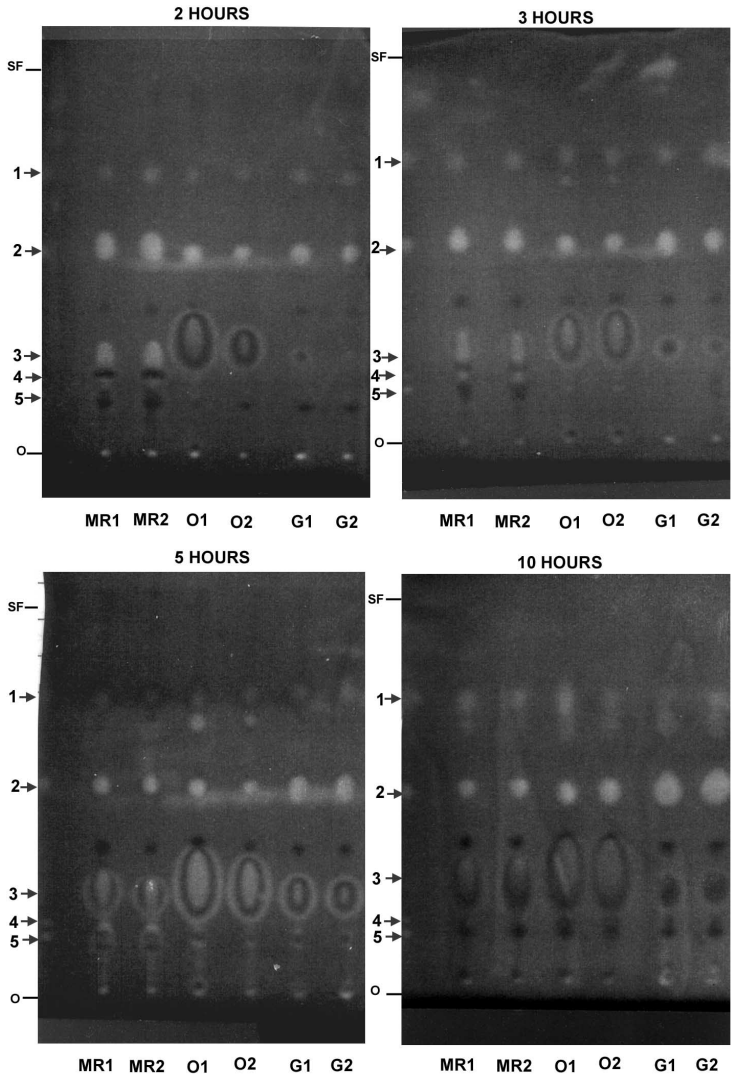


Figure 5.5: Evaluation of lipid extraction procedure by TLC. TLC separation of lipid extracted from *Y. lipolytica* P01G reveals species aligning to neutral lipid standards (NLS) DAG **(4 & 5)** and TAG **(2)**. Species aligning to free fatty acid **(3)** and steryl ester **(1)** standards were also identified. Replicate extracts were ostensibly broadly reproducible, although some variability in the abundance of free fatty acid in the lipid samples is present in some of the time points. O = origin, SF = solvent front. MR1+2, O1+2, G1+2 = two biological replicates of 1% methyl ricinoleate, 1% oleic acid and 2% glucose grown cells respectively. Solvent system = hexane:diethyl ether:acetic acid (70:30:1)

plates were removed from the solvent tank and the solvent allowed to evaporate. Plates were immersed briefly in phloxine B dye, and then visualised under UV light to identify the presence of lipid spots (see Figure 5.5).

The TLC plates revealed that neutral lipid was extracted from frozen *Y. lipolytica* cells grown on the different carbon substrates and at 2, 3, 5 and 10 hour time points. TAG molecules are present in each of the samples analysed, and their abundance is comparable within each replicate sample. Spots aligning to either one or both of the DAG isoforms in the standard are also present on the plates. There appears to be a greater abundance of DAG in the lipid extracts from cells grown on methyl ricinoleate than on glucose or oleic acid at 2, 3 and 5 hours of growth, but this difference is not noticable at 10 hours. Free fatty acid is also present within each of the lipid extracts, and is most abundant in the oleic acid fed samples. As *Y. lipolytica* is an oleaginous yeast the presence of a free fatty acid pool is expected. Its abundance in the oleic acid fed cells is also easy to rationalise as the cells were growing (and thus importing) free fatty acid. Spots aligning to a steryl ester standard were also present in each of the extracts analysed by TLC. There are spots visible at the sample origin point which will contain polar lipid molecules immobile in the hydrophobic solvent used (for example phosphatidylethanolamine and phosphatidylcholine).

The replicate extract lipid profiles are broadly reproducible. There is some variability in the abundance of the free fatty acid spots in the oleic acid extracts at 2 hours, but this decreases with subsequent time points and on the 10 hour TLC plate intensity of fatty acid spots are ostensibly the same for both oleic acid fed extracts. Furthermore, this apparent fatty acid variability was not observed in the methyl ricinoleate or glucose extracts; thus it was not deemed a significant indicator of an irreproducible preparation method.

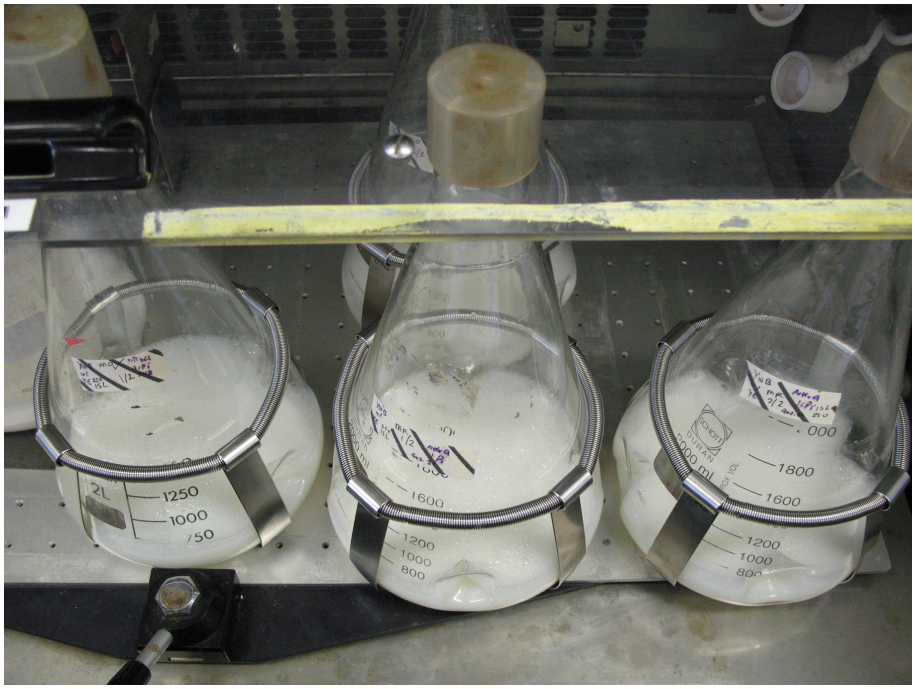


Figure 5.6: 2 L baffled conical flasks containing 500 ml of 1% lipid YNB media

5.3.7 Analysis into the effects of carbon source on the composition of storage oils in *Y. lipolytica* P01G

Y. lipolytica cells were grown in 2 litre baffled conical flasks in 500 ml YNB (minimal) media containing either 2% glucose or 1% fatty acid / fatty acid methyl ester and supplemented with 0.03% leucine to account for strain P01G's auxotrophic status (see Chapter 2.4.1, Figure 5.6). Fatty acids and FAMEs were added from a 10% stock emulsified in the presence of Tween 80 by sonication with a preparative probe sonicator. Cells were inoculated from a 2% glucose YNB parent culture that had reached exponential but not stationary phase, to form a new culture with a starting OD of 0.5. Prior to inoculation, cells were washed once with ddH_2O to remove any extrinsic carbon substrate from the cell surface. Cells were grown in the dark at 28 °C under rotation (140 rpm), harvested and washed. See Chapter 2.4 for yeast methods including cell harvest and washing procedure.

Lipid was extracted from gently thawed cells as described (Chapter 5.3.6) and derivatised for 30 minutes at 70 °C (Chapter 5.3.5) to convert any hydroxy lipid species to their silylated equivalents, which have the same migratory behaviour on TLC as non-hydroxy lipids. This was so the entire TAG pool could be analysed from a single spot on the TLC plate. Although the primary interest of an analysis of *Y. lipolytica* strains transformed with *R. communis* lipid biosynthesis genes is to quantify the amount of triricinolein production compared to an untransformed strain, it provides a simpler route to analysis and quantification as all the fatty acid components of the TAG are derived from a single spot and are compared to a single quantification standard (for example, triheptadecanoic acid).

Following derivatisation of lipid samples the neutral lipids were separated by preparative TLC and the TAG band identified by comparison with a neutral lipid standard and iodine staining. The volume of derivatised lipid sample loaded was adjusted to account for abundance observed by analytical TLC and contained between 50 μl and 140 μl of sample. Plates were developed in hexane : diethyl ether : acetic acid (70:30:1) solvent until the solvent front reached the top two centimetres of the TLC plate. The hexane : diethyl ether : acetic acid (70:30:1) solvent system was chosen as it provides good separation of TAG (including derivatised TAG) away from DAG, MAG and free fatty acid species.

TAG bands were identified on the TLC plate by iodine staining, which appear as contrasting yellow against the white background of the TLC plate allowing localisation of the region of interest. The TAG band was carefully scraped away from the glass backing of the TLC plate with a clean razor blade. The scratched plate could then be stained with Phloxine B and imaged on a transilluminator to check accuracy of TAG removal. Lipid was eluted from the scratched silica with 2:1 chloroform : methanol before being dried down and re-suspended in 1 ml 1% H_2SO_4 in CH_3OH. Samples were then kept on a heating block set to 50 °C overnight. This acid-catalysed methylation

step causes the transesterification of fatty acids and creation of fatty acid methyl esters, a volatile derivative suitable for GC analysis.

Following conversion of silylated triacylglycerols to fatty acid methyl esters (FAMEs) the next step was to investigate the fatty acid composition of the extracted TAG from each feeding type and time point. This was so the proportions and identities of different fatty acids within the TAG could be identified including potentially ricinoleic acid. During the most initial stage of lipid extraction the neutral lipid standard triheptadeconic acid (triacylglycerol of C17:0) was added at a known amount. By identifying the GC peak for the methyl ester of C17:0 the intensity was calibrated to that initial known amount. By assuming equal losses of lipid throughout the extraction, derivatisation and TLC fractionation process, and by assuming equal conversion of TAG to methyl esters, all identified peaks within the spectra had their areas compared to the C17:0 peak area. This allowed an estimation of actual amounts of different fatty acids within the TAG in the different feeding types and time points.

A Shimadzu GC-14 with an EC-Wax column was initially calibrated with a commercially available range of fatty acid standards - GC96 (Nucheck-Prep. MN, USA) and a methyl ricinoleate standard. Each fatty acid gave a peak on the GC at a particular retention time.

Once the standard data has been collected the GC spectra of the FAMEs from the feeding experiment samples were collected. The retention time, peak intensity and percentage proportion of each peak area compared to the total peak area was recorded for every run. Identities were assigned by comparing retention times to the standards' retention times.

5.3.7.1 Glucose and oleic acid

GC analysis of the TAG fraction of glucose fed *Y. lipolytica* identified C16:0 (palmitic acid), C16:1 (palmitoleic acid), C18:0 (stearic acid), C18:1 (oleic acid) and C18:2

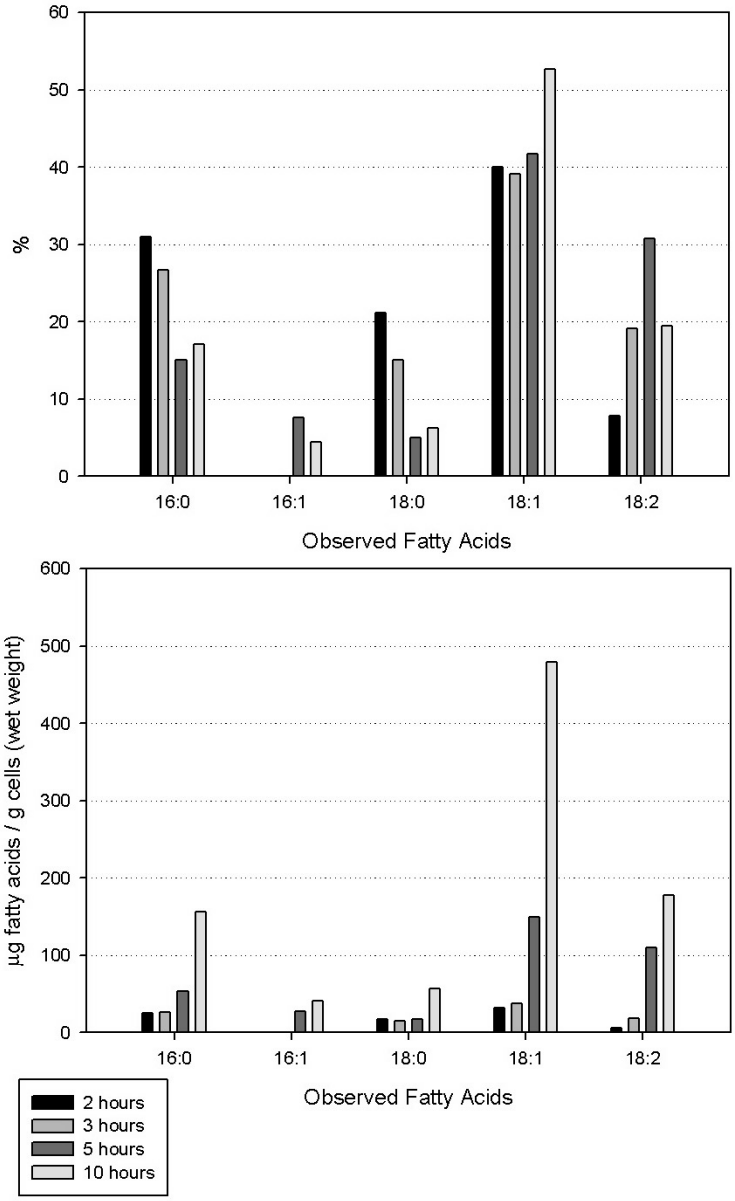


Figure 5.7: Percentage and μg abundance (per g cells, wet weight) of observed fatty acids in the TAG of 2% glucose fed *Y. lipolytica* . A general pattern of μg abundance increase is observed for all fatty acids with a significant increase in oleic acid abundance between 5 and 10 hours. This equates to a generally increasing pattern in the percentage proportion of oleic acid between 2 and 5 hours and a concomitant decrease in the percentage of palmitic and stearic acid between 2 and 5 hours.

Table 5.1: GC analysis data of TAG from *Y. lipolytica* fed on either 2% glucose or 1% oleic acid

Time	16:0		16:1		18:0		18:1		18:2		Total
point	%	μg	%	μg	%	μg	%	μg	%	μg	(μg)
Glucose											
2h	31.0	24.8	-	-	21.2	16.9	40.0	31.9	7.8	6.3	79.9
3h	26.7	26.0	-	-	15.1	14.7	39.1	38.1	19.1	18.7	97.5
5h	15.0	53.6	7.6	27.0	5.0	17.7	41.7	149.2	30.7	110.0	357.5
10h	17.1	156.0	4.5	40.8	6.3	57.2	52.7	479.7	19.4	177.1	910.8
Oleic Acid											
2h	22.7	9.3	-	-	14.8	6.1	62.5	25.7	-	-	41.1
3h	31.6	8.2	-	-	19.4	5.1	49.0	12.7	-	-	26.0
5h	28.8	3.1	-	-	24.9	2.7	46.3	5.0	-	-	10.8
10h	8.0	20.1	3.5	8.6	2.2	5.6	80.9	202.6	5.4	13.5	250.4

(linoleic acid) as the TAG fatty acid components (Table 5.1). All fatty acid components were present at each growth stage examined except C16:1 which was only present in 5 and 10 hour samples. With the exception of C18:0, a pattern of increasing abundance for each constituent fatty acid is observed with increasing growth time; the most significant increase being for oleic acid which increased from 149 μg / g to 480 μg / g cells (wet weight) at 10 hours. The second most dominant fatty acid component is linoleic acid, comprising 177 μg / g (wet weight) cells in the TAG fraction at 10 hours of growth. A graph depicting the μg quantities of observed fatty acids displays the pattern of increase for each fatty acid component (with the exception of C18:0) in the TAG (see Figure 5.7, lower). The percentage proportions of the observed fatty acid components were also calculated, and their values are described in Table 5.1 and in Figure 5.7 (upper). These values indicate that oleic acid is the dominant constituent of the TAG at each time point analysed, and that its proportion increases between 5 and 10 hours of growth to 52.7% of the TAG.

The analysis of TAG purified from lipid extracts of oleic acid grown *Y. lipolytica*

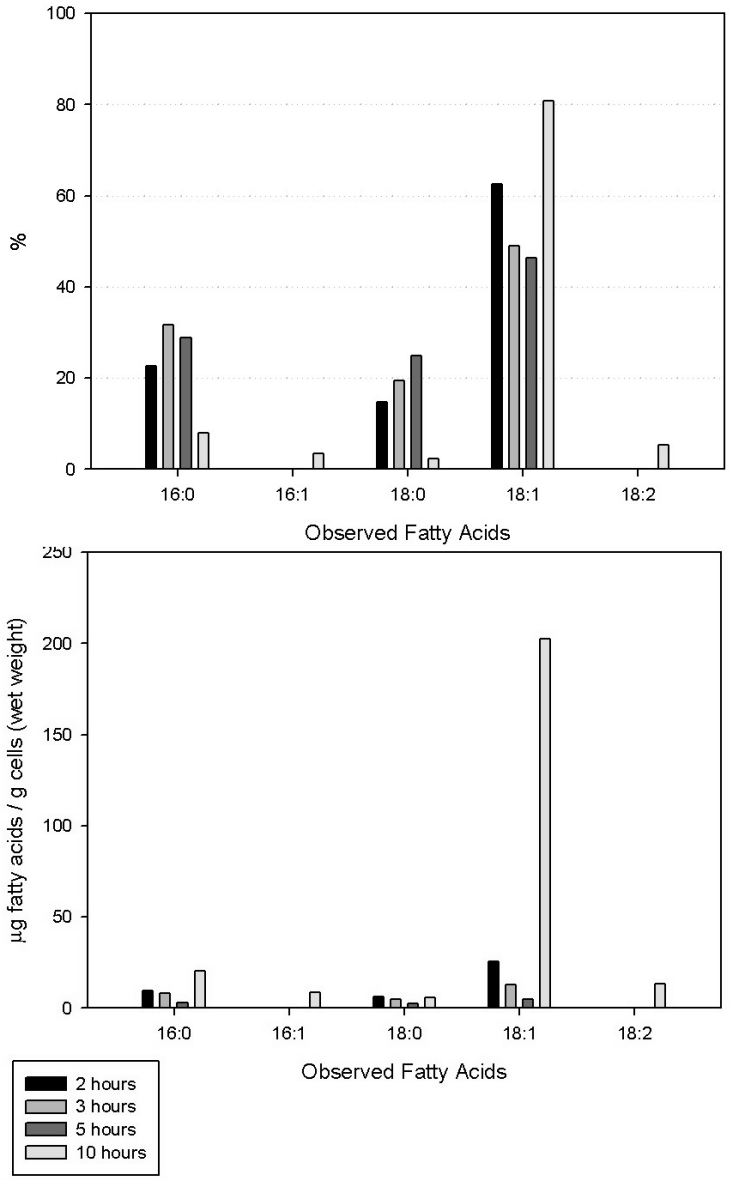


Figure 5.8: Percentage and μg abundance (per g cells, wet weight) of fatty acid observed in the TAG of *Y. lipolytica* P01G fed on 1% oleic acid (emulsified in 0.1% Tween 80). Lipid extracts were made from cells harvested at 2, 3, 5 and 10 hours of growth in oleic acid containing minimal media. The dominant fatty acid species in the TAG at each time point analysed was identified as oleic acid, although it was characterised by a very significant increase in abundance between 5 and 10 hours of growth. Linoleic acid was only identified at 10 hours of growth, indicating a delayed desaturation of oleic acid and a very different pattern of lipid accumulation compared to cells fed on glucose.

