NOTE TO USERS

University of Alberta

Proteomic analysis of flax (*Linum usitatissimum* L.) phloem fibres and stem tissues

by

Naomi Sela-Clare Hotte

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Plant Biology

Department of Biological Sciences

Edmonton, Alberta
Fall 2008

Your file Votre référence
ISBN: 978-0-494-47266-8
Our file Notre référence
ISBN: 978-0-494-47266-8

Abstract

To better understand the molecular mechanisms underlying cell wall development in flax (*Linum usitatissimum* L.) bast fibres, proteins from individually dissected phloem fibres (i.e. individual cells) at an early stage of secondary cell wall development, were extracted and compared with protein extracts from surrounding, non-fibre cells, using fluorescent (DiGE) labels and 2D-gel electrophoresis. Identities were assigned to some proteins by mass spectrometry. Phosphoprotein and glycoprotein specific staining were used to investigate post-translational modifications (PTMs). The abundance of many proteins in fibres was notably different from the surrounding non-fibre cells, with approximately 13 % of the 1,850 detectable spots being significantly (>1.5 fold, $p \leq 0.05$) enriched in fibres. A K^+ channel subunit, annexins, secretory pathway components, β-galactosidase and pectin and galactan biosynthetic enzymes were observed to be among the most highly enriched proteins detected in developing flax fibres, with many of these proteins showing electrophoretic patterns consistent with PTMs.

Table of Contents

List of Tables

List of Figures

Chapter 1: Literature review

1.1 Introduction

Flax (*Linum usitatissimum* L.) is a crop valued by historical and modern cultures for both its fibre and oil (van Zeist and Bakker-Heeres, 1975; McCorriston, 1997) and has been cultivated in Canada for almost 400 years (Johnston et al., 2002). The seed from flax, also referred to as linseed, is an oilseed with many diverse uses as both a food rich in omega fatty acids and other beneficial nutrients (Muir and Westcott, 2003; Abbadi et al., 2004), and as an industrial feedstock in the manufacture of linoleum and wood preservatives (Mohanty et al., 2000; Johnston et al., 2002). In reference to flax, the term 'fibre' usually describes the bast fibres, which are located in the phloem tissue of the stem. Flax fibre has held importance, in the past as well as today, as the textile fibre that makes linen; it has additional important uses in the pulp and paper industry and novel uses as part of industrial composites (Arbelaiz et al., 2005; Dimmock et al., 2005).

The usefulness of bast fibre depends largely on the biological characteristics of its constituent cells, such as cell wall thickness and composition, fibre cell strength, length, and number, and the degree of cellulose crystalinity (Dimmock et al., 2005). In relation to phloem fibres produced by other plants, including industrial crops like hemp (Karus and Vogt, 2004; Gutierrez et al., 2006) and jute (Sengupta and Palit, 2004), flax fibres contain one of the lowest amounts of lignin, show the highest degree of crystalinity, and have some of the longest cells (Morvan et al., 2003). The desirable qualities that flax bast fibre imparts on a woven fabric are a mixture of high strength, chemical resistance, wicking (the ability to quickly pull water from one side of the fabric to the other) and a soft, wearable feel (Hatch, 1993; Modelli et al., 2004). Conversely, flax fibres also impart some undesirable traits, like brittleness, which leads to wrinkling. Unfortunately, some of the fibre characteristics that are desirable for textiles are a bane to those farmers who grow flax as an oilseed; fibre strength and resistance to degradation make flax straw particularly difficult to manage as it does not decompose readily like other straw types (Modelli et al., 2004). Therefore, it is important to fibre and oilseed producers, as well as downstream industries, to study the molecular and cellular basis of bast fibre

characteristics, which could be altered to increase fibre value or improve production efficiencies (Dimmock et al., 2005).

In this literature review, I will first describe some structural and developmental aspects of flax stem anatomy with special emphasis on bast fibre, followed by a more detailed description of fibre development and a discussion about cell wall constituents. I will also describe general aspects of glycoproteins and protein phosphorylation, as these may be relevant to proteins involved in cell wall development.

1.2 Development and composition of fibres

In a mature flax stem, several different tissue layers are arranged in concentric rings around a central cylinder of xylem and pith. Starting from the outside and progressing inward there are: epidermis, cortex, phloem fibre, transport phloem and other phloem tissues, vascular cambium, secondary xylem and lastly, trace amounts of primary xylem and pith (Esau, 1943a). The epidermis is a single layer of cells that excrete a protective cuticle; these together function to keep pathogens out and to keep water in. Cortical tissues are green and photosynthetic; they may serve to provide assimilated carbon and energy to the adjacent tissues (Esau, 1943a). The role of phloem bast fibre in flax has not been fully elucidated, but evidence suggests a role as structural or mechanical support for the stem or in defence from herbivory or sap-sucking insects (McDougall et al., 1993). The main role of primary and secondary transport phloem is to transport photoassimilates from source tissues to sink tissues, although another important function may be the transport of other molecules like hormones, transcription factors and various RNA types (Yoo et al., 2004). The vascular cambium is a layer of meristematic cells that divides asymmetrically to produce secondary xylem to the centripetal side (toward center) and secondary phloem to the centrifugal side (away from center) (Esau, 1977; Smith, 2001). Finally, both primary and secondary xylem serve to transport water, nutrients and some hormones (Haberer and Kieber, 2002; Hartnung et al., 2002) from the roots up to the shoot tissues and serves also as a structural support (Esau, 1977).

There are two main areas of meristematic tissue that give rise to the stem patterning in flax: the shoot apical meristem (SAM) and the aforementioned vascular

cambium. The protoderm, procambium and ground meristem are each derived from the SAM and these rapidly dividing cells are progenitors of all cell types in the stem. The SAM is largely responsible for the addition of new cells in the axial direction to extend the length of the stem. In contrast, the vascular cambium is generally responsible for growth in the lateral direction (increase in stem diameter), due to the creation and subsequent growth of secondary tissue cells.

In flax, the SAM is comprised of one organized cell layer, L1, (historically referred to as the tunica) sitting a-top three or more layers, L2-L4, of slightly more visually disorganised cells, historically called the corpus (Esau, 1942). Situated at the apex of the SAM dome are a few cells from each layer that make up the meristematic initials. These cells divide occasionally to renew themselves and to create the progenitors of the future cell lines. The initial from the outermost layer, L1, divides to eventually give rise to the protoderm tissue, which subsequently divides to produce cells that differentiate in to the epidermis. The initials in the L2, L3, and L4 layers divide to eventually produce the procambium and the ground meristematic tissue. These procambial tissues go on to divide and differentiate into primary phloem, phloem parenchyma, bast fiber cells, vascular cambium and primary xylem, cortex and pith. The vascular bundles, visible in more mature tissue, containing primary phloem tissues, vascular cambium and primary xylem tissues differentiate from 'procambial strands' of elongated cells that can first be recognised in the procambium. Further, some cells that are situated between the vascular bundles and originated from the ground meristem become recruited to differentiate into vascular cambium; this creates a vascular cambium that fully encircles the stem (Esau, 1942, 1943b, a).

1.2.1 Fibre growth stages and the snap-point

In flax, bast fibre cells are have very thick cell walls (5-15μm thick), and are considered a primary phloem tissue as they originate from divisions in the protophloem within the procambium (Esau, 1943a; Morvan et al., 2003). They mature into very long, from 4 to 77 mm in length (compiled in Deyholos, 2006), multinucleate cells with extensive secondary thickenings (Ageeva et al., 2005). To arrive at their mature length, the cells expand through two main stages of elongation: coordinated growth and intrusive

growth. Coordinated growth happens near the top of the stem and it refers to how the fibre cells elongate similarly and synchronously with the surrounding cells. Basipetal to this, the older fibre cells begin intrusive growth and elongate at very rapid pace and push between the surrounding cells, which have largely completed their elongation (Gorshkova et al., 2003b). In a region approximately 5 to 9 cm down the stem from the shoot apex, the fibre cells stop elongating and begin secondary cell wall deposition. This region is referred to as the snap-point, because mechanically stressed stems tend to break at this location (Gorshkova et al., 2003a; Gorshkova et al., 2005).

There have been a number of papers published in the last 15 years that have furthered the knowledge-base on fibre elongation and secondary cell wall deposition. Using numerous microscopic and labelling techniques, it has been shown that flax fibres elongate intrusively via diffuse expansion (intercalary), as opposed to tip-growth (Ageeva et al., 2005). The mechanism driving elongation has yet to be elucidated in flax or other bast fibres but there are striking similarities to well-studied cotton cell expansion, mainly: flax fibre cells are symplastically isolated during elongation. This may allow for the build-up of turgor pressure and the use of that force to help drive elongation, as has been described in cotton (Ruan et al., 2001).

In terms of secondary cell wall deposition, there are two main types of cell wall material that can be distinguished microscopically: amorphous and crystalline. Both types appear during secondary deposition, with the amorphous, non-cellulosic material appearing first, and gradually being replaced by the more crystalline, cellulose-rich layer in a centripetal fashion (Gorshkova et al., 2003b). Thus, as the fibre cell matures, the division between the two layers moves towards the plasma membrane as each cellulosic layer is sequentially fixed into is final resting place. A similar deposition pattern of concentric circles of differently refractive material is seen in the secondary cell walls of cotton and other species (Cosgrove, 2005). The characteristics of the initial, amorphous cell wall may provide space and flexibility for the reorganization, alignment, insolubilization and crystallization of the later cell wall.

To go beyond the hierarchy of the single-cell fibre, it is important to note that flax fibre cells become tightly bound to each other in 'fibre bundles'. Each bundle can be comprised of anywhere from about 12 to 36 cells, as can be seen when visualised by

cross-section (Morvan et al., 2003). The cells adhere to each other through their middle lamellae which are especially rich in pectins and lignin (Lamblin et al., 2001; Morrison et al., 2003; Day et al., 2005). The bundles are difficult to separate by the industrial processes used to process flax straw in to usable fibre (Akin et al., 2001). Flax straw requires retting to free the useful fibre from the rest of the stem. Retting usually occurs through the activity of cell wall degrading bacteria or by treatment with enzymes (Antonov et al., 2007). In addition to separating the fibre from the stem it is important, especially for textiles, to separate the bundles into single fibre cells (Norton et al., 2006). Some retting processes are better at this than others and often retting must be followed by a mechanical process called carding to further separate the fibre cells (Akin et al., 2001).

1.2.2 Polysaccharide constituents of secondary cell walls

The main constituent of plant cell walls is cellulose; it generally makes up between 20% and 90% of the cell wall. In mature flax bast fibre cells the cellulose content is between 70 and 75% (Gorshkova et al., 1996; Gorshkova et al., 2000; Mooney et al., 2001; Day et al., 2005). Cellulose is an unbranched polysaccharide consisting of D-glucose molecules linked together by β-1,4 bonds and is polymerised by cellulose synthase in the plasma membrane (Heldt and Heldt, 1997). This differs from starches, which may or may not be branched and are mainly made of glucose molecules bonded together in α-1,4or α-1,6 bonds and are synthesised in plastids. Up to 150 cellulose chains align and form hydrogen bonds between the chains to form crystalline structures called microfibrils. Microfibrils have a high tensile strength, are impenetrable to water and are very resistant to degradation or chemical hydrolysis (Heldt and Heldt, 1997).

Hemicellulose is a term first used to describe polysaccharides that could be extracted by treatment with an alkaline solution, thus separating them categorically from cellulose, which is resistant to such chemical treatments. Hemicellulose polymers are thought to be part of the cell wall matrix, along with pectins (described below), and as such, function to bind the cellulose microfibrils together and add strength to the cell wall (Mooney et al., 2001; Morvan et al., 2003; Gorshkova and Morvan, 2006). Hemicellulose, comprising about 15 % of the cell wall weight in flax fibre, describes polysaccharides that can be branched and can be made of many different constituent

sugars (Mooney et al., 2001). Some examples of glycans that can be included in a hemicellulose molecule are fucose, arabinose, galactose, glucose, mannose and xylose. A xyloglucan, made of xylose, glucose and trace amounts of galactose, fucose and mannose has been identified in the fibre; it is suggested to be part of the primary wall (Morvan et al., 2003). Also noted in fibre, a hemicellulose made of glucose, mannose and galactose, galactoglucomannan (Gorshkova et al., 1996) which may be insolublized into the secondary cell wall.

Pectins are also a component of flax fibres where they comprise 10%-15% of fibre bundles by dry weight and are localized to the secondary cell wall, primary cell wall and middle lamella. (Davis et al., 1990; Mooney et al., 2001; Morvan et al., 2003). Pectins are made of polymers of sugars and sugar acids, like galacturonic acid; these sugar acids have a carboxyl group in place of one of the sugar hydroxyl groups appending the carbon ring. These negatively charged carboxyl groups help to hold the pectin polymers together through Ca^{2+} or Mg^{2+} mediated electrostatic bonds. In absence of doubly charged ions, pectin is water-soluble, as the polymers are not held together. EDTA, a chelator of doubly charged positive ions, is sometimes added to extraction solutions to aid the disruption of the bundles and cell wall (Davis et al., 1990); it can scavenge the Mg^{2+} and Ca^{2+} thereby solubizing the pectic compounds in the wall. Chelation allows separation of the microfibrils so as to show their highly aligned and spiral nature (Naomi Hotte, unpublished observation). The major types of pectin found in mature fibres are polygalacturonan, rhamnogalacturonans and arabinogalacturonan (Goubet et al., 1995; Andeme-Onzighi et al., 2000).

1.2.3 Non-enzymatic cell wall proteins

Glycoproteins, which can associate with the cell wall either covalently or non-covalently, are called such because they are proteins to which polysaccharides are attached via glycosidic bonds. The attachment of these polysaccharides is done in the ER and Golgi apparatus prior to excretion. In plants, the two main types of glycoproteins found in cell walls are arabinogalactans (AGPs) and hydroxyproline (Hyp) rich proteins (HRGPs) (Jose-Estanyol and Puigdomenech, 2000). However, studies attempting to characterize the proteins extracted from flax fibre secondary cell walls have shown that

most proteins are Gly-rich and Hyp-poor (Girault et al., 2000; Morvan et al., 2003). An AGP has been localized to young fibre cell walls in flax (Andeme-Onzighi et al., 2000), and it may also be Hyp-poor. The Gly-rich proteins are intriguing, as similar proteins have been found localized in either the phloem or vascular tissues of other plants (Condit et al., 1990; Girault et al., 2000).

1.3 Glycoprotein biochemistry

There are two known forms of polysaccharide-protein linkages and they are grouped as N-linked (amide bonded) and O-linked (hydroxyl bonded). Both N-linked and O-linked polysaccharides can be found appending proteins from a wide variety of genera (Hanisch, 2001). In plants, proteins can contain one or both of the N-linked and O-linked types of glycosylations (Sommer-Knudsen et al., 1996; Kishimoto et al., 1999).

For both N-linked and O-linked glycosylations, highly specific glycosyl transferases (GTs) are required to catalyze the bond formations. GTs transfer a specific sugar to its acceptor amino acid residue or carbohydate (Lim et al., 2003). The glycosyl transferases are membrane bound in the ER or Golgi apparatus.

An interesting difference between N- and O-linked glycosylation is a N-linked polysaccharide base structure is usually constructed before its addition to the protein; its formation is mediated by lipid-linked or prenyl-linked intermediates in the ER lumen (Heldt and Heldt, 1997). In contrast, sugar moieties are added to the protein sequentially by the specific glycosyl transferases for O-linked carbohydrate chains (Han and Martinage, 1992a). In N-linked polysaccharides, mannose and N-acetylglucosamine moieties combine to form the 5-sugar base-arrangement, called a trimannosyl chitobiosyl core structure; this grouping attaches to the protein exclusively through one of the N-acetylglucosamines (reviewed in Hanisch, 2001). As such, finding a combination of mannose and N-acetylglucosamine in a constituent sugar analysis is suggestive of an N-linked oligosaccharide (Sommer-Knudsen et al., 1998; Lodish et al., 2000).

An N-linked polysaccharide is attached by the amide bond to the side-chain of an Asn residue, where the Asn is found in the consensus sequence Asn-X-Ser/Thr (X can be any amino acid) (Han and Martinage, 1992b). The N-linked polysaccharides can be large

and are usually branched. They can include acetylated sugars, sugar-acids as well as sugars; the most common constituents found are glucose, galactose, fucose, mannose, and N-acetylglucosamine. Sialic acids are a group of a sugar acids found in glycosylations and its appearance is a notable difference between mammalian and plant N-linked glycoslations. A generous portion of mammalian N-linked glycosylations have terminal sialic acid residues, while this acid is rare or non-existent in plants (Takashima et al., 2006). Only since various plant genomes have been sequenced, have possible sialyltransferases been reported. At this point sialyltransferase-like sequences have been found in *A. thaliana* and *O. sativa*; but evidence of the sugar acid in plants is lacking (Shah et al., 2003; Séveno et al., 2004; Takashima et al., 2006).

The O-linked polysaccharide-protein bonds form a larger and more diverse group than the N-linked; there are a variety of amino acids and carbohydrates that can be directly involved in the bond. Plants contain O-linked bonds on Ser, Thr, and Hyp and the most common sugars involved are arabinose, galactose or N-acetylgalactosamine (Kieliszewski et al., 1995; Sommer-Knudsen et al., 1998; Jose-Estanyol and Puigdomenech, 2000; Hanisch, 2001). Ser/Thr residues are most commonly found bound directly to N-acetylgalactosamine but galactose has also been found attached to these amino acids in plants. The two main sugars attached to Hyp are arabinose and galactose.

Hyp-O-glycosylation is the most heavily studied glycosylation type in plants, due to its propensity to appear in cell wall glycoproteins. To become Hyp, Pro requires post-translational hydroxylation by prolyl hydroxylase (Lamport and Miller, 1971; Showalter and Varner, 1989; Jose-Estanyol and Puigdomenech, 2000). Prolyl hydroxylases seem to hydroxylate Pro in a sequence specific manor (Tiainen et al., 2005). If Pro is immediately preceded by either a Lys (Lys-Pro), Phe (Phe-Pro) or Tyr (Tyr-Pro) the Pro is never hydroxlated; whereas, the following sequences have been found to contain Hyp: Ser-Hyp-Hyp-Hyp-Hyp (Ser-$(Hyp)_4$), Hyp-Val, Hyp-Ala, Ala-Hyp (Kieliszewski and Lamport, 1994; Jose-Estanyol and Puigdomenech, 2000). Prolyl hydroxylases only partly control glycosylation as not all Hyps become glycosylated by glycosyl transferases (Kieliszewski and Lamport, 1994; Kieliszewski, 2001).

Hydroxyproline contigous theory, also called the O-Hyp glycosylation code, describes how the different sequence motifs that include Hyp can be used to predict the

general size and compostition of the attached polysaccharide (Fincher et al., 1974; Lamport, 1980; Smith et al., 1985). Sequences containing two or more adjacent Hyp residues attach to smaller, neutral arabinosyl chains (Kieliszewski and Lamport, 1994; Zhao et al., 2002); whereas Hyp residues found singly, even if only separated by one amino acid, generally attach to a larger polysaccharide containing both galactose and arabinose as the predominant sugars (Shpak et al., 1999; Kieliszewski, 2001; Tan et al., 2003; Sun et al., 2005). Based on the distribution of Hyp residues, the poly- or oligosaccharides usually form in clusters along the protein backbone that creates 'glycomodules' (Shpak et al., 1999; Kieliszewski, 2001; Tan et al., 2003; Sun et al., 2005); suggesting a possibility that each glycomodule may have a specific functional role.

1.4 Cell wall glycoprotein families

There are a number of glycoprotein families that are commonly associated with the plasma membrane and/or cell wall in plant cells. Four of the most common protein categories found insolublized in or associated with cell walls are arabinogalactan proteins (AGPs), extensin-like hydroxyproline-rich glycoproteins (HRGPs), proline-rich (PRPs) and non-extensin HRGPs and P-glycoproteins (PGP). Each family has certain determining factors such as characteristic consensus sequences, solubility, glycosylation amount or type, and whole molecule pKa and molecular weight. Molecular weight and pKa may be helpful for idenitifying glycoproteins based on their position on a 2-dimensional polyacrylamide gel.

1.4.1 Arabinogalactans

AGPs are a category of soluble, Hyp-rich proteins that are highly glycosylated (total molecule weight can be up to 98% carbohydrate); they have often been referred to as proteoglycans due to the predominance of carbohydrates relative to protein constituents (Jose-Estanyol and Puigdomenech, 2000). Another characteristic of AGPs is the presence of a neutral to acidic protein backbone, which may be interspersed with basic regions (Takos et al., 2000). In general the AGP backbone is rich in Hyp, Ala, Ser,

Thr and Gly, but some AGPs have been found to be low in Hyp (Baldwin et al., 1993; Du et al., 1996; Girault et al., 2000). Common motifs that are glycosylated are Ala-Hyp, Hyp-Ala, and some may contain Ser-$(Hyp)_4$ (Du et al., 1994; Kieliszewski and Lamport, 1994); as such, the protein carbohydrate connections are predominantly O-linked. Two major carbohydrates moieties for this group are galactose and arabinose; with rhamnose, glucuronic acid and others found to a much lower extent. The intra-polysaccharide galactosyl linkages can be found as β-1,3, β-1,6, β-1,3,6 or terminal, and the arabinosyl bonds can be β-1,2, β-1,3, β-1,5 or terminal. The lesser carbohydrate moieties have been found as terminally bonded rhamnose and glucuronic acid (Bacic et al., 1988; Zhao et al., 2002). The large degree of β-glycosyl bonds in AGPs enables the precipitation and visualization of them using Yariv's reagent (Yariv et al., 1962), which for a long time was a defining characteristic of this type of protein and is still used to visualize molecules with β-glycosyl bonds.

Many functions, roles, sub-cellular localizations, and expression patterns of AGPs have been proposed. AGPs have been identified in angiosperms, gymnosperms and lower plants (Clarke et al., 1979; Sommer-Knudsen et al., 1998). Many of the proteoglycans studied are either excreted extracellularly or are located in the plasma membrane (Sommer-Knudsen et al., 1998); subsequently, they can become part of the cell wall or aid in its construction (Pope, 1977; Sun et al., 2004). As a structural part of the cell wall, some AGPs are suggested to form linkages between the plasma membrane and the cell wall; additionally, roles for intercellular adhesion, cell wall loosening and cell signalling have been reported (Clarke et al., 1979; Kreuger and Van Holst, 1996; Sun et al., 2004). Arabinogalactans have been shown to be excreted from cultured cells (Sun et al., 2004), stylar and stigmatic cells (Gleeson and Clarke, 1980), sclerenchyma and vascular bundle cells (Sommer-Knudsen et al., 1998). One of the additional functions attributed to proteins in this group is signalling, perhaps having a role in cell differentiation and cellular positioning (Knox et al., 1990; Stacey et al., 1990; Schiavone et al., 1991; Schindler et al., 1995).

1.4.2 Hydroxyproline rich glycoproteins (Extensins)

A large group in the category of hydroxyproline-rich glycoproteins (HRGPs) is the extensins. The extensin group is generally defined as being insoluble, having a basic protein backbone and that is highly glycosylated, with the carbohydrates comprising about 50-60% glycoprotein weight (Sommer-Knudsen et al., 1998; Jose-Estanyol and Puigdomenech, 2000). The predominant protein-polysacharride bond is the O-linked, Hyp-arabinosyl; with some O-linked Ser-galactose bonds as well (Lamport et al., 1973; Kieliszewski and Lamport, 1994; Kieliszewski et al., 1995). In congruence with the Hyp contigous theory, the main amino acid repeating sequence is Ser-$(Hyp)_4$ (Smith et al., 1985). When hydrolysed into monosaccharide moieties, the predominant sugar is arabinose, at 90-97%, followed by galactose at 3-7% (Smith et al., 1985); this differs from the AGPs, where the arabinosyl and galactosyl moieties are generally found in similar abundance. Functionally, this high degree of arabinose in the extensins makes the overall glycoprotein basic, as the arabinose is neutral and the protein backbone is mostly basic.

Similar to many other glycoproteins, extensins display expression patterns that are tissue specific, developmentally regulated, and inducible by fungal attack and elicitors. Some extensins are thought to function in the cell wall as cross-linkers; they are excreted into the extracellular space as monomers and are there insolublized (Cooper and Varner, 1983, 1984). Their proposed function as cross-linkers is also based on work that suggests that the basic nature of the protein could non-covalently associate with pectins during cell wall formation (Cooper and Varner, 1984). Alternatively, covalent isodityrosine bonds have been suggested as permanent cross-linking in the cell wall, but this has not been fully established *in vivo* (Qi et al., 1995; Brady et al., 1996, 1998). More recently, apoplastic-peroxidase induced extensin-insolublization was shown to reduce fungal infection in a grapevine callus (Ribeiro et al., 2006) and over-expression of an extensin in arabidopsis showed decreased bacterial infection (Wei and Shirsat, 2006). Another proposed function is that the basic proteins can serve as scaffolding for the deposition of lignin (Whitmore, 1978). This is intriguing, but may not be relevant to flax fibre cells due to their low lignin content.

1.4.3 Proline-Rich Proteins and other HRGPs (non-extensin)

This category of glycoprotein is made up of various proteins that do not readily fit into the other groups. These are generally basic proteins that are minimally glycosylated, although polysaccharides can make up anywhere between 3% and 70% of total weight. Most of this group has only been studied by cDNA sequence analysis and relatively few of the proteins have been isolated and further investigated; thus, most of the glycosylation patterns are only inferred. There are some consensus sequences in this group, such as, Pro-Pro-X-Y-Lys, Pro-Pro-X-Lys, Pro-Pro-X-Y-Pro-Pro (Sommer-Knudsen et al., 1998). These proline-rich sequences can appear as single units, in clusters or may even include many repeats to cover most of the protein (Sommer-Knudsen et al., 1998). Moreover, the proline-rich repeat is a well-documented binding and interactive domain that appears in many, very diverse, protein families. Some glycosylated proteins, like the Solanaceous lectins and hybrid PRPs, contain this proline-rich repeat domain in addition to domains congruent with other glycoprotein or various other protein families (Baldwin et al., 2001; Bindschedler et al., 2006).

PRP glycoproteins studied thus far have various roles, localizations and fall under diverse regulatory regimes. Based on studies using soybean cells, some PRPs become insolublized quickly following wounding, fungal attack or elicitor application (Bradley et al., 1992; Brisson et al., 1994), which suggests a role in oxidative stress response. Similarily, a PRP categorized protein, found in pearl millet coleoptiles during fungal infection showed in vitro signs of peroxidase-mediated oxidative crosslinking; similar to that of extensin glycoproteins (Deepak et al., 2007). In relation to cell walls, a PRP found in the phloem of soybean was induced under water stress *in planta* (Battaglia et al., 2007). Studies in legumes, during nodulation, have shown the increase of an extracellular PRP nodulin (Sherrier et al., 2005), suggesting a role in symbiotic fungal relationships in addition to parasitism. Developmentally regulated PRPs have also been described (Bernhardt and Tierney, 2000). Apart from the afformentioned roles in secondary cell wall reinforcement and solidification, a developmentally controlled PRP glycoprotein, CaPRP1, has been localized to rapidly elongating tissues of young roots and leaves in *Capsicum annum* (Mang et al., 2004).

1.4.4 P-glycoproteins (PGP)

The PGP family of proteins was originally isolated from the plasma membrane of hamster cancerous-ovary cell lines (Debenham et al., 1982). PGPs were found to confer multidrug resistance to the cells in which they were over-expressed (Ambudkar et al., 2003), and PGPs are now classified as a subfamily of the very large ATP-binding cassette (ABC) transporter superfamily. These proteins are large, up to 140 kDa unglycosylated weight, have multiple transmembrane domains and are extensively glycosylated in their mature form, with the addition of 30-40 kDa of N-linked carbohydrates (Greenberger et al., 1987; Greenberger et al., 1988; Richert et al., 1988). Additionally, they also contain two nucleotide-binding domains (NBFs) (Endicott and Ling, 1989).

Over the past decade, genes homolgous to members of the original P-glycoprotein encoding family were identified in plants following the genome sequencing of arabidopsis and rice (Jasinski et al., 2003). In arabidopsis, the PGP subfamily is proportionally quite large, with 22 isoforms (Geisler and Murphy, 2006). The plant proteins are similar to their mammalian counterparts as they have 125 to 140 kDa molecular weights (Sanchez-Fernandez et al., 2001; Jasinski et al., 2003), are comprised of two similar halves, each with an integral membrane domain, NBF and predicted N-linked glycosylation sites (Theodoulou, 2000). The halves are joined by a ~60 amino acid linker domain; an area which shows the most divergence among the plant PGP homologs (Jasinski et al., 2003; Geisler et al., 2005).

An intriguing discovery is that some plant PGPs, such as AtPGP1, AtPGP2, AtPGP4 and AtPGP19, may transport the hormone auxin between cells by modulating auxin-efflux (Noh et al., 2001; Geisler et al., 2005). A proposed model suggests that PGPs are involved in non-polar auxin transport and therefore function independently of AUX/LAX and PIN proteins (Mravec et al., 2008); AUX1/LAX and PIN proteins have been previously implicated in a polar-auxin-transport role (reviewed in Geisler and Murphy, 2006). Another study suggested phosphorylation (see below for description) as regulation for PGP-related efflux by showing a relationship between auxin efflux and the phosphorylation status of the PGP (Nuhse et al., 2004).

1.5 Phosphorylation and its signalling involvement

The phosphorylation of proteins is a post-translational modification that is important for many cellular functions. Generally, the modification is reversible and is involved either directly or indirectly with protein function and/or regulation. Two of the major functions of protein phosphorylation are the transfer of a chemical communication signal, also termed signal cascade, and the transfer of energy. Signal cascades are exceptionally important to whole organisms by connecting a stimulus to an eventual response at a cellular level. Energy transfer can be involved in the signal transfer or function as an independent event. Either way, protein phosphorylation is central to the biochemical homeostasis of a cell.

Like most post-translational modifications, phosphorylations are residue-specific. In plants, the predominant phosphorylation sites are Ser and Thr; with Tyr phosphorylations found to a lesser extent (Rudrabhatla et al., 2006). When phosphorylated, the phosphate is covalently bound through the sidechain hydroxyl group of the amino acid residue (Salisbury and Ross, 1992); additionally phosphate groups can be bound through the sulfate of Cys (Aran et al., 2008). Some enzymes, like kinases, have shown an ability to auto-phosphorylate but most regulatory-related phosphorylations or dephosphorylations are catalyzed by a kinase or a phosphatase, respectively (Lodish et al., 2000). Nucleotide triphosphates (NTPs), like adenosine triphosphate (ATP), are most often the phosphate donors in phosphorylation reactions catalysed by kinases. NTPs contain a high-energy phosphoanhydride bond, where the terminal phosphate is attached to the other phosphates, which upon hydrolysis can provide the energy required for enzyme functions. For example, RuBisCO activase uses the energy from hydrolysing ATP to ADP to add a CO_2 to RuBisCO; thereby initiating RuBisCO activation (Lodish et al., 2000). RuBisCO activase is transiently phosphorylated for the production of energy that leads indirectly to its function, which is to aid in the regulation of RuBisCO (Salisbury and Ross, 1992).

