The importance of anatomy and physiology in plant metabolomics

Ute Roessner and Filomena Pettolino

Abstract

Plant metabolomics offers some unique opportunities in the assignment of biochemical pathways. The genetics of model plants is well-characterized which enables functional genomic approaches, qualitative trait loci identification and genetic engineering. Metabolomics has successfully supported the identification of gene function. As a specialized system, a number of key features of plants create challenges in sample preparation and interpretation of metabolomic data. Significantly, most plant tissues are composed of multiple cell types which are difficult to isolate, often resulting in limited numbers per cell type. This hinders spatial resolution of the analysis of metabolites. Secondly, cells are surrounded by a dynamic cell wall which is in constant turnover, interfering with the metabolome. Thirdly, green plant cells are capable of fixing carbon through photosynthesis producing metabolite-captured energy. This also implies a strong light-dependency in plant metabolism. Finally, plants are characterized by a diversity of secondary metabolites produced in response to environmental stimuli.

1 Introduction

1.1 Importance of plants

What is it about plants that make them so important to us? Apart from their visually pleasing qualities and contribution to some of the most famous landscapes in the world, plants provide the earth and its inhabitants with a large and varied set of irreplaceable resources of biological and economic importance. Plants account for 90 % of the biomass on Earth and contribute to the world's rich diversity with an estimated 350,000 species (Prance 2001). Plant importance begins with its position in the food chain as a primary producer, where energy is harvested from light. The processes of photosynthesis and respiration in plants are crucial in maintaining the life-essential balance of oxygen, carbon dioxide and water in the Earth's atmosphere. Plants provide food (either directly or indirectly), shelter and protection for animals, insects, fungi and even other plants.

A large number of industries revolve around plants and plant products. The most obvious are the agricultural, timber and paper industries which supply crops for food and textiles, building materials, and paper and packaging products. The

Topics in Current Genetics, Vol. 18
J. Nielsen, M. C. Jewett (Eds.): Metabolomics
DOI 10.1007/4735_2007_0218 / Published online: 14 April 2007

food and beverage industries use plant products to manufacture food and drink, and to modify their texture, flavor and/or color. The mining and manufacturing industries use plant products such as gums and resins (including latex for rubber) as binders, adhesives, emulsifiers and processing aids. Many of the medicines used today, including traditional herbal remedies, are either plant derived, based on natural plant products or contain plant extracts. Fossilized plants (coal) have been used as a source of energy for centuries but of increasing interest, and importance, is the use of plant biomass as a renewable energy source.

Plant metabolism is essential for the production of all of these plant products. The current focus of plant research worldwide is primarily the improvement of plants for food use. In addition, researchers are examining novel ways of generating plant products for the timber, pharmaceutical, green energy and textile industries. An understanding of plant metabolism at all levels is vital to the continued success of these research programs.

1.2 Plant metabolomics

Plant metabolism has been the target of research for a long time. Around 100 years ago the first concept of separation for plant specific compounds based on column chromatography was developed by Michael Tswett (1872-1920). The beauty of this technology was that he was able to separate chlorophyll, xanthophyll and carotene, based on their different colors, into clearly separated bands. A major step in plant research was achieved when about 50 years later Melvin Calvin and Andrew Benson discovered the carbon fixing dark reaction of photosynthesis, today commonly called the 'Calvin cycle'. Although of immense importance, the photosynthetic process has not been the only plant feature of interest as other plant specific pathways have been studied in great detail. These include the starch synthetic pathway, cell wall synthesis, vitamin production, protein and lipid metabolism. In the last century, an endless number of analytical methodologies have been developed for the extraction, detection and quantification of plant metabolites, always with the emphasis on increasing our understanding of plant metabolism, improving plant products or increasing crop yield. The exciting development of possibilities to specifically alter plant genomes by either mutations or by introduction of additional genes has opened a new opportunity in plant sciences. The release of the complete sequence of the flowering plant *Arabidopsis thaliana* at the end of the nineties has provided a great improvement in understanding not only plant biology, but also evolution and development. In parallel, novel multi-parallel and/or highly sensitive analytical tools have been developed for a comprehensive analysis of the different cell products. Most prominent amongst these new technologies has been the establishment of protocols for the determination of the expression levels of many thousands of genes in parallel (transcriptomics), the detection, identification and quantification of the protein complement (proteomics) and the determination and the simultaneous identification of a large number of metabolic compounds in a high-throughput manner (metabolomics). Metabolomics today can be considered as the accumulation and

combination of knowledge of analytical biochemistry from the last 50 years and its application towards developments of new technologies with greater sensitivity, comprehensiveness, robustness and higher throughput. Currently in the field of metabolomics, both gas and liquid chromatography coupled to various mass spectrometric detection technologies (GC- and LC-MS) are applied to analyze complex metabolite mixtures. In addition, nuclear magnetic resonance spectroscopy (NMR) has been successfully used to fingerprint plant systems. Very recently, the power of capillary electrophoresis coupled either to laser induced fluorescence detection or mass spectrometry has been discovered. The advantage of this technology is its great sensitivity allowing the analysis of a large range of metabolites in very small sample sizes. The principles, advantages and disadvantages of each of the available technologies have been described in great detail in a large number of published reviews and books (e.g. Sumner at al. 2003; Hall 2006; Saito et al. 2006; Villas-Boas et al. 2007) and will therefore not be discussed in this chapter. In addition, endless numbers of publications are available with exciting and impressive applications of metabolomic technologies in many different scientific fields. The future of research will be driven by the exponential growth of metabolomics as its own entity in the 'omics' sciences. It is important to note, that metabolomics has attracted increasing interest, not only from biologists but also from the public and politicians. Concurrent with the evolution of metabolomics is the assured confidence in the validity of the data obtained and in the way it is applied.

In the following we want to present another perspective of plant metabolomics. As plants are unique and essential members amongst all living organisms we would like to place special emphasis on the distinctiveness of plant systems and relate these back to important factors to consider when conducting metabolomics experiments in plant research.

2 Plant anatomy

2.1 Whole plant anatomy

Most plants are immobile and therefore have to quickly and efficiently adapt to changing environments. In general, plants are built of three basic organs: leaves, stems and roots, which are made of four types of tissue including the vascular, the dermal, the ground and the meristematic tissues. The roots anchor the plant in the soil and are required to absorb and transport water and nutrients from the soil to the other parts of the plant. There are 13 minerals essential to all plants, including macronutrients, such as N and P, and micronutrients, such as Na, K, B, Mn, Fe, Ca. If a plant grows in mineral deficient conditions it affects the plant growth dramatically and in the worst case can kill the plant. On the other hand, excess amounts of most of these minerals may be harmful and thus result in the presentation of toxicity symptoms which again affect growth and reproduction. In both cases, the plant has to develop mechanisms to withstand these conditions for sur-

vival. Since plant metabolism is dramatically affected by both mineral deficiency and toxicity, metabolomics approaches are currently used to monitor metabolic changes following inadequate mineral supply to gain an increased understanding of plant mechanisms for adaptation or even the development of tolerance.

The stems have two major functions, firstly, to hold up the leaves for optimal exposure to sunlight and secondly, to transport water and nutrients via the xylem and soluble carbon sources and hormones via the phloem within all parts of the plant. In contrast, the main function of leaves is to 'host' the process of photosynthesis. Photosynthesis occurs in chloroplasts, specialized green cellular compartments where light energy is captured for the production of glucose from CO_2 and water.

When analyzing plant metabolism, the anatomic complexity of plants has to be considered. Each plant organ, tissue or cell type is characterized by a specific set of metabolites in a certain distribution/concentration and is often differentially affected by external stimuli. Currently, due to the low sensitivity of analytical technologies used in metabolomics, metabolites from a sufficient amount of tissue have to be extracted for comprehensive coverage of many metabolites simultaneously. Therefore, often many different cell types and tissues may be combined and only the 'average' of the metabolite content determined. Successful attempts at single cell metabolite analysis have already been reported. Schad et al. (2005) collected enough material composed of specific cell types from cryo-preserved and laser micro-dissected tissue to analyze about 68 major metabolites by GC-MS. Unfortunately, often it is very difficult or even impossible to separate and isolate specific cell types from plant tissues. Another exciting approach for cellular as well as subcellular specific determination of metabolite abundance has been presented by Fehr et al. (2004). The authors describe the development of protein-based fluorescent-tagged nanosensors for imaging specific metabolites. One important feature of this technique is that it is almost non-invasive and can be applied to monitor dynamic changes of metabolites and also ion levels in the cells, tissues or organs of interest (Fehr et al. 2004).