Table 5.2: GC analysis data of TAG extracted from *Y. lipolytica* cells fed on either 1% methyl ricinoleate or 1% methyl oleate. Highly abundant oleic acid or ricinoleic acid peaks were observed in the TAG from methyl oleate and methyl ricinoleate fed cultures respectively.

Time	16:0		16:1		18:0		18:1		18:2		18:1-OH		Total
point	%	μg	%	μg	%	μg	%	μg	%	μg	%	μg	(μg)
Methyl Ricinoleate													
2h	1.4	34.9	-	-	1.5	38.8	3.9	100.3	4.1	106.2	89.1	2287.4	2567.6
3h	2.9	12.6	-	-	2.6	11.2	5.0	21.5	5.5	23.3	84.0	360.1	428.7
5h	8.1	3.0	-	-	7.7	2.9	7.9	3.0	-	-	76.3	28.7	37.6
10h	14.5	7.7	5.5	2.9	6.4	3.4	67.7	35.5	5.9	3.1	-	-	52.6
Methyl Oleate													
2 h	1.0	25.6	-	-	-	-	96.0	2369.9	3.0	72.9	-	-	2468.4
3 h	-	-	-	-	-	-	78.0	141.1	22.0	39.8	-	-	180.9
5 h	-	-	-	-	-	-	87.1	141.4	12.9	20.9	-	-	162.3

P01G cells identified oleic acid as the dominant component of this fraction at each time point analysed. The pattern of TAG abundance was different from that of glucose fed cells, where an increase in TAG quantity with increasing growth time was identified. With oleic acid fed cells the total amount of TAG decreased from 41.1 μg to 10.8 μg / g cells (wet weight) between 2 and 5 hours. Between 5 and 10 hours TAG accumulated significantly; reaching its maximum quantity of 250.4 μg / g cells (wet weight). The maximum quantity of oleic acid produced in the TAG was at 10 hours, where it represented 202.6 μg / g cells (wet weight), equivalent to 80.9% of the TAG. Other fatty acid components identified followed the same pattern of decreasing abundance between 2 and 5 hours, followed by an increase at 10 hours. Stearic acid drops from just under 25% of TAG to 2.2% between 5 and 10 hours, which coincides with an increase in the percentage abundance of oleic acid.

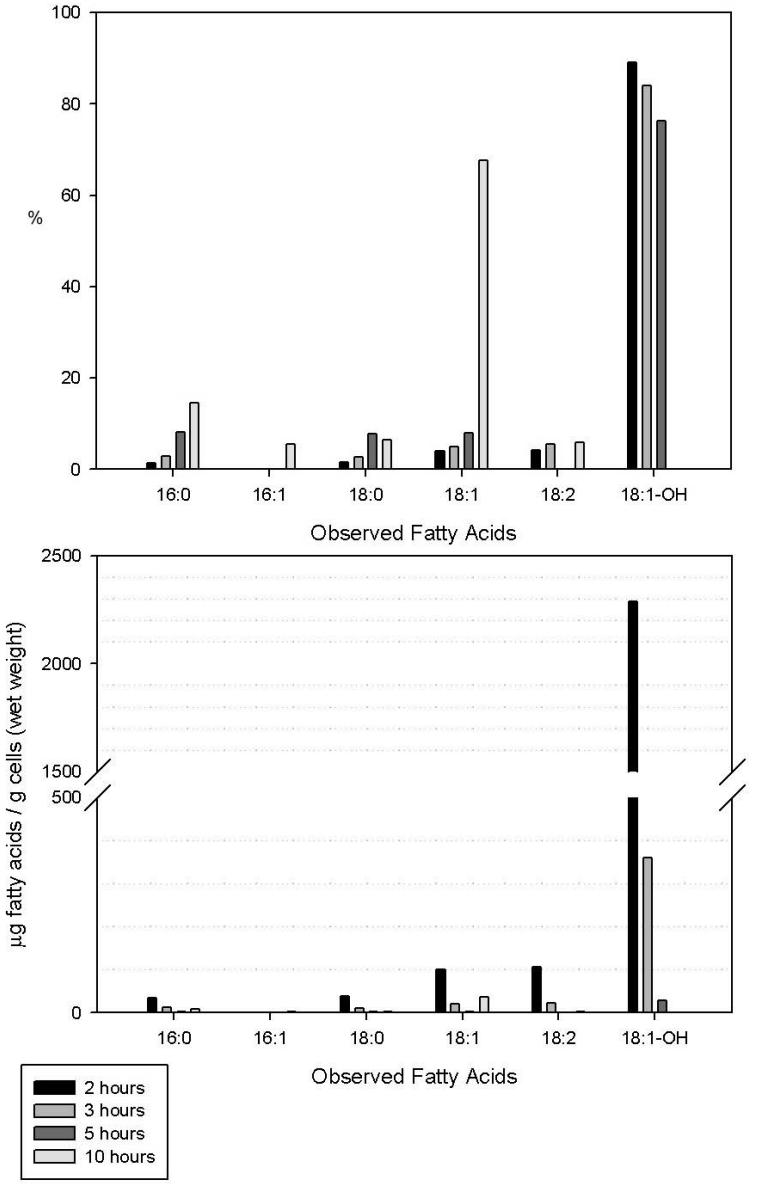


Figure 5.9: Percentage and μg values of observed fatty acids in the TAG of 1% methyl ricinoleate-fed *Y. lipolytica* . A very significant peak of with the same retention time as the C18:1-OH standard was identified at 2 hours of growth, which represented 89% of the total observed fatty acids at the TAG position of the TLC plate. This was efficiently reduced between 3 and 5 hours, and was not observed at 10 hours of growth. Other fatty acids observed at the TAG position were palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1) and linoleic acid (C18:2). Oleic acid was the dominant fatty acid at 10 hours, although its total abundance was less than at 2 hours of growth.

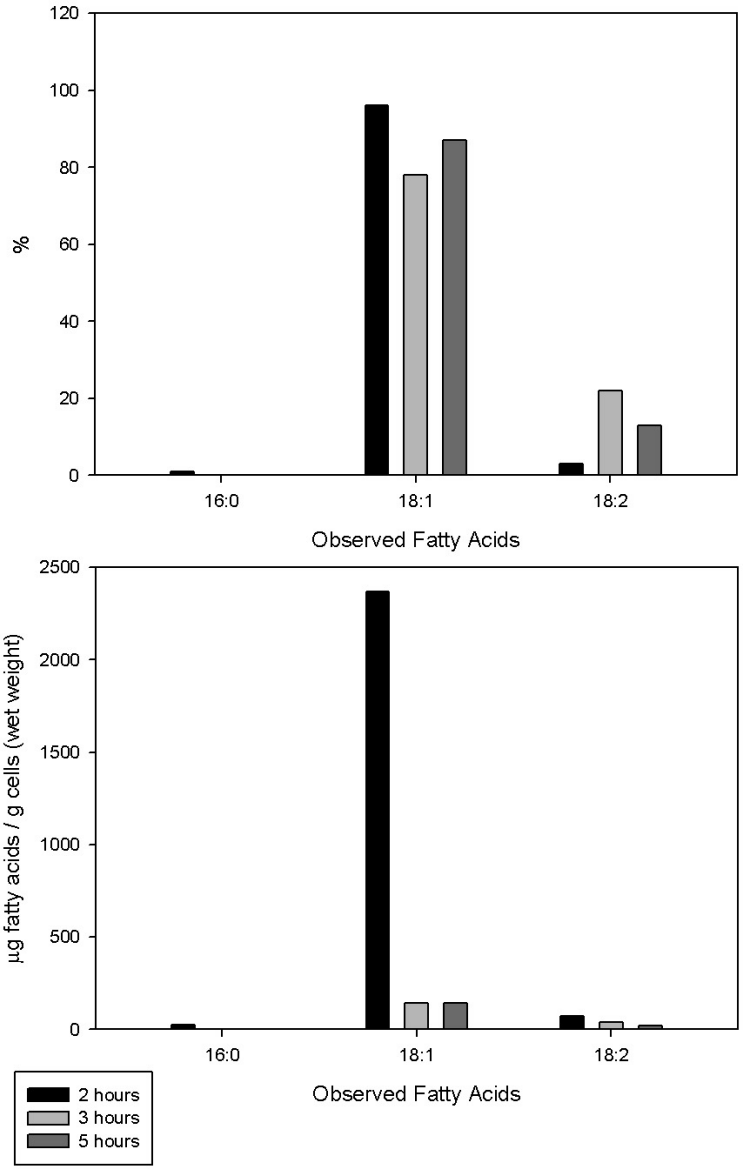


Figure 5.10: Percentage and μg values of observed fatty acids in the TAG of 1% methyl oleate-fed *Y. lipolytica* . A simple profile of fatty acids were observed at the TAG TLC position of methyl oleate-fed *Y. lipolytica* lipid extracts. Oleic acid was identified as the most abundant peak, where it represented 96% of the TAG with an abundance of 2369.9 μg/g cells (wet weight) at 2 hours. The amount of observed oleic acid was significantly reduced between 2 and 3 hours of growth.

5.3.7.2 FAMEs of ricinoleic and oleic acid

For cells growing on 1% FAME as the carbon source, the TAG profile is characterised by a significant peak on the GC of the fed lipid molecule (i.e. either methyl ricinoleate or methyl oleate), followed by its rapid reduction in abundance after 2 hours. For example a significant peak on the GC, aligning to the methyl ricinoleate standard, was found to comprise 89% of the TAG fraction with 2287.4 μg / g cells (wet weight) at 2 hours of growth. By 3 hours, this peak had reduced to 360.1 μg / g cells (wet weight), by 5 hours 28.7 μg / g cells (wet weight) and by 10 hours it was not identified on the GC trace. The percentage proportion of the peak aligning with the methyl ricinoleate standard within the TLC's TAG band was $\sim$ 80% at each time point in which it was identified, indicating no specific catabolism of hydroxy lipids present at the TAG position of the TLC plate. Other components identified include palmitic (C16:0), palmitoleic (C16:1), stearic (C18:0), oleic (C18:1) and linoleic (C18:2) acid. By 10 hours of growth, oleic acid was identified as the dominant fatty acid component of the TAG, where it comprised 67.7% of the observed fatty acids. However, although its abundance was increased from 3 and 5 hour time points, it was just over a third of that measured at 2 hours of growth.

A very similar quantity of oleic acid to that observed for ricinoleic acid was identified in the TAG band of *Y. lipolytica* cells fed on 1% methyl oleate at 2 hours (2369.9 μg compared to 2287.4 μg / g cells). Under these feeding conditions oleic acid represented 96% of the TAG band. Between 2 and 3 hours of growth a very significant decrease in the total amount of oleic acid observed, and thus the total amount of lipid analysed at the TAG band position of the TLC plate. Oleic and linoleic acid were the only fatty acids observed in the TAG profile at the 3 and 5 hour time points analysed.

The constituent fatty acid profiles of TAG from *Y. lipolytica* cells growing on oleic acid (Table 5.1) and methyl oleate (Table 5.9) are signficantly different, with a far greater abundance of an oleic acid-aligning peak present in the TAG fraction of methyl oleate fed cells at 2 hours of growth. Both analyses then show a pattern of breakdown

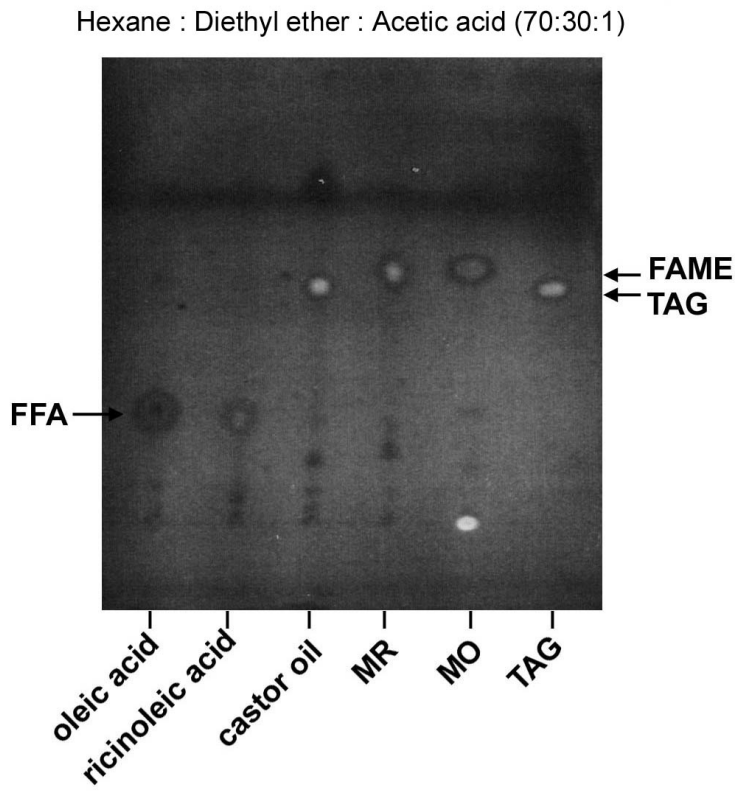


Figure 5.11: Similar migrations of TAG and FAME in the hexane : diethyl ether : acetic acid (70:30:1) solvent system. FAME molecules (i.e. derivatised methyl ricinoleate (**MR**) and the FAME of oleic acid (**MO**)) were found to have a similar migration to **TAG** in this solvent system, and thus are free to contaminate the TAG analysis should they be carried over from harvesting into lipid extraction or be present in the lipid extraction due to internalised FAME pools.

of lipid until the 5 hour time point.

Evidence emerged that the presence of abundant C18:1-OH and C18:1 in the TAG at 2 hours of growth may be an experimental artefact. This is explored in the following section.

5.3.8 Co-migration of methyl fatty acids and triacylglycerols on TLC with the hexane : diethyl ether : acetic acid (70:30:1) TLC solvent system

The evidence that cells growing on FAMEs preferentially incorporate fatty acids into their TAG stores over those growing on non-methylated fatty acid raised the possibility that FAMEs have a different fate to non-methylated fatty acids in *Y. lipolytica* cells. For example, the rate of substrate utilisation may be different between the two molecules, perhaps related to differences in polarity conferred by the different head groups. However, evidence emerged that it may be an experimental artefact: FAME molecules (FAMEs without hydroxyl groups retarding their migration, e.g. methyl oleate or silylated methyl ricinoleate) were found to have a similar migratory pattern to TAG molecules (TAGs without hydroxyl groups retarding their migration, e.g. triolein or silylated triricinolein) in the hexane:diethyl ether:acetic acid (70:30:1) solvent system (see Figure 5.11). The TLC plate compares the migration of FAME (fourth and fifth lanes from left, derivatised methyl ricinoleate (MR) and methyl oleate (MO)) to TAG (third and sixth lanes from left, derivatised castor oil and TAG standard). Although the migrations of FAME and TAG are different, they are close enough that they may merge, especially with preparative loadings of lipid. Therefore, any FAMEs within the lipid extract may contaminate and thus contribute to the observed lipids in the analysed TAG band, and would be indistinguishable from the fatty acid components of the TAG. As the lipid samples were derivatised, methyl ricinoleate is also a potential contaminant of the TAG band in methyl ricinoleate fed *Y. lipolytica* samples. Cells were routinely washed with multiple BSA and BSA/NaCl solutions previously reported to remove extrinsic media lipid from the surface of *Y. lipolytica* cells growing on oleic acid (Mlícková et al., 2004). Despite this, any extrinsic lipid which may remain on the surface of the cells would be carried through to the GC analysis. Even if no extrinsic FAMEs remained, it would be impossible to differentiate between lipid species

from TAG molecules and lipid species of internalised FAME pools. Because of this, alternative solvent systems were evaluated.

5.3.9 Evaluation of alternative TLC suitable for separation of triacylglycerol and fatty acid methyl ester molecules

5.3.9.1 Petroleum ether : diethyl ether : acetic acid (90:10:1)

New solvent systems capable of separating FAMEs from TAG were evaluated, as hexane : diethyl ether : acetic acid (H:D:A) 70:30:1 was found to be unsuitable in this regard. An examination of the literature identified petroleum ether : diethyl ether : acetic acid (PE:D:A) in the proportions 90:10:1 (Hamilton and Hamilton, 1992) as a potential alternative, giving significant separation of FAMEs (retention factor (R_f) = 0.65) and TAG (R_f = 0.35) with no other lipid components reported as migrating close to the TAG position. The suitability of the system with our TLC plates and solvents was assessed through the separation of a range of derivatised and underivatised lipid standards (see Figure 5.12). Non-hydroxy FAME molecules (i.e. methyl oleate) were found to have significantly greater R_f values than TAG (0.44 versus 0.25 respectively in our laboratory), confirming their reported separation in the literature. The highly apolar nature of the solvent system meant that polar lipid species or those containing polar functional groups such as triricinolein (present in the castor oil standard) and underivatised methyl ricinoleate remained at or near the origin. Derivatisation altered the migratory behaviour of those lipid species containing hydroxy functional groups, although 'trailing' was observed for each of the derivatised species, indicating some remaining interaction between the derivatised species and the polar TLC substrate. The derivatised methyl ricinoleate (labelled MR d on Figure 5.12) presented significant trailing in a uniform pattern of spots stretching back to the origin. The experiment was repeated with new batches of methyl ricinoleate and castor oil to discount degradation of the lipid stocks, but the pattern of trailing spots was

still present in the derivatised samples (result not shown). Such incomplete migration of derivatised hydroxy lipid on the TLC plate could pose a potential contaminant risk due to derivatised methyl ricinoleate trailing through the positions of other lipid species including TAG. Alternative TLC systems were investigated.

5.3.9.2 Hexane : diethyl ether : acetic acid (90:10:1)

To evaluate whether the derivatised methyl ricinoleate trailing pattern was peculiar to the petroleum ether system, an equivalently apolar solvent system of hexane : diethyl ether : acetic acid (90:10:1) was investigated. This gave a comparable pattern of lipid distribution and as with petroleum ether the derivatised methyl ricinoleate lipid suffered from trailing (Figure 5.13).

5.3.9.3 Hexane : diethyl ether : acetic acid (80:20:1)

Increasing the polarity of the solvent solution by increasing the ratio of diethyl ether to hexane (H:D:A 80:20:1) gave a TLC profile with good separation between the FAME and TAG positions. The derivatised methyl ricinoleate (MRD) spot still suffered from the trailing effect but its extent was less severe than in less polar solvents, with the minor trailing components having an increased migration and 'bunching up' behind the parent MR spot (Figure 5.14). However, the trailing effect was severe enough to merge with the TAG position and as such this solvent system was rejected.

5.3.9.4 Hexane : diethyl ether : acetic acid (75:25:1)

Similarly, H:D:A (75:25:1) suffered the same problem of derivatised MR trailing and had the additional problem of near-localisation of TAG and FAME bands. See Figure 5.15.

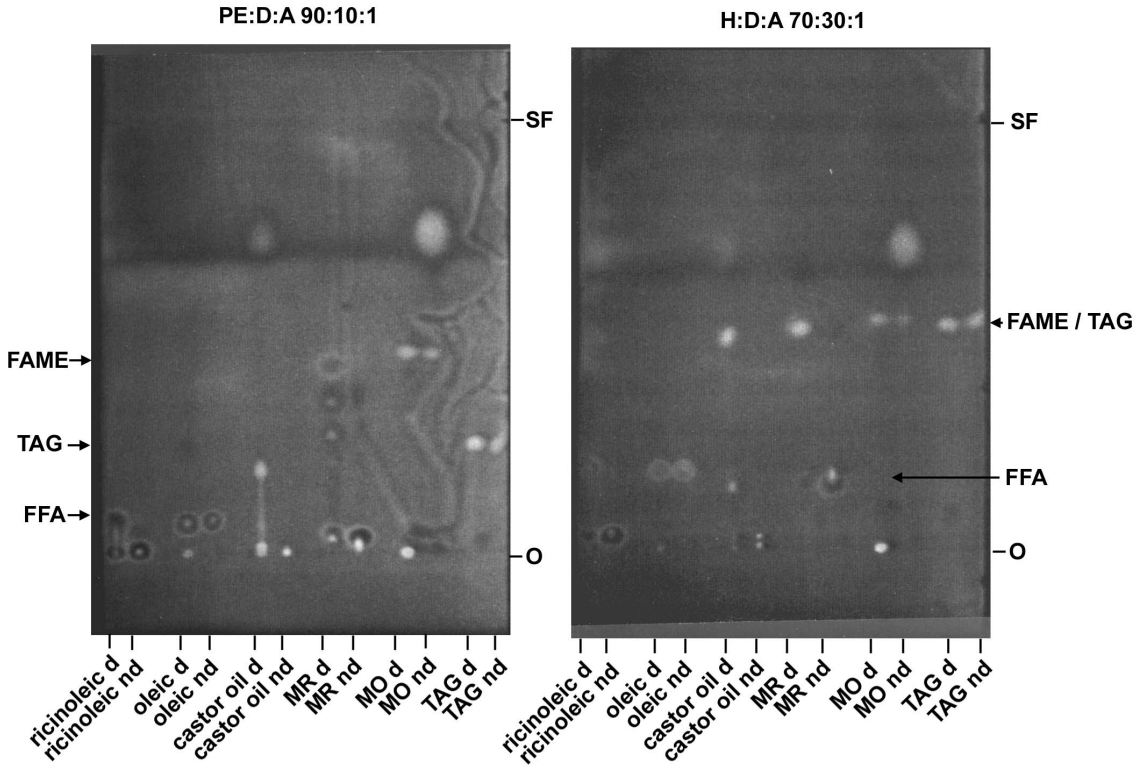


Figure 5.12: Comparison of the petroleum ether : diethyl ether : acetic acid (90:10:1) solvent system with hexane : diethyl ether : acetic acid (70:30:1) solvent system reveals increased separation of FAME and TAG lipid species. However, increased interaction between hydroxy lipids ricinoleic acid and methyl ricinoleate including their silyl derivatives in the highly apolar PE:D:A 90:10:1 solvent system causes a trailing effect which may present a contamination risk. This was most pronounced for the derivatised methyl ricinoleate sample (**MR d**). MO = methyl oleate, FFA = free fatty acid, TAG = triacylglycerol, FAME = fatty acid methyl ester, SF = solvent front, o = origin ; d and nd suffixes = derivatised and non-derivatised respectively.