Phosphorylations are well studied and very important in terms of regulation and signalling. Phosphorylation can play a key role in the synchronization of large pathways, such as keeping nitrate assimilation in tune with photosynthesis and carbon assimilation

(Huber, 2007). Key enzymes in these pathways are regulated, in part, through protein phosphorylation in response to light or dark. In the absence of light, nitrate reductase (NR) becomes phosphorylated to increase the binding specificity of an inhibitory 14-3-3 protein; upon their binding NR becomes inactivated (Bachmann et al., 1996; Huber, 2007). Also in the dark, sucrose phosphate synthase (SPS), a key enzyme involved in light-driven sucrose production, is phosphorylated at serine-158; it is inactivated simply with the phosphorylation and 14-3-3 is not involved (Huber et al., 1996). In this example, phosphorylation is used to keep the different pathways in sync but different mechanisims are used to cause the desired effect.

It is becoming clear that phosphorylation may be integral to signalling involving reactive oxygen species (ROS) (Aran et al., 2008); they are known to cause changes in the phosphorylation states of proteins. This type of signalling can happen in response to hormones, like abscisic acid, as well as biotic and abiotic stresses (Neill et al., 2002a; Neill et al., 2002b; Kolla et al., 2007). Some wound-induced cell wall reinforcements are mediated through peroxide signalling (Ribeiro et al., 2006). The current idea is the reactive oxygen species, like hydrogen peroxide and nitric oxide, are potentially oxidating Met or Cys residues near phosphorylation sites to alter the ability of the protein to be phosphorylated (Hancock et al., 2001; Huber, 2007). Regardless of how this is exactly occurring it is clear this type of signalling cascade is extremely important to defence and whole plant homeostasis (Bradley et al., 1992; Neill et al., 2002a; Neill et al., 2002b; Ribeiro et al., 2006; Kolla and Raghavendra, 2007; Kolla et al., 2007).

1.6 Literature review summary

The objective of my work was to study flax fibre development, concentrating on the biochemistry of cell wall deposition. To elucidate cellular protein biochemistry during cell wall deposition in fibres, I have used a proteomics approach. In Chapter 2, I will discuss my use of fluoresent protein labelling and gel electrophoresis to characterize the proteome of flax fibre cells during cell wall deposition. Chapter 3 builds on the flax fibre proteome by looking at protein regulation and interactions by studying the

glycoproteome and phosphoproteome of proteins localized to the flax stem. Chapter 4 is a brief conclusion of my work in flax.

1.7 Bibliography

Abbadi, A., Domergue, F., Bauer, J., Napier, J.A., Welti, R., Zaehringer, U., Cirpus, P., and Heinz, E. (2004). Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: Constraints on their accumulation. Plant Cell **16,** 2734-2748.

Ageeva, M.V., Petrovska, B., Kieft, H., Salnikov, V.V., Snegireva, A.V., van Dam, J.E.G., van Veenendaal, W.L.H., Emons, A.M.C., Gorshkova, T.A., and van Lammeren, A.A.M. (2005). Intrusive growth of flax phloem fibers is of intercalary type. Planta **222,** 565-574.

Akin, D.E., Morrison, W.H., III, Rigsby, L.L., and Dodd, R.B. (2001). Plant factors influencing enzyme retting of fiber and seed flax. Journal of Agricultural and Food Chemistry **49,** 5778-5784.

Ambudkar, S.V., Kimchi-Sarfaty, C., Sauna, Z.E., and Gottesman, M.M. (2003). P-glycoprotein: From genomics to mechanism. Oncogene **22,** 7468-7485.

Andeme-Onzighi, C., Girault, R., His, I., Morvan, C., and Driouich, A. (2000). Immunocytochemical characterization of early-developing flax fiber cell walls. Protoplasma **213,** 235-245.

Antonov, V., Marek, J., Bjelkova, M., Smirous, P., and Fischer, H. (2007). Easily available enzymes as natural retting agents. Biotechnology Journal **2,** 342-346.

Aran, M., Caporaletti, D., Senn, A.M., de Inon, M.T.T., Girotti, M.R., Llera, A.S., and Wolosiuk, R.A. (2008). ATP-dependent modulation and autophosphorylation of rapeseed 2-Cys peroxiredoxin. FEBS Journal **275,** 1450-1463.

Arbelaiz, A., Cantero, G., Fernandez, B., Mondragon, I., Ganan, P., and Kenny, J.M. (2005). Flax fiber surface modifications: Effects on fiber physico mechanical and flax/polypropylene interface properties. Polymers and Polymer Composites **26,** 324-332.

Bachmann, M., Huber, J.L., Liao, P.-C., Gage, D.A., and Huber, S.C. (1996). The inhibitor protein of phosphorylated nitrate reductase from spinach (*Spinacia oleracea*) leaves is a 14-3-3 protein. FEBS Letters **387,** 127-131.

Bacic, A., Gell, A.C., and Clarke, A.E. (1988). Arabinogalactan proteins from stigmas of *Nicotiana alata*. Phytochemistry **27,** 679-684.

Baldwin, T.C., McCann, M.C., and Roberts, K. (1993). A novel hydroxyproline-deficient arabinogalactan protein secreted by suspension-cultured cells of *Daucus carota*: Purification and partial characterization. Plant Physiology **103,** 115-123.

Baldwin, T.C., Domingo, C., Schindler, T., Seetharaman, G., Stacey, N., and Roberts, K. (2001). DcAGP1, a secreted arabinogalactan protein, is related to a family of basic proline-rich proteins. Plant Molecular Biology **45,** 421-435.

Battaglia, M., Solorzano, R.M., Hernandez, M., Cuellar-Ortiz, S., Garcia-Gomez, B., Marquez, J., and Covarrubias, A.A. (2007). Proline-rich cell wall proteins

accumulate in growing regions and phloem tissue in response to water deficit in common bean seedlings. Planta **225,** 1121-1133.

Bernhardt, C., and Tierney, M.L. (2000). Expression of AtPRP3, a proline-rich structural cell wall protein from arabidopsis, is regulated by cell-type-specific developmental pathways involved in root hair formation. Plant Physiology **122,** 705-714.

Bindschedler, L.V., Whitelegge, J.P., Millar, D.J., and Bolwell, G.P. (2006). A two component chitin-binding protein from french bean - association of a proline-rich protein with a cysteine-rich polypeptide. FEBS Letters **580,** 1541-1546.

Bradley, D.J., Kjellbom, P., and Lamb, C.J. (1992). Elicitor and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein a novel rapid defense response. Cell **70,** 21-30.

Brady, J.D., Sadler, I.H., and Fry, S.C. (1996). Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: A new potential cross-link. Biochemical Journal **315,** 323-327.

Brady, J.D., Sadler, I.H., and Fry, S.C. (1998). Pulcherosine, an oxidatively coupled trimer of tyrosine in plant cell walls: Its role in cross-link formation. Phytochemistry **47,** 349-353.

Brisson, L.F., Tenhaken, R., and Lamb, C. (1994). Function of oxidative cross-Linking of cell wall structural proteins in plant disease resistance. Plant Cell **6,** 1703-1712.

Clarke, A.E., Anderson, R.L., and Stone, B.A. (1979). Form and function of arabino galactans and arabino galactan proteins. Phytochemistry **18,** 521-540.

Condit, C.M., McLean, B.G., and Meagher, R.B. (1990). Characterization of the experession of the petunia glycine-rich protein-1 gene product. Plant Physiology **93,** 596-602.

Cooper, J.B., and Varner, J.E. (1983). Insolubilization of hydroxy proline-rich cell wall glyco protein in aerated carrot *Daucus carota* root slices. Biochemical and Biophysical Research Communications **112,** 161-167.

Cooper, J.B., and Varner, J.E. (1984). Cross-linking of soluble extension in isolated cell walls. Plant Physiology **76,** 414-417.

Cosgrove, D.J. (2005). Growth of the plant cell wall. Nature Reviews Molecular Cell Biology **6,** 850-861.

Davis, E.A., Derouet, C., Herve Du Penhoat, C., and Morvan, C. (1990). Isolation and an NMR study of pectins from flax *Linum usitatissimum* L. Carbohydrate Research **197,** 205-216.

Day, A., Ruel, K., Neutelings, G., Cronier, D., David, H., Hawkins, S., and Chabbert, B. (2005). Lignification in the flax stem: evidence for an unusual lignin in bast fibers. Planta **222,** 234-245.

Debenham, P.G., Kartner, N., Siminovitch, L., Riordan, J.R., and Ling, V. (1982). DNA mediated transfer of multiple drug resistance and plasma membrane glyco protein expression. Molecular and Cellular Biology **2,** 881-889.

Deepak, S., Shailasree, S., Sujeeth, N., Kini, R.K., Shetty, S.H., and Mithofer, A. (2007). Purification and characterization of proline/hydroxyproline-rich glycoprotein from pearl millet coleoptiles infected with downy mildew pathogen *Sclerospora graminicola*. Phytochemistry **68,** 298-305.

Deyholos, M.K. (2006). Bast fiber of flax (*Linum usitatissimum*): Biological foundations of its ancient and modern uses. Israel Journal of Plant Sciences **54,** 273-280.

Dimmock, J.P.R.E., Bennett, S.J., Wright, D., Edwards-Jones, G., and Harris, I.M. (2005). Agronomic evaluation and performance of flax varieties for industrial fibre production. Journal of Agricultural Science **143,** 299-309.

Du, H., Simpson, R.J., Clarke, A.E., and Bacic, A. (1996). Molecular characterization of a stigma-specific gene encoding an arabinogalactan-protein (AGP) from *Nicotiana alata.* Plant Journal **9,** 313-323.

Du, H., Simpson, R.J., Moritz, R.L., Clarke, A.E., and Bacic, A. (1994). Isolation of the protein backbone of an arabinogalactan-protein from the styles of *Nicotiana alata* and characterization of a corresponding cDNA. Plant Cell **6,** 1643-1653.

Endicott, J.A., and Ling, V. (1989). The biochemistry of P glycoprotein mediated multidrug resistance. Richardson, C. C., ed, pp. 137-172.

Esau, K. (1942). Vascular differentiation in the vegetative shoot of Linum.I. The procambium. American Journal of Botany **29,** 738-747.

Esau, K. (1943a). Vascular differentiation in the vegetative shoot of Linum. III. The origin of the bast fibers. American Journal of Botany **30,** 579-586.

Esau, K. (1943b). Vascular differentiation in the vegetative shoot of Linum. II. The first phloem and xylem. American Journal of Botany **30,** 248-255.

Esau, K. (1977). Anatomy of seed plants. (New York: Wiley).

Fincher, G.B., Sawyer, W.H., and Stone, B.A. (1974). Chemical and physical Properties of an arabino galactan peptide from wheat endosperm. Biochemical Journal **139,** 535-545.

Geisler, M., and Murphy, A.S. (2006). The ABC of auxin transport: The role of p-glycoproteins in plant development. FEBS Letters **580,** 1094-1102.

Geisler, M., Blakeslee, J.J., Bouchard, R., Lee, O.R., Vincenzetti, V., Bandyopadhyay, A., Titapiwatanakun, B., Peer, W.A., Bailly, A., Richards, E.L., Ejenda, K.F.K., Smith, A.P., Baroux, C., Grossniklaus, U., Mueller, A., Hrycyna, C.A., Dudler, R., Murphy, A.S., and Martinoia, E. (2005). Cellular efflux of auxin catalyzed by the arabidopsis MDR/PGP transporter AtPGP1. Plant Journal **44,** 179-194.

Girault, R., His, I., Andeme-Onzighi, C., Driouich, A., and Morvan, C. (2000). Identification and partial characterization of proteins and proteoglycans encrusting the secondary cell walls of flax fibres. Planta **211,** 256-264.

Gleeson, P.A., and Clarke, A.E. (1980). Arabino galactans of sexual and somatic tissues of *Gladiolus gandavensis* and *Lilium longiflorum.* Phytochemistry **19,** 1777-1782.

Gorshkova, T., and Morvan, C. (2006). Secondary cell-wall assembly in flax phloem fibres: role of galactans. Planta **223,** 149-158.

Gorshkova, T.A., Ageeva, M., Chemikosova, S., and Salnikov, V. (2005). Tissue-specific processes during cell wall formation in flax fiber. Plant Biosystems **139,** 88-92.

Gorshkova, T.A., Sal'nikova, V.V., Chemikosova, S.B., Ageeva, M.V., Pavlencheva, N.V., and van Dam, J.E.G. (2003a). The snap point: a transition point in *Linum usitatissimum* bast fiber development. Industrial Crops and Products **18,** 213-221.

Gorshkova, T.A., Wyatt, S.E., Salnikov, V.V., Gibeaut, D.M., Ibragimov, M.R., Lozovaya, V.V., and Carpita, N.C. (1996). Cell-wall polysaccharides of developing flax plants. Plant Physiology **110,** 721-729.

Gorshkova, T.A., Ageeva, M.V., Salnikov, V.V., Pavlencheva, N.V., Snegireva, A.V., Chernova, T.E., and Chemikosova, S.B. (2003b). Stages of bast fiber formation in *Linum usitatissimum* (Linaceae). Botanicheskii Zhurnal **88,** 1-11.

Gorshkova, T.A., Salnikov, V.V., Pogodina, N.M., Chemikosova, S.B., Yablokova, E.V., Ulanov, A.V., Ageeva, M.V., Van Dam, J.E.G., and Lozovaya, V.V. (2000). Composition and distribution of cell wall phenolic compounds in flax (*Linum usitatissimum* L.) stem tissues. Annals of Botany **85,** 477-486.

Goubet, F., Bourlard, T., Girault, R., Alexandre, C., Vandevelde, M.-C., and Morvan, C. (1995). Structural features of galactans from flax fibres. Carbohydrate Polymers **27,** 221-227.

Greenberger, L.M., Williams, S.S., and Horwitz, S.B. (1987). Biosynthesis of heterogeneous forms of multidrug resistance-associated glycoproteins. Journal of Biological Chemistry **262,** 13685-13689.

Greenberger, L.M., Lothstein, L., Williams, S.S., and Horwitz, S.B. (1988). Distinct P glycoprotein precursors are overproduced in independently isolated drug-resistant cell lines. Proceedings of the National Academy of Sciences of the United States of America **85,** 3762-3766.

Gutierrez, A., Rodriguez, I.M., and del Rio, J.C. (2006). Chemical characterization of lignin and lipid fractions in industrial hemp bast fibers used for manufacturing high-quality paper pulps. Journal of Agricultural and Food Chemistry **54,** 2138-2144.

Haberer, G., and Kieber, J.J. (2002). Cytokinins: New insights into a classic phytohormone. Plant Physiology **128,** 354-362.

Han, K.K., and Martinage, A. (1992a). Possible relationship between coding recognition amino acid sequence motif or residues and post-translational chemical modification of proteins. International Journal of Biochemistry **24,** 1349-1363.

Han, K.K., and Martinage, A. (1992b). Post-translational chemical modifications of proteins. International Journal of Biochemistry **24,** 19-28.

Hancock, J.T., Desikan, R., and Neill, S.J. (2001). Role of reactive oxygen species in cell signalling pathways. Biochemical Society Transactions **29,** 345-350.

Hanisch, F.-G. (2001). O-glycosylation of the mucin type. Biological Chemistry **382,** 143-149.

Hartnung, W., Sauter, A., and Hose, E. (2002). Abscisic acid in the xylem: Where does it come from and where does it go? Journal of Experimental Botany **53,** 27-32.

Hatch, K.L. (1993). Textile Science. (Minneapolis/Saint Paul: West Publishing).

Heldt, H.-W., and Heldt, F. (1997). Plant biochemistry & molecular biology. (Oxford: Oxford University Press).

Huber, S.C. (2007). Exploring the role of protein phosphorylation in plants: from signalling to metabolism. Biochemical Society Transactions **35,** 28-32.

Huber, S.C., Huber, J.L., Liao, P.-C., Gage, D.A., McMichael, R.W., Jr., Chourey, P.S., Hannah, L.C., and Koch, K. (1996). Phosphorylation of serine-15 of maize leaf sucrose synthase. Plant Physiology **112,** 793-802.

Jasinski, M., Ducos, E., Martinoia, E., and Boutry, M. (2003). The ATP-binding cassette transporters: Structure, function, and gene family comparison between rice and arabidopsis. Plant Physiology **131,** 1169-1177.

Johnston, A.M., Tanaka, D.L., Miller, P.R., Brandt, S.A., Nielsen, D.C., Lafond, G.P., and Riveland, N.R. (2002). Oilseed crops for semiarid cropping systems in the northern Great Plains. Agronomy Journal **94,** 231-240.

Jose-Estanyol, M., and Puigdomenech, P. (2000). Plant cell wall glycoproteins and their genes. Plant Physiology and Biochemistry **38,** 97-108.

Karus, M., and Vogt, D. (2004). European hemp industry: Cultivation, processing and product lines. Euphytica **140,** 7-12.

Kieliszewski, M.J. (2001). The latest hype on Hyp-O-glycosylation codes. Phytochemistry **57,** 319-323.

Kieliszewski, M.J., and Lamport, D.T.A. (1994). Extensin: Repetitive motifs, functional sites, post-translational codes, and phylogeny. Plant Journal **5,** 157-172.

Kieliszewski, M.J., O'Neill, M., Leykam, J., and Orlando, R. (1995). Tandem mass spectrometry and structural elucidation of glycopeptides from a hydroxyproline-rich plant cell wall glycoprotein indicate that contiguous hydroxyproline residues are the major sites of hydroxyproline O-arabinosylation. Journal of Biological Chemistry **270,** 2541-2549.

Kishimoto, T., Watanabe, M., Mitsui, T., and Hori, H. (1999). Glutelin basic subunits have a mammalian mucin-type O-linked disaccharide side chain. Archives of Biochemistry and Biophysics **370,** 271-277.

Knox, J.P., Pennell, R.I., Linstead, P.J., Stacey, N.J., Janniche, L., and Roberts, K. (1990). Extracellular matrix and cell surface markers in plant development. Journal of Cell Biology **111,** 20A.

Kolla, V.A., and Raghavendra, A.S. (2007). Nitric oxide is a signaling intermediate during bicarbonate-induced stomatal closure in *Pisum sativum*. Physiologia Plantarum **130,** 91-98.

Kolla, V.A., Vavasseur, A., and Raghavendra, A.S. (2007). Hydrogen peroxide production is an early event during bicarbonate induced stomatal closure in abaxial epidermis of arabidopsis. Planta **225,** 1421-1429.

Kreuger, M., and Van Holst, G.-J. (1996). Arabinogalactan proteins and plant differentiation. Plant Molecular Biology **30,** 1077-1086.

Lamblin, F., Saladin, G., Dehorter, B., Cronier, D., Grenier, E., Lacoux, J., Bruyant, P., Laine, E., Chabbert, B., Girault, F., Monties, B., Morvan, C., David, H., and David, A. (2001). Overexpression of a heterologous sam gene encoding S-adenosylmethionine synthetase in flax (*Linum usitatissimum*) cells: Consequences on methylation of lignin precursors and pectins. Physiologia Plantarum **112,** 223-232.

Lamport, D.T.A. (1980). Structure and function of plant glyco proteins. Preiss, P., ed, pp. P501-542.

Lamport, D.T.A., and Miller, D.H. (1971). Hydroxy proline arabinosides in the plant kingdom. Plant Physiology **48,** 454-456.

Lamport, D.T.A., Katona, L., and Roerig, S. (1973). Galactosyl serine in extensin. Biochemical Journal **133,** 125-131.

Lim, E.-K., Baldauf, S., Li, Y., Elias, L., Worrall, D., Spencer, S.P., Jackson, R.G., Taguchi, G., Ross, J., and Bowles, D.J. (2003). Evolution of substrate recognition across a multigene family of glycosyltransferases in arabidopsis. Glycobiology **13,** 139-145.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). Molecular cell biology. (New York: W. H. Freeman and Company).

Mang, H.G., Lee, J.-H., Park, J.-A., Pyee, J., Pai, H.-S., Lee, J.H., and Kim, W.T. (2004). The CaPRP1 gene encoding a putative proline-rich glycoprotein is highly expressed in rapidly elongating early roots and leaves in hot pepper (*Capsicum annuum* L. cv. Pukang). Biochimica et Biophysica Acta **1674,** 103-108.

McCorriston, J. (1997). The fiber revolution: Textile extensification, alienation, and social stratification in ancient Mesopotamia. Current Anthropology **38,** 517-549.

McDougall, G.J., Morrison, I.M., Stewart, D., Weyers, J.D.B., and Hillman, J.R. (1993). Plant fibres: Botany, chemistry and processing for industrial use. Journal of the Science of Food and Agriculture **62,** 1-20.

Modelli, A., Rondinelli, G., Scandola, M., Mergaert, J., and Cnockaert, M.C. (2004). Biodegradation of chemically modified flax fibers in soil and in vitro with selected bacteria. Feb **5,** 596-602.

Mohanty, A.K., Misra, M., and Hinrichsen, G. (2000). Biofibres, biodegradable polymers and biocomposites: An overview. Macromolecular Materials and Engineering **276,** 1-24.

Mooney, C., Stolle-Smits, T., Schols, H., and de Jong, E. (2001). Analysis of retted and non retted flax fibres by chemical and enzymatic means. Journal of Biotechnology **89,** 205-216.

Morrison, W.H., Himmelsbach, D.S., Akin, D.E., and Evans, J.D. (2003). Chemical and spectroscopic analysis of lignin in isolated flax fibers. Journal of Agricultural and Food Chemistry **51,** 2565-2568.

Morvan, C., Andeme-Onzighi, C., Girault, R., Himmelsbach, D.S., Driouich, A., and Akin, D.E. (2003). Building flax fibres: More than one brick in the walls. Plant Physiology and Biochemistry **41,** 935-944.

Mravec, J., Kubes, K., Bielach, A., Gaykova, V., Petrasek, J., Skupa, P., Chand, S., Benkova, E., Zazimalova, E., and Friml, J. (2008). Interaction of PIN and PGP transport mechanisms in auxin distribution-dependent development. Development **135,** 3345-3354.

Muir, A.D., and Westcott, N.D. (2003). Flaxseed constituents and human health. In Flax: the genus *Linum*, A.D. Muir and N.D. Westcott, eds (New York: Taylor & Francis), pp. 307.

Neill, S., Desikan, R., and Hancock, J. (2002a). Hydrogen peroxide signalling. Current Opinion in Plant Biology **5,** 388-395.

Neill, S.J., Desikan, R., Clarke, A., and Hancock, J.T. (2002b). Nitric oxide is a novel component of abscisic acid signaling in stomatal guard cells. Plant Physiology **128,** 13-16.

Noh, B., Murphy, A.S., and Spalding, E.P. (2001). Multidrug resistance-like genes of arabidopsis required for auxin transport and auxin-mediated development. Plant Cell **13,** 2441-2454.

Norton, A.J., Bennett, S.J., Hughes, M., Dimmock, J.P.R.E., Wright, D., Newman, G., Harris, I.M., and Edwards-Jones, G. (2006). Determining the physical properties of flax fibre for industrial applications: the influence of agronomic practice. Annals of Applied Biology **149,** 15-25.

Nuhse, T.S., Stensballe, A., Jensen, O.N., and Peck, S.C. (2004). Phosphoproteomics of the arabidopsis plasma membrane and a new phosphorylation site database. Plant Cell **16,** 2394-2405.

Pope, D.G. (1977). Relationships between hydroxy proline containing proteins secreted into the cell wall and medium by suspension cultured *Acer pseudoplatanus* cells. Plant Physiology **59,** 894-900.

Qi, X., Behrens, B.X., West, P.R., and Mort, A.J. (1995). Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension cultures: Evidence for a covalent cross-link between extensin and pectin. Plant Physiology **108,** 1691-1701.

Ribeiro, J.M., Pereira, C.S., Soares, N.C., Vieira, A.M., Feijo, J.A., and Jackson, P.A. (2006). The contribution of extensin network formation to rapid, hydrogen peroxide-mediated increases in grapevine callus wall resistance to fungal lytic enzymes. Journal of Experimental Botany **57,** 2025-2035.

Richert, N.D., Aldwin, L., Nitecki, D., Gottesman, M.M., and Pastan, I. (1988). Stability and covalent modification of P-glycoprotein in multidrug-resistant KB cells. Biochemistry **27,** 7607-7613.

Ruan, Y.-L., Llewellyn, D.J., and Furbank, R.T. (2001). The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K+ transporters and expansin. Plant Cell **13,** 47-60.

Rudrabhatla, P., Reddy, M.M., and Rajasekharan, R. (2006). Genome-wide analysis and experimentation of plant serine/threonine/tyrosine-specific protein kinases. Plant Molecular Biology **60,** 293-319.

Salisbury, F.B., and Ross, C.W. (1992). Plant physiology. (Belmont: Wadsworth Publishing Company).

Sanchez-Fernandez, R., Davies, T.G.E., Coleman, J.O.D., and Rea, P.A. (2001). The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. Journal of Biological Chemistry **276,** 30231-30244.

Schiavone, F.M., Roberts, K., Stacey, N., and Racusen, R.H. (1991). Appearance of cell-surface arabinogalactan proteins predicts zones of tissue remodeling in regenerating carrot embryos. Journal of Cell Biology **115,** 114A.

Schindler, T., Bergfeld, R., and Schopfer, P. (1995). Arabinogalactan proteins in maize coleoptiles: Developmental relationship to cell death during xylem differentiation but not to extension growth. Plant Journal **7,** 25-36.

Sengupta, G., and Palit, P. (2004). Characterization of a lignified secondary phloem fibre-deficient mutant of jute (*Corchorus capsularis*). Annals of Botany **93,** 211-220.

Séveno, S., Bardor, M., Paccalet, T., Gomord, V., Lerouge, P., and Faye, L. (2004). Glycoprotein sialylation in plants? Nature Biotechnology **22,** 1351–1352.

Shah, M.M., Fujiyama, K., Flynn, C.R., and Joshi, L. (2003). Sialylated endogenous glycoconjugates in plant cells. Nature Biotechnology **21,** 1470-1471.

Sherrier, D.J., Taylor, G.S., Silverstein, K.A.T., Gonzales, M.B., and VandenBosch, K.A. (2005). Accumulation of extracellular proteins bearing unique proline-rich motifs in intercellular spaces of the legume nodule parenchyma. Protoplasma **225,** 43-55.

Showalter, A.M., and Varner, J.E. (1989). In The biochemistry of plants, A. Marcus, ed (New York: Acedemic Press), pp. 485.

Shpak, E., Leykam, J.F., and Kieliszewski, M.J. (1999). Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. Proceedings of the National Academy of Sciences of the United States of America **96,** 14736-14741.

Smith, J.J., Muldoon, E.P., and Lamport, D.T.A. (1985). Extensin precursors P-1 and P-2 are highly periodic structures. Plant Physiology **77,** 62.

Smith, L.G. (2001). Plant cell division: building walls in the right places. Nature Reviews Molecular Cell Biology **2,** 33-39.

Sommer-Knudsen, J., Clarke, A.E., and Bacic, A. (1996). A galactose-rich, cell-wall glycoprotein from styles of *Nicotiana alata*. Plant Journal **9,** 71-83.

Sommer-Knudsen, J., Bacic, A., and Clarke, A.E. (1998). Hydroxyproline-rich plant glycoproteins. Phytochemistry **47,** 483-497.

Stacey, N.J., Roberts, K., and Knox, J.P. (1990). Patterns of Expression of the Jim4 arabinogalactan protein epitope in cell cultures and during somatic embryogenesis in *Daucus carota* L. Planta **180,** 285-292.

Sun, W., Zhao, Z.D., Hare, M.C., Kieliszewski, M.J., and Showalter, A.M. (2004). Tomato LeAGP-1 is a plasma membrane-bound, glycosylphosphatidylinositol-anchored arabinogalactan-protein. Physiologia Plantarum **120,** 319-327.

Sun, W., Xu, J., Yang, J., Kieliszewski, M.J., and Showalter, A.M. (2005). The lysine-rich arabinogalactan-protein subfamily in arabidopsis: Gene expression, glycoprotein purification and biochemical characterization. Plant and Cell Physiology **46,** 975-984.

Takashima, S., Abe, T., Yoshida, S., Kawahigashi, H., Saito, T., Tsuji, S., and Tsujimoto, M. (2006). Analysis of sialyltransferase-like proteins from *Oryza sativa*. Journal of Biochemistry **139,** 279-287.

Takos, A.M., Dry, I.B., and Soole, K.L. (2000). Glycosyl-phosphatidylinositol-anchor addition signals are processed in *Nicotiana tabacum*. Plant Journal **21,** 43-52.

Tan, L., Leykam, J.F., and Kieliszewski, M.J. (2003). Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. Plant Physiology **132,** 1362-1369.

Theodoulou, F.L. (2000). Plant ABC transporters. Biochimica et Biophysica Acta **1465,** 79-103.

Tiainen, P., Myllyharju, J., and Koivunen, P. (2005). Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. Journal of Biological Chemistry **280,** 1142-1148.

van Zeist, W., and Bakker-Heeres, J.A.H. (1975). Evidence for linseed cultivation before 6000 bc. Journal of Archaeological Science **2,** 215-219.

Wei, G., and Shirsat, A.H. (2006). Extensin over-expression in arabidopsis limits pathogen invasiveness. Molecular Plant Pathology **7,** 579-592.

Whitmore, F.W. (1978). Lignin carbohydrate complex formed in isolated cell walls of callus. Phytochemistry **17,** 421-426.

Yariv, J., Rapport, M.M., and Graf, L. (1962). The interaction of glycosides and saccharides with antibody to the corresponding phenylazo glycosides. Biochemical Journal **85** 383-388.

Yoo, B.-C., Kragler, F., Varkonyi-Gasic, E., Haywood, V., Archer-Evans, S., Lee, Y.M., Lough, T.J., and Lucas, W.J. (2004). A systemic small RNA signalling system in plants. Plant Cell **16,** 1979-2000.

Zhao, Z.D., Tan, L., Showalter, A.M., Lamport, D.T.A., and Kieliszewski, M.J. (2002). Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. Plant Journal **31,** 431-444.

Chapter 2: A flax fibre proteome: Identification of proteins enriched in bast fibres

2.1 Introduction

Flax (*Linum usitatissimum* L.) has attracted human attention since the beginning of agriculture (Vanzeist and Bakkerheeres, 1975; McCorriston, 1997). This is due in part to the unusual properties of the bast (i.e. phloem) fibres, which because of their great length and high tensile strength have found use in textiles and many other products (Mohanty et al., 2000). Fibre length is achieved almost entirely through intrusive growth, which is a process limited to very few cell types in plants (Esau, 1943; Lev-Yadun, 2001). The elongation stage is succeeded by a dynamic process of secondary wall deposition, in which a matrix of galactose-rich polymer in the nascent wall is gradually and centripetally replaced by highly crystalline cellulose (Gorshkova and Morvan, 2006). Because secondary wall deposition increases the tensile strength of cells, fibres which have undergone even the very first stages of cell wall thickening can be distinguished mechanically by their resistance to breakage at the "snap-point" of the stem (Gorshkova et al., 2003). The snap-point thus defines an important developmental transition from cell elongation to cell wall thickening.

Previously, others have produced libraries of cDNAs from fibre-bearing peels of flax and hemp stems (Day et al., 2005; Roach and Deyholos, 2007). In addition to containing bast fibres at various stages of development, these peels also contained many other cell types, including those associated with cambium and transport phloem. Analysis of these libraries by cDNA microarray hybridization and other techniques identified distinct patterns of expression of transcripts of polysaccharide-related enzymes in stem peels during fibre elongation and cell wall deposition. However, due to inherent technical and biological limitations, it is known that in many circumstances, abundance of transcripts and proteins for a given gene may not be highly correlated (Tian et al., 2004; Mooney et al., 2006). This well-established limitation on the biological relevance of transcriptome analysis led to the complement previous studies with a survey of the proteins present in developing flax fibres during the onset of secondary wall deposition. This is similar to a proteomics approaches used to study secondary cell wall development of other cell types in other species (Vander Mijnsbrugge et al., 2000; Blee et al., 2001;

Watson et al., 2004; Bayer et al., 2006; Yao et al., 2006). Past protein-related studies directed at the bast fibres also used fibre-rich stem peals to enrich fibre-related proteins. For this study of the proteome, I increased the specificity of this analysis by extracting proteins from phloem fibres that had been individually dissected from the snap point of growing stems, and comparing their abundance to proteins in the surrounding, non-fibre cells of the cortex from the same stems. The objective of this study is therefore to identify those proteins that contribute to the interesting pattern of cell wall deposition in flax fibres.