2.2 Cell anatomy

All eukaryotic cells share anatomical features. They are surrounded by a plasma membrane, have a nucleus containing the cell's genetic information along with a nucleolus for processing and assembly of ribonucleoprotein subunits, an endoplasmic reticulum and Golgi apparatus, mitochondria, ribosomes, peroxisomes and vacuoles. In addition, plant cells contain plastids and are surrounded by a cell wall (Fig. 1).

The plant cell wall is a rigid semi-permeable structure surrounding all plant cells. The principal component of the plant cell wall is the cellulose microfibril framework which is embedded in a matrix of non-cellulosic polysaccharides. The nature of the polysaccharide matrix is very much dependent on the plant species and the developmental stage of the cell. For higher plants, Gibeaut and Carpita (1998) have defined two types of primary walls. Type I is typical of most monocot

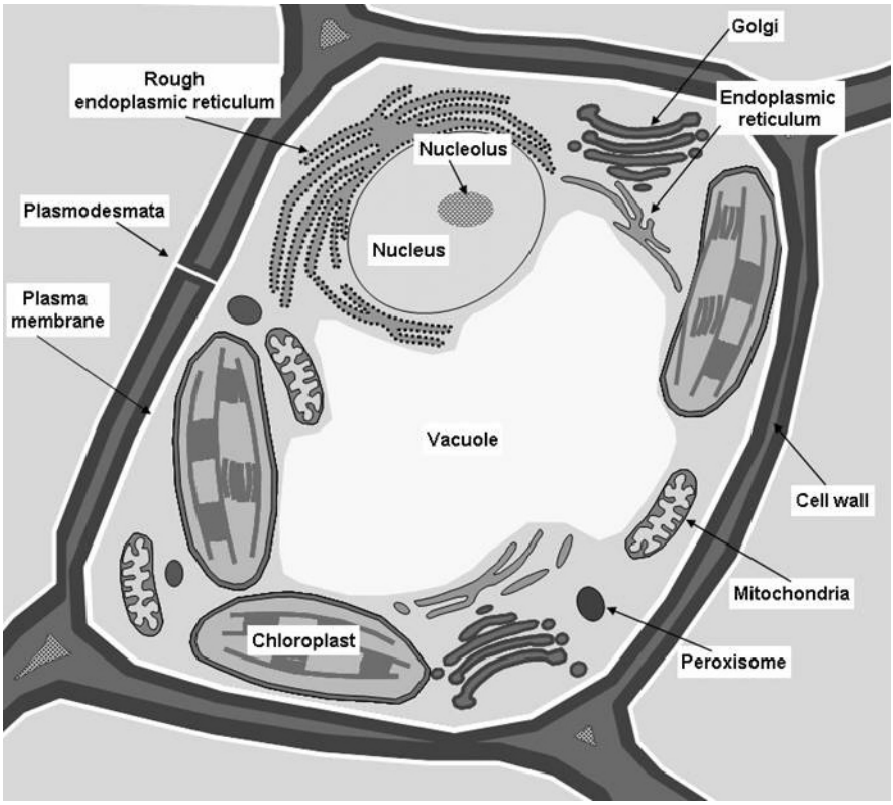

Fig. 1. A typical plant cell. Plant cells are distinguishable from animal cells by the presence of the cell wall, plastids (chloroplast), and large vacuole. Plant cells also possess plasmodesmata that allow for cell-to-cell molecular interactions.

and dicot species where xyloglucan, and/or glucomannan, associates with the cellulose microfibrils to form a framework embedded in a gel-like matrix of pectins. The Type II walls are specific to the commelinoid monocots (e.g. grasses) and contain glucuronoarabinoxylans in place of xyloglucan, and depending on the cell type and stage, also mixed-linkage β-glucans (Gibeaut and Carpita 1998). Proteins serve structural and catalytic roles in the cell wall and are involved in the strengthening and manipulation of the various components of the wall during growth and development. Secondary walls develop internal to the primary walls where further modification of the polymers is evident, including the deposition of lignin and suberin.

Living plant cells are enclosed within a plasma membrane that is restricted against the cell wall due to turgor pressure. The plasma membrane is involved in signal transduction and assists in the regulation of molecular transport into and out of the cell. In plants, particular areas of the plasma membrane combine with elements of the endoplasmic reticulum to form membranous tubes called plas-

modesmata. Plasmodesmata provide a direct physical link between adjacent cells for channeling and communication (McLean et al. 1997).

The endoplasmic reticulum (ER) forms a dynamic network of membranes involved in the synthesis, processing and sorting of targeted proteins. The ER also provides anchor sites for actin filaments and is the site of lipid synthesis and the initiation of *N*-linked glycosylation. Parts of the ER also form oil and protein storage bodies where vegetable oils and seed storage proteins that are of human nutritional value are stored (Galili et al. 1998).

Newly synthesized proteins travel through the ER via budding transport vesicles to the Golgi apparatus where they are directed to vacuoles or the cell surface. The Golgi is mobilized throughout the cell along actin filaments which undoubtedly contributes to the spatial organization and processing of cellular metabolic processes that occur through this organelle (Nebenführ and Staehelin 2001). *O*-linked glycosylation of proteins through serine, threonine and hydroxyproline (instead of hydroxylysine in animal *O*-glycosylation) occurs in the *cis*-Golgi. The carbohydrate moieties of glycoproteins that are initially *N*-glycosylated in the ER can be further processed in the Golgi. Plant proteins can have two types of *N*-linked glycans; the high mannose type consisting of the unit $(Man)_{6\text{-}9}(GlcNAc)_2$ and the complex type, which is the Golgi modified version of the high mannose glycans. The complex glycans of plants consist of the core structure $Xyl(Man)_3Fuc(GlcNAc)_2$. This differs from the core structure of mammalian complex glycans in that the Fuc attached to the proximal GlcNAc in mammals is α-1,3-linked, and in plants is α-1,6-linked. Furthermore, the Xyl that is present in plant glycoproteins is absent in mammalian glycoproteins (Sturm 1995). Additional processing of *N*-linked glycans can occur in the vacuole or extracellular compartments in transit to their final destination (Rayon et al. 1998.). In addition to containing the enzymes involved in protein and lipid glycosylation, the plant Golgi is also the site of synthesis of pectic and non-cellulosic cell wall polysaccharides.

A defining feature of plant vacuoles is their size, capable of occupying over 30 % of the cell volume. The turgor pressure of the cell is maintained by the osmotic uptake of water as solutes accumulate in the vacuole. Turgor pressure, along with cell wall extensibility, drives plant cell enlargement and expansion. Vacuoles store inorganic ions, important for pH and ionic homeostasis; organic acids, including amino acids; sugars; enzymes such as proteases, nucleases, glycosidases and lipases important for digestion; proteins; and secondary metabolites such as pigments and defensive molecules (phenolics, alkaloids, cyanogenic glycosides, saponins) (Marty 1999). Due to their chemical nature not all of these molecules are likely to be found in any one vacuole. In plant cells at least two types of vacuole have been identified; the neutral, protein-storing vacuoles and the acidic, lytic vacuoles (Staehelin and Newcomb 2000).

The role of peroxisomes in plant cells is organ or tissue specific. Peroxisomes are involved in the conversion of fixed N_2 into nitrogen-rich organic compounds in legume root nodules. Glyoxsomes are specific peroxisomes involved in lipid metabolism in germinating seeds that store fats. In leaves, peroxisomes, in conjunction with mitochondria and chloroplasts, participate in photorespiration. Per-

oxisomes serve a protective function in that the hydrogen peroxide that is liberated in each of these metabolic processes is destroyed by their resident catalases.

The mitochondria of plant cells are typical of the eukaryotic organelle responsible for the generation of ATP via the citric acid cycle and associated electron transfer chain. Plant mitochondria have a much larger genome than in other organisms, ranging in size from 200,000 to 2,600,000 nucleotides (compared with 15,000 – 18,000 in mammals). The plant mitochondrial genome, which codes for only 16 of the 20 tRNA genes required for protein synthesis, also contains some chloroplast DNA, most of which is non-functional in mitochondria (Staehelin and Newcomb 2000).

Amongst the eukaryotes, plastids are found only in plant and algal cells. There are a number of different plastids that can exist in a plant cell, each of which serves different functions. All plastids begin as proplastids that develop into, or convert from one type of plastid to another. Amyloplasts and leucoplasts are non-pigmented plastids that store starch and are involved in monoterpene synthesis, respectively. Etioplasts, which arise when chloroplast development is arrested due to the lack of light, store tubular membranes as semicrystalline structures called prolamellar bodies. The lipid membranes transform into thylakoids when the etioplast is illuminated and progresses in development to a chloroplast. Chromoplasts synthesize and store carotenes and xanthophylls giving them the yellow, orange or red coloring seen in many fruits, flowers and vegetables. Chloroplasts are the green chlorophyll containing plastids responsible for energy capture from sunlight. The photosynthetic machinery of chloroplasts resides within the thylakoid membrane system composed of stacked grana that are interconnected via the unstacked stroma.