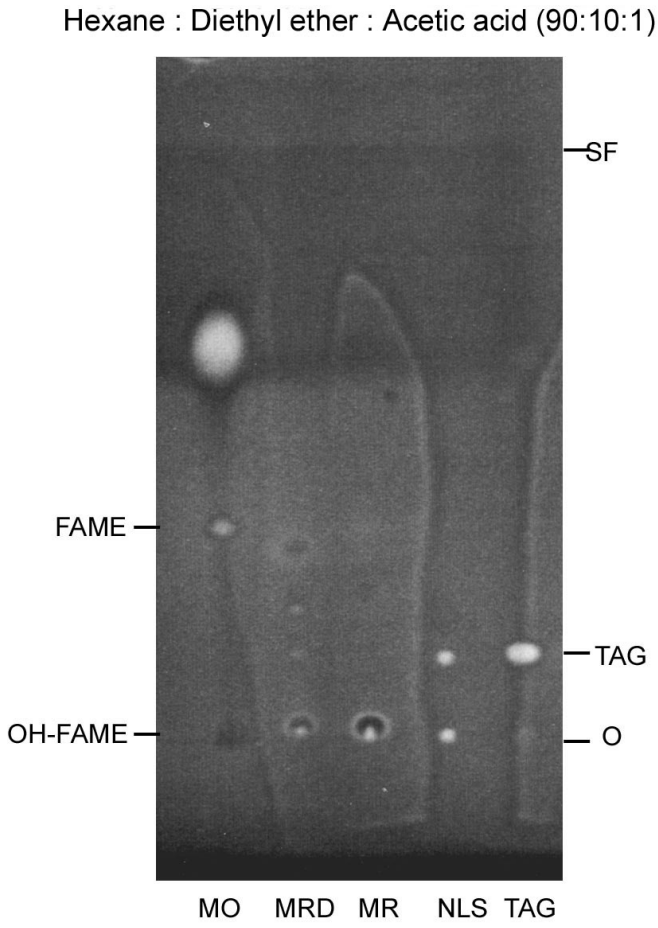


Figure 5.13: Effect of solvent ratios on lipid migration: Hexane : Diethyl Ether : Acetic Acid (90:10:1).With this solvent system good separation is obtained for FAME and TAG molecules; trailing is still observed from the derivatised methyl ricinoleate standard which may cause contamination of the TAG position. **SF = solvent front, TAG = position of triolein migration on TLC / TAG standard, O = origin, NLS = neutral lipid standard (containing MAG, DAG, TAG standards), MR = methyl ricinoleate, MRD = derivatised methyl ricinoleate, MO = methyl oleate, OH-FAME = position of methyl ricinoleate migration on TLC, FAME = position of non-hydroxy FAME migration on TLC.**

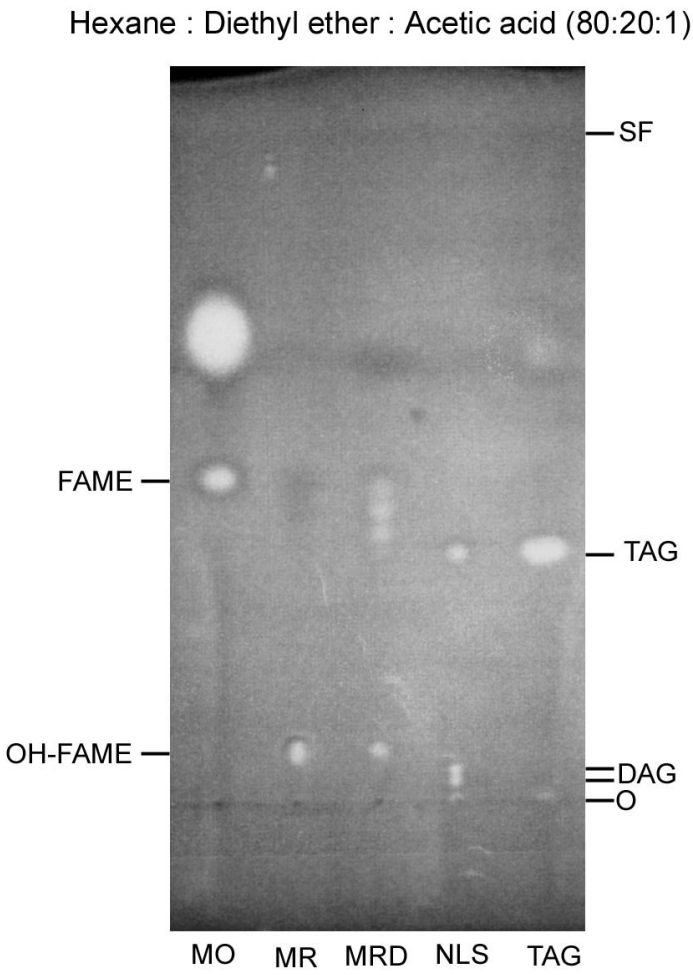


Figure 5.14: Effect of solvent ratios on lipid migration: Hexane : diethyl ether : acetic acid (80:20:1). With this solvent system although there's good separation of TAG and OH-FAME positions, trailing of the derivatised methyl ricinoleate standard contaminates the TAG position. **SF = solvent front, TAG = position of triolein migration on TLC / TAG standard, O = origin, NLS = neutral lipid standard (containing MAG, DAG, TAG standards), MR = methyl ricinoleate, MRD = derivatised methyl ricinoleate, MO = methyl oleate, OH-FAME = position of methyl ricinoleate migration on TLC, FAME = position of non-hydroxy FAME migration on TLC.**

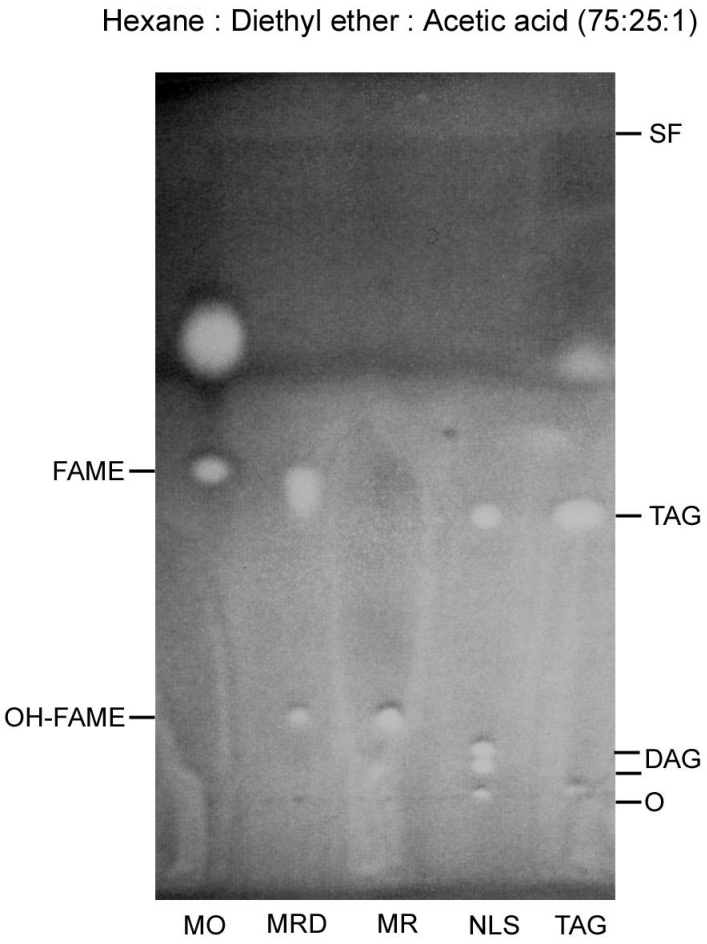


Figure 5.15: Effect of solvent ratios on lipid migration: Hexane : diethyl ether : acetic Acid (75:25:1). With this solvent system, migrations of TAG and derivatised methyl ricinoleate (OH-FAME) are too similar to allow confident separation and analysis. **SF = solvent front, TAG = position of triolein migration on TLC / TAG standard, O = origin, NLS = neutral lipid standard (containing MAG, DAG, TAG standards), MR = methyl ricinoleate, MRD = derivatised methyl ricinoleate, MO = methyl oleate, OH-FAME = position of methyl ricinoleate migration on TLC, FAME = position of non-hydroxy FAME migration on TLC.**

5.3.9.5 Hexane : diethyl ether : acetic acid (50:50:1)

Increasing the polarity of the solvent with a H:D:A (50:50:1) system resulted in a pattern of increased migration of all lipid species, whether polar or apolar, but with little alteration to the relative migration of TAG and FAME so again is unsuitable. See Figure 5.16.

5.3.9.6 Evidence of TMCS-ester degradation when silylated samples are separated by TLC

Altering the ratios of the apolar and polar solvents (hexane and diethyl ether respectively) altered the relative migration of the analysed lipid standards. In more apolar solvents, a trailing effect was observed for the derivatised methyl ricinoleate sample (**MRD**) which is a potential contaminant risk through the TLC plate. This could be reduced or nullified by increasing the polarity of the solvent, and was not observed in either H:D:A (70:30:1) or (50:50:1). However, in these solvent systems the migration of FAME and TAG was very similar and thus could not be used.

The presence of a trail of spots behind the derivatised MR position, revealed in the evaluation of new TLC systems, was a new concern. The trailing effect was not noted with hexane : diethyl ether : acetic (70:30:1) and this appears to be because with this solvent system the migrations of the trailing components are the same as the parent derivatised MR species, and thus they form a single spot. The fact that there was no evidence of trailing observed in the non-derivatised MR standard indicates it may be an artefact of the derivatisation procedure, with the derivatised FAME degrading either before or during TLC separation. It has previously been reported than the silyl esters formed by the silylating reagent trimethylchlorosilane (TMCS) are unstable and will hydrolyse slowly on TLC adsorbents (Christie, 1982). This was confirmed in a personal communication with Dr. W. Christie (July, 2007), who stated 'TMCS esters are notoriously unstable and hydrolyse rapidly unless they are stored in the silylating

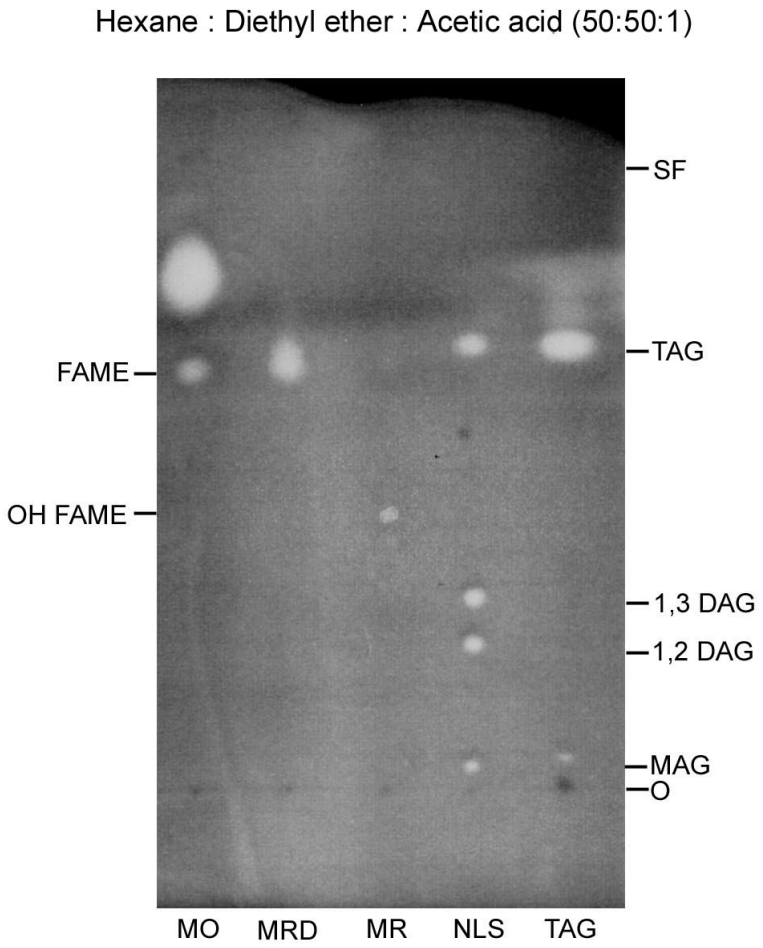


Figure 5.16: Effect of solvent ratios on lipid migration: Hexane : diethyl ether : acetic acid (50:50:1). The increased polarity of this solvent has significantly increased the migration of the analysed lipid samples, however the relative migrations of TAG and derivatised methyl ricinoleate (OH-FAME) make it unsuitable. **SF = solvent front, TAG = position of triolein migration on TLC / TAG standard, O = origin, NLS = neutral lipid standard (containing MAG, DAG, TAG standards), MR = methyl ricinoleate, MRD = derivatised methyl ricinoleate, MO = methyl oleate, OH-FAME = position of methyl ricinoleate migration on TLC, FAME = position of non-hydroxy FAME migration on TLC.**

solution. They cannot reliably be subjected to TLC or any other form of adsorption chromatography'.

If derivatised lipid species are degrading prior to their extraction from the TLC plate, it would likely lead to quantitative inaccuracies due to a selective under-representation of the silylated lipid classes. Degradation products may also cause contamination of other lipid species on the TLC plate which is highly undesirable. Without changing the main lipid separation technology, this presented two options: (1) to find an alternative derivatisation procedure which does not suffer from instability on TLC, or (2) to analyse the underivatised lipid extracts, requiring a solvent system and analysis procedure which effectively separates underivatised hydroxy fatty acids and allows the accurate quantification of hydroxy TAG components.

Alternative hydroxyl group derivatisation systems exist. Trifluoroacetate specifically derivatises lipid compounds with free hydroxyl groups (including monoacylglycerol), and are sufficiently temperature stable to be subjected to GC analysis (Christie, 1982). However, even when stored in inert solvents such as hexane they are reported to hydrolyse rapidly, and can cause selective losses of unsaturated lipid molecules (Wood and Snyder, 1966). Acetylation is an alternative procedure through which the polar hydroxyl group can be converted to a non-polar derivative to improve its chromatographic qualities. A disadvantage of this method is the overnight reaction time required for acetylation (Renkonen, 1966) which combined with the lipid extraction procedure and preferred overnight methylation step would result in undesirable extended sample handling times.

TLC-based separation and quantification of underivatised lipid extracts containing hydroxy lipids has been reported previously (Smith et al., 2003). The authors reported the use of a 70:140:3 hexane : diethyl ether : acetic acid solvent system in the separation of a range of hydroxy and non-hydroxy lipid species. The use of this TLC solvent system in our laboratory was investigated.

5.3.10 Development of alternative TLC solvent systems for the separation of non-derivatised total lipid extractions from methyl fatty acid fed *Y. lipolytica*

It has been reported that not a single solvent system will completely separate all naturally occurring neutral lipid classes in a single development (Henderson and Tocher, 1992). Where a range of hydroxy lipid species are present, each with their own R_f values, in addition to the normal range of non-hydroxy neutral lipid species, their distinct separation becomes a significant challenge.

However, TLC separations of the underivatised seed oil from *A. thaliana* transformants expressing *R. communis* oleate $\triangle$12 hydroxylase and producing a range of hydroxy lipids has previously been reported (Smith et al., 2003). This group used two separate solvent systems and silica G60 TLC plates. The first of these was hexane : diethyl ether : acetic acid (70:30:1), previously used in this report for the separation of derivatised lipid and found to be unsuitable due to the co-migration of FAME and TAG species. A second analysis used hexane : diethyl ether : acetic acid (70:140:3), which was reported to give clean separations of MAG, OH-MAG, 1-OH DAG, 2-OH DAG , 1-OH TAG, 2-OH TAG, 3-OH TAG, OH-FFA, MAG, DAG, TAG and free fatty acid (FFA). No information was given for the localisation of FAME or OH-FAME as this was not relevant to their investigation.

To investigate whether the same separation of hydroxy and non-hydroxy lipid species could be achieved in our laboratory, and to localise the migration of FAME and OH-FAME with this system, the experiment was repeated with a range of lipid standards. It is known that purity of solvents, TLC plate manufacturer, TLC plate batch, degree of hydration of TLC adsorbant particles and atmospheric pressure can all effect the migration of lipid species, so it is important to assess exactly how these combined factors influence migration within a particular laboratory. The TLC tank system was set up as described previously (Chapter 2.4.5.1) with hexane : diethyl ether

: acetic acid (70:140:3) as the solvent. Triricinolein (3-OH TAG, RRR), diricinolein monoolein (2-OH TAG, RRO), monoricinolein diolein (1-OH TAG, ROO), ricinoleic acid (OH-FFA, R), oleic acid (FFA, O), methyl ricinoleate (OH-FAME, MR), methyl oleate (FAME, MO) and castor oil were used as lipid standards alongside a neutral lipid standard containing MAG, 1,2-DAG, 1,3-DAG and TAG. Aliquots of equal loading were spotted on to a TLC plate and separated in the TLC tank until the solvent front neared the top of the plate. The plate was then removed and the solvent front position marked. The TLC plate was allowed to air dry before visualising with the fluorescent lipid-binding stain phloxine B (see Figure 5.17).