2.2 Methods and materials

2.2.1 Plant material

Fibres (i.e. individual cells) and surrounding, non-fibre cells of the cortex were isolated from the stems of *Linum usitatissimum* L., var. Norlin. A total of 495 plants were harvested from four independently grown populations. Seeds were sown two per 10 cm pot and grown as previously described (Roach and Deyholos, 2007). After 3 weeks of growth, the mean distance from the apex to snap-point was 5.9 cm, with mean plant height of 19 cm. To find the snap-point, the tip of each plant was bent over until an obvious change in the radius of curvature could be seen; below this area proved more resistant to bending under pressure than the tip. A 3 cm segment of stem, spanning from 2 cm to 5 cm below the snap-point, was further dissected to separate the individual fibres and surrounding non-fibre cells of the cortex (i.e. "the non-fibre fraction", consisting predominantly of parenchyma, sieve elements, and companion cells, but excluding epidermis, xylem and pith) for proteomic analysis. After dissection, fibres and surrounding tissues were rinsed in deionized water, blotted, then frozen in liquid nitrogen, and stored at -80 °C.

2.2.2 Protein isolation from tissues

Tissues were ground to a powder in liquid nitrogen and then further ground for one minute in 1 mL cold TCA/acetone buffer (20 mM DTT, 10 % trichloroacetic acid in cold acetone). Homogenates were transferred with an additional 1 mL of buffer to

microcentrifuge tubes and were allowed to precipitate overnight at –20 °C. After centrifugation (12000 xg, 10 °C, 15 minutes), pellets were rinsed once with 1 mL 20 mM DTT in acetone for 1 h at –20 °C, then pellets were left to dry at –20 °C for 2 h, and dissolved in 200 µL of urea/thiourea buffer (7 M urea, 2 M thiourea, 4 % (w/v) CHAPS, 30 mM Tris-Cl) by vortexing at room temperature for 30 minutes. The solution was clarified by centrifugation (12000 xg, 17 °C, 5 minutes) and supernatants were further processed by using the 2D Clean-Up Kit (GEhealthcare). Precipitates were re-dissolved in 60 µl of the urea/thiourea buffer, and concentrations of the protein samples were determined using the 2D Quant Kit (GEhealthcare) and NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies) against a BSA standard curve.

2.2.3 Fluorescent labelling of proteins

Four independent pools of approximately 125 plants each were grown in nominally identical conditions that were spatially and temporally separated from each other. Proteins were isolated separately from tissues dissected from each pool of plants, to produce four paired protein samples from fibres and the non-fibre fraction, where each pair of samples was biologically independent from every other pair. We labelled each 30 µg protein sample (pH adjusted to 8.5) with 240 pmol of Cy2, Cy3 or Cy5 fluorescent dyes, using the CyDye DiGE fluors (minimal dyes) labelling kit (GEhealthcare). Labelling reactions were stopped by the addition of 1 µl of 10 mM lysine to each tube, and after a further 10-minute incubation on ice, the volume of each sample was doubled with the addition of a sample buffer (7 M urea, 2 M thiourea, 2 % (v/v) ampholyte, 2 % (w/v) DTT, 4 % (w/v) CHAPS) to ready the samples for IEF. Labelled samples were mixed together as stated in table 2-1 to create four analytical gels, with each gel containing an internal standard and both tissue samples. The internal standard is prepared by mixing equal masses of protein extracts from fibre and non-fibre fractions of each biologically independent harvest.

2.2.4 2DE of CyDye labelled protein mixtures

All subsequent handling and separation steps for 2DE were conducted away from light. 24 cm, 3-10 NL Immobiline drystrips (GEhealthcare) were passively re-hydrated

for 10 h in (8 M urea, 4 % (w/v) CHAPS, 1 % (v/v) ampholytes 3-10, 13 mM DTT, trace bromophenol blue). A total of 60 kVh (1 h at 500v; 1 h at 1000v; 8.3 h at 8000v) at 20°C was used to focus the proteins using an IPGphor II (GEhealthcare). Paper wicks on the basic end were spiked with 13 mM DTT and were changed three times during the run. Following IEF, strips were equilibrated for SDS-PAGE separation by gentle agitation for 15 minutes in 6 M urea, 50 mM tris-Cl (pH 8.8), 30 % (v/v) glycerol, 2 % (w/v) SDS, trace bromophenol blue plus 0.5 % (w/v) DTT, followed by 15 minutes in the same solution with 4.5 % (w/v) IAA instead of DTT. After equilibration, the strips were sealed onto the top edge of self-cast, large-format, 12.5 % acrylamide gels using sealing solution (1 % low-melt agarose, trace bromophenol blue in 1X running buffer). The four analytical gels were separated by molecular weight during SDS-PAGE, simultaneously, using the Ettan Dalt *six* (GEhealthcare). The gels were run at 2 W/gel for 30 minutes then 8 W/gel until the bromophenol blue dye front just touched the end of the gels.

2.2.5 Imaging and analysis

Fluorescently labelled gels were imaged at 100 μm resolution with PMT voltage output between 50000 and 63558 counts. DeCyder 6.5 (GEhealthcare) was used to match, normalize, and statistically analyze spots. After in-gel normalization using Differential In-gel Analysis (DIA), the Biological Variation Analysis (BVA) module was used for statistical analysis and normalization across all analytical gels.

2.2.6 Spot-picking and tryptic digestion of proteins

Preparative gels, loaded with about 125 μg of protein, were post-stained with Deep Purple total protein stain (GEhealthcare) and spot-matched to the analytical gels. Gel spot excision and subsequent tryptic digestion were conducted using an Ettan Spot-picker (GEhealthcare) and ProteomeWorks MassPREP robotic handling station (Bio-Rad Laboratories and Waters corporations), resulting with peptides in a final extraction solution of 2 % ACN, 0.1 % formic acid in H_20.

2.2.7 Protein identification

LC MS/MS analysis was performed using an online 1100 series XCT Ion trap (Agilent Technologies). The autosampler injected 18 μL of each sample onto an enrichment column (Zorbax 300SB-C18 5 μm 5 x 0.3 mm) that connected to a second column (Zorbax 300SB-C18 5 μm 150 x 0.3 mm) in a peptide-separation gradient that started at 85 % solvent A (0.1 % formic acid in H_2O) and ended at 55 % solvent B (0.1 % formic acid, 5 % H_2O in ACN) over a 42 minute span. This was followed by 10 minutes of 90 % solvent B to cleanse the columns before returning to 97 % solvent A for the next sample. The MS ran a 300-2200 m/z scan followed by MS/MS analysis of the most intense ions. Raw spectral data was processed into Mascot Generic File (.mgf) format using the default method in the ChemStation Data Analysis module and ion searches were completed in MASCOT (McDougall, 1997; Loyola-Vargas et al., 2007) with the search parameters of: peptide tolerance of 2 Da, parent ion tolerance of 0.8 m/z, ion charge of +1, +2 and +3.

2.3 Results and discussion

2.3.1 Separation of fibre and non-fibre proteins

To increase my understanding of the proteins that contribute to the unique properties of flax bast fibres, I extracted proteins from ultimate fibres (i.e. individual cells) dissected from the snap-point region of vegetative stems (21-24 days post germination) (figure 2-1). The snap-point is the stem region in which secondary wall deposition begins (Gorshkova et al., 2003). I also collected the surrounding non-fibre cells (consisting predominantly of parenchyma, sieve elements, and companion cells) from the cortex of the snap-point. Throughout the remainder of this report, I will refer to the ultimate bast fibres collected from the snap-point as simply "fibres", and the surrounding, non-fibre cells of the cortex as the "non-fibre fraction". By labelling proteins from fibres and the non-fibre fraction with contrasting fluorescent dyes, and separating the mixture of the two samples simultaneously using 2D gel electrophoresis (DiGE), I was able to identify proteins that were more abundant in fibres as compared to the non-fibre fraction (figure 2-2).

In each of four replicate gels, 1850 distinct protein spots were detected from fibres, and 1695 spots from the non-fibre fraction. In total, 558 protein spots differed in fluorescent signal intensity by at least 1.5 fold ($p \leq 0.05$) between the samples, with 246 spots (13 % of total detected) enriched in fibres and 312 spots (18 % of total) enriched in the non-fibre fraction (figure 2-3). The distinctive protein profiles of fibres and the non-fibre fraction were also evident from visual inspection of the DiGE gel image (figure 2-2). Phloem fibres therefore appear to express a complement of proteins that is distinct from surrounding cell types in the stem.

2.3.2 Protein identification by LC/MSMS

I picked 190 protein spots that were enriched in fibre samples for identification by mass spectrometry. Spots were selected based on criteria of large spot volume, high fold-enrichment of signals, and well-focused spot morphology. For comparison, 50 additional spots were collected that were enriched in non-fibre fractions or that were similarly abundant in both types of protein samples. Although the patterns of fold-enrichment that are reported were reproducible within the statistical parameters indicated (table 2-2), individual ratios should not be extrapolated quantitatively to whole proteins, in part because some proteins may be represented by more than one spot.

A total of 240 spots were subjected to analysis by LC/MSMS. Of these, 126 spots produced spectra that could not be assigned to existing sequences, while spectra from the remaining 114 spots produced significant matches (i.e. MOWSE scores 40-675; two or more peptides matched per spot) to predicted spectra from Genbank protein databases (table 2-2). Four spots (#7, #41, #72, #89) contained predicted peptides that matched more than one distinct protein, indicating the presence of multiple proteins in some spots on the gel. Of the spots to which we assigned protein identities, 76 were enriched by at least 1.5 fold (i.e. 1.5×) in fibre samples, and 51 of these were statistically more abundant ($p \leq 0.05$) in fibres than the non-fibre fraction. Conversely, I was able to assign identity to 17 spots enriched 1.5-fold or more in the non-fibre fraction; at least seven of these were associated with photosynthesis (spots #44-#47, #73, #74, #81). Because photosynthesis is a process expected to dominate metabolism in the non-fibre fraction, these observations are consistent with the physical separation of fibre and non-fibre tissues I hoped to

achieve by dissection. The focus of the remainder of this thesis will be on the spots that were enriched in fibres.

The fibre-enriched proteins to which I was able to assign putative identities were classified into eight functional categories (figure 2-4). Aside from the category called "miscellaneous", which represented a diverse set of functions, most of the proteins that were identified in fibre samples could be assigned to one of three categories related to the conversion of carbohydrates for energy or glycan biosynthesis, namely: primary carbon and energy metabolism; one-carbon metabolism; and cell wall and polysaccharide metabolism (figure 2-4). The predominance of these proteins involved in the metabolism of carbohydrates and related compounds is consistent with the major biochemical activities that would be expected to observe within cells active in secondary wall biogenesis. In addition, a smaller number of proteins could be assigned to each of the remaining categories: membrane transport; cytoskeleton and secretion; ATPases; and protein and amino acid metabolism. The membership of proteins assigned to spots in each of the eight functional categories is shown in table 2-2, and is discussed in more detail in the following sections.

2.3.2.1 Primary carbon and energy metabolism

The conversion of monosaccharides and starch into energy is the inferred function of the largest proportion of proteins that were enriched (>1.5 fold) in fibres, as compared to the non-fibre fraction at the stem snap point (figure 2-4). These reactions are also summarized in figure 2-5. Two of the most highly enriched proteins that were detected in any functional category were β-amylase (spot #17; 8.8 fold enriched in fibres), and fructose kinase (#93, 6.7×; #94, 2.2×; #96, 2.0×), which catalyze the first steps in the catabolism of starch and fructose, respectively (table 2-2). The increased relative abundance of these enzymes in fibres provides some insight into the immediate sources of carbon and energy for secondary wall biogenesis. I also detected the statistically significant ($p \leq 0.05$) enrichment of enzymes of glycolysis and related processes, namely fructose-bisphosphate aldolase (#78, 2.4×), glyceraldehyde 3-phosphate dehydrogenase (#83, 2.6×; #87, 2.8×), and phosphoglucomutase (#27, 1.8×; #28, 3.7×), as well as the presence of phosphoglycerate kinase (#68, #71). Finally, fibre-enriched protein spots

putatively representing 5 of 8 enzymes of the tricarboxylic acid cycle (TCA cycle)were identified; where in further energy and metabolic precursors are generated from the products of glycolysis. The TCA cycle -associated proteins that were significantly enriched in fibres and included citrate synthase (#63, 3.7×), succinyl coA-ligase (#82, 2.3×), fumarase (#57, 2.5×), and malate dehydrogenase (#92, 3.3×).

2.3.2.2 ATPases

Many subunits of the ATPase/synthase complex were identified in either fibres or the non-fibre fraction, including an α-subunit (#35), β-subunits (#42, #43), and a γ-subunit (#99). The tissue-specific abundance patterns of these various subunits were surprisingly complex: the γ-subunit and one β-subunit (#42) were associated with equal spot intensities in both sample types, while the other ATP synthase β-subunit (#44), was 1.8× more abundant in the non-fibre fraction. Only the α-subunit was more abundant (1.6×) in fibres.

In addition to the ATPase/synthases described above, we identified peptides from several other types of putative ATPases, including three protein spots containing vacuolar-type ATPase (v-ATPase), of which, two spots (#24, 2.6×; #105, 1.8×) were significantly ($p \leq 0.05$) enriched in fibres. v-ATPases are some of the most abundant membrane proteins in the vacuole and endomembrane system, and their enrichment may reflect increased relative abundance of these organellar structures in fibres (Schumacher, 2006). We also detected a putative plasma membrane-associated AAA-ATPase (#1, 1.6×) in fibres, although this was not deemed to be more abundant in fibres by our usual statistical criteria. Both v-ATPases and AAA-ATPases have been previously demonstrated to be essential for vesicle transport, and might therefore be active in secondary wall-specific processes in developing fibres (Schumacher, 2006; Haas et al., 2007).

2.3.2.3 Cell wall and polysaccharide metabolism

Cell walls consist of many types of polymers, including cellulose, hemicellulose, and pectins. However, with the possible exception of an NAD^+-dependent epimerase/dehydratase with similarity to UDP-xylose synthases (#76, 6.1×), and GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase (GME, #101, 2.3×) almost all of the fibre-enriched, cell wall-related enzymes that were identified were most likely associated with the metabolism of pectin-like substances. For example, I identified proteins from six spots as β-galactosidases. Five of these (#12-#16) were co-located in a charge train and the sixth (#64) was an isolated spot of lower apparent molecular weight. The five spots in the charge train were significantly more intense in fibres (5.4 -9.3×), while the lower molecular weight spot was nearly similar in abundance in both types of tissues (1.15×). Within the charge train, peptides from three spots aligned with a chickpea β-galactosidase as the highest scoring match. This chickpea β-galactosidase has previously demonstrated exo- and endo- cleavage activity towards the side-chains of pectins and is found in elongating hypocotyls (Esteban et al., 2003; Esteban et al., 2005). In developing flax fibres, the deposition of a rhamnogalactan-type pectin consisting of 55-85 % galactose is known to precede establishment of the crystalline, cellulosic fibrils that characterize the mature secondary wall (Gorshkova and Morvan, 2006). Because the galactose residues of rhamnogalactans are one of the putative substrates for β-galactosidase, I speculate that the abundance of this enzyme in developing fibres is evidence of an important role for it in remodeling, removing, or recycling of galactans as part a dynamic process of cell wall deposition. However, it is also possible that the detected β-galactosidase hydrolyzes other galactosyl bonds, such as those that decorate arabinogalactan proteins (Kotake et al., 2005). Finally, the appearance of the β-galactosidase spots in a train along the axis of the first dimension separation of our electrophoretic gels is consistent with extensive post-translational modification of this abundant protein.

In addition to β-galactosidase, I also identified other spots representing one or more enzymes with possible roles in the metabolism of pectic polysaccharides. Three spots (#18, 4.1×; #19, 6.6×; #104, 3.0×) were more enriched in fibres as compared to the non-fibre fraction and share homology with UDP-rhamnose synthase. Because these

enzymes would normally be expected to contribute to the growth of rhamnogalactans, it is interesting to observe their enrichment in the same cells in which β-galactosidase might hydrolyze galactosidic bonds within these polymers. The potential co-existence of both catabolic and anabolic processes of galactan metabolism is consistent with a rapid turnover of these polymers during cell wall deposition, although the existence of the inferred enzymatic activities must still be confirmed experimentally.

2.3.2.4 One-carbon metabolism

Four enzymes associated with one-carbon (1C) metabolism were identified among the fibre-enriched protein spots in our study. Three of these: methionine synthase (#9, #10; 2.0×, 2.2× respectively), methionine adenosyltransferase (#60; 2.1×), and adenosylhomocysteinase (#41; 1.6×) are components of the S-adenosyl methionine (SAM) cycle, while the remaining protein, serine hydroxymethyltransferase (#53; 2.2×), catalyzes the transfer of carbon into the SAM cycle, via folate. Because the cumulative function of these enzymes is to provide activated methyl groups for transfer to acceptors, the identity of the major methyl transferases and their substrates in fibres is an obvious question. In plants, potential acceptors of activated methyl groups include a wide variety of molecules, among them components of pectin or lignin (Moffatt and Weretilnyk, 2001). Because the amount of lignin present in flax fibres is low in comparison to other types of schlerenchyma, particularly at the early stage of cell wall development associated with the snap point, (Gorshkova et al., 2000; Day et al., 2005), it seems unlikely that lignin is the major sink for methyl flux through the SAM cycle. Thus, pectin or other actively accumulating substances may be targets for SAM-mediated methylation in developing fibres.

2.3.2.5 Membrane transport

Only a few proteins related to transport across membranes were detected in our study. This may be due in part to the difficulty of extracting and resolving certain membrane-associated proteins. Nevertheless, I identified a K^+ channel β-subunit that was highly enriched (#97; 8.6×) in fibres, as well as two porins (#102, #102; 1.7×, 3.9×, respectively). The biological significance of the porins is unclear, however, increased

expression of K^+ channels has been previously correlated with sucrose uptake in developing cotton fibres. Thus the strong enrichment of K^+ channel proteins that were observed may reflect a similar process of the uptake of reduced carbon in flax fibres (Ruan et al., 2001; Ruan, 2007).

2.3.2.6 Cytoskeleton and secretion

Structural components of the cytoskeleton, as well as proteins related to vesicle traffic, were also relatively more abundant in fibre protein extracts as compared to surrounding tissues. The relative enrichment of at least 1.5-fold of actin (#69, #70) and tubulin (#37) in fibres was observed. These proteins may be enriched in fibres, as compared to cells of the non-fibre fraction, due in part to the differences in architecture and surface/volume ratios of these cells. Additionally, increased relative abundance of cytoskeleton proteins in fibres undergoing cell wall thickening may reflect the role of the cytoskeleton in deposition of cellulose and other cell wall components. An active secretory system, which delivers non-cellulosic polysaccharide components to the cell wall, is also expected to be present in developing flax fibres; the enrichment of myosin (#5, 2.5×; #6, 3.6×), dynamin-like proteins (#22, 3.1×), and GDP-dissociation inhibitor (#55, 2.0×; #56, 1.9×) in these cells is therefore consistent with developmental processes presumed to be active in the cells that were sampled. I also note that other components of the cytoskeleton mentioned in a structural context above (i.e. actin and tubulin) may have additional functions specifically related to secretion and other aspects of secondary wall deposition (Roberts et al., 2004; Oda and Hasezawa, 2006; Boutte et al., 2007).

2.3.2.7 Protein and amino acid metabolism

Enzymes related to protein metabolism (e.g. protein synthesis and folding) were moderately enriched (1.5×– 2.7×) in fibres as compared to the non-fibre fraction. Two translation initiation factors were more abundant in the fibre sample: eIF-4A (#62, 1.6×) and eIF-5A (#114, 2.0×). Proteins in the eIF-4A family form part of the ribosomal machinery and are involved in binding and unwinding mRNA for translation, while some eIF-5A isoform family members have more diverse functions in cell division and related processes (Thompson et al., 2004). A translational elongation factor EF2 (#4, 2.5×) was

also more abundant in fibres, while spots containing EF1α were similarly abundant (#67, 1.2×) or 2.3× fold less abundant (#54) in fibres as compared to the non-fibre fraction. Heat shock proteins HSP60 (#29, 2.7×; #30, 1.5×; #32, 2.1×), HSP70 (#20, 1.9×; #21, 1.7×; #23, 2.0×), and HSP90 (#11, 1.7×) were also enriched in fibres. These proteins may function in the processes of cytosolic protein folding and protein import into mitochondria and chloroplasts, which are commonly associated with members of the HSP60, HSP70, and HSP90 families (Young et al., 2004). Additionally, because HSP70s have been shown to have specific functions in cell wall development in yeast, the possibility that some of these proteins are active at the plasma membrane during the deposition of the flax fibre secondary wall cannot be excluded (Lopez-Ribot and Chaffin, 1996; Nombela et al., 2006).

2.3.2.8 Miscellaneous

Several of the proteins I identified could not be classified into any of the larger functional categories I have already described. Eight of these proteins were enriched by 1.5× ($p \leq 0.05$) or more in fibres, and may accordingly have specific roles in fibre development. These included annexins (#97, 2.2×; #98, 4.1×), enoyl-ACP reductase (#100, 2.1×), maturase K (#112, 3.4×), a 14-3-3 protein (#108, 2.6×), peroxidase (#85, 2.4×), and a protein kinase C inhibitor (#107, 2.8×). Among these, the enrichment of annexin in developing fibres is particularly interesting, given its previous association with cellulose synthase in structural and proteomic studies of cotton fibres (Hofmann et al., 2003; Yao et al., 2006).

2.3.3 Comparison to transcriptomic analysis

The experimental approach used in the present study differs in many ways from the previously reported microarray analysis of flax stems (Roach and Deyholos, 2007). Importantly, in the previous report, they did not dissect fibres away from other stem tissues; rather they compared transcript abundance in stem segments containing fibres at different stages of development. Therefore, a global comparison of these datasets is not warranted. Notwithstanding these limitations, it is notable that three carbohydrate-related enzymes were detected both as proteins enriched in fibres from the snap-point region of

the stem, and previously as transcripts expressed in the region of the stem containing the snap-point, including β-galactosidase (#12-16, #64), fructokinase (#93, #94), and GME (#101) (table 2-2). In the transcriptomic data, β-galactosidase and fructokinase were significantly more abundant in the region of the snap-point as compared to segments from nearer either the apex or base of the stem, while GME showed highest transcript abundance in the apical-most segment, which may be due to differences in the turnover of these various gene products. On the other hand, the previous work also identified many other snap-point enriched transcripts that were not detected as proteins in that study. These include arabinogalactan proteins and lipid transfer proteins that were further demonstrated by qRT-PCR to be enriched specifically in the phloem tissues of the snap-point, as compared to leaves or the xylem of stems. Discrepancies between transcriptomic and proteomic analyses have been previously documented, and are presumably due to differences in efficiencies of extraction and detection of various proteins, among many other technical and biological factors (Jiang et al., 2007). For example, Bayer et al. specifically noted under representation of AGPs and other cell wall proteins within their proteomic analysis, due possibly to the high degree of glycosylation of these proteins (Bayer et al., 2006). Thus, it appears likely that a comprehensive description of gene expression within developing flax fibres cannot be provided by either transcript or protein profiling, alone, but instead the results of many different experimental approaches must be considered together.

2.3.4 Technical limitations

This experiment generated some exciting results but, like most science, steps in this procedure have possible improvements. I will discuss alterations within the sample preparation, as this can provide the largest difference. In the current study a protein extraction method (TCA/acetone precipitation) was used that produced proteins mostly related to basic housekeeping actions. Although interesting, they may not be directly to cell wall biosynthesis. TCA/acetone precipitation is known to precipitate most proteins with the exception of the very hydrophobic (lost upon re-solublization) and the very small (never precipitate). This precipitation method is commonly used with plant tissues in order to concentrate the samples and remove interfering compounds, such as phenolics.

In the current study this TCA/acetone precipitation was used to precipitate proteins directly from the mechanically pulverized tissue. It is apparent now, that a few additional steps should be included. First, to remove most of the easily-extracted, soluble, metabolic proteins before precipitation, I suggest grinding the tissues with a sodium phosphate buffer to remove them from the debris; previous studies have used this sequential approach for membrane-fraction enrichment (Gifford and Taleisnik, 1994; Xu and Tsao, 1997). Secondly, I would add a membrane disruption step prior to precipitation. A traditional extraction method is boiling in the detergent SDS; although very good at extracting hydrophobic proteins, it seriously interferes with the first dimension separation and is very difficult to remove from the sample. A better solution would be to grind the remaining debris in a solution of increased salt. It is excellent at disrupting membranes and it can be removed at three different steps. Desalting can be completed prior to precipitation using desalting columns, during precipitation, and during first dimension electrophoresis by changing paper wicks regularly throughout the run; I suggest adding all three steps to the procedure to assure gel reproducibility.

Protein extraction is a very important step in the overall success of a 2DE experiment. I suggest removing 'buffer-soluble' proteins and enhancing membrane disruption prior to 'clean-up' by precipitation. These suggestions in the addition to diligent salt removal should enrich membrane bound proteins while allowing their successful labelling by CyDyes and reproducible separation on a 2D gel.

2.4 Conclusions

I have described a differential proteomic profile of a single plant cell type at a well-defined developmental stage, during which secondary cell wall biogenesis is occurring. The fibre-enriched proteins that were identified are consistent with the dynamic process of secondary wall deposition previously suggested by histological and biochemical analyses, and particularly the importance of galactans and the secretory pathway in this process (Gorshkova and Morvan, 2006). The apparent abundance of amylase suggests that starch may be an unappreciated source of materials for cell wall

biogenesis. Furthermore, observations in this study confirm previous reports that correlate accumulation of proteins such as annexins, and specific heat shock proteins with secondary cell wall deposition (Gorshkova and Morvan, 2006; Nombela et al., 2006; Yao et al., 2006). Together, the proteins that have been identified in this study provide a basis for better understanding the unique properties of phloem fibre secondary cell walls, and define targets for detailed genetic and biochemical analyses in future.

2.5 Figures and tables

Table 2-1 Experimental design relative to labelling and sample loading of analytical gels.

gel #	Cy2 labeled		Cy3 labeled		Cy5 labeled	
1	internal standard #1	30 µg	fibre sample #1	30 µg	non-fibre sample #1	30 µg
2	internal standard #2	30 µg	fibre sample #2	30 µg	non-fibre sample #2	30 µg
3	internal standard #3	30 µg	non-fibre sample #3	30 µg	fibre sample #3	30 µg
4	internal standard #4	30 µg	non-fibre sample #4	30 µg	fibre sample #4	30 µg

Note: each gel contains proteins from a unique pool (#1-#4) of independently grown plants. The Cy2-labeled internal standard is a mixture of equal masses of proteins from fibre and non-fibre samples.

Figure 2-1 A typical flax plant at the time of fibre extraction.
a) The 3 cm region of the stem from which fibres were dissected is indicated by the bracket. b) Detail of a transverse section of fresh stem tissues at the time of harvest. This hand section was obtained from just below the snap-point to demonstrate the arrangement of tissues within the stem, i.e. transverse sectioning was not used when obtaining tissues for protein analysis. A bracket indicates the region of the cortex from which the fibre and non-fibre fractions would be obtained. The position of representative fibres within the cortex is shown by arrowheads. The scale bar is 100 µm. c) Stem tissues during dissection. Fibres from which surrounding, non-fibres cells been partially removed are indicated by arrowheads. A fully dissected fibre, comprising a single cell is indicated by the arrow. This fibre is representative of the cells from which proteins were extracted. The scale bar is 100 µm.

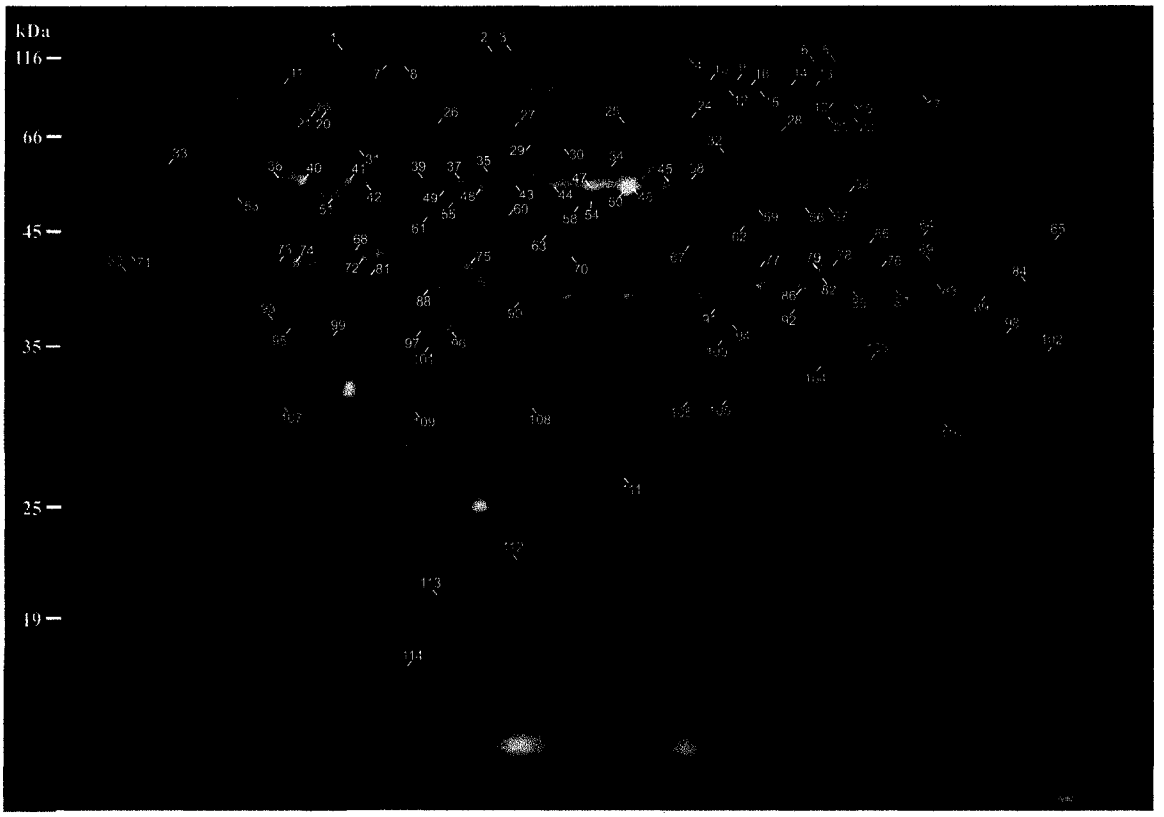


Figure 2-2 Representative analytical DiGE gel, displayed as a false-colour overlay. Proteins extracted from fibre and surrounding non-fibre tissues pre-labelled with differentially fluorescing dyes, then mixed and separated simultaneously using 2D gel electrophoresis. Labels correspond to protein spot numbers used in table 2-2 and in the text. The pH range of the first dimension separation is from 3 (left) to 10 (right). See figures A3-1 and 2 for original images.