Each of the above described compartments is characterized by their own suite of metabolites as well as concentration patterns. This is especially important to be considered as most metabolomics approaches cover metabolites extracted from whole cells, tissues, organs or even plants. Therefore no information is obtained about metabolite levels and changes within and between compartments, e.g. following environmental stimuli or genetic alteration. Recently, there have been efforts to develop metabolite analysis tools at the subcellular level. Farré et al. (2001) applied a non-aqueous fractionation technique to separate plastids, vacuoles and cytoplasm/mitochondria based on their individual density from potato tubers. The resulting fractions were characterized using compartment-specific enzyme marker assays to determine the distribution of compartments in each fraction. The percentage distribution was further correlated with levels of about 60 metabolites, analyzed using GC-MS, to give an estimation of metabolite concentrations in the different compartments (Farré et al. 2001). As mentioned above, more recent and extremely promising developments for subcellular metabolite imaging are based on fluorescent-tagged nanosensors (Fehr et al. 2004).

3 Plant physiology – Challenges for plant metabolomics

3.1 Photosynthesis

There is essentially a lot of similarity between plant primary metabolism and those of all other organisms. But, the ability of green plants to capture energy from light for the production of high-energy containing molecules, has equipped plants with a number of unique reactions. Most well known and studied is photosynthesis which is characterized by two major processes. The first is the capture of light energy for the production of ATP and the reducing equivalent NADPH. Figure 2 represents a schematic overview of the importance of photosynthesis for the supply of energy and carbon molecules for a range of metabolic processes in plant cells. The prerequisite of this so called light-reaction is the presence of chlorophyll. The second step is light-independent and produces glucose from carbon dioxide and water using ATP and NADPH, and releases oxygen. This is a very advanced process where the enzyme ribulose-1,5-bisphosphate carboxylase (Rubisco), actually the most abundant protein in green tissues, binds 6 molecules of carbon dioxide to 6 molecules of ribulose-1,5-bisphosphate producing twelve molecules of 3-phosphoglycerate. The 3-phosphoglycerates are further metabolized to release one molecule of glucose and resynthesize 6 molecules of ribulose-1,5-bisphosphate for the next cycle. The glucose is then the key metabolite for all down-stream metabolic processes, both for biosynthetic pathways or respiration via glycolysis. In most plants, sucrose is the transport form of carbon throughout the whole plant.

Photosynthesis is therefore the key process dictating the great dependency on light availability and intensity for many metabolic processes in plant cells. A large range of metabolic enzymes are regulated either directly by light or by the resulting glucose or sucrose. As a consequence, substantially different metabolite quantities are present during the day compared to the night. This has been shown in an in-depth analysis of leaf metabolites during diurnal rhythm in potato and rice (Sato et al. 2004; Urbanczyk-Wochniak et al. 2005). Therefore, special care has to be taken with the time of the day when leaf samples for metabolomics studies are harvested. Leaves however, are not the only tissue to show a light-dependent metabolite profile. As demonstrated by Roessner-Tunali et al. (2003a), even heterotrophic tissues such as potato tubers which grow in the dark in the soil, show a differential metabolite profile in the course of a day because they are dependent on the 'delivery' of sucrose from the aerial parts for starch production.

3.2 Photorespiration

Plants have to develop a specialized mechanism to survive in situations where the CO_2 levels inside a leaf become very low. This occurs in very hot and/or dry environments which cause a closure of the stomata to avoid undesired water loss, resulting in insufficient CO_2 uptake. Rubisco is a dual functional enzyme, which in

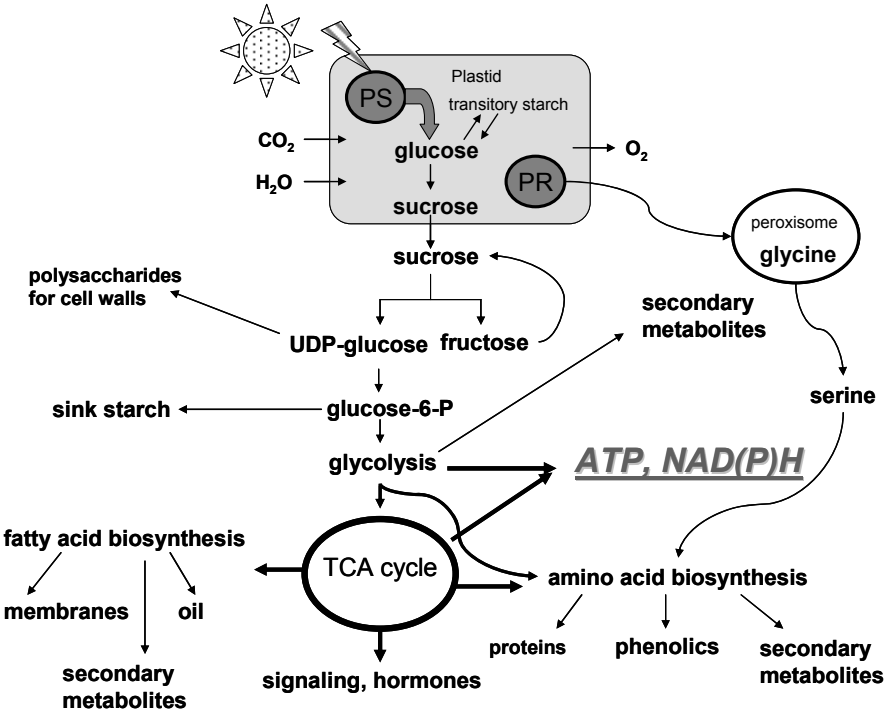

Fig. 2. Simplified scheme of metabolism in green plant tissues. PS – photosynthesis, PR – photorespiration, TCA – tricarbolic acid cycle.

low CO_2 conditions prefers to accept O_2. This leads to the production of 2-phosphoglycolate, which is toxic for the plant cell, and to reduction of ATP production. In this case, the plant uses a series of enzymatic steps to transform 2-phosphoglycolate into non-toxic and even metabolically useful compounds. The first step cleaves the phospho-group to produce glycolate. After transport of this molecule from the plastids to the peroxisomes, it is transformed into glycine, by the release of CO_2, and is then transported to the mitochondria. There, glycine is further converted to serine which can be either channeled into amino acid and protein metabolism or metabolized to form 3-phosphoglycerate, an important intermediate in glycolysis and a useful precursor for other primary metabolites. Unfortunately for the plant, all these conversions result in a net loss of CO_2 and the use of ATP and reducing equivalents. This pathway demonstrates an interesting example of how cells can control carbon flow by separating metabolic reactions into different compartments. The process of photorespiration involves the action of three compartments, the plastid, the peroxisome and the mitochondria. This means that transport proteins specific for each of the compartments and for the respective molecule requiring transport, have to be expressed and activated.

3.3 Transpiration

The evaporation of water from leaves and the stems of plants into the atmosphere are called transpiration. Water is absorbed from the soil by the roots and pumped through the vessels to the upper parts of the plant. The actual process of evaporation occurs through small pores called stomata which are located on the lower side of the leaves. The closure state of these stomata controls the amount of water released and therefore is extremely important for the balance between water gain and water loss of the plant and hence, for the actual water availability in the tissues. For instance, under water limiting conditions, the stomata are closed very rapidly in order to reduce water loss. Therefore, the opening status of the stomata can control the water content of the plant tissue which can have two effects with respect to metabolomic analysis. Firstly, water availability in plant cells has remarkable effects on all metabolic pathways and therefore metabolite levels. In addition it can result in stress-induced responses. Secondly, when samples are prepared for metabolite extraction and fresh weight is used as a way of normalization, the amount of water in the harvested tissue will influence the fresh weight and consequently the evaluation of metabolite levels. Therefore, it is most important to keep the water availability, the temperature and light intensity consistent when growing plants for comparative metabolomics as well as other 'omics' studies.