In this comparatively polar solvent system, polar lipid species such as triricinolein and methyl ricinoleate have significantly enhanced R_f values compared to, for example, H:D:A (70:30:1). A good separation of 3-OH TAG, 2-OH TAG and 1-OH TAG was achieved. Non hydroxy FAME and TAG are separated with this TLC system although their migration is similar so care would need to be taken with high loading samples so that their positions do not overlap. One of the non-hydroxy DAG isoforms (1,3-DAG) has a very similar R_f value to OH-FAME. However a more significant problem with regard to an evaluation of triricinolein production appears to be the co-migration of 3-OH TAG and OH-FFA. The presence of abundant FFA in the lipid extracts of *Y. lipolytica* cells grown on oleic acid has been reported previously (Mlícková et al., 2004) and was also identified in this study for cells growing on methyl ricinoleate, oleic acid, methyl oleate and glucose (see Section 5.3.6). It is conceivable that in *Y. lipolytica* cells grown on methyl ricinoleate, free fatty acid molecules containing hydroxyl groups would be present in the lipid extracts. In this solvent system such hydroxy free fatty acid species would co-migrate with any triricinolein present in the lipid extract. This would make it impossible to quantify triricinolein production with this TLC system, as the relative contributing effects of hydroxy free fatty acid and triricinolein are impossible to separate. In the Smith et al. (2003) report, triricinolein and hydroxy free fatty

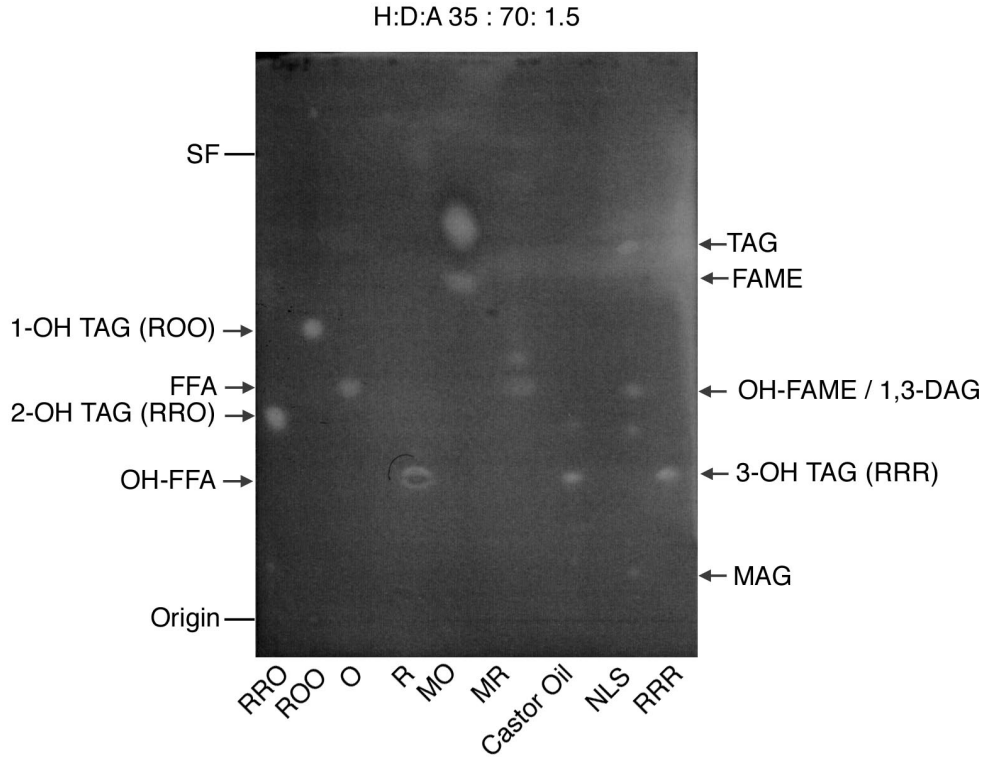


Figure 5.17: Evaluation of hexane : diethyl ether : acetic acid (70:140:3) as a solvent system capable of separating of hydroxy and non-hydroxy lipid species. The solvent system was previously reported to separate 3-OH TAG (RRR) from other lipid components (Smith et al., 2003) but in this laboratory 3-OH TAG co-migrated with OH-FFA. **Clockwise from top: TAG = triacylglycerol, FAME = fatty acid methyl ester, OH-FAME = hydroxy fatty acid methyl ester (i.e. methyl ricinoleate), DAG = diacylgylcerol, 3-OH TAG (RRR) = triricinolein, MAG = monoacylglycerol, RRR = triricinolein standard, NLS = neutral lipid standard (containing MAG, DAG, TAG), MR = methyl ricinoleate standard, MO = methyl oleate standard, R = ricinoleic acid standard, O = oleic acid standard, ROO = monoricinolein diolein, RRO = diricinolein monoolein, OH-FFA = hydroxy free fatty acid, 2-OH TAG (RRO) = diricinolein monoolein, FFA = free fatty acid, 1-OH TAG (ROO) = monoricinolein diolein, SF = solvent front.**

acid are adjacent to each other but there is clear separation. This meant that the combination of the H:D:A 70:140:3 solvent system and Whatman Silica G TLC plates were incompatible with the lipid analysis requirements of the project, and were rejected. As stated, interlaboratory reproducibility of TLC plates can be effected by a number of variables including those related to solvents, TLC plates and atmospheric pressure.

An alternative silica gel 60 TLC plate from Merck (Whitehouse Station, NJ. USA) (serial number : 1.05553.0001) was investigated. A neutral lipid standard containing MAG, 1,3-DAG, 1,2-DAG and TAG was separated alongside Castor oil, ricinoleic acid (OH-FFA) and methyl ricinoleate (OH-FAME) in the same solvent system: hexane : diethyl ether : acetic acid (70:140:3) (Figure 5.18). On evaluation, this TLC plate gave a clear separation of the ricinoleic acid and triricinolein positions. This was repeated in triplicate and was the case for each plate. Therefore, this TLC plate and solvent system combination can be used to analyse triricinolein production without contamination from other hydroxy- and non-hydroxy lipid molecules.

5.4 Discussion

5.4.1 Growth characteristics

Y. lipolytica P01G was found to grow more efficiently than *S. cerevisiae* on minimal media containing 2% glucose. The growth rates measured for *Y. lipolytica* P01G reflected those published elsewhere (Mlícková et al., 2004). The growth rates of both types of yeast are likely to be influenced by their auxotrophic status. Also, it is not a completely accurate comparison as the growth of *S. cerevisiae* is almost certainly reduced due to it growing at the preferred temperature of *Y. lipolytica*, 28 °C. Previously, a reduction in growth rate for *S. cerevisiae* growing at 27.5 °C compared to 30 °C has been described (Richards, 1928). The primary purpose of the exercise was to characterise the growth pattern of *Y. lipolytica* on 2% glucose and this has been achieved.

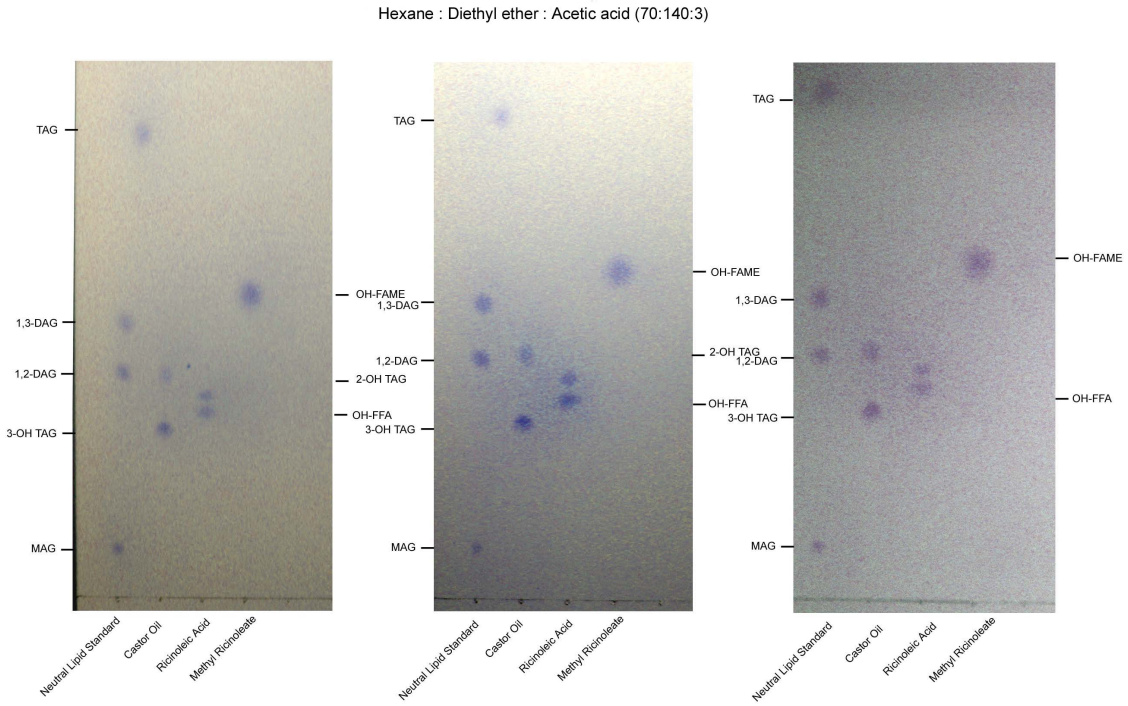


Figure 5.18: Triplicate evaluation of separation of triricinolein from ricinoleic acid in the hexane : diethyl ether : acetic acid solvent system and with Merck silica gel 60 TLC plates. Triricinolein was found to co-migrate with ricinoleic acid in the H:D:A 70:140:3 solvent system when used with Whatman Silica G 60 plates. The lipid samples analysed in triplicate here have been separated with the same solvent system but with Merck Silica Gel 60 plates and give distinct separation of triricinolein from ricinoleic. Therefore, this TLC plate and solvent system combination can be used to analyse triricinolein production without contamination from other hydroxy- and non-hydroxy lipid molecules.

The growth of *Y. lipolytica* P01G was found to be perturbed when growing on 1% methyl ricinoleate compared to 1% methyl oleate. This observation has been reported previously (Ratledge, 1984). It is hypothesised that ricinoleic acid readily incorporates into cell membranes leading to the formation of 'pores'. This may lead to depressed growth through perturbed membrane biochemistry leading to non-viable cells. Alternatively, γ-decalactone, an intermediate product of the β-oxidation of hydroxy fatty acids in certain yeasts including *Y. lipolytica* , is known to be cytotoxic (Dufossé et al., 1999; Feron et al., 1997). Although the rate of growth and total conversion of nutrients to biomass was depressed in cells growing on methyl ricinoleate compared to oleic acid, the times of growth stage transition were the same for the cells growing on either carbon substrate. This is important as it simplifies harvest timings in experiments comparing oleic and methyl ricinoleate grown cells.

Experimental evidence supported the auxotrophic status of *Y. lipolytica* strain P01G, and this was found to be a useful way of validating the strain used in this study.

5.4.2 Lipid accumulation in glucose and oleic acid fed cells

The analysis of the fatty acid profile of TAG from glucose-grown cells revealed an increase in TAG abundance between 2 and 10 hours of growth, with oleic acid (C18:1) being the dominant fatty acid at the final time point where it contributed 52.7% of the total TAG. Palmitic acid (C16:0) and linoleic acid (C18:2) contributed around 17% and 19% respectively at 10 hours, and palmitoleic (C16:1) and stearic acid (C18:0) were minor components contributing 4.5% and 6.3% respectively. A recent analysis of the fatty acid components of the lipid bodies from *Y. lipolytica* W29 (wildtype strain) grown 2% glucose YPD for 24 hours reported similar proportions of fatty acids, despite the differences in various aspects of the experimentation (Athenstaedt et al., 2006). In the Athenstaedt study (2006) oleic acid was identified as the dominant component at

24 hours, where it was reported to contribute 52.5% of the fatty acids. The general pattern of higher proportions of palmitic (C16:0) and linoleic (C18:2) and the lower proportions of palmitoleic (C16:1) and stearic acid (C18:0) described here for 10 hours were also reported for the analysis of oil bodies at 24 hours. The observation of the near identical contribution of oleic acid (C18:1) suggests that for this fatty acid it may reach its maximum proportion in the TAG by 10 hours for cells grown of 2% glucose. Also, the significant (almost 20%) proportion of linoleic acid observed in cells grown on 2% glucose supports a previous report that δ-9 desaturase is strongly expressed in glucose grown cells (Meesters and Eggink, 1996). Crucially, the similar fatty acid profile validates the concept of using TAG isolated by TLC as a method of quickly gaining insight into the lipid constituents on the lipid bodies, without requiring the extended purification procedures to isolate that organelle.

Oleic acid was identified as the dominant fatty acid species of the TAG in oleic acid grown *Y. lipolytica* P01G at 10 hours of growth, where it contributed 80.9% of the TAG. Oleic acid was reported as contributing 77.0% of the fatty acids from isolated lipid in 1% oleic acid-grown *Y. lipolytica* W29 at 24 hours of growth (Athenstaedt et al., 2006). C16:0, C16:1, C18:0 and C18:2 were also identified as minor components of the TAG in both studies, although the relative proportions of these minor components varied. As with the analysis of glucose-grown cells, this provides evidence of the 10 hours time point (mid-exponential) becoming representative of the stationary profile and again validating the analysis of the TLC-plate fractionated TAG as an alternative to analysing harvested lipid bodies.

Steryl esters are another resident neutral lipid component of yeast lipid bodies (Daum et al., 2007) which would contribute to the fatty acid profile reported in the Athenstaedt et al. (2006) paper but not the fatty acid profile reported here. However the similarities of the figures, despite the differences in the time points, support the literature base which states that steryl esters are minor components (around 5%) of

the *Y. lipolytica* oil body (Ratledge, 1982).

5.4.3 Technological development of analysis procedures compatible with hydroxy and non-hydroxy lipids

The use of TLC plates is a well established method for the separation of diverse lipid species. By altering the hydrophobicity of the TLC sovent, the migration of different lipid classes can be manipulated. In an initial analysis of total lipid extractions from *Y. lipolytica* P01G a combination a Whatman G TLC plates were used with a hexane : diethyl ether : acetic acid solvent system. The H:D:A (70:30:1) solvent system has been previously reported to provide good separation of neutral lipid classes including monoacylglycerols, diacylglycerols and triacylglycerol (Smith et al., 2003). Lipid extracted from *Y. lipolytica* P01G fed on glucose, oleic acid, methyl oleate and methyl ricinoleate was separated in this solvent system and the fatty acid methyl esters (methyl oleate and ricinoleate) were found to co-migrate with TAG, preventing the analysis of pure storage oil in this system where FAMEs may be present in the lipid extract. Alternative TLC solvent systems were investigated which allowed separation of FAMEs from TAG.

When silyl derivatives were separated in solvent systems other than H:D:A (70:30:1) there was evidence of trailing, which was hypothesised to be due to degradation of the derivative during or before TLC separation. A personal communication from Dr. W. Christie (July, 2007) confirmed an early report (Christie, 1982) that TMCS esters are unstable and hydrolyse readily if not kept in the silylating solution.

As degradation of lipid compounds would cause inaccuracies in quantification and potentially contamination of the TLC plate with breakdown products, alternative methods were investigated which negated the need for derivatisation to achieve successful identification and quantification of hydroxy lipid species. Hexane : diethyl ether : acetic acid (70:140:3) has previously been reported to effectively separate a range of hydroxy-

and non-hydroxy lipid species (Smith et al., 2003). After investigation with a range of relevant standards and Whatman Silica G TLC plates, the critical location of 3-OH TAG was found to migrate very closely to OH-FFA. An alternative Merck Silica Gel 60 TLC plate was found to allow distinct separation of OH-FFA and 3-OH TAG. 1,2-DAG and 2-OH TAG were found to have similar migratory characteristics with this solvent system and TLC plate combination however. The method described here provides the foundation to a triricinolein analysis platform for *Y. lipolytica*. Quantification can be achieved on a GC by adding a known amount of standard to the triricinolein sample after the spot has been scraped from the TLC plate. The disadvantage of this method compared to adding a standard such as triheptadecanoic acid prior to lipid extraction is it cannot account for losses during lipid extraction and separation. One way to increase accuracy of quantification would be to use triheptadecanoic acid (C17:0) added prior to lipid extraction as a standard for non-hydroxy TAG quantification. A secondary standard (e.g. C17:1) can then be added to every sample including the non-hydroxy TAG standard post TLC separation. The C17:1 peak can then be normalised against the C17:0 peak.

5.4.4 Use of methyl ricinoleate over alternatives

The problems addressed in this Chapter (co-migration of FAMEs with TAG, degradation of derivatised methyl ricinoleate when separated by TLC) could have been circumvented by using ricinoleic acid as the free fatty acid rather than its FAME-derivative. Derivatised ricinoleic acid would have remained close to the origin of the TLC plate due to the interaction of its comparatively polar carboxyl group (compared to the methyl group of the FAME molecule) with the TLC plate matrix. Although a number of alternative TLC systems were identified which separated FAMEs and TAG molecules, the additional problem of derivatised methyl ricinoleate degradation was encountered (it was due to this degradation of derivatised molecules that non-

derivatised lipid extracts were evaluated and this led to identifying a solvent system / TLC plate combination which allowed analysis of triricinolein production in FAME-fed cells). The reason that alternative methyl ricinoleate-compatible TLC systems and use of non-derivatised extracts were evaluated, rather than changing to ricinoleic acid was that the cost of the alternative would have prohibited the technique's use for routine lipid analyses. To ensure sufficient lipid is generated for lipid analyses, it is desirable to use a large culture size so as to harvest sufficient cell mass. This is especially important in the analysis of cells at an early stage of growth. For culture media containing 1% lipid substrate, a 500 ml culture fed on ricinoleic acid would require 5 g of the lipid molecule. At the time of experimentation, 5 g of ricinoleic acid from the Sigma Chemical Company cost $\sim$ 400 UKP. Therefore to perform replicate lipid analyses on multiple transformed strains of *Y. lipolytica* would incur considerable cost to the laboratory and would make alternative assay systems (e.g. *in vitro* assay or small scale feeding followed by HPLC analysis of lipid) a more sensible choice. However, Stearinerie Dubois (a specialist chemical company which supplies the cosmetic industry with a range of ricinoleic acid-derived products) was able to supply methyl ricinoleate at a significantly lower cost, and in sufficient quantities for multiple future analyses. Unfortunately, they did not stock ricinoleic acid as the free fatty acid. For this reason, development of methyl ricinoleate-compatible methodology was pursued. Although alternative fractionation techniques such as HPLC may have provided another analysis solution, lack of experimental time meant these could not be evaluated.

5.4.5 Concluding Remarks

In this chapter, the conditions for growth and lipid extraction were established for the oleaginous yeast *Y. lipolytica*. In the development of lipid analysis procedures, difficulties were encountered relating to the co-migration of lipid species by TLC and evidence of degradation of derivatised hydroxy-lipid species affecting the validity of

initial analysis attempts of methyl ricinoleate and methyl oleate fed cells. An alternative TLC / solvent system combination has been established which separates a range of lipid species including hydroxy lipid species and gives a clear separation of non-hydroxy TAG and triricinolein from other potential contaminants.

This chapter has provided a protocol for the growth and harvest of *Y. lipolytica* on different carbon substrates including methyl esters of ricinoleic acid, and the lipid extraction and analysis procedures capable of identifying and quantifying triricinolein production in this organism. Used with transformed and untransformed strains, it may provide a rapid method for the assay of *R. communis* lipid biosynthesis genes on the production of triricinolein.

Chapter 6

Final Discussion

The general hypothesis of this thesis is that the enzymatic components key to high levels of triricinolein (TRO) production in *R. communis* can be identified through the use of a proteomic analysis targeting the site of castor oil biosynthesis (the developing seed) and the contribution of these enzymes to TRO production can be rapidly assessed in yeast prior to transformations in a plant host.

To this end, a differential screen of developing and germinating *R. communis* seed ER was performed, with a view to identifying components of lipid biosynthesis which may be crucial to the production of high levels of TRO. This focussed on those proteins which are soluble in lysis buffer and separable by 2DE. 91 spots were identified as elevated in developing compared to germinating seed and thus became targets for MS analysis. A limited NCBI*nr* database comprised of independent submissions and a complete *R. communis* database were searched but no components of lipid biosynthesis were identified (a large number of storage proteins, proteins involved in protein translation, hypothetical proteins and sterol biosynthesis enzymes were, however). It is possible that key proteins involved in TRO biosynthesis were present in the 35 spots (38%) which gave no identity in the proteomic screen. However, analysis of hydrophobicity and prediction of transmembrane domains for GPAT, LPAT and oleate $\triangle 12$ hydroxylase indicate they are not present on the gels due to their poor solubility in 2DE lysis buffer or are lost due to precipitation during IEF. An alternative approach focussing on the membrane-bound components of the developing ER is utilising iTRAQ labelling and MS/MS to quantify and identify the resident proteins. This will not suffer from the 2DE limitation with membrane proteins and thus is much more likely to identify candidate proteins. Unfortunately, this work was beyond the scope of the author's industrial funding.

Y. lipolytica was hypothesised to be a suitable yeast for the rapid assay of candidates for the production of high TRO levels, due to its oleaginous nature, its ability to grow on hydrophobic substrates including ricinoleic acid (methyl ricinoleate was used in

this study) and the availability of mature genomic tools for its transformation. The central question asked was: can *Y. lipolytica* make TRO, and if so can the addition of *R. communis* complex lipid biosynthesis genes increase its production; if it can't, can the addition of these genes cause it to make it and if so by how much? Prior to answering this question, methodologies for growing the yeast, extracting its lipid and separating its storage oil through TLC had to be established. A time-course analysis of TAG composition through the exponential phase of growth of *Y. lipolytica* growing on glucose and a variety of hydrophobic substrates including methyl ricinoleate was performed. New data was obtained for glucose and oleic acid grown cells which complemented the analysis of stationary phase TAG composition in glucose and oleic acid fed *Y. lipolytica* published previously (Athenstaedt et al., 2006). Large peaks were identified for ricinoleic and oleic acid in the TAG of methyl ricinoleate and methyl oleate respectively; this was identified as an artefact due to co-migration of lipid species on the TLC plate and alternative solvent systems were investigated. An unforeseen problem as a result of this was evidence of degradation of silylated methyl ricinoleate on the TLC plate. Alternative solvent combinations were unable to identify a solvent system which prevented the silylated methyl ricinoleate degradation products running at multiple positions through the plate *and* clearly separated TAG from other lipid species. An alternative approach of not silylating the lipid extract and using a combination of high performance TLC plate and a solvent system previously shown to separate a wide range of hydroxy- and non-hydroxy complex lipids (Smith et al., 2003) was found to allow a clear separation of TRO from other lipid species.