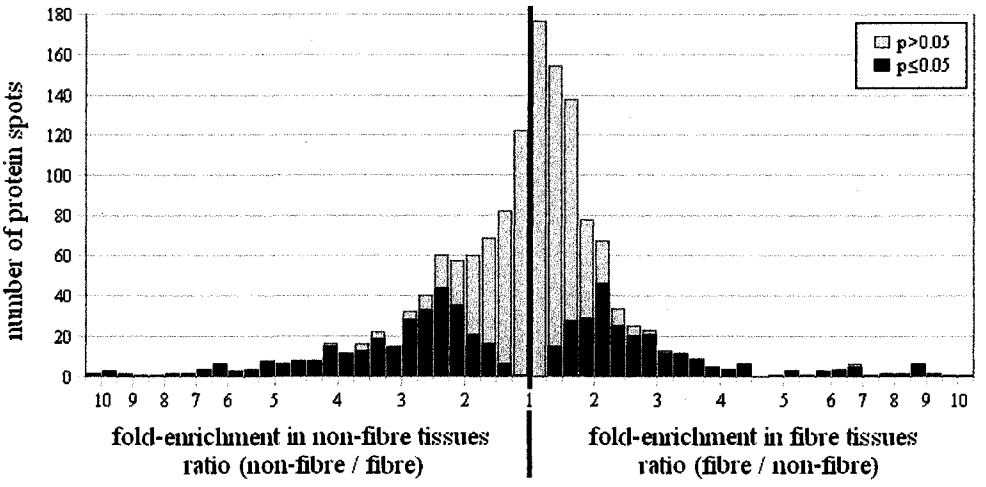


Figure 2-3 Frequency distribution of mean intensity ratios for all spots. A mean ratio near 1 meant the spot was found in equal abundance in both tissues; spots represented to the right of this point on the axis had higher signal intensity in fibre tissues, while spots represented to the left were more intense in non-fibre tissues. The grey and black regions of each bar show the portion of spots for which $p>0.05$ and $p\leq0.05$, respectively, in a t-test of the significance of differences in intensity between fibre and non-fibre tissues.

Table 2-2 Protein identities based on peptide matches to Genbank protein databases.

Spot ID #	func. cat.[a]	protein match identity	spot fold enrichment[b] fibre	non-fibre	p-value[c]	MOWSE score[d]	pept. count[e]	Genbank ID
2	C&E	aconitate hydratase	1.5		0.14	64	2	4586021
3	C&E	aconitate hydratase	1.5		0.08	68	2	4586021
17	C&E	β-amylase	8.8		<0.01	46	2	1771782
39	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	1.5		0.25	85	2	168312
40	C&E	ribulose-1,5-bisphosphate carboxylase large subunit		2.0[f]	0.08	180	4	168312
44	C&E	ribulose-1,5-bisphosphate carboxylase large subunit		6.1	<0.01	129	5	1834444
45	C&E	ribulose-1,5-bisphosphate carboxylase large subunit		5.4	<0.01	130	4	2687483
46	C&E	ribulose-1,5-bisphosphate carboxylase large subunit		2.9	<0.01	232	6	6983900
47	C&E	ribulose-1,5-bisphosphate carboxylase large subunit		3.3	<0.01	250	5	1817560
48	C&E	enolase	1.1		0.65	265	7	9581744
49	C&E	enolase		1.1	0.93	158	3	8919731
50	C&E	enolase		3.4	0.02	206	6	9581744
51	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	2.8		0.04	103	4	4098530
57	C&E	fumarate hydratase	2.5		0.01	83	2	108708038
58	C&E	fumarate hydratase	1.6		0.33	100	4	15226618
59	C&E	6-phosphogluconate dehydrogenase	1.5		0.19	100	3	2529229
63	C&E	citrate synthase	3.7		<0.01	123	4	11066954
68	C&E	phosphoglycerate kinase	1.2		0.56	257	4	1161600
71	C&E	phosphoglycerate kinase	1.7		0.06	426	7	92872324
72	C&E	ribulose-1,5-bisphosphate carboxylase large subunit				96	3	66735801
73	C&E	rubisco activase		6.1	<0.01	70	3	13430332
74	C&E	rubisco activase		5.2	<0.01	61	3	170129
75	C&E	phosphoglycerate kinase		2.9	0.02	250	6	3328122
77	C&E	fructose-bisphosphate aldolase	1.1		0.82	155	3	15227981
78	C&E	fructose-bisphosphate aldolase	2.4		0.03	102	2	20204
79	C&E	fructose-bisphosphate aldolase	1.1		0.6	116	2	15227981
80	C&E	fructose-bisphosphate aldolase	1.3		0.04	177	3	20204
81	C&E	rubisco activase		2.2	0.03	60	2	4261547
82	C&E	succinate-CoA ligase	2.3		0.02	253	5	15225353
83	C&E	glyceraldehyde-3-phosphate dehydrogenase	2.6		0.01	76	2	120666
86	C&E	glyceraldehyde-3-phosphate dehydrogenase	1.1		0.49	71	3	3023813
87	C&E	glyceraldehyde-3-phosphate dehydrogenase	3.8		<0.01	215	6	74419004
90	C&E	malate dehydrogenase	1.6		0.17	241	4	18297
91	C&E	malate dehydrogenase	1.4		0.26	138	4	18297
92	C&E	malate dehydrogenase	3.3		<0.01	296	7	10334493
93	C&E	fructokinase	6.7		<0.01	142	5	31652274
94	C&E	fructokinase	2.2		<0.01	154	3	31652274
96	C&E	kinase/ ribokinase, potential fructokinase	2.0		0.01	208	8	15224669
1	ATP	AAA-ATPase	1.6		0.24	322	10	86212372
7	ATP	ATPase, transitional endoplasmic reticulum	1.2[f]		0.65	101	4	7378614
24	ATP	vacuolar proton-ATPase	2.6		0.02	585	13	50251203
31	ATP	ATP binding		1.0	0.87	100	4	15221770
35	ATP	F1 ATPase	1.6		0.05	143	6	12986
40	ATP	ATP synthase β subunit		2.0[f]	0.08	192	4	21684923
42	ATP	ATP synthase β subunit	1.0		0.99	675	12	19685
43	ATP	ATP synthase β subunit		1.8	0.06	307	7	56784991
99	ATP	F1-ATPase gammma subunit	1.0		0.66	84	3	303626
105	ATP	vacuolar V-H$^+$ATPase subunit E	1.8		0.01	53	2	5733660
106	ATP	vacuolar V-H$^+$ATPase subunit E		1.1	0.82	100	4	5733660

continued next page

Table 2-2 continued.

Spot ID #	func. cat[a]	protein match identity	spot fold enrichment[b] fibre	non-fibre	p-value[c]	MOWSE score[d]	pept. count[e]	Genbank ID
12	CWP	β-galactosidase	8.4		<0.01	43	3	115437888
13	CWP	β-galactosidase	8.9		<0.01	42	2	3641863
14	CWP	β-galactosidase	5.4		<0.01	105	5	3641863
15	CWP	β-galactosidase	8.8		<0.01	96	5	3641863
16	CWP	β-galactosidase	9.3		<0.01	72	4	34913072
18	CWP	MUCILAGE-MODIFIED 4	4.1		<0.01	57	2	42562732
19	CWP	rhamnose biosynthetic enzyme	6.6		<0.01	100	6	108707484
27	CWP	phosphoglucomutase	1.8		0.15	170	4	12585309
28	CWP	phosphoglucomutase	3.7		0.02	122	5	6272281
36	CWP	UDP-glucose pyrophosphorylase		1.4	0.37	82	3	6136112
38	CWP	UDP-glucose pyrophosphorylase		3.5	0.01	166	6	82659609
41	CWP	UDP-glucose pyrophosphorylase	1.6[f]		0.1	129	6	9280626
64	CWP	β-galactosidase	1.2		0.51	72	2	3641863
76	CWP	NAD-dependent epimerase/dehydratase (UXS6)	6.1		<0.01	109	4	15226950
88	CWP	UDP-glucose 4-epimerase		1.1	0.84	60	2	12643850
101	CWP	GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase	2.3		<0.01	155	3	12324315
104	CWP	dTDP-D-glucose 4,6-dehydratase-like	3.0		<0.01	56	2	50253123
9	1C	Met synthase	2.0		<0.01	222	6	77556633
10	1C	Met synthase	2.2		<0.01	105	3	8439545
41	1C	S-adenosyl-L-homocysteine hydrolase	1.6[f]		0.1	174	5	1710838
53	1C	serine hydroxymethyltransferase	2.2		0.02	129	4	11762130
60	1C	Met adenosyltransferase	2.1		0.02	94	4	37051117
55	MemT	GDP dissociation inhibitor	2.0		0.13	212	5	8439465
56	MemT	GDP dissociation inhibitor	1.9		0.08	158	4	8439465
95	MemT	K^+ channel β-subunit	8.6		0.01	132	4	15219795
102	MemT	34 kDa outer mitochondrial membrane porin-like	1.7			55	2	83283993
103	MemT	36 kDa porin I	3.9		<0.01	104	4	515358
5	C&S	myosin heavy chain	2.5		0.05	46	2	108710464
6	C&S	myosin heavy chain	3.6		0.01	48	2	T00727
22	C&S	dynamin central region	3.1		0.09	83	3	92891191
25	C&S	dynamin-like		1.0	0.77	143	4	21593776
37	C&S	β-tubulin	1.8		0.06	161	6	295851
52	C&S	tubulin/FtsZ family, GTPase domain	1.7		0.06	367	12	62734655
69	C&S	actin	3.1		0.01	281	8	32186910
70	C&S	actin	1.5		0.21	459	12	15242516
4	P&AA	elongation factor EF-2	2.5		0.02	40	3	6056373
7	P&AA	ClpC protease	1.2[f]			81	4	4105131
8	P&AA	ClpC protease		1.7	0.06	286	11	18423214
11	P&AA	HSP 90	1.7		0.01	312	10	1708314
20	P&AA	HSP 70-3	1.9		0.08	404	11	38325815
21	P&AA	HSP 70	1.7		0.11	612	13	62733235
23	P&AA	HSP 70	2.0		0.04	100	3	22636
29	P&AA	chaperonin CPN60-1	2.7		0.04	139	6	108706134
30	P&AA	chaperonin CPN60-1	1.5		0.04	327	7	108706134
32	P&AA	HSP 60	2.1		0.02	140	4	16221
54	P&AA	eukaryotic elongation factor 1A		2.3	0.02	227	7	24371059
61	P&AA	26S protease regulatory subunit	2.1		<0.01	85	3	1709798
62	P&AA	translation initiation factor eIF-4A	1.6		0.09	262	9	475221

continued next page

Table 2-2 continued.

Spot ID #	func. cat.[a]	protein match identity	spot fold enrichment[b] fibre	non-fibre	p-value[c]	MOWSE score[d]	pept. count[e]	Genbank ID
65	P&AA	26S proteasome subunit P45	1.9		0.11	90	3	92870338
66	P&AA	aminomethyltransferase		3.7	<0.01	67	2	3334196
67	P&AA	elongation factor-1 alpha	1.2		0.5	54	3	396134
72	P&AA	glutamine synthetase		1.7[f]	0.26	119	4	121341
84	P&AA	PO ribosomal protein		2.5	<0.01	155	3	1143507
89	P&AA	glutamate-ammonia ligase	1.2[f]		0.5	65	3	99698
114	P&AA	eukaryotic translation initiation factor 5A	2.0		0.04	91	2	8778393
26	misc	nucleolar protein NOP5	1.4		0.37	47	2	108708132
33	misc	ferric leghemoglobin reductase	1.6		0.2	124	4	5823556
34	misc	calreticulin		1.0	0.9	78	3	3288109
85	misc	peroxidase	2.4		0.03	214	7	1389835
89	misc	type IIIa membrane protein cp-wap13				58	3	2218152
97	misc	annexin	2.2		0.01	146	4	1429207
98	misc	annexin	4.1		0.03	71	2	1429207
100	misc	enoyl-ACP reductase	2.1		0.01	44	2	2204236
107	misc	protein kinase C inhibitor	2.8		<0.01	97	5	20062
108	misc	14-3-3 protein	2.7		0.01	44	3	695767
109	misc	guanine nucleotide regulatory protein	1.5		0.31	64	2	395072
110	misc	NAD(P)H dependent 6'-deoxychalcone synthase		1.1	0.82	56	3	18728
111	misc	inorganic pyrophosphatase	2.8		0.02	148	3	48927683
112	misc	maturase K	3.4		0.01	55	2	33332553
113	misc	CBS (cystathionine β-synthase) domain-containing	1.6		0.1	92	2	15238284

a) Functional category: ATPases (ATP); Cell wall polysaccharide metabolism (CWP); Cytoskeleton and secretion (C&S); Membrane transport (MemT); Miscellaneous (misc); One-carbon metabolism (1C); Primary carbon and energy metabolism (C&E); Protein and amino acid metabolism (P&AA). Only the highest scoring protein for each spot is categorized.

b) Fold enrichment in fibre tissues or non-fibre tissues as compared to the other tissue type, expressed as linear ratio of mean signal intensities.

c) P-value for a t-test of significant differences in mean signal intensities between fibre and non-fibre tissues.

d) MOWSE scores are $-10*LOG_{10}(P)$, where P is the absolute probability the match is a random event; all listed matches are significant ($p<0.05$) with respect to the NCBInr database population sampled.

e) Peptide count, i.e. the number of peptides per spot that match the Genbank ID shown.

f) Spots in which multiple proteins were identified. The intensity ratios shown may be due to differences in abundance of more than one protein.

Protein identities are sorted by functional category, in the order in which each category is presented in the text and then alphabetically within each functional category.

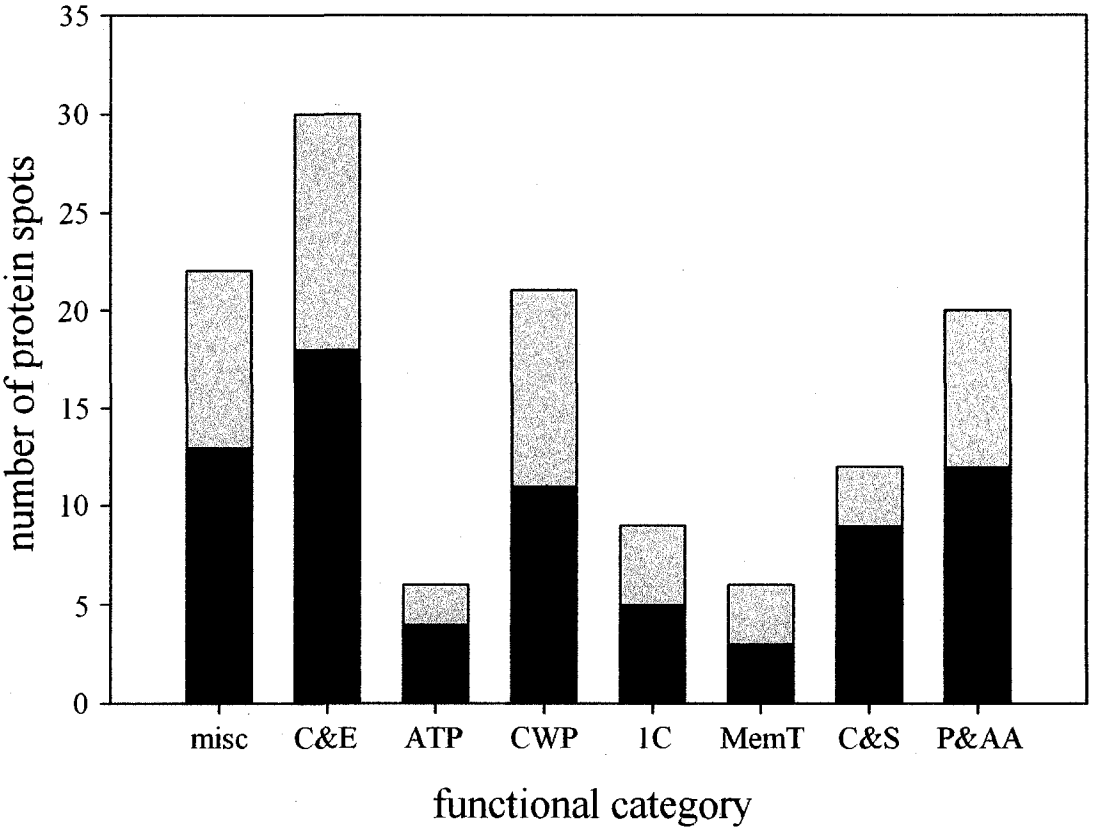


Figure 2-4 Functional categorization of fibre-enriched proteins.
All spots for which signal intensity was at least 1.5-fold greater in fibres as compared to non-fibres, and for which identity could be assigned by MS, were assigned to one of the categories shown. The grey and black regions of each bar show the portion of spots for which $p>0.05$ and $p\leq0.05$, respectively, in a t-test of the significance of differences in intensity between fibre and non-fibre tissues. ATPases (ATP); Cell wall polysaccharide metabolism (CWP); Cytoskeleton and secretion (C&S); Membrane transport (MemT); Miscellaneous (misc); One-carbon metabolism (1C); Primary carbon and energy metabolism (C&E); Protein and amino acid metabolism (P&AA).

2.6 Endnote

A version of this chapter has been published. Hotte and Deyholos, 2008. BMC Plant Biology. 8: 52.

2.7 Bibliography

Bayer, E.M., Bottrill, A.R., Walshaw, J., Vigouroux, M., Naldrett, M.J., Thomas, C.L., and Maule, A.J. (2006). Arabidopsis cell wall proteome defined using multidimensional protein identification technology. Proteomics **6,** 301-311.

Blee, K.A., Wheatley, E.R., Bonham, V.A., Mitchell, G.P., Robertson, D., Slabas, A.R., Burrell, M.M., Wojtaszek, P., and Bolwell, G.P. (2001). Proteomic analysis reveals a novel set of cell wall proteins in a transformed tobacco cell culture that synthesises secondary walls as determined by biochemical and morphological parameters. Planta **212,** 404-415.

Boutte, Y., Vernhettes, S., and Satiat-Jeunemaitre, B. (2007). Involvement of the cytoskeleton in the secretory pathway and plasma membrane organization of higher plant cells. Cell Biology International **31,** 649-654.

Day, A., Ruel, K., Neutelings, G., Cronier, D., David, H., Hawkins, S., and Chabbert, B. (2005). Lignification in the flax stem: evidence for an unusual lignin in bast fibers. Planta **222,** 234-245.

Esau, K. (1943). Vascular differentiation in the vegetative shoot of Linum. III. The origin of the bast fibers. American Journal of Botany **30,** 579-586.

Esteban, R., Labrador, E., and Dopico, B. (2005). A family of B-galactosidase cDNAs related to development of vegetative tissue in *Cicer arietinum*. Plant Science **168,** 457-466.

Esteban, R., Dopico, B., Munoz, F.J., Romo, S., Martin, I., and Labrador, E. (2003). Cloning of a *Cicer arietinum* beta-galactosidase with pectin-degrading function. Plant and Cell Physiology **44,** 718-725.

Gifford, D.J., and Taleisnik, E. (1994). Heat-shock response of *Pinus* and *Picea* seedlings. Tree Physiology **14,** 103-110.

Gorshkova, T.A., and Morvan, C. (2006). Secondary cell-wall assembly in flax phloem fibres: role of galactans. Planta **223,** 149-158.

Gorshkova, T.A., Sal'nikova, V.V., Chemikosova, S.B., Ageeva, M.V., Pavlencheva, N.V., and van Dam, J.E.G. (2003). The snap point: a transition point in *Linum usitatissimum* bast fiber development. Industrial Crops and Products **18,** 213-221.

Gorshkova, T.A., Salnikov, V.V., Pogodina, N.M., Chemikosova, S.B., Yablokova, E.V., Ulanov, A.V., Ageeva, M.V., Van Dam, J.E.G., and Lozovaya, V.V. (2000). Composition and distribution of cell wall phenolic compounds in flax (*Linum usitatissimum* L.) stem tissues. Annals of Botany **85,** 477-486.

Haas, T.J., Sliwinski, M.K., Martinez, D.E., Preuss, M., Ebine, K., Ueda, T., Nielsen, E., Odorizzi, G., and Otegui, M.S. (2007). The arabidopsis AAA ATPase SKD1

is involved in multivesicular endosome function and interacts with its positive regulator LYST-INTERACTING PROTEIN5. Plant Cell **19,** 1295-1312.

Hofmann, A., Delmer, D.P., and Wlodawer, A. (2003). The crystal structure of annexin Gh1 from *Gossypium hirsutum* reveals an unusual S-3 cluster - Implications for cellulose synthase complex formation and oxidative stress response. European Journal of Biochemistry **270,** 2557-2564.

Jiang, Y., Yang, B., Harris, N.S., and Deyholos, M.K. (2007). Comparative proteomic analysis of NaCl stress-responsive proteins in arabidopsis roots. Journal of Experimental Botany **58,** 3591-3607.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research **28,** 27-30.

Kotake, T., Dina, S., Konishi, T., Kaneko, S., Igarashi, K., Samejima, M., Watanabe, Y., Kimura, K., and Tsumuraya, Y. (2005). Molecular cloning of a beta-galactosidase from radish that specifically hydrolyzes beta-(1->3)- and beta-(1->6)-galactosyl residues of arabinogalactan protein. Plant Physiology **138,** 1563-1576.

Lev-Yadun, S. (2001). Intrusive growth - the plant analog of dendrite and axon growth in animals. New Phytologist **150,** 508-512.

Lopez-Ribot, J.L., and Chaffin, W.L. (1996). Members of the Hsp70 family of proteins in the cell wall of *Saccharomyces cerevisiae*. Journal of Bacteriology **178,** 4724-4726.

Loyola-Vargas, V.M., Broeckling, C.D., Badri, D., and Vivanco, J.M. (2007). Effect of transporters on the secretion of phytochemicals by the roots of *Arabidopsis thaliana*. Planta **225,** 301-310.

McCorriston, J. (1997). The fiber revolution: Textile extensification, alienation, and social stratification in ancient Mesopotamia. Current Anthropology **38,** 517-549.

McDougall, G.J. (1997). Procedure for selection of cell wall-associated glycoproteins. 1997 **45,** 633-636.

Moffatt, B.A., and Weretilnyk, E.A. (2001). Sustaining S-adenosyl-l-methionine-dependent methyltransferase activity in plant cells. Physiologia Plantarum **113,** 435-442.

Mohanty, A.K., Misra, M., and Hinrichsen, G. (2000). Biofibres, biodegradable polymers and biocomposites: An overview. Macromolecular Materials and Engineering **276,** 1-24.

Mooney, B.P., Miernyk, J.A., Greenlief, C.M., and Thelen, J.J. (2006). Using quantitative proteomics of arabidopsis roots and leaves to predict metabolic activity. Physiologia Plantarum **128,** 237-250.

Nombela, C., Gil, C., and Chaffin, W.L. (2006). Non-conventional protein secretionin yeast. Trends in Microbiology **14,** 15-21.

Oda, Y., and Hasezawa, S. (2006). Cytoskeletal organization during xylem cell differentiation. Journal of Plant Research **119,** 167-177.

Roach, M.J., and Deyholos, M.K. (2007). Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues Molecular Genetics and Genomics **278.**

Roberts, A.W., Frost, A.O., Roberts, E.M., and Haigler, C.H. (2004). Roles of microtubules and cellulose microfibril assembly in the localization of secondary-cell-wall deposition in developing tracheary elements. Protoplasma **224,** 217-229.

Ruan, Y.-L., Llewellyn, D.J., and Furbank, R.T. (2001). The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K+ transporters and expansin. Plant Cell **13,** 47-60.

Ruan, Y.L. (2007). Goldacre paper: Rapid cell expansion and cellulose synthesis regulated by plasmodesmata and sugar: Insights from the single-celled cotton fibre. Functional Plant Biology **34,** 1-10.

Schumacher, K. (2006). Endomembrane proton pumps: Connecting membrane and vesicle transport. Current Opinion in Plant Biology **9,** 595-600.

Thompson, J.E., Hopkins, M.T., Taylor, C., and Wang, T.-W. (2004). Regulation of senescence by eukaryotic translation initiation factor 5A: Implications for plant growth and development. Trends in Plant Science **9,** 174-179.

Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H.Y., Thorsson, V., Eng, J., Goodlett, D., Berger, J.P., Gunter, B., Linseley, P.S., Stoughton, R.B., Aebersold, R., Collins, S.J., Hanlon, W.A., and Hood, L.E. (2004). Integrated genomic and proteomic analyses of gene expression in mammalian cells. Molecular and Cellular Proteomics **3,** 960-969.

Vander Mijnsbrugge, K., Meyermans, H., Van Montagu, M., Bauw, G., and Boerjan, W. (2000). Wood formation in poplar: Identification, characterization, and seasonal variation of xylem proteins. Planta **210,** 589-598.

Vanzeist, W., and Bakkerheeres, J.A.H. (1975). Evidence for Linseed Cultivation before 6000 Bc. Journal of Archaeological Science **2,** 215-219.

Watson, B.S., Lei, Z.T., Dixon, R.A., and Sumner, L.W. (2004). Proteomics of *Medicago sativa* cell walls. Phytochemistry **65,** 1709-1720.

Xu, H.P., and Tsao, T.H. (1997). Detection and immunolocalization of glycoproteins of the plasma membrane of maize sperm cells. Protoplasma **198,** 125-129.

Yao, Y., Yang, Y.W., and Liu, J.Y. (2006). An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-DE. Electrophoresis **27,** 4559-4569.

Young, J.C., Agashe, V.R., Siegers, K., and Hartl, F.U. (2004). Pathways of chaperone-mediated protein folding in the cytosol. Nature Reviews Molecular Cell Biology **5,** 781-791.

Zhang, P.F., Foerster, H., Tissier, C.P., Mueller, L., Paley, S., Karp, P.D., and Rhee, S.Y. (2005). MetaCyc and AraCyc. Metabolic pathway databases for plant research. Plant Physiology **138,** 27-37.

Chapter 3: Analysis of the glycome and phosphoproteome of flax stems

3.1 Introduction

In general, post-translational modifications (PTMs) of proteins is integral to protein maturity, function and regulation. Two PTMs important for cell wall biosynthesis, among other cellular processes, are protein glycosylation and phosphorylation. My previous 2D proteomics study revealed possibilities of such PTMs and warranted and further study (Hotte and Deyholos, 2008).

Glycoproteins are integral components of cell wall biosynthesis. Extracellular glycoproteins can function in cell wall construction in addition to their roles as structural component. Many membrane bound proteins are also glycosylated and have the potential to be directly or indirectly involved in cell wall metabolism. The phosphorylation status of various isoforms, at a particular developmental stage, can be quite informative; these results may clarify regulation of entire pathways and thus add evidence to the isoform abundance patterns previously observed (Hotte and Deyholos, 2008).

The approach taken, in this study, to elucidate potentially interesting glycoproteins and phosphoproteins, is to complement our characterization study (Hotte and Deyholos, 2008) with gel-based glycomics and phosphoproteomics. Post-staining techniques were chosen that selectively stain for glycosylation or phosphorylation in a SDS-PAGE gel. A benefit to this approach is an "overall" look at the in-gel proteins for these PTMs and the ability to relate that information back to the isoform abundance data in (Hotte and Deyholos, 2008).

The post-stain Pro-Q Emerald is stain selective for glycosylations; it covalently binds a fluorescent molecule to carbohydrates through a two-step process. First, any 1, 2-dialcohol groups, such as those found in carbohydrates, are oxidized to aldehydes by periodic acid. The fluorescent molecule then binds to the resultant pair of aldehydes in separate reaction. The entire oxidation reaction and staining process for this commercial stain appears to be nearly identical to previously described protocols (Harrison et al., 1976) except Pro-Q Emerald appears to use a proprietary fluorescent molecule instead of dansyl hydrazine.

Phosphoprotein specific post-staining is theoretically a good way to visualize a phosphorylation difference among potential isoforms. The Pro-Q Diamond stain used allows for selective detection of phosphate moieties with a fluorescent dye and does not interfere with downstream mass spectral analysis (Schulenberg et al., 2003).

Both stains used in this study have some limitations and these will be discussed further in this appendix. In addition to glycoprotein and phosphoprotein results, I will attempt to describe plausible solutions to these limitations such that they can be used to further research in fibre cell biosynthesis.

3.2 Methods and materials

3.2.1 Plant material

Flax of the same variety, and grown under the same conditions as in chapter two and Hotte and Deyholos (2008), was also subject to sample collection from a similar area, below the snap-point. The main difference, from chapter two, was in the type of tissue collected. In the current study, stem peels consisting of cortical tissue external to the xylem, with exception of the epidermis, were isolated from the stems and not further separated into fibre and surrounding tissue. The stem peel samples are therefore composed of a heterogeneous mixture of cortical tissues including parenchyma, phloem fibre, sieve elements and companion cells, and possibly tissues from cambial and interfasicular regions as well. After tissue isolation, samples were rinsed with deionised water, blotted dry, flash frozen in liquid nitrogen, and stored at –80 °C until protein isolation.

3.2.2 Protein isolation and 2DE

Protein isolation, removal of impurities and quantitation proceeded exactly as described earlier (Hotte and Deyholos, 2008) by using the TCA/acetone extraction, 2D Clean-up kit (GEhealthcare) and 2D Quant kit (GEhealthcare) respectively. Following this, samples were doubled in volume with the addition of sample buffer (7 M urea, 2 M thiourea, 2 % (v/v) ampholyte, 2 % (w/v) DTT, 4 % (w/v) CHAPS) and kept on ice until isoelectric focusing. Samples, each consisting of 250 µg of total protein were loaded

onto two 3-10 NL Immobiline drystrips (GEhealthcare) by passive rehydration. Isoelectric focusing parameters, equilibration, and SDS-page also proceeded as previously outlined (Hotte and Deyholos, 2008). Protein molecular weight standards were added to each SDS gel prior to electrophoresis; one gel received the CandyCane™ Glycoprotein Molecular Weight Standard (Molecular Probes) and the other received the PeppermintStick™ Phosphoprotein Molecular Weight Standard (Molecular Probes).

3.2.3 Pro-Q Emerald 300 Glycoprotein Stain (Pro-Q Emerald)

Following electrophoresis, glycoproteins were detected using the standard protocol and reagents for Pro-Q Emerald (Molecular Probes). Briefly, the gel was fixed overnight in a 50 % methanol and 5 % acetic acid solution (v/v), and then washed twice (20 minutes each) with 3 % acetic acid. Next, the gel was incubated in an oxidizing solution of 20 mM periodic acid in 3 % acetic acid for 1 h with gentle agitation. After oxidizing, the gel was washed three more times, as above, then immersed in Pro-Q Emerald staining buffer (prepared as per the manufacturer's instructions) for 2.5 h in the dark and with gentle agitation. In order to destain, the gel was washed as above three more times before imaging. After imaging, the gel was prepared for total protein staining by immersion in 100 % methanol for 15 minutes.

3.2.4 Pro-Q Diamond Phosphoprotein Stain (Pro-Q Diamond)

Following electrophoresis, a gel was stained using the standard protocol and reagents for the phosphoprotein stain, Pro-Q Diamond (Molecular Probes; (Schulenberg et al., 2003)). Briefly, fixation was conducted overnight with a 50 % methanol and 5 % acetic acid solution. Followed by three washes with sterile deionised water, each for 15 minutes. The gel was then incubated in Pro-Q Diamond stain for 2 h, with gentle agitation in the dark. Destaining, commenced by three washes, each for 30 minutes in the recommended destain solution (20 % acetonitrile, 50 mM sodium acetate, pH 4). The gel was rinsed and stored in sterile deionised water before and after imaging.