3.4 Starch and other storage products

Plants are important food components as they store high-energy products, such as carbohydrates, fats and proteins. Carbohydrates can be stored as free sugars, such as hexoses in fruits or sucrose in sugar cane, or polymerized in the form of cell walls and starch. Starch is a plant specific storage product and consists of an endless number of polymerized glucose polymers. In plants, two types of starch are produced, transitory and storage starch. Transitory starch, which is a store for excess glucose made in green leaves during photosynthesis in the light, is degraded during the dark period and distributed throughout the plant for energy production via respiration or for delivery to sink organs for long-term storage. This type of starch mainly occurs in non-photosynthetically active (heterotrophic, non-green) tissues, such as tubers or grains. As mentioned earlier, starch is made of long chains of glucose molecules. The chemistry of these long chains determines the type of starch. On one hand, glucose monomers are linked by α-(1,4)-glucosidic bonds resulting in amylose, a linear, helical polymer that aggregates to form insoluble starch granules. In the other form of starch, amylopectin, these α-(1,4)-glucose chains are further substituted by α-(1,6)-glucosidic linkages forming more complex and branched structures. The biosynthetic pathway of starch starts with the formation of nucleotide-activated glucose by the enzyme ADP-glucose pyrophosphorylase. The ADP-glucose is then used as a substrate by starch synthase enzymes, which add glucose units to the end of a growing polymer chain to build up a starch molecule (releasing the ADP in the process). Branches in the chain are introduced by starch branching enzymes (SBEs), which hydrolyze α-(1,4)-

glycosidic bonds, and in their place, create α-(1,6) bonds with other glucose units. The investigation of the starch synthetic and degradation pathways, and their differences in different species and tissues, has been a target of research for many years particularly with the emphasis on increasing yield of starch-storing crops. The analysis of the intermediates of the starch biosynthetic pathway and also the many other metabolites either directly associated, e.g. from glycolysis or the TCA cycle or metabolites indirectly involved in this pathway will indicate what factors influence the flux of carbon into starch. This will result in improved and more efficient approaches to modify starch with respect to increased yields and altered structure for specific industrial applications.

3.5 Cell wall synthesis

A large proportion of the glucose generated by plant cells is directed towards cell wall synthesis. New cell walls form after nuclear division when a phragmoplast containing actin, myosin and microtubules assembles a cell plate between the nuclei. New wall components carried via Golgi vesicles are deposited at the cell plate which continues to grow from the centre towards the edges of the cell until it fuses with the existing wall. Biosynthesis of the plant cell wall is a highly regulated process due in part to the complex nature of its structure, the location of the biosynthetic machinery and the coordinated changes that take place during growth and development. Cell wall biosynthesis itself involves a number of metabolites and it is the synthesis, conversion and transport of these that are important in cell wall development. The synthesis and shuffle of carbohydrates, nucleotides, proteins, amino acids, lipids (e.g. sterols), phenolics, growth regulators and cofactors (e.g. acetyl-CoA) are critical to the growth and development of the plant cell wall, and in turn, the plant.

Polysaccharides, the major components of plant cell walls are a secondary gene product; the primary gene product being the synthases and transferases responsible for their synthesis. As a secondary gene product no template is available (unlike protein synthesis) and yet the general polysaccharide structures present in a given wall at a specific developmental stage are consistent. The identification of the cellulose synthase genes (Ces A) opened a flood gate for the discovery of a large number of genes encoding putative polysaccharide synthases. These genes have been classified according to their similarity to the CesA genes to give the CSL (cellulose synthase-like) gene families (http://cellwall.stanford.edu/). Interestingly, the genomic approaches to study cell wall biosynthesis have shown that each tissue can have multiple Ces and CSL genes, presumably to account for the different types of polysaccharides synthesized as well as providing the ability to switch on different genes to coincide with a particular developmental stage. Despite the large number of genes identified, only a handful have been unambiguously identified as polysaccharide synthase genes (Scheible and Pauly 2004) and the actual mechanisms involved in cell wall biosynthesis are not clearly defined. Metabolite profiling has the potential to shed light on some of these mechanisms.

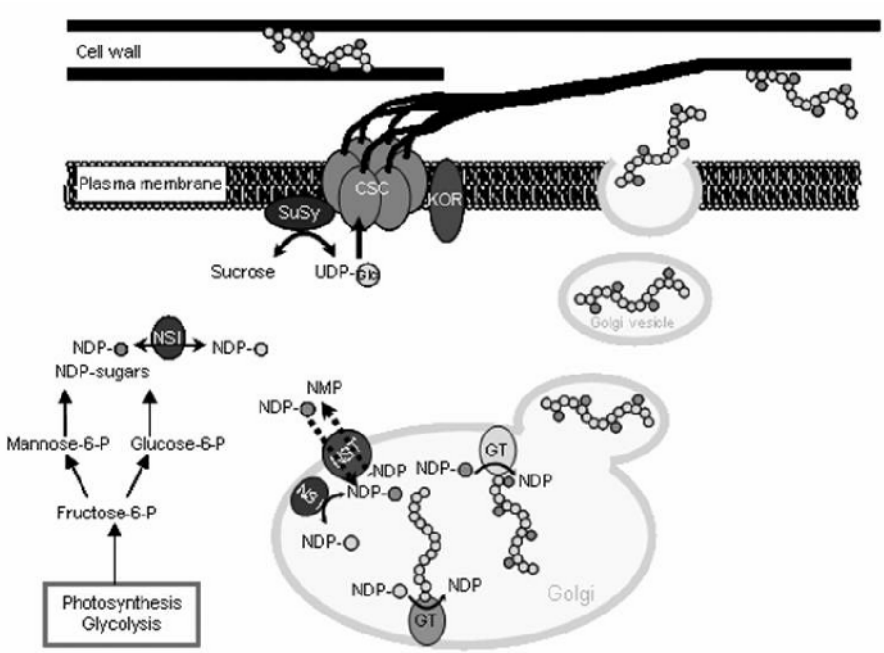

Fig. 3. Primary cell wall synthesis. Cellulose is synthesised at the plasma membrane by CSC, (Cellulose Synthase Complex) and is associated with other enzymes including SuSy (Sucrose Synthase) and KOR (KOR endoglucanase). Nucleotide sugars (NDP) are synthesised and converted in the cytosol (and most likely the Golgi) by Nucleotide Sugar Interconverting enzymes (NSI). Matrix polysaccharides are synthesised by GTs (Glycosyl Transferases) in the Golgi and transported to the wall. NDPs are transported to sites of polysaccharide synthesis including inside the Golgi by Nucleotide Sugar Transporters (NSTs).

In the synthesis of the wall polysaccharides (Fig. 3), nucleotide sugars are the donor substrates that provide the specific monosaccharides to be attached to the growing polysaccharide chain. This is also the case for glycosyltransferases which add specific sugars to a preformed polysaccharide backbone. Cellulose and callose (a developmentally regulated polysaccharide that occurs only in specialized cells and in response to wounding) are synthesized at the plasma membrane. The other cell wall polysaccharides, including xyloglucan, pectins, arabinoxylan and heteromannans are synthesized in the Golgi (Moore and Staehlin 1988) and transported to the cell surface for release into the wall space via vesicular transport mechanisms. The cellulose synthetic machinery that occurs on the plasma membrane forms rosette structures composed of cellulose synthase hexamers. A number of other enzymes appear to be closely associated with the complex, including sucrose synthase (SuSy), which presumably supplies UDP-glucose as the donor substrate to the cellulose synthase (Delmer and Amor 1995), and KOR endoglucanase. There has been some evidence to suggest that the acceptor for initiation

of cellulose synthesis is a sitosterol-linked β-glucan and it has been proposed that the KOR endo-glucanase cleaves the lipid-linked glucan chain after initiation of polymerization (Peng et al. 2002).

Nucleotide sugars are synthesized in the cytoplasm and must be transported to the sites of polysaccharide synthesis in both the Golgi and at the plasma membrane. Nucleotide sugar interconverting enzymes have also been associated with the Golgi (Baldwin et al. 2001). The evidence for the presence of nucleotide sugar transporters in Golgi membranes suggests that nucleotide sugars that are synthesized in the cytoplasm can be transported into the Golgi and perhaps converted to raise the population of various sugar nucleotides required for polysaccharide synthesis (Orellana 2005; Reiter and Vanzin 2001). With respect to metabolite profiling, the extraction, fractionation and detection of nucleotide sugars is complicated by their structural similarities and low abundances. Methods currently used are limited to the simultaneous detection of only a few metabolites, but new developments in methodology (e.g. Ramm et al. 2004) are helping to overcome these issues.

The cell wall is a dynamic structure and once deposited will undergo many changes according to developmental stage, tissue type and environmental influences. These include modification of the polysaccharides by hydrolytic, de-acetylation and de-methylesterification processes. Acetylation of polysaccharides occurs in the Golgi with acetyl-CoA as the acetyl donor and is thought to protect the polysaccharide from degradation and influence its solubility (Pauly and Scheller 2000). Homogalacturonan is believed to be synthesized in a highly-methylesterified form which is later modified by pectin methylesterases to generate esterified and unesterified regions (Willats et al. 2001), both of which are important in determining the physical properties of pectins and their associations within the wall. In addition, cross-linking of different wall components can occur through polysaccharide-polysaccharide, protein-phenolic, phenolic-polysaccharide and phenolic-phenolic interactions. Cell elongation and expansion therefore requires mechanisms that can disrupt these associations to allow for flexibility. Enzymes such as xyloglucan endotransglycosylase, peroxidases, endoglycanases and esterases and non-enzymic proteins such as expansins are involved in this process and can be influenced by regulators such as auxins and low pH (Cosgrove 2001).