6.1 Future Directions

As stated, the use of MS-based proteomic procedures which are capable of identifying membrane-bound components will complement the soluble 2DE analysis presented here

and potentially identify key TRO producing enzymes absent from this analysis.

There is increasing evidence of the importance of substrate channelling and specialised sub-domains and these may prove to have a quantitatively important role in high levels of TRO production. Blue Native gels have previously been used to dissect complexes important in mitochondria (Schagger, 2001) and may provide very interesting results in an analysis of developing ER membrane .

The protocols are now in place for the analysis of TRO production in *Y. lipolytica*, both for untransformed strains and strains expressing *R. communis* lipid biosynthesis genes. A large body of work has been published concerning *Y. lipolytica's* ability to efficiently break down hydroxy fatty acids including ricinoleic acid and methyl ricinoleate through β-oxidation. Knockout lines incapable of breaking down lipid may provide a useful background for *R. communis* lipid gene transformation experiments as the anabolic effects of the transgene on storage oil production can be assessed without the opposing catabolic process of lipid breakdown; a process known to continue into stationary phase and to act upon TAG stores (Beopoulos et al., 2008). Strains exist which have each of the known acyl CoA oxidase encoding genes knocked out, for example JMY1233 (Smit et al., 2005). Also, *Y. lipolytica* strains expressing oleate $\triangle$12 hydroxylase and capable of making ricinoleoyl CoA may provide a very useful tool for investigating the role of enzymes such as PDAT and phospholipase in the flux of ricinoleic acid out of the PC and made available for TAG biosynthesis.

Bibliography

Agrawal, G., Hajduch, M., Graham, K., and Thelen, J. (2008). In-depth investigation of soybean seed-filling proteome and comparison with a parallel study of rapeseed. *Plant Physiol.*, 148(1):504–518.

Annan, R. and Carr, S. (1997). The essential role of mass spectrometry in characterizing protein structure: mapping posttranslational modifications. *J. Protein Chem.*, 16(5):391–402.

Argon, Y. and Simen, B. (1999). GRP94, an ER chaperone with protein and peptide binding properties. *Semin. Cell Dev. Biol.*, 10(5):495–505.

Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J., and Chardot, T. (2006). Lipid particle composition of the yeast Yarrowia lipolytica depends on the carbon source. *Proteomics*, 6(5):1450–1459.

Bafor, M., Smith, M., Jonsson, L., Stobart, K., and Stymne, S. (1991). Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor-bean (Ricinus communis) endosperm. *Biochem. J.*, 280(2):507–514.

Barneche, F., Gaspin, C., Guyot, R., and Echeverria, M. (2001). Identification of 66 box C/D snoRNAs in Arabidopsis thaliana: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2'-O-methylation sites. *J. Mol. Biol.*, 311(1):57–73.

Barth, G. and Gaillardin, C. (1996). Yarrowia lipolytica. In Wolf, K., editor, *Nonconventional Yeasts in Biotechnology, a Handbook*, pages 313–388. Springer.

Bashir, M., Hubatsch, I., Leinenbach, H., Zeppezauer, M., Panzani, R., and Hussein, I. (1998). Ric c 1 and Ric c 3, the allergenic 2S albumin storage proteins of Ricinus communis: Complete primary structures and phylogenetic relationships. *Int. Arch. Allergy. Immunology*, 115(1):73–82.

Beavis, R. and Chait, E. (1989). Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.*, 3(12):432–435.

Becker, T., Vogtle, F., Stojanovski, D., and Meisinger, C. (2008). Sorting and assembly of mitochondrial outer membrane proteins. *Biochim. Biophys. Acta*, 1777(7):557–563.

Beisson, F., Koo, A., Ruuska, S., Schwender, J., Pollard, M., Thelen, J., Paddock, T., Salas, J., Savage, L., Milcamps, A., Mhaske, V., Younghee, C., and Ohlrogge, J. (2003). Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol.*, 132:681–697.

Beisson, F., Li, Y., Bonaventure, G., Pallard, M., and Ohlrogge, J. (2007). The acyltransferase GPAT5 is required for the synthesis of suberin in the seed coat and root of Arabidopsis. *Plant Cell*, 19(1):351–368.

Bell, P. and Karuso, P. (2003). Epicocconone, a novel fluorescent compound from the fungus Epicoccum nigrum. *J. Am. Chem. Soc.*, 125(31):9304–9305.

Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M., Hapala, I., Papanikolaou, S., Chardot, T., and Nicaud, J. (2008). Control of lipid accumulation in the yeast Yarrowia lipolytica. *Appl. Envion. Microbiol.*, 74(24):7779–7789.

Biemann, K. (1990). Nomenclature for peptide fragment ions (positive ions). [appendix 5]. *Methods Enzymol.*, 193:886–887.

Bjellqvist, B., Ek, K., Righetti, P., Gianazza, E., Görg, A., Westermeier, R., and Postel, W. (1982). Isoelectric focusing in immobilized pH gradients : principle, methodology and some applications. *J. Biochem. Biophys. Methods*, 6(4):317–339.

Bjellqvist, B., Sanchez, J., Pasquali, C., Ravier, F., Paquet, N., Frutiger, S., Hughes, G., and Hochstrasser, D. (1993). A nonlinear wide-range immobilized pH gradients for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis*, 14(12):1357–1365.

Borchers, C., Peter, J., Hall, M., Kunkel, T., and Tomer, K. (2000). Identification of in-gel digested proteins by complementary peptide mass fingerprinting and tandem mass spectrometry data obtained on an electrospray ionization quadrupole time-of-flight mass spectrometer. *Anal. Chem.*, 72(6):1163–1168.

Bowden, L. and Lord, J. (1976a). The cellular origin of glyoxysomal proteins in germinating castor-bean endosperm. *Biochem. J.*, 154(2):501–506.

Bowden, L. and Lord, J. M. (1976b). Similarities in the polypeptide composition of glyoxysomal and endoplasmic-reticulum membranes from castor-bean endosperm. *Biochem. J*, 154(2):491–499.

Brachmann, C., Davies, A., Cost, G., Caputo, E., Li, J., Hieter, P., and Boeke, J. (1998). Designer deletion strains derived from Saccharoymces cervesiae S288C: a

useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, 14(2):115–132.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72(248-254).

Brigham, R. (1993). Castor: Return of an old crop. In Janick, J. and Simon, J., editors, *New Crops*, pages 380–383. Wiley, New York.

Brockerhoff, H. (1971). Stereospecific analysis of triacylglycerides. *Lipids*, 6:942–956.

Broun, P., Boddupalli, S., and Somerville, C. (1998a). A bifunctional oleate 12-hydroxylase: desaturase from lesquerella fendleri. *Plant J.*, 13(2):201–210.

Broun, P., Shanklin, J., Whittle, E., and Somerville, C. (1998b). Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. *Science*, 282(5392):1315–1317.

Broun, P. and Somerville, C. (1997). Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic Arabidopsis plants that express a fatty acyl hydroxylase cDNA from castor bean. *Plant Physiol.*, 113(3):933–942.

Brown, A., Brough, C., Kroon, J., and Slabas, A. (1995). Identification of a cDNA that encodes a 1-acyl-sn-glycerol-3-phosphate acyltransferase from Limnanthes douglasii. *Plant Mol. Biol.*, 29(2):267–278.

Brown, J., Jolliffe, N., Frigerio, L., and Roberts, L. (2003). Sequence-specific, golgi-dependent vacuolar targeting of castor bean 2S albumin. *Plant J.*, 36(5):711–719.

Browse, J. (1996). Towards rational engineering of plant oils: crystal structure of the 18:0-ACP desaturase. *Trends in Plant Science*, 1(12):403–404.

Browse, J., McConn, M., James, D., and Miquel, M. (1992). Mutants of Arabidopsis deficient in the synthesis of alpha-linolenate. Biochemical and genetic characterization of the endoplasmic reticulum linoleoyl desaturase. *J. Biol. Chem.*, 268(22):16345–16351.

Browse, J. and Somerville, C. (1991). Glycerolipid synthesis: biochemistry and regulation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 42:467–506.

Bukau, B. and Horwich, A. (1998). The Hsp70 and Hsp60 chaperone machines. *Cell*, 92(3):351–366.

Cahoon, E., Dietrich, C., Meyer, K., Damude, H., Dyer, J., and Kinney, A. (2006). Conjugated fatty acids accumulate to high levels in phospholipids of metabolically engineered soybean and Arabidopsis seeds. *Phytochemistry*, 67(12):1166–1176.

Cahoon, E. and Kinney, A. (2005). Production of vegetable oils with novel properties: using genomic tools to probe and manipulate fatty acid metabolism. *Eur. J. Lipid Sci. Technol.*, 107:239–243.

Cao, Y.-Z., Oo, K.-C., and Huang, A. (1990). Lysophosphatidate acyltransfease in the microsomes from maturing seeds of meadowfoam (Limnanthes alba). *Plant Physiol.*, 94(3):1199–1206.

Cases, S., Smith, S., Zheng, Y., Myers, H., Lear, S., Sande, E., Novak, S., Collins, C., Welch, C., Lusis, A., Erickson, S., and Farese, R. (1998). Identification of a gene encoding an acyl-coa:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. Natl. Acad. Sci. USA*, 95(22):13018–13023.

Caupin, H. (1997). Products from castor oil: Past, present and future. In Gunstone, F. and Padley, F., editors, *Lipid Technologies and Applications*, pages 787–795. Marcel Dekker Inc.

Chen, F., Li, Q., and He, Z. (2007). Proteomic analysis of rice plasma membrane-associated proteins in response to chitooligosaccharide elicitors. *Journal of Integrative Plant Biology*, 49(6):863–870.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T., Higgins, D., and Thompson, J. (2003). Multiple sequence alignment with the clustal series of programs. *Nucleic Acids Res.*, 31(13):3497–3500.

Chevalier, F., Rofidal, V., Vanova, P., Bergoin, A., and Rossignol, M. (2004). Proteomic capacity of recent fluorescent dyes for protein staining. *Phytochemistry*, 65:1499–1506.

Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., and Rabilloud, T. (1998). New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis*, 19(11):1901–1909.

Chivasa, S., Hamilton, J., Pringle, R., Ndimba, B., Simon, W., Lindsey, K., and Slabas, A. (2006). Proteomic analysis of differentially expressed proteins in fungal elicitor-treated Arabidopsis cell cultures. *J. Exp. Bot.*, 57(7):1553–1562.

Christiansen, K. (1978). Triacylglycerol synthesis in lipid particles from baker's yeast (Saccharomyces cerevisiae). *Biochim. Biophys. Acta*, 530(1):78–90.

Christiansen, K. (1979). Utilization of endogenous diacylglycerol for the synthesis of triacylglycerol, phosphatidylcholine and phosphatidylethanolamine by lipid particles from baker's yeast (Saccharomyces cerevisiae). *Biochim. Biophys. Acta*, 574(3):448–460.

Christie, W. (1982). *Lipid analysis*. Pergamon Press.

Clinton, W. (1999). Developing and promoting biobased products and bioenergy. Executive Order 13134.

Clouse, S. (2002). Arabidopsis mutants reveal multiple roles for sterols in plant development. *Plant Cell*, 14(9):1995–2000.

Corbett, J., Dunn, M., Posch, A., and Görg, A. (1994). Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension: an interlaboratory comparison. *Electrophoresis*, 15(8-9):1205–1211.

Corey, E., Matsuda, S., and Bartel, B. (1993). Isolation of an Arabidopsis thaliana gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol syntase by the use of a chromatographic screen. *Proc. Natl. Acad. Sci. USA*, 90(24):11628–11632.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, 16(22):10881–10890.

Coughlan, S., Hastings, C., and Winfrey, R. (1996). Molecular characterisation of plant endoplasmic reticulum. Identification of protein disulfide-isomerase as the major reticuloplasmin. *Eur. J. Biochem.*, 235(1-2):215–224.

Coughlan, S., Hastings, C., and Winfrey, R. J. (1997). Cloning and characterization of the calreticulin gene from Ricinus communis L. *Plant Mol. Biol.*, 34(6):897–911.

Covey, T., Huang, E., and Henion, J. (1991). Structural characterisation of protein tryptic peptides via liquid chromatography/mass spectrometry and collision induced dissociation of their doubly charged molecular ions. *Anal. Chem.*, 63(13):1193–1200.

Dahlqvist, A., Ståhl, U., Lenman, A., Banas, A., Lee, M., Sandager, M., Ronne, H., and Stymne, S. (2000). Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc. Natl. Acad. Sci. USA*, 97(12):6487–6492.

Daum, G., Wagner, A., Czabany, T., and Athenstaedt, K. (2007). Dynamics of neutral lipid storage and mobilization in yeast. *Biochimie*, 89(2):243–248.

Davis, M. and Lee, T. (1997). Variable flow liquid chromatography-tandem mass spectrometry and the comprehensive analysis of complex protein digest mixtures. *J. Am. Soc. Mass Spectrom.*, 8:110–121.

Dickinson, C., Hussein, E., and Nielsen, N. (1989). Role of posttranslational cleavage in glycinin assembly. *Plant Cell*, 1(4):459–469.

Domergue, F., Chevalier, S., Creach, A., Cassagne, C., and Lessire, R. (2000). Purification of the acyl-CoA elongase complex from developing rapeseed and characterization of the 3-ketoacyl-CoA synthase and the 3-hydroxyacyl-CoA dehydratase. *Lipids*, 35:487–494.

Donaldson, R. and Beevers, H. (1977). Lipid composition of organelles from germinating castor bean endosperm. *Plant Physiol.*, 59(2):259–263.

Dubacq, J., Drapier, D., Trémolières, A., and Kader, J. (1984). Role of phospholipid transfer protein in the exchange of phospholipids between microsomes and chloroplasts. *Plant Cell Phys.*, 25(7):1197–1204.

Dufossé, L., Souchon, I., Feron, G., Latrasse, A., and Spinnler, H. (1999). In situ detoxification of the fermentation medium during γ-decalactone production production with the yeast Spordiobolus salmonicolor. *Biotechnol. Prog.*, 15:135–139.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., de Montigny, J., Marck, C., Neuveglise, C., Talla, E., Gofford, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.-M., Beyne, E., Bleykasten, C., Boiramie, A., Boyer, J., Cattolico, L., Confanioleri, F., de Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.-M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.-F., Straub, M.-L., Suleau, A., Swennene, D., Tekaia, F., Welsolowski-Louvel, M., and Westhof, E. (2004). Genome evolution in yeasts. *Nature*, 430(6995):35–44.

Dunn, M. and Görg, A. (2001). Two-dimensional polyacrylamide gel electrophoresis for proteome analysis. In Dunn, M. and Pennington, S., editors, *Proteomics. From protein sequence to function*, number ISBN 1-85996-296-3. Bios Scientific.

Dyer, J. and Mullen, R. (2001). Immunocytological localization of two plant fatty acid desaturases in the endoplasmic reticulum. *FEBS Lett.*, 494(1-2):44–47.

Eastmond, P. (2004). Cloning and characterization of the acid lipase from castor beans. *J. Biol. Chem.*, 279(44):45540–45545.

Elmore, A. (2007). Final report on the safety assessment of Ricinus communis (Castor) seed oil, hydrogenated castor oil, glyceryl ricinoleate SE, ricinoleic acid, potassium ricinoleate, sodium ricinoleate, zinc ricinoleate, cetyl ricinoleate, ethyl ricinoleate, glycol ricinoleate, isopropyl ricinoleate, methyl ricinoleate, and octyldodecyl ricinoleate. *Int. J. Toxicol.*, 26(3):31–77.

Emes, M., Bowsher, C., Hedley, C., Burrell, M., Scrase-Field, E., and Tetlow, I. (2003). Starch synthesis and carbon partitioning in developing endosperm. *J. Exp. Bot.*, 54(382):569–575.

Eng, J., McCormack, A., and Yates, J. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.*, 5:976–989.

Ephritikhine, G., Ferro, M., and Rolland, N. (2004). Plant membrane proteomics. *Plant Physiol. Biochem.*, 42(12):943–962.

Erber, A., Horstmann, C., and Schobert, C. (1997). A cDNA clone for acyl-CoA-binding protein from castor bean. *Plant Physiol.*, 114(1):396.

Eriksson, J., Chait, B., and Fenyo, D. (2000). A statistical basis for testing the significance of mass spectrometric protein identification results. *Anal. Chem.*, 72(5):999–1005.

Falkenstein, E., Schmieding, K., Lange, A., Meyer, C., Gerdes, D., Welsch, U., and Wehling, M. (1998). Localization of a putative progesterone membrane binding protein in porcine hepatocytes. *Cell Mol. Biol.*, 44(4):571–578.

Fenn, J., Mann, M., Meng, C., Wong, S., and Whitehouse, C. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71.

Feron, G., Dufossé, L., Mauvais, G., Bonnarme, P., and Spinnler, H. (1997). Fatty acid accumulation in the yeast Sporidobolus salmonicolor during batch production of γ-decalactone. *FEMS Microbiol. Lett.*, 149(1):17–24.

Fey, S., Nawrocki, A., Larsen, M., Görg, A., Roepstorff, P., Skews, G., Williams, R., and Larsen, P. (1997). Proteome analysis of Saccharomyces cerevisiae: a methodological outline. *Electrophoresis*, 18(8):1361–1372.

Fickers, P., Benetti, P., Wache, Y., Marty, A., Mauersberger, S., Smit, M., and Nicaud, J.-M. (2005). Hydrophobic substrate utilisation by the yeast Yarrowia lipolytica, and its potential applications. *FEMS Yeast Res.*, 5(6):527–543.

Filippin, L., Magalhaes, P., Di Benedetto, G., Colella, M., and Pozzan, T. (2003). Stable interactions between mitochondria and endoplasmic reticulum allow rapid accumulation of calcium in a subpopulation of mitochondria. *J. Biol. Chem.*, 278(40):39224–39234.

Fischer, K., Weber, A., Brink, S., Arbinger, B., Schünemann, D., Borchert, S., Heldt, H., Popp, B., Benz, R., and Link, T. (1994). Porins from plants. Molecular cloning and functional characterization of two new members of the porin family. *J. Biol. Chem.*, 269(41):25754–25760.

Flynn, G., Pohl, J., Flocco, M., and Rothman, J. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature*, 353(6346):726–730.

Fox, B., Shanklin, J., Ai, J., Loehr, T., and Sanders-Loehr, J. (1994). Resonance raman evidence for an Fe-O-Fe centre in stearoyl-ACP desaturase. Primary sequence identity with other diiron-oxo proteins. *Biochemistry*, 33(43):12776–12786.

Fraser, T., Waters, A., Chatrattanakunchai, S., and Stobart, K. (2000). Does triacylglycerol biosynthesis require diacylgylcerol acyltransferase (DAGAT). *Biochemical Society Transactions*, 28(6):698–700.

Galliard, T. and Stumpf, P. (1996). Enzymatic synthesis of ricinoleic acid by a microsomal preparation from developing Ricinus communis seeds. *J. Biol. Chem.*, 241(24):5806–5812.

Gasch, A., Spellman, P., Kao, C., Carmel-Harrel, O., Eisen, M., Storz, G., Botstein, D., and Brown, P. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, 11(12):4241–4257.

Geerlings, A., Ibañez, M., Memelink, J., van Der Heijden, R., and Verpoorte, R. (2000). Molecular cloning and analysis of strictosidine beta-D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in Catharanthus roseus. *J. Biol. Chem.*, 275(5):3051–3056.

Gething, M. and Sambrook, J. (1992). Protein folding in the cell. *Nature*, 355(6355):33–45.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R., Belle, A., Dephoure, N., O'Shea, E., and Weissman, J. (2003). Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741.

Glomset, J. (1968). The plasma lecithin:cholesterol acyl transferase reaction. *J. Lipid Res.*, 9:155–167.