3.2.5 SYPRO Ruby protein gel stain

After imaging the Pro-Q Emerald or Pro-Q Diamond stained gels, both gels were washed with sterile deionised water twice for 5 minutes. The gels were incubated with SYPRO Ruby gel stain solution, then destained and imaged as per the manufacturer's instructions. Firstly, the gels were immersed in staining solution, in darkness with gentle agitation, overnight; followed by destaining for 30 minutes with a 10 % methanol and 7.5 % acetic acid solution. The gels were rinsed twice with sterile deionised water prior to imaging.

3.2.6 Imaging

The Pro-Q Emerald stained gel and its subsequent SYPRO Ruby stained image were illuminated on a 365 nm transillumination platform and captured by CCD digital camera (Alpha Innotech) after filtration through a 595 nm filter.

The Pro-Q Diamond stained gel and its respective SYPRO Ruby stained image were captured on the FLA-5000 laser scanner (Fujifilm). Pro-Q Diamond stained gel was imaged using the 532 nm laser and the 575 nm emission filter, while the SYPRO Ruby image was acquired using the 473 nm laser and the 575 nm emission filter. Both images were acquired at 100 μm resolution and with maximum PMT (photon multiplier tube) voltage resulting in signal intensities values between 50000 and 63558 units.

3.2.7 Image analysis

DeCyder differential analysis software version 6.5 (GEhealthcare, DeCyder 6.5) was used for image analysis including spot-matching, spot volume calculations and back-matching to chapter two images. For the Pro-Q Diamond and respective SYPRO Ruby images, spot volumes were exported from DeCyder 6.5 for manual calculation of a spot volume ratio (spot volume in Pro-Q Diamond image to spot volume in SYPRO Ruby image).

3.2.8 Spot-picking and tryptic digestion of proteins

Spots for subsequent tryptic digestion were excised from the gels by hand using sterile, tip-removed, PCR tubes and the Alpha Imager for illumination. Excised gel plugs were transferred to a 96 well plate for further processing. Tryptic digestion was

completed as described elsewhere (Cao et al., 2008). LC/MSMS mass spectral analysis of peptides proceeded as per described previously (Cao et al., 2008; Hotte and Deyholos, 2008; Sharma et al., 2008).

3.3 Results and discussion

3.3.1 Pro-Q Emerald Glycoprotein stained gel

Based on my visual inspection of the Pro-Q Emerald stained gel, I identified 31 spots that were unambiguously stained and selected these for further analysis using LC/MSMS and database searches (figure 3-1).

Periodic acid is a strong oxidizer and has been used for decades in carbohydrate research for the conversion of dialcohols to aldehydes (Clancy and Whelan, 1959; Harrison et al., 1976; Kim et al., 2000). At high periodic acid concentrations (e.g. 40 mM) all dialcohols should be oxidized (Clancy and Whelan, 1967). At lower concentrations (0.4 mM), this acid appears to preferentially oxidize acyclic glycols, like glycerol and sialic acid (White et al., 1974; Roberts, 1977), as opposed to the cyclic variety like other sugars and sugar alcohols common to known secondary cell wall glycoproteins. To my knowledge, polysaccharide moieties of most plant-derived glycoproteins contain only cyclic glycols, like galactose, arabinose, rhamnose, glucose, mannose, N-acetylglucosamine, N-acetylgalactosamine and fucose as opposed to sugar acids with acyclic appendages. It should be noted that plants do produce sugar and sugar acids with acyclic appendages, but these have yet to be found attached to proteins (O'Neill and York, 2004).

With respect to the current study, the inherent oxidizing preference of periodic acid for acyclic glycols in combination with in-gel, non-protein-bound, acyclic molecules, may affect the staining sensitivity for plant-derived glycoproteins. Firstly, the suggested protocol was designed for mammalian derived glycoproteins, which have oligosaccharide side chains that often terminate with the 9-carbon sugar alcohol, sialic acid; these are therefore easily and preferentially oxidized by periodic acid. In plants, relatively few (if any) protein glycosylations contain sialic acid (Shah et al., 2003; Séveno et al., 2004; Takashima et al., 2006) or acyclic dialcohols. Secondly, because of

the lack of acyclic glycols in plant glycoproteins, periodic acid may have oxidized other dialcohols that were originally in the equilibration buffer but then remain in the gel after electrophoresis, rather than glycosylated proteins (Clancy and Whelan, 1967). Thirdly, the protocol suggests diluting the periodic acid below optimum concentration, from 40 mM to 20 mM to cover the area of a large format 2D gel. Thus, all things combined, the Pro-Q Emerald stain may have inherently low sensitivity in 2D gel analysis of plant glycoproteins.

If poor oxidiziation is the main problem decreasing sensitivity of the Pro-Q Emerald stain in the conditions used in the current study, I propose a few protocol alterations. Firstly, removal of potentially problematic acyclic dialcohols, like glycerol, is required. One would need to determine which molecule remains in the gel with enough concentration to be problematic. If the molecule cannot be removed or substituted, cut the top inch off the gel prior to periodic acid oxidation, as the problematic molecule appears to remain in this area of the gel after electrophoresis (figure 3-1). Secondly, if poor oxidation is the problem, allow for longer incubation times with the periodic acid to ensure that even the more difficult to oxidize cyclic dialcohols are oxidized as well. Thirdly, ensure that the concentration of periodic acid used is no less than 40 mM as this is the concentration at which previous research suggests oxidation of all dialcohols occurs (Clancy and Whelan, 1959; Roberts, 1977).

Other potential PTMs that would be in the 2D gel, like prenylation, acetylation, lipidation, methylation, and phosphorylation, lack dialcohol groups. Therefore, the proposed harsh conditions should not decrease selectivity for glycosylations. The exception may be proteins covalently bound to RNA (dialcohol located at the 5' end) and those bound to NAD(P)H or FADH (dialcohol in the sugar moiety); this is based on chemical structure similarity only and has not been suggested in literature previously. Both plant and mammal-derived proteins may have these post-translational modifications. Since these have not been noted previously as producing false-positive results under similar conditions as I propose, I suggest these modifications will not affect sensitivity either.

To test these protocol alterations, I suggest running 1D or 2D gels to establish the usefulness of this stain for plant-derived glycoproteins. It would be important to test

using crude protein extracts known to contain cell wall glycoproteins. I can not unequivocally say the lack of staining seen in my gels is not due to a lack of glycosylated proteins; this problem has been reported previously (Bayer et al., 2006). It is likely not to see mature cell wall glycoproteins in my study due to the protein extraction method used. The TCA/acetone method in combination with mortar and pestle and liquid nitrogen does not appear to provide the disruption required to release cell wall bound glycoproteins. Despite this, I would expect to see unfinished cell wall glycoproteins in this preparation as they are sequentially glycosylated in the golgi.

3.3.2 Pro-Q Diamond Phosphoprotein stained gel

Pro-Q Diamond stain allows for selective detection of phosphate moieties with a fluorescent dye and does not interfere with downstream mass spectral analysis (Schulenberg et al., 2003). It is often used to determine relative protein phosphorylation status and is used in conjunction with a gel-based proteomics approach. Pro-Q Diamond stained approximately 85% of the total spots in this experiment.

It is important to note, there are many limitations to the use of Pro-Q Diamond stain as a definitive marker for phosphorylation. Firstly, the stain can stain non-specifically under some conditions, particularly when it is nearing its expiration date (as per manufacturer's protocol). Secondly, to ensure accurate estimates of phosphorylation a quantitative phosphoprotein standard would need to be included. The manufacturer recommended set of standards contains a mixture of phosphorylated and non-phosphorylated proteins so there should be a difference between how Pro-Q Diamond stains them. The phosphorylations in this set are only quantified per 'band' and not per 'spot' so as such are only useable in a 1D gel format; it is not particularly accurate to quantitatively compare 1D and 2D spot volumes. Therefore, without a 2D-worthy set of empirical standards, it is difficult to determine a threshold at which one can say that a spot is or is not phosphorylated, on a 2D gel.

Due to these specificity problems, it is important to stain with a total-protein stain, such as SYPRO Ruby, after the Pro-Q Diamond stain in order to calculate a phosphorylation-stain to total-stain ratio (Schulenberg et al., 2003). This ratio calculation enables discussion of the relative differences among protein spot phosphorylation. For

example, highly abundant spots with few phosphorylations will have a low ratio and highly phosphorylated, low-abundance spots will have a high ratio. Additionally, ratio calculation for every spot in the gel will produce a 'gradient' of ratios, which can be used to guestimate phosphorylation thresholds, *in lieu* of a physically determined threshold.

To determine plausible thresholds in this study, a ratio was calculated for every confirmed spot and plotted against spot volume (figure 3-2a). The plotted ratios appear as distribution spanning from a ratio of 0.1 to 9.5. Based on this distribution of all spots within this study and previous work (Schulenberg et al., 2003), I determined that a ratio of 1 is a good indicator of clearly phosphorylated proteins. Further, (Schulenberg et al., 2003) describes proteins with a ratio down to 0.1 as possibly phosphorylated. It is clear that solid, empirically derived thresholds are needed, but *in lieu* of this I will cautiously describe my data. For descriptive purposes I will use "highly phosphorylated" to refer to spots with a ratio >1, "moderately phosphorylated" for spots with a ratio between 1 and 0.4, and "few, minimal or no phosphorylations" for spots with a ratio less than 0.4 (figure 3-2a, b). The calculated spot-volume ratio for each database-matched spot is listed in table 3-1; a graphical distribution of the listed ratios is in figure 3-2b.

Another note, with respect to this gel specifically, the background staining was particularly variable, which made the calculated ratios vary enormously. A visual inspection of each spot was made to determine the extent of variation and inconclusiveness of the resulting ratio; this is noted where applicable in table 3-1 and in the text. This problem may have been inconsequential with the use of different software; DeCyder 6.5 does not allow manual background subtraction but others may.

3.3.3 Protein identification by LC/MSMS

Protein spots were chosen based on differential phosphorylation or glycosylation status; a total of 38 spots were satisfactorily matched to proteins in the NCBInr database using data from LC/MSMS. Matches are significant ($p<0.05$) based on probability-based statistical calculations provided by MASCOT; the MOWSE score for each significant match is listed in table 3-1 for each spot. Of these 38 spots, nine were glycosylated, one was highly phosphorylated, eleven moderately phosphorylated, nineteen minimally phosphorylated and seven were inconclusively phosphorylated due to variable

background. Protein specific glycosylation and phosphorylation results from post-staining 2D gels are also listed in table 3-1. Spots with database matches were categorized into the following functional groups: Primary carbon and energy metabolism, light harvest, protein and amino acid metabolism, reactive oxygen transfer, secondary metabolism, ATPases, and cytoskeleton and secretion (table 3-1 and figure 3-3). Spots with matches to multiple different proteins are categorized for each hit separately.

3.3.3.1 Primary carbon and energy metabolism

There were five glycosylated protein spots that fit into category of carbon and energy transfer; all matching the large subunit of ribulose 1,5-bisphosphate carboxylase (rubisco lg) and were minimally to moderately phosphorylated. Three spots (#1, #2, and #3) all significantly matched to rubisco lg. #3 is the most abundant of the three and it also shows the most intense staining by the glycoprotein specific stain. Interestingly, #3 appears to stain more intensely, compared to its rubisco lg neighbours, than might be expected based on protein abundance alone as the contributing factor (figure 3-1). Rubisco lg is a chloroplast-encoded and translated protein; therefore it should theoretically not be glycosylated because it does not enter into the ER system. Other than glycosylation, rubisco lg has been noted to have many other PTMs like, phosphorylation, acetylation, deformylation, methylation, carbamylation, reduction and oxidation of cysteines, glutamidation, 'tight' non-covalent bonding to substrates and covalent binding to RNA (Bruce-Carver et al., 1990; Kuehn et al., 1991; Gutteridge and Gatenby, 1995; Portis and Salvucci, 2002; Yosef et al., 2004; Serafini-Fracassini and Del Duca, 2008); perhaps any one of these can explain the result.

The other two glycosylated protein spots, #4 and #5, that also matched to rubisco lg are situated at an estimated molecular weight of 24 and 23 kDa; much to small to be a common, active form of rubisco lg. Nonetheless, spots #4 and #5 are glycosylated and minimally phosphorylated. Additionally, they appear to be up regulated in the green tissue, as compared to fibre when matched back to a re-analysis of my chapter two data.

Three spots, #6, #7, and #8, were only matched rubisco lg; they were not glycosylated and were minimally to moderately phosphorylated, with ratios of 0.29, 0.50, and 0.33 respectively. Spots, #9, #10 and #11 were matched to rubisco lg as well, but

also to other proteins; these will be discussed further with the other protein hits. Two rubisco activases were matched to #12 and #13, both are unglycosylated, have minimal and moderate phosphoryations, respectively and are located in a charge train together. Rubisco activase becomes phosphorylated in response to light and this altered phosphorlyation status enables its interaction with rubisco (Portis and Salvucci, 2002; Houtz et al., 2008). Three different protein spots #14, #15, and #16, had hits to glyceraldehyde-3-phosphatase. Of these three, #14 is situated in the basic half of the gel at a pI of about 7.9 while #15 and #16 are at pH 5.5 and 5.6 respectively. #14 is minimally phosphorylated and #15 and #16 have inconclusive phosphorylations due to a neighbouring highly phosphorylated protein, see figure 3-4a. As well, #15 was matched to two different proteins; due to the positioning of this spot in a "crowded" area of the gel I suggest this spot is a result of different proteins co migrating to the same place. A phosphoglycerate kinase was matched to #17; it has only minimal phosphorylations.

3.3.3.2 Light harvesting

With relation to light harvesting, #18 was matched to a photosystem 1 oxygen evolving protein (PSBP-1) from *A. thaliana*. In this flax stem peel sample, the oxygen evolving protein appears stained by the glycoprotein stain and is only minimally phosphorylated. The molecular weight of the *A. thaliana* PSBP-1 protein is 28 kDa and the empirical weight of this spot is about 24 kDa. Additionally, this spot was also easily matched to chapter two gels and was found to be significantly more abundant (2.3x, p=0.0001) in the green tissue sample, as would be expected for a component of the chloroplasts.

Two other spots were matched to proteins normally associated with photosynthesis or light harvesting; these appeared to be unglycosylated and only minimally phosphorylated. A predicted manganese-stabilizing protein, spot #19, is normally associated with photosystem II. In regulation data generated in chapter two, this spot was significantly up regulated in the green tissue sample (2.5x, p=0.001). A Chlorophyll a/b binding protein from *Pinus sylvestris*, which is involved in light harvesting was best matched to spot #20 (this spot could not be accurately matched to chapter two gels). Previous work in spinach suggests the manganese-binding protein has

the capacity to bind GTP (Spetea et al., 2004) and a recent study suggests chlorophyll a/b binding proteins are reversibly phosphorylated as part of the regulation of light excitation balance in photosystems I and II (Kargul and Barber, 2008).

3.3.3.3 Protein and amino acid metabolism

Two spots related to protein metabolism were found to be glycosylated and also found to be either minimally phosphorylated, spot #21, or moderately phosphorylated, spot #11. Spot #21 was best matched to a protein disulfide isomerase (PDI) from *Brassica carinata* that may be involved in protein folding in the ER lumen (Marchler-Bauer et al., 2007). This protein spot was clearly glycosylated; past studies suggest glycosylations may help keep the protein in the lumen or provide physical positioning for activity (Freedman et al., 1994; Wilkinson and Gilbert, 2004). There are many reports of glycosylation of this PDI-type but very few reports of protein phosphorylation for protein disulfide isomerases in general. Nucleotide-triphosphates are integral to the reaction catalyzed by PDI, to provide energy for the reaction, but there is little evidence discussing phosphorylation as regulation for this protein family; one from rat spleen shows three tyrosine phosphorylations for that particular PDI (Donella-Deana et al., 1996). Spot #11, was identified as both rubisco lg and to calreticulin. It should be noted that this spot was annotated as calreticulin in an *A. thaliana* 2D gel in the SWISS 2D-PAGE database (O04151) (SWISS-2DPAGE, 2004). In the current analysis the possibility of rubisco lg being a contaminant is high based on the location of this spot; it is located at a similar MW as the main group of rubisco lg spots. Calreticulin has a major role in protein folding and proper glycosylation in the ER lumen and is itself a glycoprotein (Coppolino and Dedhar, 1998). Calreticulin acts as a glycosylation editor, it binds to improperly glycosylated proteins to flag them for downstream degradation (Helenius et al., 1997). It is also known to be involved in signal transduction (Williams et al., 1997) and phosphorylation can alter its activity (Mueller et al., 2008).

Peptides from three spots, #22, #23 and #24, were all matched to HSP 70 proteins. These spots were all unglycosylated and were found to have minimal or moderate phosphorylations. #23 and #24 were matched to different HSP 70 proteins from *Solanum lycopersicum* and #22 was best matched to a putative HSP 70 from *Vitis vinifera.*

Spatially, #23 and #24 were located near to each other in a spot cluster near pI 5.5, while #22 was located in a separate cluster near pI 5.2. Interestingly, #24 showed slightly lower total abundance than #23 but was clearly more phosphorylated, (figure 3-4b). An analysis of *A. thaliana* HSP proteins suggests specific expression patterns for each HSP 70 protein within the genome (Sung et al., 2001); additionally, abundance appears positively correlated with activity. A study in yeast shows that phosphorylation of HSP 70 can cause a conformational change that decreases the binding affinity of the HSP for its substrate (Bonner et al., 2000). Taken together, this suggests that the isoform or protein in spot #24 could have reduced activity, due to both its lower relative abundance and higher phosphorylation status.

I identified a further four spots that were putatively related to protein metabolism, but which did not show evidence of glycosylation and have inconclusive, minimal or moderate phosphorylations. Two spots had significant hits to ribosomal subunits: spot #25 to a 20S ribosomal subunit protease and spot #15 to a 60S large ribosomal subunit (commonly referred to as PO in eukaryotes). Based on its observed migration, the 20S ribosomal spot has an estimated molecular weight of about 28 kDa, which is similar to its theoretical a molecular weight of 27 kDa. On the other hand, the 60S subunit has an empirical weight of 42 kDa and its match has a weight of 35 kDa. With respect to phosphorylations, #25 is inconclusive due to background variation and #15 is inconclusive due overshadowing by a nearby phosphoprotein and multiple protein hits to that spot. Two spots were matched to amino acid related transferases, #26 was matched to a general aminomethyltransferase from a core eudicot plant, *flaveria trinervia,* and #27 was matched to alaninetransferase, specifically AOAT1, from *A. thaliana*. Spots #26 and #27 both have low or moderate phosphorylation ratios, 0.21 and 0.40, suggesting minimal or moderate phosphorylations for each. Pyridoxal-phosphate is a cofactor in reactions catalyzed by alanineaminotransferase, and it can be covalently bound to the protein (Heldt and Heldt, 1997; Liepman and Olsen, 2003).

3.3.3.4 Reactive oxygen interaction and transfer

There were six spots with significant database hits to proteins possibly involved in interactions with reactive oxygen species. One of these appears stained by the Pro-Q

Emerald glycoprotein stain but its phosphorylation status is inconclusive due to gel variation. This spot (#28), was matched to a secretory peroxidase protein from *Linum usitatissimum*. The theoretical pI and mw of this protein are 8 and 39 kDa; empirically this spot was observed in the basic region of the gel, around pH 8.8, and around 41 kDa in weight. A fibre-enriched (2.4x) spot was also identified as *L. usitatissimum* peroxidase in Hotte and Deyholos (2008) (spot #85 therein); that protein spot also appears at 41 kDa, as this one, but it appeared to be in a more acidic region. Studies in peanut suspension cells have isolated four different secreted peroxidases, each with N-linked glycosylations, including the interaction of a secreted β-galactosidase acting on glycans of a secreted peroxidase, but with no suggestion as to its purpose (Wan et al., 1994; Watson et al., 1998).

Spots #29 and #30 both matched to proteins in the peroxiredoxin (PRX) family and both spots appear to be unglycosylated. Spot #29 was significantly matched to a 2-cys peroxiredoxin-like protein and showed minimal phosphorylations, while #30 had a hit to chain A of peroxiredoxin D (PRX D) and had a very high phosphorylation status (figure 3-4d for spot #30). These two peroxiredoxins are from the 2-cys-PRX family, in which members tend to function as antioxidants and oxygen radical scavengers or are involved in cell signalling (Dietz, 2007). A recent study on a 2-cys-PRX protein from *Brassica*, describes both nucleotide-triphosphate binding and cysteine phosphorylation for this protein (Aran et al., 2008). Spot #31 is unglycosylated and its phosphorylation status is inconclusive due to background staining. It was best matched to chloroplast drought-induced stress protein (CDSP32) in *Solanum tuberosum*; this protein is categorically a protein disulfide oxidoreductase.

Spot #32 matched a superoxide dismutase from the core eudicot plant *Raphanus sativus*. This spot appears minimally or moderately phosphorylated, but the distinction between the two is inconclusive due to background variation. In general, this enzyme acts as a catalyst that converts superoxide radicals and hydrogen into oxygen and hydrogen peroxide (Heldt and Heldt, 1997). It is involved in numerous processes *in planta*, such as photosynthesis and signalling (Cohu and Pilon, 2007). Studies with non-plant superoxide dismutases suggest that they can be regulated themselves by

phosphorylation or can interfere with the phosphorylation status of other proteins (Csar et al., 2001; Archambaud et al., 2006)

3.3.3.5 Secondary metabolism

Spot #33 was matched to a geranylgeranyl (GG) reductase from *A. thaliana.* This spot appears unglycosylated and inconclusively phosphorylated. The molecular weight of the *A. thaliana* protein is 53 kDa with a pI of 9.0; spot #33 is situated similarly at 49 kDa with a pI of about 8.8. The *A. thaliana* GG reductase is an enzyme in the isoprenoid pathway that ultimately converts a geranylgeranyl diphosphate to a phytyl diphosphate. In *A. thaliana* it is known to be involved in the biosynthesis of chlorophyll a, phylloquinones and tocopherols; all of which have a role, to some degree, in photosynthesis (Keller et al., 1998). It is also involved in the biosynthesis of many different isoprenoid compounds such as gibberellins, carotenoids, protein glycosylation-mediating dolichols, the prenyl compound for protein prenylation, and many compounds related to plant defence (Heldt and Heldt, 1997).

3.3.3.6 ATPases

All six spots that were best matched to various ATPases or ATPase domain-containing proteins were unglycosylated. Spot #34 matched an AAA-ATPase as well as matching an AAA-ATPase in Hotte and Deyholos, 2008 (spot #1 therein); this spot was up regulated 1.6 fold in the fibre sample. In the current analysis it was found to be moderately phosphorylated. Three spots, #35, #9,and #36, all matched a F1 ATPase β-subunit. All three spots appear minimally to moderately phosphorylated with ratios of 0.38, 0.27 and 0.58, respectively. Spot #37 was best matched to subunit E of a V-H^+ATPase and it is moderately phosphorylated. Similarly, this spot was also matched to subunit E of a V-H^+ATPase in Hotte and Deyholos, 2008 (spot #105 therein); it was 1.8 fold up-regulated in the fibre sample. An adenosine kinase was matched to #15; its phosphorylation status was masked by a highly phosphorylated neighbouring protein spot (figure 3-4b).

3.3.3.7 Cytoskeleton and secretion

Two spots were matched to proteins involved in the cytoskeletion. Spot #10 matched an α-tubulin from *Prunis dulcis* and spot #38 matched an actin from *Glycine max*. Both protein spots are unglycosylated; #10 is moderately phosphorylated and #38 is minimally phosphorylated.

3.3.3.8 Spots without any significant database matches

Spots that were stained by the glycoprotein specific stain but resulted in no significant database hits after LC/MSMS and subsequent analysis are shown in figure 3-1. Of these unelucidated spots x1, x2, x3, x4, x5, x6, x7, x8 and x9 are stained by Pro-Q Emerald and can be confidently matched to the SYPRO Ruby stained gel. Some of the other unidentified spots show probable staining with the glycoprotein stain but may be only faintly stained or unstained by the SYPRO ruby total protein stain. This effect, in addition to the general lack of proteins that were stained with the glycoprotein stain, made accurate spot-matching difficult between the glycoprotein stain and Sypro ruby stain, even within the same gel. Spots x10, x11, x12, x13, x14, x15, x16, x17, x18, x19, x20, x21, and x22 were very difficult to match and pick due to these problems (figure 3-1). That aside, these spots may actually be quite highly glycosylated while not very abundant or simply not very stainable by SYPRO ruby.

With respect to phosphoprotein staining, this stain is able to clearly detect individual spots that are both low abundance and highly phosphorylated. Some interesting staining patterns were noticeable but the protein spots involved were not identified. I will show and describe just two small areas as examples. One is situated near 14 kDa and a pI less than 4.6, an obvious pattern of high and low phosphorylations in neighbouring spots are apparent (figure 3-4c). The more abundant spots, x23, and x24, are obviously less phosphorylated than spots, x25, x26 and x27; x25, x26 and x27 are also of very low total-protein abundance. Another interesting spot pattern is situated near pI 5.6 and 17 kDa (figure 3-4d). Spot #30 was in fact matched to a peroxiredoxin but the surrounding spots, such as x28 and x29, were unmatched.

3.4 Conclusions

In this study, Pro-Q Diamond and Pro-Q Emerald stains had problems with either sensitivity or selectivity. It was difficult to ascertain if Pro-Q Diamond is staining non-selectively for many of the less-stained spots or if it is accurately staining minimally phosphorylated proteins. I suggested that a better set of quantified markers be used to help differentiate between the two possibilities. Pro-Q Emerald staining resulted in fewer than expected stained spots and those that were stained were quite faint. I suggested protocol alterations, such as increased periodic acid oxidation, to ensure full staining of potential glycoproteins. There were shortcomings with the approach taken but the results do warrant and require replication.

Even though there were obvious staining problems I was able to unambiguously attain results on the glycosylation and phosphorylation status of a few protein spots. Protein disulfide isomerase, calreticulin and peroxidase were stained by the glycoprotein stain, which was consistent with previous literature about the glycosylation status of these proteins. Similarly, rubisco activase, rubisco lg, calreticulin, alanine aminotransferase, an isoform of HSP70, peroxidase, peroxiredoxin and various ATPases were stained with the phosphoprotein stain; all of which are consistent with previous literature.

An interesting result from the phosphorylation stain was the relative differences among in phosphorylation status of HSP70 isoforms. This observation is consistent with the theory that HSP70 proteins can be regulated by phosphorylation (Bonner et al., 2000). Another interesting result was a highly-phosphorylated peroxiredoxin 5-like protein spot; this result has not been reported before in the flax stem proteome. More work should be completed to characterize this spot and to realize its niche.

3.5 Figures and tables

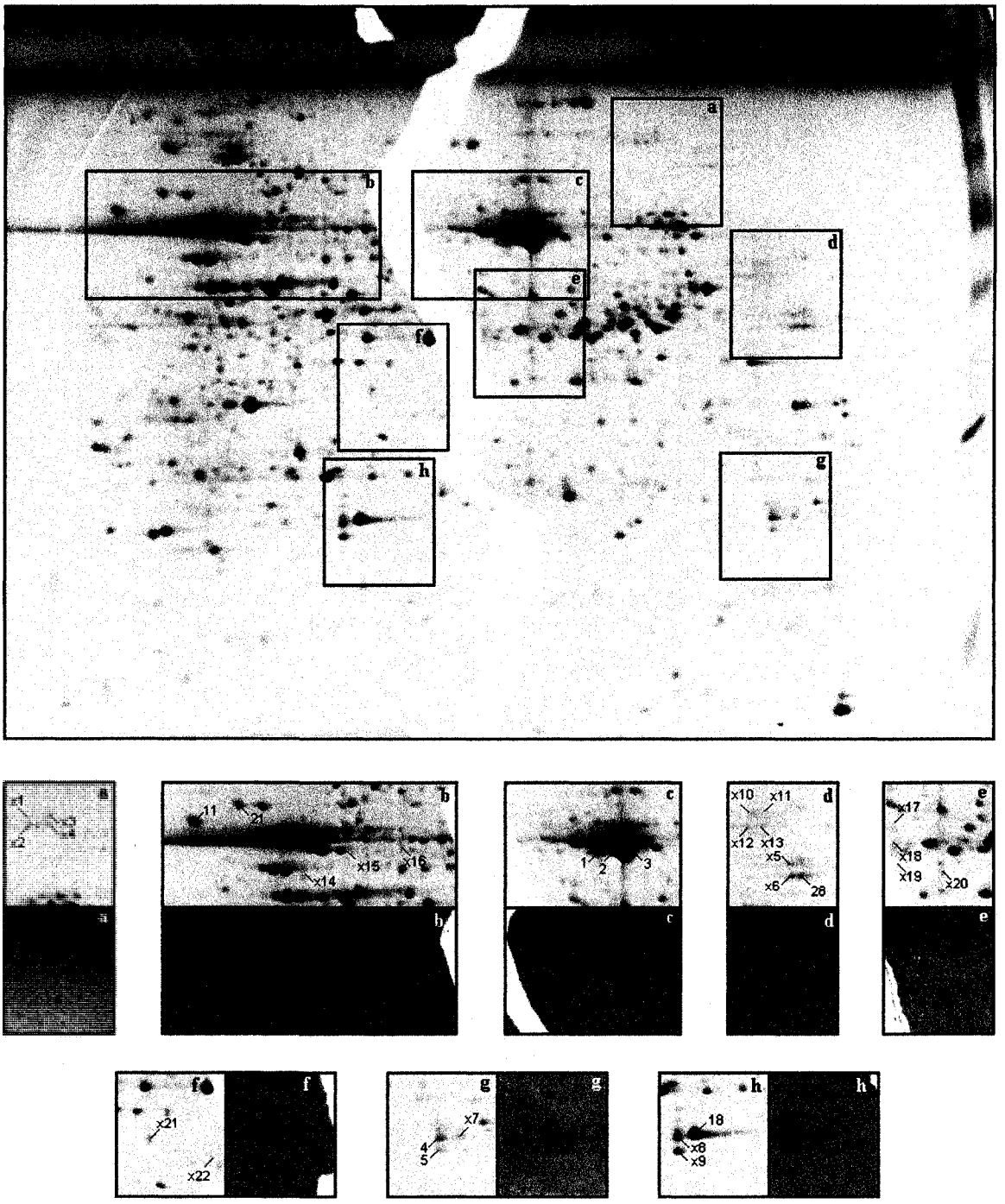


Figure 3-1 Results of Pro-Q Emerald 300 glycoprotein stain and Sypro Ruby total protein stain. Spot numbers 1, 2, 3, 4, 5, 11, 18, 21, and 28 appear glycosylated and have associated Genbank database matches. Spot numbers x1 through x22 appear glycosylated but have no associated database matches. Framed crops have coresponding letters to their placement on the large gel. The SYPRO Ruby cropped images are on top or to the left (black letters); the respective Pro-Q Emerald stained images are directly bellow and or the right (white letters). **Table 3-1** Protein

identities based on peptide matches to Genbank protein databases.