Once the cell has ceased growing, further layers are added to the wall to form the secondary wall thickenings. Secondary walls form three layers inside the primary wall layer (S_1, S_2 and S_3 from outer to inner layers) and are composed of polysaccharides (mostly cellulose) and lignin, although lignin is rarely found in the S_3 layer. Lignin is synthesized by the free radical-driven polymerization of phenylpropanoid monomers including ferulic, coumaric, sinapic, cinnamic and hydoxybenzoic acids.

Biosynthesis of the plant cell wall involves a number of key metabolic processes. Whether or not polysaccharides are strictly defined as metabolites remains to be resolved. However, polysaccharides and oligosaccharides should not be ignored when it comes to plant metabolite profiling since much of plant metabolism revolves around carbohydrate shuffling and incorporation into cell wall and storage polysaccharides. A major challenge facing analysts is the high throughput

measurement of complex carbohydrates due to the structural similarity of a vast and diverse range of possible structures. For example, the same hexose units can join together to form 20 different disaccharide structures and 448 trisaccharide structures, while three different hexoses joined together can potentially give rise to 2688 isomers, but in the case of peptides, three different amino acids will give rise to only six different peptides (Oxley et al. 2004).

Recent developments using molecular genetic approaches have contributed significantly to the current understanding of cell wall biosynthesis, but many aspects of its regulation remain a mystery. The use of metabolomics techniques will help to unravel some of these mysteries particularly when developments in spatial resolution of metabolites come to fruition.

3.6 Secondary metabolites

Plant secondary metabolites, organic compounds that are produced by plants but are not directly involved in their growth and development, include an extremely varied and complex array of molecules. In our everyday lives we use these metabolites, either native or chemically modified, in items such as dyes and food colorings, polymers, fiber, adhesives, oils and waxes, flavoring agents, fragrances and drugs. The function of many secondary metabolites *in planta* is not known, but for others, they serve roles in defense against herbivorous and microbial attack, pollination and seed dispersal and act as allelopathic agents. Plant secondary metabolites are classed into three major groups; the terpenoids, the alkaloids and the phenylpropanoids and related phenolic compounds, all of which are produced by extremely complex biosynthetic pathways.

The terpenoid group of compounds are not restricted to plants but are also produced by animals and microorganisms. Plants however possess a much wider variety of terpenoids than other organisms and by developing highly specialized cells (such as glandular epidermal cells) are able to produce and store them in large quantities. Over 22,000 different compounds belong to the plant group of terpenoids, all of which are derived from the 5-carbon (C_5) precursor isopentenyl diphosphate (IPP) (McGarvey and Croteau 1995). IPP is generated from the fusion of 3 acetyl-CoA molecules and gives rise to mevalonic acid which is in turn phosphorylated and decarboxylated to IPP. Repetitive addition of IPP units gives rise to a series of prenyl diphosphate molecules. These are processed by specific terpenoid synthases to yield terpenoid skeletons that are further modified enzymatically to deliver the vast array of terpenoids that exist in nature. The simplest terpenoid is isoprene which is composed of a single C_5 unit (the terpenoid monomer). Examples of the monoterpenes (2 C_5 units) include the components found in the essential oils of flowers, herbs and spices. Sesquiterpenes (3 C_5 units) can also be found in essential oils and are involved in defense against herbivores and microorganisms. The diterpenes (4 C_5 units) include phytol (component of chlorophyll), the giberellin growth regulators, phytoallexins and taxol (an anticancer agent). Brassinosteroids, some wax components and membrane phytosterols (such as β-sitosterol, campesterol and stigmasterol) belong to the triterpene group, which are

composed of two C_{15} chains linked together. The tetraterpenes (8 C_5 units) include the carotenoids which are essential for photosynthesis, and examples of polyterpenes (more than 8 C_5 units) include rubber, dolichol (essential for sugar transfer reactions), plastiquinone and ubiquinone (electron carriers).

The alkaloids were historically defined as 'pharmacologically active, nitrogen-containing basic compounds of plant origin' based on the therapeutic use of these compounds in traditional medicines. Alkaloids have since been isolated from animal and insect sources (usually toxins) but we continue to use the plant alkaloids in modern medicine. Over 12,000 alkaloid structures from plants have been described, with approximately 20% of plant species known to accumulate alkaloids (DeLuca and Pierre 2000). These include caffeine, camptothecin, cocaine, codeine, morphine, nicotine, quinine and strychnine. The role of alkaloids in plants is generally a defensive one, with evidence for their involvement in wound responses. Not all the biosynthetic pathways for alkaloid biosynthesis have been elucidated, however, it is known that they are mostly derived from amino acids such as tryptophan, tyrosine, phenylalanine, lysine and ornithine, sometimes in combination with steroid or terpenoid-like groups.

The phenylpropanoids and related phenolic compounds (>2500 compounds) are generated through the shikimic acid or malonite/acetate biochemical pathway. Whilst many of the phenolic compounds serve structural roles in the plant cell wall, others have also been ascribed roles in plant defense, flower color and plant flavors and aromas. Lignins are deposited in secondary cell walls to strengthen and reinforce the wall, while suberin acts to protect tissue from dehydration and pathogen attack. The flavanoid group of compounds includes the anthocyanins that impart color in the way of pigments; condensed tannins; and isoflavanoids that serve as defense and signaling molecules.

Metabolomics of secondary metabolites is complicated by their vast numbers and diverse chemistries. Techniques are continuously developing to incorporate as many secondary metabolites in profiling analysis as possible. Recently, von Roepenack-Lahaye et al. (2004) successfully used capillary liquid chromatography coupled to ESI-QqTOF-MS profile to obtain approximately 2000 mass signals from *Arabidopsis* tissue that covered a large number of secondary metabolites but not mono- and sesquiterpenoids, triterpenoid alcohols, phytosterols, waxes or carotenoids. This is typical of all methods currently being used where only a subset of secondary metabolites is detectable mostly due to extraction procedures and low resolution. Furthermore, this group of compounds is so complex that it is possible that previously unidentified structures exist but are being overlooked.

4 Unique aspects of plant research

4.1 Functional genomics

The ever developing area of functional genomics aims to assign function to the multitude of genes that have been identified by genomic analyses of biological

systems. In the "post-genomic" era, the profiling of biological systems at the levels of RNA (transcriptomics), protein (proteomics) and metabolite (metabolomics) is essential to functional genomics. Functional genomics is the ultimate tool for the rational improvement of plants for food, fiber and other commodities that are essential to life as we know it today.

In plants, much of the genetic information gained has been from model systems such as *Arabidopsis thaliana* and commercially important crop plants such as rice, potato and maize. *Arabidopsis* and rice (*Arabidopsis* genome initiative 2000; Yu et al. 2002) have now been fully sequenced. These sequences provide an essential tool for plant functional genomics because of the similarity in gene sequences within the plant kingdom. Based on comparative sequences alone, 54 % of genes in higher plants can be assigned a function (Somerville and Somerville 1999). Useful genetic information has also been gained from other sources where the entire genome is not necessarily available. Expressed sequence tag (EST) libraries have been used to correlate gene expression with developmental processes in plants. For example in potato, ESTs were used to identify genes involved in tuber initiation, dormancy and sprouting (Ronning et al. 2003). Insertion mutant libraries, which are available for *Arabidopsis*, maize, petunia and snapdragon (Somerville and Somerville 1999), and gene silencing by double stranded RNA production, allow for the phenotypic analysis of plants where particular genes have been interrupted or silenced. Developments in gene chip and microarray technologies have also provided essential information by quantitative analysis of gene expression associated with particular treatments or developmental stages (Schena et al. 1997).

Molecular genetic techniques have assisted in identifying entire genomes and transcriptomes because, to a large extent, it is possible to assign gene function based on orthology. However, this does not necessarily help in describing gene function at a cellular level. For example, knowing that a gene codes for a particular enzyme does not provide information on how the enzyme is regulated or what chain of events are triggered by it, nor does it take into account gene duplication. Proteomic technologies are advancing with the development of separation and mass spectrometric platforms for functional genomics. Metabolomic approaches will go one step further in filling the gaps and addressing some of the questions raised by the identification of the vast number of genes discovered from genomic analyses.