Goffeau, A., Barrell, B., Bussey, H., Davis, R., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J., Jacq, C., Johnston, M., Louis, E., Mewes, H., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. (1996). Life with 6000 genes. *Science*, 274(5287):563–567.

Gonzalez, E. (1982). Aggregated forms of malate and citrate synthase are localized in endoplasmic reticulum of endosperm of germinating castor bean. *Plant Physiol.*, 69(1):83–87.

Gonzalez, E. and Beevers, H. (1976). Role of the endoplasmic reticulum in glyoxysome formation in castor bean endosperm. *Plant Physiol.*, 57(3):406–409.

Gorbet, D. and Knauft, D. (1997). Registration of 'sunoleic 95r' peanut. *Crop Sci.*, 37:1392.

Görg, A., Boguth, G., Obermaier, C., and Weiss, W. (1998a). Two-dimensional electrophoresis electrophoresis of proteins in an immobilized pH 4-12 gradient. *Electrophoresis*, 19(8-9):1516–1519.

Görg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J., and Madjar, J. (1997). Very alkaline immobilized ph gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis*, 18(328-337).

Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, 21(6):1037–1053.

Görg, A., Postel, W., Gunther, S., Weser, J., Strahler, R., Hanash, S., Somerlot, L., and Kuick, R. (1998b). Approach to stationary two-dimensional pattern: influence of focusing time and immobiline/carrier ampholytes concentrations. *Electrophoresis*, 9(1):37–46.

Goshe, M., Blonder, B., and Smith, R. (2003). Affinity labelling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J. Proteome Res.*, 2:153–161.

Greenwood, J. and Bewley, J. (1982). Seed development in Ricinus communis (castor been). I. Descriptive morphology. *Can. J. Bot.*, 60(2):1751–1760.

Griffiths, G., Stobart, A., and Stymne, S. (1985). The acylation of sn-glycerol 3-phosphate and the metabolism of phosphatidate in microsomal preparations from the developing cotyledons of safflower (Carthamus tinctorius L.) seed. *Biochem. J.*, 230(2):379–388.

Griffiths, G., Walsh, M., and Harwood, J. (1993). Acyl-thioesterase activity in developing seeds of cocoa. *Phytochemistry*, 32(6):1403–1406.

Gunstone, F. (1979). Lipids. In Barton, D. and Ollis, W., editors, *In "Comprehensive Organic Chemistry"*. Pergamon, Oxford.

Guy, J., Abreu, I., Moche, M., Lindqvist, Y., Whittle, E., and Shanklin, J. (2006). A single mutation in the castor Delta9-18:0-desaturase changes reaction partitioning from desaturation to oxidase chemistry. *Proc. Natl. Acad. Sci. USA*, 103(46):17220–17224.

Gygi, S. and Aebersold, R. (2000). Mass spectrometry and proteomics. *Curr. Opin. Chemical Biology*, 4:489–494.

Gygi, S., Rist, B., Gerber, S., Turecek, F., Gelb, M., and Aebersold, R. (2003). Quantitative analysis of complex protein mixtures using istope-coded affinity tags. *Nat. Biotechnol.*, 17(10):994–999.

Gygi, S., Rochon, Y., Franza, B., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.*, 19(3):1720–1730.

Hajduch, M., Casteel, J., Hurrelmeyer, K., Song, Z., Agrawal, G., and Thelen, J. (2006). Proteomic analysis of seed filling in Brassica napus. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. *Plant Physiol.*, 141(1):32–46.

Hajduch, M., Ganapathy, A., Stein, J., and Thelen, J. (2005). A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiol.*, 137(4):1397–1419.

Hall, J. (2004). *Proteomic analysis of the heat shock and acclimation responses of cyanobacteria*. PhD thesis, Biological and Biomedical Sciences, Durham University.

Hamilton, R. and Hamilton, S. (1992). *Lipid analysis: A practical approach*. Oxford University Press.

Han, D., Eng, J., Zhou, H., and Aebersold, R. (2001). Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.*, 19(10):946–951.

Hara-Nishimura, I., Shimada, T., Hatano, K., Takeuchi, Y., and Nishimura, M. (1998). Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. *Plant Cell*, 10(5):825–836.

Harwood, J. (1996). Recent advances in the biosynthesis of plant fatty acids. *Biochim. Biophys. Acta*, 1301(1-2):7–56.

He, X., Turner, C., Chen, G., Lin, J., and McKeon, T. (2004). Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. *Lipids*, 39(4):311–318.

Heazlewood, J., Howell, K., and Millar, A. (2003). Mitochondrial complex I from Arabidopsis and rice: orthologs of mammalian and fungal components coupled with plant-specific subunits. *Biochim. Biophys. Acta*, 1604(3):159–169.

Hedge, R. and Podder, S. (1992). Studies on the variants of the protein toxins ricin and abrin. *Eur. J. Biochem.*, 204:155–164.

Heinemeyer, J., Eubel, H., Wehmhoner, D., Jansch, L., and Braun, H.-P. (2004). Proteomic approach to characterize the supramolecular organization of photosystems in higher plants. *Phytochemistry*, 65(12):1683–1692.

Heinz, E. and Roughan, P. (1983). Similarities and differences in lipid metabolism of chloroplasts isolated from 18:3 and 16:3 plants. *Plant Physiol.*, 72(2):173–179.

Hellyer, A., Leadlay, P., and Slabas, A. (1992). Induction, purification and characterisation of acyl-ACP thioesterase from developing seeds of oil seed rape (Brassica napus). *Plant Mol. Biol.*, 20(5):763–780.

Hendershot, L., Bole, D., Kohler, G., and Kearney, J. F. (1987). Assembly and secretion of heavy chains that do not associate posttranslationally with immunoglobulin heavy chain-binding protein. *J. Cell Biol.*, 104(3):761–767.

Henderson, R. and Tocher, D. (1992). Thin-layer chromatography. In Hamilton, R. and Hamilton, S., editors, *Lipid analysis: A practical approach*, pages 65–111. IRL Press, Oxford.

Heukeshoven, J. and Dernick, R. (1988). Improved silver staining procedure for fast staining in PhastSystem Development Unit. 1. Staining of sodium dodecyl sulfate gels. *Electrophoresis*, 9(1):28–32.

Hobbs, D., Lu, C., and Hills, M. (1999). Cloning of a cDNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana and its functional expression. *FEBS Lett.*, 452(3):145–149.

Huang, A. and Beevers, H. (1973). Localization of enzymes within microbodies. *J. Cell Biol.*, 58(2):379–389.

Hurkman, W. and Beevers, L. (1982). Sequestration of pea reserve proteins by rough microsomes. *Plant Physiol.*, 69(6):1414–1417.

Hwang, C., Sinskey, A., and Lodish, H. (1992). Oxidized redox state of glutathione in the endoplasmic reticulum. *Science*, 257(5076):1496–1502.

Ichihara, K. (1984). sn-Glycerol-3-phosphate acyltransferase in a particulate fraction from maturing safflower seeds. *Arch. Biochem. Biophys.*, 232(2):685–698.

Ichihara, K. and Noda, M. (1980). Fatty acid composition and lipid synthesis in developing safflower seed. *Phytochemistry*, 19:49–54.

Irwin, S., Keen, J., Findlay, J., and Lord, J. (1990). The Ricinus communis 2S albumin precursor: a single preproprotein may be processed into two different heterodimeric storage proteins. *Mol. Gen. Genet.*, 222(2-3):400–408.

Ishizaki-Nishizawa, O., Fujii, T., Azuma, M., Sekiguchi, K., Murata, N., Ohtani, I., and Toguri, T. (1996). Low-temperature resistance of higher plants is significantly enhanced by a nonspecific cyanobacterial desaturase. *Nat. Biotechnol.*, 14(8):1003–1006.

Jain, A. (2004). Cloning and structural analysis of a cDNA clone encoding glycinin (Gly-1) seed storage protein of peanut. *Electronic J. Biotech.*, 7(3):189–199.

Jako, C., Kumar, A., Yangdou, W., Zou, J., Barton, D., Giblin, E., Covello, P., and Taylor, D. (2001). Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiol.*, 126(2):861–874.

James, W. and Johnson, D. (1999). Vegetable oil lubricants for internal combustion engines and total loss lubricants. *US Patent*, 5,888,947.

Jensen, O., Podtelejnikov, A., and Mann, M. (1997). Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. *Anal. Chem.*, 69(23):4741–4750.

Johnson, A. and van Waes, M. (1999). The translocon: a dynamic gateway at the ER membrane. *Annu. Rev. Cell. Dev. Biol.*, 15:799–842.

Kainuma, K., Ookura, T., and Kawamura, Y. (1995). Purification and characterization of protein disulfide isomerase from soybean. *J. Biochem.*, 117(1):208–215.

Kaplan, N., Lowenstein, J., and Colowick, N., editors (1975). *Methods in Enzymology: Lipids*, volume 35. Academic Press.

Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal. Chem.*, 60(20):2299–2301.

Karssen, C., Brinkhorst-van deer Swan, D., Breekland, A., and Koorneef, H. (1983). Induction of dormancy during seed development by endogenous abscisic acid: studies on abscisic acid deficient genotypes of Arabidopsis thaliana (L.) Heynh. *Planta*, 157(2):158–165.

Katavic, V., Reed, D., Taylor, D., Giblin, E., Barton, D., Zou, J., Mackenzie, S., Covello, P., and Kunst, L. (1995). Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity. *Plant Physiol.*, 108(1):399–409.

Kates, M. (1955). Hydrolysis of lecithin by plant plastid enzymes. *Can. J. Biochem. Physiol.*, 33:575–589.

Kennedy, E. (1961). Biosynthesis of complex lipids. *Fed. Proc. Am. Soc. Exp. Biol.*, 20:934–940.

Kerbarle, P. (2000). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J. Mass Spectrom.*, 35(7):804–817.

Kerbarle, P. and Peshke, M. (2000). On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Fresenius J. Anal. Chem.*, 406:11–35.

Kim, H., Li, Y., and Huang, A. (2005). Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in Arabidopsis. *Plant Cell*, 17(4):1073–1089.

Kim, J., Cheon, S., Park, S., Song, Y., and Kim, J. (2000). Serum induced hypha formation in the dimorphic yeast Yarrowia lipolytica. *FEMS Microbiol. Lett.*, 190:9–12.

Kinjarde, S. and Pant, A. (2002). Emulsifier from a tropical marine yeast, Yarrowia lipolytica NCIM 3589. *J. Basic Microbiol.*, 42(1):67–73.

Kinney, A. (1994). Genetic modification of the storage lipids of plants. *Curr. Opin. Biotechnol.*, 5:144.

Knutzon, D., Thompson, G., Radke, S., Johnson, W., Knauf, V., and Kridl, J. (1992). Modification of Brassica seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. *Proc. Natl. Acad. Sci. USA*, 89(7):2624–2628.

Koch, G. (2005). The endoplasmic reticulum and calcium storage. *Bioessays*, 11(527-531).

Koch, G., Smith, M., Macer, D., Webster, P., and Mortara, R. (1986). Endoplasmic reticulum contains a common, abundant calcium-binding glycoprotein, endoplasmin. *J. Cell Sci.*, 86(1):217–232.

Kohlwein, S. and Paltauf, F. (1984). Uptake of fatty acids by the yeasts Saccharomyces uvarum and Saccharomycopsis lipolytica. *Biochim. Biophys. Acta*, 792(3):310–317.

Koller, A., Washburn, M., Lange, B., Andon, N., Deciu, C., Haynes, P., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., and Yates, J. (2002). Protein survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. USA*, 99(18):11969–11974.

Kornfeld, R. and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.*, 54:631–664.

Kroon, J., Wei, W., Simon, J., and Slabas, A. (2006). Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of plants and animals. *Phytochemistry*, 67(23):2541–2549.

Kyte, J. and Doolittle, R. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157(1):105–132.

Lacey, D. and Hills, M. (1999). Heterogeneity of the endoplasmic reticulum with respect to lipid synthesis in developing seeds of Brassica napus. *Phytochemistry*, 50(4):915–917.

Laemmli, U. (1970a). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature*, 22(5259):680–685.

Laemmli, U. (1970b). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–685.

Lander, E., Linton, L., Birren, B., Nusbaum, C., Zody, M., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough,

R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R., Wilson, R., Hillier, L., McPherson, J., Marra, M., Mardis, E., Fulton, L., Chinwalla, A., Pepin, K., Gish, W., Chissoe, S., Wendl, M., Delehaunty, K., Miner, T., Delehaunty, A., Kramer, J., Cook, L., Fulton, R., Johnson, D., Minx, P., Clifton, S., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R., Muzny, D., Scherer, S., Bouck, J., Sodergren, E., Worley, K., Rives, C., Gorrell, J., Metzker, M., Naylor, S., Kucherlapati, R., Nelson, D., Weinstock, G., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R., Federspiel, N., Abola, A., Proctor, M., Myers, R., Schmutz, J., Dickson, M., Grimwood, J., Cox, D., Olson, M., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G., Athanasiou, M., Schultz, R., Roe, B., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W., de la Bastide, M., Dedhia, N., Blcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D., Burge, C., Cerutti, L., Chen, H., Church, D., Clamp, M., Copley, R., Doerks, T., Eddy, S., Eichler, E., Furey, T., Galagan, J., Gilbert, J., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L., Jones, T., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W., Kitts, P., Koonin, E., Korf, I., Kulp, D., Lancet, D., Lowe, T., McLysaght, A., Mikkelsen, T., Moran, J., Mulder, N., Pollara, V., Ponting, C., Schuler, G., Schultz, J., Slater, G., Smit, A., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y., Wolfe, K., Yang, S., Yeh, R., Collins, F., Guyer, M., Peterson, J., Felsenfeld, A., Wetterstrand, K., Patrinos, A., Morgan, M., de Jong, P., Catanese, J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y., and Szustakowki, J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409:860–921.

Lardizabal, K., Mai, J., Wagner, N., Wyrick, A., Voelker, T., and Hawkins, D. (2001). DGAT2 is a new diacylglycerol acyltransferase gene family: purification, cloning and expression in insect cells of two polypeptides from Mortierella ramanniana with diacylglycerol acyltransferase activity. *J. Biol. Chem.*, 276(42):38862–38869.

Laureti, D., Fedeli, A., Scarpa, G., and Marras, G. (1998). Performance of castor

(Ricinus communis L.) cultivars in Italy. *Industrial Crops and Products*, 7(2):91–93.

Lee, Y., Tang, C.-J., and Litzinger, T. (1998). A triple quadrupole mass spectrometer system for studies of gas-phase combustion chemistry of energetic materials. *Meas. Sci. Technol.*, 9:1576–1586.

Lehner, R. and Kuksis, A. (1996). Biosynthesis of triacylgycerols. *Prog. Lipid Res.*, 35:169–201.

Lei, Z., Elmer, A., Watson, B., Dixon, R., Mendes, P., and Sumner, L. (2005). A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume Medicago truncatula. *Mol. Cell. Proteomics*, 4(11):1812–1825.

Li, H. and Chye, M. (2003). Membrane localization of Arabidopsis acyl-CoA binding protein ACBP2. *Plant Mol. Biol.*, 51(4):483–492.

Lin, X., Kaul, S., Rounsley, S., Shea, T., Benito, M., Town, C., Fujii, C., Mason, T., Bowman, C., Barnstead, M., Feldblyum, T., Buell, C., Ketchum, K., Lee, J., Ronning, C., Koo, H., Moffat, K., Cronin, L., Shen, M., Pai, G., Van Aken, S., Umayam, L., Tallon, L., Gill, J., Adams, M., Carrera, A., Creasy, T., Goodman, H., Somerville, C., Copenhaver, G., Preuss, D., Nierman, W., White, O., Eisen, J., Salzberg, S., Fraser, C., and Venter, J. (1999). Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. *Nature*, 402:761–768.

Lindqvist, Y., Huang, W., Schneider, G., and Shanklin, J. (1996). Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *EMBO J.*, 15(16):4081–4092.

Lindsey, K., Pullen, M., and Topping, J. (2003). Importance of plant sterols in pattern formation and hormone signalling. *Trends Plant Sci.*, 8(11):521–525.

Loo, J., Edmonds, C., and Smith, R. (1993). Tandem mass spectrometry of very large molecules. 2. dissociation of multiply charged proline-containing proteins from electrospray ionization. *Anal. Chem.*, 65(4):425–438.

Lopez, M., Breggen, K., Robinson, M., and Patton, W. (2000). A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis*, 21:3673–3683.

Lu, C., Wallis, J., and Browse, J. (2007). An analysis of expressed sequence tags of developing castor endosperm using a full-length cDNA library. *BMC Plant Biol.*, 7(42):1–9.

Lycett, G., Croy, R., Shirsat, A., and Boulter, D. (1984). The complete nucleotide sequence of a legumin gene from pea (Pisum sativum L.). *Nucleic Acids Res.*, 12(11):4493–4506.

Maltman, D., Gadd, S., Simon, W., and Slabas, A. (2007). Differential proteomic analysis of the endoplasmic reticulum from developing and germinating seeds of castor (Ricinus communis) identifies seed protein precursors as significant components of the endoplasmic reticulum. *Proteomics*, 7(9):1513–1528.

Maltman, D., Simon, W., Wheeler, C., Dunn, M., Wait, R., and Slabas, A. (2002). Proteomic analysis of the endoplasmic reticulum from developing and germinating seed of castor (Ricinus communis). *Electrophoresis*, 23(4):626–639.

Man, W., Miyazaki, M., Chu, K., and Ntambi, J. (2006). Colocalization of SCD1 and DGAT2: implying preference for endogenous monounsaturated fatty acids in triglyceride synthesis. *J. Lipid Res.*, 47:1928–1939.

Mann, M. and Wilm, M. (1994). Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.*, 66:4390–4399.

Maor, R., Jones, A., Nuhse, T., Studholme, D., Peck, S., and Shirasu, K. (2007). Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Mol. Cell. Proteomics*, 6(4):601–610.

Matera, A., Terns, R., and Terns, M. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.*, 8(3):209–220.

Mauersberger, S., Ohkuma, M., Schunck, W., and Takagi, M. (1996). Candida maltosa. In Wolf, K., editor, *Nonconventional Yeasts in Biotechnology, a Handbook*, pages 411–580. Springer-Verlag.

Mayer, K., Schller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Dsterhft, A., Stiekema, W., Entian, K., Terryn, N., Harris, B., Ansorge, W., Brandt, P., Grivell, L., Rieger, M., Weichselgartner, M., de Simone, V., Obermaier, B., Mache, R., Mller, M., Kreis, M., Delseny, M., Puigdomenech, P., Watson, M., Schmidtheini, T., Reichert, B., Portatelle, D., Perez-Alonso, M., Boutry, M., Bancroft, I., Vos, P., Hoheisel, J., Zimmermann, W., Wedler, H., Ridley, P., Langham, S., McCullagh, B., Bilham, L., Robben, J., Van der Schueren, J., Grymonprez, B., Chuang, Y., Vandenbussche, F., Braeken, M., Weltjens, I., Voet, M., Bastiaens, I., Aert, R., Defoor, E., Weitzenegger, T., Bothe, G., Ramsperger, U., Hilbert, H., Braun, M., Holzer, E., Brandt, A., Peters, S., van Staveren, M., Dirske, W., Mooijman, P., Klein Lankhorst, R., Rose, M., Hauf, J., Ktter, P., Berneiser, S., Hempel, S., Feldpausch, M., Lamberth, S., Van den Daele, H., De Keyser, A., Buysshaert, C., Gielen, J., Villarroel, R., De Clercq, R., Van Montagu, M., Rogers, J., Cronin, A., Quail, M., Bray-Allen, S., Clark, L., Doggett, J., Hall, S., Kay, M., Lennard, N., McLay, K., Mayes, R., Pettett, A., Rajandream, M., Lyne, M., Benes, V.,

Rechmann, S., Borkova, D., Blcker, H., Scharfe, M., Grimm, M., Lhnert, T., Dose, S., de Haan, M., Maarse, A., Schfer, M., Mller-Auer, S., Gabel, C., Fuchs, M., Fartmann, B., Granderath, K., Dauner, D., Herzl, A., Neumann, S., Argiriou, A., Vitale, D., Liguori, R., Piravandi, E., Massenet, O., Quigley, F., Clabauld, G., Mndlein, A., Felber, R., Schnabl, S., Hiller, R., Schmidt, W., Lecharny, A., Aubourg, S., Chefdor, F., Cooke, R., Berger, C., Montfort, A., Casacuberta, E., Gibbons, T., Weber, N., Vandenbol, M., Bargues, M., Terol, J., Torres, A., Perez-Perez, A., Purnelle, B., Bent, E., Johnson, S., Tacon, D., Jesse, T., Heijnen, L., Schwarz, S., Scholler, P., Heber, S., Francs, P., Bielke, C., Frishman, D., Haase, D., Lemcke, K., Mewes, H., Stocker, S., Zaccaria, P., Bevan, M., Wilson, R., de la Bastide, M., Habermann, K., Parnell, L., Dedhia, N., Gnoj, L., Schutz, K., Huang, E., Spiegel, L., Sehkon, M., Murray, J., Sheet, P., Cordes, M., Abu-Threideh, J., Stoneking, T., Kalicki, J., Graves, T., Harmon, G., Edwards, J., Latreille, P., Courtney, L., Cloud, J., Abbott, A., Scott, K., Johnson, D., Minx, P., Bentley, D., Fulton, B., Miller, N., Greco, T., Kemp, K., Kramer, J., Fulton, L., Mardis, E., Dante, M., Pepin, K., Hillier, L., Nelson, J., Spieth, J., Ryan, E., Andrews, S., Geisel, C., Layman, D., Du, H., Ali, J., Berghoff, A., Jones, K., Drone, K., Cotton, M., Joshu, C., Antonoiu, B., Zidanic, M., Strong, C., Sun, H., Lamar, B., Yordan, C., Ma, P., Zhong, J., Preston, R., Vil, D., Shekher, M., Matero, A., Shah, R., Swaby, I., O'Shaughnessy, A., Rodriguez, M., Hoffmann, J., Till, S., Granat, S., Shohdy, N., Hasegawa, A., Hameed, A., Lodhi, M., Johnson, A., Chen, E., Marra, M., Martienssen, R., and McCombie, W. (1999). Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. *Nature*, 402:769–777.