spot ID #	func. cat.[a]	protein match identity	Pro-Q Diamond ratio[b]	Pro-Q Emerald staining[c]	empirical[d] kDa	empirical[d] pI	theoretical[e] kDa	theoretical[e] pI	MOWSE score[f]	pept. count[g]	Genbank ID
1	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.40	Yes	53	6.6	51	6.5	254	5	50262298
2	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.36	Yes	53	6.7	50	6.3	138	4	33636029
3	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.43	Yes	53	6.8	53	6.2	297	6	2065057
4	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.27	Yes	24	8.7	53	6.1	56	2	6017859
5	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.28	Yes	23	8.7	52	6.1	52	2	24634964
6	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.29	-	53	6.5	51	6.1	102	2	6561367
7	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.50	-	55	5.2	53	6.3	182	5	1750356
8	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.33	-	54	5.8	53	6.1	246	4	523330
9	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.27 *	-	55	5.6	36	6.4	75	2	108951200
10	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.46 *	-	53	5.5	52	6.1	59	2	6983900
11	C&E	ribulose-1,5-bisphosphate carboxylase large subunit	0.84 *	Yes	57	4.6	51	6.2	52	2	18140565
12	C&E	rubisco activase	0.36	-	46	5.5	40	6.7	95	2	62733168
13	C&E	rubisco activase	0.56	-	46	5.5	51	8.6	108	2	10720247
14	C&E	glyceraldehyde-3-phosphate dehydrogenase	0.26	-	41	7.9	37	8.3	72	2	120666
15	C&E	glyceraldehyde-3-phosphate dehydrogenase	inc. *	-	42	5.5	45	7.6	82	3	435103
16	C&E	glyceraldehyde-3-phosphate dehydrogenase	inc.	-	42	5.6	34	6.0	55	2	18072795
17	C&E	phosphoglycerate kinase	0.27	-	45	6.5	24	5.1	193	3	1022803
18	PS&LH	oxygen-evolving enhancer protein	0.36	Yes	24	6.1	28	6.9	103	2	15222166
19	PS&LH	predicted manganese-stabilizing protein	0.30	-	33	5.6	39	5.4	172	5	125526093
20	PS&LH	chlorophyll a/b binding, type 1	0.17	-	22	5.5	27	7.7	54	1	20788
21	P&AA	protein disulfide isomerase	0.24	Yes	60	5.1	56	5.0	57	2	77999357
11	P&AA	calreticulin	0.84 *	Yes	57	4.6	50	4.6	47	2	9858814
22	P&AA	probable HSP 70	0.13	-	71	5.2	74	5.4	112	3	147860809
23	P&AA	HSP 70	0.18	-	68	5.5	71	5.1	288	5	123620
24	P&AA	HSP 70-BiP	0.72	-	70	5.5	73	5.1	235	4	1346172
25	P&AA	20S proteasome subunit	inc.	-	28	6.7	27	5.9	82	2	12229948
15	P&AA	PO ribosomal protein	inc. *	-	42	5.5	35	5.4	57	2	77552928
26	P&AA	aminomethyltransferase	0.21	-	45	8.2	45	8.9	78	1	3334196
27	P&AA	alanine aminotransferase	0.40	-	52	7.4	54	6.9	117	2	13430566
28	ROI	peroxidase	inc.	Yes	41	8.8	39	8.1	55	1	1389835
29	ROI	peroxiredoxin, 2-Cys	0.23	-	24	5.1	22	4.9	103	2	47027073
30	ROI	peroxiredoxin, PRX5-like	3.01	-	17	5.6	18	5.6	64	1	66360171
31	ROI	chloroplast drought-induced stress protein of 32kDa	inc.	-	32	6.5	34	8.1	96	2	2582822
32	ROI	superoxide dismutase	inc.	-	22	7.5	25	8.8	80	1	3108345
33	2°met	geranylgeranyl reductase	inc.	-	49	8.8	53	9.0	52	2	2828267
34	ATP	AAA-ATPase	0.59	-	101	5.6	90	5.2	105	3	1705678
35	ATP	ATP synthase beta subunit	0.38	-	55	5.5	51	5.2	157	2	66276267
9	ATP	ATP synthase beta subunit	0.27 *	-	55	5.6	60	6.1	292	5	3893824
36	ATP	ATP synthase beta subunit	0.60	-	55	5.5	51	5.2	78	3	66276267
37	ATP	vacuolar V-H+ATPase subunit E	0.83	-	31	7.4	27	6.9	100	2	3334411
15	ATP	adenosine kinase	inc. *	-	42	5.5	38	5.1	51	1	51949800
10	C&S	tubulin alpha	0.46 *	-	53	5.5	50	4.9	232	6	464849
38	C&S	actin	0.36	-	47	5.7	42	5.2	93	2	18532

a) Functional category: Primary carbon and energy metabolism (C&E); Protein and amino acid metabolism (P&AA); Reactive oxygen interaction (ROI); Photosynthesis and light harvest (PS&LH); Secondary metabolism (2° met); ATPases (ATP); Cytoskeleton and secretion (C&S).

b) Ratios compare spot volumes generated from Pro-Q Diamond stained and SYPRO Ruby stained gel images. Inconclusive

c) Staining by Pro-Q Emerald glycoprotein stain: Spot was stained (Yes); Spot was not stained (-).

d) Empirical molecular weight (kDa) and isoelectric point (pI) refers to the physical placement within gels of spots analysed.

e) Theoretical molecular weight (kDa) and isoelectric point (pI) refers to info for the Genbank ID match.

f) MOWSE scores are $-10*\text{LOG}_{10}(P)$, where P is the absolute probability the match is a random event; all listed matches are significant ($p<0.05$) with respect to the NCBInr database population searched.

g) Peptide count refers to the number of peptides per spot that match the Genbank ID shown

*) *Spots with multiple proteins were identified. The phosphorylation ratios shown may be due to more than on protein.*

Proteins are numbered by their appearance in the text and grouped by functional category in the table.

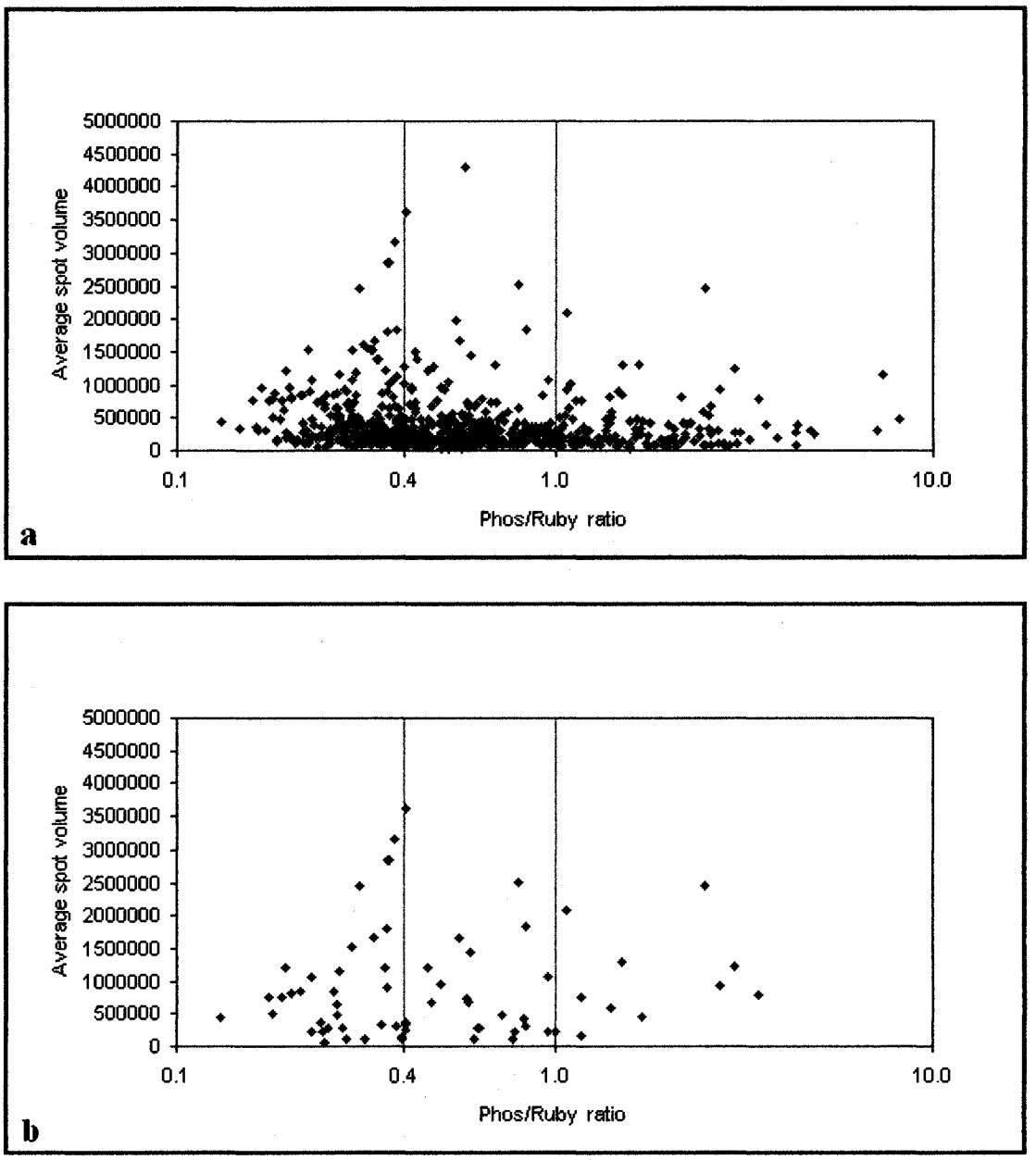


Figure 3-2 Distribution of Pro-Q Diamond to SYPRO Ruby ratio with relation to average spot volume. (a) Distribution of all protein spots in gel. (b) Distribution of all spots refered to in thesis. Spots located between 0.1 and 0.4 are considered minimally phosphorylated; between 0.4 and 1.0 are moderately phosphorylated; above a ratio of 1.0 they are considered clearly or highly phosphorylated. 0.4 and 1.0 are arbitrary cut-off points determined from the distribution in (a). Average spot volume refers to an average between the SYPRO Ruby and Pro-Q Diamond spot volumes for each spot.

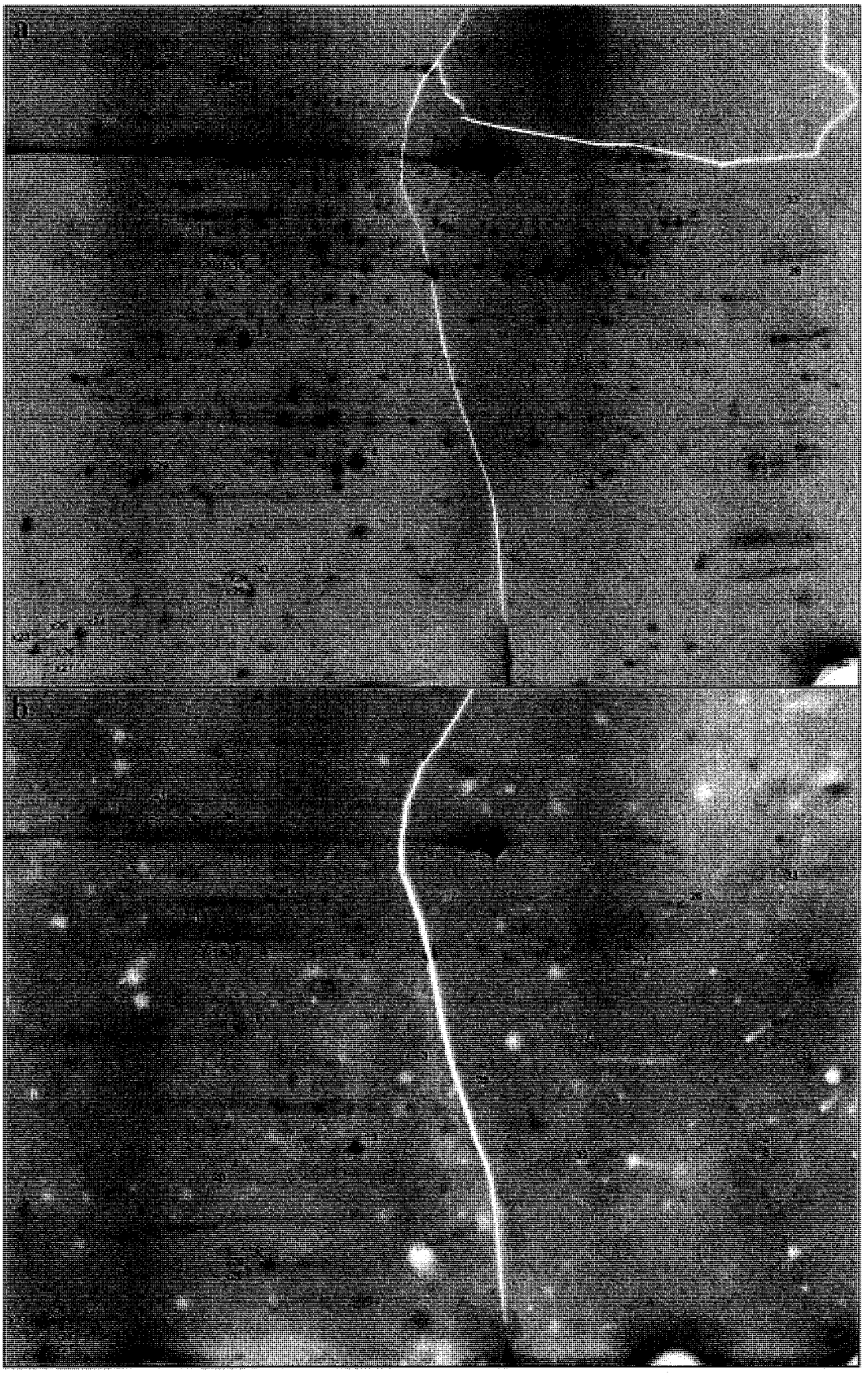

a
b

Figure 3-3 (previous page) Comparison of SYPRO Ruby image (a) and Pro-Q Diamond (b). Numbered spots have database matches and 'x'+numbered spots do not. Orientation of images is acidic (left) to basic (right).

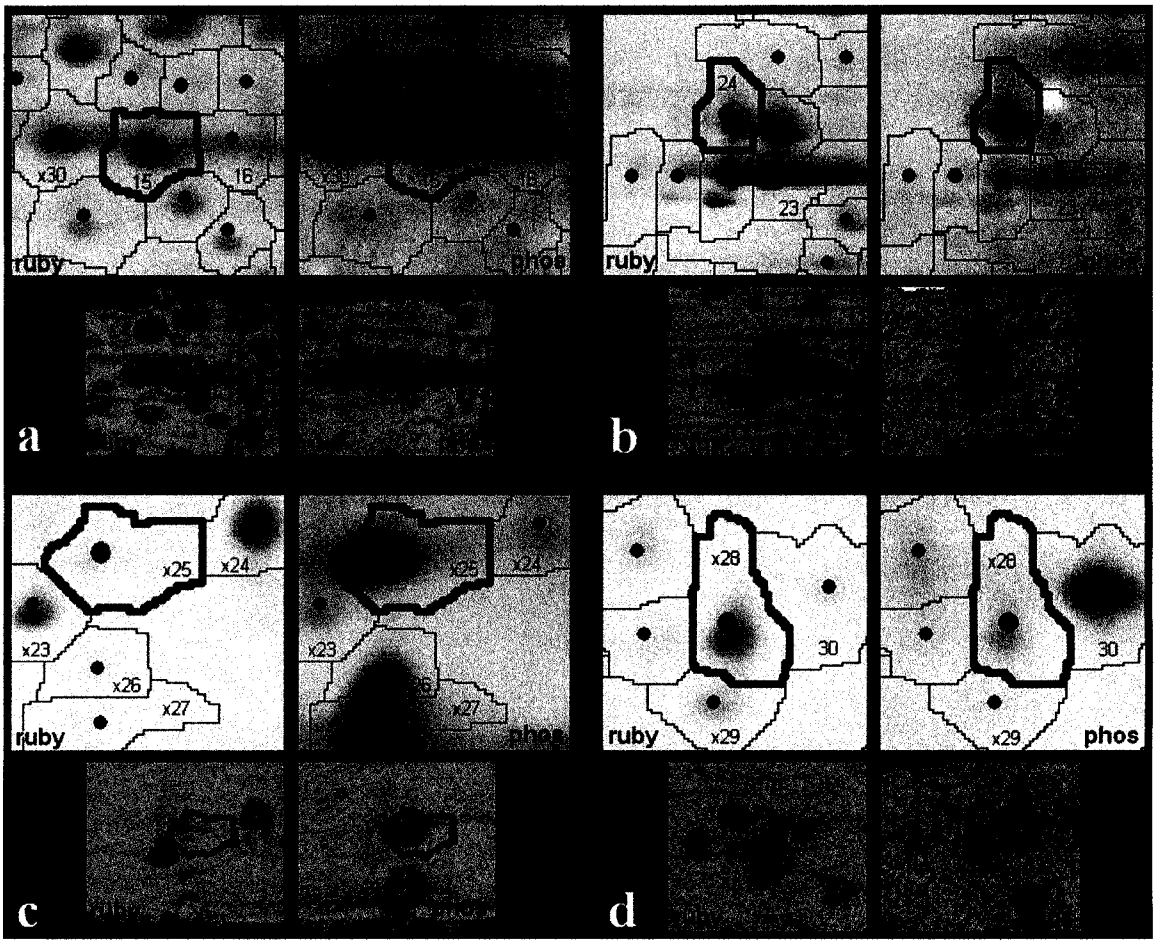


Figure 3-4 Close-up of phosphorylation results for select protein spots. (a) shows a highly phosphorylated protein just to the left of spots 15 and 16. (b) the difference between phosphorylation status of HSP proteins, spot 23 and 24. (c) and (d) show obvious phosphorylation differences between protein spots. Spots x23 through x29 have no database match. Each grouping of four pictures is comprised of a 2D image from the SYPRO Ruby stained gel and Pro-Q Diamond stained gel along with the respective 3D version directly below. 2D and 3D images were created in Decyder 6.5.

3.6 Bibliography

Aran, M., Caporaletti, D., Senn, A.M., de Inon, M.T.T., Girotti, M.R., Llera, A.S., and Wolosiuk, R.A. (2008). ATP-dependent modulation and autophosphorylation of rapeseed 2-Cys peroxiredoxin. FEBS Journal **275,** 1450-1463.

Archambaud, C., Nahori, M.-A., Pizarro-Cerda, J., Cossart, P., and Dussurget, O. (2006). Control of *Listeria* superoxide dismutase by phosphorylation. Journal of Biological Chemistry **281,** 31812-31822.

Bayer, E.M., Bottrill, A.R., Walshaw, J., Vigouroux, M., Naldrett, M.J., Thomas, C.L., and Maule, A.J. (2006). Arabidopsis cell wall proteome defined using multidimensional protein identification technology. Proteomics **6,** 301-311.

Bonner, J.J., Carlson, T., Fackenthal, D.L., Paddock, D., Storey, K., and Lea, K. (2000). Complex regulation of the yeast heat shock transcription factor. Molecular Biology of the Cell **11,** 1739-1751.

Bruce-Carver, M.R., Margosiak, S.A., Guzman, E., and Kuehn, G.D. (1990). In-vivo posttranslational covalent modification of ribulose bisphosphate carboxylase oxygenase by transglutaminase in plants. FASEB Journal **4,** A2129.

Cao, T., Srivastava, S., Rahman, M.H., Kav, N.N.V., Hotte, N., Deyholos, M.K., and Strelkov, S.E. (2008). Proteome-level changes in the roots of *Brassica napus* as a result of Plasmodiophora brassicae infection. Plant Science **174,** 97-115.

Clancy, M.J., and Whelan, W.J. (1959). Selective periodate oxidation of reducing-end groups in oligosaccharides. Chemistry and Industry, 673-675.

Clancy, M.J., and Whelan, W.J. (1967). Enzymatic polymerization of monosaccharides. Archives of Biochemistry and Biophysics **118,** 730-735.

Cohu, C.M., and Pilon, M. (2007). Regulation of superoxide dismutase expression by copper availability. Physiologia Plantarum **129,** 747-755.

Coppolino, M.G., and Dedhar, S. (1998). Calreticulin. International Journal of Biochemistry and Cell Biology **30,** 553-558.

Csar, X.F., Wilson, N.J., Strike, P., Sparrow, L., McMahon, K.A., Ward, A.C., and Hamilton, J.A. (2001). Copper/zinc superoxide dismutase is phosphorylated and modulated specifically by granulocyte-colony stimulating factor in myeloid cells. Proteomics **1,** 435-443.

Dietz, K.-J. (2007). The dual function of plant peroxiredoxins in antioxidant defence and redox signaling. In Subcellular biochemistry, pp. 267-294.

Donella-Deana, A., James, P., Staudenmann, W., Cesaro, L., Marin, O., Brunati, A.M., Ruzzene, M., and Pinna, L.A. (1996). Isolation from spleen of a 57-kDa protein substrate of the tyrosine kinase Lyn: Identification as a protein related to protein disulfide-isomerase and localisation of the phosphorylation sites. European Journal of Biochemistry **235,** 18-25.

Freedman, R.B., Hirst, T.R., and Tuite, M.F. (1994). Protein disulphide isomerase: building bridges in protein folding. Trends in Biochemical Sciences **19,** 331-336.

Gutteridge, S., and Gatenby, A.A. (1995). Rubisco synthesis, assembly, mechanism and regulation. Plant Cell **7,** 809-819.

Harrison, F.W., Weber, P., and Hof, L. (1976). The fluorescent demonstration of tissue aldehydes with dansyl hydrazine. Histochemistry **49,** 349-351.

Heldt, H.-W., and Heldt, F. (1997). Plant biochemistry & molecular biology. (Oxford: Oxford University Press).

Helenius, A., Trombetta , E.S., Hebert , D.N., and Simons, J.F. (1997). Calnexin, calreticulin and the folding of glycoproteins. Trends in Cell Biology **7,** 193-200.

Hotte, N.S., and Deyholos, M.K. (2008). A flax fibre proteome: identification of proteins enriched in bast fibres. BMC Plant Biology **8,** 52.

Houtz, R.L., Magnani, R., Nayak, N.R., and Dirk, L.M.A. (2008). Co- and post-translational modifications in Rubisco: unanswered questions. Journal of Experimental Botany **59,** 1635-1645.

Kargul, J., and Barber, J. (2008). Photosynthetic acclimation: Structural reorganisation of light harvesting antenna - role of redox-dependent phosphorylation of major and minor chlorophyll a/b binding proteins. FEBS Journal **275,** 1056-1068.

Keller, Y., Bouvier, F., d'Harlingue, A., and Camara, B. (1998). Metabolic compartmentation of plastid prenyllipid biosynthesis. European Journal of Biochemistry **251,** 413-417.

Kim, U.J., Kuga, S., Wada, M., Okano, T., and Kondo, T. (2000). Periodate oxidation of crystalline cellulose. Biomacromolecules **1,** 488-492.

Kuehn, G.D., Sotelo, M., Morales, T., Bruce-Carver, M.R., Guzman, E., and Margosiak, S.A. (1991). Purification and properties of transglutaminase from *Medicago sativa* L. Alfalfa. FASEB Journal **5,** A1510.

Liepman, A.H., and Olsen, L.J. (2003). Alanine aminotransferase homologs catalyze the glutamate:glyoxylate aminotransferase reaction in peroxizomes of arabidopsis. Plant Physiology **131,** 215-227.

Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Krylov, D., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Thanki, N., Yamashita, R.A., Yin, J.J., Zhang, D., and Bryant, S.H. (2007). CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Research **35,** D237-240.

Mueller, C.F.H., Wassmann, K., Berger, A., Holz, S., Wassmann, S., and Nickenig, G. (2008). Differential phosphorylation of calreticulin affects AT1 receptor mRNA stability in VSMC. Biochemical and Biophysical Research Communications **370,** 669-674.

O'Neill, M.A., and York, W.S. (2004). The composition and structure of plant primary cell walls. In The plant cell wall, J.K.C. Rose, ed (New York: Wiley), pp. 3-4.

Portis, A.R., Jr., and Salvucci, M.E. (2002). The discovery of rubisco activase: Yet another story of serendipity. Photosynthesis Research **73,** 257-264.

Roberts, G. (1977). Histochemical detection of sialic acid residues using periodate oxidation. Histochemical Journal **9,** 97-102.

Schulenberg, B., Aggeler, R., Beechem, J.M., Capaldi, R.A., and Patton, W.F. (2003). Analysis of steady-state protein phosphorylation in mitochondria using a novel fluoresent phosphosensor dye. Journal of Biological Chemistry **278,** 27251-27255.

Serafini-Fracassini, D., and Del Duca, S. (2008). Transglutaminases: Widespread cross-linking enzymes in plants. Annals of Botany **102,** 145-152.

Séveno, S., Bardor, M., Paccalet, T., Gomord, V., Lerouge, P., and Faye, L. (2004). Glycoprotein sialylation in plants? Nature Biotechnology **22,** 1351–1352.

Shah, M.M., Fujiyama, K., Flynn, C.R., and Joshi, L. (2003). Sialylated endogenous glycoconjugates in plant cells. Nature Biotechnology **21,** 1470-1471.

Sharma, N., Hotte, N., Rahman, M.H., Mohammadi, M., Deyholos, M.K., and Kav, N.N.V. (2008). Towards identifying *Brassica* proteins involved in mediating resistance to *Leptosphaeria maculans*: A proteomics-based approach. Proteomics **8,** 3516-3535.

Spetea, C., Hundal, T., Lundin, B., Heddad, M., Adamska, I. and Andersson, B. (2004). Multiple evidence for nucleotide metabolisim in the chloroplast thykaloid lumen. Proceedings of the National Academy of Science of the United States of America **101,** 1409-1414.

Sung, D.Y., Vierling, E., and Guy, C.L. (2001). Comprehensive expression profile analysis of the arabidopsis hsp70 gene family. Plant Physiology **126,** 789-800.

SWISS-2DPAGE. (2004). Two-dimensional polyacrylamide gel electrophoresis database from the Geneva University Hospital. Proteomics **4,** 2352-2356.

Takashima, S., Abe, T., Yoshida, S., Kawahigashi, H., Saito, T., Tsuji, S., and Tsujimoto, M. (2006). Analysis of sialyltransferase-like proteins from *Oryza sativa*. Journal of Biochemistry **139,** 279-287.

Wan, L., Gijzen, M., and Van Huystee, R.B. (1994). Heterogeneous glycosylation of cationic peanut peroxidase. Biochemistry and Cell Biology **72,** 411–417

Watson, I.N., Watson, L.M., Murray, T.A., Lige, B., van Huystee, R.B., and McManus, M.T. (1998). Identification of two further cationic peroxidase isoenzymes secreted by peanut cells in suspension culture. Plant Physiology and Biochemistry **36,** 591-599.

White, D.A., Miyada, D.S., and Nakamura, R.M. (1974). Characterization of the periodate oxidation of glycerol and related compounds, and application toward determination of serum triglyceride concentrations. Clinical Chemistry **20,** 645-648.

Wilkinson, B., and Gilbert, H.F. (2004). Protein disulfide isomerase. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1699,** 35-44.

Williams, C.M., Zhang, G., Michalak, M., and Cass, D.D. (1997). Calcium-induced protein phosphorylation and changes in levels of calmodulin and calreticulin in maize sperm cells. Sexual Plant Reproduction **10,** 83-88.

Yosef, I., Irihimovitch, V., Knopf, J.A., Cohen, I., Orr-Dahan, I., Nahum, E., Keasar, C., and Shapira, M. (2004). RNA binding activity of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit from *Chlamydomonas reinhardtii*. Journal of Biological Chemistry **279,** 10148-10156.

Chapter 4: Conclusions

4.1 Conclusions

This proteomic study of excised flax fibres and stem peels has yielded interesting details of fibre development; all of which are good foundations for further study. The identification of a β-galactosidase highly upregulated in fibre cells is interesting as it may have a role in galactan turnover during cell wall biosynthesis. Additionally, rhamnose synthase enzymes were also noted to be up regulated in fibre cells and they may be involved in the generation of monosaccharides destined for the cell wall. Galactan turnover in fibre cells undergoing cell wall development has been reported previously; galactans were localized to golgi bodies, golgi vesicles and vacuoles in addition to the cell wall. In the current study, proteins related to vesicle movement and secretory pathways were found to be upregulated in fibre cells. Further study should involve determination of the localization of this β-galactosidase; as it is equally plausible that it is secreted or alternatively, located within the cells in the golgi system or vacuoles. Experimental evidence of the β-galactosidase substrate(s) would also be important.

Potential carbon sources (other than recycled galactans) were also noted; proteins related to sucrose utilization and starch degradation were both up regulated in fibre cells. Interestingly, an increase in a K^+ channel was also noticed; a similar K^+ channel has previously been reported as involved in ion balance associated with sucrose uptake. I would be interested to know which is the main carbon source feeding the large cell wall carbon requirements.

Energy production would be another avenue requiring fixed carbon. Many TCA cycle proteins were found to be upregulated in the fibre sample. Interestingly, proteins involved in processes that require carbon were found to be upregulated in fibre cells but a similar upregulation of fixed-carbon producing proteins was not seen. In fact, many photosynthesis and light-harvesting proteins were found to be down regulated in fibre cells. This suggests that developing fibres are a carbon sink and require external carbon sources, such as sucrose.

I identified two calcium-binding proteins, annexin and calreticulin, both of which

are upregulated in fibre cells. An annexin has been shown to co-localize with cellulose synthase and is presumed to play a regulatory role therein but no clear experimental evidence has clearly demonstrated such a role (Hofmann et al., 2003; Yao et al., 2006). Calreticulin was found in a glycosylated and phosphorylated state; it may be involved in the editing of the glycosylated proteins, such as arabinogalactan proteins previously found in bast fibre cell walls (Andeme-Onzighi et al., 2000).

Two peroxidases were identified and matched to *L. usitatissimum* proteins. One was upregulated in fibre and the other was clearly glycosylated. Two peroxiredoxin proteins were also noted, with one being highly phosphorylated in stem peal tissue. It is difficult to attach clear roles to these proteins without further experimental evidence, as they have been associated with many different roles.

This study did not adequately sample the membrane-bound fraction of the proteome but I did suggest ways to improve the appearance of such proteins in a similar experiment. Other reports suggest similar difficulties with protein extraction from cotton fibres (Yao et al., 2006); their approach, although intriguing, did not show evidence of extracted membrane-bound proteins either.

In addition to sample preparation, I explored ways to increase the usefulness of both Pro-Q Diamond phosphoprotein stain and Pro-Q Emerald glycoprotein stains. Pro-Q Diamond requires the use of a quantitative, 2D-specific marker to help establish in-gel thresholds of phosphorylation. Pro-Q Emerald may be staining with less sensitivity in plant-derived proteins than in mammalian-derived samples.

This proteome characterization of flax bast fibres provides important cell-specific information about fibre development. It provides proteomics-based information on basic metabolism and cell wall biosynthesis within the flax bast fibre and the surrounding stem tissues.

4.2 Bibliography

Andeme-Onzighi, C., Girault, R., His, I., Morvan, C., and Driouich, A. (2000). Immunocytochemical characterization of early-developing flax fiber cell walls. Protoplasma **213,** 235-245.

Hofmann, A., Delmer, D.P., and Wlodawer, A. (2003). The crystal structure of annexin Gh1 from *Gossypium hirsutum* reveals an unusual S-3 cluster - Implications for cellulose synthase complex formation and oxidative stress response. European Journal of Biochemistry **270,** 2557-2564.

Yao, Y., Yang, Y.W., and Liu, J.Y. (2006). An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-DE. Electrophoresis **27,** 4559-4569.