4.2 Breeding and QTL analysis

In order to create a novel variety of genotypes and phenotypes, plant genomes can be manipulated in a targeted fashion using breeding. Classical plant breeding deliberately crosses closely or even distantly related species to produce new crops with desired features by introducing genes, and therefore traits, from one species into another genetic background. Basically, plant breeding has been performed since the start of agricultural practices thousands of years ago, but today is approached in a much more sophisticated and organized manner to ensure food secu-

rity and sustain agriculture. Classical breeding relies on the homologous recombination process between two genomes creating novel genetic diversity. Currently, large breeding programs worldwide for many different plant species are aiming for the development of better crops, e.g. with increased yield and quality of the crop, increased tolerance levels to environmental challenges, resistance to viruses, bacteria, fungi or insects, as well as increased tolerance to certain herbicides. In future, metabolomics technologies may become an important tool as a high-throughput method to screen for desired features of a crossed progeny, e.g. for vitamin, acid and/or sugar contents in fruits.

Another great potential tool for identifying novel genetic variety and new genes involved in plant performance is quantitative trait locus mapping. A quantitative trait locus (QTL) is an interval across a chromosome that is associated with a particular feature of the plant, a trait. QTLs are not necessarily genes themselves but are stretches of DNA that are closely linked to the genes controlling the desired trait. The statistical investigation of the alleles which occur in a locus and the produced trait is called QTL mapping. QTL mapping aims to identify the loci, decipher the genes within these loci and ultimately, to identify the functions of the underlying genes. In the past, most QTL analysis was done on single traits, such as yield, plant height or stress tolerance. More recently, with the development of novel technologies for high-throughput simultaneous analysis of transcripts or metabolites, a great potential for multi-trait analysis has become available. Combining for instance the techniques of QTL analysis with those of metabolomics will offer identification of novel QTLs affecting either the level of a single metabolite or even the levels of many metabolites simultaneously. Two primary and exciting examples of this approach have been presented very recently by Schauer et al. (2006) and Keurentjes et al. (2006). Schauer et al. (2006) utilized a GC-MS based metabolite profiling method to analyze the metabolite profiles of fruits from a cultivated tomato species (*Solanum lycopersicon*) in which marker-defined genome regions were introgressed with homologous regions of a wild and non-ripening tomato species (*Solanum penellii*). The authors describe the identification of a large number of single metabolite QTLs as well as many QTLs affecting whole pathways and/or the metabolic network. The work of Keurentjes et al. (2006) demonstrated the investigation of the variation of metabolite composition in plants by analyzing 14 *Arabidopsis thaliana* accessions using a non-targeted LC-QTOF-MS method for the simultaneous detection of more than 2000 individual mass peaks. In addition, the analysis of the metabolomes of a recombinant inbred line (RIL) population of the two most divergent accessions allowed the detection of respective QTLs for about 75 % of all mass peaks. Both examples can be seen as the pioneer work for future QTL identification and mapping as they demonstrate the potential that metabolomics approaches are offering. Once metabolomics technologies become more robust, faster and easier to automate, it will be one of the most promising and informative methods to study genetic segregation and identify novel genes.

Another, sometimes quicker way of introducing new genetic variety into a species is the use of genetic techniques for the production of genetically altered organisms based on transgenesis. These techniques enable the specific introduction

or deletion of targeted genes rather than the random approach used in plant breeding. Several methods are established for doing this but the most common methods include the "gene gun" and the *Agrobacterium* based method. Today, the *Agrobacterium*-mediated genetic transformation is the most commonly used technology for the production of transgenic plants and protocols for a large range of different species have been established. Extensive research in this technology is aiming at improvement of the efficiency of the actual gene transfer. Since the first transgenic plant was created more than 20 years ago, scientists have wanted to analyze the intended as well as unintended effects that the introduced gene has on the plants performance, including the visible phenotype or the abundance of certain cell products. Huge efforts have been made in the development of strategies for risk assessment of genetically modified organisms and metabolomics has been identified as the tool of choice for comprehensive analysis of transgenesis, including effects on plant metabolism as well as on potential interactions with human health and the environment (Rischer and Oksman-Caldentey, 2006).

4.3 Genetic engineering

Plants provide an ideal system for the expression of both foreign and non-foreign genes, either for improved qualities or for the production of selected compounds such as plant secondary metabolites. Worldwide, 90 million hectares of crops are biotech approved across 21 different countries. These include (in order of acreage) USA (49.8 million hectares), Argentina, Brazil, Canada, China, Paraguay, India, South Africa, Uruguay, Australia, Mexico, Romania, the Philippines, Spain, Colombia, Iran, Honduras, Portugal, Germany, France and the Czech Republic (<0.1 million hectares) (James 2005). The industry is worth $5.25 billion, which equates to 18 % of the global commercial seed market. The majority of crops have been modified for pathogen resistance by the introduction of *Bacillus thuringiensis* toxin (BT) (e.g. maize, cotton, canola, rice, potato, tomato) and/or herbicide resistance (e.g. soybean, maize, cotton, canola, rice, sugar beet, tomato). Viral resistance is also available for some plant species (e.g. in squash and papaya). The modified traits that are most common in crops currently focus on farming practices to reduce pesticide use and increase crop yields. The future will see the increase in the introduction of genes into crops to modify nutritional qualities. For example, biotech soybean with high oleic acid content, tomato with increased lycopene levels, and potatoes, maize and wheat with modified starch.

In addition to the modification of crop plants for selected traits, it is possible to use plants as protein or secondary metabolite factories both in the field and in tissue culture. Although potentially expensive, plant tissue culture can offer some advantages over traditional field growing practices. These advantages include the growth of metabolically active cells from rare plants, or plants that are either difficult to cultivate or have long maturation periods. Furthermore, a culture system can be manipulated to occlude the influence of environmental factors such as climate, nutrient availability and disease. Plant-based systems offer a feasible alternative to microbial or mammalian cell culture systems for the production of re-

combinant proteins and can offer some advantages for the production of medically important proteins. Plants don't carry human pathogens or produce endotoxins and they have the necessary machinery for post-translational modification of proteins. Although glycosylation in plants differs only slightly from mammalian glycosylation, it is different enough to cause potential immunogenic and efficacy concerns. These factors are being overcome by the introduction of mammalian glycosyltransferases into plants to produce proteins with mammalian glycosylation patterns (Palacpac et al. 1999).

A number of medically relevant proteins have been produced in plant cells. These include various immunoglobulins and immunoglobulin fragments, human erythropoietin, interleukins and granulocyte macrophage stimulating factor (see Hellwig et al. 2004 and references therein). Furthermore, the effectiveness of recombinant proteins as oral vaccines has been demonstrated. Hepatitis B surface antigen (HBsAg) expressed in potato and fed to mice elicited an immune response to the antigen (Kong et al. 2001). Similarly, in humans, Norwalk virus capsid protein (NVCP) expressed in potato was shown to stimulate an antibody response against the antigen upon its oral administration (Tacket et al. 2000.)

In addition to protein production, plant cell culture provides a controlled method for the production of secondary metabolites for medicinal purposes. Examples of plant cell culture production of secondary metabolites include codeine and morphine by poppy (*Papaver somniferum*), ginsenosides by *Panax ginseng* and capsaicin by *Capsicum frutescens*. Perhaps the most successful example is the production of taxol by plant tissue culture. Taxol is an alkaloid anticancer agent found in the bark of Pacific yew trees (*Taxus brevifolia*). The species does not grow abundantly and taxol is collected only from trees that are over 50 years old. The natural yields of taxol are low (0.001% by dry weight of bark) and it is difficult and expensive to chemically synthesize. Plant tissue culture techniques have enabled reasonable levels (14 mg/L) of taxol to be produced and accumulated in the medium of *Taxus* cultures (Ketchum and Gibson 1996).

The moss, *Physcomitrella patens*, provides a unique system for the establishment of recombinant technologies relevant to higher plants. The lifecycle of *Physcomitrella* is dominated by the haploid gametophytic stage, which means that there are no dominant/recessive traits that can complicate interpretation of genetic screens through the influence of a second allele. Furthermore, *Physcomitrella* has an extremely efficient homologous recombination system making it a far superior system to any other seed plant and twice as efficient as mouse embryonic stem cells for gene targeting (Reski and Frank 2005). Cultures of *Physcomitrella*, which have the advantage of genetic stability compared to tissue cultured cells of higher plants, have been used to produce a humanized antibody for deep-vein thrombosis prevention (Decker and Reski 2004) and human vascular endothelial growth factor (Baur et al. 2005)

The use of plants as "factories" offers an extremely promising approach to the production of recombinant proteins and secondary metabolites. However, a number of obstacles, particularly low yield, must be overcome before this technology can truly advance. Functional genomic approaches, including metabolomics, will certainly allow this to proceed by describing biosynthetic and regulatory pathways

(see Oksman-Caldentey and Inze 2004, and references therein). These approaches will enable rational engineering of biosynthetic pathways to produce metabolites of interest on demand.