McCaskill, D. and Croteau, R. (1998). Some caveats for bioengineering terpenoid metabolism in plants. *Trends in Biotechnology*, 16(8):349–355.

McDonnell, M., Bailey, S., Griffiths, G., and Watts, P. (2000). Characterization of ricin heterogeneity by electrospray mass spectrometry, capillary electrophoresis, and resonant mirror. *Anal. Biochem.*, 279(1):23–36.

McKay, I. (1992). Growth of fermentative and non-fermentative yeasts in natural yoghurt, stored in polystyrene cartons. *Int. J. Food Microbiol.*, 15:383–388.

McKeon, T. (2005). Genetic modification of seed oils. In Erhan, S., editor, *Industrial uses of vegetable oils*, pages 1–13. AOCS Press.

McKeon, T., Lin, J., Goodrich-Tanrikulu, M., and Stafford, A. (1997). Ricinoleate biosynthesis in castor microsomes. *Industrial Crops and Products*, 6(3-4):383–389.

McKeon, T., Lin, J.-T., and Chen, G. (2002). Developing a safe source of castor oil. *Inform - Industrial Oils*, 13:381–385.

Meesters, P. and Eggink, G. (1996). Isolation and characterization of a delta-9 fatty acid desaturase gene from the oleaginous yeast Cryptococcus curvatus CBS 570. *Yeast*, 12(8):723–730.

Meldolesi, J. and Pozzan, T. (1998). The heterogeneity of ER Ca2+ stores has a key role in nonmuscle cell signaling and function. *J. Cell Biol.*, 142(6):1395–1398.

Melnick, J., Dul, J., and Argon, Y. (1994). Sequential interaction of the chaperones BiP and GRP94 with immunoglobulin chains in the endoplasmic reticulum. *Nature*, 370(373-375).

Mhaske, V., Beldjilali, K., Ohlrogge, J., and Pollard, M. (2005). Isolation and characterisation of an Arabidopsis thaliana knockout line for phospholipid:diacylglycerol transacylase gene (At5g13640). *Plant Physiol.*, 43(413-417).

Michalak, M., Milner, R., Burns, K., and Opas, M. (1992). Calreticulin. *Biochem. J.*, 285(3):681–692.

Minglin, L., Yuxiu, Z., and Tuanyao, C. (2005). Identification of genes up-regulated in response to Cd exposure in Brassica juncea L. *Gene*, 363:151–158.

Miquel, M. and Browse, J. (1994). High-oleate oilseeds fail to develop at low temperature. *Plant Physiol.*, 106:421.

Mlícková, K., Roux, E., Athenstaedt, K., d'Andrea, S., Daum, G., Chardot, T., and Nicaud, J. (2004). Lipid accumulation, lipid body formation, and acyl coenzyme A oxidases of the yeast Yarrowia lipolytica. *Appl. Envion. Microbiol.*, 70(7):3918–3924.

Mlickova, K., Roux, E., Athenstaedt, K., d'Andrea, S., Daum, G., Chardot, T., and Nicaud, J.-M. (2004). Lipid accumulation, lipid body formation, and acyl coenzyme A oxidases of the yeast Yarrowia lipolytica. *Appl. Environ. Microbiol.*, 70(7):3918–3924.

Moire, L., Rezzonico, E., Goepfert, S., and Poirier, Y. (2004). Impact of unusual fatty acid synthesis on futile cycling through beta-oxidation and on gene expression in transgenic plants. *Plant Physiol.*, 134:432–442.

Molloy, M. (2000). Two-dimensional electrophoresis of membrane proteins using immobilized ph gradients. *Anal. Biochem.*, 280:1–10.

Molloy, M., Herbert, B., Walsh, B., Tyler, M., Traini, M., Sanchez, J., Hochstrasser, D., Williams, K., and Gooley, A. (1998). Extraction of membrane proteins by differential solubilization for separation using two-dimensional electrophoresis. *Electrophoresis*, 19:837–844.

Morris, L. (1967). The mechanism of ricinoleic acid biosynthesis in Ricinus communis seeds. *Biochem. Biophys. Res. Commun.*, 17:311–315.

Müllner, H. and Daum, G. (2004). Dynamics of neutral lipid storage in yeast. *Acta Biochimica Polonica*, 51(2):323–347.

Nakamura, Y., Tsuchiya, M., and Ohta, H. (2007). Plastidic phosphatidic acid phosphatases identified in a distinct subfamily of lipid phosphate phosphatases with prokaryotic origin. *J. Biol. Chem.*, 282:29013–29021.

Ndimba, B., Chivasa, S., Simon, W., and Slabas, A. (2005). Identification of Arabidopsis salt and osmotic stress responsive proteins using two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics*, 5(6):4185–4196.

Nielsen, N., Dickinson, C., Cho, T., Thanh, V., Scallon, B., Fischer, R., Sims, T., Drews, G., and Goldberg, R. (1989). Characterization of the glycinin gene family in soybean. *Plant Cell*, 1(3):313–328.

Novak-Hofer, I. and Siegenthaler, P. (1977). Two-dimensional separation of chloroplast membrane proteins by isoelectric focusing and electrophoresis in sodium dodecyl sulphate. *Biochim. Biophys. Acta*, 468(3):461–471.

Oeklers, P., Cromley, D., Padamsee, M., Billheimer, J., and Sturley, S. (2002). The DGA1 gene determines a second triglyceride synthetic pathway in yeast. *J. Biol. Chem.*, 277:8877–8881.

Oeklers, P., Tinelenberg, A., Erdeniz, N., Cromley, D., Billheimer, J., and Sturley, S. (2000). A lecithin cholesterol acyltransferase-like gene mediates diacylglycerol esterification in yeast. *J. Biol. Chem.*, 275:15609–15612.

Oelkers, P., Behari, A., Cromley, D., Billheimer, J., and Sturley, S. (1998). Characterization of two human genes encoding acyl coenzyme A:cholesterol acyltransferase-related enzymes. *J. Biol. Chem.*, 273:26765–26771.

O'Farrel, P. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, 250:4007–4021.

O'Farrel, P., Goodman, H., and O'Farrel, P. (1977). High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell*, 12:1133–1141.

Ohlrogge, J. and Jaworski, J. (1997). Regulation of fatty acid synthesis. *Annu. Rev. Plant Physiol. Mol. Biol.*, 48:109–136.

Okorokov, L., Silva, F., and Okorokova Facanha, A. (2001). Ca^{2+} and H^+ homeostasis in fission yeast: a role of Ca^{2+}/H^+ exchange and distinct V-H^+-ATPases of the secretory pathway organelles. *FEBS Lett.*, 505(2):321–324.

Oo, K., Chew, Y., and Ong, A. (1989). In Biacs, P., Gruiz, K., and Kremmer, T., editors, *Biological role of plant lipids*, pages 93–94. Akademial Kiado.

Pacheco-Moises, F., Valencia-Turcotte, L., Altuzar-Martinez, M., and Rodriguez-Sotres, R. (1997). Regulation of acyltransferase activity in immature maize embryos by abscisic acid and the osmotic environment. *Plant Physiol.*, 114(3):1095–1101.

Pandey, A. and Mann, M. (2000). Proteomics to study genes and genomes. *Nature*, 405(6788):837–846.

Papanikolaou, S., Muniglia, L., Chevalot, I., Aggelis, G., and Marc, I. (2003). Accumulation of a cocoa-butter-like lipid by Yarrowia lipolytica cultivated on agro-industrial residues. *Curr. Microbiol.*, 46(1):124–130.

Pappin, D., Horjup, P., and Bleasby, A. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.*, 3(6):327–332.

Pardo, M., Ward, M., Pitrach, A., Sanchez, M., Nombela, C., Blackstock, W., and Gil, C. (2000). Cross-species identification of novel Candida albicans immunogenic proteins by combination of two-dimensional polyacrylamide gel electrophoresis and mass spectrometry. *Electrophoresis*, 21(13):2651–2659.

Patterson, S. and Aebersold, R. (2003). Proteomics: the first decade and beyond. *Nat. Genet.*, 33:311–323.

Patterson, S., Aebersold, R., and Goodlet, D. (2001). Mass spectrometry-based methods for protein identification and phosphorylation site analysis. In Pennington, S. and Dunn, M., editors, *Proteomics. From protein sequence to function*, pages 87–130. Bios Scientific.

Pearce, M. and Slabas, A. (1998). Phosphatidate phosphatase from avocado (Persea americana) - purification, substrate selectivity and possible metabolic implications for the Kennedy pathway and cell signalling in plants. *Plant J.*, 14(5):555–564.

Perdew, G., Schaup, H., and Selivonchick, D. (1983). The use of a zwitterionic detergent in two-dimensional gel electrophoresis and proteomics. *Anal. Biochem.*, 135:453–455.

Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567.

Peterson, J., Ora, A., Van, P., and Helenius, A. (1995). Transient, lectin-like association of calreticulin with folding intermediates of cellular and viral glycoproteins. *Mol. Gen. Biol.*, 6(9):1173–1184.

Pierrugues, O., Brutesco, C., Oshiro, J., Gouy, M., Deveaux, Y., Carman, G., Thuriaux, P., and Kazmaier, M. (2001). Lipid phosphate phosphatases in Arabidopisis: Regulation of the ArLPP1 gene in response to stress. *J. Biol. Chem.*, 276(23):20300–20308.

Pollard, M., Anderson, L., Fan, C., Hawkins, D., and Davies, H. (1991). A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of Umbellularia californica. *Arch. Biochem. Biophys.*, 284:306–312.

Rabilloud, T., Adessi, C., Giraudel, A., and Lunardi, J. (1997). Improvement of the solubilization of proteins in two-dimensional electrphoresis with immobilized pH gradients. *Electrophoresis*, 18:307–316.

Rabilloud, T. and Chevallet, M. (2000). Solubilisation of proteins in two-dimensional electrophoresis. In *Two-dimensional gel electrophoresis and identification methods*. Springer.

Ramagli, L., Capetillo, S., Becker, F., and Rodriguez, L. (1985). Alterations in nonhistone chromatin proteins during hepatocarcinogenesis induced by diverse acting carcinogens. *Carcinogenesis*, 6(3):367–375.

Ramli, U., Salas, J., Quant, P., and Harwood, J. (2005). Metabolic control analysis reveals an important role for diacylglycerol acyltransfase in olive but not in oil palm lipid accumulation. *FEBS J.*, 272:5764–5770.

Ratledge, C. (1982). Microbial oils and fat: An assessment of their commercial potential. *Prog. Ind. Microbiol.*, 16:119–206.

Ratledge, C. (1984). Microbial conversions of alkanes and fatty acids. *J. Am. Oil Chem. Soc.*, 61(2):447–453.

Renkonen, O. (1966). Individual molecular species of phospholipids III. molecular species of ox-brain lecithins. *Biochim. Biophys. Acta*, 125:288–309.

Richards, O. (1928). The growth of the yeast saccharomyces cerevisiae. *Annals of Botany*, 42(165):271–283.

Righetti, P. (1990). Recent developments in electrophoretic methods. *J. Chomatography*, 7:516.

Righetti, P., Castagna, A., Antonioli, P., and Boschetti, E. (2005). Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis*, 26(2):297–319.

Roberts, L., Lamb, F., Pappin, D., and Lord, J. (1985). The primary sequence of Ricinus communis agglutinin. Comparison with ricin. *J. Biol. Chem.*, 260:15682–15686.

Rock, C. and Cronan, J. (1996). Escherichia coli as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim. Biophys. Acta*, 1302:1–16.

Rodriguez-Sotres, R. and Black, M. (1993). Osmotic potential and abscisic acid regulate triacylglycerol synthesis in developing wheat embryos. *Planta*, 192(1):9–15.

Roetheli, J., Glaser, L., and Brigham, R. (1991). Castor: Assessing the feasibility of U.S. production. In *Workshop Summary*, pages 18–19. USDA Office of Agr. Materials. Growing Ind. Material Ser.

Roostita, R. and Fleet, G. (1996). The occurence and growth of yeasts in Camembert and blue-veined cheeses. *Int. J. Food Microbiol.*, 28:393–404.

Ross, P., Huang, Y., Marchese, J., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Barlet-Jones, M., He, F., Jacobson, A., and Pappin, D. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics*, 3:1154–1169.

Rowland, J. (2006). *Analysis of the androgen-regulated proteome in human prostate cancer*. PhD thesis, University of Newcastle.

Rowland, J., Robson, J., Simon, W., Leung, H., and Slabas, A. (2007). Evaluation of an in vitro model of androgen ablation and identification of the androgen responsive proteome in LNCaP cells. *Proteomics*, 7(1):47–63.

Ruiz-Herrera, J. and Sentandreu, R. (2002). Different effectors of dimorphism in Yarrowia lipolytica. *Arch. Microbiol.*, 178(6):477–483.

Rutschow, H., Ytterberg, A., Friso, G., Nilsson, R., and van Wijk, K. (2008). Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in Arabiopsis thaliana. *Plant Physiol.*, 148(1):156–175.

Sabounchi-Schutt, F., Astrom, J., Olsson, I., Eklund, A., Grunewald, J., and Bjellqvist, B. (2000). An immobiline DryStrip application method enabling high-capacity two-dimensional gel electrophoresis. *Electrophoresis*, 21(17):3649–3656.

Saint-Jore-Dupas, C., Nebenführ, A., Boulaflous, A., Follet-Gueye, M., Plasson, C., Hawes, C., Driouich, A., Faye, L., and Gomord, V. (2006). Plant N-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. *Plant Cell*, 18(11):3182–3200.

Sakaguchi, M., Hachiya, N., Mihara, K., and Omura, T. (1992). Mitochondrial porin can be translocated across both endoplasmic reticulum and mitochondrial membranes. *J. Biochem.*, 112(2):243–248.

Sakamotoa, H., Ukenaa, K., Takemorid, H., Okamotod, M., Kawatac, M., , and Tsutsui, K. (2004). Expression and localization of 25-Dx, a membrane-associated putative progesterone-binding protein, in the developing Purkinje cell. *Neuroscience*, 126(2):325–334.

Salanoubat, M., Lemcke, K., Rieger, M., Ansorge, W., Unseld, M., Fartmann, B., Valle, G., Blcker, H., Perez-Alonso, M., Obermaier, B., Delseny, M., Boutry, M., Grivell, L., Mache, R., Puigdomnech, P., De Simone, V., Choisne, N., Artiguenave, F., Robert, C., Brottier, P., Wincker, P., Cattolico, L., Weissenbach, J., Saurin, W., Qutier, F., Schfer, M., Mller-Auer, S., Gabel, C., Fuchs, M., Benes, V., Wurmbach, E., Drzonek, H., Erfle, H., Jordan, N., Bangert, S., Wiedelmann, R., Kranz,

H., Voss, H., Holland, R., Brandt, P., Nyakatura, G., Vezzi, A., D'Angelo, M., Pallavicini, A., Toppo, S., Simionati, B., Conrad, A., Hornischer, K., Kauer, G., Lhnert, T., Nordsiek, G., Reichelt, J., Scharfe, M., Schn, O., Bargues, M., Terol, J., Climent, J., Navarro, P., Collado, C., Perez-Perez, A., Ottenwlder, B., Duchemin, D., Cooke, R., Laudie, M., Berger-Llauro, C., Purnelle, B., Masuy, D., de Haan, M., Maarse, A., Alcaraz, J., Cottet, A., Casacuberta, E., Monfort, A., Argiriou, A., flores, M., Liguori, R., Vitale, D., Mannhaupt, G., Haase, D., Schoof, H., Rudd, S., Zaccaria, P., Mewes, H., Mayer, K., Kaul, S., Town, C., Koo, H., Tallon, L., Jenkins, J., Rooney, T., Rizzo, M., Walts, A., Utterback, T., Fujii, C., Shea, T., Creasy, T., Haas, B., Maiti, R., Wu, D., Peterson, J., Van Aken, S., Pai, G., Militscher, J., Sellers, P., Gill, J., Feldblyum, T., Preuss, D., Lin, X., Nierman, W., Salzberg, S., White, O., Venter, J., Fraser, C., Kaneko, T., Nakamura, Y., Sato, S., Kato, T., Asamizu, E., Sasamoto, S., Kimura, T., Idesawa, K., Kawashima, K., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakayama, S., Nakazaki, N., Shinpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (2000). Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana. *Nature*, 408:820–822.

Salas, J., Sanchez, J., Ramli, U., Manaf, A., Williams, M., and Harwood, J. (2000). Biochemistry of lipid metabolism in olive and other oil fruits. *Prog. Lipid Res.*, 39:151–180.

Sandager, L., Gustavsson, M., Ståhl, U., Dahlqvist, A., Wiberg, E., Banas, A., Lenman, M., Ronne, H., and Stymne, S. (2002). Storage lipid synthesis is non-essential in yeast. *J. Biol. Chem.*, 277:6478–6482.

Sandager, M., Dahlqvist, A., Banas, A., Ståhl, U., Lenman, M., Gustavsson, M., and Stymne, S. (2000). An acyl-CoA:cholesterol acyltransfease (ACAT)-related gene is involved in the accumulation of triacylglycerols in Saccharomyces cerevisiae. *Biochemical Society Transactions*, 28:700–702.

Sanger, F., Air, G., Barrell, B., Brown, N., Coulson, A., Fiddes, C., Hutchison, C., Slocombe, P., and Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 265(5596):687–695.

Santoni, V., Doumas, P., Rouquie, D., Mansion, M., Rabilloud, T., and Rossignol, M. (1999). Large scale characterization of plant plasma membrane proteins. *Biochimie*, 81:655–661.

Santoni, V., Molloy, M., and Rabilloud, T. (2000). Membrane proteins and proteomics: Un amour impossible? *Electrophoresis*, 21:1054–1070.

Sasaki, Y. and Nagano, Y. (2004). Plant acetyl-CoA carboxylase: Structure, biosynthesis, regulation and gene manipulation for plant breeding. *Biosci. Biotechnol. Biochem.*, 68(6):1175–1184.

Schagger, H. (2001). Blue-native gels to isolate protein complexes from mitochondria. *Methods Cell Biol.*, 65:231–244.

Schagger, H. and von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.*, 199(2):223–231.

Shanklin, J. and Somerville, C. (1991). Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. *Proc. Natl. Acad. Sci. USA*, 88:2510–2514.

Sharief, F. and Li, S. (1982). Amino acid sequence of small and large subunits of seed storage protein from Ricinus communis. *J. Biol. Chem.*, 24(14753-14759).

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996a). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.*, 65(5):850–858.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996b). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.*, 68(5):850–858.

Shimakata, T. and Stumpf, P. (1982). Purification and characterizations of beta-Ketoacyl-[acyl-carrier-protein] reductase, beta-hydroxyacyl-[acyl-carrier-protein] dehydrase, and enoyl-[acyl-carrier-protein] reductase from Spinacia oleracea leaves. *Arch. Biochem. Biophys.*, 218:77–91.