Appendix 1: At4g01870 promoter deletion experiment

A1.1 Introduction

Previous gene expression analyses have identified a novel gene arabidopsis gene that is highly induced (17-fold) at the transcript level following exposure to 2,4,6-trinitrotoluene (TNT) (Ekman et al., 2003). This gene has no known function, and is identified only by its AGI (Arabidopsis Genome Initiative) code, At4g01870. As part of efforts to understand plant responses to environmental toxins, I studied the regulation of At4g01870 transcription using reporter gene fusions.

At4g01870 is located on chromosome 4, in a similar orientation to both the surrounding At4g01860 and At4g01880 genes (Arabidopsis Genome, 2000). The predicted translation start site is 378 bp downstream of the predicted stop translation site of At4g01880 and the predicted un-translated regions (UTRs) of both genes are separated by only 252 bp (figure A1-1). At4g01870 is noted as having only one exon of 2071 bp and no introns. The coding region is 1959 bp with a predicted 15 bp 5'UTR and a predicted 97 bp 3'UTR; it codes for a predicted protein of 652 amino acids, molecular weight of 72.8 kDa and isoelectric point of 5.85 (www.arabidopsis.org).

The At4g01870 coding sequence shows homology to the C-terminal domain of TolB, a protein that has been structurally well characterized in *Escherichia coli* (Abergel et al., 1999). The C-terminal TolB domain is considered a β-propeller and belongs to the larger WD domain family; in which, the proteins often have Trp-Arg repeats (Ponting and Pallen, 1999). The β-propeller domain in TolB is made of many, larger, repeating units that subsequently fold into a six-pronged propeller shape with a central tunnel (Abergel et al., 1999; Carr et al., 2000). In β-propellers, many use the tunnel or bowl-shaped tunnel entrance to coordinate a ligand or carry out a catalytic function. For example, in oligopeptidase, the tunnel only allows access to the catalytic site for peptides up to 30 amino acids long, thereby positioning the appropriate ligand. The β-propellor region appears to be a conserved and useful structure that can be utilised for many different purposes, as evidenced by the diversity of substrates and ligands that it interacts with (Carr et al., 2000).

The At14g01870 putative protein is noted as containing the TolB C-terminal domain (InterPro:IPR011042) and WD domain (InterPro:IPR011659); the closest homology found to a bacterial WD/ β-propeller domain is to *Solibacter usitatus* with a 17 % amino acid identity over the total protein (figure A1-2b and appendix 2: figure A2-1). Although there are obvious similarities and conserved regions with bacterial TolB proteins, it seems to be a very distant relationship. I have found much closer homologies within the plant kingdom.

Within the *A. thaliana* genome, At4g01870 shows the most sequence homology to At1g21670 and At1g21680, as found using both blastn and blastp searches within *A. thaliana* restricted databases (Altschul et al., 1997). At the nucleotide level, At1g21670 and At1g21680 coding sequences show 47 % and 48 % identical nucleotides with At4g01870, respectively (figure A1-3 and appendix 2: figure A2-2), which was calculated from a multiple sequence alignment using DIALIGN (Morgenstern, 2004). Both found genes code for putative proteins of unknown function. Notably, At1g21670 and At1g21680 share 67 % identities with each other and they are also physical neighbours on chromosome 1; suggesting that both of the chromosome 1 genes are related more closely to each other than to At1g01870 and they may have been involved in a past tandem duplication event.

Although At1g21670 and At1g21680 are the best same species homologs, there are potential orthologs from other plants. To find the following genes, I used both blastp and blastn searches of the NCBI non-redundant database with all organisms included; some sequences appeared in both types of searches and others appeared only in one type. The overall best hit was to a *Vitis vinifera* cv. 'Pinot noir' gene, CAN71280, with 68 % sequence identity and 74 % coding region coverage (blastp, score: 888 bits, expect: 0.0); this gene also codes for an unknown protein and was sequenced as part of a *V. vinifera* genome sequencing effort (Velasco et al., 2005; Velasco, 2006). The next best, and close second, homology was to a sequenced mRNA from *Solanum lycopersicum,* BT013533, at about 65 % identities and 70 % coverage (Blastn, score: 233, expect: 2e-57); again, this codes for a protein of unknown function (Kirkness et al., 2004). My searches also found *Oryza sativa* genes, most notably, Os07g0638100 that At4g01870 is said to share homology with, it shares 55 % identical amino acids and 12 % similar amino acids with

At4g01870 (Ohyanagi et al., 2006). Lastly, an interesting homology is to a 'resistance gene marker' in *Gossypium hirsutum,* AY627706 (Hinchliffe et al., 2005), which shows 71 % nucleotide identity with 13 % of the At4g01870 coding region and could be the only database 'lead' to protein function. See figures A1-2 a and b for a graphic depiction of the amino acid 'identities' for each prospect, as compared to At4g01870; also, see appendix 2: figure A2-3 for the multiple protein-sequence alignments that the graphs are based on and the accession information for all sequences.

A promoter deletion experiment was designed to establish the native expression patterns of At4g01870. At4g01870 codes for a protein of unknown function. This appendix describes the procedure taken to create transgenic *Arabidopsis thaliana* from the original point of promoter fragment isolation. I will also show preliminary GUS-staining results.

A1.2 Methods and materials

A1.2.1 Plant material

Wildtype *Arabidopsis thaliana* 'Columbia' ecotype was used for all transformations and subsequent experiments in this chapter. For seed sterilization all seeds were treated for 10 minutes in bleach diluted with 0.1 % SDS solution (1 part: 9 parts), followed by 10 rinses with sterile water. After sterilization, seeds were re-suspended in 0.05% Phytablend, and distributed onto solid growth media. Seeds were sown on 15mm x 100 mm petri plates filled with 20 ml of media or 15 mm x 150 mm plates filled with 50 ml of media and subsequently stratified at 4 °C for 2 days. After stratification seeds were germinated and grown under 16hr day and 8hr night (long-day) in standard growth chamber conditions at 22 °C. All arabidopsis growth media mixtures contained ½ concentration Murashige and Skoog (1962) basal media (½ MS) and were solidified with 3 g/L Phytablend (Caisson Laboratories). Some media were supplemented with a final concentration of 0.5 % sucrose (Suc), 25 μg/ml Hygromycin b (Hyg 25) and/or 150 μg/ml Timentin (Tim 150). For arabidopsis growth to maturity, seedlings were transferred from plates 10 cm pots, at 4 seedlings per pot, and grown under the above conditions until post seed-set senescence.

A1.2.2 Primers and promoter fragment isolation

Arabidopsis genomic DNA (isolated with DNeasy plant mini kit from Qiagen) was used as a template for the isolation of four promoter fragments from the 5' upstream region of At4g01870. The four forward primers used to create the progressive 5' deletions each have an introduced XbaI site for subcloning and their positions and sequences are:

-381-F, 5'- TGGTCTAGATGTATTCTTGTTGGT___-3';

-267-F, 5'-TGGTCTAGAATTCTCTGTCTTTGGTCA___-3';

-167-F, 5'-TGGTCTAGAAATTAACGATAAAAGAT___-3';

-100-F, 5'-TGGTCTAGAGTTCACACTTCACACTCA___-3'.

One reverse primer, with an introduced Nco I site, was used in combination with the above four primers to amplify the four promoter fragments (figure A1-4 and A1-5 a); its relative-to-start-site position and sequence are:

+3-R, 5'-GTCACCATGGAGTTTTTGATGA-3'. All above primers were designed by Dr. Deyholos and were synthesized by Integrated DNA Technologies (IDT).

A1.2.3 Blunt cloning into Topo vectors

The blunt fragments produced by PCR were ligated into pCR-Blunt II-Topo vectors (Invitrogen) and subsequently used to transform One Shot Competent Cells (*E. coli*) using the protocols supplied with the kit. Resulting colonies were selected based on their growth overnight at 37 °C on standard Luria-Bertani (LB) agar plates, supplemented with kanamycin (50 μg/ml; LB Kan 50). Colonies were selected and further cultured in 5 ml liquid LB kan 50 broth, overnight at 180 rpm, and the resulting cultures were used for PCR (to check for inserts) and plasmid isolation. Plasmids were isolated using the QIAprep spin miniprep kit (Qiagen) as instructed by the supplier. Isolated plasmids were digested using Nco I and Xba I (New England Biolabs) to release the inserts and to produce overhangs for directional cloning. The appropriately sized DNA fragments were excised from an agarose gel (1.2 % low-melt agar gel and 1x TAE buffer; figure A1-5 b).

Excised gel pieces were ligated into pre-digested and gel-purified pCambia 1303 vectors; vectors were double-digested with Nco I and Xba I and gel purified prior to

ligation to remove the CaMv 35s promoter insert. Both pCambia and isolated fragment gel slices were prepared for ligation using QIAquick get extraction kit (Qiagen). Ligation was carried out with a sticky-end concentration ratio of 1 (vector) to 3 (fragment) using T4 ligase and the accompanying buffer and protocol (Fermentas). Ligation was followed by adding 1 µl of the ligation solution to competent *E.coli* cells in a standard freeze-thaw procedure. Potentially transformed cells were streaked onto LB kan 50 plates, grown overnight at 37 °C, and subsequently selected based on their growth. Chosen colonies were than used to inoculate 5 ml LB kan 50 liquid cultures and used for PCR insert validation. Overnight liquid cultures were pelleted at 8000 xg and the resultant pellet was used to isolate the transformed pCambia vectors; the plasmids were isolated from cellular debris and genomic DNA using GeneJET Plasmid Miniprep kit (Fermentas). Some purified vector was used for sequencing and in a triple digest with Hindi III, BstE II, Nco I to show proper insertion, and insert directionality. Otherwise, purified plasmids were used to transform competent *Agrobacterium tumefaciens* (strain GV3101) using a standard freeze-thaw transformation protocol. Potentially transformed *A. tumefaciens* were selected for growth on LB Gent 100, Rif 34, Kan 50 plates after 2 days at 30 °C. Two chosen colonies from each of the four fragment insert lines were chosen to inoculate 5 ml LB cultures for subsequent PCR validation, glycerol stock creation and arabidopsis transformation.

Each arabidopsis line (made up of eight plants) was transformed four times (once a week for 4 weeks) with one of the eight transformed *A. tumefaciens* cultures (two cultures for each original DNA fragment; figure A1-5 c). 2 ml of each 5 ml *A. tumefaciens* culture were centrifuged at 8000 xg to pellet the cells. These pelleted cells were resuspended in infiltration liquid (½ MS + 0.5 % suc + 0.5 µl/ml Silwett L-77) and 5 µl of this was used to inoculate each closed-bud inflorescence on each plant. The *A. tumefaciens* used for the weekly repeat arabidopsis transformations were grown in 5ml cultures inoculated from the original glycerol stock made on the first week. After the final inoculation these T0 arabidopsis plants were grown up to seed and the T1 generation seeds were harvested from them.

Approximately 3000 T1 generation seeds for each line were surface sterilized. All sterilized seeds from each line were plated on large (15 cm) selection plates (½ MS

Hyg 25 + Tim 150). Plates were sealed with medical tape and stratified for two days at 4 °C in the dark, after which, they were placed in standard growth chamber conditions for 10-14 days. Plants were selected as transformed from these Hyg plates, if they had a healthy root system in addition to nice green true leaves (non-transformed were yellowish, had only cotyledons and usually had no roots). Approximately 30-50 healthy, non-clustered, transformants were noted on each plate. Healthy, potentially transformed plants were transplanted to soil and grown up to harvest the T2 seed from each T1 individual. T2 seed was treated similarly to the T1 seed, except that 50 seeds from 6 individuals from each of 8 lines were sterilized and plated on selection plates. Four healthy T2 transformants from each 6 individuals were chosen to subsequently plant to soil and harvest T3 seed. Approximately 50 T3 seeds from each T2 individual were sterilized and plated on ½ MS + 0.5 % suc + Hyg 25 selection plates. To check segregation ratios of the lines, after 2 weeks growth these seedlings were scored for transformation and the frequencies tallied.

A1.2.4 Experimental treatments and GUS assays with transformants

Two main experiments utilizing qualitative GUS staining were completed: (1) seedlings grown on non-selective media and GUS stained after 8 DAI, 16 DAI and 20 DAI (untreated) and (2) seedlings grown on non-selective media then treated with either the 'treatment' 1,3-dinitrobenzene (13DNB; 15 mg/L) + dimethylsulfoxide (DMSO; 0.1%) or the 'control' DMSO (0.1%) on 19 DAI and GUS stained on 20 DAI. All GUS staining assays were considered preliminary screens of promoter expression.

Surface sterilized seeds were placed on non-selective ½ MS agar + 0.5 % suc. After stratification, seeds were grown under 16 h day 8 h light cycle at 25 °C until 8 DAI, 16 DAI or 20 DAI. Plated seedlings were treated with applications of DMSO (control) or 13DNB + DMSO (treated). On 19 DAI, 4 ml of the appropriate treatment or control solution was added to the solid media in each plate (final concentrations were based on total media used, ie. 20 ml solid agar + 4 ml added treatment liquid). Plants were kept under growth conditions for 24 h, then on 20 DAI, plants were removed from the treatment, control and untreated plates for GUS staining. Previous tests (data not shown) indicate that this treatment with 13DNB + DMSO onto solid media, causes obvious and

visible detrimental effects to the seedlings after 48 h. Further, the DMSO-only treatment showed no such visible effect.

The arabidopsis lines used in these GUS assays originate from transformations with 5 different At4g01870 promoter region constructs, transformations with native pCambia 1303 plasmid and untransformed wild type arabidopsis. Four of the promoter constructs were put together by me and are described above. The remaining AT4g01870 promoter construct and arabidopsis transformants were previously created by Miao Min Min and Aaron Donahue. This construct is similar in all respects to my constructs except the promoter fragment is longer; it covers from –550 to +3 of the native At4g01780 gene promoter (original primers used to create this promoter fragment were:
+3 miao R, 5'- TCCACCATGGAGTTTTTGATGAATTGTTGGT -3';
-539 miao F, 5'- GACATCTAGAGCTGTTTTAGGATTCAAATGC -3'). The original pCambia 1303 vector, which has CaMv 35s promoter driving the GUS gene, was used to make arabidopsis transformants; these plants were created by Yuanqing Jiang and are T4 generation seedlings. These original pCambia vector transformed arabidopsis were included in the above experiments as controls and for a qualitative comparison of GUS staining.

A1.2.5 GUS staining

A standard GUS staining procedure was followed and the GUS staining solution (50 mM NaPO4, pH 7.0, 0.5 mM K3Fe(CN)6, 0.5 mM K4Fe(CN)6, 1 mM X-gluc) was mixed fresh for each staining from stock solutions. All seedlings from one transgenic line and from one treatment were stained together. After removal from the agar plate, seedlings were immediately placed in 5-10 ml of ice-cold 90 % acetone for 10 minutes. To stain, they were rinsed with 5 ml of 50 mM NaPO4, pH 7.0, then immersed in 2 ml of staining solution and left at 37 °C for 3, 4 or 16 h. Some lines were removed from the staining solution early (3-4 h after staining start) to prevent over-staining but most were left for 16 hours. To stop GUS staining reaction, remove chlorophyll and preserve the plant tissue, stained seedlings were transferred to 70% ethanol for 24 hours then transferred to 95% ethanol indefinitely.

A1.3 Results

A1.3.1 Cloning and proof of inserts

The four promoter fragments were isolated from arabidopsis by PCR. The original isolations were 111 bp, 181 bp, 281 bp, and 392 bp (figure A1-5 a); fragments will be referred to as A, B, C and D, respectively. The resulting PCR fragments were used for blunt-end cloning into Topo vectors. This was followed by insert release from the vector with Nco I and Xba I to enable directional cloning (digestion pictured in figure A1-5 b). Inserts replaced the CaMv 35s promoter in pCambia 1303.

After plasmid amplification in *E. coli*, the plasmids were transferred to *A. tumefaciens* and sequenced. For sequencing, primers matched to a multiple cloning site upstream of the insert and to the GUS gene were used. Sequencing results indicate that all fragments were inserted into the plasmids, one time and in the correct orientation. Additionally, this sequencing PCR amplification of inserts, from plasmids in *A. tumefaciens,* resulted in the correct sized bands for each fragment (figure A1-5c).

A1.3.2 Seedling counts on selective media

Hygromycin selection for potentially transformed arabidopsis can be difficult to discern due to poor growth of transformants and similar poor growth of non-transformants. To help clarify the difference between transformed and non-transformed I tested adding 0.5% suc to the selection media. It is notable that Tim 150 was added instead of sucrose to the first round of selections, wherein there may be carry-over contamination with transgenic *A. tumefaciens*. Results of this test suggested that the addition of 0.5 % suc could help clarify selection results (table A1-1). Briefly, the wildtype plants and a non-transformed sibling (non-transformed as per PCR for respective insert) all showed no large healthy seedlings on either the Hyg 25 + 0.5 % plates or those with just Hyg 25. The transformed (judged by PCR for insert) T2 generation seedlings (heterozygous) grew much better on the sucrose supplemented plates. The T2 generation seedlings all had 4 to 6 (out of about 30) big healthy seedlings grow on the selection plates lacking suc, with the rest of the seedlings germinated but produced no further growth. On suc supplemented selection media the heterozygous

transformants showed growth similar to a 3:1 ratio of growth:germination-only. I went on the assumption that Hyg resistence is purely dominant and therefore the heterozygotic siblings and homozygotic transformants should have the same phenotype. Also, I did not PCR amplify inserts from each plant involved to ensure my visual observations were correct. Thereafter, I preformed all selection counts on Hyg 25 + 0.5 % suc plates.

Seven different transformation events for the 'A' insert were assayed for segregation ratios; similarly eleven for insert 'B', eleven for insert 'C' and five for insert 'D'. Numerous more original transformation events for each insert were carried to the selfed T3 generation but with no characterization.

Plants were selected for the preliminary GUS assays described above based on the segregation ratios. Lines were chosen that had 100 % strong healthy growth on selection plates, with a preference for lines that had sibling lines with a 3:1 segregation ratio; this was to skew my line choices towards single insertion events. Unfortunately, I do not have any clear results that this selection logic worked.

A1.3.3 Preliminary GUS staining assays

All patterns observed did not fluctuate due to seedling age, fragment length or 13DNB treatment. Some lines showed very stable and similar staining across all the plantlets assayed while other lines ranged from seedlings with very dark staining to those with no staining. The pattern most often encountered was even staining of the vascular tissue in the roots and shoots and leaves (figure A1.6 a,c,d,e). This vascular-specific pattern was also seen in the extended-promoter lines (-550 bp), which I included to re-evaluate. Other patterns that emerged in all lines were: staining in outer halves of true leaves (figure A1.6 b,f,h), a combination of both vascular staining and outer leaf staining (figure A1.6 g), and just the leaf lateral meristems. Interestingly, of the plant lines transformed with only pCambia 1303, only few showed staining and those that did only very faint staining of the lateral meristems were observed.

Several possibilities exist to explain why any of the GUS staining patterns observed may not be inducible by 13DNB or could differ from the native gene

expression. The *GUS* reporter gene expression may not correlate with the native gene mRNA accumulation. Additionally, GUS expression also may not accurately depict the native At4g01870 mRNA accumulation noted in previous studies. Firstly, native mRNA accumulation may be strongly affected by post-transcriptional regulation like mRNA degradation-rate. Secondly, limitations associated with pCambia vectors could alter the GUS expression pattern; the CaMv 35s promoter driving the gene for Hyg resistance may affect or overpower the promoter fragment driving the expression of GUS. A comparison between the amounts of native gene transcripts, through PCR techniques, to the protein-activity of GUS does not take into account the possibility of post-transcriptional regulation. Both explanations may also describe why numerous transformation lines from each promoter fragment length construct show the same patterning.

Variation in patterning among all of the lines also lead to difficult interpretation of the GUS staining results. Additional possibilities to explain the variety are inherent problems with plant transformations such as, multiple construct insertions into the genome, gene silencing, and interference from nearby native promoters to the insertion site.

Based on the above technical problems and difficulties with GUS interpretation I suggest extra complementary techniques be used in addition to any replication of my experiments. To clarify the GUS reporter protein results I would use rt-pcr and *in situ* hybridization to compare native gene expression to *GUS* or *GFP* transcript expression of these constructs. If differences in expression were apparent between the native gene and reporter gene transcripts, I would look at expanding the area of where the native gene regulation may be coming from; such as, before the upstream gene (At4g01880), in the upstream gene's introns or in At4g01870's 3'UTR. I might try creating new reporter gene constructs like *GUS* flanked by any of the current promoter fragments and the 3'UTR (-100pro::*GUS/GFP*::3'UTR).

To add to the promoter characterization, I would look at the *in situ* immunoprecipitation of the native gene protein and compare it to GUS protein expression results. Some of the main challenges of this would be the creation of purified native protein or protein fragments in addition to the downstream creation of a good, working,

specific antibody. Luckily, this protein is quite different at the amino acid level from the known closest homologies in arabidopsis (At1g21670 and At1g21680), which should aid in the creation of a specific antibody. Additionally, based on information from protein structure prediction servers, the protein from At4g01870 lacks known transit sequences, is predicted to be a cytoplasmic protein (Hirokawa et al., 1998; Horton et al., 2007) and unglycosylated (lacks transit into ER system); this should make it easy to work with and to create an antibody. To create a purified population of this protein I would create a construct of the coding region attached to the nucleotide sequence for a histidine tag for protein isolation. Due to the 'simplicity' of this gene and probable lack of glycosylations, this construct could be transformed into *E. coli* for recombinant protein expression.

A1.4 Conclusions

I created four different promoter fragment constructs and successfully introduced these back into arabidopsis. To generate more accurate seedling counts on selection media I tested the addition of sucrose to the antibiotic-containing media; this addition of sucrose was found to be very helpful for seedling growth discrimination. Preliminary GUS staining results show that these four constructs produce similar expression patterns to past constructs. Additionally, the main expression pattern observed was total-plant vascular tissue staining.

A1. 5 Figures and tables

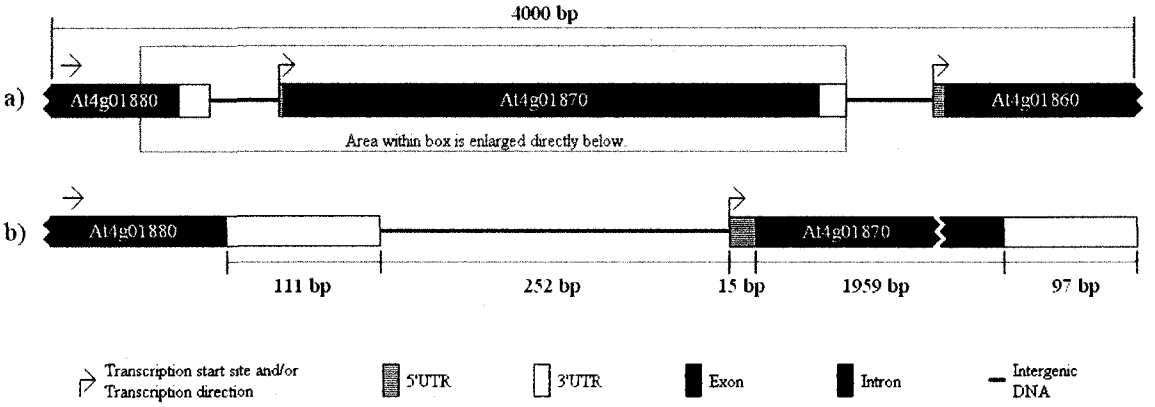


Figure A1-1 Genomic structure of At4g01870 and surrounding genes (At4g01860 and At4g01880).
A) A reverse oriented 4000 bp segment of chromosome 4 *A. thaliana.* B) A closer look at UTRs and potenial promoter region of At4g01870, also showing the close proximity between At4g01880 and At4g01870.

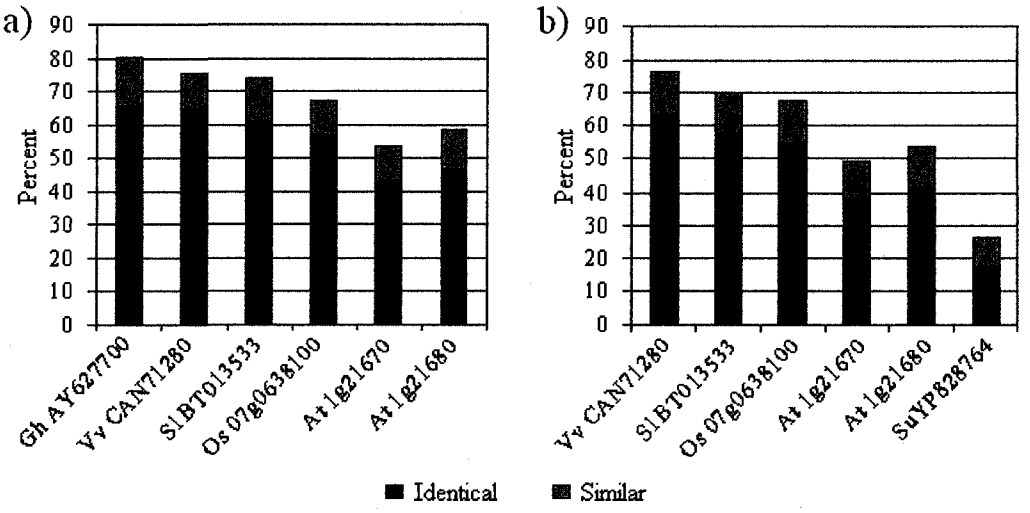


Figure A1-2 Amino acid homologies of various proteins as compared to the At4g01870 putative protein sequence.
A) Comparison of amino acid homology compared to the 97 amino acid portion of At4g01870 putative protein sequence to which the translated *G. hirsutium* resistance gene marker matches. B) Comparison of amino acid homology for related proteins as compared to the whole At4g01870 putative protein sequence. Percent refers to the fraction of amino acids that are identical or similar to the amino acids in At4g01870 when compiled in a multiple sequence alignment (appendix 2: figures A2-1 and A2-3). 'Identical' means exactly the same amino acid and 'similar' describes an amino acid that could be derived through one nucleotide substitution. Gh = *Gossypium hirsutium*, Vv = *Vitis Vinifera*, Sl = *Solanum lycopersicum*, Os = *Oryza Sativa*, At =

Arabidopsis thaliana, Su = *Solibacter usitatus*. The unique gene identifier follows each organism classifier.

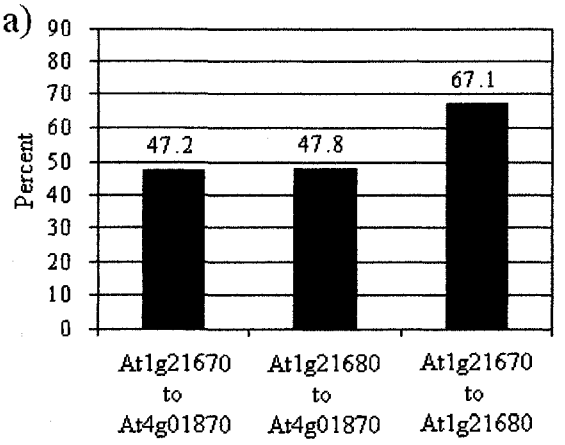


Figure A1-3 The percentage of identical nucleotides found between At4g01870, At1g21670 and At1g21680.
Calculated from a multiple sequence alignment that was created in DIALIGN using coding sequences from each gene (appendix 2: figure A2-2).

Table A1-1 Seedling counts on various selection media

	Wildtype			Tol 21.12			Tol 32.22			D2.1			D2.2			D2.3		
Media type	big	small	seeded	big	small	seeded	big	small	seeded	big	small	seeded	big	small	seeded	big	small	seeded
MS	3	46	49	6	4	30	3	14	17	7	25	34	6	27	34	24	21	49
MS+suc	44	0	45	31	0	32	49	1	51	30	0	32	33	0	33	42	3	46
MS+Hyg	0	42	42	0	33	33	4	27	33	4	32	36	5	26	33	6	0	29
MS+suc+Hyg	0	34	34	0	39	39	23	10	34	35	11	50	29	12	42	21	4	25

Note: 'big' is the number of seedlings that grew roots and green, true leaves. 'small' is the number of seedlings that remained rootless and only had cotyledons. 'seeded' is the total amount of seeds that were plated. Tol 21.12 , Tol 32.22, D2.1, D2.2, and D2.3 all refer to the T1 generation parent of the plated seeds and arise from independent transformation events (seeds plated are the T2 generation)

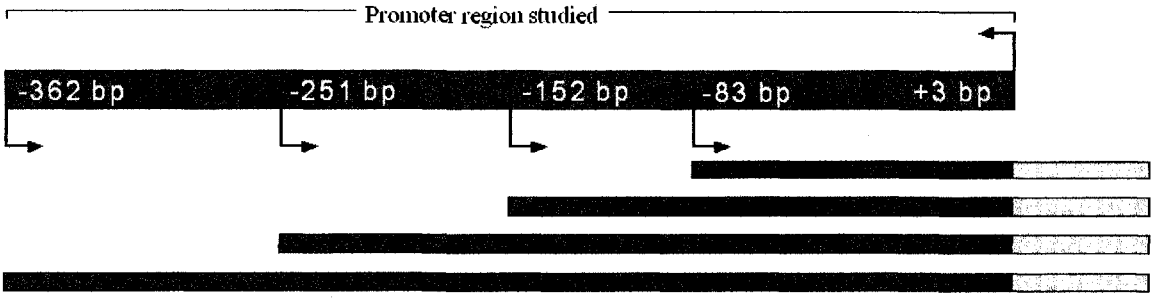


Figure A1-4 Diagram of promoter fragments and constructs created.

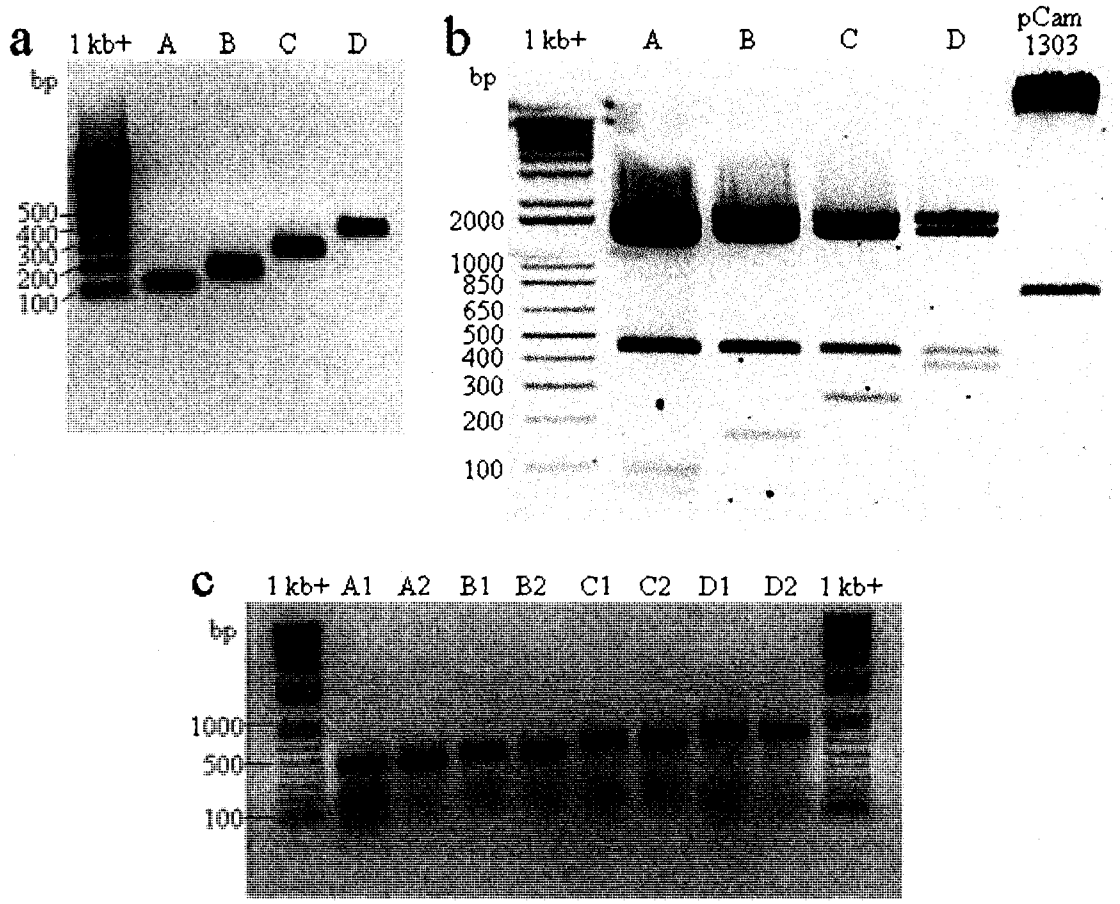


Figure A1-5 Agarose gels showing results of cloning At4g01870 promoter fragments (a) PCR using primers designed to amplify At4g01870 promoter from arabidopsis genomic DNA. (b) A double restriction digest using Nco I and Xba I to release fragments and the CaMv 35s promoter from pCambia 1303. (c) A final PCR using *A. tumefaciens* DNA to check the for the appropriate fragment lengths.