5 Recent, current and future of plant metabolomics

5.1 Successful applications

Metabolomics as a technology allows a large number of metabolites of different compound classes to be analyzed and has already been successfully applied to a range of fields in plant sciences. As a tool, metabolomics is applied to answer biological questions that range from the simple to the complex, and to increase our understanding of plant biology and physiology. Metabolomics is also used for the comprehensive phenotyping of genetic varieties or genetically altered plants, for gene identification in functional genomics approaches and to monitor plant behavior and responses to challenging environmental conditions.

The model plant *Arabidopsis thaliana* has been the target of extensive metabolomics studies with different emphases. For example, the metabolome of a large range of mutants has already been analyzed and is the future focus of a large functional genomics program funded by NSF aiming for the identification of the function of all genes in this plant by 2010. In addition, an interesting investigation of the metabolic differences of a range of different accessions of *Arabidopsis* revealed that there exists a large, unexpected diversity between these accessions not only in the amounts of individual metabolites but also in the appearance of certain metabolites (Keutentjes et al. 2006). This study also demonstrated the applicability of metabolomics for high-resolution QTL analysis by untargeted LC-QTOF-MS of the metabolomes of a recombinant inbred line population (RIL) from a cross between two divergent accessions for the identification of QTLs for more than three quarters of the detected mass signals (Keutentjes et al. 2006). As the model plant, the knowledge about *Arabidopsis* genetics and physiology is immense, and large studies have been conducted to investigate cellular responses to a number of different environmentally challenging conditions. For example, the effects on the metabolite profiles have been determined for plants grown in sulphur deficient conditions. Most importantly, these measurements were done in conjunction with transcriptomics analysis to demonstrate the first attempts of integration of both types of datasets (Hirai et al. 2005, Nikiforova et al. 2005). These examples have shown the power of combining metabolomics and transcriptomics analyses for a systems biology approach towards understanding cellular responses and adaptation. Another important stress factor for plants is varying temperature as shown in the detailed characterization of metabolic adaptations to low and high temperatures (Kaplan et al. 2004; Cook et al. 2004). Interestingly, it could be shown that low temperatures have more profound effects than heat, and novel findings of metabolic adaptation to temperature stress were identified (Kaplan et al. 2004). One approach to identify adaptation mechanisms to abiotic stress in plants is to

compare the cellular responses of native, stress-tolerant species and ecotypes. Gong et al. (2005) compared transcript and metabolite abundances between a *Arabidopsis* and a highly salt-tolerant related species, *Thellungiella halophila*, in response to salt stress. Some responses were similar in both species but there were also a range of differences identified in how they responded to the increased salt. These differences will lead to the identification of novel mechanisms that are either constitutively or inductively operating in stress tolerance.

Of great importance, from an agricultural point of view, will be the in-depth analysis of economically important crop plants. A number of interesting metabolomics applications have been demonstrated which have resulted in increased understanding of crop development and physiology and deciphered the impacts of certain external factors on crop quality and quantity. Tomato is one of the major crops under investigation and metabolomics methodologies based on GC-MS have been used to analyze fruit metabolites during development, and following transgenic overexpression of an *Arabidopsis*-derived hexokinase (Roessner-Tunali et al. 2003b). Tikunov et al. (2005) have focused their analysis on volatile compounds produced by the fruit resulting in new insights into fruit metabolism. Another pioneer example by Schauer et al. (2006), as previously mentioned, combined metabolomics with conventional QTL analysis to identify metabolic trait loci.

Cereal grains play an important role in nutrition. They are very carbohydrate-rich but also contain high-value proteins. Increasingly, efforts are being undertaken to understand grain development and quality in order to improve yield and nutritional value. Rice is the major primary food for most nations and because of its importance, has been the target for research for many years. To date, rice is chosen as the model plant for cereal and monocot genetics and physiology which has driven the initiative to sequence its whole genome. The application of metabolomics technologies in rice has started only recently with just a few published examples. Tarpley et al. (2005) have monitored metabolite levels in different tissue sections of developing rice seedlings, allowing the identification of biomarker metabolites being influenced by development, environment or genotype. A combination of different metabolomics techniques based on capillary electrophoresis were used to examine diurnal differences in metabolite concentrations in rice leaves (Sato et al. 2004), demonstrating the dependency of a large range of metabolites on light availability that result in changing patterns throughout the day. In addition, wheat, barley and maize are some of the most important cereal crops and a huge amount of genetic information is available and accumulating on these species from previous and ongoing breeding programs aimed at the development of stronger and higher yielding cultivars. Metabolomics as a tool to investigate metabolite levels of e.g. wheat and barley has only just begun and is mainly used to monitor responses to abiotic stress conditions (Roessner et al. 2006, Roessner, unpublished results). Abiotic stress is the major cause of substantial yield losses because tolerance mechanisms are not very well developed in commercial cultivars. The comparison of metabolite responses of these commercial cultivars with those of landraces exhibiting greater tolerance to certain stresses should lead to the determination of the role of both metabolites and genes in stress tolerance, and

thus provide new strategies for breeding and genetic engineering of novel stress-resistant crops.

Legumes play a critical role in natural agriculture because of their ability to fix nitrogen in symbiotic interactions which makes them economically and environmentally important crop species. Nodule formation occurs in most legume species once a compatible *Rhizobium* bacteria strain is present in the soil. This process has been investigated in detail using a metabolomics approach by Colabatch et al. (2004) and Desbrosses et al. (2005). These reports provided novel insights into nodule formation processes but are also important examples of studying plant-microbe interactions using metabolomics.

6 Future

In recent years it has become apparent that metabolomics will be one of the most important tools in biological sciences. In the near future, many institutions and laboratories worldwide will have established the physical and intellectual capacities to apply metabolomics in their research programs. In plant research, potential applications for metabolomics are enormous and the outcomes overwhelming. Although the technologies employed in metabolomic analyses are uncovering a huge amount of new knowledge in biology, a range of challenges are still to be faced. One bottleneck in metabolomic analysis is the identification of novel compounds. Additionally, in order to allow greatest spatial resolution, the sensitivity and selectivity of currently available technologies has to be increased. Multi-parallel and high-throughput analyses result in large data sets which need be evaluated, extracted and interpreted. To do this we need to work closely together with computer scientists and bioinformaticians to improve and develop bioinformatics methodologies to extract useful and novel information out of the data flow. One step toward this would be the establishment of an open source database for metabolomics data which will attract bioinformaticians and computer scientists to use the huge data sets for large scale statistical analysis, comparative metabolomics and the development of new methodologies for data analysis, mining, visualization and interpretation. This kind of database will also allow us to compare data between labs which will ultimately lead to a better understanding of our own data. An additional major challenge in the metabolomics field is the integration of metabolic data with genomic and proteomic datasets. The ultimate goal is to comprehensively describe complex biological systems and as such, metabolomics will become an important player in systems biology.

References

Arabidospis genome initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815

Baldwin TC, Handford MG, Yuseff M-I, Orellana A, Dupree P. (2001) Identification and characterization of GONST1, a Golgi-localised GDP-mannose transporter in *Arabidopsis*. Plant Cell 13:2283-2295

Baur A, Reski R, Gorr G (2005) Enhanced recovery of a secreted recombinant human growth factor using stabilizing additives and by co-expression of human serum albumin in the moss *Physcomitrella patens*. Plant Biotech J 3:331-340

Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J,

Udvardi MK (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus.* Plant J 39:487-512

Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proc Natl Acad Sci USA 101:15243-15248

Cosgrove DJ (2001) Wall Structure and Wall Loosening. A Look Backwards and Forwards. Plant Physiol 125:131–134

Decker EL, Reski R (2004) The moss bioreactor. Curr Opin Plant Biol 7:166-170

DeLuca V and St Pierre B 2000. The cell and developmental biology of alkaloid biosynthesis. Trends Plant Sci 5: 168-173

Delmer DP, Amor Y (1995) Cellulose biosynthesis. Plant Cell 7:987-100

Desbrosses GG, Kopka J, Udvardi MK. (2005) *Lotus japonicus* metabolic profiling. Development of gas chromatography-mass spectrometry resources for the study of plant-microbe interactions. Plant Physiol. 137:1302-1318

Farré EM, Tiessen A, Roessner U, Geigenberger P, Trethewey RN, Willmitzer L (2001) Analysis of the compartmentation of glycolytic intermediates, nucleotides, sugars, organic acids, amino acids and sugar alcohols in potato tubers using a non-aqueous fractionation method. Plant Physiol. 127:685-700

Fehr M, Ehrhardt DW, Lalonde S, Frommer WB. (2004) Minimally invasive dynamic imaging of ions and metabolites in living cells. Curr Opin Plant Biol. 7:345-351