Shockey, J., Fulda, M., and Browse, J. (2002). Arabidopsis contains nine long-chain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerollipid metabolism. *Plant Physiol.*, 129:1710–1722.

Shockey, J., Fulda, M., and Browse, J. (2003). Arabidopsis contains a superfamily of acyl-activatiing enzymes. Phylogentic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. *Plant Physiol.*, 132:1065–1076.

Shockey, J., Gidda, S., Chapital, D., Kuan, J., Dhanoa, P., Bland, J., Rothstein, S., Mullen, R., and Dyer, J. (2006). Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell*, 18:2294–2313.

Shoshan-Barmatz, V., Zalk, R., Gincel, D., and Vardi, N. (1994). Subcellular localization of VDAC in mitochondria and ER in the cerebellum. *Biochim. Biophys. Acta*, 1657(2-3):105–114.

Shotwell, M. and Larkins, B. (1989). *The biochemistry of plants : a comprehensive treatise. Molecular biology*, volume 15. Academic Pr.

Simcox, P., Garland, W., De Luca, V., Canvin, D., and Dennis, D. (1979). Respiratory pathways and fat synthesis in the developing castor oil seed. *Can. J. Bot.*, 57:1008–1014.

Singh, S., Thomaeus, S., Lee, M., Stymne, S., and Green, A. (2001). Transgenic expression of a delta12-expoygenase gene in Arabidopsis seeds inhibits accumulation of linoleic acid. *Planta*, 212:872–879.

Slabas, A., Simon, J., and Brown, A. (2001). Biosynthesis and regulation of fatty acids and triglycerides in oil seed rape. Current status and future trends. *Eur. J. Lipid Sci. Technol.*, 103:455–466.

Slabas, A., Simon, J., Schierer, T., Kroon, J., Fawcett, T., Hayman, M., Gilroy, J., Nishida, I., Murata, N., Rafferty, J., Turnbull, A., and Rice, D. (2000). Plant glycerol-3-phosphate-1-acyltransferase (GPAT): structure selectivity studies. *Biochemical Society Transactions*, 28(6):677–679.

Slack, C., Roughan, P., Browse, J., and Gardiner, S. (1985). Some properties of cholinephosphotransferase from developing safflower cotyledons. *Biochim. Biophys. Acta*, 833:10–20.

Smit, M., Mokgoro, M., Setati, E., and Nicaud, J. (2005). alpha,omega-dicarboxylic acid accumulation by acyl-CoA oxidase deficient mutants of Yarrowia lipolytica. *Biotechnol. Lett.*, 27(12):859–864.

Smith, M., Jonsson, L., Stymne, S., and Stobart, K. (1992). Evidence for cytochrome b5 as an electron donor in ricinoleic acid biosynthesis in microsomal preparations from developing castor bean (Ricinus communis L.). *Biochem. J.*, 287:141–144.

Smith, M., Moon, H., Chowrira, G., and Kunst, L. (2003). Heterologous expression of a fatty acid hydroxylase gene in developing seeds of Arabidopsis thaliana. *Planta*, 217:507–516.

Sorger, D. and Daum, G. (2002). Synthesis of triacylglycerols by the acyl-coenzyme A:diacyl-glycerol acyltransferase Dga1p in lipid particles of the yeast Saccharomyces cerevisiae. *J. Bacteriol.*, 184:519–524.

Ståhl, U., Carlsson, A., Lenman, M., Dahlqvist, A., Huang, B., Banas, A., Banas, A., and Stymne, S. (2004). Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from Arabidopsis. Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from Arabidopsis. *Plant Physiol.*, 135:1324–1335.

Stalling, D., Gehrke, C., and Zumwalt, R. (1968). A new silylation reagent for amino acids bis(trimethylsilyl)trifluoroacetamide (BSTFA). *Biochem. Biophys. Res. Comm.*, 31(4):616.

Staswick, P., Hermodson, M., and Nielsen, N. (1981). Identification of the acidic and basic subunit complexes of glycinin. *J. Biol. Chem.*, 256(16):8752–8755.

Staswick, P., Hermodson, M., and Nielsen, N. (1984). The amino acid sequence of the A2Bla subunit of glycinin. *J. Biol. Chem.*, 259(21):13431–13435.

Steinberg, T., Haugland, R., and Singer, V. (1996a). Applications of SYPRO orange and SYPRO red protein gel stains. *Anal. Biochem.*, 239(2):238–245.

Steinberg, T., Jones, L., Haugland, R., and Singer, V. (1996b). SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal. Biochem.*, 239(2):223–237.

Stobart, K., Mancha, M., Lenman, M., Dahlqvist, A., and Stymne, S. (1997). Triacylglycerols are synthesized and utilized by transacylation reactions in microsomal preparations of developing safflower (Carthamus tinctorius L.) seeds. *Planta*, 203:58–66.

Stymne, S. and Stobart, K. (1987). Triacylglycerol biosynthesis. In Stumpf, P. and Conn, E., editors, *The Biochemistry of Plants*, volume 9. Academic Press.

Suzuki, I., Los, D., and Murata, N. (2000). Perception and transduction of low-temperature signals to induce desaturation of fatty acids. *Biochemical Society Transactions*, 28(6):628–630.

Szabadkai, G., Bianchi, K., Várnai, P., De Stefani, D., Wieckowski, M., Cavagna, D., Nagy, A., Balla, T., and Rizzuto, R. (2006). Chaperone-mediated coupling of endoplasmic reticulum and mitochondrial Ca2+ channels. *J. Cell Biol.*, 175(6):901–911.

Szabo, R. (1999). Dimorphism in Yarrowia lipolytica: filament formation is supressed by nitrogen starvation and inhibition of respiration. *Folia Microbiol.*, 44(1):19–24.

Szczesna-Skorupa, E., Chen, C., and Rogers, S. (1998). Mobility of cytochrome p450 in the endoplasmic reticulum membrane. *Proc. Natl. Acad. Sci. USA*, 95(25):14793–14798.

Tabata, S., Kaneko, T., Nakamura, Y., Kotani, H., Kato, T., Asamizu, E., Miyajima, N., Sasamoto, S., Kimura, T., Hosouchi, T., Kawashima, K., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakayama, S., Nakazaki, N., Naruo, K., Okumura, S., Shinpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Sato, S., de la Bastide, M., Huang, E., Spiegel, L., Gnoj, L., O'Shaughnessy, A., Preston, R., Habermann, K., Murray, J., Johnson, D., Rohlfing, T., Nelson, J., Stoneking, T., Pepin, K., Spieth, J., Sekhon, M., Armstrong, J., Becker, M., Belter, E., Cordum, H., Cordes, M., Courtney, L., Courtney, W., Dante, M., Du, H., Edwards, J., Fryman, J., Haakensen, B., Lamar, E., Latreille, P., Leonard, S., Meyer, R., Mulvaney, E., Ozersky, P., Riley,

A., Strowmatt, C., Wagner-McPherson, C., Wollam, A., Yoakum, M., Bell, M., Dedhia, N., Parnell, L., Shah, R., Rodriguez, M., See, L., Vil, D., Baker, J., Kirchoff, K., Toth, K., King, L., Bahret, A., Miller, B., Marra, M., Martienssen, R., McCombie, W., Wilson, R., Murphy, G., Bancroft, I., Volckaert, G., Wambutt, R., Dsterhft, A., Stiekema, W., Pohl, T., Entian, K., Terryn, N., Hartley, N., Bent, E., Johnson, S., Langham, S., McCullagh, B., Robben, J., Grymonprez, B., Zimmermann, W., Ramsperger, U., Wedler, H., Balke, K., Wedler, E., Peters, S., van Staveren, M., Dirkse, W., Mooijman, P., Lankhorst, R., Weitzenegger, T., Bothe, G., Rose, M., Hauf, J., Berneiser, S., Hempel, S., Feldpausch, M., Lamberth, S., Villarroel, R., Gielen, J., Ardiles, W., Bents, O., Lemcke, K., Kolesov, G., Mayer, K., Rudd, S., Schoof, H., Schueller, C., Zaccaria, P., Mewes, H., Bevan, M., and Fransz, P. (2000). Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. *Nature*, 408:823–826.

Takishima, K., Watanabe, S., Yamada, M., and Mamiya, G. (1986). The amino-acid sequence of the non-specific lipid transfer protein from germinated castor bean endosperms. *Biochim. Biophys. Acta*, 870(2):248–255.

Takishima, K., Watanabe, S., Yamada, M., Suga, T., and Mamiya, G. (1988). Amino acid sequences of two nonspecific lipid-transfer proteins from germinated castor bean. *Eur. J. Biochem.*, 177(2):241–249.

Tal, M., Silberstein, A., and Nusser, E. (1985). Why does Coomassie Brilliant Blue R interact differently with different proteins? A partial answer. *J. Biol. Chem.*, 260:9976–9980.

Tan, X., Meyers, B., Kozik, A., West, M., Morgante, M., St Clair, D., Bent, A., and Michelmore, R. (2007). Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. *BMC Plant Biol.*, 7(1):56.

Theologis, A., Ecker, J., Palm, C., Federspiel, N., Kaul, S., White, O., Alonso, J., Altafi, H., Araujo, R., Bowman, C., Brooks, S., Buehler, E., Chan, A., Chao, Q., Chen, H., Cheuk, R., Chin, C., Chung, M., Conn, L., Conway, A., Conway, A., Creasy, T., Dewar, K., Dunn, P., Etgu, P., Feldblyum, T., Feng, J., Fong, B., Fujii, C., Gill, J., Goldsmith, A., Haas, B., Hansen, N., Hughes, B., Huizar, L., Hunter, J., Jenkins, J., Johnson-Hopson, C., Khan, S., Khaykin, E., Kim, C., Koo, H., Kremenetskaia, I., Kurtz, D., Kwan, A., Lam, B., Langin-Hooper, S., Lee, A., Lee, J., Lenz, C., Li, J., Li, Y., Lin, X., Liu, S., Liu, Z., Luros, J., Maiti, R., Marziali, A., Militscher, J., Miranda, M., Nguyen, M., Nierman, W., Osborne, B., Pai, G., Peterson, J., Pham, P., Rizzo, M., Rooney, T., Rowley, D., Sakano, H., Salzberg, S., Schwartz, J., Shinn, P., Southwick, A., Sun, H., Tallon, L., Tambunga, G., Toriumi, M., Town, C., Utterback, T., Van Aken, S., Vaysberg, M., Vysotskaia, V., Walker, M., Wu, D., Yu, G., Fraser, C., Venter, J., and Davis, R. (2000). Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. *Nature*, 408:816–820.

Thomaeus, S., Carlsson, A., and Stymne, S. (2001). Distribution of fatty acids in polar and neutral lipids during seed development in Arabidopsis thaliana genetically engineered to produce acetylenic, epoxy and hydroxy fatty acids. *Plant Sci.*, 161:997–1003.

Thomson, S. and Williams, D. (2005). Delineation of the lectin site of the molecular chaperone calreticulin. *Cell Stress Chaperones*, 10(3):242–251.

Thorpe, S. C., Kemeny, D. M., Panzani, R. C., McGurl, B., and Lord, M. J. (1988). Allergy to castor bean. II. Identification of the major allergens in castor bean seeds. *J. Allergy Clin. Immunol.*, 82(1):62–66.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognam, F., Hawkins, E., Currie, I., and Davison, M. (2001a). Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, 1(3):377–396.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001b). Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, 1(3):377–396.

Tsuboi, S., Suga, T., Takishima, K., Mamiya, G., Matsui, K., Ozeki, Y., and Yamada, M. (1991). Organ-specific occurrence and expression of the isoforms of nonspecific lipid transfer protein in castor bean seedlings, and molecular cloning of a full-length cDNA for a cotyledon-specific isoform. *J. Biochem.*, 110(5):823–831.

Tsuji, A., Yuasa, K., and Matsuda, Y. (2004). Identification of oligopeptidase B in higher plants. Purification and characterization of oligopeptidase B from quiescent wheat embyro, Tritcum aestivum. *J. Biochem.*, 136(5):673–681.

Unlu, M., Morgan, M., and Minden, J. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*, 18(11):2071–2077.

van Anken, E. and Braakman, I. (2005). Versatility of the endoplasmic reticulum protein folding factory. *Crit. Rev. Biochem. Mol. Biol.*, 40(4):191–228.

Van de Loo, F., Broun, P., Turner, S., and Somerville, C. (1995). An oleate 12-hydroxylase from Ricinus communis L. is a fatty acyl desaturase homolog. *Proc. Natl. Acad. Sci. USA*, 92(15):6743–6747.

van de Loo, F., Broun, P., Turner, S., and Somerville, C. (1995). An oleate 12-hydroxylase from Ricinus communis L. is a fatty acyl desaturase homolog. *Proc. Natl. Acad. Sci. USA*, 92(15):6743–6747.

van de Loo, F. and Somerville, C. (1994). Plasmid omega-3 fatty acid desaturase cDNA from Ricinus communis. *Plant Physiol.*, 105(1):443–444.

van Leeuwen, J. and Kearse, K. (1996). Deglucosylation of N-linked glycans is an important step in the dissociation of calreticulin-class I-TAP complexes. *Proc. Natl. Acad. Sci. USA*, 93(24):13997–14001.

Venter, J., Adams, M., Myers, E., Li, P., Mural, R., Sutton, G., Smith, H., Yandell, M., Evans, C., Holt, R., Gocayne, J., Amanatides, P., Ballew, R., Huson, D., Wortman, J., Zhang, Q., Kodira, C., Zheng, X., Chen, L., Skupski, M., Subramanian, G., Thomas, P., Zhang, J., Gabor Miklos, G., Nelson, C., Broder, S., Clark, A., Nadeau, J., McKusick, V., Zinder, N., Levine, A., Roberts, R., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T., Higgins, M., Ji, R., Ke, Z., Ketchum, K., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G., Milshina, N., Moore, H., Naik, A., Narayan, V., Neelam, B., Nusskern, D., Rusch, D., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J., Guig, R., Campbell, M., Sjolander, K., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu,

D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. *Science*, 291:1304–1351.

Vernieri, P., Perata, P., Lorenzi, R., and Ceccarelli, N. (1989). Abscisic acid levels during early seed development in Sechium edule Sw. *Plant Physiol.*, 91(4):1351–1355.

Vigil, E. (1973). Structure and function of plant microbodies. *Subcell. Biochem.*, 2(3):237–285.

Vijay, I. and Stumpf, P. (1972). Fat Metabolism in Higher Plants. XLVIII. Properties of oleoyl Coenzyme A a desaturase of Carthamus tinctorius. *J. Biol. Chem.*, 247(2):360–366.

Waché, Y., Aguedo, M., Choquet, A., Gatfield, I., Nicaud, J.-M., and Belin, J.-M. (2001). Role of β-oxidation enzymes in γ-decalactone production by the yeast Yarrowia lipolytica. *Appl. Environ. Microbiol.*, 67(12):5700–5704.

Waché, Y., Aguedo, M., Nicaud, J.-M., and Belin, J.-M. (2003). Catabolism of hydroxyacids and biotechnological production of lactones by Yarrowia lipolytica. *Appl. Microbiol. Biotechnol.*, 61(5-6):393–404.

Waché, Y., Laroche, C., Bergmark, K., Møller-Andersen, C., Aguedo, M., Le Dall, M.-T., Wang, H., Nicaud, J.-M., and Belin, J.-M. (2000). Involvement of Acyl Coenzyme A Oxidase Isozymes in Biotransformation of Methyl Ricinoleate into γ-Decalactone by Yarrowia lipolytica. *Appl. Environ. Microbiol.*, 66(3):1233–1236.

Walburg, G. and Larkins, B. (1983). Oat seed globulin: Subunit characterization and demonstration of its synthesis as a precursor. *Plant Physiol.*, 72(1):161–165.

Walther-Larsen, H., Brandt, J., Collinge, D. B., and Thordal-Christensen, H. (1993). A pathogen-induced gene of barley encodes a HSP90 homologue showing striking similarity to vertebrate forms resident in the endoplasmic reticulum. *Plant Mol. Biol.*, 21(6):1097–1108.

Wan, X. and Liu, J. (2008). Comparative proteomics analysis reveals an intimate protein network provoked by hydrogen peroxide stress in rice seedling leaves. *Mol. Cell. Proteomics*, 7(8):1469–1488.

Washburn, M., Wolters, D., and Yates, J. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.*, 19(3):242–247.

Wasinger, V., Cordwell, S., Cerpa-Poljak, A., Yan, J., Gooley, A., Wilkins, M., Duncan, M., Harris, R., Williams, K., and Humphery-Smith, I. (1995a). Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis*, 16(7):1090–1094.

Wasinger, V., Cordwell, S., Cerpa-Poljak, A., Yan, J., Gooley, A., Wilkins, M., Duncan, M., Harris, R., Williams, K., and Humphrey-Smith, I. (1995b). Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis*, 16(7):1090–1094.

Watanabe, D., Yamada, K., Nishina, Y., Tajima, Y., Koshimizu, U., Nagata, A., and Nishimune, Y. (1994). Molecular cloning of a novel Ca(2+)-binding protein (calmegin) specifically expressed during male meiotic germ cell development. *J. Biol. Chem.*, 269(10):7744–7749.

Weig, A. and Komor, E. (1992). The lipid-transfer protein C of Ricinus communis L.: isolation of two cDNA sequences which are strongly and exclusively expressed in cotyledons after germination. *Planta*, 187(3):367–371.

Whitelegge, J. (2002). Plant proteomics: BLASTing out of a MudPIT. *Proc. Natl. Acad. Sci. USA*, 99(18):11564–11566.

Wickerham, L. (1966). Validation of the species Pichia guilliermondii. *J. Bacteriol.*, 92(4):1269.

Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fet, S., Larsen, P., and Görg, A. (2000). Towards higher resolution: 2D-Electrophoresis of Saccharomyces cerevisiae proteins using overlapping narrow IPGs. *Electrophoresis*, 21(13):2610–2616.

Wood, R. and Snyder, F. (1966). Gas-liquid chromatographic analysis of long chain isomeric glyceryl monoethers. *Lipids*, 1(1):62–72.

Xu, L., Paulsen, A., Ryu, S., and Wang, X. (1996a). Intracellular localization of phospholipase D in leaves and seedling tissues of castor bean. *Plant Physiol.*, 111(1):101–107.

Xu, L., Zheng, L., Coughlan, S., and Wang, X. (1996b). Structure and analysis of phospholipase D gene from Ricinus communis L. *Plant Mol. Biol.*, 32(4):767–771.

Yang, H., Bard, M., Bruner, D., Gleeson, A., Deckelbaum, R., Aljinovic, G., Pohl, T., Rothstein, R., and Sturley, S. (1996). Sterol esterification in yeast: a two-gene process. *Science*, 272(5266):1353–1356.

Yarmush, M. and Jayaraman, A. (2002). Advances in proteomic technologies. *Annu. Rev. Biomed. Eng.*, 4:349–373.

Yermanos, D. (1975). Composition of jojoba seed during development. *J. Am. Oil Chem. Soc.*, 52(4):115–117.

Youle, R. and Huang, A. (1978). Evidence that the castor bean allergens are the albumin storage proteins in the protein bodies of castor bean. *Plant Physiol.*, 61(6):1040–1042.

Yu, C., Kennedy, N., Change, C., and Rothblatt, J. (1996). Molecular cloning and characterization of two isoforms of Saccharomyces cervesiae acyl-CoA:sterol acyltransferase. *J. Biol. Chem.*, 271(39):24157–24163.

Zheng, Z., Xia, Q., Dauk, M., Shen, W., Selvaraj, G., and Zou, J. (2003). Arabidopsis AtGPAT1, a member of the membrane-bound glycerol-3-phosphate acyltransferase gene family, is essential for tapetum differentiation and male fertility. *Plant Cell*, 15(8):1872–1887.

Zou, J., Wei, Y., Jako, C., Kumar, A., Selvaraj, G., and Taylor, D. (1999). The Arabidopsis thaliana TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene. *Plant J.*, 19(6):645–653.

Zubarev, R., Hakansson, P., and Sundqvist, B. (1996). Accuracy requirements for peptide characterisation by monoisotopic molecular mass measurements. *Anal. Chem.*, 68(22):4060–4063.

Zuo, X. and Speicher, D. (2002). Comprehensive analysis of complex proteomes using microscale solution isoelectrofocusing prior to narrow pH range two-dimensional electrophoresis. *Proteomics*, 2(1):58–68.