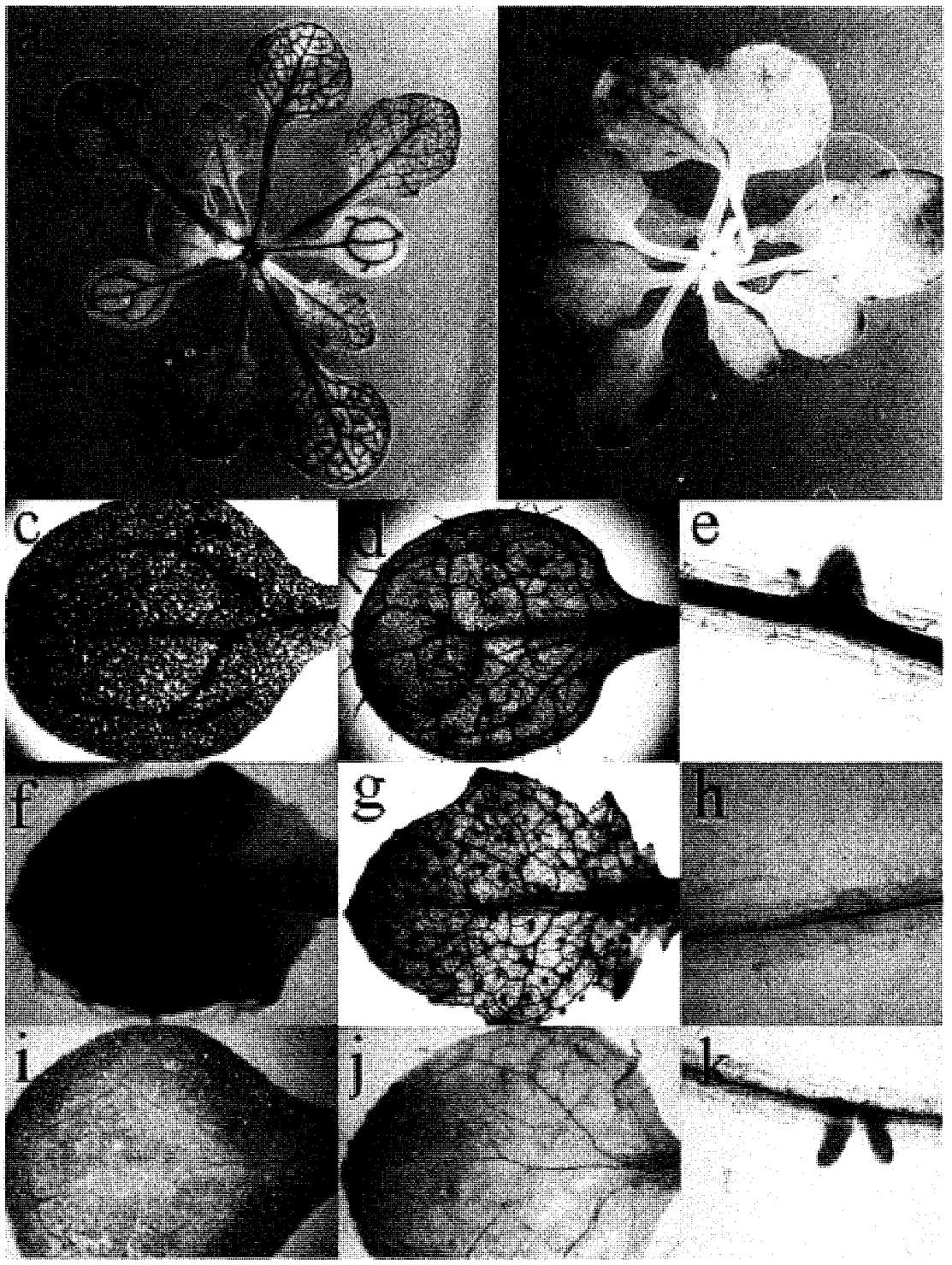


Figure A1-6 Photographs of representative GUS staining results for At4g01870 promoter deletion. (a) 20 DAI seedling with most common pattern observed. (b) 20 DAI seedling with alternative pattern of outer leaf staining. (c-e) close-up of vascular tissue staining in most common pattern; cotyledon, true leaf, and root respectively. (f) close-up of true leaf with outer-leaf staining only. (g) close-up of true leaf with both prominent vascular tissue and outer-leaf staining. (h) representative root of seedling with outer-leaf staining. (I-k) representative cotyledon, true leaf and root of unstained transgenic seedling.

A1.6 Bibliography

Abergel, C., Bouveret, E., Claverie, J.M., Brown, K., Rigal, A., Lazdunski, C., and Benedetti, H. (1999). Structure of the *Escherichia coli* TolB protein determined by MAD methods at 1.95 A resolution. Structure **7,** 1291-1300.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research **25,** 3389-3402.

Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature **408,** 796-815.

Carr, S., Penfold, C.N., Bamford, V., James, R., and Hemmings, A.M. (2000). The structure of TolB, an essential component of the tol-dependent translocation system, and its protein-protein interaction with the translocation domain of colicin E9. Structure **8,** 57-66.

Ekman, D.R., Lorenz, W.W., Przybyla, A.E., Wolfe, N.L., and Dean, J.F.D. (2003). SAGE analysis of transcriptome responses in arabidopsis roots exposed to 2,4,6-trinitrotoluene. Plant Physiology **133,** 1397-1406.

Hinchliffe, D.J., Lu, Y., Potenza, C., Segupta-Gopalan, C., Cantrell, R.G., and Zhang, J. (2005). Resistance gene analogue markers are mapped to homeologous chromosomes in cultivated tetraploid cotton. Theoretical and Applied Genetics **110,** 1074-1085.

Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics **14(4),** 378-379.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Research **35** (supp), W585-W587.

Kirkness, E.F., Wang, W., and Vazeille, A. (2004). Direct submission.

Morgenstern, B. (2004). DIALIGN: Multiple DNA and protein sequence alignment at BiBiServ. Nucleic Acids Research **32,** W33-W36.

Ohyanagi, H., Tanaka, T., Sakai, H., Shigemoto, Y., Yamaguchi, K., Habara, T., Fujii, Y., Antonio, B.A., Nagamura, Y., Imanishi, T., Ikeo, K., Itoh, T., Gojobori, T., and Sasaki, T. (2006). The Rice Annotation Project Database (RAP-DB): Hub for *Oryza sativa* ssp. japonica genome information. Nucleic Acids Research **34,** D741-D744.

Ponting, C.P., and Pallen, M.J. (1999). A beta-propeller domain within TolB. Molecular Microbiology **31,** 739-740.

Velasco, R. (2006). Direct submission.

Velasco, R., Zharkikh, A., Troggio, M., Bhatnagar, S., Pindo, M., Cartwright, D., Coppola, G., Eldredge, G., Vezzulli, S., Malacarne, G., Mitchell, J., Stefanini, M., Segala, C., Gutin, N., Grando, M.S., Pruss, D., Dematte', L., Cestaro, A., Toppo, S., Fontana, P., Skolnick, M., Gutin, A., Salamini, F., and Viola, R. (2005). The first genome sequence of an elite grapevine cultivar (Pinot noir Vitis vinifera L.): coping with a highly heterozygous genome. GenomeProject:18357

Appendix 2: At4g01870 sequence alignments

A2.1 Figures and tables

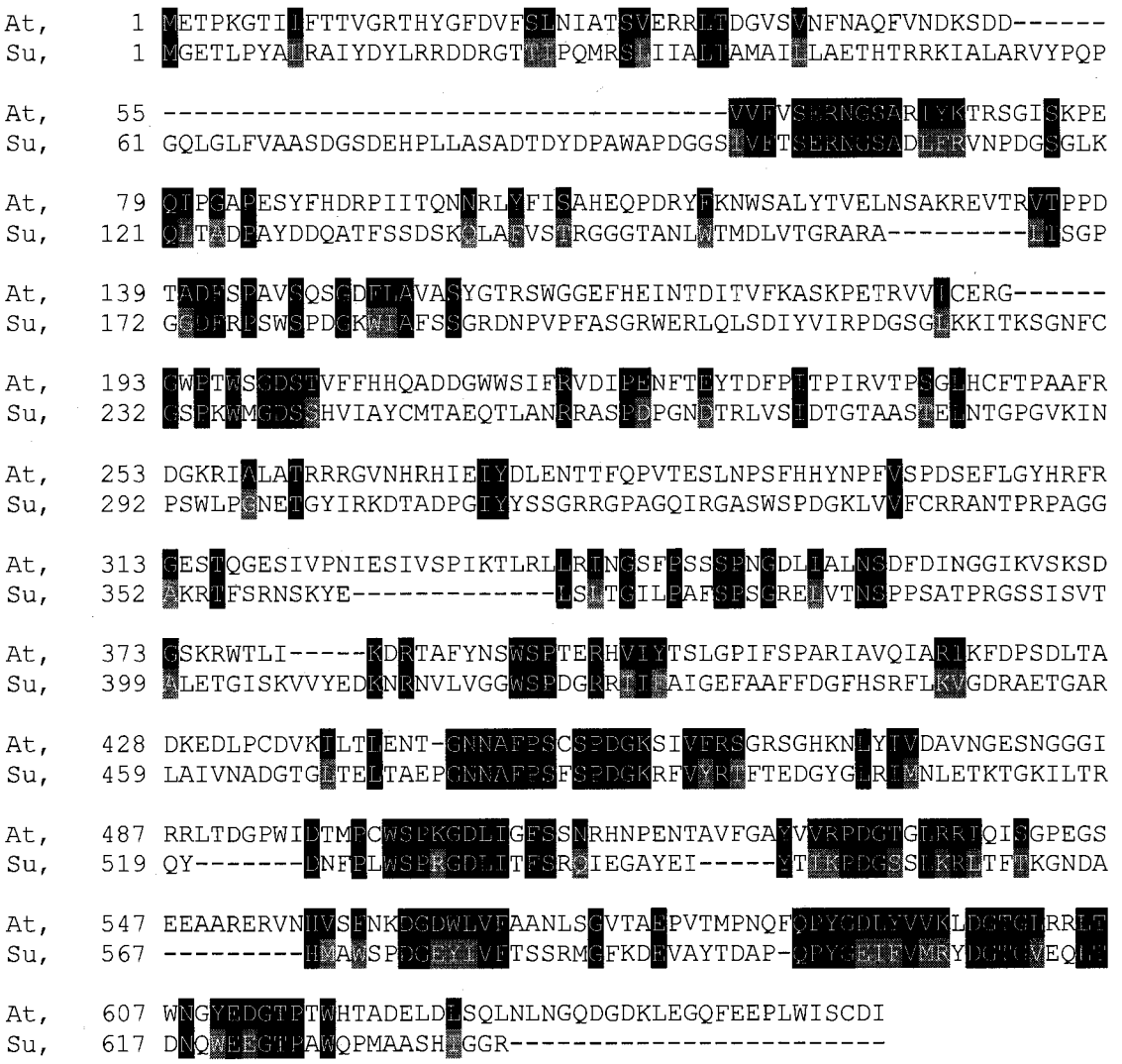
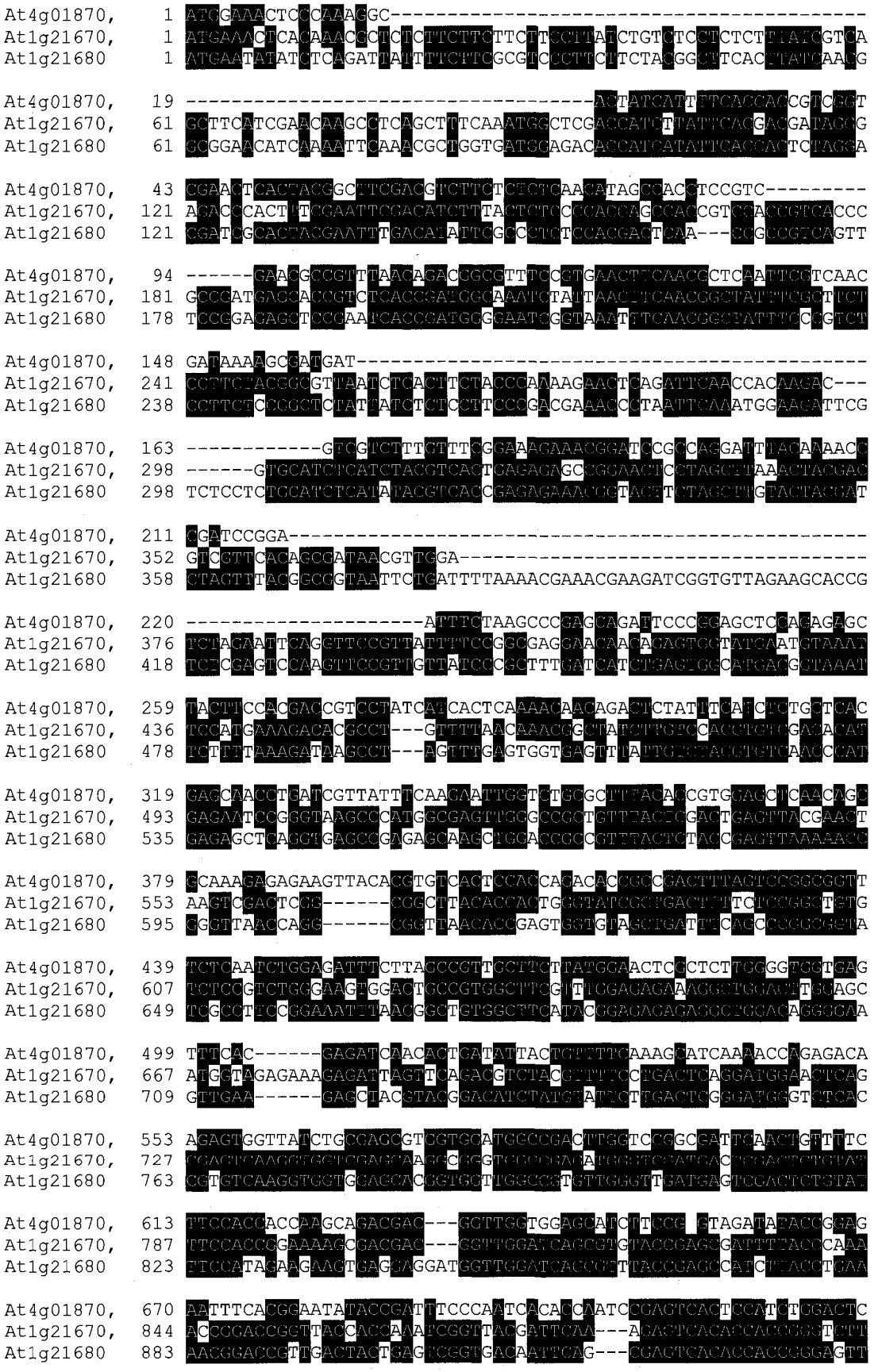
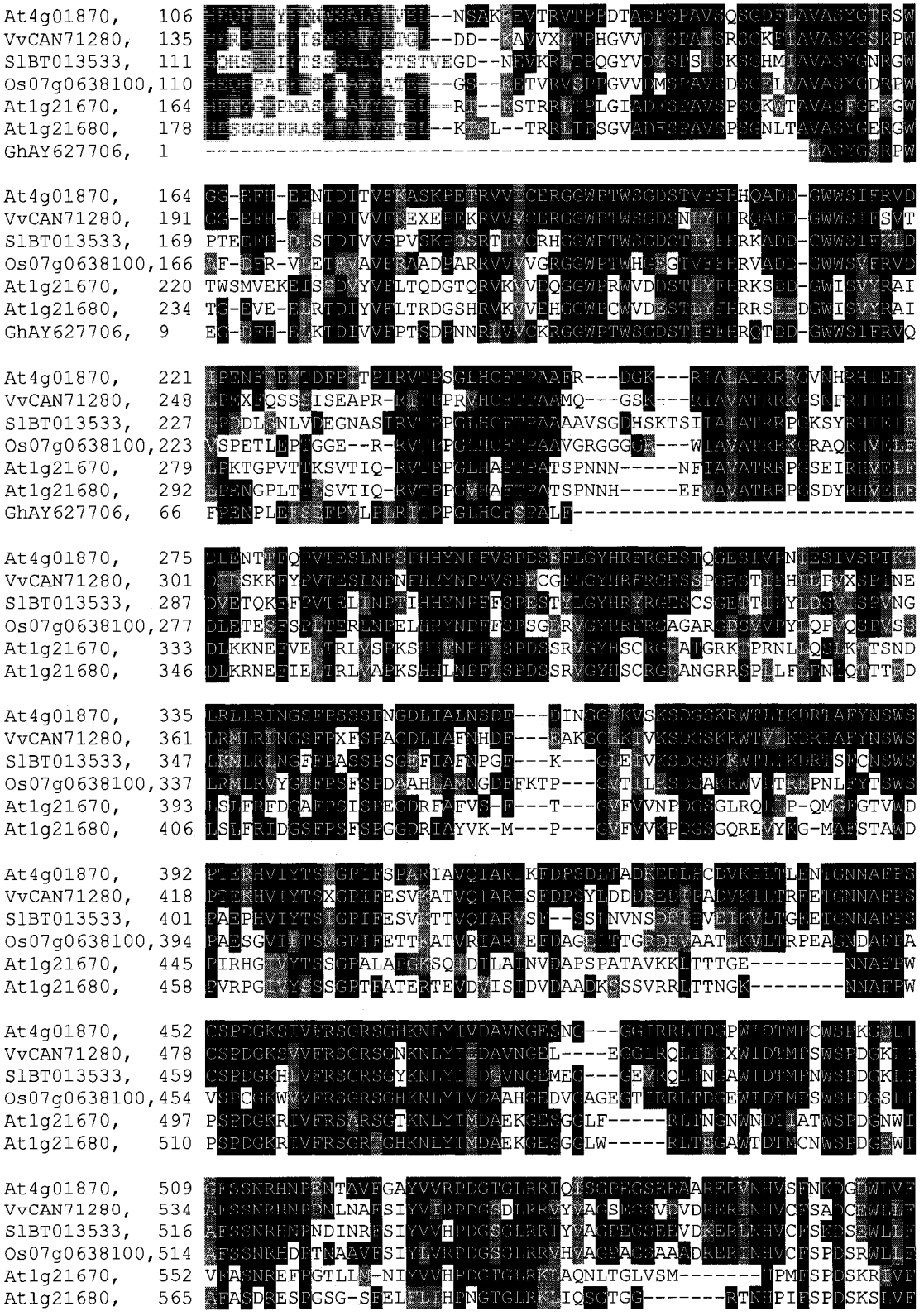


Figure A2-1 Protein sequence alignment between arabidopsis gene At4g01870 (At) and *Solibacter usitatus* strain 'Ellin6076' WD/ β-propeller domain-containing gene Su YP_828764 (Su). 'Black highlighting' means exactly the same amino acid and 'gray highlighting' describes an amino acid that could be derived through one nucleotide substitution. Alignment made with DIALIGN (Morgenstern, 2004). *A. thaliana* sequence information: TAIRAccession AASequence: 1009125680 (www.arabidopsis.org). *S. usitatus* sequence information: Accession and version YP_828764.1, GI:116626608.


At4g01870,
At1g21670,
At1g21680


At4g01870, 730
At1g21670, 901
At1g21680 940
At4g01870, 787
At1g21670, 961
At1g21680 1000
At4g01870, 847
At1g21670, 1021
At1g21680 1060
At4g01870, 907
At1g21670, 1081
At1g21680 1120
At4g01870, 967
At1g21670, 1141
At1g21680 1180
At4g01870, 1027
At1g21670, 1201
At1g21680 1240
At4g01870, 1087
At1g21670, 1255
At1g21680 1294
At4g01870, 1147
At1g21670, 1309
At1g21680 1348
At4g01870, 1207
At1g21670, 1366
At1g21680 1405
At4g01870, 1267
At1g21670, 1426
At1g21680 1465
At4g01870, 1327
At1g21670, 1471
At1g21680 1510
At4g01870, 1387
At1g21670, 1522
At1g21680 1561
At4g01870, 1447
At1g21670, 1579
At1g21680 1615
At4g01870, 1507
At1g21670, 1636
At1g21680 1675
At4g01870, 1567
At1g21670, 1696
At1g21680 1735

Figure A2-2 Coding sequence nucleotide multiple sequence alignment of arabidopsis genes. *A. thaliana* genes At4g01870 (TAIRAccession Sequence: 2141407), At1g21670 (TAIRAccession Sequence: 2036984) and At1g21680 (TAIRAccession Sequence: 2036918). Alignment made with DIALIGN (Morgenstern, 2004).



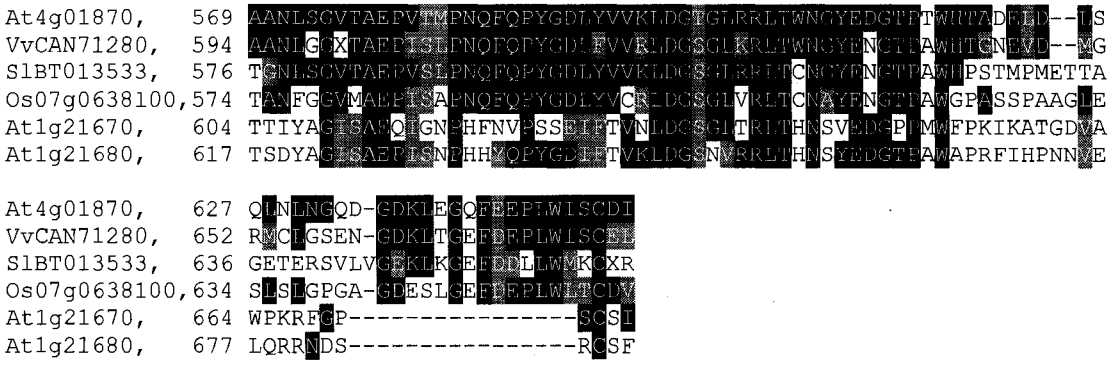


Figure A2-3 Amino acid sequence alignment comparing plant sequences that showed the most homology to At4g01870. 'Black highlighting' means exactly the same amino acid and 'gray highlighting' describes an amino acid that could be derived through one nucleotide substitution. *Arabidopsis thaliana* genes At4g01870 (TAIRAccession AASequence:1009125680), At1g21670 (TAIRAccession AASequence: 1009108523) and At1g21680 (AASequence:1009108524); *Vitis vinifera* gene Vv_CAN71280 (Accession and version: AM426496.2 GI:147767868); *Solanum lycopersicum* mRNA SlBT013533 (Accession and version: BT013533.1 GI:47104948); *Oryza Sativa* gene locus Os07g0638100 (Accession and version: NM_001066941.1 GI:115473614); *Gossypium hirsutum* gene marker GhAY627706 (Accession and version: AY627706.1 GI:49234795). Alignment made with DIALIGN (Morgenstern, 2004) and *S. lycopersicum* mRNA translated using Translate (Gasteiger et al., 2003) (www.expasy.org). For translation of SlBT013533 5' to 3' ORF 2 was chosen, aligned from first Met (M) to first Stop (X).

A2.2 Bibliography

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., R.D., A., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research **31,** 3784-3788.

Morgenstern, B. (2004). DIALIGN: Multiple DNA and protein sequence alignment at BiBiServ. Nucleic Acids Research **32,** W33-W36.

Appendix 3: Additional figures

A3.1 Figures and tables

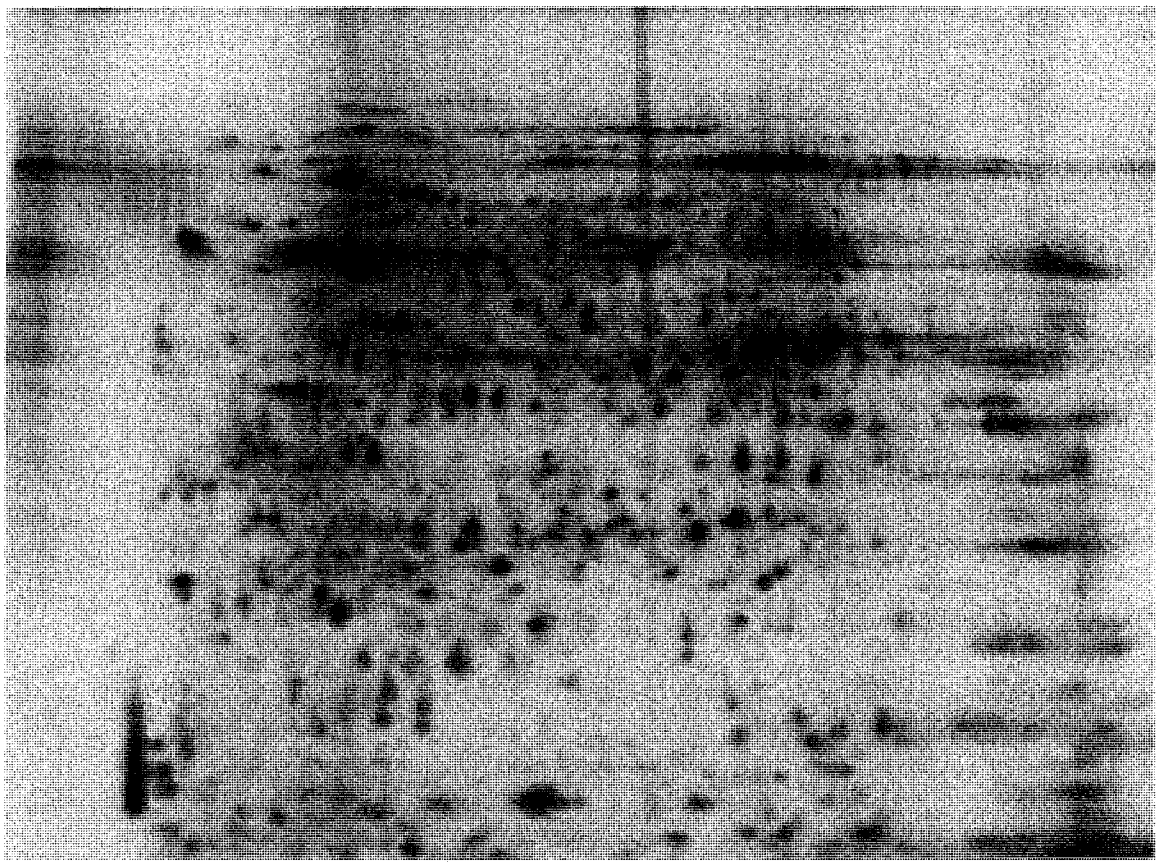


Figure A3-1 Representative fibre-sample, CyDye-labelled, DiGE image in a non-overlay and non-false-color format. This image forms the 'red' layer in the false-colour overlay (Figure 2-2).

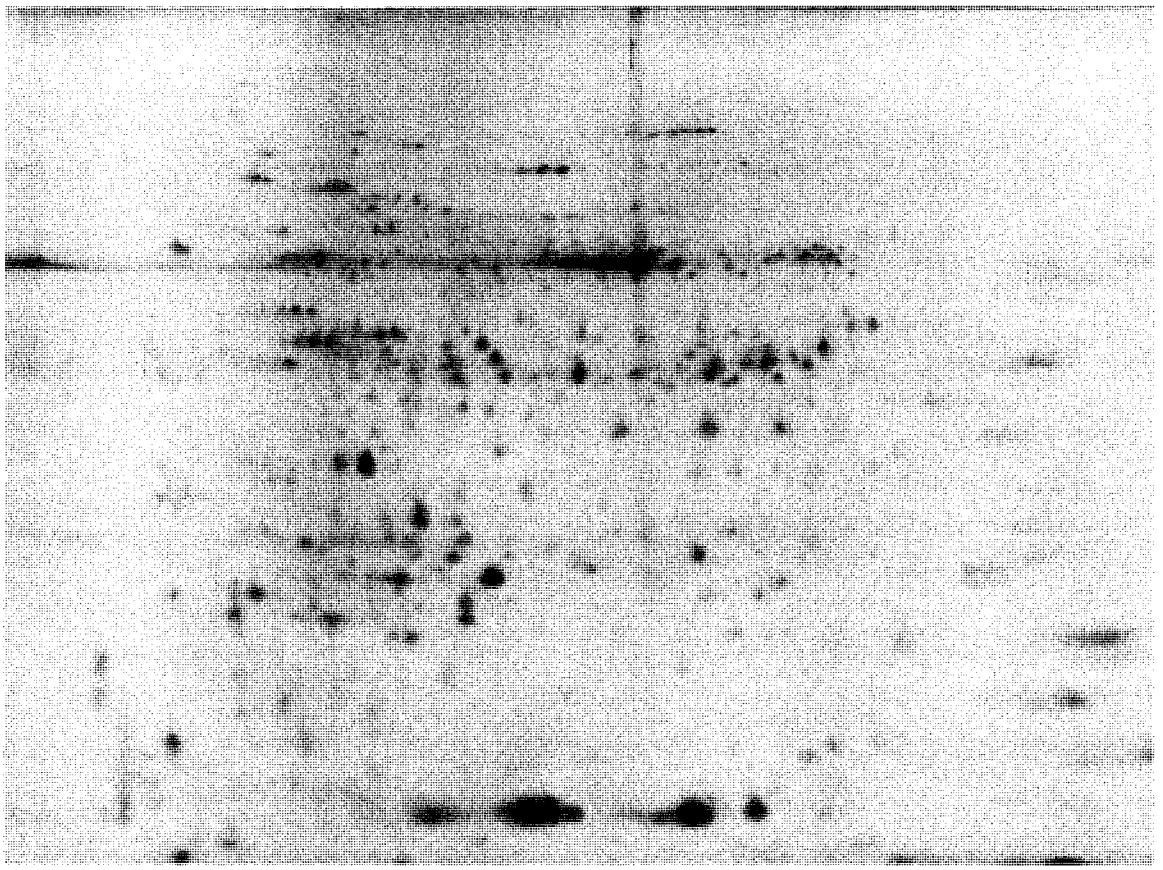


Figure A3-2 Representative surrounding-tissue-sample, CyDye-labelled, DiGE image in a non-overlay and non-false-colour format. This image forms the 'green' layer in the false-colour overlay (figure 2-3)