Galili G, Sengupta-Gopalan C, Ceriotti A (1998) The endoplasmic reticulum of plant cells and its role in protein maturation and biogenesis of oil bodies. Plant Mol Biol 38:1-29

Gibeaut DM, Carpita NC (1994) Biosynthesis of plant cell wall polysaccharides. FASEB J 8:904-915

Gong Q, Li P, Ma S, Indu Rupassara S, Bohnert HJ. (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. Plant J 44:826-839

Hall RD. (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol 169:453-68

Hellwig S, Drossard J, Twyman RM, Fischer R (2004) Plant cell cultures for the production of recombinant proteins. Nature Biotech 22:1415-1422

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K. (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. J Biol Chem 280:25590-25595

James C (2005) Executive Summary of Global Status of Commercialized Biotech/GM Crops: 2005. *ISAAA Briefs* No 34 ISAAA: Ithaca, NY

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol 136:4159-4168

Ketchum, REB, DM Gibson (1996) Paclitaxel production in suspension cell cultures of *Taxus*. Plant Cell Tiss Org Cult 46:9-16

Keurentjes JJ, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LH,

Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nature Genet 38:842-849

Kong Q, Richter L, Yang YF, Arntzen CJ, Mason HS, Thanavala Y (2001) Oral immunization with hepatitis B surface antigen expressed in transgenic plants. Proc Natl Acad Sci USA 98: 11539-11544

Marty F (1999) Plant vacuoles. Plant Cell 11:587-599

McGarvey DJ, Croteau R (1995) Terpenoid metabolism. Plant Cell 7:1015-1026

McLean BG, Hempel FD, Zambryski PC (1997) Plant intercellular communication via plasmodesmata. Plant Cell 9:1043-1054

Moore PJ, Staehelin LA (1988) Immunogold localization of the cell-wall-matrix polysaccharides rhamnogalacturonan I and xyloglucan during cell expansion and cytokinesis in *Trifolium prantese* L.; implication for secretory pathways. Planta 174: 433– 445

Nebenführ A, Staehelin LA (2001) Mobile factories: Golgi dynamics in plant cells. Trends Plant Sci 6:160-167

Nikiforova VJ, Kopka J, Tolstikov V, Fiehn O, Hopkins L, Hawkesford MJ, Hesse H, Hoefgen R. (2005) Systems rebalancing of metabolism in response to sulphur deprivation, as revealed by metabolome analysis of *Arabidopsis* plants. Plant Physiol 138:304-318

Oksman-Caldentey K-M, Inze D (2004) Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci 9:433-440

Orellana A (2005) Biosynthesis of non-cellulosic polysaccharides in the Golgi apparatus. Topological considerations. Plant Biosystems 139:42-45

Oxley D, Currie G, Bacic A (2004) In: Purifying Proteins for Proteomics. A laboratory manual. Pp 579-637. R.J. Simpson (Ed) Cold Spring Harbour Laboratory Press New York

Palacpac NQ, Yoshida S, Sakai H, Kimura Y, Fujiyama K, Yoshida T, Seki T (1999) Stable expression of human β1,4-galactosyltransferase in plant cells modifies N-linked glycosylation patterns. Proc Natl Acad Sci USA 96:4692-4697

Pauly M, Scheller HV (2000) O-Acetylation of plant cell wall polysaccharides: identification and partial characterization of a rhamnogalacturonan O-acetyl-transferase from potato suspension-cultured cells. Planta 210:659-667

Peng L, Kawagoe Y, Hogan P, Delmer D (2002) Sitosterol-β-glucoside as primer for cellulose synthesis in plants. Sci 295:147-150

Prance GT (2001) Discovering the plant world. Taxon 50:345-359

Ramm M, Wolfender J-L, Queiroz EF, Hostettmann K, Hamburger M (2004) Rapid analysis of nucleotide-activated sugars by high-performance liquid chromatography coupled with diode-array detection, electrospray ionization mass spectrometry and nuclear magnetic resonance. J Chromatogr A 1034:139-148

Rayon C, Lerouge P, Faye L (1998) The protein N-glycosylation in plants. J Exp Bot 49:1463–1472

Reiter W-D, Vanzin GF (2001) Molecular genetics of nucleotide sugar interconversion pathways in plants. Plant Mol Biol 47:95-113

Reski R and Frank W (2005) Moss (*Physcomitrella patens*) functional genomics – Gene discovery and tool development, with implications for crop plants and human health. Briefings in Functional Genomics and Proteomics 4:48-57

Rischer H, Oksman-Caldentey KM. (2006) Unintended effects in genetically modified crops: revealed by metabolomics? Trends Biotech 24:102-104

Roessner-Tunali U, Urbanczyk-Wochniak E, Czechowski T, Kolbe A, Willmitzer ., Fernie AR (2003a) *De novo* amino acid biosynthesis in plant storage tissues is regulated by sucrose levels. Plant Physiol 133:683-692

Roessner-Tunali U, Hegemann B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, Fernie AR (2003b) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. Plant Physiol 133:84-99

Roessner U, Patterson J, Forbes MG, Fincher G, Langridge P, Bacic A. (2006) An investigation of boron toxicity in barley using metabolomics. Plant Physiol 142:1087-101

Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD., Fry WE, van der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML, Baker BJ, Buell CR (2003) Comparative analyses of potato expressed sequence tag libraries. Plant Physiol 131: 419-429

Saito K, Dixon RA, Willmitzer L. (Eds) (2006) Plant Metabolomics. Vol. 57 of the series Biotechnology in agriculture and forestry. Eds. Nagata T., Lörz H. and Widholm JM. Publ. Springer-Verlag Berlin-Heidelberg, Germany

Sato S, Soga T, Nishioka T, Tomita M (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. Plant J 40:151-163

Schad M, Mungur R, Fiehn O, Kehr J. 2005. Metabolic profiling of laser microdissected vascular bundles of *Arabidopsis thaliana*. Plant Methods 1:2

Schauer N, Semel Y, Roessner U, Gurb A, Balbo I, Carrari F, Pleban T , Perez-Melisa A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Quantitative genetics of metabolite accumulation in intraspecific introgressions of tomato. Nature Biotechnol 24: 447-454

Scheible W-R, Pauly M (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. Curr Opin Plant Biol 7:285-295

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-470

Somerville C, Somerville S (1999) Plant functional genomics. Science 285:380-383

Staehelin AL, Newcomb EH (2000) Membrane structure and membranous organelles. In Biochemistry and Molecular Biology of Plants, B Buchanan, W Gruissem, R Jones, Eds. American Society of Plant Physiologists, USA

Sturm A. 1995. N-glycosylation of plant proteins. *New Compr. Biochem.* 29a:521–542

Sumner LW, Mendes P, Dixon RA. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochem 62:817-836

Tacket CO, Mason HS, Losonsky G, Estes MK, Levine MM, Arntzen CJ (2000) Human immune responses to a novel norwalk virus vaccine delivered in transgenic potatoes. J Infect Diseases 182:302–305

Tarpley L, Duran AL, Kebrom TH, Sumner LW (2005) Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. BMC Plant Biol 5:8

Tikunov Y, Lommen A, de Vos CH, Verhoeven HA, Bino RJ, Hall RD, Bovy AG. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant Physiol. 139:1125-1137

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J 42:218-235

Urbanczyk-Wochniak E, Baxter C, Kolbe A, Kopka J, Sweetlove LJ, Fernie AR. (2005) Profiling of diurnal patterns of metabolite and transcript abundance in potato (*Solanum tuberosum*) leaves. Planta 221:891-903

Willats WGT, Orfila C, Limberg G, Buchholti HC, van Alebeek G-JWM, Voragen AG J, Marcus SE,. Christensen TMIE, Mikkelsen JD, Murray BS, Knox JP (2001) Modulation of the degree and pattern of methyl-esterification of pectic homogalacturonan in plant cell walls. J Biol Chem 276:19404–19413

Villas-Boas S.G., Roessner U., Hansen M., Smedsgaard J., Nielsen J. (2006) Metabolome Analysis. Publ. Wiley & Sons, New Jersey, NJ, USA (in press)

von Roepenack-Lahaye E, Degenkolb T, Zerjeski M, Franz M, Roth U, Wessjohann L, Schmidt J, Scheel D, Clemens S (2004) Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. Plant Physiol 134:548–559

Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Sci 296:79-92

Pettolino, Filomena
School of Botany, The University of Melbourne, 3010 Victoria, Australia

Roessner, Ute
Australian Centre for Plant Functional Genomics, School of Botany, The University of Melbourne, 3010 Victoria, Australia
ute.roessner@acpfg.com.au